wwPDB X-ray Structure Validation Summary Report (i)

Feb 19, 2024 – 03:49 PM EST

PDB ID	:	4KIS
Title	:	Crystal Structure of a LSR-DNA Complex
Authors	:	Rutherford, K.; Yuan, P.; Perry, K.; Van Duyne, G.D.
Deposited on	:	2013-05-02
Resolution	:	3.20 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:

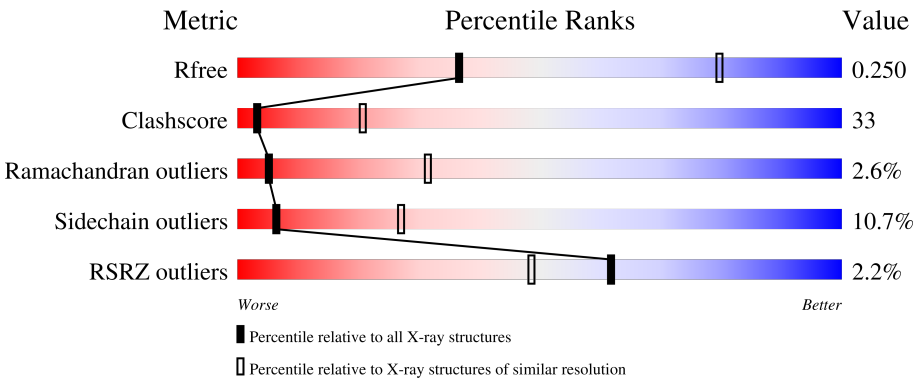
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation 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 3.20 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	1133 (3.20-3.20)
Clashscore	141614	1253 (3.20-3.20)
Ramachandran outliers	138981	1234 (3.20-3.20)
Sidechain outliers	138945	1233 (3.20-3.20)
RSRZ outliers	127900	1095 (3.20-3.20)

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	328	3%, 42%, 49%, 5%, •
1	B	328	%, 43%, 39%, 10%, 7%
1	C	328	4%, 36%, 50%, 8%, •, 6%
1	D	328	2%, 44%, 37%, 7%, 12%
2	E	26	23%, 77%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
2	G	26	58% 42%
2	I	26	23% 77%
2	K	26	23% 77%
3	F	26	35% 58% 8%
3	H	26	27% 69% .
3	J	26	23% 77%
3	L	26	19% 73% 8%

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
5	CA	A	505	-	-	-	X

2 Entry composition (i)

There are 6 unique types of molecules in this entry. The entry contains 14347 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 Putative integrase [Bacteriophage A118].

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	315	2651	1683	464	493	11	0	0	0
1	B	305	2557	1630	439	477	11	0	0	0
1	C	309	2589	1646	448	484	11	0	0	0
1	D	287	2289	1454	413	413	9	0	0	0

There are 32 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	453	LEU	-	expression tag	UNP Q928V6
A	454	GLU	-	expression tag	UNP Q928V6
A	455	HIS	-	expression tag	UNP Q928V6
A	456	HIS	-	expression tag	UNP Q928V6
A	457	HIS	-	expression tag	UNP Q928V6
A	458	HIS	-	expression tag	UNP Q928V6
A	459	HIS	-	expression tag	UNP Q928V6
A	460	HIS	-	expression tag	UNP Q928V6
B	453	LEU	-	expression tag	UNP Q928V6
B	454	GLU	-	expression tag	UNP Q928V6
B	455	HIS	-	expression tag	UNP Q928V6
B	456	HIS	-	expression tag	UNP Q928V6
B	457	HIS	-	expression tag	UNP Q928V6
B	458	HIS	-	expression tag	UNP Q928V6
B	459	HIS	-	expression tag	UNP Q928V6
B	460	HIS	-	expression tag	UNP Q928V6
C	453	LEU	-	expression tag	UNP Q928V6
C	454	GLU	-	expression tag	UNP Q928V6
C	455	HIS	-	expression tag	UNP Q928V6
C	456	HIS	-	expression tag	UNP Q928V6
C	457	HIS	-	expression tag	UNP Q928V6

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
C	458	HIS	-	expression tag	UNP Q928V6
C	459	HIS	-	expression tag	UNP Q928V6
C	460	HIS	-	expression tag	UNP Q928V6
D	453	LEU	-	expression tag	UNP Q928V6
D	454	GLU	-	expression tag	UNP Q928V6
D	455	HIS	-	expression tag	UNP Q928V6
D	456	HIS	-	expression tag	UNP Q928V6
D	457	HIS	-	expression tag	UNP Q928V6
D	458	HIS	-	expression tag	UNP Q928V6
D	459	HIS	-	expression tag	UNP Q928V6
D	460	HIS	-	expression tag	UNP Q928V6

- Molecule 2 is a DNA chain called DNA (26-MER).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
2	E	26	522	254	76	167	25	0	0	0
2	G	26	522	254	76	167	25	0	0	0
2	I	26	522	254	76	167	25	0	0	0
2	K	26	522	254	76	167	25	0	0	0

- Molecule 3 is a DNA chain called DNA (26-MER).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
3	F	26	538	255	117	141	25	0	0	0
3	H	26	538	255	117	141	25	0	0	0
3	J	26	538	255	117	141	25	0	0	0
3	L	26	538	255	117	141	25	0	0	0

- Molecule 4 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	Zn		
4	A	3	3	3	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	B	1	Total Zn 1 1	0	0
4	C	1	Total Zn 1 1	0	0
4	D	1	Total Zn 1 1	0	0

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

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	A	2	Total Ca 2 2	0	0
5	E	1	Total Ca 1 1	0	0
5	B	2	Total Ca 2 2	0	0
5	C	1	Total Ca 1 1	0	0
5	D	1	Total Ca 1 1	0	0

- Molecule 6 is water.

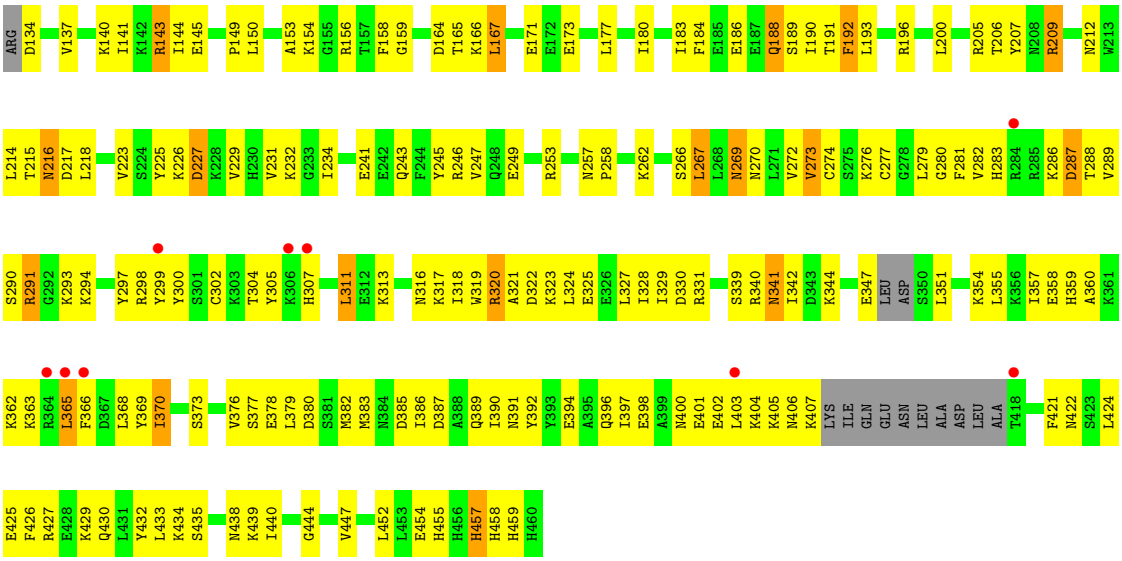
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
6	E	2	Total O 2 2	0	0
6	F	1	Total O 1 1	0	0
6	G	2	Total O 2 2	0	0
6	H	1	Total O 1 1	0	0
6	I	1	Total O 1 1	0	0
6	J	1	Total O 1 1	0	0

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.

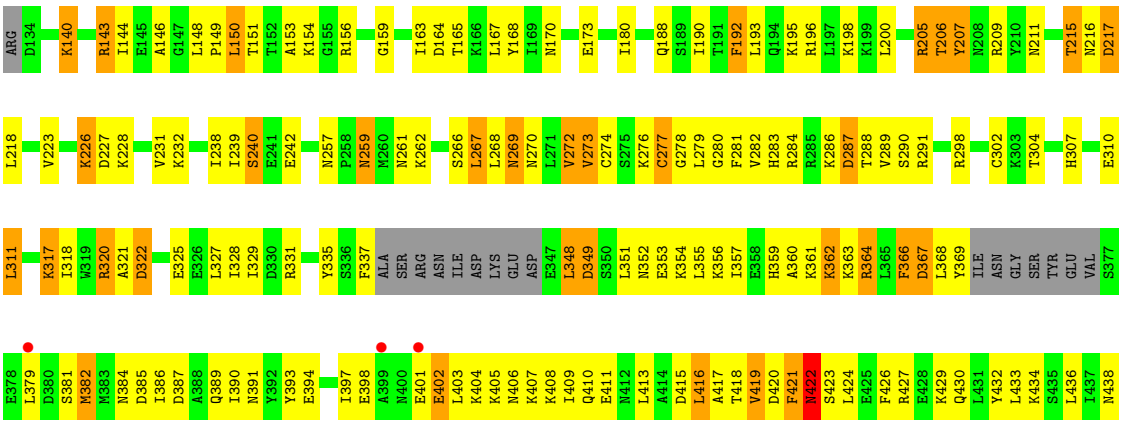
• Molecule 1: Putative integrase [Bacteriophage A118]

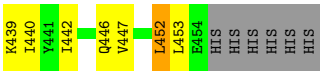
Chain A:



• Molecule 1: Putative integrase [Bacteriophage A118]

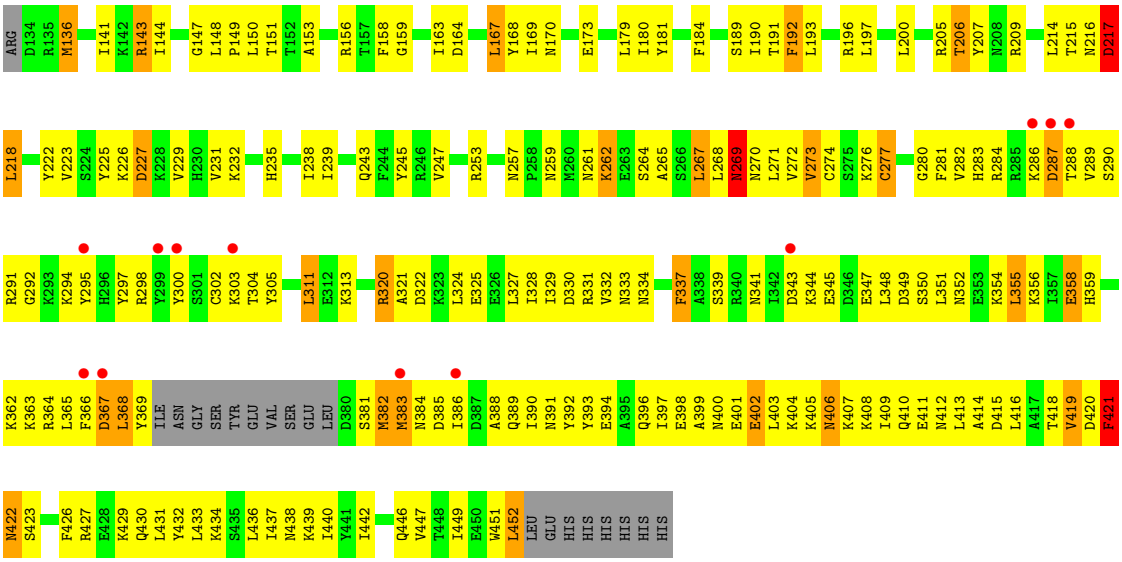
Chain B:





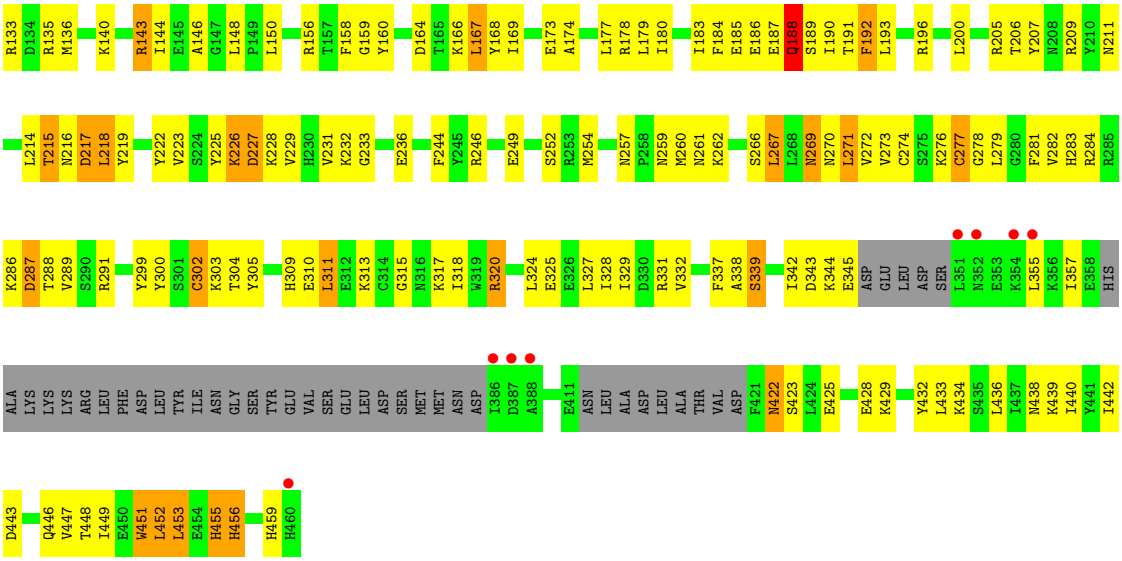
• Molecule 1: Putative integrase [Bacteriophage A118]

Chain C:



• Molecule 1: Putative integrase [Bacteriophage A118]

Chain D:



• Molecule 2: DNA (26-MER)

Chain E:

• Molecule 2: DNA (26-MER)

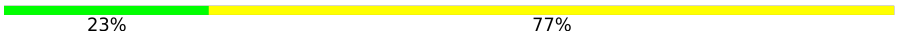
Chain G:

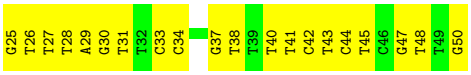
• Molecule 2: DNA (26-MER)

Chain I:

• Molecule 2: DNA (26-MER)

Chain K:



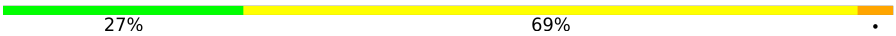


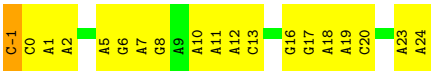
• Molecule 3: DNA (26-MER)

Chain F:

• Molecule 3: DNA (26-MER)

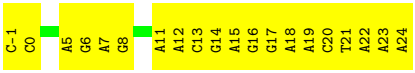
Chain H:





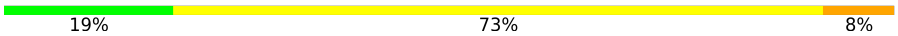
• Molecule 3: DNA (26-MER)

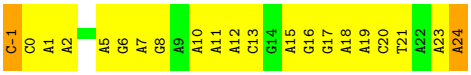
Chain J:



• Molecule 3: DNA (26-MER)

Chain L:




C-1
C0
A1
A2
A5
G6
A7
G8
A9
A10
A11
A12
C13
G14
A15
G16
G17
A18
A19
C20
T21
A22
A23
A24

4 Data and refinement statistics (i)

Property	Value	Source
Space group	I 2 3	Depositor
Cell constants a, b, c, α, β, γ	290.75Å 290.75Å 290.75Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	49.86 – 3.20 49.86 – 3.15	Depositor EDS
% Data completeness (in resolution range)	(Not available) (49.86-3.20) 96.7 (49.86-3.15)	Depositor EDS
R_{merge}	0.12	Depositor
R_{sym}	0.12	Depositor
$< I/\sigma(I) >$ [1]	1.24 (at 3.12Å)	Xtriage
Refinement program	CNS 1.3	Depositor
R, R_{free}	0.236 , 0.256 0.219 , 0.250	Depositor DCC
R_{free} test set	3353 reflections (4.94%)	wwPDB-VP
Wilson B-factor (Å^2)	92.3	Xtriage
Anisotropy	0.000	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.27 , 66.5	EDS
L-test for twinning[2]	$< \|L\| > = 0.47$, $< L^2 > = 0.30$	Xtriage
Estimated twinning fraction	0.002 for -l,-k,-h	Xtriage
F_o,F_c correlation	0.93	EDS
Total number of atoms	14347	wwPDB-VP
Average B, all atoms (Å^2)	100.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 8.19% 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, CA

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	0/2702	0.69	0/3620
1	B	0.49	0/2601	0.72	0/3485
1	C	0.41	0/2634	0.66	0/3530
1	D	0.44	0/2333	0.70	1/3136 (0.0%)
2	E	0.57	0/579	0.83	0/892
2	G	0.60	0/579	0.84	0/892
2	I	0.42	0/579	0.75	0/892
2	K	0.57	0/579	0.80	0/892
3	F	0.52	0/609	0.71	0/938
3	H	0.60	0/609	0.76	0/938
3	J	0.39	0/609	0.70	0/938
3	L	0.53	0/609	0.74	0/938
All	All	0.47	0/15022	0.72	1/21091 (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	B	0	1
1	D	0	1
3	F	0	2
3	H	0	1
3	L	0	2
All	All	0	7

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	D	451	TRP	N-CA-C	5.61	126.16	111.00

There are no chirality outliers.

5 of 7 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	B	168	TYR	Sidechain
1	D	168	TYR	Sidechain
3	F	-1	DC	Sidechain
3	F	16	DG	Sidechain
3	H	-1	DC	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	2651	0	2628	179	0
1	B	2557	0	2565	169	0
1	C	2589	0	2594	213	0
1	D	2289	0	2170	149	0
2	E	522	0	302	44	0
2	G	522	0	302	17	0
2	I	522	0	302	35	0
2	K	522	0	302	29	0
3	F	538	0	289	28	0
3	H	538	0	289	25	0
3	J	538	0	289	24	0
3	L	538	0	289	33	0
4	A	3	0	0	0	0
4	B	1	0	0	0	0
4	C	1	0	0	0	0
4	D	1	0	0	0	0
5	A	2	0	0	0	0
5	B	2	0	0	0	0
5	C	1	0	0	0	0
5	D	1	0	0	0	0
5	E	1	0	0	0	0
6	E	2	0	0	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
6	F	1	0	0	0	0
6	G	2	0	0	0	0
6	H	1	0	0	0	0
6	I	1	0	0	0	0
6	J	1	0	0	0	0
All	All	14347	0	12321	880	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 33.

The worst 5 of 880 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:C:415:ASP:HA	1:C:418:THR:HG23	1.36	1.08
3:J:18:DA:H2"	3:J:19:DA:H5"	1.37	1.07
2:E:30:DG:H2"	2:E:31:DT:H5"	1.37	1.02
2:K:25:DG:H2"	2:K:26:DT:H5'	1.39	1.01
2:I:44:DC:H2"	2:I:45:DT:H5"	1.43	1.00

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	309/328 (94%)	263 (85%)	43 (14%)	3 (1%)	15 54
1	B	299/328 (91%)	256 (86%)	34 (11%)	9 (3%)	4 28
1	C	305/328 (93%)	251 (82%)	47 (15%)	7 (2%)	6 34
1	D	279/328 (85%)	228 (82%)	39 (14%)	12 (4%)	2 20
All	All	1192/1312 (91%)	998 (84%)	163 (14%)	31 (3%)	5 31

5 of 31 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	341	ASN
1	B	227	ASP
1	B	278	GLY
1	B	419	VAL
1	B	421	PHE

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	291/302 (96%)	266 (91%)	25 (9%)	10	38
1	B	281/302 (93%)	246 (88%)	35 (12%)	4	21
1	C	284/302 (94%)	251 (88%)	33 (12%)	5	24
1	D	228/302 (76%)	205 (90%)	23 (10%)	7	29
All	All	1084/1208 (90%)	968 (89%)	116 (11%)	6	27

5 of 116 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	416	LEU
1	D	311	LEU
1	C	273	VAL
1	D	302	CYS
1	D	217	ASP

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	B	384	ASN
1	D	270	ASN
1	C	270	ASN
1	D	230	HIS
1	C	269	ASN

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 13 ligands modelled in this entry, 13 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	315/328 (96%)	-0.01	9 (2%)	51	36	45, 87, 156, 171	3 (0%)
1	B	305/328 (92%)	-0.24	3 (0%)	82	72	43, 81, 149, 167	2 (0%)
1	C	309/328 (94%)	0.08	12 (3%)	39	25	52, 117, 167, 190	0
1	D	287/328 (87%)	0.00	8 (2%)	53	37	51, 91, 166, 184	1 (0%)
2	E	26/26 (100%)	-0.45	0	100	100	68, 87, 115, 116	0
2	G	26/26 (100%)	-0.50	0	100	100	64, 82, 99, 109	0
2	I	26/26 (100%)	-0.41	0	100	100	89, 116, 157, 165	0
2	K	26/26 (100%)	-0.39	0	100	100	77, 93, 118, 121	0
3	F	26/26 (100%)	-0.30	0	100	100	69, 96, 128, 135	0
3	H	26/26 (100%)	-0.52	0	100	100	69, 85, 108, 111	0
3	J	26/26 (100%)	-0.43	0	100	100	96, 120, 174, 188	0
3	L	26/26 (100%)	-0.42	0	100	100	72, 96, 123, 131	0
All	All	1424/1520 (93%)	-0.10	32 (2%)	62	48	43, 94, 161, 190	6 (0%)

The worst 5 of 32 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	D	386	ILE	4.4
1	C	366	PHE	3.8
1	D	460	HIS	3.7
1	C	295	TYR	3.6
1	A	365	LEU	3.5

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
5	CA	A	505	1/1	0.28	0.50	160,160,160,160	0
5	CA	B	503	1/1	0.81	0.61	121,121,121,121	0
5	CA	A	504	1/1	0.82	0.39	132,132,132,132	0
5	CA	C	502	1/1	0.84	0.65	145,145,145,145	0
5	CA	D	502	1/1	0.88	0.98	145,145,145,145	0
5	CA	E	101	1/1	0.92	0.47	118,118,118,118	0
5	CA	B	502	1/1	0.95	0.66	125,125,125,125	0
4	ZN	A	502	1/1	0.96	0.14	87,87,87,87	0
4	ZN	C	501	1/1	0.97	0.08	144,144,144,144	0
4	ZN	D	501	1/1	0.99	0.12	116,116,116,116	0
4	ZN	A	503	1/1	0.99	0.13	110,110,110,110	0
4	ZN	B	501	1/1	1.00	0.15	94,94,94,94	0
4	ZN	A	501	1/1	1.00	0.10	106,106,106,106	0

6.5 Other polymers (i)

There are no such residues in this entry.