



# wwPDB X-ray Structure Validation Summary Report ⓘ

Aug 22, 2017 – 05:08 PM EDT

PDB ID : 5D39  
Title : Transcription factor-DNA complex  
Authors : Li, J.; Niu, F.; Ouyang, S.; Liu, Z.  
Deposited on : unknown  
Resolution : 3.20 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
Mogul	:	1.7.2 (RC1), CSD as538be (2017)
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20029824
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-20029824

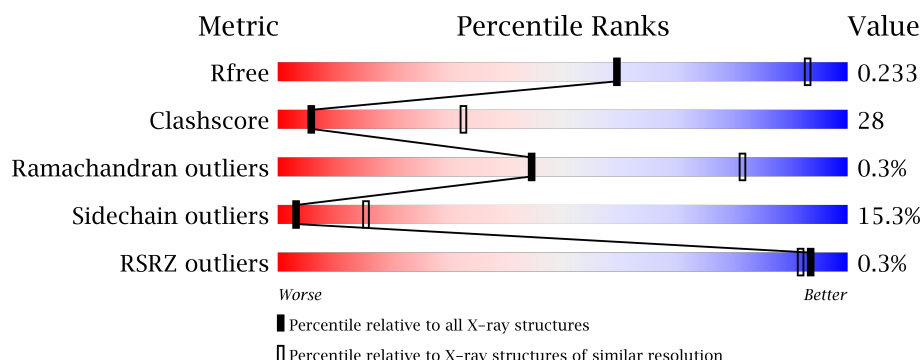
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 3.20 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.




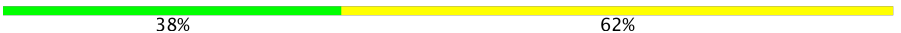

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
$R_{free}$	100719	1015 (3.22-3.18)
Clashscore	112137	1009 (3.20-3.20)
Ramachandran outliers	110173	1118 (3.22-3.18)
Sidechain outliers	110143	1117 (3.22-3.18)
RSRZ outliers	101464	1020 (3.22-3.18)

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	539	<div> <div>51%</div> <div>33%</div> <div>6%</div> <div>9%</div> </div>
1	B	539	<div> <div>52%</div> <div>32%</div> <div>6%</div> <div>10%</div> </div>
1	C	539	<div> <div>52%</div> <div>30%</div> <div>8%</div> <div>10%</div> </div>
1	D	539	<div> <div>52%</div> <div>29%</div> <div>9%</div> <div>10%</div> </div>
2	F	21	<div> <div>62%</div> <div>29%</div> <div>10%</div> </div>

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
2	N	21	 62%29%10%
3	E	21	 38%62%
3	M	21	 52%38%10%

## 2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 17397 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 Signal transducer and activator of transcription 6.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	489	Total	C	N	O	P	S	0	0	0
			3891	2478	683	713	1	16			
1	B	484	Total	C	N	O	P	S	0	0	0
			3850	2451	674	708	1	16			
1	C	484	Total	C	N	O	P	S	0	0	0
			3851	2454	673	707	1	16			
1	D	487	Total	C	N	O	P	S	0	0	0
			3877	2468	678	714	1	16			

There are 12 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	120	SER	-	expression tag	UNP P42226
A	121	ASN	-	expression tag	UNP P42226
A	122	ALA	-	expression tag	UNP P42226
B	120	SER	-	expression tag	UNP P42226
B	121	ASN	-	expression tag	UNP P42226
B	122	ALA	-	expression tag	UNP P42226
C	120	SER	-	expression tag	UNP P42226
C	121	ASN	-	expression tag	UNP P42226
C	122	ALA	-	expression tag	UNP P42226
D	120	SER	-	expression tag	UNP P42226
D	121	ASN	-	expression tag	UNP P42226
D	122	ALA	-	expression tag	UNP P42226

- Molecule 2 is a DNA chain called DNA (5'-D(P\*TP\*CP\*TP\*GP\*TP\*CP\*TP\*TP\*CP\*CP\*AP\*GP\*GP\*AP\*AP\*AP\*TP\*CP\*CP\*AP\*T)-3').

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
2	N	21	Total	C	N	O	P		0	0	0
			425	204	72	128	21				

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	F	21	Total	C	N	O	P	0	0	0
			425	204	72	128	21			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	M	21	Total	C	N	O	P	0	0	0
			436	207	84	124	21			
3	E	21	Total	C	N	O	P	0	0	0
			436	207	84	124	21			

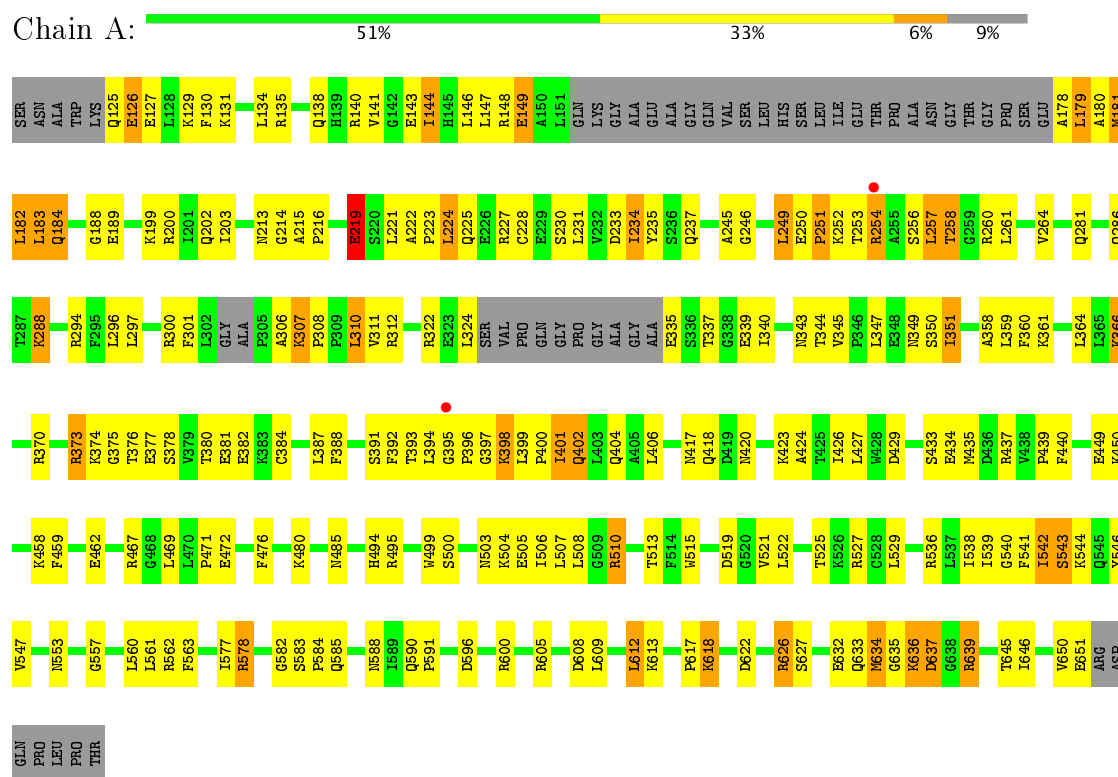
- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	61	Total	O	0	0
			61	61		
4	B	55	Total	O	0	0
			55	55		
4	C	50	Total	O	0	0
			50	50		
4	D	31	Total	O	0	0
			31	31		
4	N	4	Total	O	0	0
			4	4		
4	M	2	Total	O	0	0
			2	2		
4	E	1	Total	O	0	0
			1	1		
4	F	2	Total	O	0	0
			2	2		

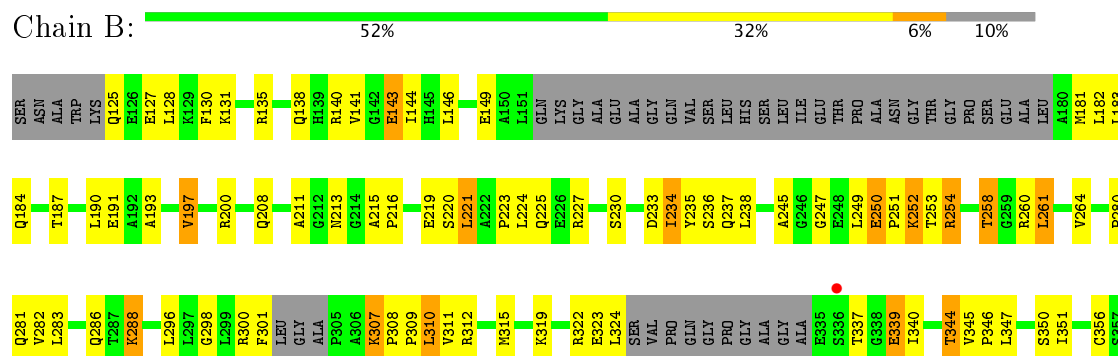
### 3 Residue-property plots

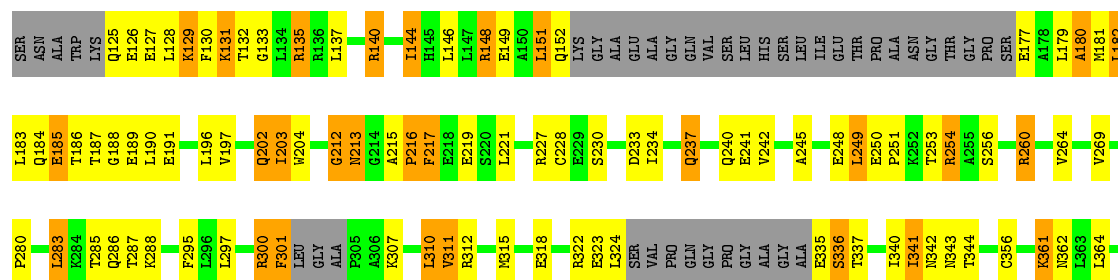
These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are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

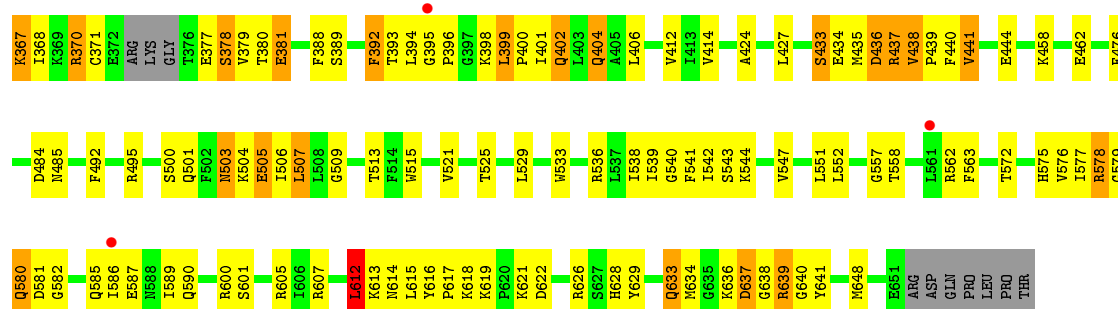
- Molecule 1: Signal transducer and activator of transcription 6



- Molecule 1: Signal transducer and activator of transcription 6

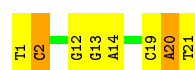






- Molecule 2: DNA (5'-D(P\*TP\*CP\*TP\*GP\*TP\*CP\*TP\*TP\*CP\*CP\*AP\*GP\*GP\*AP\*AP\*A P\*TP\*CP\*CP\*AP\*T)-3')

Chain N: 62% 29% 10%



- Molecule 2: DNA (5'-D(P\*TP\*CP\*TP\*GP\*TP\*CP\*TP\*TP\*CP\*CP\*AP\*GP\*GP\*AP\*AP\*A P\*TP\*CP\*CP\*AP\*T)-3')

Chain F: 62% 29% 10%



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

Chain M: 52% 38% 10%



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

Chain E: 38% 62%





## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	68.39Å 94.70Å 145.64Å 79.62° 78.31° 89.58°	Depositor
Resolution (Å)	44.12 – 3.20 44.12 – 3.20	Depositor EDS
% Data completeness (in resolution range)	97.1 (44.12-3.20) 89.4 (44.12-3.20)	Depositor EDS
$R_{merge}$	0.08	Depositor
$R_{sym}$	0.08	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.90 (at 3.19Å)	Xtriage
Refinement program	PHENIX 1.9_1692, REFMAC	Depositor
R, $R_{free}$	0.210 , 0.236 0.228 , 0.233	Depositor DCC
$R_{free}$ test set	2622 reflections (4.84%)	DCC
Wilson B-factor (Å <sup>2</sup> )	92.8	Xtriage
Anisotropy	0.304	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.29 , 68.9	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.47$ , $\langle L^2 \rangle = 0.30$	Xtriage
Estimated twinning fraction	0.095 for h,-k,h-l	Xtriage
$F_o, F_c$ correlation	0.94	EDS
Total number of atoms	17397	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	118.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.27% 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 ⓘ

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

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	0.33	1/3944 (0.0%)	0.61	0/5316
1	B	0.33	1/3902 (0.0%)	0.61	2/5259 (0.0%)
1	C	0.33	0/3902	0.63	3/5258 (0.1%)
1	D	0.32	0/3929	0.62	3/5296 (0.1%)
2	F	0.80	0/474	1.18	2/728 (0.3%)
2	N	0.97	2/474 (0.4%)	1.28	4/728 (0.5%)
3	E	0.68	0/490	1.05	0/755
3	M	0.71	0/490	1.18	3/755 (0.4%)
All	All	0.41	4/17605 (0.0%)	0.71	17/24095 (0.1%)

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
1	B	0	1
1	D	0	1
All	All	0	3

All (4) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	N	1	DT	C1'-N1	6.80	1.58	1.49
2	N	1	DT	C2-O2	-6.56	1.17	1.22
1	B	216	PRO	N-CD	5.22	1.55	1.47
1	A	251	PRO	N-CD	5.11	1.55	1.47

The worst 5 of 17 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed( $^{\circ}$ )	Ideal( $^{\circ}$ )
2	N	20	DA	O4'-C1'-N9	11.72	116.20	108.00
1	B	508	LEU	CA-CB-CG	9.04	136.10	115.30
3	M	1	DA	O4'-C1'-N9	8.43	113.90	108.00
1	C	151	LEU	CA-CB-CG	7.03	131.46	115.30
2	N	20	DA	C1'-O4'-C4'	-6.41	103.69	110.10

There are no chirality outliers.

All (3) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	219	GLU	Peptide
1	B	245	ALA	Peptide
1	D	180	ALA	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	3891	0	3973	241	0
1	B	3850	0	3918	182	0
1	C	3851	0	3925	246	0
1	D	3877	0	3946	241	0
2	F	425	0	239	16	0
2	N	425	0	239	13	0
3	E	436	0	237	12	0
3	M	436	0	237	13	0
4	A	61	0	0	47	0
4	B	55	0	0	36	0
4	C	50	0	0	78	0
4	D	31	0	0	32	0
4	E	1	0	0	1	0
4	F	2	0	0	4	0
4	M	2	0	0	3	0
4	N	4	0	0	0	0
All	All	17397	0	16714	943	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 943 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:D:179:LEU:HD23	1:D:182:LEU:CB	1.44	1.46
1:D:179:LEU:CD2	1:D:182:LEU:HD22	1.46	1.43
1:D:179:LEU:HD21	1:D:182:LEU:CD2	1.50	1.40
1:A:246:GLY:HA2	1:A:249:LEU:CD1	1.50	1.39
1:A:245:ALA:O	1:A:249:LEU:CD2	1.69	1.39

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	480/539 (89%)	461 (96%)	17 (4%)	2 (0%)	38	77
1	B	473/539 (88%)	455 (96%)	16 (3%)	2 (0%)	38	77
1	C	471/539 (87%)	458 (97%)	12 (2%)	1 (0%)	51	86
1	D	476/539 (88%)	462 (97%)	13 (3%)	1 (0%)	51	86
All	All	1900/2156 (88%)	1836 (97%)	58 (3%)	6 (0%)	44	81

5 of 6 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	637	ASP
1	A	377	GLU
1	A	396	PRO
1	B	149	GLU
1	D	216	PRO

### 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	424/459 (92%)	360 (85%)	64 (15%)	3	16
1	B	420/459 (92%)	363 (86%)	57 (14%)	4	20
1	C	420/459 (92%)	355 (84%)	65 (16%)	3	14
1	D	423/459 (92%)	351 (83%)	72 (17%)	2	11
All	All	1687/1836 (92%)	1429 (85%)	258 (15%)	3	15

5 of 258 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	632	GLU
1	C	300	ARG
1	D	433	SER
1	B	646	ILE
1	C	196	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (5) such sidechains are listed below:

Mol	Chain	Res	Type
1	B	138	GLN
1	B	237	GLN
1	C	184	GLN
1	C	240	GLN
1	D	628	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](#)

4 non-standard protein/DNA/RNA residues are modelled in this entry.

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	$\# Z  > 2$
1	PTR	A	641	1	15,16,17	1.28	2 (13%)	19,22,24	0.67	0
1	PTR	B	641	1	15,16,17	1.34	2 (13%)	19,22,24	0.74	0
1	PTR	C	641	1	15,16,17	1.32	1 (6%)	19,22,24	0.71	0
1	PTR	D	641	1	15,16,17	1.38	2 (13%)	19,22,24	0.66	0

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
1	PTR	A	641	1	-	0/9/11/13	0/1/1/1
1	PTR	B	641	1	-	0/9/11/13	0/1/1/1
1	PTR	C	641	1	-	0/9/11/13	0/1/1/1
1	PTR	D	641	1	-	0/9/11/13	0/1/1/1

The worst 5 of 7 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	D	641	PTR	OH-CZ	-4.25	1.30	1.40
1	C	641	PTR	OH-CZ	-4.23	1.30	1.40
1	A	641	PTR	OH-CZ	-4.16	1.31	1.40
1	B	641	PTR	OH-CZ	-4.15	1.31	1.40
1	B	641	PTR	CA-C	2.21	1.53	1.50

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

3 monomers are involved in 3 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
1	B	641	PTR	1	0
1	C	641	PTR	1	0
1	D	641	PTR	1	0

## 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 ⓘ

### 6.1 Protein, DNA and RNA chains ⓘ

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	488/539 (90%)	-0.02	2 (0%) 92 89	56, 96, 164, 227	0
1	B	483/539 (89%)	-0.02	1 (0%) 94 93	54, 98, 164, 212	0
1	C	483/539 (89%)	-0.09	0 100 100	83, 125, 179, 217	0
1	D	486/539 (90%)	-0.09	3 (0%) 89 83	78, 122, 183, 219	0
2	F	21/21 (100%)	-0.61	0 100 100	90, 133, 214, 249	0
2	N	21/21 (100%)	-0.76	0 100 100	93, 133, 203, 228	0
3	E	21/21 (100%)	-0.74	0 100 100	111, 128, 201, 227	0
3	M	21/21 (100%)	-0.68	0 100 100	107, 124, 202, 222	0
All	All	2024/2240 (90%)	-0.08	6 (0%) 93 92	54, 114, 178, 249	0

The worst 5 of 6 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	D	395	GLY	3.4
1	D	586	ILE	3.0
1	A	395	GLY	2.4
1	B	336	SER	2.2
1	A	254	ARG	2.2

### 6.2 Non-standard residues in protein, DNA, RNA chains ⓘ

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. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. 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	LLDF	B-factors( $\text{\AA}^2$ )	Q<0.9
1	PTR	A	641	16/17	0.94	0.18	-	103,120,156,159	0
1	PTR	B	641	16/17	0.94	0.15	-	95,112,142,151	0
1	PTR	C	641	16/17	0.97	0.21	-	62,86,109,118	0
1	PTR	D	641	16/17	0.97	0.21	-	61,93,117,118	0

### 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.