



# wwPDB X-ray Structure Validation Summary Report ⓘ

Oct 10, 2017 – 06:43 AM EDT

PDB ID : 2G4C  
Title : Crystal Structure of human DNA polymerase gamma accessory subunit  
Authors : Fan, L.; Farr, C.L.; Kaguni, L.S.; Tainer, J.A.  
Deposited on : unknown  
Resolution : 3.15 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
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20030345
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-20030345

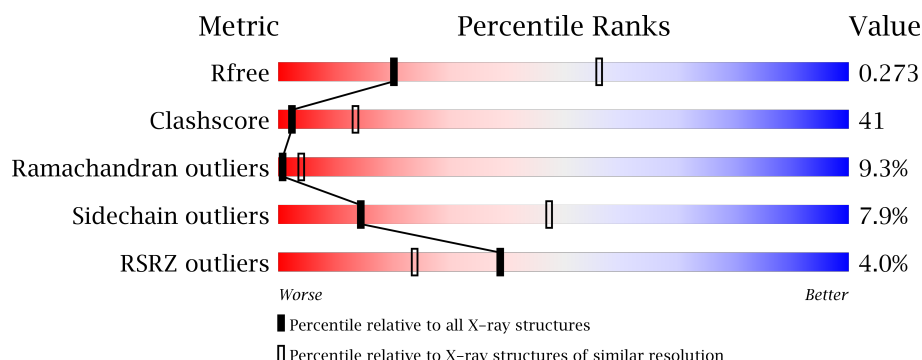
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 3.15 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	1259 (3.20-3.12)
Clashscore	112137	1397 (3.20-3.12)
Ramachandran outliers	110173	1368 (3.20-3.12)
Sidechain outliers	110143	1367 (3.20-3.12)
RSRZ outliers	101464	1264 (3.20-3.12)

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	474	<div> <div>2%</div> <div> <div></div> <div>34%</div> <div>41%</div> <div>8%</div> <div>16%</div> </div> </div>
1	B	474	<div> <div>2%</div> <div> <div></div> <div>34%</div> <div>41%</div> <div>8%</div> <div>16%</div> </div> </div>
1	C	474	<div> <div>5%</div> <div> <div></div> <div>31%</div> <div>43%</div> <div>9%</div> <div>16%</div> </div> </div>
1	D	474	<div> <div>5%</div> <div> <div></div> <div>31%</div> <div>43%</div> <div>9%</div> <div>16%</div> </div> </div>

## 2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 12847 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 polymerase gamma subunit 2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	397	Total	C	N	O	S	0	0	0
			3202	2050	558	578	16			
1	B	397	Total	C	N	O	S	0	0	0
			3202	2050	558	578	16			
1	C	397	Total	C	N	O	S	0	0	0
			3206	2053	560	577	16			
1	D	397	Total	C	N	O	S	0	0	0
			3206	2053	560	577	16			

There are 60 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	12	MET	-	INITIATING METHIONINE	UNP Q9UHN1
A	13	ARG	-	CLONING ARTIFACT	UNP Q9UHN1
A	14	GLY	-	CLONING ARTIFACT	UNP Q9UHN1
A	15	SER	-	CLONING ARTIFACT	UNP Q9UHN1
A	16	HIS	-	EXPRESSION TAG	UNP Q9UHN1
A	17	HIS	-	EXPRESSION TAG	UNP Q9UHN1
A	18	HIS	-	EXPRESSION TAG	UNP Q9UHN1
A	19	HIS	-	EXPRESSION TAG	UNP Q9UHN1
A	20	HIS	-	EXPRESSION TAG	UNP Q9UHN1
A	21	HIS	-	EXPRESSION TAG	UNP Q9UHN1
A	22	GLY	-	CLONING ARTIFACT	UNP Q9UHN1
A	23	SER	-	CLONING ARTIFACT	UNP Q9UHN1
A	24	THR	-	CLONING ARTIFACT	UNP Q9UHN1
A	25	MET	-	CLONING ARTIFACT	UNP Q9UHN1
A	169	THR	ALA	SEE REMARK 999	UNP Q9UHN1
B	12	MET	-	INITIATING METHIONINE	UNP Q9UHN1
B	13	ARG	-	CLONING ARTIFACT	UNP Q9UHN1
B	14	GLY	-	CLONING ARTIFACT	UNP Q9UHN1
B	15	SER	-	CLONING ARTIFACT	UNP Q9UHN1
B	16	HIS	-	EXPRESSION TAG	UNP Q9UHN1
B	17	HIS	-	EXPRESSION TAG	UNP Q9UHN1

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
B	18	HIS	-	EXPRESSION TAG	UNP Q9UHN1
B	19	HIS	-	EXPRESSION TAG	UNP Q9UHN1
B	20	HIS	-	EXPRESSION TAG	UNP Q9UHN1
B	21	HIS	-	EXPRESSION TAG	UNP Q9UHN1
B	22	GLY	-	CLONING ARTIFACT	UNP Q9UHN1
B	23	SER	-	CLONING ARTIFACT	UNP Q9UHN1
B	24	THR	-	CLONING ARTIFACT	UNP Q9UHN1
B	25	MET	-	CLONING ARTIFACT	UNP Q9UHN1
B	169	THR	ALA	SEE REMARK 999	UNP Q9UHN1
C	12	MET	-	INITIATING METHIONINE	UNP Q9UHN1
C	13	ARG	-	CLONING ARTIFACT	UNP Q9UHN1
C	14	GLY	-	CLONING ARTIFACT	UNP Q9UHN1
C	15	SER	-	CLONING ARTIFACT	UNP Q9UHN1
C	16	HIS	-	EXPRESSION TAG	UNP Q9UHN1
C	17	HIS	-	EXPRESSION TAG	UNP Q9UHN1
C	18	HIS	-	EXPRESSION TAG	UNP Q9UHN1
C	19	HIS	-	EXPRESSION TAG	UNP Q9UHN1
C	20	HIS	-	EXPRESSION TAG	UNP Q9UHN1
C	21	HIS	-	EXPRESSION TAG	UNP Q9UHN1
C	22	GLY	-	CLONING ARTIFACT	UNP Q9UHN1
C	23	SER	-	CLONING ARTIFACT	UNP Q9UHN1
C	24	THR	-	CLONING ARTIFACT	UNP Q9UHN1
C	25	MET	-	CLONING ARTIFACT	UNP Q9UHN1
C	169	THR	ALA	SEE REMARK 999	UNP Q9UHN1
D	12	MET	-	INITIATING METHIONINE	UNP Q9UHN1
D	13	ARG	-	CLONING ARTIFACT	UNP Q9UHN1
D	14	GLY	-	CLONING ARTIFACT	UNP Q9UHN1
D	15	SER	-	CLONING ARTIFACT	UNP Q9UHN1
D	16	HIS	-	EXPRESSION TAG	UNP Q9UHN1
D	17	HIS	-	EXPRESSION TAG	UNP Q9UHN1
D	18	HIS	-	EXPRESSION TAG	UNP Q9UHN1
D	19	HIS	-	EXPRESSION TAG	UNP Q9UHN1
D	20	HIS	-	EXPRESSION TAG	UNP Q9UHN1
D	21	HIS	-	EXPRESSION TAG	UNP Q9UHN1
D	22	GLY	-	CLONING ARTIFACT	UNP Q9UHN1
D	23	SER	-	CLONING ARTIFACT	UNP Q9UHN1
D	24	THR	-	CLONING ARTIFACT	UNP Q9UHN1
D	25	MET	-	CLONING ARTIFACT	UNP Q9UHN1
D	169	THR	ALA	SEE REMARK 999	UNP Q9UHN1

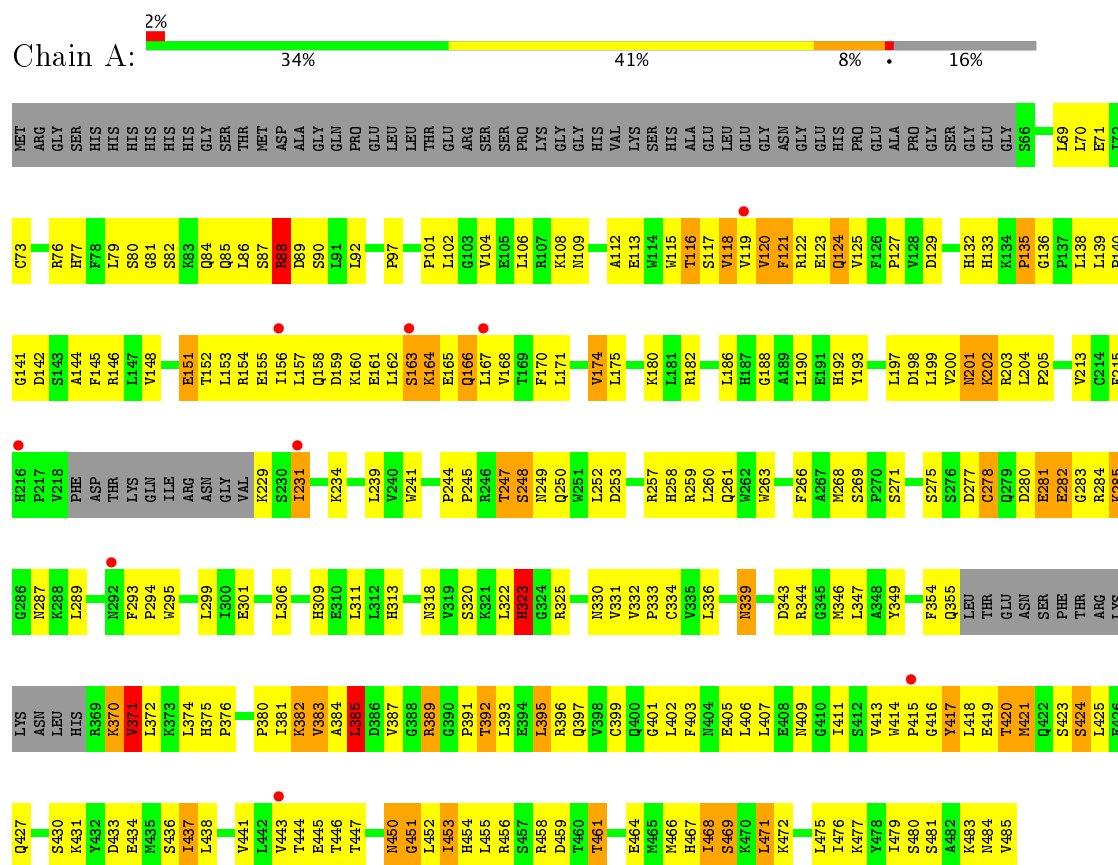
- Molecule 2 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	12	Total 12	O 12	0	0
2	B	5	Total 5	O 5	0	0
2	C	5	Total 5	O 5	0	0
2	D	9	Total 9	O 9	0	0

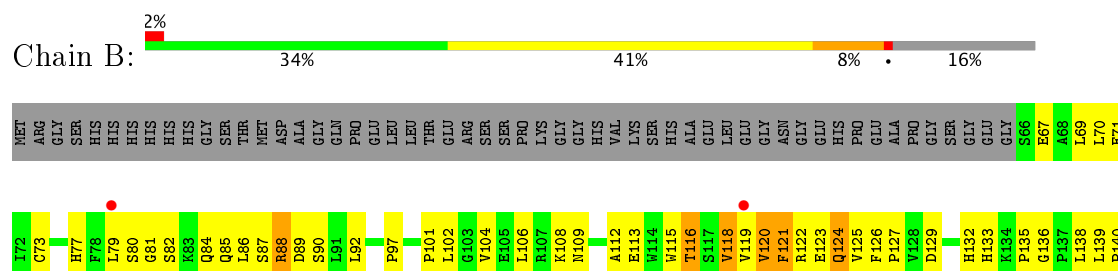
### 3 Residue-property plots

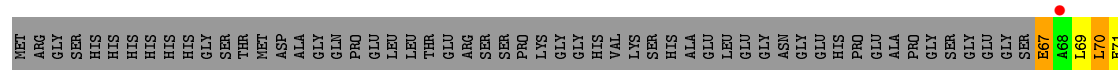
These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are 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 polymerase gamma subunit 2



#### • Molecule 1: DNA polymerase gamma subunit 2





S424	L425	E426	Q427	L428	Y429	S430	K431	Y432	D433	E434	M435	S436	I437	L438	F439	V440	V441	L442	V443	T444	E445	T446	T447	L448	G451	L452	I453	H454	L455	R456	S457	T460	T461	M462	K463	M466	H467	I468	S469	K470	L471	K472	D473	F474	L475	I476	K477	Y478	I479	S480	S481	A482	K483	V484	V485	
PHE	THR	ARG	LYS	LYS	ASN	LEU	H368	R369	K370	V371	L372	H375	P376	C377	L378	L381	K382	V383	A384	L385	D386	V387	G388	R389	G390	P391	T392	L393	E394	L395	R396	C399	Q400	Q401	L402	F403	M404	E405	L406	L407	E408	M409	G410	I411	S412	M413	M414	P415	G416	Y417	L418	P419	T420	M421	Q422	S423
E282	G283	R284	K285	G286	N287	K288	L289	W295	L299	I300	E301	T302	L303	L306	G307	D308	H309	E310	L311	L312	H313	M314	Y315	P316	G317	N318	V319	S320	K321	L322	H323	G324	R325	V331	V332	P333	C334	V335	L336	S337	V338	M346	L350	Y351	D352	S353	F354	Q355	LEU	THR	GLU	ASN	SER			
C214	F215	H216	P217	V218	PHE	ASP	THR	LYS	GLN	ILE	ARG	ASN	GLY	V44	K229	D159	K160	E161	L162	S163	K164	E165	Q166	A237	S238	L239	V240	W241	F242	T243	P244	T247	Q250	W251	L252	D253	F254	W255	L256	R257	L260	R264	K265	F266	V200	P270	S271	N272	S275	S276	D277	C278	Q279	D280	ASN	E281
D142	R146	L147	V148	S149	A150	E151	T152	L153	R154	E155	I156	L157	Q158	R159	K160	E161	L162	S163	K164	E165	Q166	L167	Y168	T169	F170	L171	L175	S178	G179	K180	R182	L186	A189	L190	E191	H192	Y193	L197	D198	L199	V200	K202	R203	L204	P205	A209	Q210	I211	G212	V213						
I72	C73	Q74	R75	R76	H77	F78	L79	S82	K83	Q84	Q85	L86	S87	R88	D89	S90	L91	L92	S93	C94	C95	H96	P97	L102	G103	V104	E105	L106	H107	K108	E113	T116	S117	V118	V119	F120	F121	R122	E123	Q124	V125	F126	P127	V128	D129	H132	H133	K134	P135	G136	F137	L138	G141			



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 32	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	101.79Å 101.79Å 170.28Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	50.00 – 3.15 48.76 – 3.17	Depositor EDS
% Data completeness (in resolution range)	(Not available) (50.00-3.15) 95.9 (48.76-3.17)	Depositor EDS
$R_{merge}$	0.11	Depositor
$R_{sym}$	0.11	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.09 (at 3.19Å)	Xtriage
Refinement program	CNS 1.1	Depositor
R, $R_{free}$	0.232 , 0.285 0.222 , 0.273	Depositor DCC
$R_{free}$ test set	1610 reflections (4.99%)	DCC
Wilson B-factor (Å <sup>2</sup> )	91.6	Xtriage
Anisotropy	0.063	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.33 , 75.6	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.32$	Xtriage
Estimated twinning fraction	0.467 for -h,-k,l 0.054 for h,-h-k,-l 0.054 for -k,-h,-l	Xtriage
$F_o, F_c$ correlation	0.93	EDS
Total number of atoms	12847	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	87.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 13.37% 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	0.39	0/3278	0.74	3/4432 (0.1%)
1	B	0.39	0/3278	0.74	3/4432 (0.1%)
1	C	0.39	0/3283	0.66	1/4439 (0.0%)
1	D	0.39	0/3283	0.66	1/4439 (0.0%)
All	All	0.39	0/13122	0.70	8/17742 (0.0%)

There are no bond length outliers.

The worst 5 of 8 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	88	ARG	NE-CZ-NH1	-13.41	113.59	120.30
1	B	88	ARG	NE-CZ-NH2	-13.18	113.71	120.30
1	A	88	ARG	NE-CZ-NH2	13.12	126.86	120.30
1	B	88	ARG	NE-CZ-NH1	13.01	126.80	120.30
1	B	88	ARG	CD-NE-CZ	6.97	133.35	123.60

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	3202	0	3209	276	0
1	B	3202	0	3209	267	0
1	C	3206	0	3211	289	0
1	D	3206	0	3211	277	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	A	12	0	0	0	0
2	B	5	0	0	0	0
2	C	5	0	0	0	0
2	D	9	0	0	0	0
All	All	12847	0	12840	1063	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 41.

The worst 5 of 1063 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:450:ASN:HD22	1:B:452:LEU:HD23	1.11	1.08
1:A:450:ASN:HD22	1:A:452:LEU:HD23	1.11	1.07
1:A:88:ARG:HD2	1:A:88:ARG:H	1.24	0.98
1:D:306:LEU:HB2	1:D:335:VAL:HG23	1.46	0.98
1:C:306:LEU:HB2	1:C:335:VAL:HG23	1.43	0.95

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	391/474 (82%)	284 (73%)	66 (17%)	41 (10%)	0	3
1	B	391/474 (82%)	284 (73%)	64 (16%)	43 (11%)	0	3
1	C	391/474 (82%)	288 (74%)	72 (18%)	31 (8%)	1	6
1	D	391/474 (82%)	288 (74%)	73 (19%)	30 (8%)	1	7
All	All	1564/1896 (82%)	1144 (73%)	275 (18%)	145 (9%)	1	4

5 of 145 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	121	PHE
1	A	199	LEU
1	A	202	LYS
1	A	282	GLU
1	A	309	HIS

### 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	356/419 (85%)	331 (93%)	25 (7%)	18	53
1	B	356/419 (85%)	330 (93%)	26 (7%)	16	51
1	C	356/419 (85%)	325 (91%)	31 (9%)	12	41
1	D	356/419 (85%)	325 (91%)	31 (9%)	12	41
All	All	1424/1676 (85%)	1311 (92%)	113 (8%)	14	47

5 of 113 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	67	GLU
1	C	241	TRP
1	D	386	ASP
1	C	70	LEU
1	C	122	ARG

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 48 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	192	HIS
1	B	313	HIS
1	D	292	ASN
1	B	201	ASN
1	B	258	HIS

### 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 [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	397/474 (83%)	0.19	9 (2%) 61 45	23, 73, 155, 193	0
1	B	397/474 (83%)	0.20	9 (2%) 61 45	26, 74, 161, 194	0
1	C	397/474 (83%)	0.28	24 (6%) 23 12	28, 80, 163, 202	0
1	D	397/474 (83%)	0.25	22 (5%) 26 14	26, 78, 169, 202	0
All	All	1588/1896 (83%)	0.23	64 (4%) 39 24	23, 77, 162, 202	0

The worst 5 of 64 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	161	GLU	5.3
1	B	406	LEU	4.2
1	D	443	VAL	4.1
1	C	439	PHE	4.0
1	C	281	GLU	3.8

### 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.