W O R L D W I D E

wwPDB

P R O T E I N   D A T A   B A N K

wwPDB X-ray Structure Validation Summary Report (i)

Oct 12, 2021 – 03:17 PM EDT

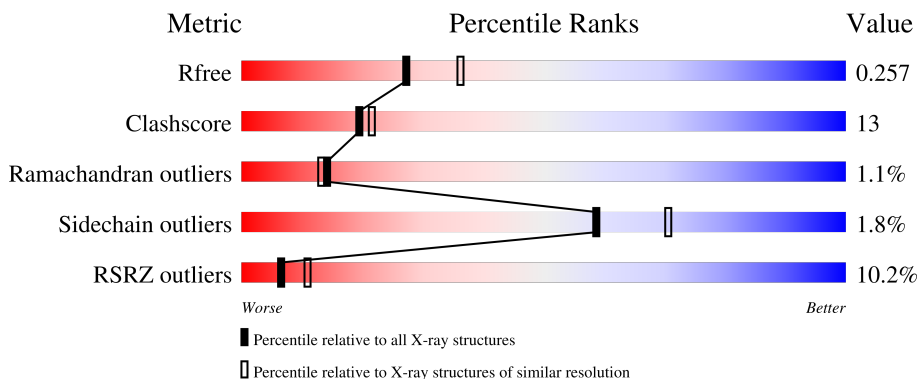|		
PDB ID	:	1XHX
Title	:	Phi29 DNA Polymerase, orthorhombic crystal form
Authors	:	Kamtekar, S.; Berman, A.J.; Wang, J.; Lazaro, J.M.; de Vega, M.; Blanco, L.; Salas, M.; Steitz, T.A.
Deposited on	:	2004-09-21
Resolution	:	2.35 Å(reported)

---

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 2.35 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	Percentile Ranks	Value
Rfree		0.257
Clashscore		13
Ramachandran outliers		1.1%
Sidechain outliers		1.8%
RSRZ outliers		10.2%

Worse   *Better*

▮ Percentile relative to all X-ray structures

▯ Percentile relative to X-ray structures of similar resolution

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	1164 (2.36-2.36)
Clashscore	141614	1232 (2.36-2.36)
Ramachandran outliers	138981	1211 (2.36-2.36)
Sidechain outliers	138945	1212 (2.36-2.36)
RSRZ outliers	127900	1150 (2.36-2.36)

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	575	14% / 73% / 25% / ••
1	B	575	11% / 72% / 26% / ••
1	C	575	11% / 73% / 25% / ••
1	D	575	5% / 79% / 19% / ••

2   Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 20193 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 polymerase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	571	Total	C	N	O	S	0	0	0
			4668	3041	754	852	21			
1	B	571	Total	C	N	O	S	0	0	0
			4668	3041	754	852	21			
1	C	571	Total	C	N	O	S	0	0	0
			4668	3041	754	852	21			
1	D	571	Total	C	N	O	S	0	0	0
			4668	3041	754	852	21			

There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	12	ALA	ASP	engineered mutation	UNP P03680
A	66	ALA	ASP	engineered mutation	UNP P03680
B	12	ALA	ASP	engineered mutation	UNP P03680
B	66	ALA	ASP	engineered mutation	UNP P03680
C	12	ALA	ASP	engineered mutation	UNP P03680
C	66	ALA	ASP	engineered mutation	UNP P03680
D	12	ALA	ASP	engineered mutation	UNP P03680
D	66	ALA	ASP	engineered mutation	UNP P03680

- Molecule 2 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	C	2	Total	Mg	0	0
			2	2		

- Molecule 3 is SULFATE ION (three-letter code: SO4) (formula: O_4S).

SO4

O1

O

O3 O⁻ —— S —— O⁻ O4

S

O

O2

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
3	D	1	Total 5	O 4	S 1	0	0
3	D	1	Total 5	O 4	S 1	0	0

- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	295	Total 295	O 295	0	0
4	B	388	Total 388	O 388	0	0
4	C	382	Total 382	O 382	0	0
4	D	444	Total 444	O 444	0	0

3 Residue-property plots ⓘ

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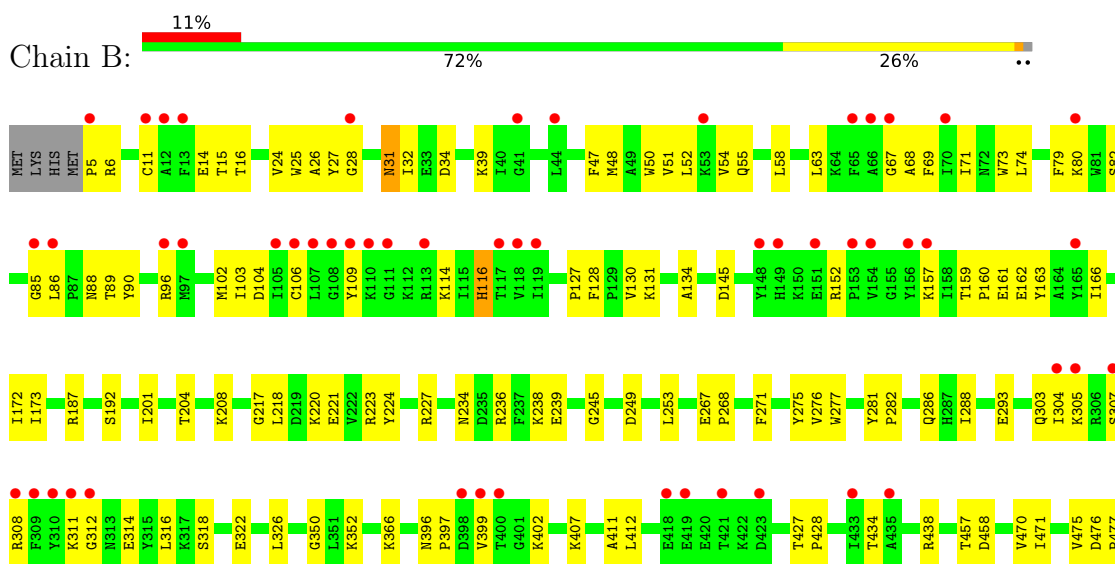 and electron density. Residues are color-coded according to the number of geometric quality criteria for which they contain at least one outlier: green = 0, yellow = 1, orange = 2 and red = 3 or more. A red dot above a residue indicates a poor fit to the electron density (RSRZ > 2). Stretches of 2 or 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.
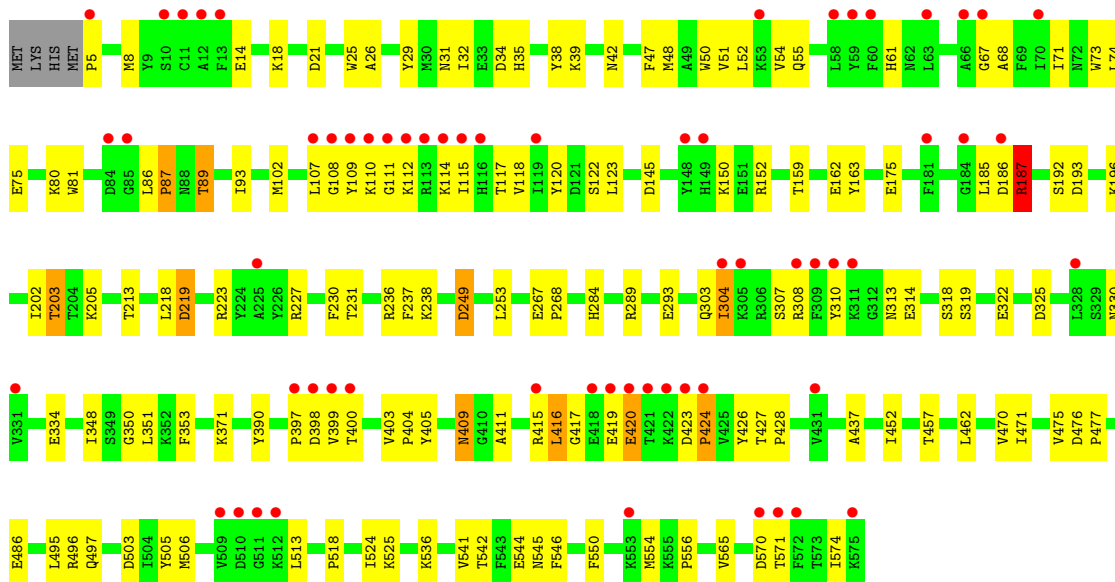
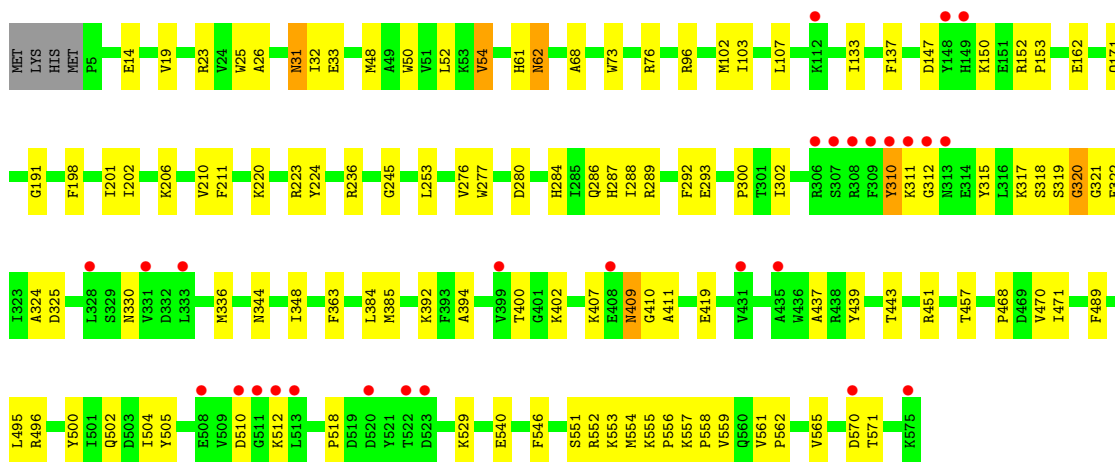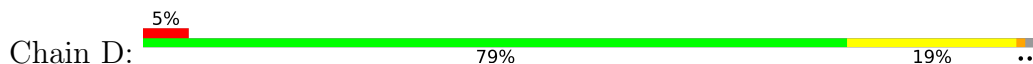• Molecule 1: DNA polymerase

Chain A:

• Molecule 1: DNA polymerase

Chain B:

G481 F489 L495 K507 E508 V509 D510 G511 K512 E515 G516 S517 P518 T522 D523 G532 M533 T534 K538 T542 N545 F546 F550 S551 R552 K553 M554 K555 P556 K557 P558 V561 P562 V565 V566 L567 V568 D569 D570 T571 F572 T573 I574 K575

• Molecule 1: DNA polymerase

Chain C:

11%

73%   25%

MET LYS HIS MET P5 M8 Y9 S10 C11 A12 F13 E14 K18 D21 W25 A26 Y29 M30 N31 I32 E33 D34 H35 Y38 K39 N42 F47 M48 A49 W50 V51 L52 K53 V54 Q55 L58 Y59 F60 H61 N62 L63 A66 G67 A68 F69 I70 I71 N72 W73 L74

E75 K80 W81 D84 G85 L86 P87 N88 T89 I93 M102 L107 G108 Y109 K110 G111 K112 R113 K114 I115 H116 T117 V118 I119 Y120 D121 S122 L123 D145 Y148 H149 K150 E151 R152 T159 E162 Y163 E175 F181 G184 L185 D186 H187 S192 D193 K196

I202 T203 T204 K205 T213 L218 D219 R223 Y224 A225 Y226 R227 F230 T231 R236 F237 K238 D249 L253 E267 P268 H284 R289 E293 Q303 I304 K305 R306 S307 R308 F309 Y310 K311 G312 N313 E314 S318 S319 E322 D325 L328 S329 N330

V331 E334 I348 S349 G350 L351 K352 F353 K371 Y390 P397 D398 V399 T400 V403 P404 Y405 N409 G410 A411 R415 L416 G417 E418 E419 E420 T421 K422 D423 P424 V425 Y426 T427 P428 V431 A437 I452 T457 L462 V470 I471 V475 D476 P477

E486 L495 R496 Q497 D503 I504 Y505 M506 V509 D510 G511 K512 L513 P518 I524 K525 K536 V541 T542 E543 E544 N545 F546 F550 K553 M554 K555 P556 V565 D570 T571 F572 T573 I574 K675

• Molecule 1: DNA polymerase

Chain D:

5%

79%   19%

MET LYS HIS MET P5 E14 V19 R23 V24 W25 A26 N31 I32 E33 M48 A49 W50 V51 L52 K53 V54 H61 N62 A68 W73 R76 R96 M102 I103 L107 K112 I133 F137 D147 Y148 H149 K150 E151 R152 P153 E162 Q171

G191 F198 I201 I202 K206 V210 F211 K220 R223 Y224 R236 G245 L253 V276 W277 D280 H284 I285 Q286 H287 I288 R289 F292 E293 P300 T301 I302 R306 S307 R308 F309 Y310 K311 G312 N313 E314 Y315 L316 K317 S318 S319 G320 G321 E322

I323 A324 D325 L328 S329 N330 V331 D332 L333 M336 N344 I348 F363 L384 M385 K392 F393 A394 V399 T400 G401 K402 K407 E408 N409 G410 A411 E419 V431 A435 W436 A437 R438 Y439 T443 R451 T457 P468 D469 V470 I471 F489

L495 R496 Y500 I501 Q502 D503 I504 Y505 E508 V509 D510 G511 K512 L513 P518 D519 D520 Y521 T522 D523 K529 E540 F546 S551 R552 K553 M554 K555 P556 K557 P558 V559 Q560 V561 P562 V565 D570 T571 K575

4 Data and refinement statistics ⓘ

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	96.26Å 149.91Å 199.02Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	19.96 – 2.35 30.72 – 2.35	Depositor EDS
% Data completeness (in resolution range)	99.7 (19.96-2.35) 99.8 (30.72-2.35)	Depositor EDS
R_{merge}	0.10	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.16 (at 2.36Å)	Xtriage
Refinement program	CNS 1.0	Depositor
R, R_{free}	0.206 , 0.257 0.207 , 0.257	Depositor DCC
R_{free} test set	12006 reflections (10.01%)	wwPDB-VP
Wilson B-factor (Å^2)	42.7	Xtriage
Anisotropy	0.251	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.31 , 47.3	EDS
L-test for twinning[2]	$< |L| > = 0.50, < L^2 > = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	20193	wwPDB-VP
Average B, all atoms (Å^2)	54.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.72% 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 ⓘ

5.1   Standard geometry ⓘ

Bond lengths and bond angles in the following residue types are not validated in this section: SO4, MG

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	0.35	0/4788	0.65	3/6459 (0.0%)
1	B	0.36	0/4788	0.62	0/6459
1	C	0.36	0/4788	0.61	0/6459
1	D	0.38	0/4788	0.62	0/6459
All	All	0.36	0/19152	0.63	3/25836 (0.0%)

There are no bond length outliers.

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	5	PRO	CA-N-CD	-13.35	92.81	111.50
1	A	309	PHE	CA-CB-CG	-6.95	97.21	113.90
1	A	309	PHE	CA-C-N	-5.65	104.77	117.20

There are no chirality outliers.

There are no planarity outliers.

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	4668	0	4676	134	0
1	B	4668	0	4676	116	0
1	C	4668	0	4676	141	0
1	D	4668	0	4676	97	0

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	C	2	0	0	0	0
3	D	10	0	0	0	0
4	A	295	0	0	9	0
4	B	388	0	0	4	0
4	C	382	0	0	12	0
4	D	444	0	0	7	0
All	All	20193	0	18704	478	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 13.

The worst 5 of 478 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:310:TYR:OH	1:A:322:GLU:HG3	1.37	1.23
1:C:223:ARG:NH2	1:C:424:PRO:HG2	1.70	1.06
1:C:110:LYS:HG2	1:C:115:ILE:HD11	1.36	1.04
1:B:89:THR:HG22	1:B:90:TYR:H	1.22	1.03
1:A:75:GLU:HB3	1:A:406:LEU:HD11	1.45	0.98

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	569/575 (99%)	518 (91%)	44 (8%)	7 (1%)	13	11
1	B	569/575 (99%)	521 (92%)	44 (8%)	4 (1%)	22	23
1	C	569/575 (99%)	525 (92%)	36 (6%)	8 (1%)	11	9
1	D	569/575 (99%)	536 (94%)	28 (5%)	5 (1%)	17	17
All	All	2276/2300 (99%)	2100 (92%)	152 (7%)	24 (1%)	14	13

5 of 24 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	62	ASN
1	A	311	LYS
1	B	311	LYS
1	D	62	ASN
1	A	85	GLY

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	502/506 (99%)	492 (98%)	10 (2%)	55	66
1	B	502/506 (99%)	493 (98%)	9 (2%)	59	70
1	C	502/506 (99%)	491 (98%)	11 (2%)	52	63
1	D	502/506 (99%)	496 (99%)	6 (1%)	71	82
All	All	2008/2024 (99%)	1972 (98%)	36 (2%)	59	70

5 of 36 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	409	ASN
1	D	409	ASN
1	C	420	GLU
1	D	310	TYR
1	B	88	ASN

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 28 such sidechains are listed below:

Mol	Chain	Res	Type
1	C	31	ASN
1	D	485	HIS
1	C	171	GLN
1	D	287	HIS
1	C	55	GLN

5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

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

There are no non-standard protein/DNA/RNA residues in this entry.

5.5 Carbohydrates ⓘ

There are no monosaccharides in this entry.

5.6 Ligand geometry ⓘ

Of 4 ligands modelled in this entry, 2 are monoatomic - leaving 2 for Mogul analysis.

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	SO4	D	3001	-	4,4,4	1.92	1 (25%)	6,6,6	0.86	0
3	SO4	D	3000	-	4,4,4	1.84	2 (50%)	6,6,6	0.95	0

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	D	3001	SO4	O1-S	3.29	1.63	1.46
3	D	3000	SO4	O1-S	2.92	1.61	1.46
3	D	3000	SO4	O3-S	-2.12	1.30	1.47

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 ⓘ

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, 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	571/575 (99%)	0.71	81 (14%)	2	4	25, 55, 108, 135	0
1	B	571/575 (99%)	0.58	62 (10%)	5	9	23, 50, 95, 149	0
1	C	571/575 (99%)	0.55	62 (10%)	5	9	22, 47, 106, 147	0
1	D	571/575 (99%)	0.27	28 (4%)	29	42	22, 43, 84, 150	0
All	All	2284/2300 (99%)	0.53	233 (10%)	6	11	22, 48, 103, 150	0

The worst 5 of 233 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	D	310	TYR	11.5
1	D	306	ARG	8.7
1	A	148	TYR	8.5
1	B	310	TYR	8.4
1	A	309	PHE	8.1

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

There are no non-standard protein/DNA/RNA residues in this entry.

6.3 Carbohydrates ⓘ

There are no monosaccharides in this entry.

6.4 Ligands ⓘ

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
2	MG	C	5001	1/1	0.88	0.19	53,53,53,53	0
3	SO4	D	3001	5/5	0.93	0.19	98,99,102,103	0
2	MG	C	5000	1/1	0.95	0.22	47,47,47,47	0
3	SO4	D	3000	5/5	0.97	0.16	67,75,77,83	0

6.5   Other polymers ⓘ

There are no such residues in this entry.