



wwPDB X-ray Structure Validation Summary Report ⓘ

Nov 21, 2016 – 05:16 PM EST

PDB ID : 4Y5U
Title : Transcription factor
Authors : Li, J.; Niu, F.; Ouyang, S.; Liu, Z.
Deposited on : 2015-02-12
Resolution : 2.71 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

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 : 1.7.1 (RC1), CSD as537be (2016)
Xtriage (Phenix) : 1.9-1692
EDS : rb-20028320
Percentile statistics : 20151230.v01 (using entries in the PDB archive December 30th 2015)
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-20028320

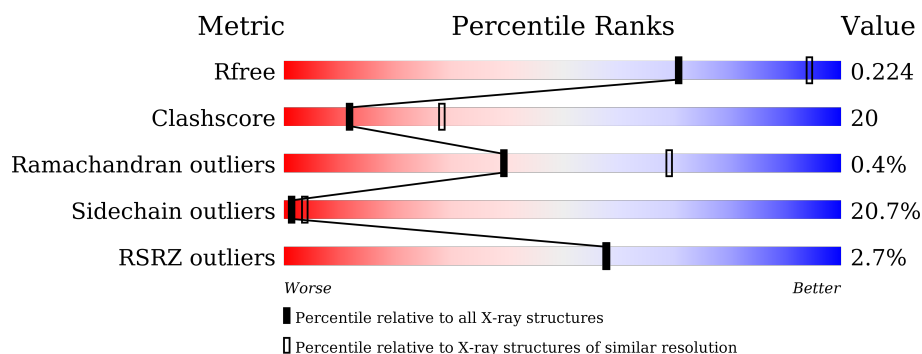
1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.71 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}	91344	2103 (2.70-2.70)
Clashscore	102246	2422 (2.70-2.70)
Ramachandran outliers	100387	2382 (2.70-2.70)
Sidechain outliers	100360	2382 (2.70-2.70)
RSRZ outliers	91569	2107 (2.70-2.70)

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 ≥ 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	549	<div> <div>2%</div> <div> <div></div> <div>45%</div> <div>26%</div> <div>9%</div> <div>•</div> <div>19%</div> </div> </div>
1	B	549	<div> <div>2%</div> <div> <div></div> <div>42%</div> <div>28%</div> <div>9%</div> <div>•</div> <div>20%</div> </div> </div>

2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 7227 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	445	Total	C	N	O	P	S	0	0	0
			3550	2268	622	645	1	14			
1	B	438	Total	C	N	O	P	S	0	0	0
			3500	2240	611	634	1	14			

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	110	SER	-	expression tag	UNP P42226
A	111	ASN	-	expression tag	UNP P42226
A	112	ALA	-	expression tag	UNP P42226
B	110	SER	-	expression tag	UNP P42226
B	111	ASN	-	expression tag	UNP P42226
B	112	ALA	-	expression tag	UNP P42226

- Molecule 2 is NICKEL (II) ION (three-letter code: NI) (formula: Ni).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	1	Total	Ni	0	0
			1	1		

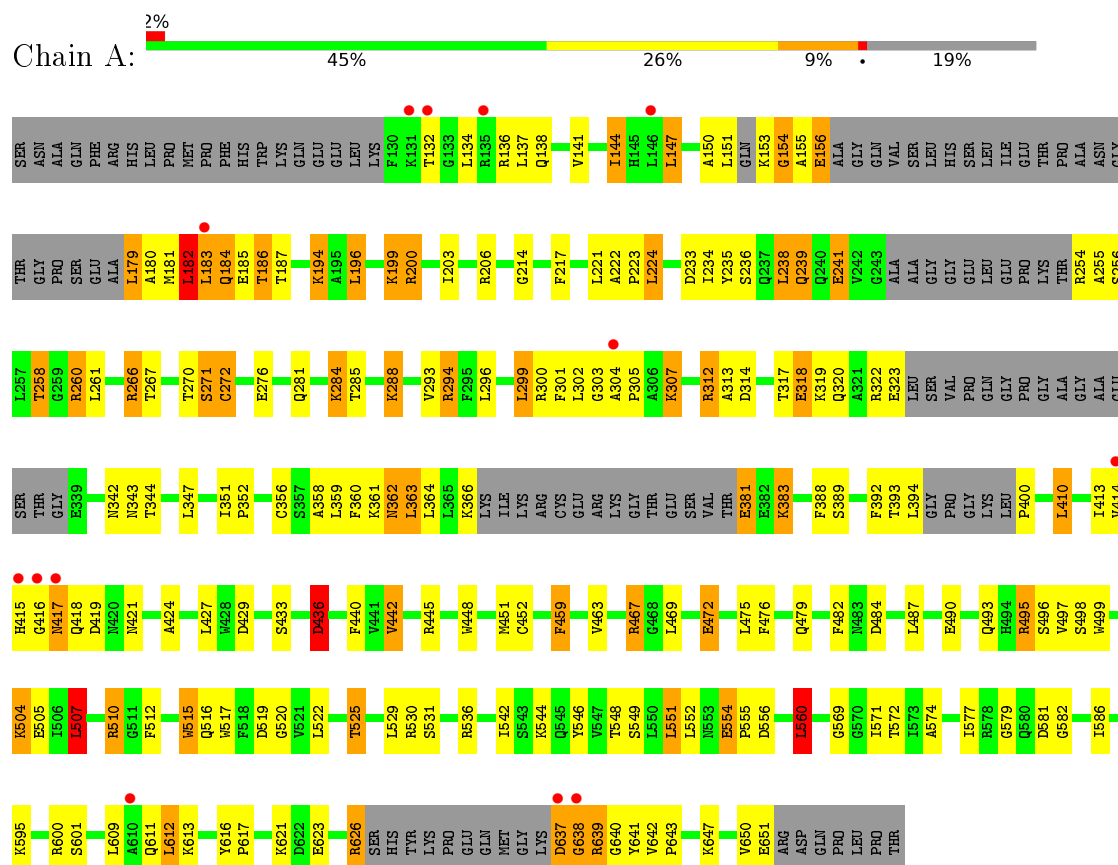
- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	98	Total	O	0	0
			98	98		
3	B	78	Total	O	0	0
			78	78		

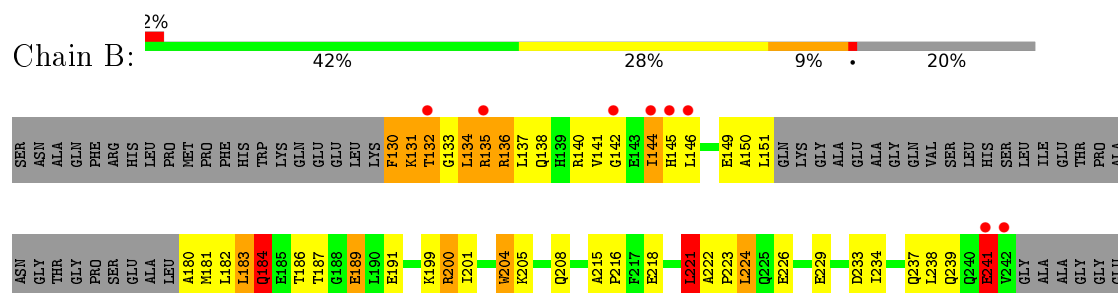
3 Residue-property plots

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors 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



LEU	PRO	GLY	GLN	PRO	GLY	PRO	LYS	THR	ARG	A255	S256	L257	R260	L261	D262	E263	R266	S271	Q278	P279	P280	Q281	V282	L283	K284	T285	Q286	T287	K288	F289	V293	R294	L297	G298	L299	L302	G303	A304	P305	A306	K307	L310	V311	R312	V316	K319	Q320	A321	R322	E323	LEU	SER	VAL				
PRO	GLN	GLY	PRO	LYS	GLY	ALA	GLY	ALA	GLU	SER	THR	GLY	E339	I340	I341	N342	N343	T344	V345	S350	I351	P352	G353	A358	L359	F360	K361	N362	L363	L364	L365	K366	LYS	ILE	LYS	ARG	CYS	GLU	ARG	LYS	GLY	THR	GLU	SER	VAL	THR	E381	E382	K383	C384	A385	S389	A390	S391	F392	T393	L394
GLY	PRO	GLY	LYS	L399	P400	I401	Q402	L406	L410	V414	H415	G416	N417	Q418	D419	K423	A424	T425	I426	L427	W428	D429	F432	S433	R437	V438	P439	F440	V441	V442	A443	E444	R445	V446	P447	W448	C452	E453	T454	R467	G468	L469	L470	P471	E472	S486	E490	Q493									
H494	R495	S496	V497	S498	W499	S500	K504	E505	I506	L507	L508	R509	R510	G511	W515	Q516	W517	G520	V521	L522	D523	L524	T525	L529	R530	S531	Y532	R536	I542	S543	K544	Q545	W546	V547	L550	L551	L552	R553	E554	P555	D556	G557	T558	P559	L560	L561	R562	S566	E567	I571							
R578	I586	Q590	K595	S601	R607	D608	L609	A610	Q611	L612	K613	N614	L615	Y616	P617	K618	K619	P620	K621	D622	E623	R626	SER	HIS	TYR	LYS	PRO	GLU	GLN	MET	GLY	LYS	ASP	G638	R639	V642	T645	I646	K647	V650	E651	ARG	ASP	GLN	PRO	LEU	PRO	THR									

4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α , β , γ	66.15Å 94.32Å 179.36Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	46.36 – 2.71 46.36 – 2.71	Depositor EDS
% Data completeness (in resolution range)	90.8 (46.36-2.71) 89.1 (46.36-2.71)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.09	Depositor
$\langle I/\sigma(I) \rangle$ ¹	5.10 (at 2.73Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.8.2_1309)	Depositor
R, R_{free}	0.185 , 0.253 0.196 , 0.224	Depositor DCC
R_{free} test set	1951 reflections (6.99%)	DCC
Wilson B-factor (Å ²)	23.2	Xtriage
Anisotropy	0.090	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.33 , 48.5	EDS
L-test for twinning ²	$\langle L \rangle = 0.43$, $\langle L^2 \rangle = 0.26$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.90	EDS
Total number of atoms	7227	wwPDB-VP
Average B, all atoms (Å ²)	21.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.39% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

²Theoretical 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: NI, 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	1.77	18/3594 (0.5%)	1.07	9/4843 (0.2%)
1	B	1.72	18/3545 (0.5%)	1.06	6/4782 (0.1%)
All	All	1.74	36/7139 (0.5%)	1.06	15/9625 (0.2%)

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	2
1	B	0	2
All	All	0	4

The worst 5 of 36 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	554	GLU	CD-OE1	-6.66	1.18	1.25
1	A	517	TRP	CE3-CZ3	-6.52	1.27	1.38
1	B	498	SER	CB-OG	-6.14	1.34	1.42
1	B	505	GLU	CD-OE1	-6.10	1.19	1.25
1	A	520	GLY	C-O	-6.05	1.14	1.23

The worst 5 of 15 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	182	LEU	CA-CB-CG	7.26	132.00	115.30
1	A	560	LEU	CA-CB-CG	6.82	130.98	115.30
1	A	507	LEU	CA-CB-CG	6.74	130.79	115.30
1	A	294	ARG	NE-CZ-NH1	6.33	123.46	120.30

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	609	LEU	CB-CG-CD2	-5.98	100.84	111.00

There are no chirality outliers.

All (4) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	303	GLY	Peptide
1	A	436	ASP	Peptide
1	B	184	GLN	Peptide
1	B	418	GLN	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	3550	0	3614	145	1
1	B	3500	0	3568	147	0
2	A	1	0	0	0	0
3	A	98	0	0	8	0
3	B	78	0	0	12	0
All	All	7227	0	7182	288	1

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 20.

The worst 5 of 288 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:183:LEU:O	1:B:186:THR:OG1	1.56	1.23
1:B:144:ILE:CG2	1:B:183:LEU:HD12	1.73	1.17
1:A:153:LYS:O	1:A:155:ALA:N	1.86	1.09
1:A:233:ASP:OD1	1:A:300:ARG:NH1	1.85	1.08
1:A:153:LYS:C	1:A:155:ALA:H	1.50	1.02

All (1) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:180:ALA:CB	1:A:344:THR:O[3_544]	1.90	0.30

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	428/549 (78%)	413 (96%)	13 (3%)	2 (0%)	34	63
1	B	423/549 (77%)	416 (98%)	6 (1%)	1 (0%)	52	80
All	All	851/1098 (78%)	829 (97%)	19 (2%)	3 (0%)	39	69

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	154	GLY
1	B	415	HIS
1	A	638	GLY

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	386/469 (82%)	311 (81%)	75 (19%)	2	4
1	B	382/469 (81%)	298 (78%)	84 (22%)	1	3
All	All	768/938 (82%)	609 (79%)	159 (21%)	1	4

5 of 159 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	626	ARG
1	B	200	ARG
1	B	578	ARG
1	A	647	LYS
1	B	136	ARG

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (4) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	342	ASN
1	A	415	HIS
1	A	420	ASN
1	A	611	GLN

5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains ⓘ

2 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	13,16,17	2.73	4 (30%)	19,22,24	1.38	3 (15%)
1	PTR	B	641	1	13,16,17	2.44	3 (23%)	19,22,24	2.03	7 (36%)

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

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

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	641	PTR	P-O2P	-5.59	1.35	1.54
1	A	641	PTR	P-O3P	-5.47	1.35	1.54
1	B	641	PTR	P-O1P	-4.90	1.35	1.50
1	B	641	PTR	P-O2P	-4.87	1.38	1.54
1	B	641	PTR	P-O3P	-4.52	1.39	1.54

The worst 5 of 10 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	641	PTR	O2P-P-OH	-3.74	93.56	105.47
1	B	641	PTR	O-C-CA	-3.06	117.51	125.72
1	A	641	PTR	O-C-CA	-2.37	119.37	125.72
1	B	641	PTR	CE1-CD1-CG	-2.15	118.00	121.02
1	B	641	PTR	P-OH-CZ	2.63	131.32	123.85

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
1	A	641	PTR	1	0

5.5 Carbohydrates ⓘ

There are no carbohydrates in this entry.

5.6 Ligand geometry ⓘ

Of 1 ligands modelled in this entry, 1 is monoatomic - leaving 0 for Mogul analysis.

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 ⓘ

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, 95th 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 ²)	Q<0.9
1	A	444/549 (80%)	-0.24	13 (2%) 55 55	1, 17, 50, 73	0
1	B	437/549 (79%)	-0.21	11 (2%) 61 61	2, 18, 58, 90	0
All	All	881/1098 (80%)	-0.22	24 (2%) 58 58	1, 18, 54, 90	0

The worst 5 of 24 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	146	LEU	4.0
1	B	132	THR	3.9
1	B	304	ALA	3.9
1	B	415	HIS	3.4
1	B	144	ILE	3.2

6.2 Non-standard residues in protein, DNA, RNA chains [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. 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, 95th 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(Å ²)	Q<0.9
1	PTR	A	641	16/17	0.98	0.12	-	6,11,17,24	0
1	PTR	B	641	16/17	0.97	0.11	-	5,10,17,19	0

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. 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, 95th 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(\AA^2)	Q<0.9
2	NI	A	701	1/1	1.00	0.11	-3.89	11,11,11,11	0

6.5 Other polymers [i](#)

There are no such residues in this entry.