WORLDWIDE wwPDB PROTEIN DATA BANK

wwPDB X-ray Structure Validation Summary Report (i)

Feb 1, 2016 – 04:42 AM GMT

PDB ID	:	2NXO
Title	:	Crystal structure of protein SCO4506 from Streptomyces coelicolor, Pfam DUF178
Authors	:	Tyagi, R.; Eswaramoorthy, S.; Burley, S.K.; Swaminathan, S.; New York SGX Research Center for Structural Genomics (NYSGXRC)
Deposited on	:	2006-11-17
Resolution	:	2.04 Å(reported)

The following versions of software and data (see references (i)) 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 ⓘ

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

The reported resolution of this entry is 2.04 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.

■ Percentile relative to all X-ray structures

▯ Percentile relative to X-ray structures of similar resolution
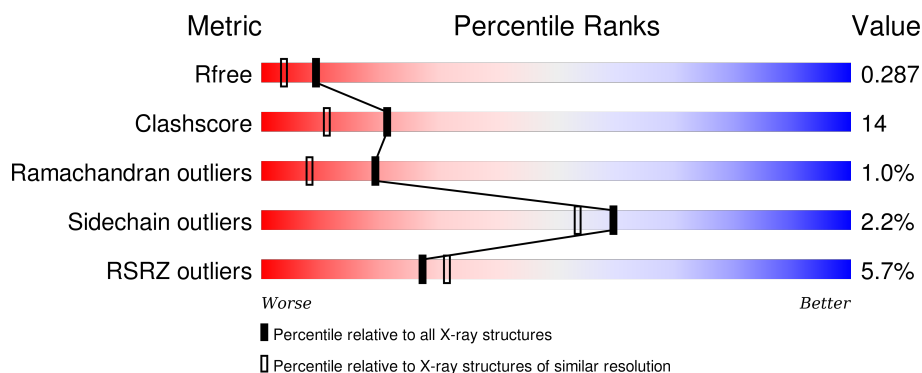
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	91344	1192 (2.04-2.04)
Clashscore	102246	1269 (2.04-2.04)
Ramachandran outliers	100387	1258 (2.04-2.04)
Sidechain outliers	100360	1258 (2.04-2.04)
RSRZ outliers	91569	1194 (2.04-2.04)

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	291	4% 68% 24% · 7%
1	B	291	5% 71% 20% · 9%
1	C	291	2% 72% 18% · 8%
1	D	291	9% 59% 20% · 19%

2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 8700 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 Hypothetical protein SCO4506.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	272	Total	C	N	O	S	0	0	0
			2119	1360	359	393	7			
1	B	266	Total	C	N	O	S	0	0	0
			2082	1333	353	389	7			
1	C	269	Total	C	N	O	S	0	0	0
			2095	1341	356	391	7			
1	D	236	Total	C	N	O	S	0	0	0
			1854	1193	313	341	7			

There are 40 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-1	SER	-	CLONING ARTIFACT	UNP Q9L0T8
A	0	LEU	-	CLONING ARTIFACT	UNP Q9L0T8
A	282	GLU	-	EXPRESSION TAG	UNP Q9L0T8
A	283	GLY	-	EXPRESSION TAG	UNP Q9L0T8
A	284	HIS	-	EXPRESSION TAG	UNP Q9L0T8
A	285	HIS	-	EXPRESSION TAG	UNP Q9L0T8
A	286	HIS	-	EXPRESSION TAG	UNP Q9L0T8
A	287	HIS	-	EXPRESSION TAG	UNP Q9L0T8
A	288	HIS	-	EXPRESSION TAG	UNP Q9L0T8
A	289	HIS	-	EXPRESSION TAG	UNP Q9L0T8
B	-1	SER	-	CLONING ARTIFACT	UNP Q9L0T8
B	0	LEU	-	CLONING ARTIFACT	UNP Q9L0T8
B	282	GLU	-	EXPRESSION TAG	UNP Q9L0T8
B	283	GLY	-	EXPRESSION TAG	UNP Q9L0T8
B	284	HIS	-	EXPRESSION TAG	UNP Q9L0T8
B	285	HIS	-	EXPRESSION TAG	UNP Q9L0T8
B	286	HIS	-	EXPRESSION TAG	UNP Q9L0T8
B	287	HIS	-	EXPRESSION TAG	UNP Q9L0T8
B	288	HIS	-	EXPRESSION TAG	UNP Q9L0T8
B	289	HIS	-	EXPRESSION TAG	UNP Q9L0T8
C	-1	SER	-	CLONING ARTIFACT	UNP Q9L0T8

Continued on next page...

Continued from previous page...

| Chain	Residue	Modelled	Actual	Comment	Reference
C	0	LEU	-	CLONING ARTIFACT	UNP Q9L0T8
C	282	GLU	-	EXPRESSION TAG	UNP Q9L0T8
C	283	GLY	-	EXPRESSION TAG	UNP Q9L0T8
C	284	HIS	-	EXPRESSION TAG	UNP Q9L0T8
C	285	HIS	-	EXPRESSION TAG	UNP Q9L0T8
C	286	HIS	-	EXPRESSION TAG	UNP Q9L0T8
C	287	HIS	-	EXPRESSION TAG	UNP Q9L0T8
C	288	HIS	-	EXPRESSION TAG	UNP Q9L0T8
C	289	HIS	-	EXPRESSION TAG	UNP Q9L0T8
D	-1	SER	-	CLONING ARTIFACT	UNP Q9L0T8
D	0	LEU	-	CLONING ARTIFACT	UNP Q9L0T8
D	282	GLU	-	EXPRESSION TAG	UNP Q9L0T8
D	283	GLY	-	EXPRESSION TAG	UNP Q9L0T8
D	284	HIS	-	EXPRESSION TAG	UNP Q9L0T8
D	285	HIS	-	EXPRESSION TAG	UNP Q9L0T8
D	286	HIS	-	EXPRESSION TAG	UNP Q9L0T8
D	287	HIS	-	EXPRESSION TAG	UNP Q9L0T8
D	288	HIS	-	EXPRESSION TAG	UNP Q9L0T8
D	289	HIS	-	EXPRESSION TAG	UNP Q9L0T8

- Molecule 2 is water.

| Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	151	Total 151 O 151	0	0
2	B	158	Total 158 O 158	0	0
2	C	121	Total 121 O 121	0	0
2	D	120	Total 120 O 120	0	0

3 Residue-property plots ⓘ

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.
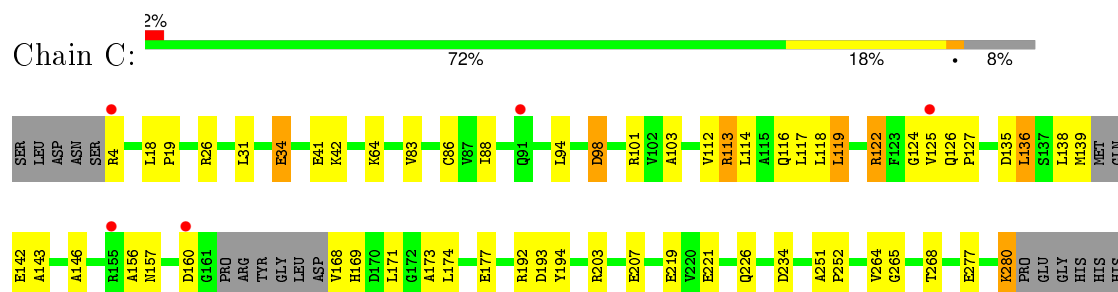
• Molecule 1: Hypothetical protein SCO4506
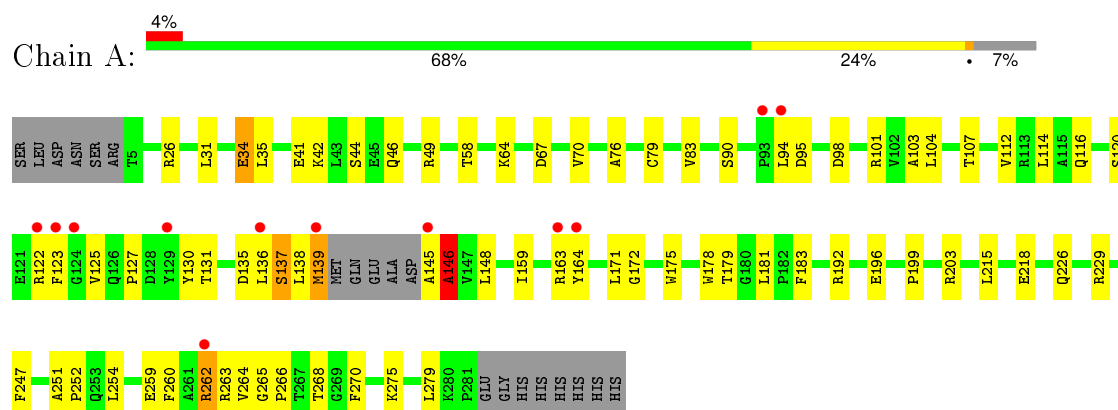
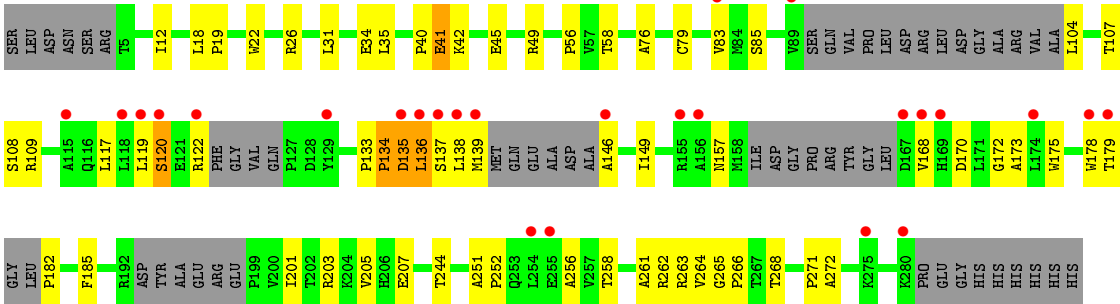• Molecule 1: Hypothetical protein SCO4506
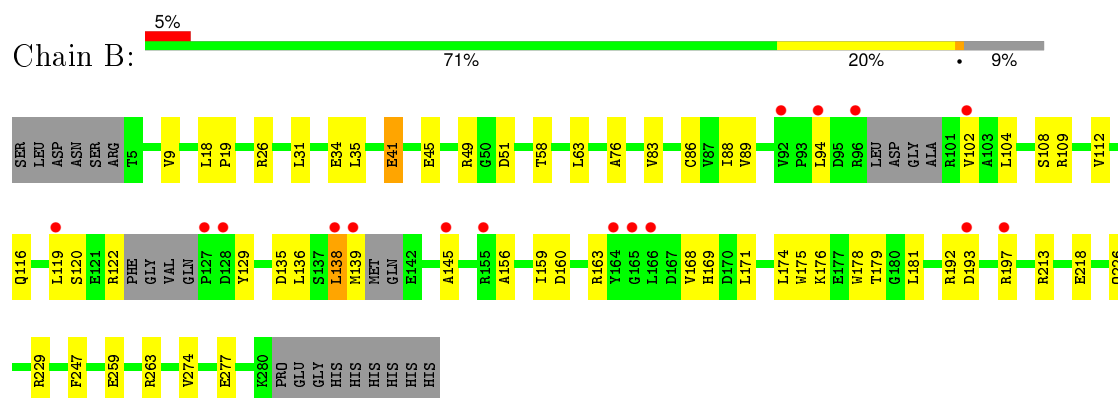
• Molecule 1: Hypothetical protein SCO4506

HIS
HIS
HIS

● **Molecule 1: Hypothetical protein SCO4506**

Chain D:

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	75.79Å 97.37Å 86.57Å 90.00° 106.48° 90.00°	Depositor
Resolution (Å)	30.28 – 2.04 48.69 – 2.04	Depositor EDS
% Data completeness (in resolution range)	96.8 (30.28-2.04) 97.0 (48.69-2.04)	Depositor EDS
R_{merge}	0.04	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	4.65 (at 2.05Å)	Xtriage
Refinement program	CNS 1.1	Depositor
R, R_{free}	0.249 , 0.280 0.250 , 0.287	Depositor DCC
R_{free} test set	2263 reflections (3.05%)	DCC
Wilson B-factor (Å^2)	26.8	Xtriage
Anisotropy	0.208	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.34 , 50.2	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning[2]	$< |L| > = 0.50$, $< L^2 > = 0.33$	Xtriage
Outliers	0 of 76009 reflections	Xtriage
F_o,F_c correlation	0.92	EDS
Total number of atoms	8700	wwPDB-VP
Average B, all atoms (Å^2)	31.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.40% 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	0.36	0/2167	0.66	2/2953 (0.1%)
1	B	0.37	0/2126	0.63	1/2892 (0.0%)
1	C	0.34	0/2139	0.62	1/2911 (0.0%)
1	D	0.33	0/1891	0.60	1/2565 (0.0%)
All	All	0.35	0/8323	0.63	5/11321 (0.0%)

There are no bond length outliers.

All (5) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	D	34	GLU	N-CA-C	-6.42	93.66	111.00
1	B	34	GLU	N-CA-C	-6.05	94.66	111.00
1	A	34	GLU	N-CA-C	-5.65	95.73	111.00
1	C	34	GLU	N-CA-C	-5.59	95.89	111.00
1	A	146	ALA	N-CA-C	5.39	125.56	111.00

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	2119	0	2108	73	0
1	B	2082	0	2069	49	0
1	C	2095	0	2081	47	0
1	D	1854	0	1857	65	0

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	A	151	0	0	11	0
2	B	158	0	0	12	0
2	C	121	0	0	7	0
2	D	120	0	0	19	0
All	All	8700	0	8115	231	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 14.

The worst 5 of 231 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:A:139:MET:SD	1:A:146:ALA:HB2	1.98	1.02
1:B:138:LEU:HD22	2:B:441:HOH:O	1.71	0.90
1:A:98:ASP:HA	1:A:125:VAL:HG13	1.51	0.90
1:A:262:ARG:HH11	1:A:262:ARG:HB3	1.38	0.86
1:A:179:THR:HB	2:A:388:HOH:O	1.78	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	268/291 (92%)	255 (95%)	11 (4%)	2 (1%)	26	15
1	B	258/291 (89%)	247 (96%)	10 (4%)	1 (0%)	39	28
1	C	263/291 (90%)	248 (94%)	11 (4%)	4 (2%)	13	3
1	D	222/291 (76%)	211 (95%)	8 (4%)	3 (1%)	14	4
All	All	1011/1164 (87%)	961 (95%)	40 (4%)	10 (1%)	19	8

5 of 10 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	C	98	ASP
1	C	136	LEU
1	D	120	SER
1	A	137	SER
1	A	146	ALA

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	224/243 (92%)	217 (97%)	7 (3%)	47	39
1	B	221/243 (91%)	217 (98%)	4 (2%)	66	62
1	C	221/243 (91%)	216 (98%)	5 (2%)	58	51
1	D	199/243 (82%)	196 (98%)	3 (2%)	72	70
All	All	865/972 (89%)	846 (98%)	19 (2%)	60	53

5 of 19 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	136	LEU
1	B	139	MET
1	C	280	LYS
1	B	41	GLU
1	D	35	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	169	HIS

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	272/291 (93%)	0.28	12 (4%)	38	43	13, 28, 52, 68	0
1	B	266/291 (91%)	0.16	16 (6%)	25	28	14, 24, 51, 66	0
1	C	269/291 (92%)	0.13	5 (1%)	70	75	17, 29, 49, 63	0
1	D	236/291 (81%)	0.62	26 (11%)	7	8	16, 38, 55, 64	0
All	All	1043/1164 (89%)	0.29	59 (5%)	27	31	13, 29, 52, 68	0

The worst 5 of 59 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	127	PRO	9.7
1	A	139	MET	6.8
1	D	118	LEU	5.8
1	B	96	ARG	5.7
1	D	137	SER	5.3

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.