WORLDWIDE PDB PROTEIN DATA BANK

wwPDB X-ray Structure Validation Summary Report ⓘ

Nov 5, 2023 – 01:55 PM EST

PDB ID	:	4YXA
Title	:	Complex of SpaO(SPOA1,2 SeMet) and OrgB(APAR)::T4lysozyme fusion protein
Authors	:	Notti, R.Q.; Stebbins, C.E.
Deposited on	:	2015-03-22
Resolution	:	2.35 Å(reported)

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

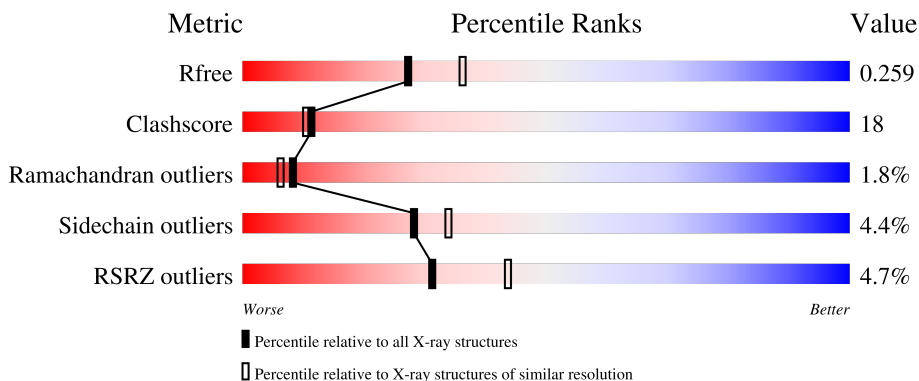|		
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36

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 2.35 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.
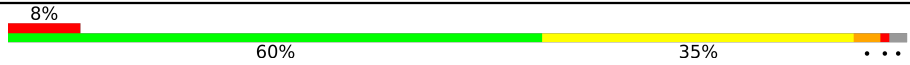
Metric	Percentile Ranks	Value
Rfree		0.259
Clashscore		18
Ramachandran outliers		1.8%
Sidechain outliers		4.4%
RSRZ outliers		4.7%

Worse   *Better*

▌ 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$_{free}$	130704	1164 (2.36-2.36)
Clashscore	141614	1232 (2.36-2.36)
Ramachandran outliers	138981	1211 (2.36-2.36)
Sidechain outliers	138945	1212 (2.36-2.36)
RSRZ outliers	127900	1150 (2.36-2.36)

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	73	5% — 47% / 36% / • / 15%
1	D	73	3% — 63% / 22% / 15%
2	B	70	4% — 47% / 43% / • • •
2	E	70	57% / 29% / 9% / 6%
3	C	197	3% — 52% / 32% / • / 15%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
3	F	197	

2 Entry composition ⓘ

There are 4 unique types of molecules in this entry. The entry contains 4940 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 Surface presentation of antigens protein SpaO.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	62	Total	C	N	O	S	0	0	0
			477	304	87	85	1			
1	D	62	Total	C	N	O	S	0	1	0
			486	311	89	85	1			

There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	GLY	-	expression tag	UNP P40699
A	2	PRO	-	expression tag	UNP P40699
A	3	VAL	-	expression tag	UNP P40699
A	4	ASP	-	expression tag	UNP P40699
D	1	GLY	-	expression tag	UNP P40699
D	2	PRO	-	expression tag	UNP P40699
D	3	VAL	-	expression tag	UNP P40699
D	4	ASP	-	expression tag	UNP P40699

- Molecule 2 is a protein called Surface presentation of antigens protein SpaO.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	67	Total	C	N	O	Se	0	0	0
			522	332	85	102	3			
2	E	66	Total	C	N	O	Se	0	0	0
			515	327	84	101	3			

There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	1	GLY	-	expression tag	UNP P40699
B	2	PRO	-	expression tag	UNP P40699
B	3	VAL	-	expression tag	UNP P40699
B	4	ASP	-	expression tag	UNP P40699

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
E	1	GLY	-	expression tag	UNP P40699
E	2	PRO	-	expression tag	UNP P40699
E	3	VAL	-	expression tag	UNP P40699
E	4	ASP	-	expression tag	UNP P40699

- Molecule 3 is a protein called Oxygen-regulated invasion protein OrgB,Endolysin.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	C	168	Total 1319	C 845	N 236	O 233	S 5	0	0	0
3	F	193	Total 1499	C 954	N 266	O 274	S 5	0	0	0

There are 24 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	1	GLY	-	expression tag	UNP P0CL45
C	2	PRO	-	expression tag	UNP P0CL45
C	3	VAL	-	expression tag	UNP P0CL45
C	4	ASP	-	expression tag	UNP P0CL45
C	45	GLY	ARG	conflict	UNP P00720
C	53	ASN	ASP	engineered mutation	UNP P00720
C	87	THR	CYS	engineered mutation	UNP P00720
C	130	ALA	CYS	engineered mutation	UNP P00720
C	170	ARG	ILE	conflict	UNP P00720
C	195	ALA	LYS	engineered mutation	UNP P00720
C	196	ALA	ASN	engineered mutation	UNP P00720
C	197	ALA	LEU	engineered mutation	UNP P00720
F	1	GLY	-	expression tag	UNP P0CL45
F	2	PRO	-	expression tag	UNP P0CL45
F	3	VAL	-	expression tag	UNP P0CL45
F	4	ASP	-	expression tag	UNP P0CL45
F	45	GLY	ARG	conflict	UNP P00720
F	53	ASN	ASP	engineered mutation	UNP P00720
F	87	THR	CYS	engineered mutation	UNP P00720
F	130	ALA	CYS	engineered mutation	UNP P00720
F	170	ARG	ILE	conflict	UNP P00720
F	195	ALA	LYS	engineered mutation	UNP P00720
F	196	ALA	ASN	engineered mutation	UNP P00720
F	197	ALA	LEU	engineered mutation	UNP P00720

- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	10	Total 10	O 10	0	0
4	B	10	Total 10	O 10	0	0
4	C	47	Total 47	O 47	0	0
4	D	10	Total 10	O 10	0	0
4	E	13	Total 13	O 13	0	0
4	F	32	Total 32	O 32	0	0

3 Residue-property plots ⓘ

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: Surface presentation of antigens protein SpaO
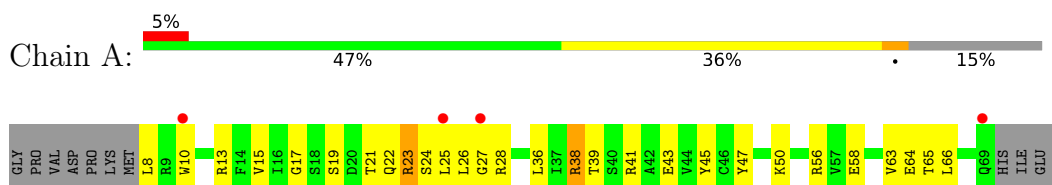
• Molecule 1: Surface presentation of antigens protein SpaO
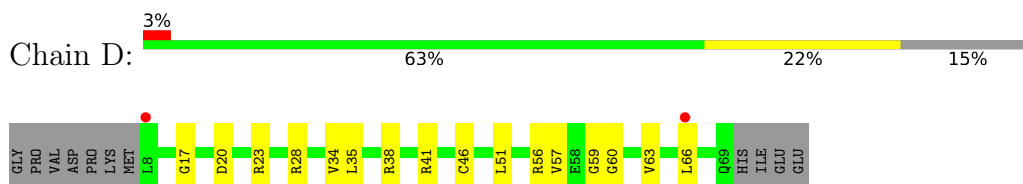
• Molecule 2: Surface presentation of antigens protein SpaO
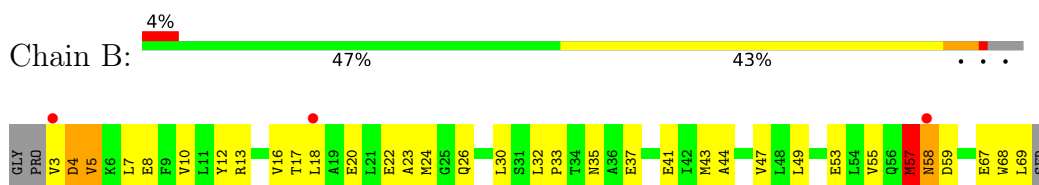
• Molecule 2: Surface presentation of antigens protein SpaO

• Molecule 3: Oxygen-regulated invasion protein OrgB,Endolysin

ASP LYS ALA ILE GLY ARG ASN THR ASN GLY VAL ILE THR LYS ASP GLU A96 E97 K98 L99 F100 N101 Q102 A107 V108 R109 G110 I111 K116 L117 K118 P119 V120 Y121 D122 S123 A126 V127 R128 R129 M135 V136 F137 Q138 M139 G140 V144 T148 R152 Q156 K157 R158

W159 D160 K168 P176 N177 R181 W191 Y194 A195 A196 ALA

• Molecule 3: Oxygen-regulated invasion protein OrgB,Endolysin

Chain F:

G1 P2 V3 D4 M5 L6 K7 N8 P12 S13 P14 L15 S16 P17 K24 R25 K26 E29 R30 Y31 F37 D43 E44 G45 L46 R47 L48 K49 I50 Y51 K52 N53 T54 E55 G56 Y57 Y58 T59 I60 G61 I62 G63 H64 L65 S69 P70 S71 L72 N73 A74 S77 E78 L79

D80 LYS ALA I83 V90 I91 T92 K93 D94 E95 A96 E97 K98 L99 F100 D103 V104 R113 L117 K118 P119 V120 Y121 D122 S123 L124 D125 R128 R129 L132 I133 N134 G140 A145 N149 L154 K157 R158 W159 D160 E161 L166 S169 P176 K180

R181 R187 A193 Y194 A195 ALA ALA

4   Data and refinement statistics ⓘ

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	62.88Å 88.50Å 63.32Å 90.00° 116.07° 90.00°	Depositor
Resolution (Å)	45.80 – 2.35 45.80 – 2.35	Depositor EDS
% Data completeness (in resolution range)	98.8 (45.80-2.35) 96.8 (45.80-2.35)	Depositor EDS
R_{merge}	0.09	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.28 (at 2.34Å)	Xtriage
Refinement program	PHENIX	Depositor
R, R_{free}	0.198 , 0.262 0.200 , 0.259	Depositor DCC
R_{free} test set	1600 reflections (6.22%)	wwPDB-VP
Wilson B-factor (Å^2)	43.4	Xtriage
Anisotropy	0.194	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.31 , 34.8	EDS
L-test for twinning[2]	$< |L| > = 0.44, < L^2 > = 0.26$	Xtriage
Estimated twinning fraction	0.248 for l,-k,h	Xtriage
Reported twinning fraction	0.280 for l,-k,h	Depositor
Outliers	1 of 25740 reflections (0.004%)	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	4940	wwPDB-VP
Average B, all atoms (Å^2)	46.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.22% 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 ⓘ

5.1 Standard geometry ⓘ

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	0.51	0/484	0.72	0/653
1	D	0.44	0/493	0.69	0/666
2	B	0.53	0/524	0.73	0/706
2	E	0.48	0/517	0.80	2/696 (0.3%)
3	C	0.48	0/1343	0.67	1/1817 (0.1%)
3	F	0.51	0/1525	0.72	1/2065 (0.0%)
All	All	0.50	0/4886	0.71	4/6603 (0.1%)

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand. A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
2	B	0	1
3	F	0	3
All	All	0	4

There are no bond length outliers.

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	E	34	THR	C-N-CA	6.76	138.61	121.70
3	F	15	LEU	CA-CB-CG	5.80	128.63	115.30
3	C	58	TYR	CA-CB-CG	5.66	124.15	113.40
2	E	34	THR	CA-C-N	5.02	128.25	117.20

There are no chirality outliers.

All (4) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
2	B	57	MSE	Peptide

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Group
3	F	13	SER	Peptide
3	F	74	ALA	Peptide
3	F	93	LYS	Peptide

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	477	0	480	30	1
1	D	486	0	493	18	1
2	B	522	0	529	35	0
2	E	515	0	520	21	0
3	C	1319	0	1348	47	1
3	F	1499	0	1511	51	1
4	A	10	0	0	4	0
4	B	10	0	0	3	0
4	C	47	0	0	5	0
4	D	10	0	0	2	0
4	E	13	0	0	0	0
4	F	32	0	0	5	0
All	All	4940	0	4881	173	2

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 173 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)
3:F:43:ASP:OD1	3:F:181:ARG:NH1	2.03	0.92
3:F:158:ARG:HD2	3:F:161:GLU:OE1	1.79	0.81
3:F:187:ARG:NH1	4:F:201:HOH:O	2.05	0.81
1:A:8:LEU:N	4:A:101:HOH:O	2.14	0.79
3:C:116:LYS:HE2	3:C:148:THR:HG22	1.63	0.79

All (2) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
3:C:56:GLY:O	1:D:38:ARG:NH1[2_455]	2.18	0.02
1:A:39:THR:OG1	3:F:55:GLU:OE1[2_554]	2.19	0.01

5.3 Torsion angles ⓘ

5.3.1 Protein backbone ⓘ

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	60/73 (82%)	48 (80%)	10 (17%)	2 (3%)	4	2
1	D	61/73 (84%)	55 (90%)	6 (10%)	0	100	100
2	B	65/70 (93%)	55 (85%)	8 (12%)	2 (3%)	4	2
2	E	64/70 (91%)	60 (94%)	2 (3%)	2 (3%)	4	2
3	C	164/197 (83%)	150 (92%)	13 (8%)	1 (1%)	25	27
3	F	189/197 (96%)	171 (90%)	14 (7%)	4 (2%)	7	4
All	All	603/680 (89%)	539 (89%)	53 (9%)	11 (2%)	8	6

5 of 11 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	23	ARG
2	B	58	ASN
3	F	14	PRO
3	F	94	ASP
2	E	35	ASN

5.3.2 Protein sidechains ⓘ

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	49/63 (78%)	48 (98%)	1 (2%)	55	66
1	D	50/63 (79%)	50 (100%)	0	100	100
2	B	58/57 (102%)	56 (97%)	2 (3%)	37	46
2	E	57/57 (100%)	53 (93%)	4 (7%)	15	15
3	C	136/165 (82%)	131 (96%)	5 (4%)	34	42
3	F	155/165 (94%)	145 (94%)	10 (6%)	17	18
All	All	505/570 (89%)	483 (96%)	22 (4%)	28	34

5 of 22 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	F	49	LYS
3	F	69	SER
3	F	65	LEU
3	F	90	VAL
3	C	123	SER

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (4) such sidechains are listed below:

Mol	Chain	Res	Type
2	B	26	GLN
2	B	58	ASN
3	F	149	ASN
3	F	165	ASN

5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

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

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

5.5 Carbohydrates ⓘ

There are no monosaccharides in this entry.

5.6 Ligand geometry ⓘ

There are no ligands in this entry.

5.7 Other polymers ⓘ

There are no such residues in this entry.

5.8 Polymer linkage issues ⓘ

There are no chain breaks in this entry.

6 Fit of model and data ⓘ

6.1 Protein, DNA and RNA chains ⓘ

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	62/73 (84%)	0.46	4 (6%)	18	27	29, 52, 66, 73	0
1	D	62/73 (84%)	0.27	2 (3%)	47	59	31, 47, 65, 75	0
2	B	64/70 (91%)	0.33	3 (4%)	31	44	27, 43, 69, 77	0
2	E	63/70 (90%)	0.09	0	100	100	31, 42, 62, 69	0
3	C	168/197 (85%)	0.20	5 (2%)	50	61	23, 37, 69, 84	0
3	F	193/197 (97%)	0.41	15 (7%)	13	19	26, 49, 76, 87	0
All	All	612/680 (90%)	0.30	29 (4%)	31	44	23, 45, 73, 87	0

The worst 5 of 29 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
3	C	58	TYR	8.5
2	B	18	LEU	6.9
3	F	56	GLY	5.4
1	A	69	GLN	4.7
1	D	66	LEU	4.6

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

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

6.3 Carbohydrates ⓘ

There are no monosaccharides in this entry.

6.4 Ligands ⓘ

There are no ligands in this entry.

6.5 Other polymers ⓘ

There are no such residues in this entry.