



# wwPDB X-ray Structure Validation Summary Report ⓘ

Aug 22, 2023 – 07:58 AM EDT

PDB ID : 2RFE  
Title : Crystal structure of the complex between the EGFR kinase domain and a Mig6 peptide  
Authors : Zhang, X.; Pickin, K.A.; Bose, R.; Jura, N.; Cole, P.A.; Kuriyan, J.  
Deposited on : 2007-09-28  
Resolution : 2.90 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

<https://www.wwpdb.org/validation/2017/XrayValidationReportHelp>

with specific help available everywhere you see the ⓘ symbol.

The types of validation reports are described at

<http://www.wwpdb.org/validation/2017/FAQs#types>.

---

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

MolProbity : 4.02b-467  
Xtriage (Phenix) : 1.13  
EDS : 2.35  
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) : 2.35

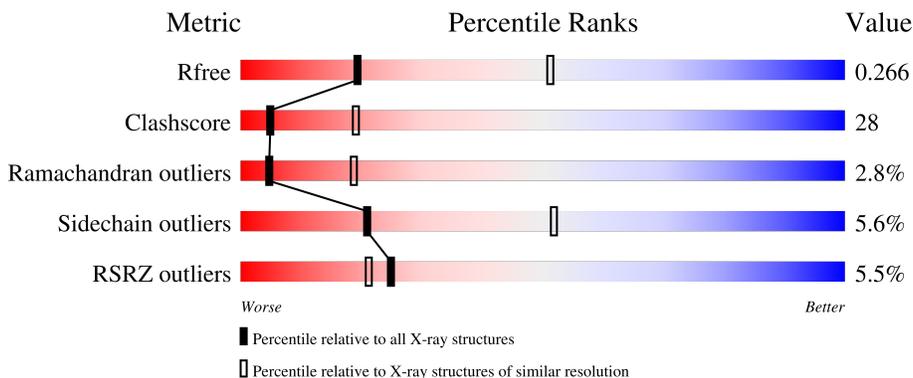
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.90 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.



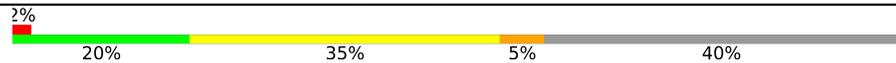
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
$R_{free}$	130704	1957 (2.90-2.90)
Clashscore	141614	2172 (2.90-2.90)
Ramachandran outliers	138981	2115 (2.90-2.90)
Sidechain outliers	138945	2117 (2.90-2.90)
RSRZ outliers	127900	1906 (2.90-2.90)

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 of the lower bar indicate the fraction of residues that contain outliers for  $\geq 3$ , 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  $\leq 5\%$ . 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	324	 2% 45% 40% 12%
1	B	324	 3% 45% 35% 16%
1	C	324	 8% 43% 35% 19%
1	D	324	 6% 49% 35% 14%
2	E	40	 2% 38% 22% 5% 35%

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
2	F	40	 <p>A horizontal bar chart representing the quality of chain. The bar is divided into five segments: a small red segment (2%), a green segment (20%), a yellow segment (35%), a small orange segment (5%), and a grey segment (40%).</p>

## 2 Entry composition i

There are 3 unique types of molecules in this entry. The entry contains 8889 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 Epidermal growth factor receptor.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	286	2210	1426	371	398	15	0	0	0
1	B	273	2116	1367	358	376	15	0	0	0
1	C	262	2028	1309	344	360	15	0	0	0
1	D	278	2130	1376	358	381	15	0	0	0

There are 16 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	675	GLY	-	expression tag	UNP P00533
A	676	ALA	-	expression tag	UNP P00533
A	677	MET	-	expression tag	UNP P00533
A	799	GLU	LYS	engineered mutation	UNP P00533
B	675	GLY	-	expression tag	UNP P00533
B	676	ALA	-	expression tag	UNP P00533
B	677	MET	-	expression tag	UNP P00533
B	799	GLU	LYS	engineered mutation	UNP P00533
C	675	GLY	-	expression tag	UNP P00533
C	676	ALA	-	expression tag	UNP P00533
C	677	MET	-	expression tag	UNP P00533
C	799	GLU	LYS	engineered mutation	UNP P00533
D	675	GLY	-	expression tag	UNP P00533
D	676	ALA	-	expression tag	UNP P00533
D	677	MET	-	expression tag	UNP P00533
D	799	GLU	LYS	engineered mutation	UNP P00533

- Molecule 2 is a protein called ERBB receptor feedback inhibitor 1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	E	26	Total	C	N	O	S	0	0	0
			190	122	29	38	1			
2	F	24	Total	C	N	O	S	0	0	0
			179	116	27	35	1			

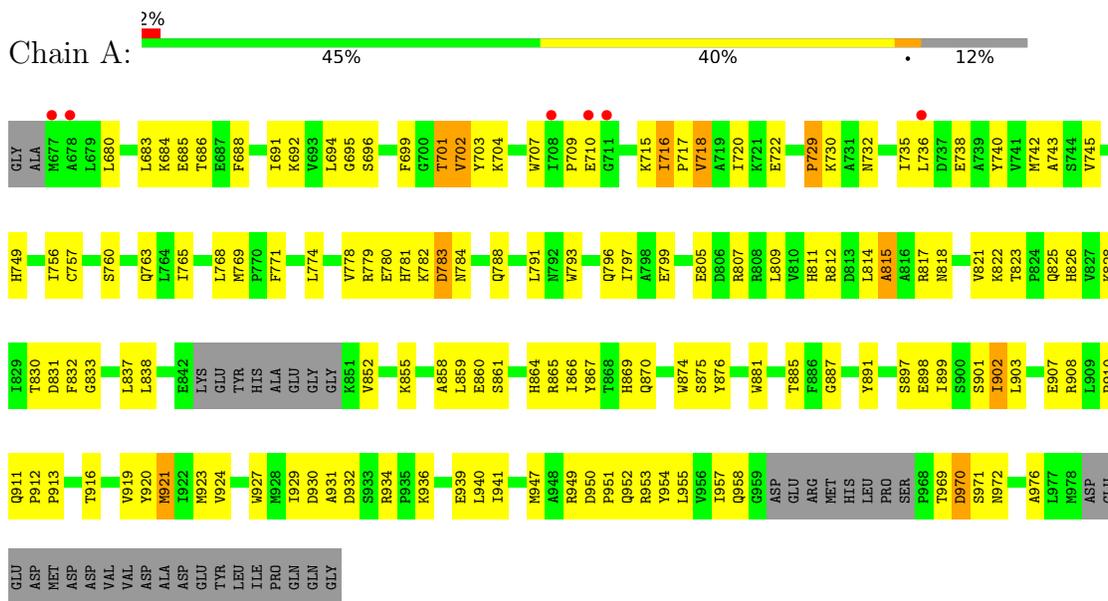
- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	12	Total	O	0	0
			12	12		
3	B	7	Total	O	0	0
			7	7		
3	C	9	Total	O	0	0
			9	9		
3	D	7	Total	O	0	0
			7	7		
3	E	1	Total	O	0	0
			1	1		

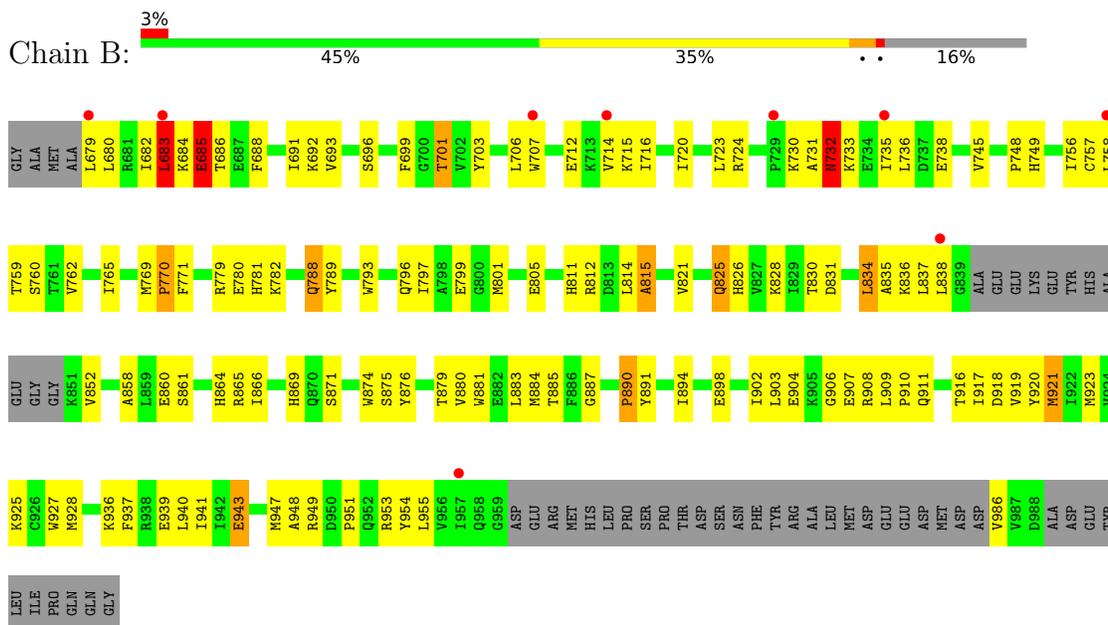
### 3 Residue-property plots i

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.

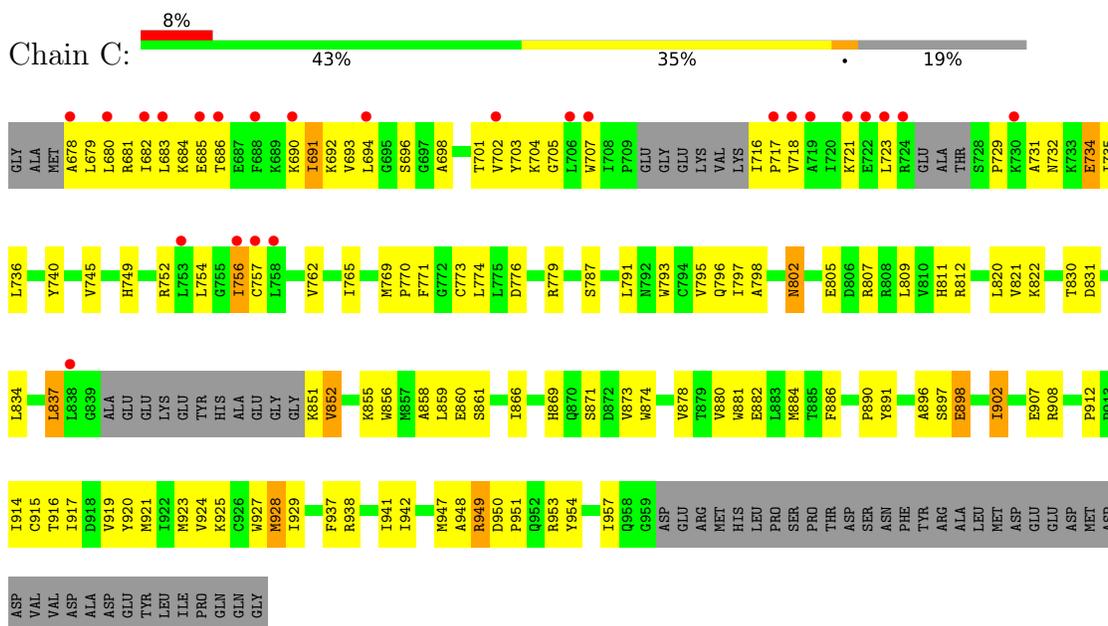
- Molecule 1: Epidermal growth factor receptor



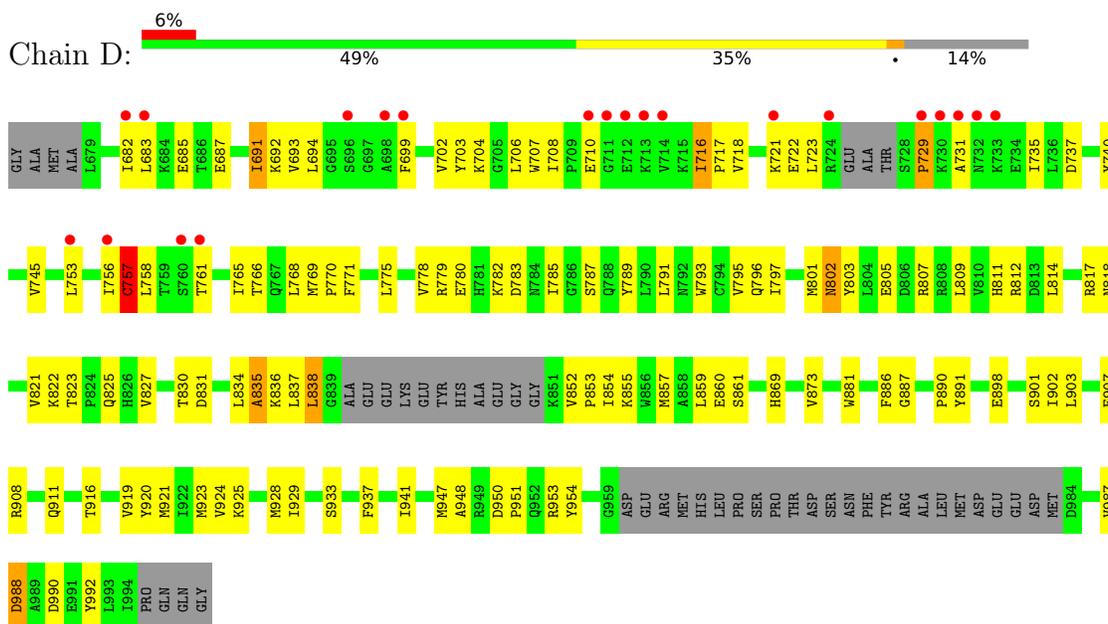
- Molecule 1: Epidermal growth factor receptor



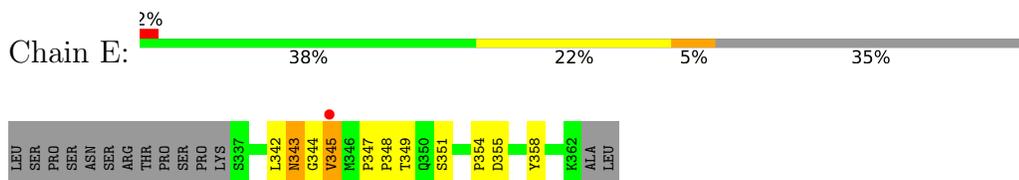
- Molecule 1: Epidermal growth factor receptor



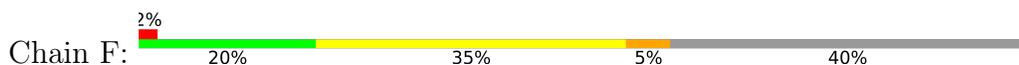
- Molecule 1: Epidermal growth factor receptor



- Molecule 2: ERBB receptor feedback inhibitor 1



- Molecule 2: ERBB receptor feedback inhibitor 1



LEU	SER	PRO	SER	ASN	SER	ARG	THR	PRO	SER	PRO	LYS	S337	L338	P339	S340	Y341	L342	M343	G344	V345	M346	P347	F348	T349	Q350	A353	P354	K357	Y358	V359	S360	SER	LYS	ALA	LEU
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	-----	-----	-----	-----

## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	90.51Å 98.42Å 101.50Å 90.00° 112.59° 90.00°	Depositor
Resolution (Å)	49.85 – 2.90 49.84 – 2.91	Depositor EDS
% Data completeness (in resolution range)	82.6 (49.85-2.90) 83.4 (49.84-2.91)	Depositor EDS
$R_{merge}$	0.08	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.83 (at 2.91Å)	Xtrriage
Refinement program	CNS	Depositor
R, $R_{free}$	0.229 , 0.271 0.227 , 0.266	Depositor DCC
$R_{free}$ test set	1611 reflections (4.95%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	46.2	Xtrriage
Anisotropy	0.864	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.33 , 58.4	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.48$ , $\langle L^2 \rangle = 0.31$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
$F_o, F_c$ correlation	0.91	EDS
Total number of atoms	8889	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	61.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.49% 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	0.45	0/2258	0.67	1/3069 (0.0%)
1	B	0.43	0/2161	0.68	0/2935
1	C	0.45	0/2071	0.68	1/2812 (0.0%)
1	D	0.41	0/2175	0.64	1/2959 (0.0%)
2	E	0.53	0/197	0.82	0/271
2	F	0.49	0/186	0.67	0/256
All	All	0.44	0/9048	0.67	3/12302 (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	C	729	PRO	N-CA-CB	5.83	110.30	103.30
1	A	729	PRO	N-CA-CB	5.71	110.15	103.30
1	D	729	PRO	N-CA-CB	5.57	109.98	103.30

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	2210	0	2171	117	0
1	B	2116	0	2108	131	0
1	C	2028	0	2000	119	0
1	D	2130	0	2078	111	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	E	190	0	174	9	0
2	F	179	0	167	19	0
3	A	12	0	0	0	0
3	B	7	0	0	0	0
3	C	9	0	0	0	0
3	D	7	0	0	0	0
3	E	1	0	0	0	0
All	All	8889	0	8698	483	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 28.

The worst 5 of 483 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:A:949:ARG:HG2	1:B:826:HIS:HB2	1.41	1.03
1:D:769:MET:HE1	1:D:822:LYS:HB2	1.45	0.98
1:A:716:ILE:HD13	1:A:716:ILE:H	1.31	0.94
1:D:716:ILE:HG22	1:D:987:VAL:O	1.69	0.93
1:D:707:TRP:HB3	1:D:716:ILE:HG13	1.54	0.89

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	280/324 (86%)	229 (82%)	45 (16%)	6 (2%)	<b>7</b> <b>26</b>
1	B	267/324 (82%)	224 (84%)	36 (14%)	7 (3%)	<b>5</b> <b>20</b>
1	C	254/324 (78%)	222 (87%)	27 (11%)	5 (2%)	<b>7</b> <b>27</b>
1	D	270/324 (83%)	231 (86%)	31 (12%)	8 (3%)	<b>4</b> <b>17</b>

*Continued on next page...*

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
2	E	24/40 (60%)	17 (71%)	4 (17%)	3 (12%)	0	0
2	F	22/40 (55%)	17 (77%)	3 (14%)	2 (9%)	1	1
All	All	1117/1376 (81%)	940 (84%)	146 (13%)	31 (3%)	5	19

5 of 31 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	729	PRO
1	B	685	GLU
1	B	890	PRO
1	C	685	GLU
1	C	698	ALA

### 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	230/284 (81%)	215 (94%)	15 (6%)	17	45
1	B	223/284 (78%)	209 (94%)	14 (6%)	18	46
1	C	211/284 (74%)	200 (95%)	11 (5%)	23	55
1	D	220/284 (78%)	211 (96%)	9 (4%)	30	64
2	E	22/37 (60%)	21 (96%)	1 (4%)	27	61
2	F	21/37 (57%)	19 (90%)	2 (10%)	8	26
All	All	927/1210 (77%)	875 (94%)	52 (6%)	21	52

5 of 52 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	943	GLU
1	C	852	VAL
2	E	349	THR
1	C	732	ASN
1	C	779	ARG

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 19 such sidechains are listed below:

Mol	Chain	Res	Type
1	D	781	HIS
1	D	818	ASN
1	D	952	GLN
1	D	792	ASN
1	B	792	ASN

### 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 monosaccharides 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 [i](#)

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	286/324 (88%)	0.22	6 (2%) 63 61	28, 50, 96, 104	0
1	B	273/324 (84%)	0.25	9 (3%) 46 41	30, 55, 93, 107	0
1	C	262/324 (80%)	0.46	25 (9%) 8 6	22, 56, 115, 125	0
1	D	278/324 (85%)	0.47	21 (7%) 13 10	29, 66, 103, 118	0
2	E	26/40 (65%)	0.45	1 (3%) 40 36	48, 61, 87, 90	0
2	F	24/40 (60%)	0.54	1 (4%) 36 32	66, 71, 88, 90	0
All	All	1149/1376 (83%)	0.35	63 (5%) 25 21	22, 57, 104, 125	0

The worst 5 of 63 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	723	LEU	4.6
1	B	714	VAL	4.6
1	D	699	PHE	4.3
1	C	683	LEU	4.1
1	D	683	LEU	4.0

### 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 monosaccharides 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.