



# wwPDB X-ray Structure Validation Summary Report ⓘ

May 25, 2020 – 06:08 pm BST

PDB ID : 4RR2  
Title : Crystal structure of human primase  
Authors : Baranovskiy, A.G.; Gu, J.; Suwa, Y.; Babayeva, N.D.; Tahirov, T.H.  
Deposited on : 2014-11-05  
Resolution : 2.65 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 following versions of software and data (see [references ⓘ](#)) were used in the production of this report:

MolProbity : 4.02b-467  
Mogul : 1.8.5 (274361), CSD as541be (2020)  
Xtriage (Phenix) : 1.13  
EDS : 2.11  
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) : 2.11

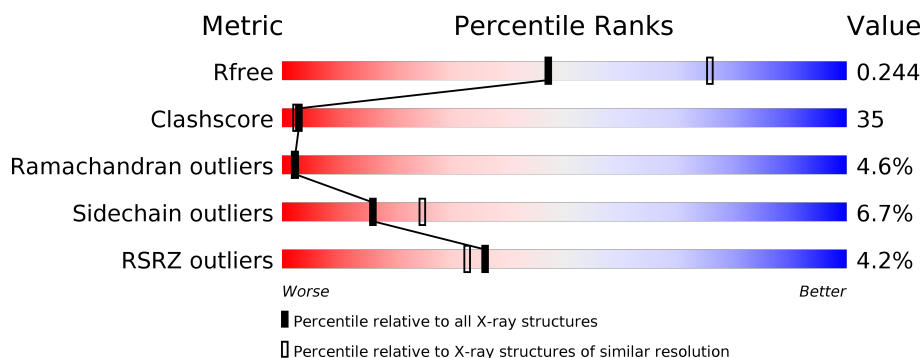
# 1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 2.65 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	1332 (2.68-2.64)
Clashscore	141614	1374 (2.68-2.64)
Ramachandran outliers	138981	1349 (2.68-2.64)
Sidechain outliers	138945	1349 (2.68-2.64)
RSRZ outliers	127900	1318 (2.68-2.64)

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 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	420	<div> <div>%</div> <div> <div></div> <div>46%</div> <div>43%</div> <div>•</div> <div>7%</div> </div> </div>
1	C	420	<div> <div>%</div> <div> <div></div> <div>44%</div> <div>43%</div> <div>6%</div> <div>7%</div> </div> </div>
2	B	509	<div> <div>6%</div> <div> <div></div> <div>14%</div> <div>23%</div> <div>5%</div> <div>58%</div> </div> </div>
2	D	509	<div> <div>4%</div> <div> <div></div> <div>40%</div> <div>37%</div> <div>6%</div> <div>•</div> <div>16%</div> </div> </div>

## 2 Entry composition

There are 5 unique types of molecules in this entry. The entry contains 11896 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 DNA primase small subunit.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	389	Total	C	N	O	S	0	0	0
			3267	2101	566	585	15			
1	C	389	Total	C	N	O	S	0	0	0
			3263	2099	566	583	15			

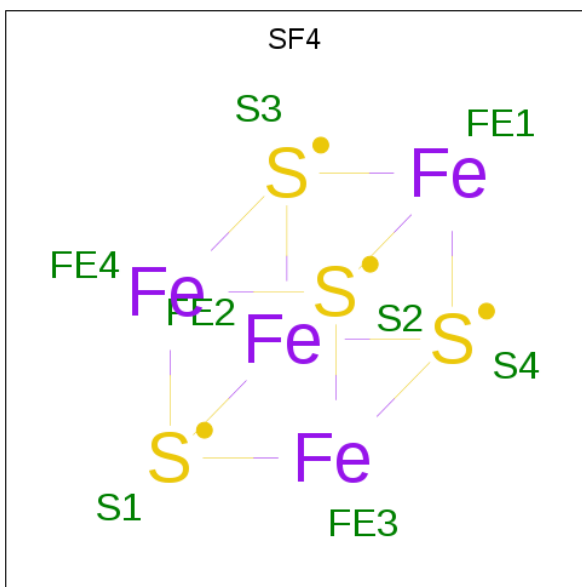
- Molecule 2 is a protein called DNA primase large subunit.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	213	Total	C	N	O	S	0	0	0
			1755	1133	299	321	2			
2	D	429	Total	C	N	O	S	0	0	0
			3513	2249	608	643	13			

- Molecule 3 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	1	Total	Zn	0	0
			1	1		
3	C	1	Total	Zn	0	0
			1	1		

- Molecule 4 is IRON/SULFUR CLUSTER (three-letter code: SF4) (formula: Fe<sub>4</sub>S<sub>4</sub>).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
4	D	1	Total	Fe	S	0	0
			8	4	4		

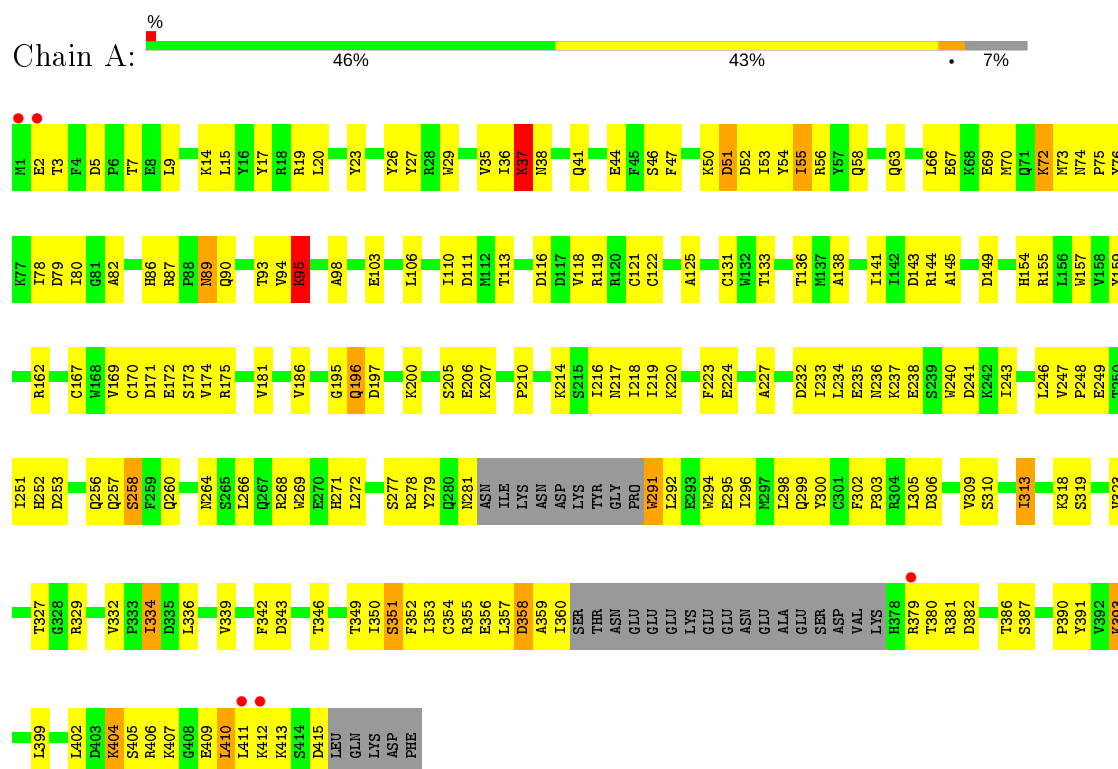
- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	23	Total	O	0	0
			23	23		
5	B	4	Total	O	0	0
			4	4		
5	C	40	Total	O	0	0
			40	40		
5	D	21	Total	O	0	0
			21	21		

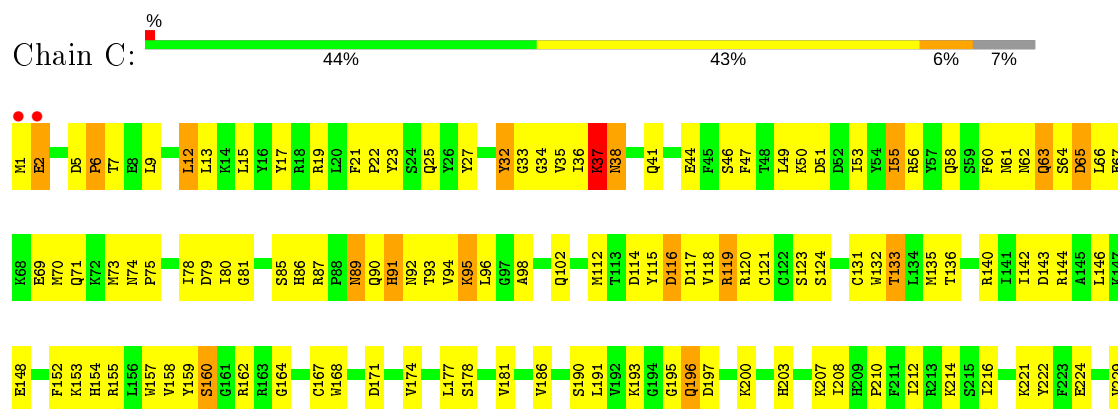
### 3 Residue-property plots [i](#)

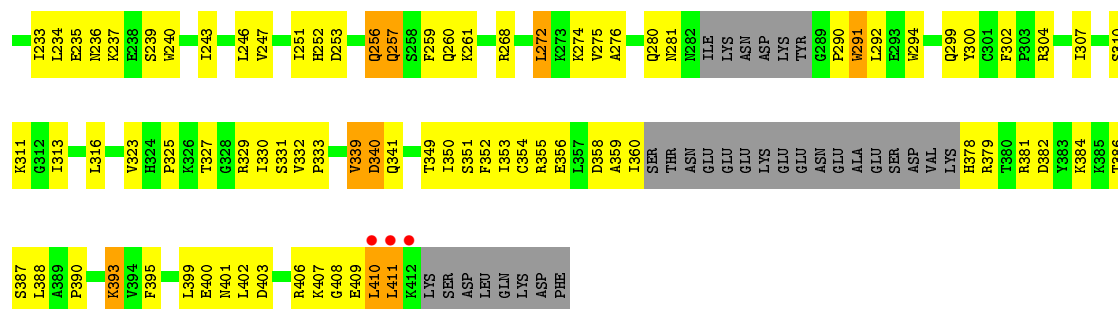
These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are 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: DNA primase small subunit

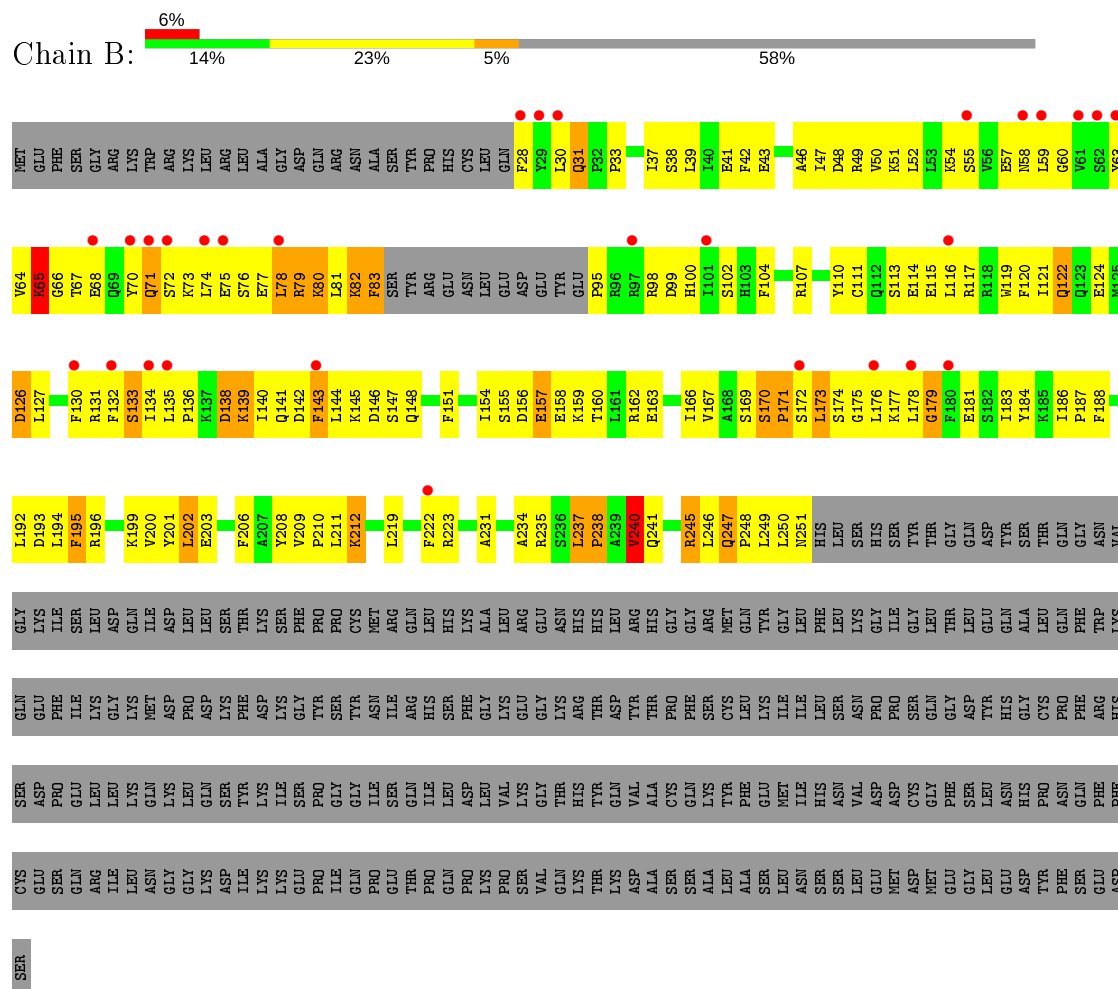


- Molecule 1: DNA primase small subunit

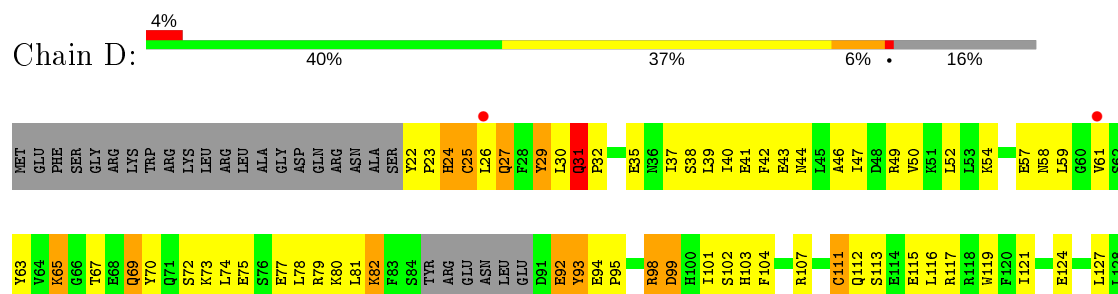




• Molecule 2: DNA primase large subunit



• Molecule 2: DNA primase large subunit



GLY	T360	Q275	K212	R129
LYS	D361	L276	D213	F130
ASP	Y362		I214	R131
ILE	T363	S280		I134
LYS	P364		I218	L135
LYS	F365	F284	L219	P136
GLU	S366	P285	N220	K137
PRO		P286	E221	D138
ILE	K369	C287	F222	K139
GLN	I370	N288		I140
PRO		N289	K225	
GLU	S373	Q290	L226	L144
THR	N374	L291	S227	
PRO	P375	N292	K228	S147
GLN	P376	K293	A229	Q148
PRO	S377	L294	L230	
LYS	Q378	L295	A231	I154
PRO		N296	L232	S155
SER	H382	E297	T233	D156
VAL			K234	E157
GLN	P385	I311	R235	E158
LYS	F386	F312	S236	K159
THR	R387		L237	
LYS		I316	P238	
ASP	E392		N239	E163
ALA	L393	L320	V240	Q164
SER	L394	E321	Q241	E165
SER		Q322	S242	I166
ALA	L414	A323	D243	V167
LEU	V415	L324	P244	A168
ALA	K416	Q325	R245	S169
SER	G417	F326	L246	S170
LEU	T418	N327	Q247	P171
ASN	H419	K328	P248	S172
SER	Y420	Q329	L249	L173
SER	Q421	E330	L250	S174
LEU	V422		N251	G175
GLU	A423	K335	P252	L176
MET	C424	N336	L253	A177
ASP	Q425	D337	S254	L178
MET			H255	G179
GLU	E429	K340	S256	
GLY	M430	F341	Y257	I183
LEU	L431	D342	T258	Y184
GLU	H432	K343	G259	
ASP		G344	D261	L192
TYR	C437	Y345	S260	
PHE	G438	S346	Y262	F195
SER	F439		S263	R196
GLU	S440		T264	G197
GLU	L441	I349	Q265	A198
ASP	N442	R350	G266	K199
SER	H443	S352	N267	V200
	P444	F353	V268	Y201
	N445	G354	G269	L202
	Q446	K355	K270	E203
	F447	E356	I271	
		G357	S272	V209
		K358	L273	P210
		R359	D274	L211
				GLY

## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	86.19 Å 88.90 Å 94.68 Å 93.82° 96.57° 111.72°	Depositor
Resolution (Å)	48.48 – 2.65 48.47 – 2.60	Depositor EDS
% Data completeness (in resolution range)	88.5 (48.48-2.65) 88.1 (48.47-2.60)	Depositor EDS
$R_{merge}$	0.04	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.89 (at 2.61 Å)	Xtriage
Refinement program	CNS 1.1	Depositor
R, $R_{free}$	0.229 , 0.271 0.206 , 0.244	Depositor DCC
$R_{free}$ test set	3691 reflections (5.10%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	71.2	Xtriage
Anisotropy	0.392	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.30 , 51.9	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.50$ , $\langle L^2 \rangle = 0.33$	Xtriage
Estimated twinning fraction	0.006 for -k,-h,-l	Xtriage
$F_o, F_c$ correlation	0.95	EDS
Total number of atoms	11896	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	53.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.24% 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](#)

Bond lengths and bond angles in the following residue types are not validated in this section: ZN, SF4

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	0/3349	0.70	0/4514
1	C	0.49	0/3346	0.74	0/4512
2	B	0.39	0/1787	0.63	0/2402
2	D	0.44	0/3595	0.67	1/4838 (0.0%)
All	All	0.45	0/12077	0.69	1/16266 (0.0%)

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	D	252	HIS	N-CA-C	5.33	125.39	111.00

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	3267	0	3249	206	0
1	C	3263	0	3243	198	0
2	B	1755	0	1801	184	0
2	D	3513	0	3494	232	0
3	A	1	0	0	0	0
3	C	1	0	0	0	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
4	D	8	0	0	0	0
5	A	23	0	0	11	0
5	B	4	0	0	0	0
5	C	40	0	0	11	0
5	D	21	0	0	3	0
All	All	11896	0	11787	815	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 35.

The worst 5 of 815 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:D:82:LYS:HE2	2:D:98:ARG:HD3	1.36	1.07
2:D:117:ARG:HG2	2:D:230:LEU:HD13	1.38	1.05
2:B:68:GLU:HA	2:B:71:GLN:HB3	1.42	1.02
2:B:82:LYS:HG2	2:B:98:ARG:HD3	1.42	1.01
1:C:50:LYS:HD2	1:C:50:LYS:H	1.25	1.00

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	383/420 (91%)	331 (86%)	45 (12%)	7 (2%)	8	12
1	C	383/420 (91%)	327 (85%)	36 (9%)	20 (5%)	2	1
2	B	209/509 (41%)	156 (75%)	37 (18%)	16 (8%)	1	0
2	D	425/509 (84%)	346 (81%)	58 (14%)	21 (5%)	2	2
All	All	1400/1858 (75%)	1160 (83%)	176 (13%)	64 (5%)	2	2

5 of 64 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	37	LYS
1	A	50	LYS
1	A	95	LYS
1	A	196	GLN
1	A	380	THR

### 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	364/393 (93%)	347 (95%)	17 (5%)	26	40
1	C	363/393 (92%)	341 (94%)	22 (6%)	18	29
2	B	194/459 (42%)	171 (88%)	23 (12%)	5	7
2	D	389/459 (85%)	363 (93%)	26 (7%)	16	25
All	All	1310/1704 (77%)	1222 (93%)	88 (7%)	16	25

5 of 88 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	B	247	GLN
1	C	119	ARG
2	D	288	MET
1	C	2	GLU
1	C	55	ILE

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 63 such sidechains are listed below:

Mol	Chain	Res	Type
1	C	86	HIS
1	C	257	GLN
2	D	329	GLN
1	C	89	ASN
1	C	102	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 carbohydrates in this entry.

## 5.6 Ligand geometry [i](#)

Of 3 ligands modelled in this entry, 2 are monoatomic - leaving 1 for Mogul analysis.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the Chemical Component Dictionary. The Link column lists molecule types, if any, to which the group is linked. 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  $|Z| > 2$  is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\# Z  > 2$	Counts	RMSZ	$\# Z  > 2$
4	SF4	D	1000	2	0,12,12	0.00	-	-		

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the Chemical Component Dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
4	SF4	D	1000	2	-	-	0/6/5/5

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 [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	389/420 (92%)	-0.05	5 (1%) 77 75	13, 42, 78, 100	0
1	C	389/420 (92%)	-0.07	5 (1%) 77 75	14, 37, 75, 95	0
2	B	213/509 (41%)	0.61	29 (13%) 3 2	23, 77, 100, 109	0
2	D	429/509 (84%)	0.11	20 (4%) 31 28	15, 61, 95, 107	0
All	All	1420/1858 (76%)	0.09	59 (4%) 36 33	13, 50, 93, 109	0

The worst 5 of 59 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	B	29	TYR	9.1
1	C	1	MET	6.8
2	D	361	ASP	6.7
1	A	1	MET	6.7
2	B	74	LEU	6.5

### 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](#)

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, 95<sup>th</sup> 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( $\text{\AA}^2$ )	Q<0.9
4	SF4	D	1000	8/8	0.98	0.21	17,26,32,38	0
3	ZN	A	800	1/1	0.98	0.19	34,34,34,34	0
3	ZN	C	900	1/1	1.00	0.21	26,26,26,26	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.