WORLDWIDE PDB PROTEIN DATA BANK

wwPDB X-ray Structure Validation Summary Report (i)

Aug 20, 2023 – 03:44 PM EDT

|		
PDB ID	:	2IOU
Title	:	Major Tropism Determinant P1 (Mtd-P1) Variant Complexed with Bordetella brochiseptica Virulence Factor Pertactin extracellular domain (Prn-E).
Authors	:	Miller, J.L.; Ghosh, P.
Deposited on	:	2006-10-10
Resolution	:	3.16 Å(reported)

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 3.16 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.
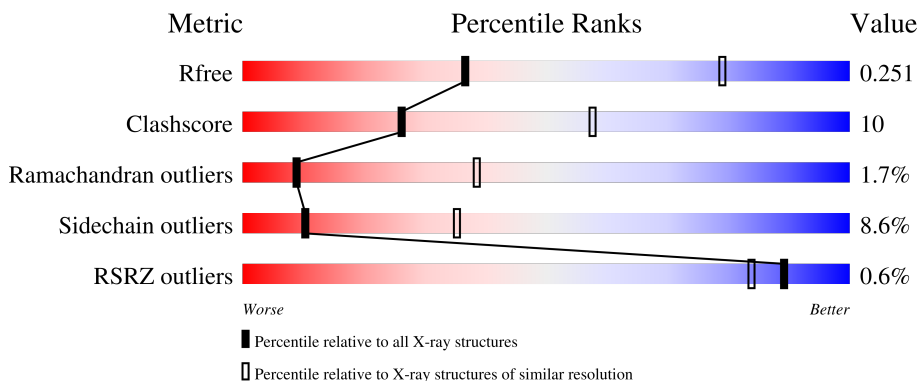
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	1665 (3.20-3.12)
Clashscore	141614	1804 (3.20-3.12)
Ramachandran outliers	138981	1770 (3.20-3.12)
Sidechain outliers	138945	1769 (3.20-3.12)
RSRZ outliers	127900	1616 (3.20-3.12)

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	376	80% 19% ••
1	B	376	84% 14% ••
1	C	376	84% 15% •
1	D	376	85% 15% •
1	E	376	83% 16% •

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	F	376	78% 20%
2	G	535	63% 24% 6% 5%
2	H	535	63% 24% 5% 5%

2   Entry composition ⓘ

There are 3 unique types of molecules in this entry. The entry contains 23900 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 Major Tropism Determinant P1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	376	Total	C	N	O	S	0	0	0
			2758	1736	477	539	6			
1	B	376	Total	C	N	O	S	0	0	0
			2758	1736	477	539	6			
1	C	376	Total	C	N	O	S	0	0	0
			2758	1736	477	539	6			
1	D	376	Total	C	N	O	S	0	0	0
			2758	1736	477	539	6			
1	E	376	Total	C	N	O	S	0	0	0
			2758	1736	477	539	6			
1	F	376	Total	C	N	O	S	0	0	0
			2758	1736	477	539	6			

- Molecule 2 is a protein called Pertactin Extracellular Domain.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	G	508	Total	C	N	O	S	54	0	0
			3674	2265	686	717	6			
2	H	508	Total	C	N	O	S	54	0	0
			3674	2265	686	717	6			

- Molecule 3 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	1	Total	Mg	0	0
			1	1		
3	B	1	Total	Mg	0	0
			1	1		
3	D	2	Total	Mg	0	0
			2	2		

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: Major Tropism Determinant P1

Chain A:

• Molecule 1: Major Tropism Determinant P1

Chain B:

• Molecule 1: Major Tropism Determinant P1

Chain C:

• Molecule 1: Major Tropism Determinant P1

Chain D:

• Molecule 1: Major Tropism Determinant P1

Chain E:

• Molecule 1: Major Tropism Determinant P1

Chain F:

• Molecule 2: Pertactin Extracellular Domain
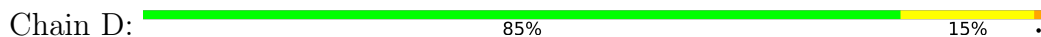
Chain G:

E465 R468 G481 L482 A489 D490 L497 H507 S516 E517 P518 A521 N522 V527 Q528 T529 P530 R531 T536 L539 K542 D543 G544 K545 V546 D547 I548 Y551 R552 Y553 R554 L555 A556 A557 N558 W563 G567 A568 K569 P572

• Molecule 2: Pertactin Extracellular Domain

Chain H:

2%

63%   24%   5% • 5%

D38 I44 I45 K46 R50 Q51 Q58 S59 D60 G63 V64 R65 T70 T71 R77 Q78 A79 Q80 L83 L84 E85 Q94 S97 V98 T99 S100 S101 L104 F105 G108 V109 R110 R111 F112 L113 G114 T115 V116 T117 D126 L130 V133 S134 D135 T136 R137

D138 D139 I142 A143 L144 Y145 V146 A147 I155 S158 T159 L160 E169 R170 T175 V176 Q177 R178 S179 T180 I181 T190 L191 Q192 P193 L194 Q195 P196 E197 D198 L199 V205 L206 G207 D208 T209 S210 V211 T212 A213 V214 P215 G218 A219 P220 T239 G240 G241 R242 A243 A244

M249 A252 T260 G264 ASP ALA PRO ALA GLY GLY ALA VAL PRO GLY GLY ALA VAL PRO GLY GLY PHE GLY PRO LEU LEU ASP G292 W293 Y294 V298 S299 D300 S301 Q314 L315 G322 A325 R326 S330 P337 T344 G345 G346 G347

A348 R349 A355 S356 P357 L358 L373 L374 Y375 R376 V377 L378 P379 V382 Q391 G392 Q393 G394 D395 I396 V397 A398 T399 E400 L401 P402 P405 G406 A407 S408 S409 V414 A415 L416 A417 S418 Q419 A420 R421 W422 T423 G424 A425 T426 R427 A428 V429 D430 S431 L432 S433 I434 D435

T438 W439 V440 M441 N446 V447 G448 S454 D455 G481 L497 H507 N513 S516 E517 S520 N522 T529 P530 R531 G532 S533 T536 K542 D543 G544 K545 T550 Y551 R552 N558 G559 N560 V566 G567 A568 K569 A570 P571 P572

4   Data and refinement statistics ⓘ

Property	Value	Source
Space group	P 61	Depositor
Cell constants a, b, c, α, β, γ	413.46Å 413.46Å 98.86Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	49.63 – 3.16 49.60 – 3.16	Depositor EDS
% Data completeness (in resolution range)	99.9 (49.63-3.16) 99.9 (49.60-3.16)	Depositor EDS
R_{merge}	0.16	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.35 (at 3.19Å)	Xtriage
Refinement program	REFMAC	Depositor
R, R_{free}	0.224 , 0.249 0.227 , 0.251	Depositor DCC
R_{free} test set	8362 reflections (5.03%)	wwPDB-VP
Wilson B-factor (Å^2)	72.6	Xtriage
Anisotropy	0.063	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.34 , 29.8	EDS
L-test for twinning[2]	$< |L| > = 0.51, < L^2 > = 0.35$	Xtriage
Estimated twinning fraction	0.011 for h,-h-k,-l	Xtriage
F_o,F_c correlation	0.92	EDS
Total number of atoms	23900	wwPDB-VP
Average B, all atoms (Å^2)	69.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.34% 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 ⓘ

Bond lengths and bond angles in the following residue types are not validated in this section: MG

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	0.45	2/2828 (0.1%)	0.59	1/3856 (0.0%)
1	B	1.67	4/2828 (0.1%)	0.79	7/3856 (0.2%)
1	C	0.44	0/2828	0.55	0/3856
1	D	0.44	0/2828	0.57	2/3856 (0.1%)
1	E	0.43	0/2828	0.54	0/3856
1	F	0.87	5/2828 (0.2%)	0.83	8/3856 (0.2%)
2	G	1.57	12/3731 (0.3%)	1.31	26/5073 (0.5%)
2	H	3.57	23/3731 (0.6%)	1.64	48/5073 (0.9%)
All	All	1.68	46/24430 (0.2%)	0.98	92/33282 (0.3%)

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
1	B	0	2
1	F	1	0
2	G	0	9
2	H	0	7
All	All	1	18

The worst 5 of 46 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	H	408	SER	CB-OG	153.18	3.41	1.42
2	H	299	SER	CB-OG	96.94	2.68	1.42
1	B	349	SER	CB-OG	-67.15	0.55	1.42
2	G	377	VAL	CB-CG2	64.35	2.88	1.52
2	H	379	PRO	N-CD	63.04	2.36	1.47

The worst 5 of 92 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	H	300	ASP	CB-CG-OD1	-39.30	82.93	118.30
2	H	300	ASP	CB-CG-OD2	38.97	153.38	118.30
2	G	377	VAL	CG1-CB-CG2	-35.82	53.59	110.90
2	H	378	LEU	C-N-CD	-33.40	47.13	120.60
2	G	377	VAL	CA-CB-CG2	-29.63	66.46	110.90

All (1) chirality outliers are listed below:

Mol	Chain	Res	Type	Atom
1	F	349	SER	CA

5 of 18 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	B	346	ASN	Mainchain
1	B	349	SER	Mainchain
2	G	111	ARG	Mainchain,Peptide
2	G	197	GLU	Mainchain,Peptide
2	G	198	ASP	Sidechain

5.2 Too-close contacts ⓘ

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	2758	0	2615	51	0
1	B	2758	0	2614	52	0
1	C	2758	0	2615	40	0
1	D	2758	0	2615	37	0
1	E	2758	0	2615	45	0
1	F	2758	0	2614	66	0
2	G	3674	0	3666	120	0
2	H	3674	0	3665	128	0
3	A	1	0	0	0	0
3	B	1	0	0	0	0
3	D	2	0	0	0	0
All	All	23900	0	23019	487	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 10.

The worst 5 of 487 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)
2:H:401:LEU:CG	2:H:401:LEU:CB	1.84	1.51
2:G:399:THR:CG2	2:G:399:THR:CB	1.99	1.39
1:F:348:THR:OG1	1:F:348:THR:CB	1.74	1.35
2:H:379:PRO:CB	2:H:379:PRO:CG	2.12	1.28
1:F:348:THR:CB	1:F:348:THR:CG2	2.12	1.27

There are no symmetry-related clashes.

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	374/376 (100%)	350 (94%)	22 (6%)	2 (0%)	29	65
1	B	374/376 (100%)	350 (94%)	21 (6%)	3 (1%)	19	55
1	C	374/376 (100%)	352 (94%)	20 (5%)	2 (0%)	29	65
1	D	374/376 (100%)	348 (93%)	24 (6%)	2 (0%)	29	65
1	E	374/376 (100%)	346 (92%)	28 (8%)	0	100	100
1	F	374/376 (100%)	348 (93%)	23 (6%)	3 (1%)	19	55
2	G	504/535 (94%)	439 (87%)	42 (8%)	23 (5%)	2	15
2	H	504/535 (94%)	444 (88%)	41 (8%)	19 (4%)	3	19
All	All	3252/3326 (98%)	2977 (92%)	221 (7%)	54 (2%)	9	38

5 of 54 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	G	112	PHE

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
2	G	199	LEU
2	G	300	ASP
2	G	355	ALA
2	G	378	LEU

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	271/271 (100%)	254 (94%)	17 (6%)	18	49
1	B	271/271 (100%)	261 (96%)	10 (4%)	34	66
1	C	271/271 (100%)	260 (96%)	11 (4%)	30	63
1	D	271/271 (100%)	256 (94%)	15 (6%)	21	54
1	E	271/271 (100%)	252 (93%)	19 (7%)	15	45
1	F	271/271 (100%)	254 (94%)	17 (6%)	18	49
2	G	375/388 (97%)	316 (84%)	59 (16%)	2	11
2	H	375/388 (97%)	318 (85%)	57 (15%)	3	12
All	All	2376/2402 (99%)	2171 (91%)	205 (9%)	10	35

5 of 205 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	G	239	THR
2	G	531	ARG
2	H	533	SER
2	G	299	SER
2	G	395	ASP

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 47 such sidechains are listed below:

Mol	Chain	Res	Type
1	F	317	ASN

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
2	G	369	GLN
1	F	377	HIS
2	G	187	HIS
2	G	507	HIS

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 ⓘ

Of 4 ligands modelled in this entry, 4 are monoatomic - leaving 0 for Mogul analysis.

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

5.7 Other polymers ⓘ

There are no such residues in this entry.

5.8 Polymer linkage issues ⓘ

The following chains have linkage breaks:

Mol	Chain	Number of breaks
2	H	4
2	G	3
1	F	2
1	A	1
1	B	1

The worst 5 of 11 chain breaks are listed below:

Model	Chain	Residue-1	Atom-1	Residue-2	Atom-2	Distance (Å)
1	A	348:THR	C	349:SER	N	1.19
1	B	349:SER	C	350:LEU	N	1.19
1	F	350:LEU	C	351:SER	N	1.19
1	H	399:THR	C	400:GLU	N	1.15
1	G	111:ARG	C	112:PHE	N	1.08

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	376/376 (100%)	-0.22	0	100	100	47, 61, 75, 83	0
1	B	376/376 (100%)	-0.13	1 (0%)	94	92	46, 65, 75, 80	0
1	C	376/376 (100%)	-0.22	0	100	100	44, 58, 71, 80	0
1	D	376/376 (100%)	-0.13	1 (0%)	94	92	52, 66, 75, 80	0
1	E	376/376 (100%)	-0.00	1 (0%)	94	92	58, 75, 89, 97	0
1	F	376/376 (100%)	-0.08	0	100	100	51, 69, 83, 91	0
2	G	502/535 (93%)	0.06	8 (1%)	72	59	57, 71, 92, 109	0
2	H	502/535 (93%)	0.17	10 (1%)	65	50	63, 81, 105, 120	0
All	All	3260/3326 (98%)	-0.05	21 (0%)	89	84	44, 68, 92, 120	0

The worst 5 of 21 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	G	558	ASN	3.7
2	H	193	PRO	3.7
2	H	550	THR	2.9
1	E	284	THR	2.7
2	G	198	ASP	2.6

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.

WORLDWIDE
wwPDB
PROTEIN DATA BANK

6.4 Ligands (i)

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95th percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å^2)	Q<0.9
3	MG	A	3	1/1	0.91	0.55	51,51,51,51	0
3	MG	D	4	1/1	0.96	0.59	57,57,57,57	0
3	MG	D	2	1/1	0.97	0.22	49,49,49,49	0
3	MG	B	1	1/1	0.97	0.23	50,50,50,50	0

6.5 Other polymers (i)

There are no such residues in this entry.