wwPDB X-ray Structure Validation Summary Report (i)

Feb 1, 2016 – 01:32 AM GMT

PDB ID	:	2DG8
Title	:	Crystal structure of the putative trasncriptional regulator SCO7518 from Streptomyces coelicolor A3(2)
Authors	:	Hayashi, T.; Watanabe, N.; Sakai, N.; Tamura, T.; Yao, M.; Tanaka, I.
Deposited on	:	2006-03-08
Resolution	:	2.21 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
http://wwpdb.org/validation/2016/XrayValidationReportHelp
with specific help available everywhere you see the (i) symbol.

The following versions of software and data (see references (1)) were used in the production of this report:

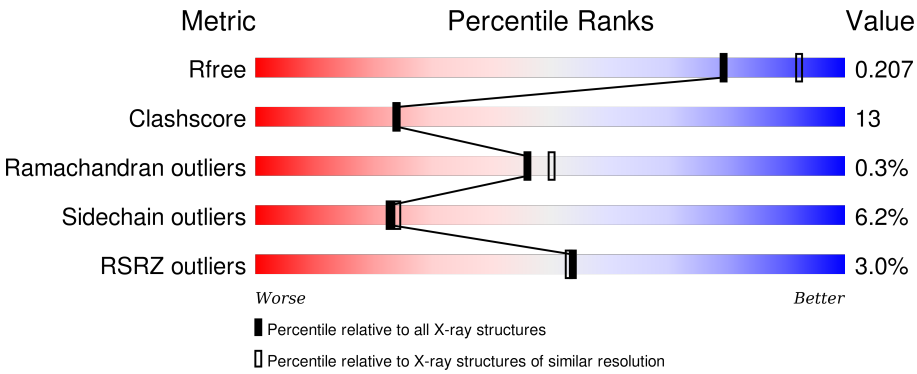
MolProbity	:	4.02b-467
Mogul	:	1.7 (RC4), CSD as536be (2015)
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20026688
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	trunk26865

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 2.21 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}	91344	4405 (2.24-2.20)
Clashscore	102246	5146 (2.24-2.20)
Ramachandran outliers	100387	5065 (2.24-2.20)
Sidechain outliers	100360	5066 (2.24-2.20)
RSRZ outliers	91569	4414 (2.24-2.20)

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 on the lower bar indicate the fraction of residues that contain outliers for >=3, 2, 1 and 0 types of geometric quality criteria. 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	193	2%, 69%, 21%, •, 9%
1	B	193	5%, 57%, 31%, •, 9%
1	C	193	2%, 64%, 25%, •, 10%
1	D	193	2%, 65%, 23%, •, 10%

2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 5739 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 putative tetR-family transcriptional regulatory protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	175	1400	867	270	261	2	0	0	0
1	B	176	1410	874	274	260	2	0	0	0
1	C	174	1396	864	272	258	2	0	0	0
1	D	173	1384	859	268	255	2	0	0	0

There are 32 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	186	LEU	-	EXPRESSION TAG	GB 14495028
A	187	GLU	-	EXPRESSION TAG	GB 14495028
A	188	HIS	-	EXPRESSION TAG	GB 14495028
A	189	HIS	-	EXPRESSION TAG	GB 14495028
A	190	HIS	-	EXPRESSION TAG	GB 14495028
A	191	HIS	-	EXPRESSION TAG	GB 14495028
A	192	HIS	-	EXPRESSION TAG	GB 14495028
A	193	HIS	-	EXPRESSION TAG	GB 14495028
B	186	LEU	-	EXPRESSION TAG	GB 14495028
B	187	GLU	-	EXPRESSION TAG	GB 14495028
B	188	HIS	-	EXPRESSION TAG	GB 14495028
B	189	HIS	-	EXPRESSION TAG	GB 14495028
B	190	HIS	-	EXPRESSION TAG	GB 14495028
B	191	HIS	-	EXPRESSION TAG	GB 14495028
B	192	HIS	-	EXPRESSION TAG	GB 14495028
B	193	HIS	-	EXPRESSION TAG	GB 14495028
C	186	LEU	-	EXPRESSION TAG	GB 14495028
C	187	GLU	-	EXPRESSION TAG	GB 14495028
C	188	HIS	-	EXPRESSION TAG	GB 14495028
C	189	HIS	-	EXPRESSION TAG	GB 14495028
C	190	HIS	-	EXPRESSION TAG	GB 14495028

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
C	191	HIS	-	EXPRESSION TAG	GB 14495028
C	192	HIS	-	EXPRESSION TAG	GB 14495028
C	193	HIS	-	EXPRESSION TAG	GB 14495028
D	186	LEU	-	EXPRESSION TAG	GB 14495028
D	187	GLU	-	EXPRESSION TAG	GB 14495028
D	188	HIS	-	EXPRESSION TAG	GB 14495028
D	189	HIS	-	EXPRESSION TAG	GB 14495028
D	190	HIS	-	EXPRESSION TAG	GB 14495028
D	191	HIS	-	EXPRESSION TAG	GB 14495028
D	192	HIS	-	EXPRESSION TAG	GB 14495028
D	193	HIS	-	EXPRESSION TAG	GB 14495028

- Molecule 2 is water.

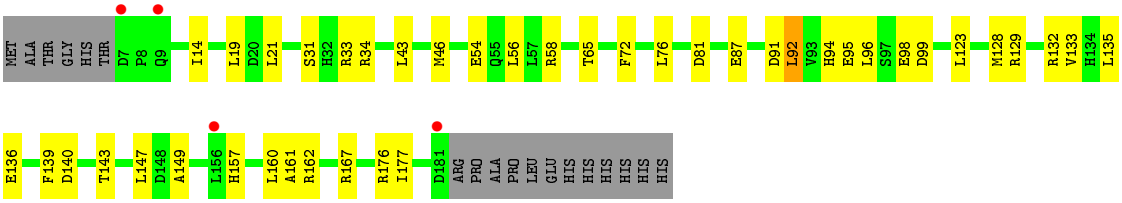
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
2	A	56	56	56	0	0
			Total	O		
2	B	25	25	25	0	0
			Total	O		
2	C	30	30	30	0	0
			Total	O		
2	D	38	38	38	0	0

3 Residue-property plots (i)

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors 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.

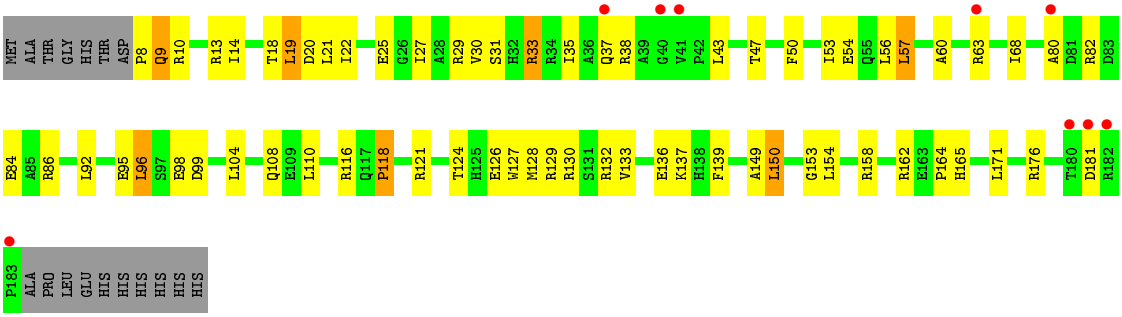
- Molecule 1: putative tetR-family transcriptional regulatory protein

Chain A:



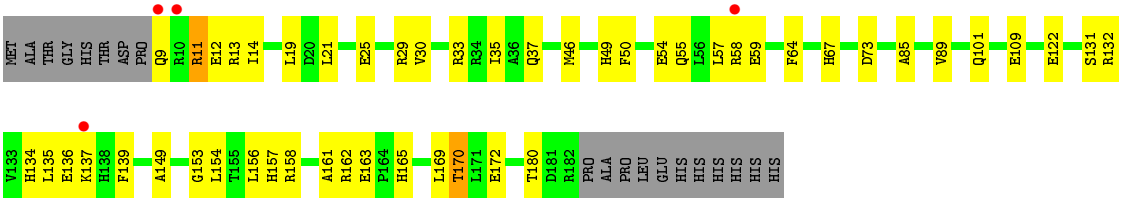
- Molecule 1: putative tetR-family transcriptional regulatory protein

Chain B:



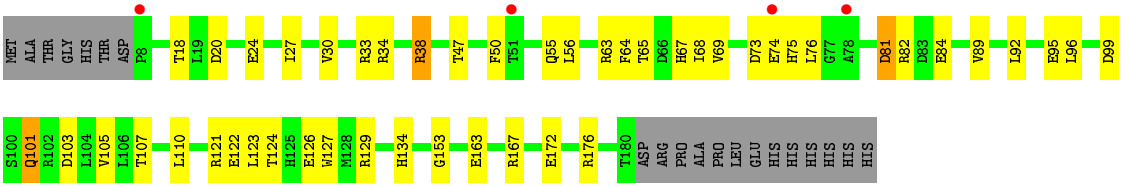
- Molecule 1: putative tetR-family transcriptional regulatory protein

Chain C:



- Molecule 1: putative tetR-family transcriptional regulatory protein

Chain D:


MET
ALA
THR
GLY
HIS
THR
ASP
P8
T18
L19
D20
E24
I27
V30
R33
R34
R38
T47
F50
T51
Q55
L56
R63
F64
T65
D66
H67
I68
V69
D73
E74
H75
L76
G77
A78
D81
R32
D83
E84
V89
L92
E95
L96
D99

S100
Q101
R102
D103
L104
V105
L106
T107
L110
R121
E122
L123
T124
H125
E126
W127
M128
R129
H134
G153
E163
R167
E172
R176
T180
ASP
ARG
PRO
ALA
PRO
LEU
GLU
HIS
HIS
HIS
HIS
HIS
HIS

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	51.34Å 101.26Å 76.91Å 90.00° 98.60° 90.00°	Depositor
Resolution (Å)	19.79 – 2.21 35.85 – 2.21	Depositor EDS
% Data completeness (in resolution range)	97.6 (19.79-2.21) 97.7 (35.85-2.21)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.06	Depositor
$< I/\sigma(I) >$ [1]	21.08 (at 2.22Å)	Xtriage
Refinement program	CNS 1.1	Depositor
R, R_{free}	0.218 , 0.247 0.219 , 0.207	Depositor DCC
R_{free} test set	1916 reflections (5.05%)	DCC
Wilson B-factor (Å^2)	41.8	Xtriage
Anisotropy	0.502	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.34 , 32.6	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning[2]	$< \|L\| > = 0.47$, $< L^2 > = 0.30$	Xtriage
Outliers	0 of 37993 reflections	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	5739	wwPDB-VP
Average B, all atoms (Å^2)	44.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.33% 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, 0.375 respectively for untwinned datasets, and 0.333, 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	0.33	0/1424	0.57	0/1927
1	B	0.29	0/1435	0.57	0/1941
1	C	0.32	0/1419	0.55	0/1918
1	D	0.33	0/1408	0.58	0/1904
All	All	0.32	0/5686	0.57	0/7690

There are no bond length outliers.

There are no bond angle outliers.

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	1400	0	1385	30	0
1	B	1410	0	1402	50	0
1	C	1396	0	1387	36	0
1	D	1384	0	1378	46	0
2	A	56	0	0	1	0
2	B	25	0	0	1	0
2	C	30	0	0	0	0
2	D	38	0	0	2	0
All	All	5739	0	5552	146	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 13.

The worst 5 of 146 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:33:ARG:H	1:B:33:ARG:HD3	1.23	1.02
1:A:143:THR:HG22	1:D:176:ARG:HH12	1.26	1.00
1:B:121:ARG:HH12	1:C:101:GLN:NE2	1.72	0.87
1:B:25:GLU:HB3	1:B:29:ARG:HD3	1.60	0.84
1:C:165:HIS:CE1	1:C:170:THR:HG22	2.13	0.84

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	173/193 (90%)	168 (97%)	5 (3%)	0	100	100
1	B	174/193 (90%)	171 (98%)	2 (1%)	1 (1%)	30	29
1	C	172/193 (89%)	166 (96%)	5 (3%)	1 (1%)	30	29
1	D	171/193 (89%)	168 (98%)	3 (2%)	0	100	100
All	All	690/772 (89%)	673 (98%)	15 (2%)	2 (0%)	46	50

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	C	11	ARG
1	B	118	PRO

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	142/157 (90%)	135 (95%)	7 (5%)	31	35
1	B	143/157 (91%)	129 (90%)	14 (10%)	10	8
1	C	141/157 (90%)	132 (94%)	9 (6%)	22	23
1	D	140/157 (89%)	135 (96%)	5 (4%)	42	51
All	All	566/628 (90%)	531 (94%)	35 (6%)	23	24

5 of 35 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	96	LEU
1	B	171	LEU
1	D	75	HIS
1	B	110	LEU
1	B	137	LYS

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 16 such sidechains are listed below:

Mol	Chain	Res	Type
1	C	9	GLN
1	C	49	HIS
1	C	134	HIS
1	B	146	GLN
1	C	146	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 carbohydrates 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	175/193 (90%)	0.23	4 (2%)	64	63	27, 38, 54, 84	0
1	B	176/193 (91%)	0.52	9 (5%)	32	31	27, 47, 71, 92	0
1	C	174/193 (90%)	0.22	4 (2%)	64	63	30, 42, 59, 80	0
1	D	173/193 (89%)	0.17	4 (2%)	64	63	30, 43, 59, 78	0
All	All	698/772 (90%)	0.29	21 (3%)	54	53	27, 42, 65, 92	0

The worst 5 of 21 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	10	ARG	6.1
1	B	182	ARG	4.9
1	B	183	PRO	4.5
1	C	9	GLN	4.0
1	D	8	PRO	3.8

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 carbohydrates 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.