wwPDB X-ray Structure Validation Summary Report (i)

Sep 17, 2023 – 10:10 AM EDT

PDB ID	:	4Y97
Title	:	Crystal Structure of human Pol alpha B-subunit in complex with C-terminal domain of catalytic subunit
Authors	:	Suwa, Y.; Gu, J.; Baranovskiy, A.G.; Babayeva, N.D.; Tahirov, T.H.
Deposited on	:	2015-02-17
Resolution	:	2.51 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 (1)) were used in the production of this report:

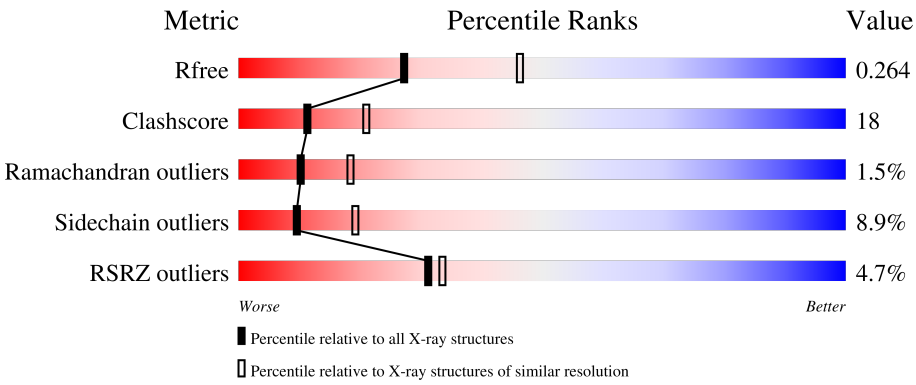
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.35.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.35.1

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 2.51 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	4661 (2.50-2.50)
Clashscore	141614	5346 (2.50-2.50)
Ramachandran outliers	138981	5231 (2.50-2.50)
Sidechain outliers	138945	5233 (2.50-2.50)
RSRZ outliers	127900	4559 (2.50-2.50)

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	598	2%, 50%, 21%, •, 26%
1	C	598	2%, 51%, 19%, •, •, 26%
1	E	598	%, 53%, 17%, •, 27%
1	G	598	2%, 47%, 21%, 5%, •, 27%
2	B	180	10%, 56%, 35%, 6%, •

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
2	D	180	8% 66% 27% 6%
2	F	180	4% 61% 33%
2	H	180	20% 57% 38%

2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 19724 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 alpha subunit B.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	442	3439	2187	574	663	15	0	0	0
1	C	441	3435	2187	573	660	15	0	0	0
1	E	439	3422	2178	571	658	15	0	0	0
1	G	436	3400	2164	568	653	15	0	0	0

- Molecule 2 is a protein called DNA polymerase alpha catalytic subunit.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	B	173	1426	902	240	269	15	0	0	0
2	D	178	1466	926	248	277	15	0	0	0
2	F	176	1451	919	244	273	15	0	0	0
2	H	179	1475	931	250	279	15	0	0	0

- Molecule 3 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	Zn		
3	B	2	2	2	0	0
3	D	2	2	2	0	0
3	F	2	2	2	0	0
3	H	2	2	2	0	0

- Molecule 4 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	A	37	Total O 37 37	0	0
4	B	8	Total O 8 8	0	0
4	C	63	Total O 63 63	0	0
4	D	24	Total O 24 24	0	0
4	E	39	Total O 39 39	0	0
4	F	14	Total O 14 14	0	0
4	G	13	Total O 13 13	0	0
4	H	4	Total O 4 4	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 polymerase alpha subunit B

Chain A: 2% 50% 21% • 26%

MET SER ALA SER ALA GLN GLN LEU ALA GLU GLU LEU GLN ILE PHE GLY LEU ASP CYS GLU GLU ALA LEU ILE GLU LYS LEU VAL GLU LEU CYS VAL GLN TYR GLY GLN ASN GLU GLU GLY MET VAL GLY GLU LEU ILE ALA PHE CYS THR SER THR HIS LYS VAL GLY LEU THR SER GLU

ILE LEU ASN SER PHE GLU HIS GLU PHE LEU SER LYS ARG LEU SER LYS ALA ARG HIS SER THR CYS LYS ASP SER GLY HIS ALA GLY ALA ARG ASP ILE VAL SER ILE GLN GLU LEU ILE GLU VAL GLU GLU GLU GLU GLU ILE LEU LEU ASN SER TYR THR THR PRO SER LYS GLY SER

GLN LYS ARG ALA ILE SER THR PRO GLU THR PRO LEU THR LYS ARG SER VAL SER THR ARG SER PRO HIS GLN LEU LEU SER PRO SER SER PHE SER PRO SER ALA THR P157 S158 Q159 K160 Y161 N162 S163 R164 S173 V180 S181 W182 R186 G187 A188 G189 N190 I191 S192 L193

K194 E200 A201 L202 T203 M209 F210 Q211 K212 L213 T221 C222 E226 L227 K232 I237 L243 L253 N266 I271 R276 P285 L296 F297 Q300 V301 I307 N308 T309 K313 V323 P324 T331 E332 E333 D334 A335 D336 F337 E338 Q339

L343 V344 Y349 S352 D353 S354 L360 L361 D362 L363 I364 A365 V366 I367 D370 D373 H386 E387 Q388 N391 C392 L393 L394 T395 E399 D400 I401 F402 K403 R407 T408 I409 I410 T413 L420 P424 S425 L426 R427 D428 V429 P433 V434 Y435 L445 S446

R447 E448 D449 K450 E458 I464 S473 L476 L477 F478 H479 L480 S487 S488 S489 G490 T491 F495 L499 K500 R506 S507 Y508 Y509 P510 L511 Y512 P513 P514 Q515 E516 D517 M518 Y522 F525 L531 D536 I539 I540 L544 K549 L552 R561

Q566 V567 G568 G569 T570 R573 R577 R578 P579 A580 A581 D582 G583 A584 E585 R586 Q587 Q594 R597 I598

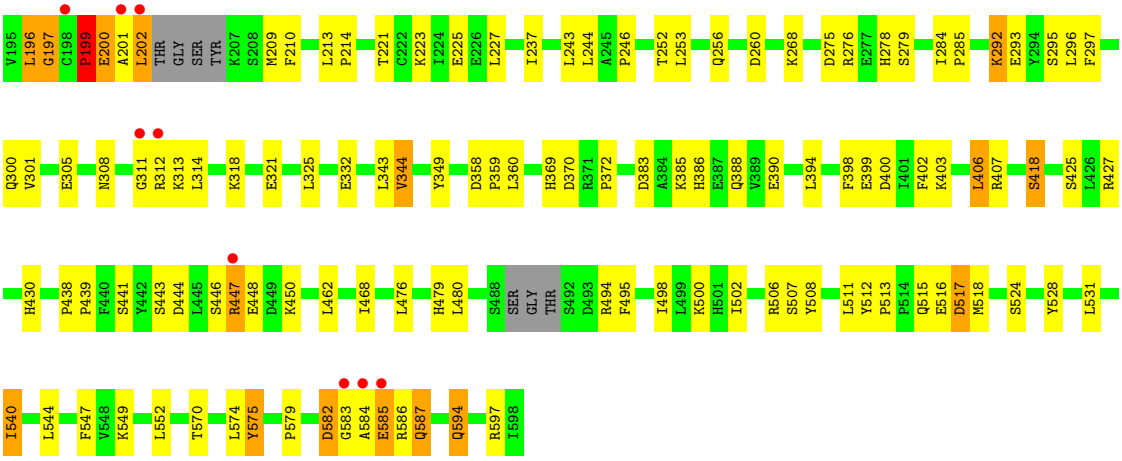
• Molecule 1: DNA polymerase alpha subunit B

Chain C: 2% 51% 19% • • 26%

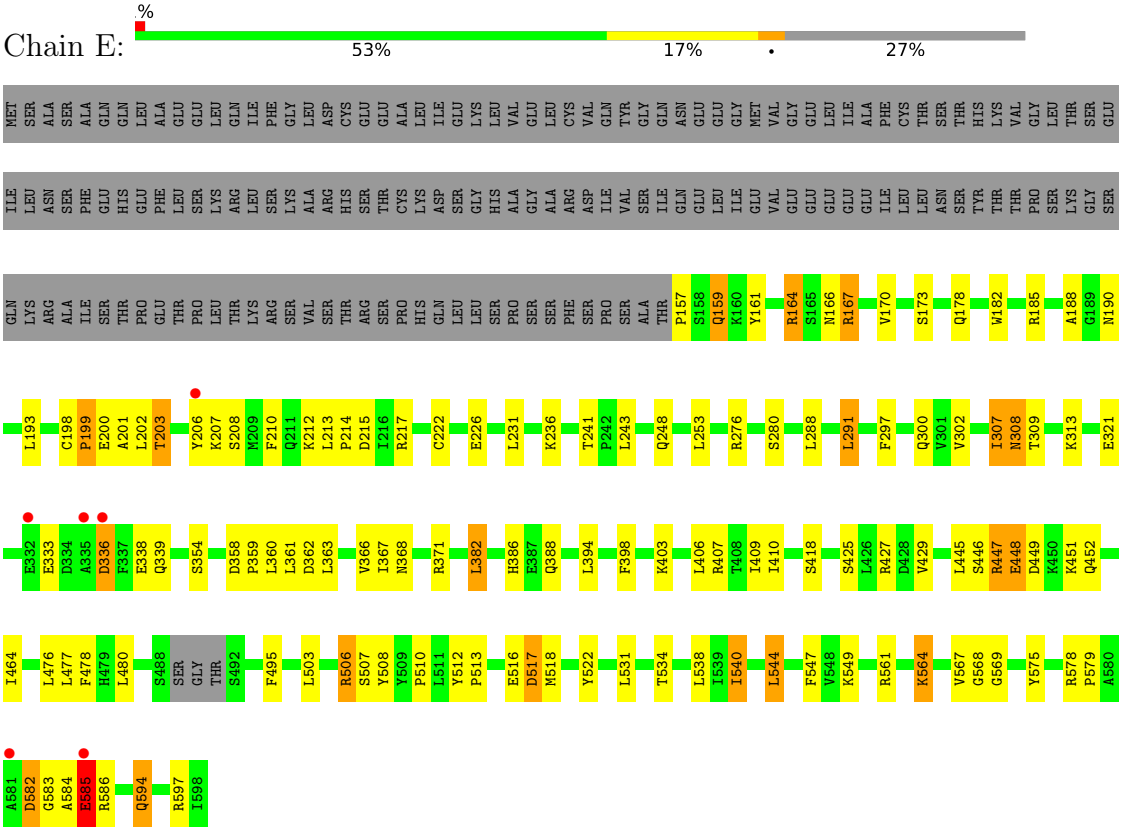
MET SER ALA SER ALA GLN GLN LEU ALA GLU GLU LEU GLN ILE PHE GLY LEU ASP CYS GLU GLU ALA LEU ILE GLU LYS LEU VAL GLU LEU CYS VAL GLN TYR GLY GLN ASN GLU GLU GLY MET VAL GLY GLU LEU ILE ALA PHE CYS THR SER THR HIS LYS VAL GLY LEU THR SER GLU

ILE LEU ASN SER PHE GLU HIS GLU PHE LEU SER LYS ARG LEU SER LYS ALA ARG HIS SER THR CYS LYS ASP SER GLY HIS ALA GLY ALA ARG ASP ILE VAL SER ILE GLN GLU LEU ILE GLU VAL GLU GLU GLU GLU GLU ILE LEU LEU ASN SER TYR THR THR PRO SER LYS GLY SER

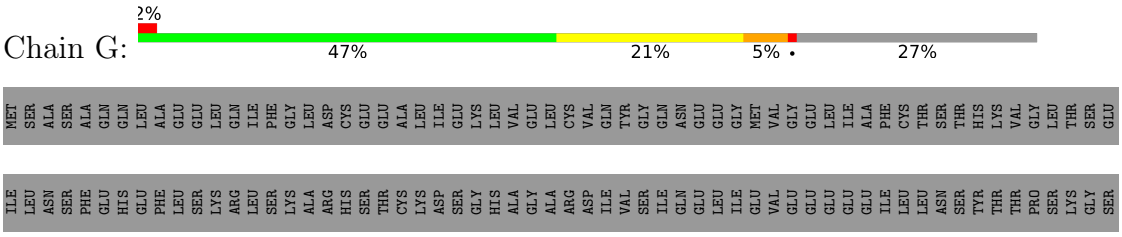
GLN LYS ARG ALA ILE SER THR PRO GLU THR PRO LEU THR LYS ARG SER VAL SER THR ARG SER PRO HIS GLN LEU LEU SER PRO SER SER F151 S152 P153 S154 A155 T156 P157 S158 Q159 R164 S165 N166 R167 G168 E169 Q178 G179 V180 S181 W182 R185 A188 G189 N190 K194

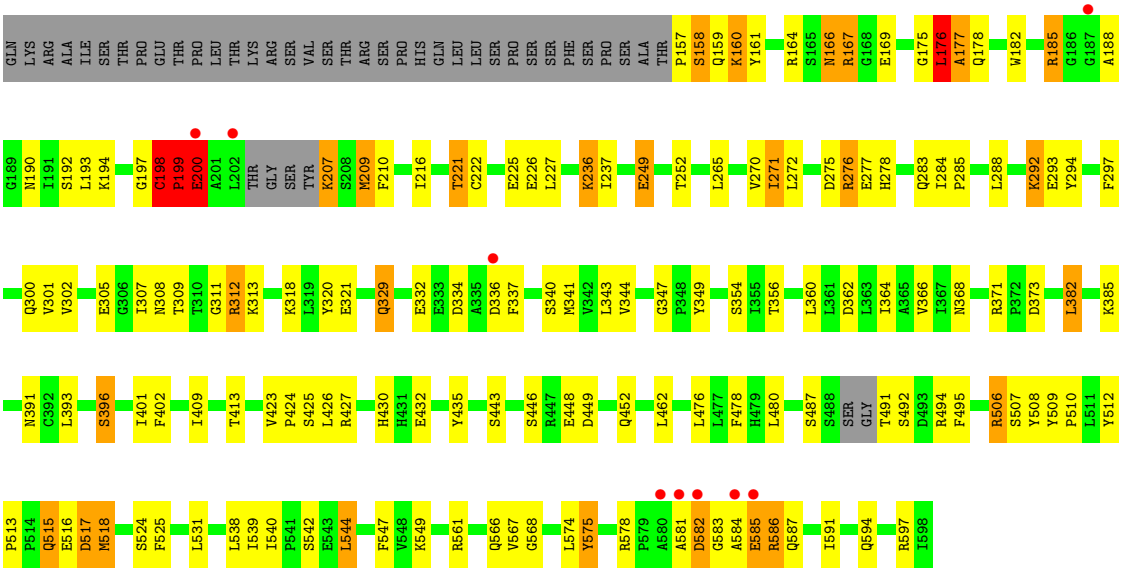


• Molecule 1: DNA polymerase alpha subunit B



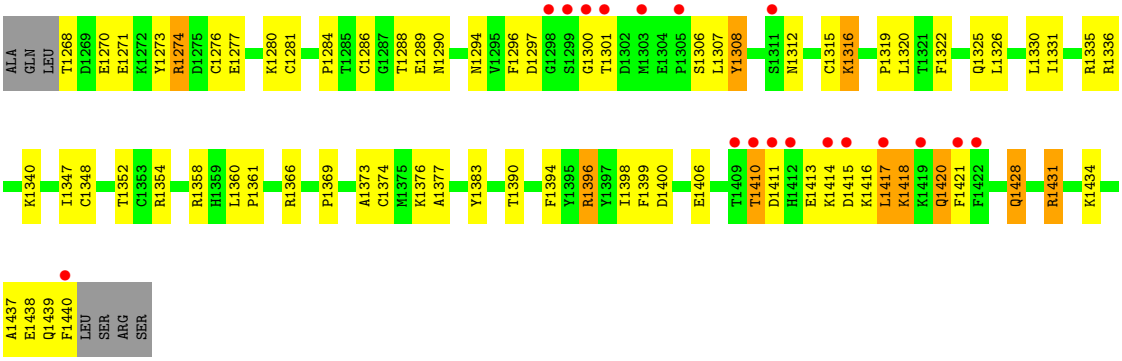
• Molecule 1: DNA polymerase alpha subunit B





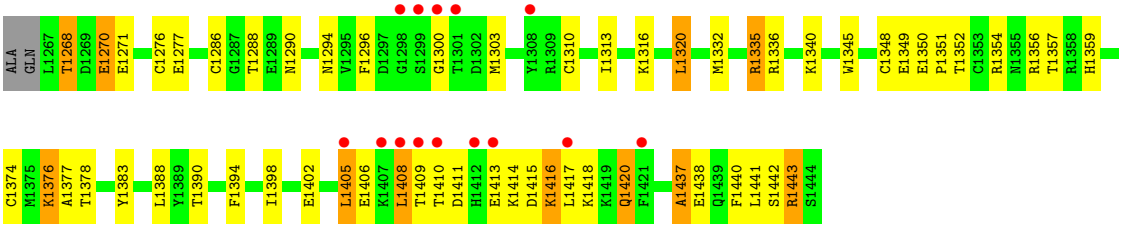
• Molecule 2: DNA polymerase alpha catalytic subunit

Chain B:



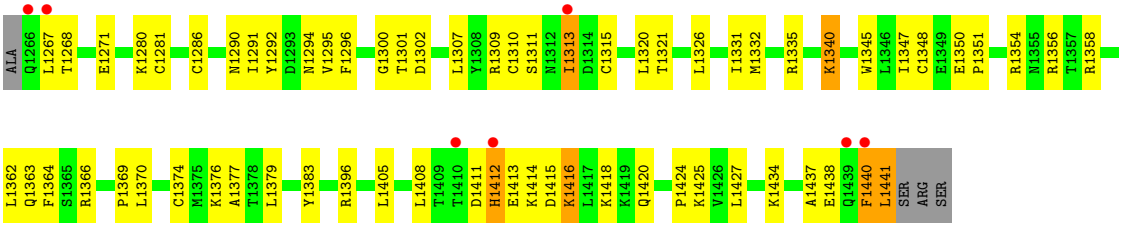
• Molecule 2: DNA polymerase alpha catalytic subunit

Chain D:



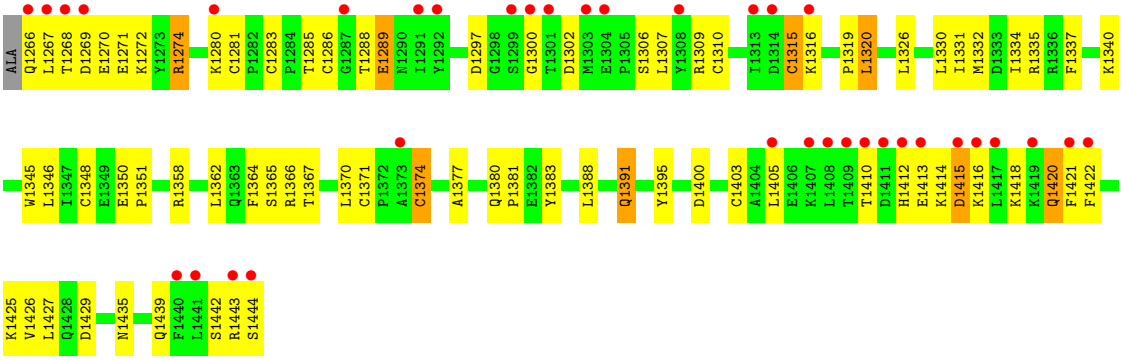
• Molecule 2: DNA polymerase alpha catalytic subunit

Chain F:



- Molecule 2: DNA polymerase alpha catalytic subunit

Chain H:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	103.59Å 137.13Å 132.09Å 90.00° 100.97° 90.00°	Depositor
Resolution (Å)	47.68 – 2.51 47.68 – 2.51	Depositor EDS
% Data completeness (in resolution range)	94.2 (47.68-2.51) 94.3 (47.68-2.51)	Depositor EDS
R_{merge}	0.08	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.07 (at 2.51Å)	Xtriage
Refinement program	CNS 1.1	Depositor
R, R_{free}	0.218 , 0.257 0.228 , 0.264	Depositor DCC
R_{free} test set	5894 reflections (5.04%)	wwPDB-VP
Wilson B-factor (Å^2)	31.2	Xtriage
Anisotropy	0.639	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.30 , 53.3	EDS
L-test for twinning[2]	$< \|L\| > = 0.48$, $< L^2 > = 0.31$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.93	EDS
Total number of atoms	19724	wwPDB-VP
Average B, all atoms (Å^2)	50.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.36% 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: ZN

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	0.43	0/3517	0.74	1/4777 (0.0%)
1	C	0.49	0/3512	0.84	7/4769 (0.1%)
1	E	0.46	0/3499	0.77	3/4751 (0.1%)
1	G	0.39	0/3475	0.78	8/4717 (0.2%)
2	B	0.43	0/1459	0.69	2/1967 (0.1%)
2	D	0.44	0/1499	0.67	1/2019 (0.0%)
2	F	0.42	0/1484	0.71	1/2001 (0.0%)
2	H	0.40	0/1508	0.68	1/2031 (0.0%)
All	All	0.44	0/19953	0.76	24/27032 (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	E	0	1

There are no bond length outliers.

The worst 5 of 24 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	C	156	THR	C-N-CD	-20.28	75.98	120.60
1	G	198	CYS	N-CA-C	-10.39	82.94	111.00
1	C	156	THR	C-N-CA	9.90	163.59	122.00
1	G	198	CYS	C-N-CD	-9.72	99.21	120.60
1	G	197	GLY	N-CA-C	7.01	130.63	113.10

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	E	522	TYR	Sidechain

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	3439	0	3414	134	0
1	C	3435	0	3410	132	0
1	E	3422	0	3398	78	0
1	G	3400	0	3380	144	0
2	B	1426	0	1389	53	0
2	D	1466	0	1434	50	0
2	F	1451	0	1419	63	0
2	H	1475	0	1442	72	0
3	B	2	0	0	0	0
3	D	2	0	0	0	0
3	F	2	0	0	0	0
3	H	2	0	0	0	0
4	A	37	0	0	3	0
4	B	8	0	0	0	0
4	C	63	0	0	3	0
4	D	24	0	0	2	0
4	E	39	0	0	4	0
4	F	14	0	0	2	0
4	G	13	0	0	2	0
4	H	4	0	0	0	0
All	All	19724	0	19286	691	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 18.

The worst 5 of 691 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:E:161:TYR:O	1:E:164:ARG:HD2	1.42	1.16
1:G:209:MET:HB2	2:H:1443:ARG:HB3	1.30	1.11
1:C:185:ARG:HH11	1:C:185:ARG:HB3	1.06	1.08

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:C:166:ASN:H	1:C:166:ASN:ND2	1.46	1.08
1:G:169:GLU:HB3	1:G:597:ARG:HD2	1.40	1.03

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	440/598 (74%)	412 (94%)	22 (5%)	6 (1%)	11	20
1	C	435/598 (73%)	409 (94%)	18 (4%)	8 (2%)	8	14
1	E	435/598 (73%)	414 (95%)	16 (4%)	5 (1%)	14	26
1	G	430/598 (72%)	400 (93%)	19 (4%)	11 (3%)	5	8
2	B	171/180 (95%)	163 (95%)	7 (4%)	1 (1%)	25	43
2	D	176/180 (98%)	166 (94%)	8 (4%)	2 (1%)	14	26
2	F	174/180 (97%)	164 (94%)	7 (4%)	3 (2%)	9	16
2	H	177/180 (98%)	168 (95%)	8 (4%)	1 (1%)	25	43
All	All	2438/3112 (78%)	2296 (94%)	105 (4%)	37 (2%)	10	18

5 of 37 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	158	SER
1	A	517	ASP
1	A	585	GLU
1	C	157	PRO
1	C	199	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	389/527 (74%)	361 (93%)	28 (7%)	14	28
1	C	389/527 (74%)	356 (92%)	33 (8%)	10	21
1	E	387/527 (73%)	350 (90%)	37 (10%)	8	16
1	G	385/527 (73%)	347 (90%)	38 (10%)	8	15
2	B	162/168 (96%)	147 (91%)	15 (9%)	9	17
2	D	167/168 (99%)	152 (91%)	15 (9%)	9	19
2	F	165/168 (98%)	149 (90%)	16 (10%)	8	16
2	H	168/168 (100%)	154 (92%)	14 (8%)	11	22
All	All	2212/2780 (80%)	2016 (91%)	196 (9%)	9	19

5 of 196 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	E	452	GLN
2	F	1441	LEU
1	E	538	LEU
2	F	1311	SER
1	G	190	ASN

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 50 such sidechains are listed below:

Mol	Chain	Res	Type
1	E	419	HIS
1	G	256	GLN
2	H	1439	GLN
1	E	452	GLN
2	F	1363	GLN

5.3.3 RNA (i)

There are no RNA molecules in this entry.

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

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

5.5 Carbohydrates (i)

There are no monosaccharides in this entry.

5.6 Ligand geometry (i)

Of 8 ligands modelled in this entry, 8 are 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	442/598 (73%)	-0.07	12 (2%)	54	58	18, 42, 87, 110	0
1	C	441/598 (73%)	-0.12	14 (3%)	47	51	16, 35, 86, 106	0
1	E	439/598 (73%)	-0.14	6 (1%)	75	77	20, 38, 86, 105	0
1	G	436/598 (72%)	0.03	9 (2%)	63	66	31, 53, 91, 115	0
2	B	173/180 (96%)	0.45	18 (10%)	6	6	15, 55, 104, 116	0
2	D	178/180 (98%)	0.22	14 (7%)	12	12	17, 50, 96, 110	0
2	F	176/180 (97%)	0.03	7 (3%)	38	41	26, 48, 89, 104	0
2	H	179/180 (99%)	0.89	36 (20%)	1	0	39, 71, 105, 118	0
All	All	2464/3112 (79%)	0.06	116 (4%)	31	33	15, 46, 94, 118	0

The worst 5 of 116 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	B	1300	GLY	8.4
1	C	152	SER	8.3
1	G	580	ALA	7.4
2	D	1300	GLY	6.5
2	H	1412	HIS	5.8

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

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

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
3	ZN	H	1501	1/1	0.93	0.05	88,88,88,88	0
3	ZN	F	1502	1/1	0.96	0.10	62,62,62,62	0
3	ZN	B	1501	1/1	0.96	0.04	82,82,82,82	0
3	ZN	H	1502	1/1	0.98	0.09	64,64,64,64	0
3	ZN	F	1501	1/1	0.99	0.07	53,53,53,53	0
3	ZN	D	1501	1/1	0.99	0.06	66,66,66,66	0
3	ZN	B	1502	1/1	1.00	0.14	37,37,37,37	0
3	ZN	D	1502	1/1	1.00	0.10	43,43,43,43	0

6.5 Other polymers (i)

There are no such residues in this entry.