wwPDB X-ray Structure Validation Summary Report (i)

Feb 14, 2024 – 04:02 AM EST

PDB ID	:	3MLQ
Title	:	Crystal structure of the Thermus thermophilus transcription-repair coupling factor RNA polymerase interacting domain with the Thermus aquaticus RNA polymerase beta1 domain
Authors	:	Darst, S.A.; Westblade, L.F.; Campbell, E.A.
Deposited on	:	2010-04-18
Resolution	:	2.91 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:

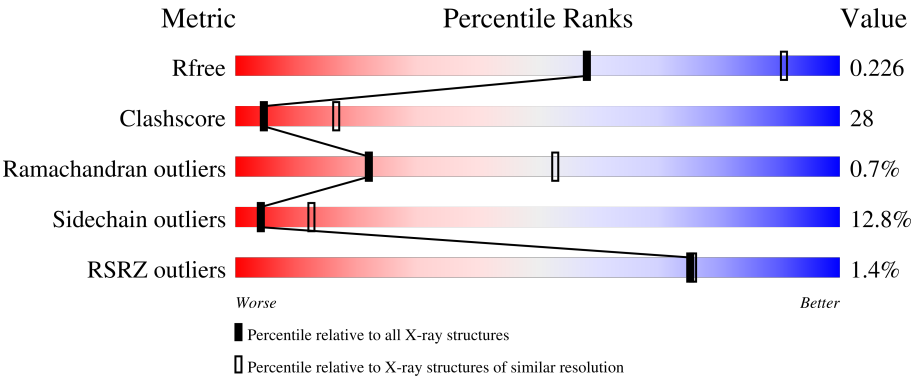
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, 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 2.91 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.



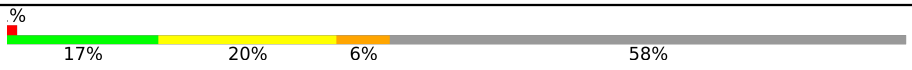
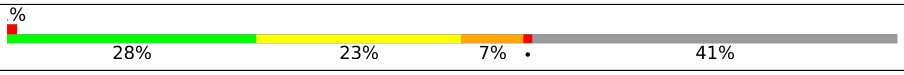
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	2307 (2.94-2.90)
Clashscore	141614	2531 (2.94-2.90)
Ramachandran outliers	138981	2462 (2.94-2.90)
Sidechain outliers	138945	2464 (2.94-2.90)
RSRZ outliers	127900	2248 (2.94-2.90)

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	188	56% 34% 8%
1	B	188	54% 36% 5%
1	C	188	59% 32% 5%
1	D	188	56% 36% 6%
2	E	71	3% 20% 24% 7% 46%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
2	F	71	
2	G	71	
2	H	71	

2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 6863 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-directed RNA polymerase subunit beta.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	186	Total	C	N	O	S	0	0	0
			1450	915	254	276	5			
1	B	180	Total	C	N	O	S	0	0	0
			1413	893	246	269	5			
1	C	182	Total	C	N	O	S	0	0	0
			1419	896	248	270	5			
1	D	185	Total	C	N	O	S	0	0	0
			1423	899	244	274	6			

There are 12 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	16	MET	-	initiating methionine	UNP Q9KWU7
A	332	GLY	-	linker	UNP Q9KWU7
A	333	GLY	-	linker	UNP Q9KWU7
B	16	MET	-	initiating methionine	UNP Q9KWU7
B	332	GLY	-	linker	UNP Q9KWU7
B	333	GLY	-	linker	UNP Q9KWU7
C	16	MET	-	initiating methionine	UNP Q9KWU7
C	332	GLY	-	linker	UNP Q9KWU7
C	333	GLY	-	linker	UNP Q9KWU7
D	16	MET	-	initiating methionine	UNP Q9KWU7
D	332	GLY	-	linker	UNP Q9KWU7
D	333	GLY	-	linker	UNP Q9KWU7

- Molecule 2 is a protein called Transcription-repair coupling factor.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
2	E	38	Total	C	N	O	0	0	0
			290	192	50	48			
2	F	30	Total	C	N	O	0	0	0
			235	154	39	42			

Continued on next page...

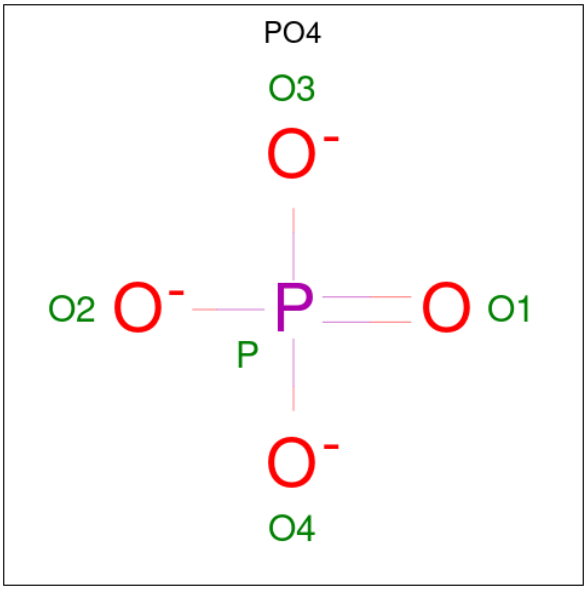
Continued from previous page...

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
2	G	40	Total	C	N	O	0	0	0
			296	191	50	55			
2	H	42	Total	C	N	O	0	0	0
			332	217	58	57			

There are 16 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
E	317	GLY	-	expression tag	UNP Q72KB4
E	318	PRO	-	expression tag	UNP Q72KB4
E	319	HIS	-	expression tag	UNP Q72KB4
E	320	MET	-	expression tag	UNP Q72KB4
F	317	GLY	-	expression tag	UNP Q72KB4
F	318	PRO	-	expression tag	UNP Q72KB4
F	319	HIS	-	expression tag	UNP Q72KB4
F	320	MET	-	expression tag	UNP Q72KB4
G	317	GLY	-	expression tag	UNP Q72KB4
G	318	PRO	-	expression tag	UNP Q72KB4
G	319	HIS	-	expression tag	UNP Q72KB4
G	320	MET	-	expression tag	UNP Q72KB4
H	317	GLY	-	expression tag	UNP Q72KB4
H	318	PRO	-	expression tag	UNP Q72KB4
H	319	HIS	-	expression tag	UNP Q72KB4
H	320	MET	-	expression tag	UNP Q72KB4

- Molecule 3 is PHOSPHATE ION (three-letter code: PO4) (formula: O_4P).



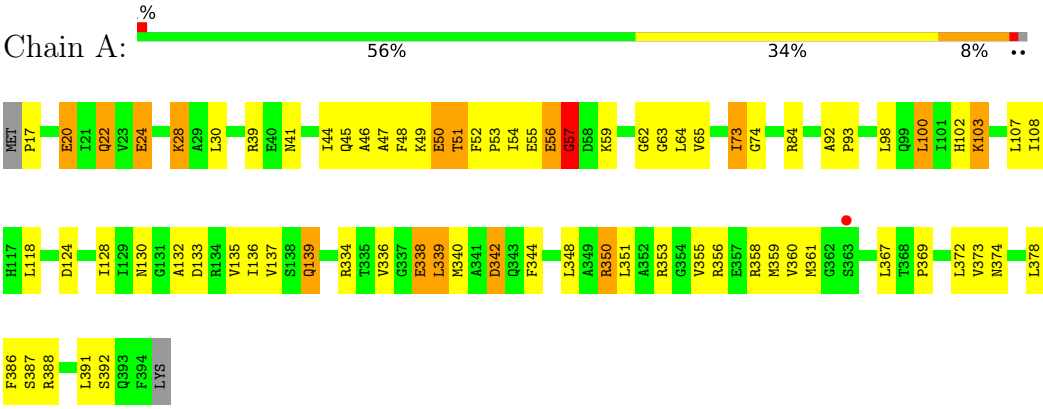
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
			Total	O	P		
3	A	1	5	4	1	0	0

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.

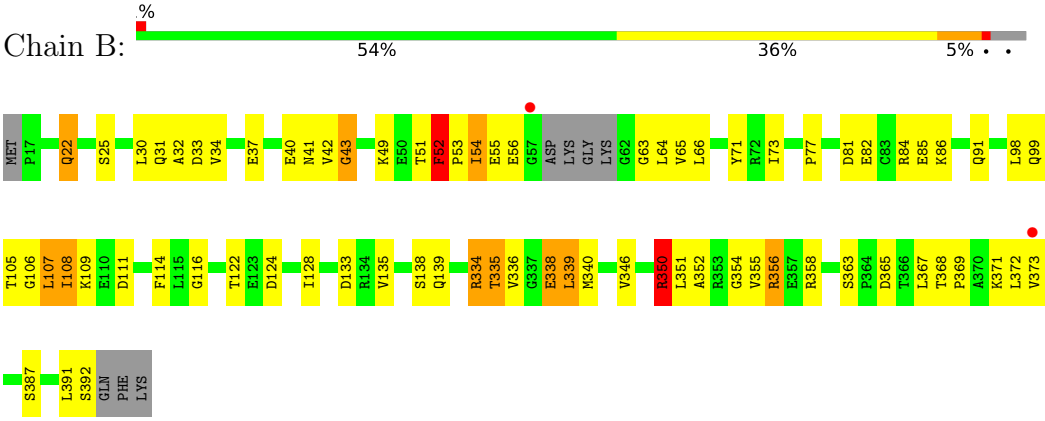
• Molecule 1: DNA-directed RNA polymerase subunit beta

Chain A:



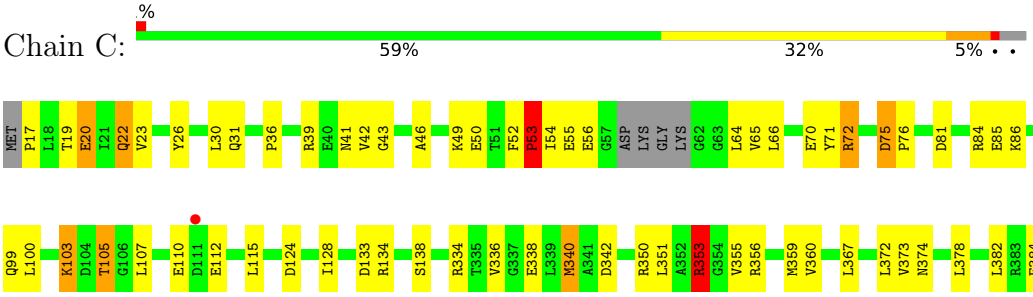
• Molecule 1: DNA-directed RNA polymerase subunit beta

Chain B:



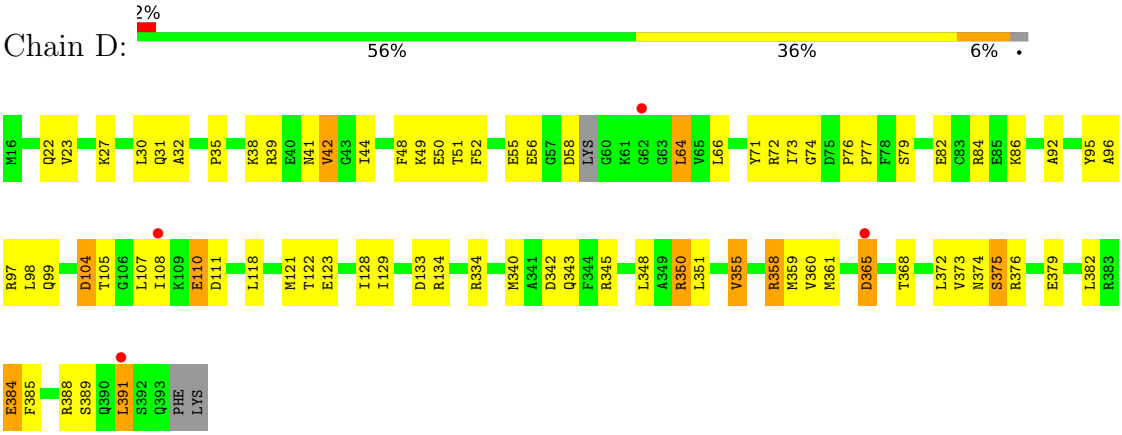
• Molecule 1: DNA-directed RNA polymerase subunit beta

Chain C:



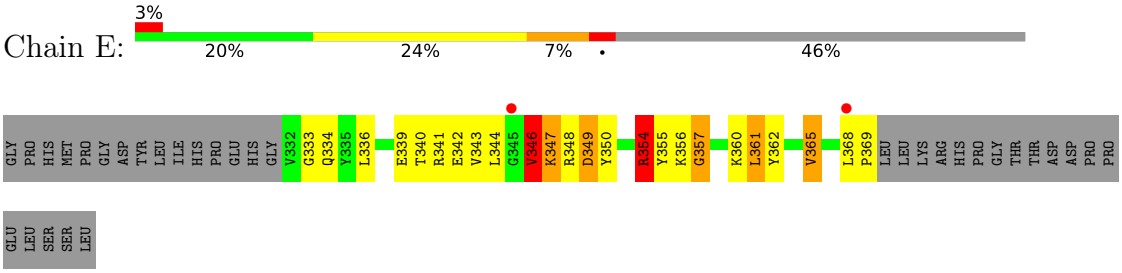
• Molecule 1: DNA-directed RNA polymerase subunit beta

Chain D:



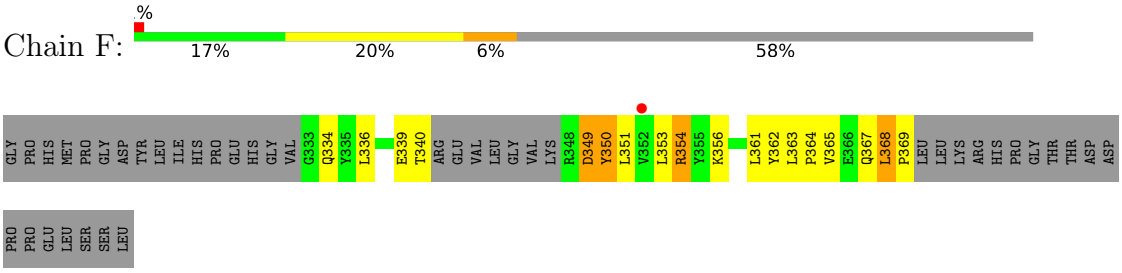
• Molecule 2: Transcription-repair coupling factor

Chain E:



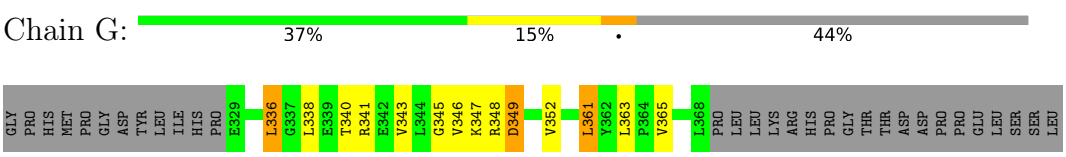
• Molecule 2: Transcription-repair coupling factor

Chain F:



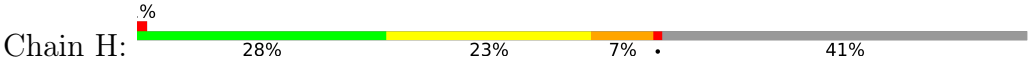
• Molecule 2: Transcription-repair coupling factor

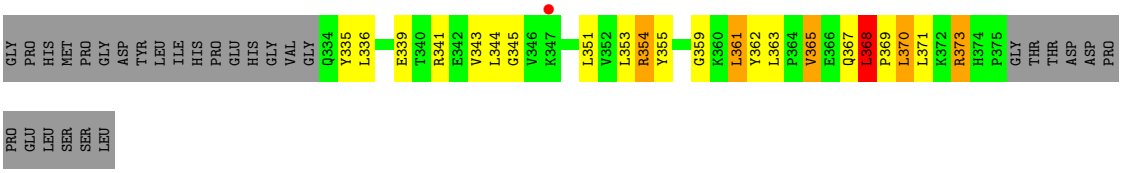
Chain G:



• Molecule 2: Transcription-repair coupling factor

Chain H:




GLY PRO HIS MET PRO GLY ASP TYR LEU ILE HIS PRO GLU HIS GLY VAL GLY Q334 Y335 L336 E339 T340 R341 E342 V343 L344 G345 V346 K347 L351 V352 L353 R354 Y355 G359 K360 L361 Y362 L363 P364 V365 E366 Q367 L368 P369 L370 L371 K372 R373 H374 P375 GLY THR THR ASP ASP PRO
PRO GLU LEU SER SER LEU

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 43	Depositor
Cell constants a, b, c, α, β, γ	106.58Å 106.58Å 122.30Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	29.56 – 2.91 29.56 – 2.91	Depositor EDS
% Data completeness (in resolution range)	88.8 (29.56-2.91) 88.9 (29.56-2.91)	Depositor EDS
R_{merge}	0.09	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	3.12 (at 2.90Å)	Xtriage
Refinement program	REFMAC 5.5.0088, CNS	Depositor
R, R_{free}	0.227 , 0.250 0.227 , 0.226	Depositor DCC
R_{free} test set	1144 reflections (4.28%)	wwPDB-VP
Wilson B-factor (Å^2)	54.4	Xtriage
Anisotropy	0.575	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.31 , 10.6	EDS
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.32$	Xtriage
Estimated twinning fraction	0.467 for h,-k,-l	Xtriage
Reported twinning fraction	0.522 for H, K, L 0.478 for -H, K, -L	Depositor
Outliers	0 of 26709 reflections	Xtriage
F_o,F_c correlation	0.91	EDS
Total number of atoms	6863	wwPDB-VP
Average B, all atoms (Å^2)	52.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.29% 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: PO4

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	0.47	0/1475	0.67	1/1989 (0.1%)
1	B	0.48	0/1437	0.68	1/1937 (0.1%)
1	C	0.50	0/1443	0.62	1/1947 (0.1%)
1	D	0.51	0/1447	0.63	1/1955 (0.1%)
2	E	0.53	0/295	0.96	3/400 (0.8%)
2	F	0.38	0/239	0.81	0/321
2	G	0.34	0/301	0.56	0/408
2	H	0.37	0/338	0.75	0/458
All	All	0.48	0/6975	0.68	7/9415 (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	A	0	3
1	B	0	3
1	C	0	2
2	E	0	2
2	F	0	1
2	H	0	1
All	All	0	12

There are no bond length outliers.

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

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	E	357	GLY	N-CA-C	-7.54	94.25	113.10
1	C	353	ARG	NE-CZ-NH1	6.49	123.55	120.30

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	E	354	ARG	NE-CZ-NH1	6.20	123.40	120.30
1	B	350	ARG	NE-CZ-NH1	6.08	123.34	120.30
1	A	57	GLY	C-N-CA	5.59	135.68	121.70

There are no chirality outliers.

5 of 12 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	139	GLN	Peptide
1	A	52	PHE	Peptide
1	A	57	GLY	Peptide
1	B	51	THR	Peptide
1	B	52	PHE	Peptide

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	1450	0	1450	81	1
1	B	1413	0	1419	76	1
1	C	1419	0	1414	82	0
1	D	1423	0	1401	95	0
2	E	290	0	297	30	0
2	F	235	0	226	25	0
2	G	296	0	269	15	0
2	H	332	0	336	24	0
3	A	5	0	0	0	0
All	All	6863	0	6812	384	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 28.

The worst 5 of 384 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:47:ALA:O	1:A:51:THR:HG22	1.32	1.29

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:B:105:THR:HG21	1:B:107:LEU:HD12	1.31	1.11
2:E:343:VAL:HG23	2:E:344:LEU:HD12	1.36	1.07
1:D:351:LEU:HD13	1:D:374:ASN:O	1.53	1.06
1:D:64:LEU:HD11	1:D:359:MET:HB3	1.32	1.06

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:353:ARG:NH2	1:B:392:SER:OG[4_665]	1.95	0.25

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	184/188 (98%)	173 (94%)	10 (5%)	1 (0%)	29 60
1	B	176/188 (94%)	170 (97%)	4 (2%)	2 (1%)	14 41
1	C	178/188 (95%)	171 (96%)	6 (3%)	1 (1%)	25 57
1	D	181/188 (96%)	168 (93%)	13 (7%)	0	100 100
2	E	36/71 (51%)	33 (92%)	2 (6%)	1 (3%)	5 17
2	F	26/71 (37%)	23 (88%)	3 (12%)	0	100 100
2	G	38/71 (54%)	35 (92%)	3 (8%)	0	100 100
2	H	40/71 (56%)	37 (92%)	2 (5%)	1 (2%)	5 20
All	All	859/1036 (83%)	810 (94%)	43 (5%)	6 (1%)	22 53

5 of 6 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	53	PRO

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	B	43	GLY
1	B	53	PRO
1	C	53	PRO
2	H	368	LEU

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	154/159 (97%)	136 (88%)	18 (12%)	5 15
1	B	152/159 (96%)	134 (88%)	18 (12%)	5 15
1	C	151/159 (95%)	131 (87%)	20 (13%)	4 11
1	D	150/159 (94%)	131 (87%)	19 (13%)	4 12
2	E	28/62 (45%)	23 (82%)	5 (18%)	2 5
2	F	22/62 (36%)	18 (82%)	4 (18%)	1 4
2	G	26/62 (42%)	23 (88%)	3 (12%)	5 16
2	H	33/62 (53%)	28 (85%)	5 (15%)	3 8
All	All	716/884 (81%)	624 (87%)	92 (13%)	4 12

5 of 92 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	D	42	VAL
1	D	376	ARG
1	D	64	LEU
1	D	350	ARG
2	E	346	VAL

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 19 such sidechains are listed below:

Mol	Chain	Res	Type
1	D	31	GLN

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
2	E	334	GLN
2	H	334	GLN
1	D	343	GLN
1	B	91	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)

1 ligand 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	PO4	A	1	-	4,4,4	0.80	0	6,6,6	0.52	0

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

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, 95^{th} 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	186/188 (98%)	0.20	1 (0%) 91 91	38, 48, 62, 74	0
1	B	180/188 (95%)	0.21	2 (1%) 80 81	39, 46, 65, 69	0
1	C	182/188 (96%)	0.18	1 (0%) 91 91	38, 48, 61, 70	0
1	D	185/188 (98%)	0.24	4 (2%) 62 60	39, 47, 68, 78	0
2	E	38/71 (53%)	0.36	2 (5%) 26 23	54, 62, 83, 89	0
2	F	30/71 (42%)	0.48	1 (3%) 46 42	58, 65, 81, 85	0
2	G	40/71 (56%)	0.18	0 100 100	52, 61, 84, 89	0
2	H	42/71 (59%)	0.10	1 (2%) 59 57	58, 72, 118, 135	0
All	All	883/1036 (85%)	0.22	12 (1%) 75 76	38, 49, 72, 135	0

The worst 5 of 12 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	E	368	LEU	3.3
1	D	365	ASP	2.8
1	A	363	SER	2.5
2	H	347	LYS	2.4
2	E	345	GLY	2.4

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, 95^{th} 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	PO4	A	1	5/5	0.98	0.17	53,53,54,54	0

6.5 Other polymers

There are no such residues in this entry.