



# wwPDB X-ray Structure Validation Summary Report

Aug 26, 2024 – 04:04 PM EDT

PDB ID : 9CRW  
Title : Crystal structure of the Candida albicans kinesin-8 proximal tail domain  
Authors : Trofimova, D.; Doubleday, C.; Hunter, B.; Serrano Arevalo, J.; Davison, E.;  
Wen, E.; Munro, K.; Allingham, J.S.  
Deposited on : 2024-07-22  
Resolution : 2.49 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) : 1.20.1  
EDS : 3.0  
Percentile statistics : 20231227.v01 (using entries in the PDB archive December 27th 2023)  
CCP4 : 9.0.002 (Gargrove)  
Density-Fitness : 1.0.11  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : 2.38.3

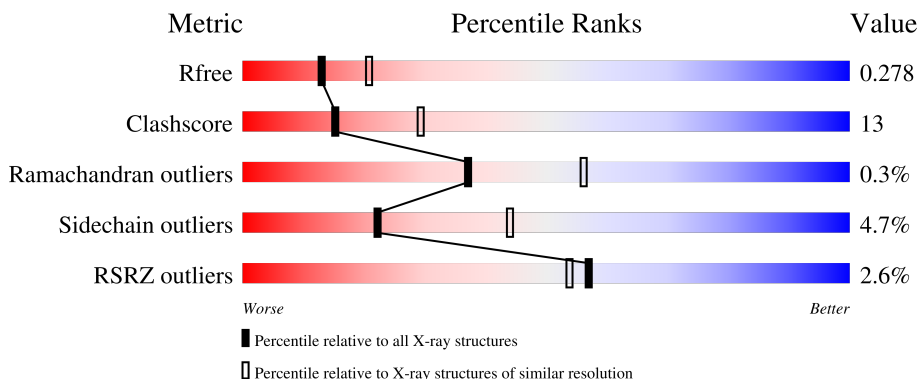
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.49 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.



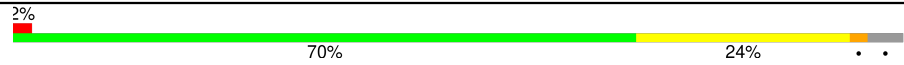

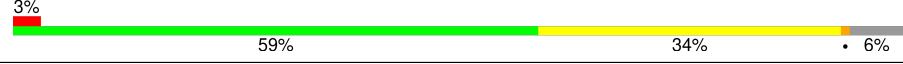
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
$R_{free}$	164625	5504 (2.50-2.50)
Clashscore	180529	6282 (2.50-2.50)
Ramachandran outliers	177936	6191 (2.50-2.50)
Sidechain outliers	177891	6193 (2.50-2.50)
RSRZ outliers	164620	5504 (2.50-2.50)

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 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	232	 2% 64% 31% ..
1	B	232	 3% 66% 23% • 9%
1	C	232	 2% 67% 25% • 6%
1	D	232	 % 75% 21% ..
1	E	232	 3% 63% 31% • 6%

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
1	F	232	
1	G	232	
1	H	232	

## 2 Entry composition

There is only 1 type of molecule in this entry. The entry contains 14593 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.

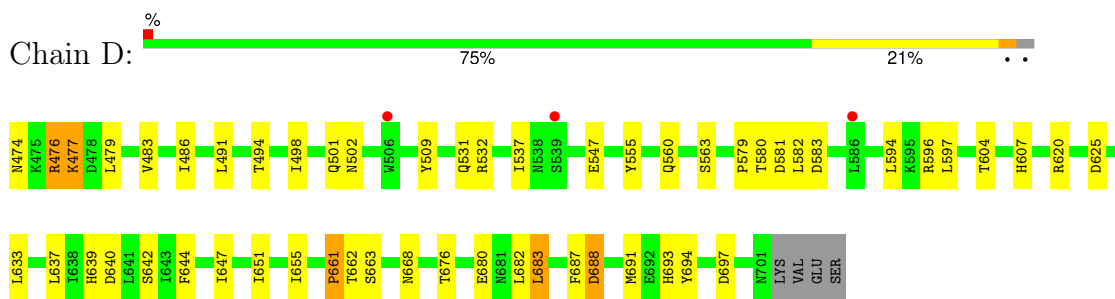
- Molecule 1 is a protein called Kinesin-like protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	D	228	Total 1887	C 1197	N 310	O 373	S 7	0	0	0
1	C	217	Total 1811	C 1154	N 296	O 354	S 7	0	0	0
1	A	226	Total 1870	C 1187	N 306	O 370	S 7	0	0	0
1	B	212	Total 1768	C 1128	N 286	O 347	S 7	0	0	0
1	E	219	Total 1822	C 1160	N 295	O 360	S 7	0	0	0
1	F	222	Total 1841	C 1171	N 299	O 364	S 7	0	0	0
1	G	214	Total 1786	C 1140	N 292	O 347	S 7	0	0	0
1	H	217	Total 1808	C 1153	N 294	O 354	S 7	0	0	0

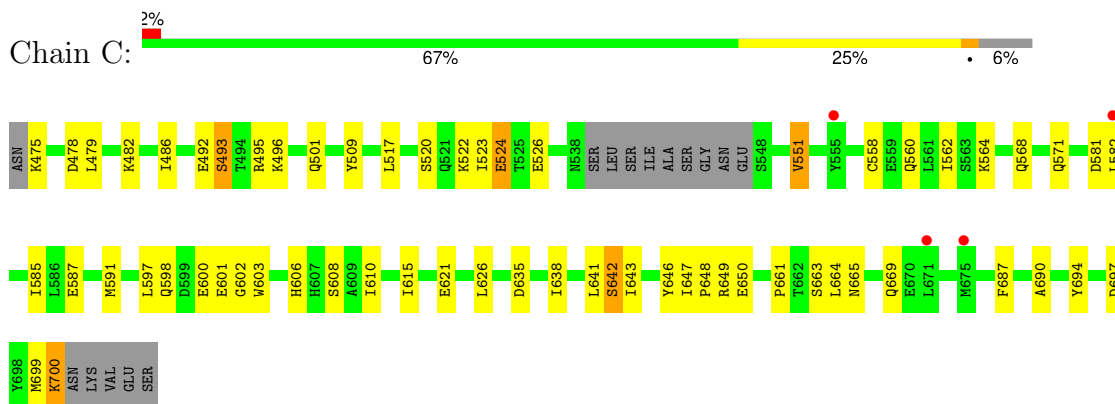
### 3 Residue-property plots [i](#)

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.

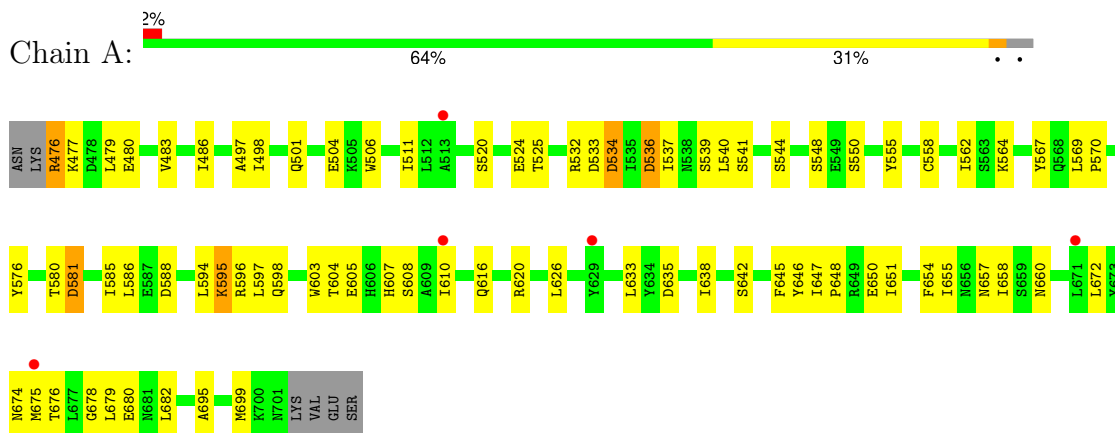
- Molecule 1: Kinesin-like protein



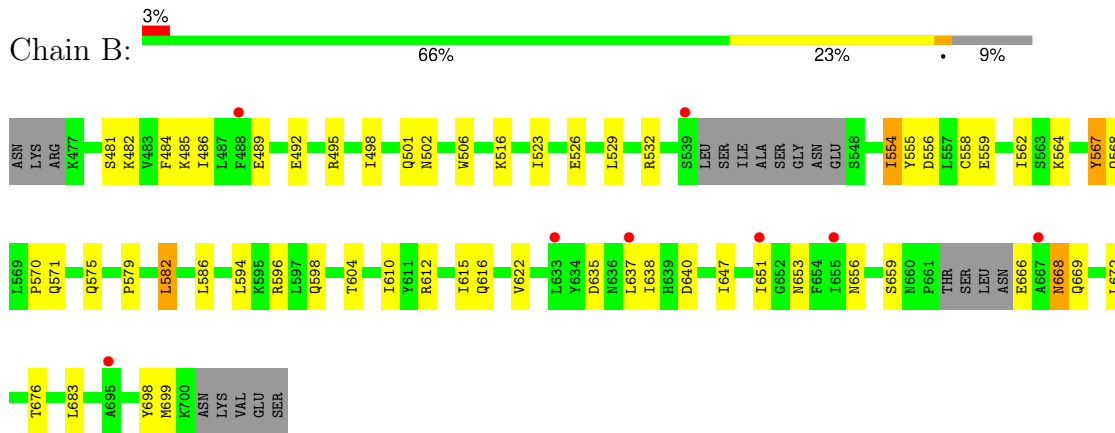
- Molecule 1: Kinesin-like protein



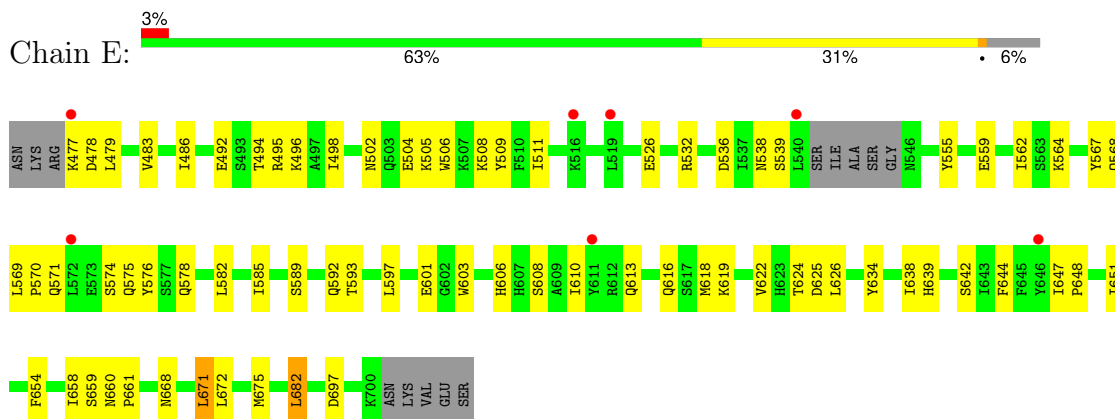
- Molecule 1: Kinesin-like protein



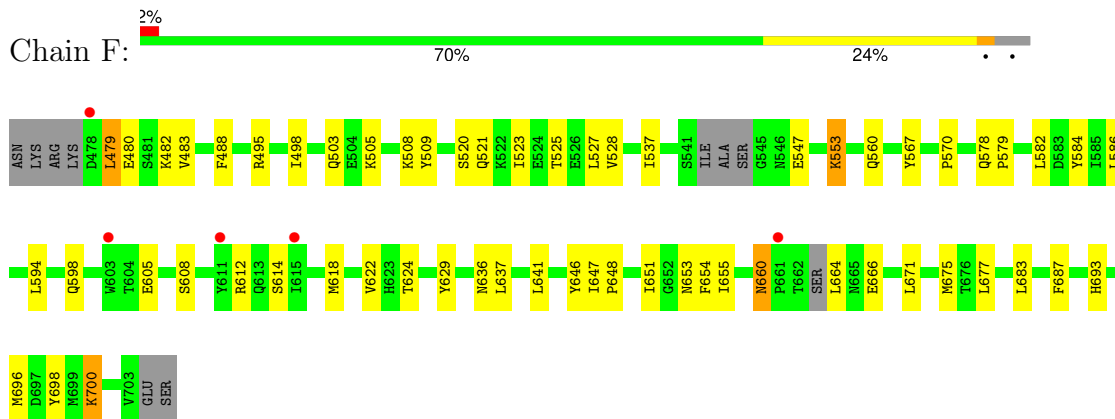
• Molecule 1: Kinesin-like protein



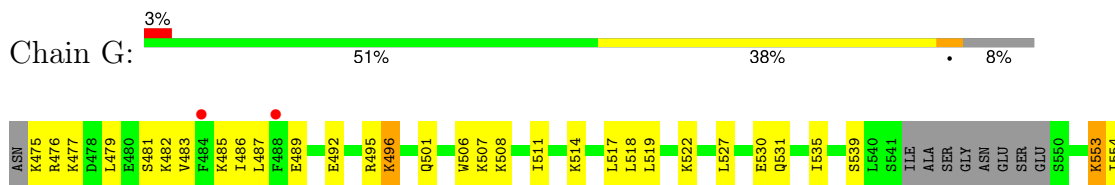
• Molecule 1: Kinesin-like protein

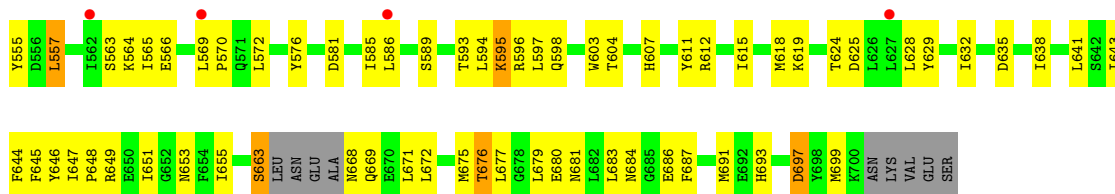


• Molecule 1: Kinesin-like protein

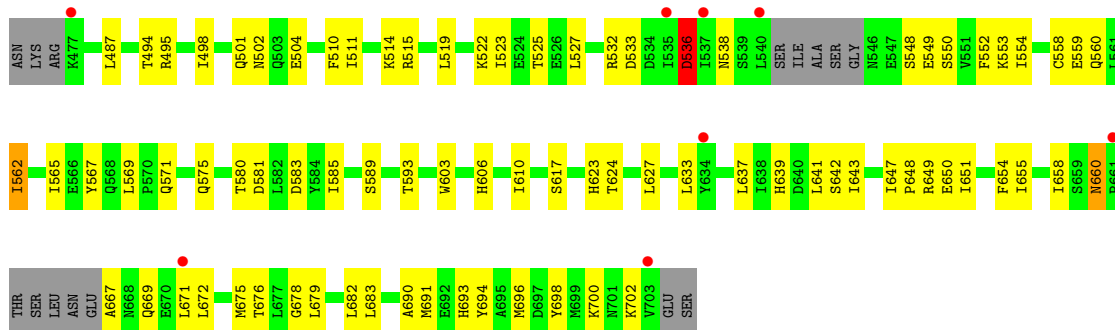


• Molecule 1: Kinesin-like protein





● Molecule 1: Kinesin-like protein



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	83.96Å 104.57Å 118.78Å 90.00° 93.37° 90.00°	Depositor
Resolution (Å)	47.84 – 2.49 47.84 – 2.49	Depositor EDS
% Data completeness (in resolution range)	98.7 (47.84-2.49) 86.3 (47.84-2.49)	Depositor EDS
$R_{merge}$	0.07	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	0.52 (at 2.48Å)	Xtrriage
Refinement program	PHENIX (1.21_5207: ???)	Depositor
R, $R_{free}$	0.234 , 0.278 0.234 , 0.278	Depositor DCC
$R_{free}$ test set	69308 reflections (2.77%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	67.0	Xtrriage
Anisotropy	0.255	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.27 , 56.6	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.50$ , $\langle L^2 \rangle = 0.33$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
$F_o, F_c$ correlation	0.95	EDS
Total number of atoms	14593	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	101.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.44% 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	0.52	0/1902	0.65	0/2563
1	B	0.58	0/1798	0.71	2/2419 (0.1%)
1	C	0.56	1/1842 (0.1%)	0.68	0/2479
1	D	0.55	0/1919	0.64	0/2585
1	E	0.51	0/1853	0.64	0/2496
1	F	0.54	0/1871	0.64	1/2519 (0.0%)
1	G	0.45	0/1816	0.69	2/2442 (0.1%)
1	H	0.52	0/1838	0.65	1/2473 (0.0%)
All	All	0.53	1/14839 (0.0%)	0.66	6/19976 (0.0%)

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	C	524	GLU	CG-CD	5.21	1.59	1.51

The worst 5 of 6 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	637	LEU	CA-CB-CG	5.69	128.40	115.30
1	G	557	LEU	CA-CB-CG	5.63	128.24	115.30
1	G	519	LEU	CA-CB-CG	5.51	127.97	115.30
1	H	562	ILE	CG1-CB-CG2	-5.13	100.11	111.40
1	F	700	LYS	CD-CE-NZ	5.06	123.34	111.70

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	1870	0	1846	51	0
1	B	1768	0	1744	43	0
1	C	1811	0	1795	48	0
1	D	1887	0	1865	36	0
1	E	1822	0	1797	57	0
1	F	1841	0	1814	47	0
1	G	1786	0	1776	83	0
1	H	1808	0	1789	74	0
All	All	14593	0	14426	388	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 13.

The worst 5 of 388 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:H:554:ILE:HD11	1:H:676:THR:HG22	1.52	0.90
1:H:647:ILE:HG23	1:H:648:PRO:HD3	1.55	0.86
1:H:643:ILE:HA	1:H:649:ARG:HH22	1.39	0.84
1:G:553:LYS:O	1:G:557:LEU:HD22	1.77	0.84
1:B:669:GLN:HA	1:B:672:LEU:HD12	1.60	0.83

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	224/232 (97%)	215 (96%)	9 (4%)	0	100 100
1	B	206/232 (89%)	196 (95%)	10 (5%)	0	100 100

*Continued on next page...*

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	C	213/232 (92%)	202 (95%)	11 (5%)	0	100	100
1	D	226/232 (97%)	217 (96%)	7 (3%)	2 (1%)	14	28
1	E	215/232 (93%)	209 (97%)	4 (2%)	2 (1%)	14	28
1	F	216/232 (93%)	208 (96%)	8 (4%)	0	100	100
1	G	208/232 (90%)	193 (93%)	15 (7%)	0	100	100
1	H	211/232 (91%)	196 (93%)	13 (6%)	2 (1%)	14	28
All	All	1719/1856 (93%)	1636 (95%)	77 (4%)	6 (0%)	37	56

5 of 6 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	D	662	THR
1	E	478	ASP
1	H	533	ASP
1	D	661	PRO
1	H	536	ASP

### 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	212/218 (97%)	198 (93%)	14 (7%)	14	28
1	B	200/218 (92%)	193 (96%)	7 (4%)	31	57
1	C	205/218 (94%)	197 (96%)	8 (4%)	27	52
1	D	214/218 (98%)	206 (96%)	8 (4%)	29	55
1	E	207/218 (95%)	199 (96%)	8 (4%)	27	52
1	F	209/218 (96%)	197 (94%)	12 (6%)	17	35
1	G	203/218 (93%)	190 (94%)	13 (6%)	14	30
1	H	205/218 (94%)	198 (97%)	7 (3%)	32	58
All	All	1655/1744 (95%)	1578 (95%)	77 (5%)	22	44

5 of 77 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	G	496	LYS
1	H	536	ASP
1	G	539	SER
1	G	653	ASN
1	H	660	ASN

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 13 such sidechains are listed below:

Mol	Chain	Res	Type
1	E	684	ASN
1	F	531	GLN
1	H	538	ASN
1	G	501	GLN
1	H	501	GLN

### 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 oligosaccharides 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	226/232 (97%)	0.13	5 (2%) 62 59	52, 89, 132, 161	0
1	B	212/232 (91%)	0.16	8 (3%) 44 41	53, 83, 126, 152	0
1	C	217/232 (93%)	-0.05	4 (1%) 67 64	60, 86, 127, 165	0
1	D	228/232 (98%)	-0.05	3 (1%) 74 71	63, 87, 122, 172	0
1	E	219/232 (94%)	0.11	7 (3%) 50 47	60, 96, 133, 177	0
1	F	222/232 (95%)	0.11	5 (2%) 61 58	57, 91, 148, 177	0
1	G	214/232 (92%)	0.39	6 (2%) 55 51	74, 136, 165, 184	0
1	H	217/232 (93%)	0.23	8 (3%) 45 42	79, 109, 160, 184	0
All	All	1755/1856 (94%)	0.13	46 (2%) 57 54	52, 96, 154, 184	0

The worst 5 of 46 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	H	661	PRO	3.9
1	B	667	ALA	3.5
1	H	537	ILE	3.5
1	B	539	SER	3.4
1	E	540	LEU	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 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.