



# Full wwPDB X-ray Structure Validation Report ⓘ

Jun 16, 2024 – 10:12 PM EDT

PDB ID : 5NW9  
Title : Crystal structure of the complex of Tdp1 with duplex DNA  
Authors : Richardson, J.M.; Ruksenaite, E.; Morris, E.R.  
Deposited on : 2017-05-05  
Resolution : 2.04 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

<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	:	2.37.1
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)	:	2.37.1

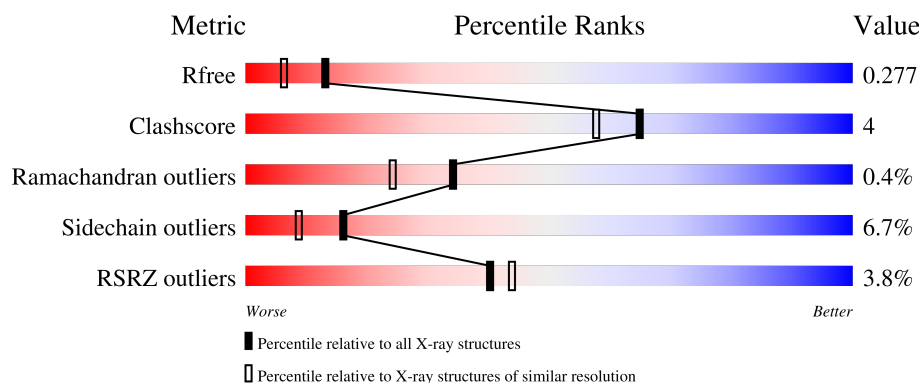
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.04 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	1692 (2.04-2.04)
Clashscore	141614	1773 (2.04-2.04)
Ramachandran outliers	138981	1752 (2.04-2.04)
Sidechain outliers	138945	1752 (2.04-2.04)
RSRZ outliers	127900	1672 (2.04-2.04)

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	485	 2% 78% 11% 11%
1	B	485	 5% 76% 12% 11%
2	C	9	 67% 33%

## 2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 7120 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 Tyrosyl-DNA phosphodiesterase 1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	431	Total	C	N	O	S	0	0	0
			3447	2235	583	618	11			
1	B	431	Total	C	N	O	S	0	0	0
			3447	2235	583	618	11			

There are 56 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	124	MET	-	initiating methionine	UNP Q9NUW8
A	125	GLY	-	expression tag	UNP Q9NUW8
A	126	SER	-	expression tag	UNP Q9NUW8
A	127	SER	-	expression tag	UNP Q9NUW8
A	128	HIS	-	expression tag	UNP Q9NUW8
A	129	HIS	-	expression tag	UNP Q9NUW8
A	130	HIS	-	expression tag	UNP Q9NUW8
A	131	HIS	-	expression tag	UNP Q9NUW8
A	132	HIS	-	expression tag	UNP Q9NUW8
A	133	HIS	-	expression tag	UNP Q9NUW8
A	134	SER	-	expression tag	UNP Q9NUW8
A	135	SER	-	expression tag	UNP Q9NUW8
A	136	GLY	-	expression tag	UNP Q9NUW8
A	137	LEU	-	expression tag	UNP Q9NUW8
A	138	VAL	-	expression tag	UNP Q9NUW8
A	139	PRO	-	expression tag	UNP Q9NUW8
A	140	ARG	-	expression tag	UNP Q9NUW8
A	141	GLY	-	expression tag	UNP Q9NUW8
A	142	SER	-	expression tag	UNP Q9NUW8
A	143	HIS	-	expression tag	UNP Q9NUW8
A	144	MET	-	expression tag	UNP Q9NUW8
A	145	LEU	-	expression tag	UNP Q9NUW8
A	146	GLU	-	expression tag	UNP Q9NUW8
A	147	ASP	-	expression tag	UNP Q9NUW8
A	148	PRO	-	expression tag	UNP Q9NUW8

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
A	322	ASN	ASP	conflict	UNP Q9NUW8
A	328	THR	MET	conflict	UNP Q9NUW8
A	548	LEU	PHE	conflict	UNP Q9NUW8
B	124	MET	-	initiating methionine	UNP Q9NUW8
B	125	GLY	-	expression tag	UNP Q9NUW8
B	126	SER	-	expression tag	UNP Q9NUW8
B	127	SER	-	expression tag	UNP Q9NUW8
B	128	HIS	-	expression tag	UNP Q9NUW8
B	129	HIS	-	expression tag	UNP Q9NUW8
B	130	HIS	-	expression tag	UNP Q9NUW8
B	131	HIS	-	expression tag	UNP Q9NUW8
B	132	HIS	-	expression tag	UNP Q9NUW8
B	133	HIS	-	expression tag	UNP Q9NUW8
B	134	SER	-	expression tag	UNP Q9NUW8
B	135	SER	-	expression tag	UNP Q9NUW8
B	136	GLY	-	expression tag	UNP Q9NUW8
B	137	LEU	-	expression tag	UNP Q9NUW8
B	138	VAL	-	expression tag	UNP Q9NUW8
B	139	PRO	-	expression tag	UNP Q9NUW8
B	140	ARG	-	expression tag	UNP Q9NUW8
B	141	GLY	-	expression tag	UNP Q9NUW8
B	142	SER	-	expression tag	UNP Q9NUW8
B	143	HIS	-	expression tag	UNP Q9NUW8
B	144	MET	-	expression tag	UNP Q9NUW8
B	145	LEU	-	expression tag	UNP Q9NUW8
B	146	GLU	-	expression tag	UNP Q9NUW8
B	147	ASP	-	expression tag	UNP Q9NUW8
B	148	PRO	-	expression tag	UNP Q9NUW8
B	322	ASN	ASP	conflict	UNP Q9NUW8
B	328	THR	MET	conflict	UNP Q9NUW8
B	548	LEU	PHE	conflict	UNP Q9NUW8

- Molecule 2 is a DNA chain called DNA (5'-D(P\*TP\*GP\*CP\*GP\*CP\*AP\*GP\*TP\*A)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	C	9	Total	C	N	O	P	0	0	0
			169	78	30	52	9			

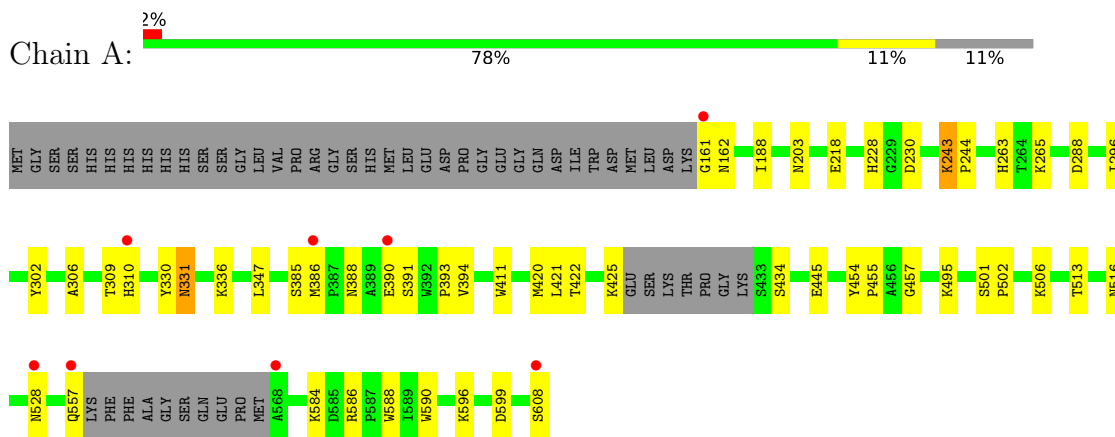
- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	38	Total 38	O 38	0	0
3	B	18	Total 18	O 18	0	0
3	C	1	Total 1	O 1	0	0

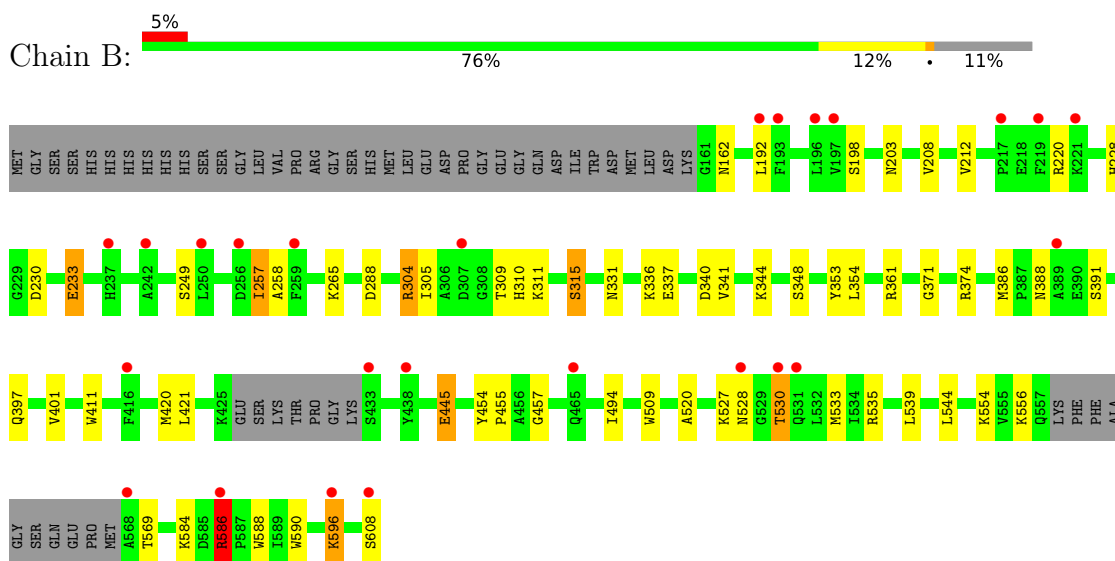
### 3 Residue-property plots [i](#)

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: Tyrosyl-DNA phosphodiesterase 1



- Molecule 1: Tyrosyl-DNA phosphodiesterase 1



- Molecule 2: DNA (5'-D(P\*TP\*GP\*CP\*GP\*CP\*AP\*GP\*TP\*A)-3')



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	108.01Å 195.31Å 50.83Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	52.05 – 2.04 52.05 – 2.04	Depositor EDS
% Data completeness (in resolution range)	99.2 (52.05-2.04) 99.2 (52.05-2.04)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.77 (at 2.05Å)	Xtriage
Refinement program	REFMAC 5.8.0107	Depositor
R, $R_{free}$	0.232 , 0.271 0.236 , 0.277	Depositor DCC
$R_{free}$ test set	3336 reflections (4.84%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	46.4	Xtriage
Anisotropy	0.217	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.32 , 40.3	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	0.95	EDS
Total number of atoms	7120	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	56.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.28% 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

### 5.1 Standard geometry

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	0.56	0/3559	0.64	0/4837
1	B	0.50	0/3559	0.60	1/4837 (0.0%)
2	C	0.54	0/188	0.89	0/289
All	All	0.53	0/7306	0.63	1/9963 (0.0%)

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand. A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
1	A	0	1

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	586	ARG	NE-CZ-NH1	5.79	123.19	120.30

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	161	GLY	Peptide

### 5.2 Too-close contacts

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	3447	0	3393	27	0
1	B	3447	0	3393	28	0
2	C	169	0	90	3	0
3	A	38	0	0	0	0
3	B	18	0	0	0	0
3	C	1	0	0	0	0
All	All	7120	0	6876	57	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 4.

All (57) 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:586:ARG:HH11	1:B:586:ARG:HG2	1.37	0.89
2:C:-8:DT:H2''	2:C:-7:DG:C8	2.22	0.75
1:A:495:LYS:HE2	1:A:516:ASN:OD1	1.86	0.74
1:B:586:ARG:HH11	1:B:586:ARG:CG	2.03	0.71
1:B:340:ASP:O	1:B:344:LYS:HD3	1.98	0.63
1:A:388:ASN:OD1	1:A:391:SER:HB2	1.99	0.62
1:A:228:HIS:HD2	1:A:230:ASP:H	1.48	0.61
1:B:337:GLU:O	1:B:341:VAL:HG23	2.01	0.61
1:B:233:GLU:H	1:B:233:GLU:CD	2.05	0.60
1:A:306:ALA:O	1:A:309:THR:HG22	2.04	0.57
1:B:527:LYS:O	1:B:530:THR:HB	2.05	0.56
1:B:586:ARG:HG2	1:B:586:ARG:NH1	2.15	0.56
1:B:445:GLU:HA	1:B:445:GLU:OE2	2.06	0.56
1:B:208:VAL:O	1:B:212:VAL:HG23	2.07	0.55
1:A:306:ALA:HB3	1:A:309:THR:HG22	1.88	0.54
1:A:421:LEU:HD12	1:A:421:LEU:C	2.30	0.53
1:A:243:LYS:HB3	1:A:244:PRO:HD3	1.90	0.52
1:B:554:LYS:HB2	1:B:569:THR:HG22	1.90	0.52
2:C:-8:DT:C2'	2:C:-7:DG:C8	2.93	0.52
1:A:457:GLY:HA3	1:A:588:TRP:CZ2	2.47	0.50
1:B:397:GLN:HE22	1:B:494:ILE:HG23	1.76	0.49
1:B:228:HIS:HD2	1:B:230:ASP:H	1.61	0.49
1:A:188:ILE:HD11	1:A:296:ILE:HD13	1.94	0.49
1:A:390:GLU:OE1	1:A:434:SER:OG	2.32	0.47
1:A:495:LYS:HB2	1:A:513:THR:O	2.16	0.47
1:B:304:ARG:HG3	1:B:305:ILE:N	2.30	0.46
1:A:394:VAL:CG1	1:A:420:MET:HE2	2.45	0.46
1:B:509:TRP:HB3	1:B:544:LEU:HD23	1.96	0.46

*Continued on next page...*

*Continued from previous page...*

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:394:VAL:HG11	1:A:420:MET:HE2	1.97	0.46
1:A:306:ALA:HB3	1:A:309:THR:CG2	2.46	0.46
1:B:228:HIS:HD2	1:B:230:ASP:N	2.13	0.46
1:A:421:LEU:HD12	1:A:422:THR:N	2.30	0.45
1:B:257:ILE:HG13	1:B:258:ALA:N	2.31	0.45
1:A:228:HIS:HD2	1:A:230:ASP:N	2.12	0.45
1:B:457:GLY:HA3	1:B:588:TRP:CZ2	2.52	0.44
1:B:586:ARG:CG	1:B:586:ARG:NH1	2.70	0.44
1:A:586:ARG:CZ	1:A:586:ARG:HB2	2.46	0.43
1:B:315:SER:HB2	1:B:348:SER:O	2.19	0.43
1:B:354:LEU:HD11	1:B:539:LEU:HD11	2.00	0.43
1:A:391:SER:O	1:A:393:PRO:HD3	2.19	0.42
1:A:336:LYS:HA	1:A:336:LYS:HD2	1.81	0.42
1:A:495:LYS:HE2	1:A:516:ASN:CG	2.40	0.42
1:B:454:TYR:N	1:B:455:PRO:CD	2.82	0.42
1:A:306:ALA:O	1:A:309:THR:CG2	2.67	0.42
1:A:330:TYR:O	1:A:331:ASN:CB	2.68	0.42
1:B:371:GLY:O	1:B:374:ARG:HB3	2.20	0.41
1:B:353:TYR:CE1	1:B:544:LEU:HD12	2.55	0.41
1:A:302:TYR:CD1	1:A:347:LEU:HA	2.55	0.41
1:A:454:TYR:CD2	1:A:599:ASP:HB3	2.55	0.41
1:A:501:SER:HB2	1:A:502:PRO:CD	2.51	0.41
1:B:421:LEU:C	1:B:421:LEU:HD12	2.41	0.41
1:B:520:ALA:HB2	1:B:535:ARG:HD2	2.03	0.41
1:A:263:HIS:NE2	2:C:0:DA:P	2.94	0.40
1:B:257:ILE:CD1	1:B:533:MET:SD	3.09	0.40
1:B:596:LYS:HA	1:B:596:LYS:HE2	2.03	0.40
1:B:586:ARG:NH1	1:B:586:ARG:HB3	2.36	0.40
1:A:454:TYR:N	1:A:455:PRO:CD	2.84	0.40

There are no symmetry-related clashes.

## 5.3 Torsion angles ⓘ

### 5.3.1 Protein backbone ⓘ

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	425/485 (88%)	411 (97%)	13 (3%)	1 (0%)	47	39
1	B	425/485 (88%)	407 (96%)	16 (4%)	2 (0%)	29	18
All	All	850/970 (88%)	818 (96%)	29 (3%)	3 (0%)	34	24

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	528	ASN
1	A	411	TRP
1	B	411	TRP

### 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	375/421 (89%)	356 (95%)	19 (5%)	24	15
1	B	375/421 (89%)	344 (92%)	31 (8%)	11	5
All	All	750/842 (89%)	700 (93%)	50 (7%)	16	9

All (50) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	162	ASN
1	A	203	ASN
1	A	218	GLU
1	A	243	LYS
1	A	265	LYS
1	A	288	ASP
1	A	310	HIS
1	A	331	ASN
1	A	385	SER
1	A	386	MET
1	A	425	LYS
1	A	445	GLU
1	A	506	LYS

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	A	528	ASN
1	A	557	GLN
1	A	584	LYS
1	A	590	TRP
1	A	596	LYS
1	A	608	SER
1	B	162	ASN
1	B	192	LEU
1	B	198	SER
1	B	203	ASN
1	B	220	ARG
1	B	233	GLU
1	B	249	SER
1	B	257	ILE
1	B	265	LYS
1	B	288	ASP
1	B	304	ARG
1	B	309	THR
1	B	310	HIS
1	B	311	LYS
1	B	315	SER
1	B	331	ASN
1	B	336	LYS
1	B	361	ARG
1	B	386	MET
1	B	388	ASN
1	B	391	SER
1	B	401	VAL
1	B	420	MET
1	B	445	GLU
1	B	530	THR
1	B	556	LYS
1	B	584	LYS
1	B	586	ARG
1	B	590	TRP
1	B	596	LYS
1	B	608	SER

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (7) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	228	HIS

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	A	397	GLN
1	A	528	ASN
1	B	228	HIS
1	B	263	HIS
1	B	388	ASN
1	B	397	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 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	431/485 (88%)	0.13	8 (1%) 66 71	31, 48, 73, 116	0
1	B	431/485 (88%)	0.43	25 (5%) 23 25	38, 59, 93, 115	0
2	C	9/9 (100%)	-0.31	0 100 100	52, 63, 69, 95	0
All	All	871/979 (88%)	0.28	33 (3%) 40 44	31, 53, 89, 116	0

All (33) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	608	SER	5.2
1	B	193	PHE	4.0
1	B	221	LYS	4.0
1	B	192	LEU	3.7
1	A	528	ASN	3.6
1	A	608	SER	3.4
1	B	197	VAL	3.4
1	B	219	PHE	3.2
1	B	307	ASP	3.1
1	A	310	HIS	3.0
1	B	530	THR	2.9
1	B	531	GLN	2.9
1	B	528	ASN	2.9
1	B	217	PRO	2.7
1	B	196	LEU	2.6
1	B	465	GLN	2.5
1	B	259	PHE	2.5
1	B	416	PHE	2.5
1	B	250	LEU	2.5
1	B	596	LYS	2.5
1	A	568	ALA	2.4
1	B	242	ALA	2.4
1	B	389	ALA	2.4

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
1	B	438	TYR	2.4
1	B	568	ALA	2.4
1	B	586	ARG	2.4
1	B	256	ASP	2.3
1	A	386	MET	2.2
1	A	161	GLY	2.2
1	A	390	GLU	2.1
1	B	433	SER	2.1
1	B	237	HIS	2.1
1	A	557	GLN	2.1

## 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.