W O R L D W I D E
wwPDB
P R O T E I N   D A T A   B A N K

wwPDB X-ray Structure Validation Summary Report (i)

Feb 1, 2016 – 02:49 PM GMT

PDB ID	:	4AKZ
Title	:	CRYSTAL STRUCTURE OF VIRB8 FROM BRUCELLA SUIS
Authors	:	Coincon, M.; Smith, M.A.; Sygusch, J.; Baron, C.
Deposited on	:	2012-02-29
Resolution	:	2.25 Å(reported)

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 2.25 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.
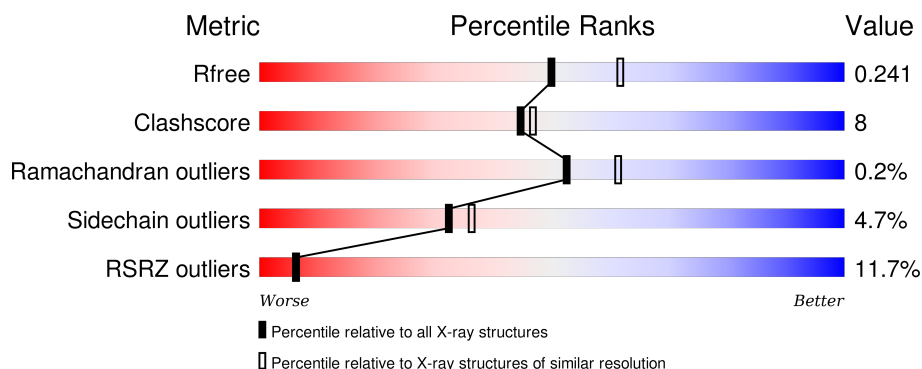
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	91344	1640 (2.28-2.24)
Clashscore	102246	1095 (2.26-2.26)
Ramachandran outliers	100387	1063 (2.26-2.26)
Sidechain outliers	100360	1063 (2.26-2.26)
RSRZ outliers	91569	1647 (2.28-2.24)

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	138	9% / 81% 15% •
1	B	138	8% / 87% 9% • •
1	C	138	12% / 76% 18% • •
1	D	138	9% / 75% 19% • •
1	E	138	19% / 78% 17% • •

2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 6291 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 TYPE IV SECRETION SYSTEM PROTEIN VIRB8.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	138	Total	C	N	O	S	0	0	0
			1094	688	181	222	3			
1	B	134	Total	C	N	O	S	0	0	0
			1066	673	176	214	3			
1	C	134	Total	C	N	O	S	0	0	0
			1066	673	176	214	3			
1	D	133	Total	C	N	O	S	0	0	0
			1058	669	175	211	3			
1	E	134	Total	C	N	O	S	0	0	0
			1066	673	176	214	3			

- Molecule 2 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	218	Total	O	0	0
			218	218		
2	B	204	Total	O	0	0
			204	204		
2	C	152	Total	O	0	0
			152	152		
2	D	180	Total	O	0	0
			180	180		
2	E	187	Total	O	0	0
			187	187		

3 Residue-property plots ⓘ

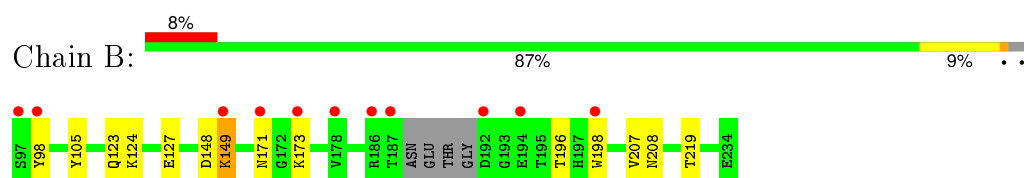These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are color-coded according 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.
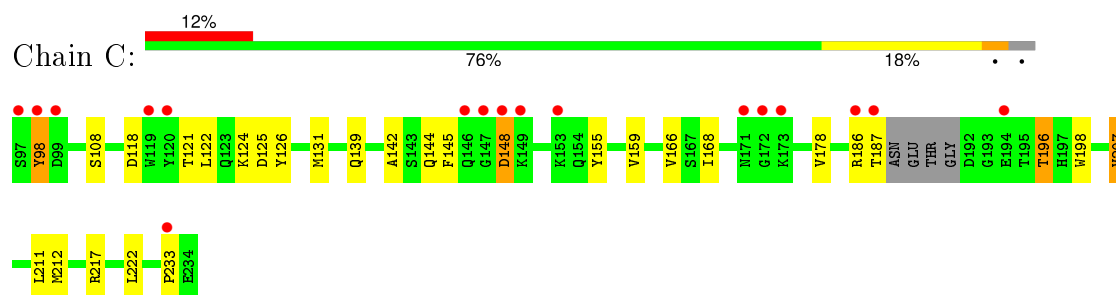
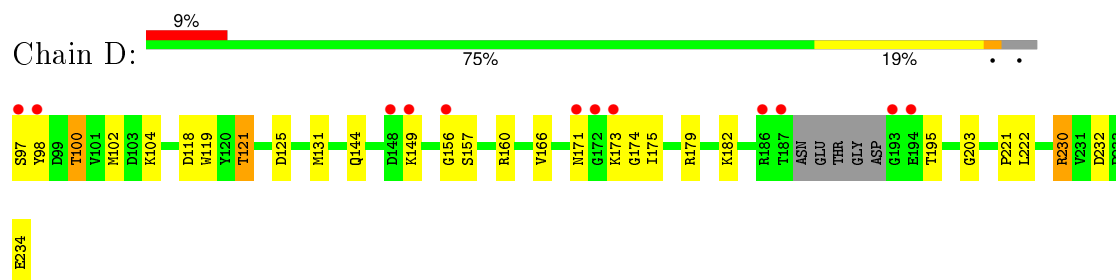- Molecule 1: TYPE IV SECRETION SYSTEM PROTEIN VIRB8

Chain A:

9%
81%
15%

S97 Y98 T121 D125 M131 S136 Q139 S140 Y141 Q144 D148 K153 N171 G172 K173 R179 K182 R186 T187 N188 E189 T190 G191 D192 G193 E194 T195 T196 H197 W198 Y206 P209 M212 T219 G223 R230 E234

- Molecule 1: TYPE IV SECRETION SYSTEM PROTEIN VIRB8

Chain B:

8%
87%
9%

S97 Y98 Y105 Q123 K124 E127 D148 K149 N171 G172 K173 V178 R186 T187 ASN GLU THR GLY D192 G193 E194 T195 T196 H197 W198 V207 N208 T219 E234

- Molecule 1: TYPE IV SECRETION SYSTEM PROTEIN VIRB8

Chain C:

12%
76%
18%

S97 Y98 D99 S108 D118 W119 Y120 T121 L122 Q123 K124 D125 Y126 M131 Q139 A142 S143 Q144 F145 Q146 G147 D148 K149 K153 Q154 Y155 V159 V166 S167 I168 N171 G172 K173 V178 R186 T187 ASN GLU THR GLY D192 G193 E194 T195 T196 H197 W198 V207

L211 M212 R217 L222 P233 E234

- Molecule 1: TYPE IV SECRETION SYSTEM PROTEIN VIRB8

Chain D:

9%
75%
19%

S97 Y98 D99 T100 V101 M102 D103 K104 D118 W119 Y120 T121 D125 M131 Q144 D148 K149 G156 S157 R160 V166 N171 G172 K173 G174 I175 R179 K182 R186 T187 ASN GLU THR GLY ASP G193 E194 T195 G203 P221 L222 R230 V231 D232 P233

E234

- Molecule 1: TYPE IV SECRETION SYSTEM PROTEIN VIRB8

Chain E:

19%
78%
17%

4 Data and refinement statistics (i)

Property	Value	Source
Space group	I 41 2 2	Depositor
Cell constants a, b, c, α, β, γ	198.44Å 198.44Å 103.22Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	19.94 – 2.25 19.84 – 2.24	Depositor EDS
% Data completeness (in resolution range)	97.6 (19.94-2.25) 97.7 (19.84-2.24)	Depositor EDS
R_{merge}	0.04	Depositor
R_{sym}	(Not available)	Depositor
$<I/\sigma(I)>$ [1]	6.22 (at 2.23Å)	Xtriage
Refinement program	PHENIX (PHENIX.REFINE)	Depositor
R, R_{free}	0.190 , 0.217 0.223 , 0.241	Depositor DCC
R_{free} test set	3592 reflections (8.10%)	DCC
Wilson B-factor (Å^2)	37.6	Xtriage
Anisotropy	0.305	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.37 , 77.3	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning[2]	$<|L|> = 0.48$, $<L^2> = 0.30$	Xtriage
Outliers	0 of 48076 reflections	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	6291	wwPDB-VP
Average B, all atoms (Å^2)	50.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.25% 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	0.73	1/1119 (0.1%)	0.63	0/1522
1	B	0.61	0/1090	0.54	0/1481
1	C	0.60	0/1090	0.50	0/1481
1	D	0.52	0/1082	0.55	0/1470
1	E	0.49	0/1090	0.49	0/1481
All	All	0.60	1/5471 (0.0%)	0.54	0/7435

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	206	TYR	CD2-CE2	-5.95	1.30	1.39

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	1094	0	1046	24	0
1	B	1066	0	1023	15	0
1	C	1066	0	1023	16	0
1	D	1058	0	1019	21	1
1	E	1066	0	1023	15	0
2	A	218	0	0	4	1
2	B	204	0	0	1	2

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	C	152	0	0	1	0
2	D	180	0	0	0	0
2	E	187	0	0	2	0
All	All	6291	0	5134	83	3

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 8.

The worst 5 of 83 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:149:LYS:HD3	1:B:149:LYS:H	1.06	1.15
1:B:149:LYS:CD	1:B:149:LYS:H	1.71	1.04
1:B:149:LYS:CD	1:B:149:LYS:N	2.39	0.84
1:B:149:LYS:HD3	1:B:149:LYS:N	1.91	0.82
1:D:230:ARG:HD3	1:D:232:ASP:OD2	1.82	0.80

All (3) 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)
2:B:2077:HOH:O	2:B:2077:HOH:O[5_554]	0.46	1.74
1:D:131:MET:CG	2:A:2189:HOH:O[5_555]	2.09	0.11
2:B:2105:HOH:O	2:B:2105:HOH:O[5_554]	2.17	0.03

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	136/138 (99%)	135 (99%)	0	1 (1%)	26	26
1	B	130/138 (94%)	130 (100%)	0	0	100	100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	C	130/138 (94%)	127 (98%)	3 (2%)	0	100	100
1	D	129/138 (94%)	127 (98%)	2 (2%)	0	100	100
1	E	130/138 (94%)	126 (97%)	4 (3%)	0	100	100
All	All	655/690 (95%)	645 (98%)	9 (1%)	1 (0%)	52	61

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	189	GLU

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	121/121 (100%)	114 (94%)	7 (6%)	25	25
1	B	118/121 (98%)	115 (98%)	3 (2%)	55	66
1	C	118/121 (98%)	110 (93%)	8 (7%)	20	18
1	D	117/121 (97%)	111 (95%)	6 (5%)	29	32
1	E	118/121 (98%)	114 (97%)	4 (3%)	44	54
All	All	592/605 (98%)	564 (95%)	28 (5%)	32	36

5 of 28 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	131	MET
1	C	148	ASP
1	E	196	THR
1	C	139	GLN
1	C	144	GLN

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

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 carbohydrates 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 (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	138/138 (100%)	0.51	13 (9%)	11	12	20, 32, 79, 111	0
1	B	134/138 (97%)	0.61	11 (8%)	14	15	23, 39, 78, 95	0
1	C	134/138 (97%)	0.73	17 (12%)	5	5	36, 54, 87, 110	0
1	D	133/138 (96%)	0.58	12 (9%)	12	13	22, 36, 70, 91	0
1	E	134/138 (97%)	0.95	26 (19%)	1	1	37, 52, 87, 106	0
All	All	673/690 (97%)	0.68	79 (11%)	6	6	20, 45, 84, 111	0

The worst 5 of 79 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	E	187	THR	8.4
1	C	97	SER	7.3
1	A	190	THR	7.1
1	A	188	ASN	6.9
1	E	192	ASP	6.2

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.