



# Full wwPDB X-ray Structure Validation Report ⓘ

Aug 24, 2017 – 01:18 PM EDT

PDB ID : 5VXJ  
Title : 2.50 Å resolution structure of IpaD from Shigella flexneri in complex with single-domain antibody JMK-E3  
Authors : Barta, M.L.; Lovell, S.; Battaile, K.P.; Picking, W.D.; Picking, W.L.  
Deposited on : unknown  
Resolution : 2.50 Å(reported)

This is a Full wwPDB X-ray Structure Validation 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
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20029824
Percentile statistics	:	20161228.v01 (using entries in the PDB archive December 28th 2016)
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)	:	rb-20029824

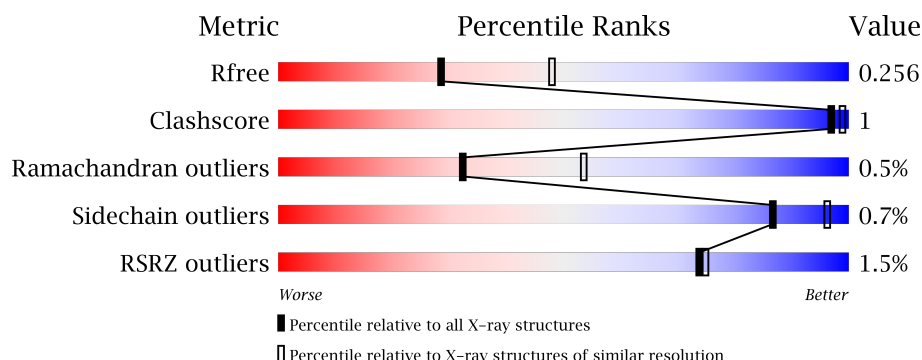
# 1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 2.50 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))
$R_{free}$	100719	3846 (2.50-2.50)
Clashscore	112137	4554 (2.50-2.50)
Ramachandran outliers	110173	4463 (2.50-2.50)
Sidechain outliers	110143	4465 (2.50-2.50)
RSRZ outliers	101464	3876 (2.50-2.50)

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	289	<div> <div>94%</div> <div> <div></div> <div>3%</div> <div>0%</div> <div>0%</div> </div> <div>...</div> </div>
1	C	289	<div> <div>93%</div> <div> <div>3%</div> <div>0%</div> <div>0%</div> <div>0%</div> </div> <div>6%</div> </div>
1	E	289	<div> <div>93%</div> <div> <div>0%</div> <div>0%</div> <div>0%</div> <div>0%</div> </div> <div>...</div> </div>
1	G	289	<div> <div>88%</div> <div> <div>3%</div> <div>0%</div> <div>0%</div> <div>0%</div> </div> <div>7% 5%</div> </div>
1	I	289	<div> <div>87%</div> <div> <div>2%</div> <div>0%</div> <div>0%</div> <div>0%</div> </div> <div>10%</div> </div>

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
2	B	146	 80% • 18%
2	D	146	 79% • 19%
2	F	146	 79% • 19%
2	H	146	 79% • 18%
2	J	146	 77% • 20%

## 2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 29721 atoms, of which 14591 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 Invasin IpaD.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	280	Total	C	H	N	O	S	0	0	0
			4313	1353	2144	371	439	6			
1	C	273	Total	C	H	N	O	S	0	0	0
			4250	1332	2122	361	428	7			
1	E	278	Total	C	H	N	O	S	0	0	0
			4244	1335	2102	366	435	6			
1	G	274	Total	C	H	N	O	S	0	0	0
			4109	1300	2025	358	419	7			
1	I	259	Total	C	H	N	O	S	0	0	0
			3918	1238	1938	337	399	6			

There are 25 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	34	GLY	-	expression tag	UNP P18013
A	35	SER	-	expression tag	UNP P18013
A	36	THR	-	expression tag	UNP P18013
A	37	GLY	-	expression tag	UNP P18013
A	322	SER	CYS	engineered mutation	UNP P18013
C	34	GLY	-	expression tag	UNP P18013
C	35	SER	-	expression tag	UNP P18013
C	36	THR	-	expression tag	UNP P18013
C	37	GLY	-	expression tag	UNP P18013
C	322	SER	CYS	engineered mutation	UNP P18013
E	34	GLY	-	expression tag	UNP P18013
E	35	SER	-	expression tag	UNP P18013
E	36	THR	-	expression tag	UNP P18013
E	37	GLY	-	expression tag	UNP P18013
E	322	SER	CYS	engineered mutation	UNP P18013
G	34	GLY	-	expression tag	UNP P18013
G	35	SER	-	expression tag	UNP P18013
G	36	THR	-	expression tag	UNP P18013
G	37	GLY	-	expression tag	UNP P18013

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
G	322	SER	CYS	engineered mutation	UNP P18013
I	34	GLY	-	expression tag	UNP P18013
I	35	SER	-	expression tag	UNP P18013
I	36	THR	-	expression tag	UNP P18013
I	37	GLY	-	expression tag	UNP P18013
I	322	SER	CYS	engineered mutation	UNP P18013

- Molecule 2 is a protein called single-domain antibody JMK-E3.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
2	B	119	Total	C	H	N	O	S	0	0	0
			1740	551	851	161	172	5			
2	D	118	Total	C	H	N	O	S	0	0	0
			1741	550	852	163	171	5			
2	F	118	Total	C	H	N	O	S	0	0	0
			1756	553	863	164	171	5			
2	H	119	Total	C	H	N	O	S	0	0	0
			1752	553	857	164	173	5			
2	J	117	Total	C	H	N	O	S	0	0	0
			1712	543	837	161	166	5			

- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	36	Total	O	0	0
			36	36		
3	B	14	Total	O	0	0
			14	14		
3	C	28	Total	O	0	0
			28	28		
3	D	7	Total	O	0	0
			7	7		
3	E	16	Total	O	0	0
			16	16		
3	F	10	Total	O	0	0
			10	10		
3	G	36	Total	O	0	0
			36	36		
3	H	11	Total	O	0	0
			11	11		
3	I	17	Total	O	0	0
			17	17		

*Continued on next page...*

*Continued from previous page...*

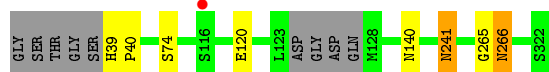
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	J	11	Total	O	0	0
			11	11		

### 3 Residue-property plots [i](#)

These plots are drawn for all protein, RNA and DNA 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: Invasin IpaD

Chain A:  94% ...

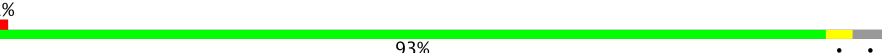


#### • Molecule 1: Invasin IpaD

Chain C:  93% ... 6%




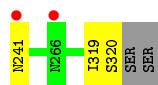
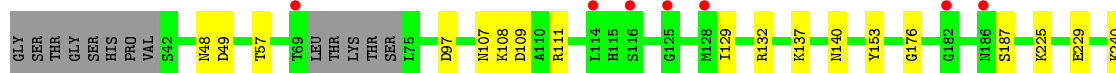
#### • Molecule 1: Invasin IpaD

Chain E:  93% ...




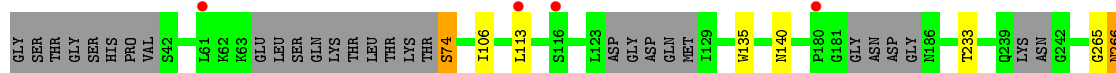
#### • Molecule 1: Invasin IpaD

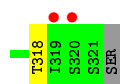
Chain G:  88% 7% 5%



#### • Molecule 1: Invasin IpaD

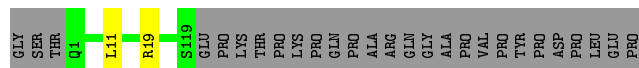
Chain I:  87% ... 10%





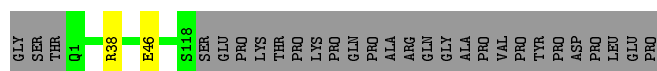
- Molecule 2: single-domain antibody JMK-E3

Chain B: 80% 18%



- Molecule 2: single-domain antibody JMK-E3

Chain D: 79% 19%



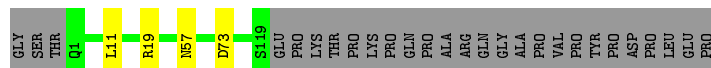
- Molecule 2: single-domain antibody JMK-E3

Chain F: 79% 19%



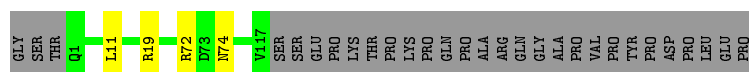
- Molecule 2: single-domain antibody JMK-E3

Chain H: 79% 18%



- Molecule 2: single-domain antibody JMK-E3

Chain J: 77% 20%





## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	79.01Å 163.35Å 216.02Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	45.05 – 2.50 48.62 – 2.50	Depositor EDS
% Data completeness (in resolution range)	98.4 (45.05-2.50) 98.6 (48.62-2.50)	Depositor EDS
$R_{merge}$	0.14	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.14 (at 2.51Å)	Xtriage
Refinement program	PHENIX 1.12rc1_2803	Depositor
R, $R_{free}$	0.196 , 0.259 0.190 , 0.256	Depositor DCC
$R_{free}$ test set	4813 reflections (5.01%)	DCC
Wilson B-factor (Å <sup>2</sup> )	39.5	Xtriage
Anisotropy	0.446	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.38 , 45.4	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.50$ , $\langle L^2 \rangle = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.94	EDS
Total number of atoms	29721	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	54.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.81% 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, 0.333 respectively for untwinned datasets, and 0.375, 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	0.24	0/2201	0.38	0/2980
1	C	0.24	0/2158	0.39	0/2918
1	E	0.24	0/2172	0.39	0/2943
1	G	0.24	0/2114	0.39	0/2864
1	I	0.23	0/2005	0.38	0/2713
2	B	0.26	0/906	0.48	0/1226
2	D	0.25	0/906	0.48	0/1226
2	F	0.26	0/910	0.48	0/1230
2	H	0.26	0/912	0.49	0/1234
2	J	0.26	0/892	0.48	0/1208
All	All	0.24	0/15176	0.42	0/20542

There are no bond length outliers.

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	2169	2144	2144	7	0
1	C	2128	2122	2122	3	0
1	E	2142	2102	2102	4	0
1	G	2084	2025	2025	11	0
1	I	1980	1938	1938	5	0
2	B	889	851	853	1	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	D	889	852	854	1	0
2	F	893	863	865	2	0
2	H	895	857	859	4	0
2	J	875	837	839	2	0
3	A	36	0	0	0	0
3	B	14	0	0	0	0
3	C	28	0	0	0	0
3	D	7	0	0	0	0
3	E	16	0	0	1	0
3	F	10	0	0	0	0
3	G	36	0	0	1	0
3	H	11	0	0	0	0
3	I	17	0	0	1	0
3	J	11	0	0	0	0
All	All	15130	14591	14601	33	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 1.

All (33) 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:G:107:ASN:O	1:G:109:ASP:N	2.10	0.83
1:E:140:ASN:OD1	2:F:19:ARG:NH1	2.20	0.74
1:G:176:GLY:O	3:G:401:HOH:O	2.06	0.73
1:I:74:SER:N	3:I:401:HOH:O	2.20	0.73
1:G:111:ARG:NH2	1:G:129:ILE:O	2.25	0.70
1:A:241:ASN:O	1:A:241:ASN:ND2	2.27	0.68
1:G:132:ARG:NH2	2:H:57:ASN:OD1	2.34	0.61
1:C:266:ASN:HB3	1:C:267:GLY:HA2	1.82	0.60
1:G:137:LYS:NZ	2:H:73:ASP:OD2	2.34	0.60
2:F:106:ARG:NH1	2:F:108:TYR:OH	2.35	0.59
1:I:140:ASN:OD1	2:J:19:ARG:NH1	2.37	0.58
1:G:153:TYR:OH	1:G:229:GLU:OE2	2.27	0.53
1:A:140:ASN:OD1	2:B:19:ARG:NH1	2.42	0.53
1:E:185:GLY:O	1:E:187:SER:N	2.44	0.51
1:A:265:GLY:O	1:A:266:ASN:CB	2.60	0.49
1:I:265:GLY:O	1:I:266:ASN:HB3	2.14	0.48
1:G:97:ASP:OD1	2:H:19:ARG:NH2	2.47	0.48
1:E:88:ASP:OD2	3:E:401:HOH:O	2.20	0.47
1:A:39:HIS:N	1:A:40:PRO:HD2	2.29	0.47

*Continued on next page...*

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:I:106:ILE:HD11	1:I:135:TRP:CE2	2.50	0.46
1:A:265:GLY:O	1:A:266:ASN:HB3	2.16	0.46
1:A:120:GLU:OE1	1:A:120:GLU:N	2.44	0.44
1:I:113:LEU:HB2	1:I:318:THR:HG22	1.99	0.44
1:A:39:HIS:N	1:A:40:PRO:CD	2.80	0.43
1:C:126:ASP:O	1:C:129:ILE:HG22	2.18	0.43
1:E:227:LEU:HD22	1:E:237:VAL:HG23	2.00	0.43
1:G:48:ASN:O	1:G:49:ASP:CB	2.67	0.43
1:G:140:ASN:OD1	2:H:19:ARG:NH1	2.52	0.42
1:C:124:ASP:N	1:C:124:ASP:OD1	2.51	0.42
1:G:48:ASN:O	1:G:49:ASP:HB2	2.18	0.42
2:J:72:ARG:HD3	2:J:74:ASN:OD1	2.20	0.42
2:D:38:ARG:NE	2:D:46:GLU:OE2	2.46	0.41
1:G:319:ILE:O	1:G:320:SER:CB	2.68	0.40

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	276/289 (96%)	270 (98%)	4 (1%)	2 (1%)	25	43
1	C	269/289 (93%)	258 (96%)	11 (4%)	0	100	100
1	E	274/289 (95%)	263 (96%)	8 (3%)	3 (1%)	17	29
1	G	270/289 (93%)	257 (95%)	10 (4%)	3 (1%)	17	29
1	I	249/289 (86%)	245 (98%)	3 (1%)	1 (0%)	38	59
2	B	117/146 (80%)	113 (97%)	4 (3%)	0	100	100
2	D	116/146 (80%)	111 (96%)	5 (4%)	0	100	100
2	F	116/146 (80%)	107 (92%)	9 (8%)	0	100	100
2	H	117/146 (80%)	110 (94%)	7 (6%)	0	100	100

Continued on next page...

*Continued from previous page...*

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
2	J	115/146 (79%)	110 (96%)	5 (4%)	0	100	100
All	All	1919/2175 (88%)	1844 (96%)	66 (3%)	9 (0%)	32	53

All (9) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	E	186	ASN
1	G	108	LYS
1	A	266	ASN
1	E	241	ASN
1	G	57	THR
1	A	241	ASN
1	G	241	ASN
1	I	266	ASN
1	E	183	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	244/254 (96%)	243 (100%)	1 (0%)	93	98
1	C	239/254 (94%)	238 (100%)	1 (0%)	93	98
1	E	238/254 (94%)	237 (100%)	1 (0%)	93	98
1	G	226/254 (89%)	223 (99%)	3 (1%)	73	90
1	I	218/254 (86%)	216 (99%)	2 (1%)	82	94
2	B	90/115 (78%)	89 (99%)	1 (1%)	78	92
2	D	90/115 (78%)	90 (100%)	0	100	100
2	F	91/115 (79%)	91 (100%)	0	100	100
2	H	91/115 (79%)	90 (99%)	1 (1%)	78	92
2	J	87/115 (76%)	86 (99%)	1 (1%)	78	92
All	All	1614/1845 (88%)	1603 (99%)	11 (1%)	87	96

All (11) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	74	SER
2	B	11	LEU
1	C	124	ASP
1	E	258	LYS
1	G	187	SER
1	G	225	LYS
1	G	240	LYS
2	H	11	LEU
1	I	74	SER
1	I	233	THR
2	J	11	LEU

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 ⓘ

### 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, 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	280/289 (96%)	0.01	1 (0%) 92 92	24, 39, 70, 99	0
1	C	273/289 (94%)	0.14	10 (3%) 42 44	30, 47, 85, 150	0
1	E	278/289 (96%)	0.03	3 (1%) 80 81	32, 49, 77, 111	0
1	G	274/289 (94%)	0.12	9 (3%) 47 50	28, 48, 83, 121	0
1	I	259/289 (89%)	0.09	6 (2%) 61 63	30, 55, 88, 104	0
2	B	119/146 (81%)	-0.10	0 100 100	29, 38, 64, 82	0
2	D	118/146 (80%)	-0.01	0 100 100	32, 46, 75, 101	0
2	F	118/146 (80%)	-0.20	0 100 100	29, 41, 69, 86	0
2	H	119/146 (81%)	-0.06	0 100 100	29, 41, 73, 89	0
2	J	117/146 (80%)	-0.12	0 100 100	26, 46, 72, 99	0
All	All	1955/2175 (89%)	0.02	29 (1%) 74 75	24, 46, 79, 150	0

All (29) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	126	ASP	11.4
1	C	125	GLY	10.9
1	C	73	THR	4.1
1	E	185	GLY	3.9
1	G	266	ASN	3.6
1	C	266	ASN	3.5
1	G	241	ASN	3.4
1	C	71	THR	3.3
1	I	61	LEU	3.2
1	G	182	GLY	2.9
1	C	265	GLY	2.7
1	G	186	ASN	2.7
1	G	128	MET	2.6

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
1	I	180	PRO	2.6
1	C	70	LEU	2.3
1	I	319	ILE	2.3
1	E	266	ASN	2.2
1	I	320	SER	2.2
1	C	124	ASP	2.2
1	A	116	SER	2.2
1	C	127	GLN	2.1
1	G	114	LEU	2.1
1	I	113	LEU	2.1
1	I	116	SER	2.1
1	G	69	THR	2.1
1	E	322	SER	2.1
1	G	125	GLY	2.1
1	G	116	SER	2.1
1	C	69	THR	2.0

## 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.