



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 1, 2016 – 06:33 AM GMT

PDB ID : 2XI9  
Title : PILUS-PRESENTED ADHESIN, SPY0125 (CPA), P1 FORM  
Authors : Pointon, J.A.; Smith, W.D.; Saalbach, G.; Crow, A.; Kehoe, M.A.; Banfield, M.J.  
Deposited on : 2010-06-28  
Resolution : 1.90 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](mailto:validation@mail.wwpdb.org)  
A user guide is available at  
<http://wwpdb.org/validation/2016/XrayValidationReportHelp>  
with specific help available everywhere you see the ⓘ symbol.

---

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

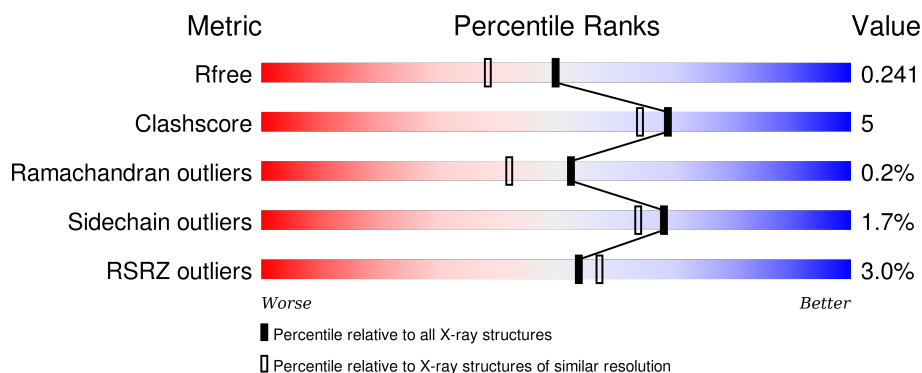
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

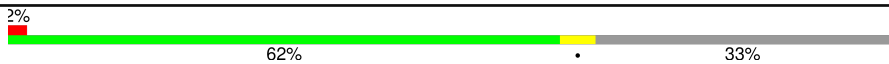
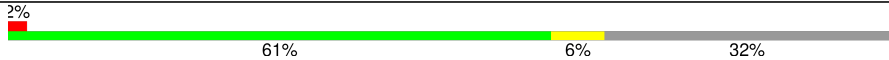
The reported resolution of this entry is 1.90 Å.

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	4755 (1.90-1.90)
Clashscore	102246	5398 (1.90-1.90)
Ramachandran outliers	100387	5338 (1.90-1.90)
Sidechain outliers	100360	5339 (1.90-1.90)
RSRZ outliers	91569	4766 (1.90-1.90)

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  $\geq 3$ , 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  $\leq 5\%$ . 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	457	
1	B	457	

## 2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 5423 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 ANCILLARY PROTEIN 1.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	304	Total	C	N	O	S	Se	0	0	0
			2426	1547	387	488	1	3			
1	B	310	Total	C	N	O	S	Se	0	0	0
			2470	1571	397	498	1	3			

There are 38 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	267	MSE	-	EXPRESSION TAG	UNP Q8GRA2
A	268	GLY	-	EXPRESSION TAG	UNP Q8GRA2
A	269	SER	-	EXPRESSION TAG	UNP Q8GRA2
A	270	SER	-	EXPRESSION TAG	UNP Q8GRA2
A	271	HIS	-	EXPRESSION TAG	UNP Q8GRA2
A	272	HIS	-	EXPRESSION TAG	UNP Q8GRA2
A	273	HIS	-	EXPRESSION TAG	UNP Q8GRA2
A	274	HIS	-	EXPRESSION TAG	UNP Q8GRA2
A	275	HIS	-	EXPRESSION TAG	UNP Q8GRA2
A	276	HIS	-	EXPRESSION TAG	UNP Q8GRA2
A	277	SER	-	EXPRESSION TAG	UNP Q8GRA2
A	278	SER	-	EXPRESSION TAG	UNP Q8GRA2
A	279	GLY	-	EXPRESSION TAG	UNP Q8GRA2
A	280	LEU	-	EXPRESSION TAG	UNP Q8GRA2
A	281	VAL	-	EXPRESSION TAG	UNP Q8GRA2
A	282	PRO	-	EXPRESSION TAG	UNP Q8GRA2
A	283	ARG	-	EXPRESSION TAG	UNP Q8GRA2
A	284	GLY	-	EXPRESSION TAG	UNP Q8GRA2
A	285	SER	-	EXPRESSION TAG	UNP Q8GRA2
B	267	MSE	-	EXPRESSION TAG	UNP Q8GRA2
B	268	GLY	-	EXPRESSION TAG	UNP Q8GRA2
B	269	SER	-	EXPRESSION TAG	UNP Q8GRA2
B	270	SER	-	EXPRESSION TAG	UNP Q8GRA2
B	271	HIS	-	EXPRESSION TAG	UNP Q8GRA2
B	272	HIS	-	EXPRESSION TAG	UNP Q8GRA2

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
B	273	HIS	-	EXPRESSION TAG	UNP Q8GRA2
B	274	HIS	-	EXPRESSION TAG	UNP Q8GRA2
B	275	HIS	-	EXPRESSION TAG	UNP Q8GRA2
B	276	HIS	-	EXPRESSION TAG	UNP Q8GRA2
B	277	SER	-	EXPRESSION TAG	UNP Q8GRA2
B	278	SER	-	EXPRESSION TAG	UNP Q8GRA2
B	279	GLY	-	EXPRESSION TAG	UNP Q8GRA2
B	280	LEU	-	EXPRESSION TAG	UNP Q8GRA2
B	281	VAL	-	EXPRESSION TAG	UNP Q8GRA2
B	282	PRO	-	EXPRESSION TAG	UNP Q8GRA2
B	283	ARG	-	EXPRESSION TAG	UNP Q8GRA2
B	284	GLY	-	EXPRESSION TAG	UNP Q8GRA2
B	285	SER	-	EXPRESSION TAG	UNP Q8GRA2

- Molecule 2 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	262	Total O 262 262	0	0
2	B	265	Total O 265 265	0	0



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	35.48 Å 66.46 Å 101.80 Å 86.06° 90.06° 75.08°	Depositor
Resolution (Å)	34.11 – 1.90 34.11 – 1.90	Depositor EDS
% Data completeness (in resolution range)	93.6 (34.11-1.90) 78.2 (34.11-1.90)	Depositor EDS
$R_{merge}$	0.07	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	6.62 (at 1.89 Å)	Xtriage
Refinement program	REFMAC 5.5.0102	Depositor
R, $R_{free}$	0.206 , 0.232 0.211 , 0.241	Depositor DCC
$R_{free}$ test set	2946 reflections (5.38%)	DCC
Wilson B-factor (Å <sup>2</sup> )	18.5	Xtriage
Anisotropy	0.388	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.29 , 23.3	EDS
Estimated twinning fraction	0.428 for h,h-k,-l	Xtriage
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.51$ , $\langle L^2 \rangle = 0.34$	Xtriage
Outliers	0 of 57731 reflections	Xtriage
$F_o, F_c$ correlation	0.93	EDS
Total number of atoms	5423	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	26.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 8.45% of the height of the origin peak. No significant pseudotranslation is detected.*

<sup>1</sup>Intensities estimated from amplitudes.

<sup>2</sup>Theoretical values of  $\langle |L| \rangle$ ,  $\langle L^2 \rangle$  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	0.70	0/2474	0.72	1/3347 (0.0%)
1	B	0.70	0/2519	0.72	2/3408 (0.1%)
All	All	0.70	0/4993	0.72	3/6755 (0.0%)

There are no bond length outliers.

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	555	ASP	CB-CG-OD2	-5.99	112.91	118.30
1	B	555	ASP	CB-CG-OD1	5.52	123.27	118.30
1	A	471	ARG	NE-CZ-NH1	5.10	122.85	120.30

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	2426	0	2363	20	0
1	B	2470	0	2396	24	0
2	A	262	0	0	4	0
2	B	265	0	0	3	1
All	All	5423	0	4759	44	1

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 5.

The worst 5 of 44 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:444:MSE:CE	1:B:449:THR:HG22	1.77	1.14
1:B:444:MSE:HE2	1:B:449:THR:CG2	1.80	1.10
1:A:566:PHE:HE1	1:A:568:ILE:HD11	1.28	0.97
1:A:566:PHE:CE1	1:A:568:ILE:HD11	2.00	0.96
1:B:336:ILE:HG22	1:B:338:LEU:HD13	1.61	0.83

All (1) 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:2169:HOH:O	2:B:2200:HOH:O[1_655]	2.07	0.13

## 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	300/457 (66%)	290 (97%)	9 (3%)	1 (0%)	46	35
1	B	306/457 (67%)	298 (97%)	8 (3%)	0	100	100
All	All	606/914 (66%)	588 (97%)	17 (3%)	1 (0%)	52	42

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	599	VAL

### 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	268/399 (67%)	263 (98%)	5 (2%)	65	59
1	B	272/399 (68%)	268 (98%)	4 (2%)	72	69
All	All	540/798 (68%)	531 (98%)	9 (2%)	68	64

5 of 9 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	586	ASP
1	B	586	ASP
1	B	397	PHE
1	A	373	ILE
1	B	338	LEU

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

### 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 ⓘ

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, 95<sup>th</sup> 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 <sup>2</sup> )	Q<0.9
1	A	301/457 (65%)	0.08	11 (3%) 45 49	7, 21, 57, 116	0
1	B	307/457 (67%)	0.04	7 (2%) 64 67	7, 21, 60, 136	0
All	All	608/914 (66%)	0.06	18 (2%) 54 57	7, 21, 59, 136	0

The worst 5 of 18 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	604	HIS	4.6
1	A	602	VAL	4.2
1	A	600	ILE	3.9
1	B	323	GLN	3.5
1	B	603	THR	3.4

### 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.