Full wwPDB X-ray Structure Validation Report ⓘ

Mar 8, 2026 – 09:31 AM UTC

PDB ID	:	6VRG / pdb_00006vrg
Title	:	Structure of HIV-1 integrase with native amino-terminal sequence
Authors	:	Eilers, G.; Gupta, K.; Allen, A.; Zhou, J.; Hwang, Y.; Cory, M.; Bushman, F.D.; Van Duyne, G.D.
Deposited on	:	2020-02-07
Resolution	:	2.40 Å(reported)

This is a Full wwPDB X-ray Structure Validation 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 ⓘ 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 ⓘ) were used in the production of this report:

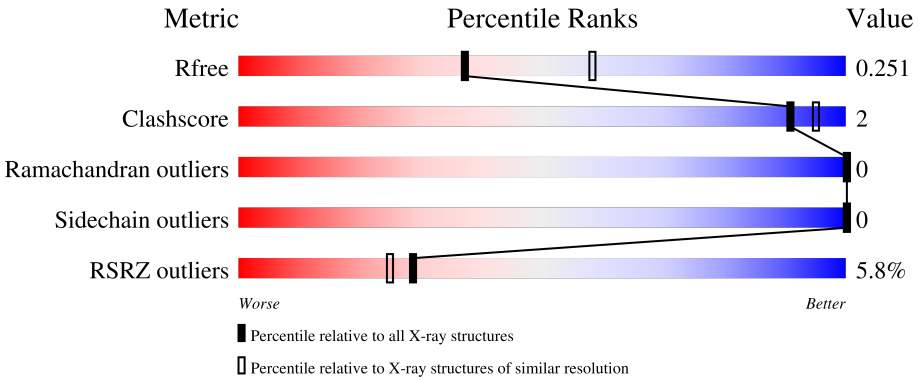
MolProbity	:	4-5-2 with Phenix2.0
Mogul	:	2022.3.0, CSD as543be (2022)
Xtriage (Phenix)	:	2.0
EDS	:	3.0
Percentile statistics	:	20250101.v01 (using entries in the PDB archive January 1st 2025)
CCP4	:	9.0.010 (Gargrove)
Density-Fitness	:	1.0.12
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.49

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 2.40 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}	180053	4912 (2.40-2.40)
Clashscore	190562	5391 (2.40-2.40)
Ramachandran outliers	187476	5320 (2.40-2.40)
Sidechain outliers	187428	5321 (2.40-2.40)
RSRZ outliers	180081	4916 (2.40-2.40)

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	212	2% 85% 5% 10%
1	B	212	4% 86% · 11%
1	C	212	7% 83% 5% 12%
1	D	212	8% 86% · 11%

2 Entry composition (i)

There are 5 unique types of molecules in this entry. The entry contains 11916 atoms, of which 5832 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 Integrase.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	191	Total	C	H	N	O	S	0	0	0
			2953	933	1478	256	278	8			
1	B	189	Total	C	H	N	O	S	0	0	0
			2925	925	1463	253	276	8			
1	C	186	Total	C	H	N	O	S	0	0	0
			2874	909	1436	250	271	8			
1	D	188	Total	C	H	N	O	S	0	0	0
			2910	921	1455	252	274	8			

There are 12 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	131	ASP	TRP	engineered mutation	UNP F2WR39
A	139	ASP	PHE	engineered mutation	UNP F2WR39
A	185	LYS	PHE	engineered mutation	UNP F2WR39
B	131	ASP	TRP	engineered mutation	UNP F2WR39
B	139	ASP	PHE	engineered mutation	UNP F2WR39
B	185	LYS	PHE	engineered mutation	UNP F2WR39
C	131	ASP	TRP	engineered mutation	UNP F2WR39
C	139	ASP	PHE	engineered mutation	UNP F2WR39
C	185	LYS	PHE	engineered mutation	UNP F2WR39
D	131	ASP	TRP	engineered mutation	UNP F2WR39
D	139	ASP	PHE	engineered mutation	UNP F2WR39
D	185	LYS	PHE	engineered mutation	UNP F2WR39

- Molecule 2 is ZINC ION (CCD ID: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	1	Total	Zn	0	0
			1	1		
2	B	1	Total	Zn	0	0
			1	1		

Continued on next page...

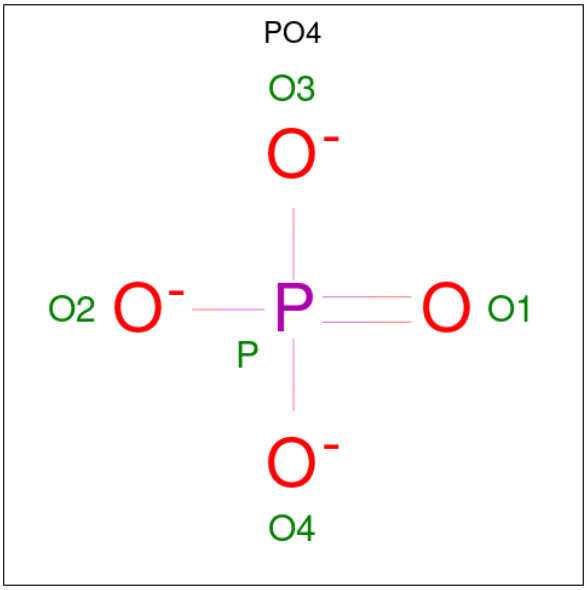
Continued from previous page...

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

- Molecule 3 is POTASSIUM ION (CCD ID: K) (formula: K).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	1	Total K 1 1	0	0
3	B	1	Total K 1 1	0	0
3	C	1	Total K 1 1	0	0
3	D	1	Total K 1 1	0	0

- Molecule 4 is PHOSPHATE ION (CCD ID: PO4) (formula: O_4P).



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

Continued on next page...

Continued from previous page...

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

- Molecule 5 is water.

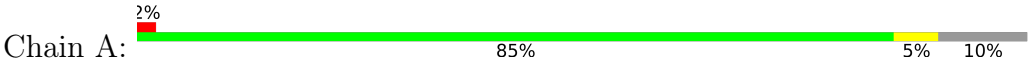
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	A	72	Total O 72 72	0	0
5	B	63	Total O 63 63	0	0
5	C	46	Total O 46 46	0	0
5	D	45	Total O 45 45	0	0

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.

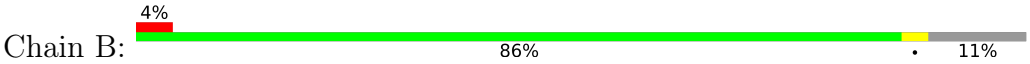
• Molecule 1: Integrase

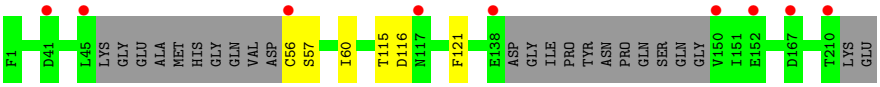
Chain A:



• Molecule 1: Integrase

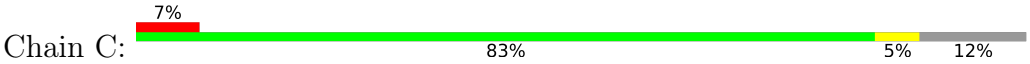
Chain B:

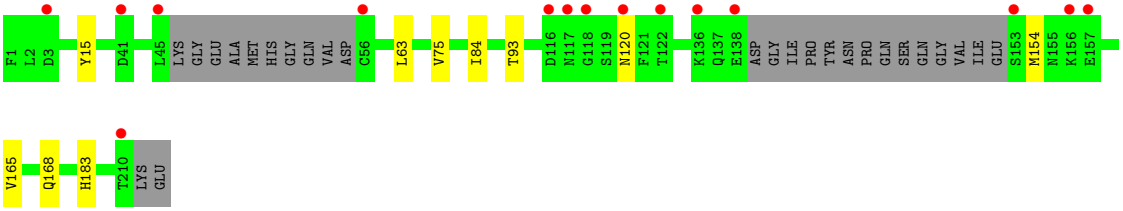




• Molecule 1: Integrase

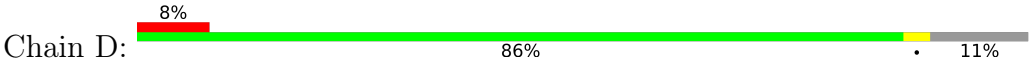
Chain C:

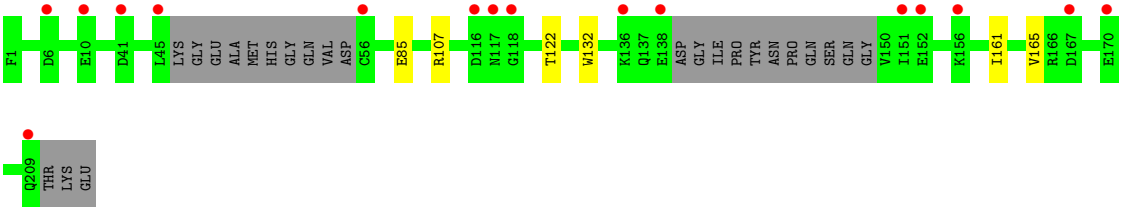




• Molecule 1: Integrase

Chain D:





4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 43 21 2	Depositor
Cell constants a, b, c, α, β, γ	102.92Å 102.92Å 279.20Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	96.57 – 2.40 96.57 – 2.40	Depositor EDS
% Data completeness (in resolution range)	100.0 (96.57-2.40) 99.9 (96.57-2.40)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.82 (at 2.40Å)	Xtriage
Refinement program	PHENIX 1.11.1_2575	Depositor
R, R_{free}	0.225 , 0.253 0.224 , 0.251	Depositor DCC
R_{free} test set	2000 reflections (1.96%)	wwPDB-VP
Wilson B-factor (Å^2)	23.7	Xtriage
Anisotropy	0.331	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.41 , 40.3	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	0.92	EDS
Total number of atoms	11916	wwPDB-VP
Average B, all atoms (Å^2)	31.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.20% 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, K, 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	0.21	0/1501	0.40	0/2026
1	B	0.24	0/1488	0.46	1/2010 (0.0%)
1	C	0.18	0/1464	0.35	0/1977
1	D	0.16	0/1481	0.35	0/2000
All	All	0.20	0/5934	0.39	1/8013 (0.0%)

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	B	116	ASP	CA-CB-CG	5.24	117.84	112.60

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	1475	1478	1476	8	0
1	B	1462	1463	1460	3	0
1	C	1438	1436	1434	6	0
1	D	1455	1455	1453	4	0
2	A	1	0	0	0	0
2	B	1	0	0	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	C	1	0	0	0	0
2	D	1	0	0	0	0
3	A	1	0	0	0	0
3	B	1	0	0	0	0
3	C	1	0	0	0	0
3	D	1	0	0	0	0
4	A	5	0	0	0	0
4	B	5	0	0	0	0
4	C	5	0	0	0	0
4	D	5	0	0	0	0
5	A	72	0	0	0	0
5	B	63	0	0	0	0
5	C	46	0	0	0	0
5	D	45	0	0	0	0
All	All	6084	5832	5823	20	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 2.

All (20) 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:B:56:CYS:SG	1:B:60:ILE:HD12	2.33	0.68
1:A:63:LEU:HD13	1:A:113:VAL:HG13	1.77	0.66
1:B:115:THR:HG21	1:B:121:PHE:CD2	2.32	0.64
1:D:161:ILE:O	1:D:165:VAL:HG22	2.04	0.57
1:C:84:ILE:HD11	1:C:154:MET:SD	2.45	0.56
1:A:62:GLN:C	1:A:63:LEU:HD12	2.32	0.55
1:A:62:GLN:O	1:A:63:LEU:HD12	2.06	0.54
1:D:85:GLU:OE2	1:D:107:ARG:NH1	2.39	0.53
1:C:165:VAL:HG23	1:C:168:GLN:NE2	2.26	0.50
1:A:101:LEU:HD21	1:A:130:CYS:SG	2.56	0.46
1:C:93:THR:C	1:C:120:ASN:OD1	2.59	0.45
1:B:56:CYS:SG	1:B:57:SER:N	2.90	0.45
1:C:84:ILE:HD12	1:C:183:HIS:CD2	2.52	0.45
1:A:14:LYS:HD2	1:A:15:TYR:CE1	2.53	0.44
1:A:149:GLY:HA3	1:A:152:GLU:OE1	2.19	0.43
1:C:15:TYR:OH	1:D:132:TRP:O	2.37	0.43
1:A:101:LEU:HD12	1:A:101:LEU:HA	1.91	0.41
1:D:122:THR:O	1:D:122:THR:HG22	2.20	0.41
1:C:63:LEU:HD23	1:C:75:VAL:O	2.20	0.40

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:74:LEU:HD23	1:A:100:PHE:CG	2.57	0.40

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	185/212 (87%)	181 (98%)	4 (2%)	0	100	100
1	B	183/212 (86%)	181 (99%)	2 (1%)	0	100	100
1	C	180/212 (85%)	176 (98%)	4 (2%)	0	100	100
1	D	182/212 (86%)	181 (100%)	1 (0%)	0	100	100
All	All	730/848 (86%)	719 (98%)	11 (2%)	0	100	100

There are no Ramachandran outliers to report.

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	157/174 (90%)	157 (100%)	0	100	100
1	B	156/174 (90%)	156 (100%)	0	100	100
1	C	153/174 (88%)	153 (100%)	0	100	100
1	D	155/174 (89%)	155 (100%)	0	100	100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
All	All	621/696 (89%)	621 (100%)	0	100 100

There are no protein residues with a non-rotameric sidechain to report.

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (3) such sidechains are listed below:

Mol	Chain	Res	Type
1	B	18	ASN
1	C	168	GLN
1	D	171	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)

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

5.5 Carbohydrates (i)

There are no oligosaccharides in this entry.

5.6 Ligand geometry (i)

Of 12 ligands modelled in this entry, 8 are monoatomic - leaving 4 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	$\#\|Z\| > 2$
4	PO4	D	303	-	4,4,4	0.94	0	6,6,6	0.47	0
4	PO4	A	303	-	4,4,4	0.97	0	6,6,6	0.46	0
4	PO4	B	303	-	4,4,4	0.99	0	6,6,6	0.42	0

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\#\|Z\| > 2$	Counts	RMSZ	$\#\|Z\| > 2$
4	PO4	C	303	-	4,4,4	0.94	0	6,6,6	0.48	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	191/212 (90%)	-0.07	4 (2%)	63	59	13, 25, 50, 65	0
1	B	189/212 (89%)	-0.03	9 (4%)	35	32	12, 25, 53, 73	0
1	C	186/212 (87%)	0.23	15 (8%)	18	14	14, 29, 62, 90	0
1	D	188/212 (88%)	0.29	16 (8%)	16	13	17, 32, 60, 81	0
All	All	754/848 (88%)	0.10	44 (5%)	29	25	12, 27, 56, 90	0

All (44) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	138	GLU	6.9
1	A	138	GLU	4.8
1	B	210	THR	4.6
1	C	153	SER	4.4
1	C	117	ASN	4.3
1	D	116	ASP	4.2
1	D	45	LEU	4.0
1	B	167	ASP	4.0
1	D	117	ASN	3.8
1	C	120	ASN	3.6
1	C	45	LEU	3.3
1	D	41	ASP	3.3
1	D	167	ASP	3.3
1	B	56	CYS	3.2
1	C	41	ASP	3.1
1	C	116	ASP	3.1
1	C	156	LYS	3.0
1	D	138	GLU	2.9
1	C	136	LYS	2.8
1	B	41	ASP	2.8
1	A	211	LYS	2.6

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	D	209	GLN	2.6
1	B	45	LEU	2.6
1	B	117	ASN	2.6
1	B	152	GLU	2.6
1	C	118	GLY	2.5
1	D	6	ASP	2.5
1	D	56	CYS	2.5
1	D	170	GLU	2.5
1	C	157	GLU	2.3
1	D	10	GLU	2.3
1	C	3	ASP	2.2
1	C	210	THR	2.2
1	D	118	GLY	2.2
1	A	152	GLU	2.2
1	A	45	LEU	2.2
1	D	156	LYS	2.2
1	D	152	GLU	2.1
1	C	56	CYS	2.1
1	C	122	THR	2.1
1	B	150	VAL	2.0
1	D	136	LYS	2.0
1	D	151	ILE	2.0
1	B	138	GLU	2.0

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 oligosaccharides 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, 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.

Continued on next page...

Continued from previous page...

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å^2)	Q<0.9
3	K	C	302	1/1	0.88	0.11	60,60,60,60	0
4	PO4	D	303	5/5	0.88	0.11	47,60,70,72	0
4	PO4	C	303	5/5	0.89	0.10	45,54,56,60	0
3	K	B	302	1/1	0.90	0.10	50,50,50,50	0
3	K	A	302	1/1	0.96	0.07	46,46,46,46	0
3	K	D	302	1/1	0.96	0.07	48,48,48,48	0
4	PO4	A	303	5/5	0.98	0.07	26,27,31,41	0
4	PO4	B	303	5/5	0.98	0.08	26,33,35,47	0
2	ZN	D	301	1/1	0.99	0.02	30,30,30,30	0
2	ZN	A	301	1/1	1.00	0.01	16,16,16,16	0
2	ZN	B	301	1/1	1.00	0.01	18,18,18,18	0
2	ZN	C	301	1/1	1.00	0.01	25,25,25,25	0

6.5 Other polymers (i)

There are no such residues in this entry.