wwPDB X-ray Structure Validation Summary Report (i)

Dec 13, 2023 – 10:41 am GMT

PDB ID	:	4A0A
Title	:	Structure of hsDDB1-drDDB2 bound to a 16 bp CPD-duplex (pyrimidine at D-1 position) at 3.6 A resolution (CPD 3)
Authors	:	Scrima, A.; Fischer, E.S.; Iwai, S.; Gut, H.; Thoma, N.H.
Deposited on	:	2011-09-08
Resolution	:	3.60 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
https://www.wwpdb.org/validation/2017/XrayValidationReportHelp
with specific help available everywhere you see the (i) 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 (i)) were used in the production of this report:

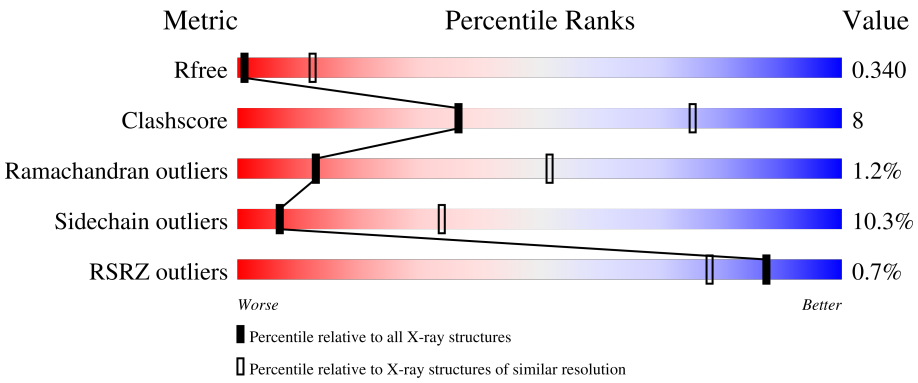
MolProbity	:	4.02b-467
Mogul	:	1.8.4, CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

The reported resolution of this entry is 3.60 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	1257 (3.70-3.50)
Clashscore	141614	1353 (3.70-3.50)
Ramachandran outliers	138981	1307 (3.70-3.50)
Sidechain outliers	138945	1307 (3.70-3.50)
RSRZ outliers	127900	1161 (3.70-3.50)

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 >=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 <=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	1159	52% 13% • 33%
2	B	382	2% 65% 25% • 7%
3	C	15	33% 47% 20%
4	D	16	56% 38% 6%

2 Entry composition (i)

There are 5 unique types of molecules in this entry. The entry contains 9619 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 DAMAGE-BINDING PROTEIN 1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	780	6129	3901	1025	1169	34	0	0	0

There are 20 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-18	MET	-	expression tag	UNP Q16531
A	-17	HIS	-	expression tag	UNP Q16531
A	-16	HIS	-	expression tag	UNP Q16531
A	-15	HIS	-	expression tag	UNP Q16531
A	-14	HIS	-	expression tag	UNP Q16531
A	-13	HIS	-	expression tag	UNP Q16531
A	-12	HIS	-	expression tag	UNP Q16531
A	-11	VAL	-	expression tag	UNP Q16531
A	-10	ASP	-	expression tag	UNP Q16531
A	-9	GLU	-	expression tag	UNP Q16531
A	-8	ASN	-	expression tag	UNP Q16531
A	-7	LEU	-	expression tag	UNP Q16531
A	-6	TYR	-	expression tag	UNP Q16531
A	-5	PHE	-	expression tag	UNP Q16531
A	-4	GLN	-	expression tag	UNP Q16531
A	-3	GLY	-	expression tag	UNP Q16531
A	-2	GLY	-	expression tag	UNP Q16531
A	-1	GLY	-	expression tag	UNP Q16531
A	0	ARG	-	expression tag	UNP Q16531
A	224	LYS	GLU	engineered mutation	UNP Q16531

- Molecule 2 is a protein called DNA DAMAGE-BINDING PROTEIN 2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	B	355	2843	1806	499	527	11	0	0	0

There are 20 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	76	MET	-	expression tag	UNP Q2YDS1
B	77	HIS	-	expression tag	UNP Q2YDS1
B	78	HIS	-	expression tag	UNP Q2YDS1
B	79	HIS	-	expression tag	UNP Q2YDS1
B	80	HIS	-	expression tag	UNP Q2YDS1
B	81	HIS	-	expression tag	UNP Q2YDS1
B	82	HIS	-	expression tag	UNP Q2YDS1
B	83	ARG	-	expression tag	UNP Q2YDS1
B	84	ARG	-	expression tag	UNP Q2YDS1
B	85	LEU	-	expression tag	UNP Q2YDS1
B	86	VAL	-	expression tag	UNP Q2YDS1
B	87	PRO	-	expression tag	UNP Q2YDS1
B	88	ARG	-	expression tag	UNP Q2YDS1
B	89	GLY	-	expression tag	UNP Q2YDS1
B	90	SER	-	expression tag	UNP Q2YDS1
B	91	GLY	-	expression tag	UNP Q2YDS1
B	92	GLY	-	expression tag	UNP Q2YDS1
B	93	ARG	-	expression tag	UNP Q2YDS1
B	180	GLN	LEU	variant	UNP Q2YDS1
B	214	ARG	TRP	variant	UNP Q2YDS1

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
3	C	15	334	159	66	94	15	0	0	0

- Molecule 4 is a DNA chain called 5'-D(*CP*CP*TP*GP*CP*TP*CP*CP*TP*TP*TP*CP*AP*CP*CP*C)-3'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
4	D	16	312	151	50	96	15	0	0	0

- Molecule 5 is CALCIUM ION (three-letter code: CA) (formula: Ca).

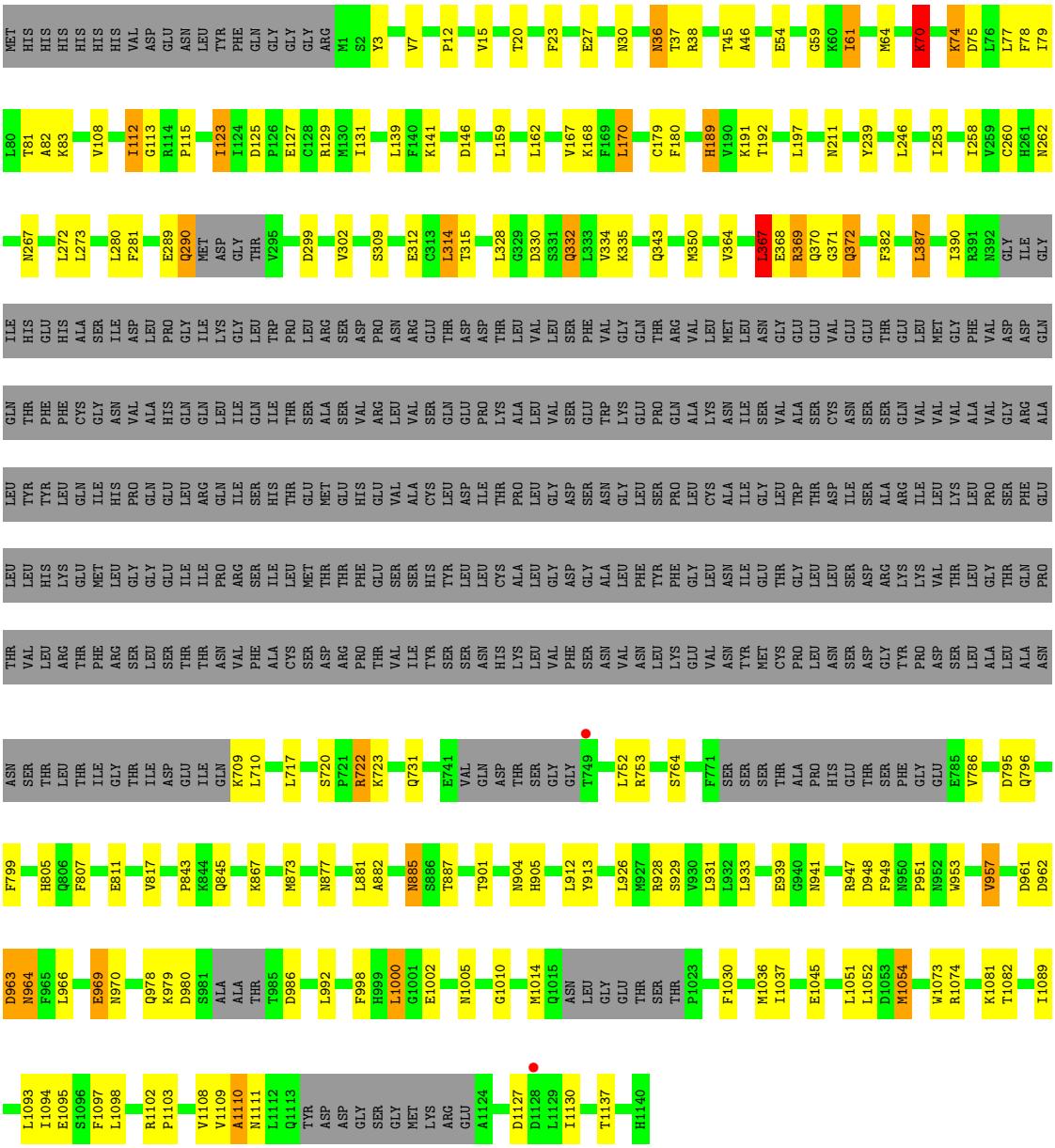
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	Ca		
5	C	1	1	1	0	0

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: DNA DAMAGE-BINDING PROTEIN 1

Chain A: 52% 13% • 33%



• Molecule 2: DNA DAMAGE-BINDING PROTEIN 2

Chain B: 2% 65% 25% • 7%

MET HIS HIS HIS HIS HIS HIS ARG ARG LEU VAL PRO ARG GLY SER GLY GLY ARG THR GLY GLY GLN LYS LYS VAL G101 S117 Q121 Q124 Q127 L134 K135 S136 Y137 H140 S144 R148 T158 H159 S167 I172 I173 L174 W175 D178 V179 Q180 M181

K182 T183 I186 M189 D193 A194 I195 T196 N201 Q202 F203 N204 Q207 L208 F209 V210 S211 S212 I213 R214 F222 S225 V226 I227 Q228 V229 F230 T233 D234 D237 Y238 W239 V243 S246 R249 Q250 M251 G255 L262 L263 L266 D267 G268 H269

F272 A279 K280 V281 A284 E285 F286 N287 P288 R289 C290 D291 W292 L293 M294 A295 V299 L308 K312 M321 K325 P326 A329 A330 Y331 F332 T338 K339 L340 L341 T342 T343 D344 Q345 R346 N347 S354 Y355 I364 I365 H366 P367 H368 R369 L374 I377

I388 V389 R392 Y393 D396 Q397 L398 L399 K403 R404 T405 I408 N412 H418 Q419 N432 L441 A442 S443 N448 W452 N453 R454 E455 ASP THR

• Molecule 3: 5'-D(*GP*GP*TP*GP*AP*AP*AP*(TTD)P*AP*GP*CP*AP*GP*DGP)-3'

Chain C: 33% 47% 20%

G1 G2 G3 T4 G5 T8 T9 A11 G12 C13 A14 G15 G16

• Molecule 4: 5'-D(*CP*CP*TP*GP*CP*TP*CP*CP*TP*TP*TP*CP*AP*CP*CP*C)-3'

Chain D: 56% 38% 6%

C1 C5 T6 C7 C8 A9 T10 T11 C12 A13 C16

4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 2 2 21	Depositor
Cell constants a, b, c, α, β, γ	155.64Å 227.14Å 114.32Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	46.25 – 3.60 46.25 – 3.60	Depositor EDS
% Data completeness (in resolution range)	100.0 (46.25-3.60) 99.5 (46.25-3.60)	Depositor EDS
R_{merge}	0.23	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.23 (at 3.57Å)	Xtriage
Refinement program	REFMAC 5.5.0102	Depositor
R, R_{free}	0.263 , 0.347 0.259 , 0.340	Depositor DCC
R_{free} test set	1188 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å^2)	71.3	Xtriage
Anisotropy	0.127	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.24 , 2.4	EDS
L-test for twinning[2]	$< \|L\| > = 0.45$, $< L^2 > = 0.28$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.85	EDS
Total number of atoms	9619	wwPDB-VP
Average B, all atoms (Å^2)	73.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.24% of the height of the origin peak. No significant pseudotranslation is detected.*

[1]Intensities estimated from amplitudes.

[2]Theoretical values of $< |L| >$, $< L^2 >$ 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: CA, TTD

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	0/6236	0.55	1/8429 (0.0%)
2	B	0.36	0/2917	0.55	0/3962
3	C	0.61	0/331	1.00	2/509 (0.4%)
4	D	0.66	0/346	1.14	1/529 (0.2%)
All	All	0.39	0/9830	0.60	4/13429 (0.0%)

There are no bond length outliers.

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
4	D	9	DA	P-O3'-C3'	5.89	126.77	119.70
1	A	367	LEU	CA-CB-CG	5.49	127.93	115.30
3	C	15	DG	P-O3'-C3'	5.29	126.04	119.70
3	C	12	DG	P-O3'-C3'	5.14	125.87	119.70

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	6129	0	6105	90	0
2	B	2843	0	2788	55	0
3	C	334	0	182	7	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
4	D	312	0	182	4	0
5	C	1	0	0	0	0
All	All	9619	0	9257	149	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 8.

The worst 5 of 149 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:A:928:ARG:HG2	1:A:928:ARG:HH11	1.40	0.84
1:A:969:GLU:HG2	1:A:970:ASN:N	1.99	0.78
1:A:369:ARG:HG2	1:A:370:GLN:N	1.98	0.78
1:A:38:ARG:HH12	1:A:54:GLU:HB3	1.48	0.77
1:A:127:GLU:HB2	1:A:129:ARG:HD2	1.66	0.77

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	764/1159 (66%)	681 (89%)	73 (10%)	10 (1%)	12 50
2	B	353/382 (92%)	314 (89%)	36 (10%)	3 (1%)	19 59
All	All	1117/1541 (72%)	995 (89%)	109 (10%)	13 (1%)	13 51

5 of 13 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	962	ASP
1	A	112	ILE

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	A	1110	ALA
2	B	189	MET
1	A	70	LYS

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	679/1015 (67%)	617 (91%)	62 (9%)	9	39
2	B	313/335 (93%)	273 (87%)	40 (13%)	4	24
All	All	992/1350 (74%)	890 (90%)	102 (10%)	7	34

5 of 102 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	1082	THR
2	B	203	PHE
2	B	412	ASN
1	A	1093	LEU
2	B	158	THR

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 31 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	941	ASN
2	B	373	HIS
1	A	990	GLN
2	B	432	ASN
2	B	269	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)

1 non-standard protein/DNA/RNA residue is 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	$\#\|Z\| > 2$
3	TTD	C	9	3	42,45,46	1.82	10 (23%)	62,74,77	2.39	16 (25%)

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
3	TTD	C	9	3	-	12/22/109/110	0/5/6/6

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

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	C	9	TTD	C1R-N1T	4.74	1.51	1.45
3	C	9	TTD	C1'-N1	4.59	1.51	1.45
3	C	9	TTD	C5'-C4R	3.64	1.62	1.51
3	C	9	TTD	C2-N1	3.49	1.43	1.36
3	C	9	TTD	C6-N1	3.48	1.52	1.46

The worst 5 of 16 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
3	C	9	TTD	O4R-C1R-N1T	8.41	118.62	108.65
3	C	9	TTD	C5T-C4T-N3T	6.58	121.79	116.06
3	C	9	TTD	N3-C2-N1	6.07	122.99	116.69
3	C	9	TTD	C5-C4-N3	6.07	121.35	116.06
3	C	9	TTD	N3T-C2T-N1T	4.41	121.27	116.69

There are no chirality outliers.

5 of 12 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	C	9	TTD	C2'-C3R-O3R-PB
3	C	9	TTD	O4'-C1'-N1-C2
3	C	9	TTD	O4R-C4'-C5R-O5R
3	C	9	TTD	O4'-C1'-N1-C6
3	C	9	TTD	O4'-C4R-C5'-O5'

There are no ring outliers.

1 monomer is involved in 2 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	C	9	TTD	2	0

5.5 Carbohydrates (i)

There are no monosaccharides in this entry.

5.6 Ligand geometry (i)

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 (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	780/1159 (67%)	-0.31	2 (0%)	94	88	54, 68, 96, 117	0
2	B	355/382 (92%)	-0.25	6 (1%)	70	55	54, 73, 97, 102	0
3	C	14/15 (93%)	-0.68	0	100	100	82, 105, 114, 119	0
4	D	16/16 (100%)	-0.46	0	100	100	67, 91, 124, 127	0
All	All	1165/1572 (74%)	-0.29	8 (0%)	87	78	54, 69, 100, 127	0

The worst 5 of 8 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	1128	ASP	2.5
2	B	225	SER	2.4
2	B	227	ILE	2.2
2	B	233	THR	2.2
2	B	234	ASP	2.1

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. 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	B-factors(Å^2)	Q<0.9
3	TTD	C	9	40/41	0.92	0.16	85,92,96,96	0

6.3 Carbohydrates (i)

There are no monosaccharides 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. 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	B-factors(Å^2)	Q<0.9
5	CA	C	1017	1/1	0.85	0.16	77,77,77,77	0

6.5 Other polymers (i)

There are no such residues in this entry.