



wwPDB X-ray Structure Validation Summary Report ⓘ

Nov 21, 2016 – 04:13 PM EST

PDB ID : 4Y5W
Title : Transcription factor-DNA complex
Authors : Li, J.; Niu, F.; Ouyang, S.; Liu, Z.
Deposited on : 2015-02-12
Resolution : 3.10 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.

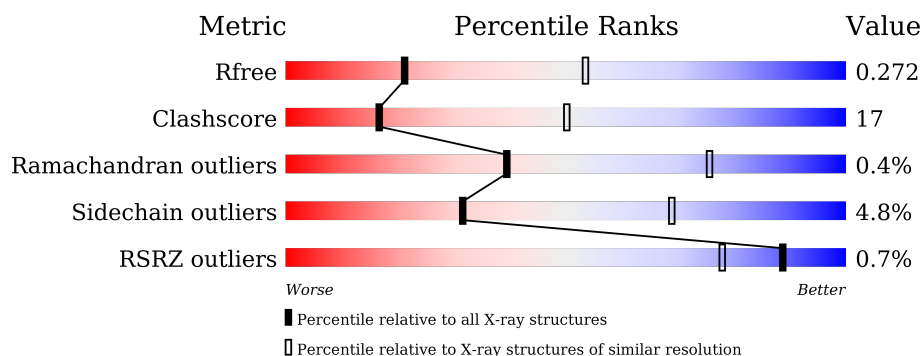
The following versions of software and data (see [references ⓘ](#)) were used in the production of this report:

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

i





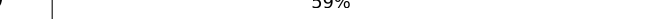

X-RAY DIFFRACTION

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	1114 (3.14-3.06)
Clashscore	102246	1222 (3.14-3.06)
Ramachandran outliers	100387	1174 (3.14-3.06)
Sidechain outliers	100360	1174 (3.14-3.06)
RSRZ outliers	91569	1119 (3.14-3.06)

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	
1	B	549	
1	C	549	
1	D	549	
2	E	22	
2	M	22	

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
3	F	22	<div> <div>5%</div> <div>36%</div> <div>64%</div> </div>
3	N	22	<div> <div>5%</div> <div>36%</div> <div>64%</div> </div>

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
1	PTR	A	641	-	-	X	-

2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 17083 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	478	Total	C	N	O	P	S	0	0	0
			3802	2421	671	693	1	16			
1	B	477	Total	C	N	O	P	S	0	1	0
			3804	2425	671	691	1	16			
1	D	472	Total	C	N	O	P	S	0	0	0
			3745	2389	656	683	1	16			
1	C	465	Total	C	N	O	P	S	0	0	0
			3697	2361	649	670	1	16			

There are 12 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
D	110	SER	-	expression tag	UNP P42226
D	111	ASN	-	expression tag	UNP P42226
D	112	ALA	-	expression tag	UNP P42226
C	110	SER	-	expression tag	UNP P42226
C	111	ASN	-	expression tag	UNP P42226
C	112	ALA	-	expression tag	UNP P42226

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

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
2	E	22	Total	C	N	O	P		0	0	0
			457	217	89	129	22				

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	M	22	Total	C	N	O	P	0	0	0
			457	217	89	129	22			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	F	22	Total	C	N	O	P	0	0	0
			445	214	74	135	22			
3	N	22	Total	C	N	O	P	0	0	0
			445	214	74	135	22			

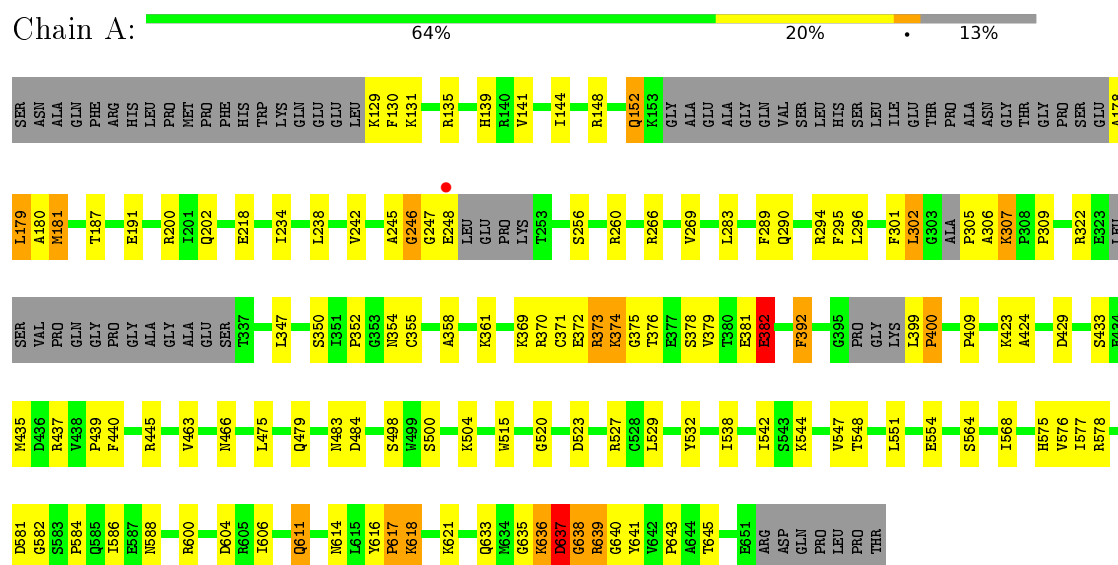
- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	102	Total	O	0	0
			102	102		
4	B	54	Total	O	0	0
			54	54		
4	D	24	Total	O	0	0
			24	24		
4	F	2	Total	O	0	0
			2	2		
4	M	5	Total	O	0	0
			5	5		
4	N	2	Total	O	0	0
			2	2		
4	C	42	Total	O	0	0
			42	42		

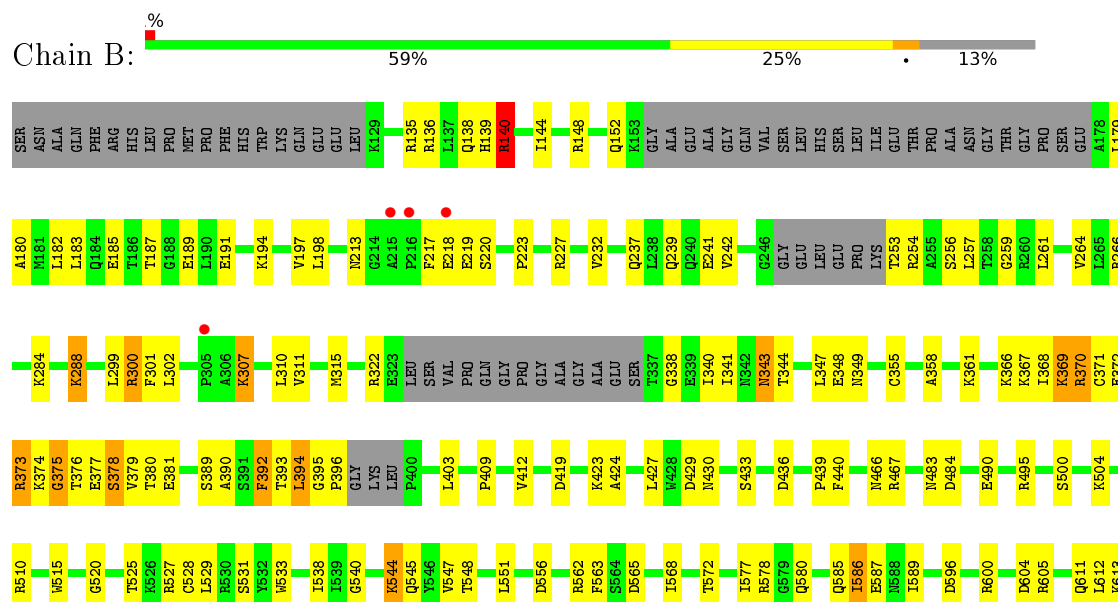
3 Residue-property plots

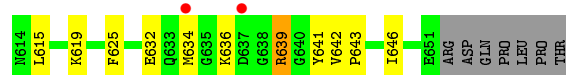
These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are color-coded according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



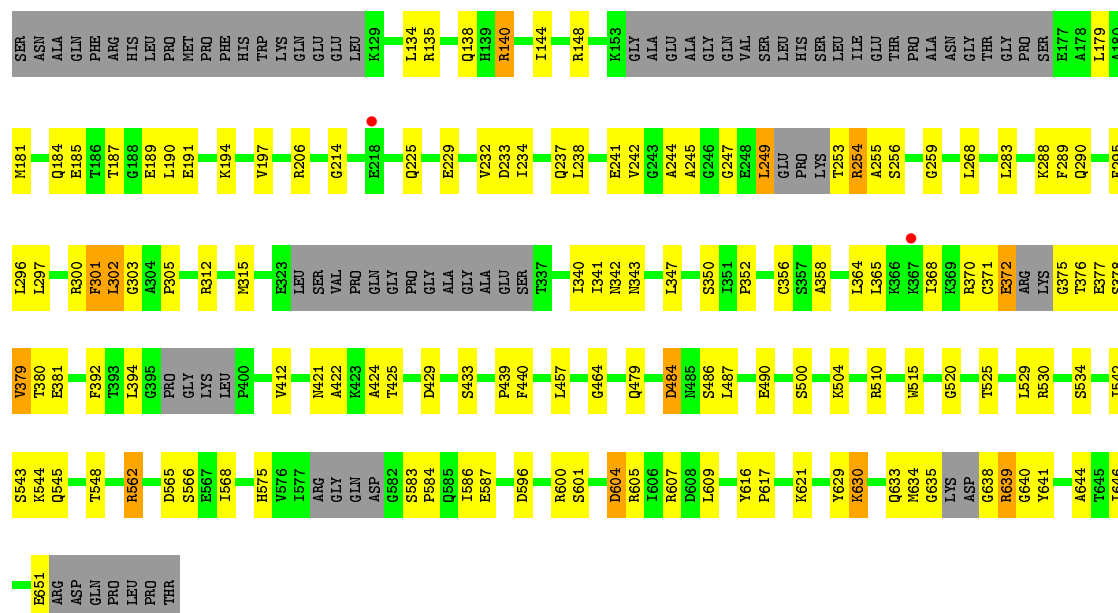
- Molecule 1: Signal transducer and activator of transcription 6





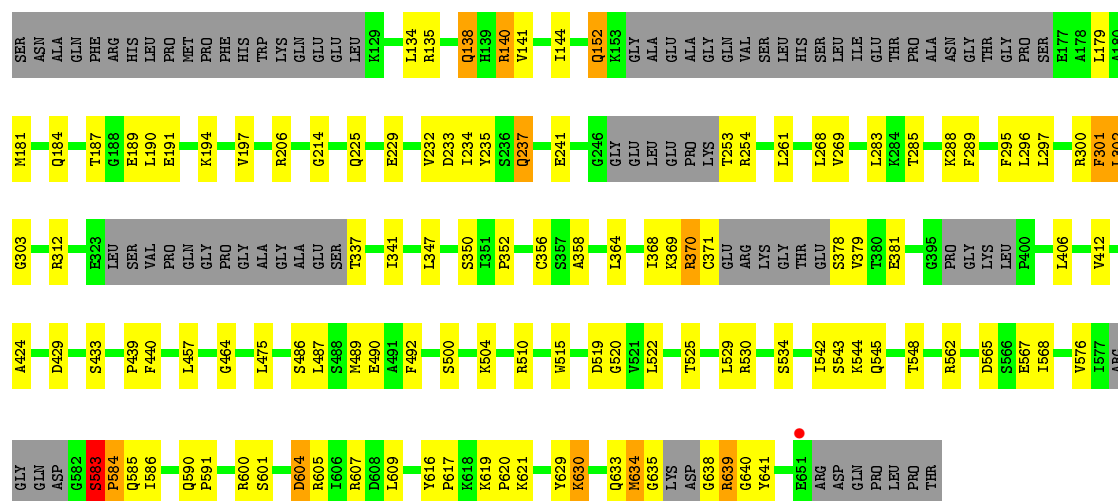
- Molecule 1: Signal transducer and activator of transcription 6

Chain D: 61% 23% 14%



- Molecule 1: Signal transducer and activator of transcription 6

Chain C: 63% 19% 15%

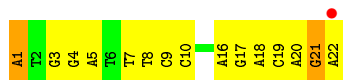


- Molecule 2: DNA (5'-D(P*AP*TP*GP*GP*AP*TP*TP*TP*CP*CP*TP*AP*GP*GP*AP*A P*GP*AP*CP*AP*A)-3')

Chain E: 5% 32% 68%



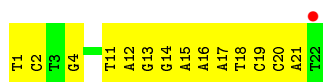
- Molecule 2: DNA (5'-D(P*AP*TP*GP*GP*AP*TP*TP*TP*CP*CP*TP*AP*GP*GP*AP*A P*GP*AP*CP*AP*A)-3')



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



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



4 Data and refinement statistics

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, α , β , γ	68.30Å 94.24Å 147.78Å 99.85° 101.73° 89.93°	Depositor
Resolution (Å)	38.81 – 3.10 38.81 – 3.10	Depositor EDS
% Data completeness (in resolution range)	89.5 (38.81-3.10) 83.8 (38.81-3.10)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.11	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.55 (at 3.12Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.8.1_1168)	Depositor
R, R_{free}	0.250 , 0.278 0.245 , 0.272	Depositor DCC
R_{free} test set	2765 reflections (5.06%)	DCC
Wilson B-factor (Å ²)	61.7	Xtriage
Anisotropy	0.099	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.29 , 27.0	EDS
L-test for twinning ²	$\langle L \rangle = 0.50$, $\langle L^2 \rangle = 0.33$	Xtriage
Estimated twinning fraction	0.088 for h,-k,-h-l	Xtriage
F_o, F_c correlation	0.88	EDS
Total number of atoms	17083	wwPDB-VP
Average B, all atoms (Å ²)	69.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 9.63% 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: 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	0.36	2/3851 (0.1%)	0.55	3/5186 (0.1%)
1	B	0.32	1/3856 (0.0%)	0.51	2/5196 (0.0%)
1	C	0.31	1/3744 (0.0%)	0.53	2/5043 (0.0%)
1	D	0.31	1/3792 (0.0%)	0.55	2/5107 (0.0%)
2	E	0.63	0/514	1.00	0/792
2	M	0.66	0/514	1.03	2/792 (0.3%)
3	F	0.57	0/496	0.95	0/762
3	N	0.63	0/496	1.00	0/762
All	All	0.37	5/17263 (0.0%)	0.62	11/23640 (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	2
1	B	0	1
1	C	0	1
1	D	0	1
All	All	0	5

All (5) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	382	GLU	C-N	-5.59	1.21	1.34
1	A	617	PRO	N-CD	5.46	1.55	1.47
1	B	396	PRO	N-CD	5.27	1.55	1.47
1	C	584	PRO	N-CD	5.09	1.54	1.47
1	D	584	PRO	N-CD	5.00	1.54	1.47

The worst 5 of 11 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	M	21	DG	P-O3'-C3'	-7.83	110.31	119.70
1	A	369	LYS	O-C-N	7.52	134.73	122.70
2	M	1	DA	O5'-P-OP2	-7.18	99.24	105.70
1	C	583	SER	C-N-CD	6.08	141.16	128.40
1	D	583	SER	C-N-CD	6.04	141.08	128.40

There are no chirality outliers.

All (5) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	218	GLU	Peptide
1	A	638	GLY	Peptide
1	B	375	GLY	Peptide
1	C	301	PHE	Peptide
1	D	301	PHE	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	3802	0	3882	161	1
1	B	3804	0	3884	133	1
1	C	3697	0	3778	84	0
1	D	3745	0	3815	118	0
2	E	457	0	248	22	0
2	M	457	0	248	23	3
3	F	445	0	251	31	0
3	N	445	0	251	12	0
4	A	102	0	0	16	0
4	B	54	0	0	5	0
4	C	42	0	0	3	0
4	D	24	0	0	1	0
4	F	2	0	0	0	0
4	M	5	0	0	2	0
4	N	2	0	0	0	0
All	All	17083	0	16357	547	4

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 17.

The worst 5 of 547 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)
3:F:13:DG:H2"	3:F:14:DG:C8	1.30	1.61
3:F:13:DG:H2"	3:F:14:DG:N7	1.26	1.48
1:A:302:LEU:HD22	1:A:306:ALA:CB	1.54	1.36
1:A:302:LEU:CD2	1:A:306:ALA:HB3	1.56	1.34
1:A:302:LEU:CG	1:A:306:ALA:HB3	1.59	1.33

All (4) 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)
2:M:1:DA:OP2	2:M:22:DA:OP2[1_655]	1.50	0.70
2:M:1:DA:OP2	2:M:22:DA:P[1_655]	2.08	0.12
1:A:466:ASN:ND2	1:B:483:ASN:O[1_655]	2.13	0.07
2:M:1:DA:P	2:M:22:DA:OP2[1_655]	2.18	0.02

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	465/549 (85%)	448 (96%)	14 (3%)	3 (1%)	30	68
1	B	467/549 (85%)	448 (96%)	18 (4%)	1 (0%)	52	84
1	C	448/549 (82%)	436 (97%)	10 (2%)	2 (0%)	39	75
1	D	455/549 (83%)	438 (96%)	15 (3%)	2 (0%)	39	75
All	All	1835/2196 (84%)	1770 (96%)	57 (3%)	8 (0%)	39	75

5 of 8 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	246	GLY
1	B	338	GLY
1	D	371	CYS
1	A	637	ASP
1	C	583	SER

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	413/469 (88%)	395 (96%)	18 (4%)	35	71
1	B	413/469 (88%)	387 (94%)	26 (6%)	22	58
1	C	402/469 (86%)	385 (96%)	17 (4%)	36	73
1	D	406/469 (87%)	389 (96%)	17 (4%)	36	73
All	All	1634/1876 (87%)	1556 (95%)	78 (5%)	31	69

5 of 78 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	528	CYS
1	B	639	ARG
1	C	583	SER
1	B	531	SER
1	B	585	GLN

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 7 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	628	HIS
1	C	585	GLN
1	D	237	GLN
1	B	213	ASN
1	D	290	GLN

5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

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

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	13,16,17	0.69	0	19,22,24	0.98	1 (5%)
1	PTR	B	641	1	13,16,17	1.24	1 (7%)	19,22,24	0.69	0
1	PTR	C	641	1	13,16,17	1.21	1 (7%)	19,22,24	0.64	0
1	PTR	D	641	1	13,16,17	1.21	1 (7%)	19,22,24	0.68	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

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	641	PTR	OH-CZ	-4.32	1.30	1.40
1	D	641	PTR	OH-CZ	-4.20	1.30	1.40
1	C	641	PTR	OH-CZ	-4.17	1.30	1.40

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed($^{\circ}$)	Ideal($^{\circ}$)
1	A	641	PTR	O-C-CA	-2.18	119.88	125.72

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

4 monomers are involved in 23 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
1	A	641	PTR	17	0
1	B	641	PTR	2	0
1	C	641	PTR	3	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 [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	477/549 (86%)	-0.16	1 (0%) 95 91	25, 55, 102, 137	0
1	B	476/549 (86%)	-0.09	6 (1%) 79 62	23, 56, 105, 141	0
1	C	464/549 (84%)	-0.09	1 (0%) 95 91	43, 72, 114, 146	0
1	D	471/549 (85%)	-0.16	2 (0%) 93 85	46, 72, 112, 148	0
2	E	22/22 (100%)	-0.25	1 (4%) 37 17	74, 86, 103, 142	0
2	M	22/22 (100%)	-0.37	1 (4%) 37 17	74, 86, 111, 144	0
3	F	22/22 (100%)	-0.36	1 (4%) 37 17	48, 89, 114, 129	0
3	N	22/22 (100%)	-0.36	1 (4%) 37 17	50, 90, 114, 155	0
All	All	1976/2284 (86%)	-0.14	14 (0%) 89 78	23, 67, 110, 155	0

The worst 5 of 14 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	E	22	DA	4.5
3	N	22	DT	4.2
1	B	305	PRO	4.0
1	B	215	ALA	3.1
1	C	651	GLU	3.1

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(\AA^2)	Q<0.9
1	PTR	D	641	16/17	0.96	0.20	-	45,50,52,52	0
1	PTR	A	641	16/17	0.82	0.26	-	70,75,78,85	0
1	PTR	C	641	16/17	0.97	0.18	-	43,46,49,49	0
1	PTR	B	641	16/17	0.88	0.19	-	65,68,75,77	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.