Full wwPDB X-ray Structure Validation Report ⓘ

Dec 8, 2020 – 11:33 PM EST

PDB ID	:	7KLJ
Title	:	Crystal structure of the WD-repeat domain of human KIF21A
Authors	:	Zeng, H.; Dong, A.; Loppnau, P.; Hutchinson, A.; Seitova, A.; Edwards, A.M.; Arrowsmith, C.H.; Halabelian, L.; Structural Genomics Consortium (SGC)
Deposited on	:	2020-10-30
Resolution	:	1.52 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 following versions of software and data (see references ⓘ) were used in the production of this report:

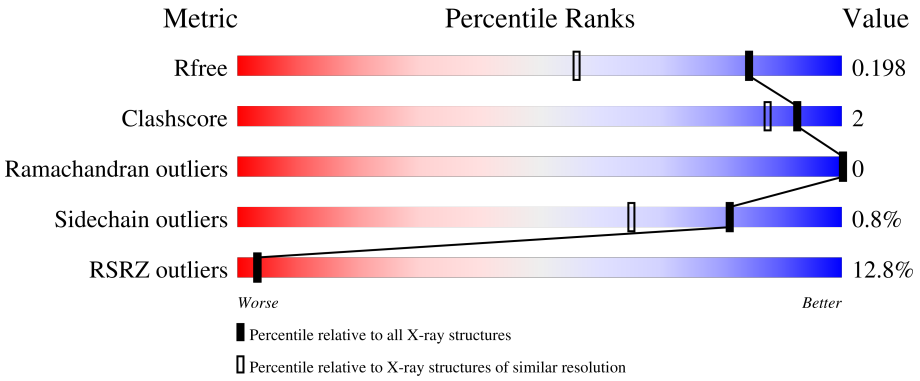
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.15.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.15.1

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 1.52 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	4009 (1.54-1.50)
Clashscore	141614	4249 (1.54-1.50)
Ramachandran outliers	138981	4148 (1.54-1.50)
Sidechain outliers	138945	4146 (1.54-1.50)
RSRZ outliers	127900	3943 (1.54-1.50)

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	335	13% 93%
1	B	335	11% 90% 6%

2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 5573 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 Isoform 2 of Kinesin-like protein KIF21A.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	324	Total	C	N	O	S	0	11	0
			2530	1598	430	484	18			
1	B	315	Total	C	N	O	S	0	12	0
			2474	1560	424	472	18			

There are 36 discrepancies between the modelled and reference sequences:

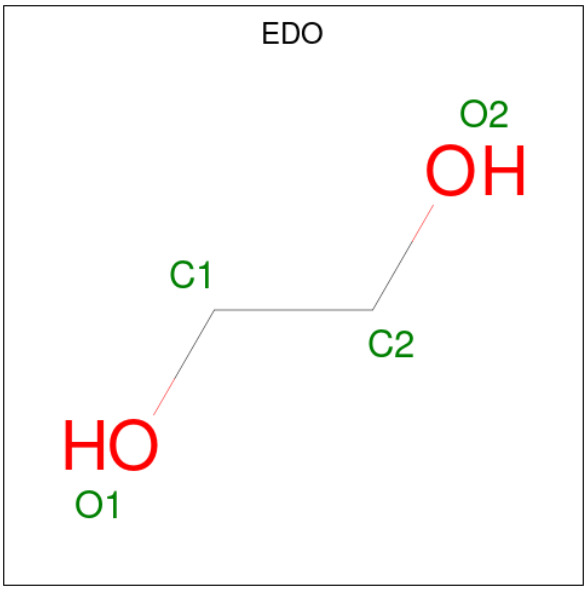
Chain	Residue	Modelled	Actual	Comment	Reference
A	1319	MET	-	expression tag	UNP Q7Z4S6
A	1320	HIS	-	expression tag	UNP Q7Z4S6
A	1321	HIS	-	expression tag	UNP Q7Z4S6
A	1322	HIS	-	expression tag	UNP Q7Z4S6
A	1323	HIS	-	expression tag	UNP Q7Z4S6
A	1324	HIS	-	expression tag	UNP Q7Z4S6
A	1325	HIS	-	expression tag	UNP Q7Z4S6
A	1326	SER	-	expression tag	UNP Q7Z4S6
A	1327	SER	-	expression tag	UNP Q7Z4S6
A	1328	GLY	-	expression tag	UNP Q7Z4S6
A	1329	ARG	-	expression tag	UNP Q7Z4S6
A	1330	GLU	-	expression tag	UNP Q7Z4S6
A	1331	ASN	-	expression tag	UNP Q7Z4S6
A	1332	LEU	-	expression tag	UNP Q7Z4S6
A	1333	TYR	-	expression tag	UNP Q7Z4S6
A	1334	PHE	-	expression tag	UNP Q7Z4S6
A	1335	GLN	-	expression tag	UNP Q7Z4S6
A	1336	GLY	-	expression tag	UNP Q7Z4S6
B	1319	MET	-	expression tag	UNP Q7Z4S6
B	1320	HIS	-	expression tag	UNP Q7Z4S6
B	1321	HIS	-	expression tag	UNP Q7Z4S6
B	1322	HIS	-	expression tag	UNP Q7Z4S6
B	1323	HIS	-	expression tag	UNP Q7Z4S6
B	1324	HIS	-	expression tag	UNP Q7Z4S6
B	1325	HIS	-	expression tag	UNP Q7Z4S6

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	1326	SER	-	expression tag	UNP Q7Z4S6
B	1327	SER	-	expression tag	UNP Q7Z4S6
B	1328	GLY	-	expression tag	UNP Q7Z4S6
B	1329	ARG	-	expression tag	UNP Q7Z4S6
B	1330	GLU	-	expression tag	UNP Q7Z4S6
B	1331	ASN	-	expression tag	UNP Q7Z4S6
B	1332	LEU	-	expression tag	UNP Q7Z4S6
B	1333	TYR	-	expression tag	UNP Q7Z4S6
B	1334	PHE	-	expression tag	UNP Q7Z4S6
B	1335	GLN	-	expression tag	UNP Q7Z4S6
B	1336	GLY	-	expression tag	UNP Q7Z4S6

- Molecule 2 is 1,2-ETHANEDIOL (three-letter code: EDO) (formula: $C_2H_6O_2$).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	1	Total C O 4 2 2	0	0
2	B	1	Total C O 4 2 2	0	0

- Molecule 3 is UNKNOWN ATOM OR ION (three-letter code: UNX) (formula: X).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	B	1	Total X 1 1	0	0

Continued on next page...

Continued from previous page...

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

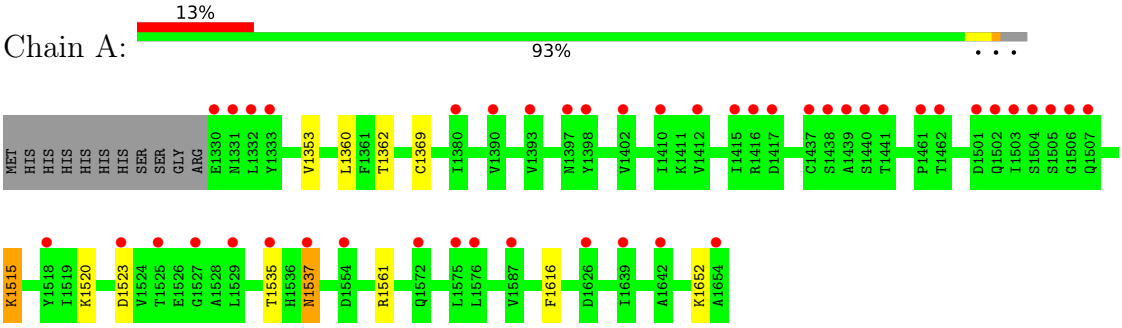
- Molecule 4 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	A	277	Total O 280 280	0	3
4	B	271	Total O 276 276	0	5

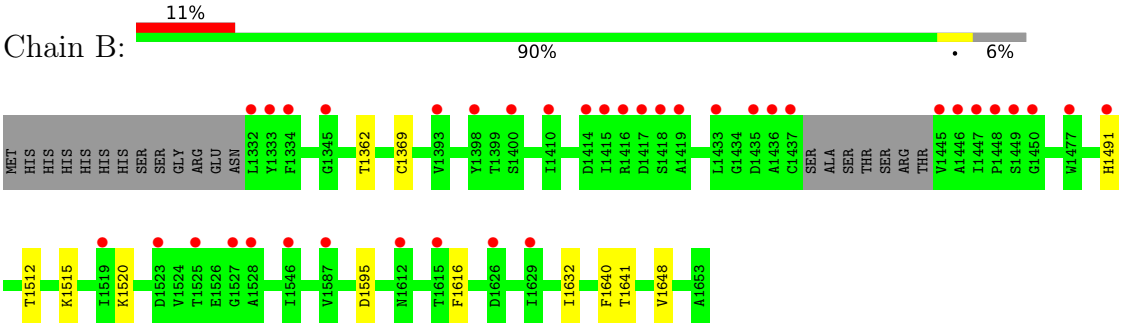
3 Residue-property plots (i)

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: Isoform 2 of Kinesin-like protein KIF21A



• Molecule 1: Isoform 2 of Kinesin-like protein KIF21A



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	55.27Å 67.07Å 79.88Å 90.00° 94.43° 90.00°	Depositor
Resolution (Å)	20.20 – 1.52 20.19 – 1.52	Depositor EDS
% Data completeness (in resolution range)	99.2 (20.20-1.52) 99.2 (20.19-1.52)	Depositor EDS
R_{merge}	0.07	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	3.58 (at 1.52Å)	Xtriage
Refinement program	REFMAC 5.8.0258	Depositor
R, R_{free}	0.170 , 0.185 0.184 , 0.198	Depositor DCC
R_{free} test set	1307 reflections (1.47%)	wwPDB-VP
Wilson B-factor (Å^2)	19.3	Xtriage
Anisotropy	0.107	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.36 , 44.8	EDS
L-test for twinning[2]	$< \|L\| > = 0.50$, $< L^2 > = 0.34$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.97	EDS
Total number of atoms	5573	wwPDB-VP
Average B, all atoms (Å^2)	21.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.67% 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: UNX, EDO

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	0.65	1/2586 (0.0%)	0.76	0/3521
1	B	0.61	0/2527	0.77	0/3436
All	All	0.63	1/5113 (0.0%)	0.76	0/6957

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

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	1652	LYS	C-N	8.91	1.54	1.34

There are no bond angle outliers.

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	B	1595[B]	ASP	Peptide

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	2530	0	2455	8	0
1	B	2474	0	2389	7	0
2	A	4	0	6	0	0
2	B	4	0	6	0	0
3	A	4	0	0	0	0
3	B	1	0	0	0	0
4	A	280	0	0	1	0
4	B	276	0	0	2	0
All	All	5573	0	4856	15	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 (15) 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:1515:LYS:CE	1:A:1515:LYS:HA	2.30	0.61
1:B:1515:LYS:HG2	4:B:2021:HOH:O	2.01	0.60
1:A:1515:LYS:HZ3	1:A:1561:ARG:CZ	2.19	0.55
1:A:1515:LYS:HE3	1:A:1515:LYS:HA	1.90	0.54
1:A:1520:LYS:NZ	1:A:1537:ASN:OD1	2.39	0.54
1:A:1362:THR:O	1:A:1369:CYS:HA	2.09	0.53
1:B:1515:LYS:NZ	4:B:1806:HOH:O	2.42	0.52
1:B:1362:THR:O	1:B:1369:CYS:HA	2.13	0.48
1:A:1515:LYS:HE2	4:A:1988:HOH:O	2.12	0.48
1:B:1491[A]:HIS:ND1	1:B:1520:LYS:HE2	2.35	0.42
1:B:1641:THR:O	1:B:1648:VAL:HA	2.20	0.41
1:A:1523:ASP:OD2	1:A:1535:THR:HG21	2.19	0.41
1:B:1491[A]:HIS:HE2	1:B:1512:THR:HG1	1.65	0.41
1:B:1632:ILE:HA	1:B:1640:PHE:O	2.21	0.40
1:A:1353[B]:VAL:CG2	1:A:1360:LEU:HD11	2.51	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	333/335 (99%)	318 (96%)	15 (4%)	0	100	100
1	B	323/335 (96%)	309 (96%)	14 (4%)	0	100	100
All	All	656/670 (98%)	627 (96%)	29 (4%)	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	275/288 (96%)	272 (99%)	3 (1%)	73	52
1	B	269/288 (93%)	268 (100%)	1 (0%)	91	82
All	All	544/576 (94%)	540 (99%)	4 (1%)	81	69

All (4) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	1515	LYS
1	A	1537	ASN
1	A	1616	PHE
1	B	1616	PHE

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

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 7 ligands modelled in this entry, 5 are unknown - 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	$\#\|Z\| > 2$
2	EDO	B	1701	-	3,3,3	0.18	0	2,2,2	0.05	0
2	EDO	A	1701	-	3,3,3	0.10	0	2,2,2	0.11	0

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the Chemical Component Dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
2	EDO	B	1701	-	-	0/1/1/1	-
2	EDO	A	1701	-	-	0/1/1/1	-

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	324/335 (96%)	1.03	45 (13%)	2	2	12, 18, 35, 55	0
1	B	315/335 (94%)	0.92	37 (11%)	4	4	12, 19, 32, 44	0
All	All	639/670 (95%)	0.97	82 (12%)	3	3	12, 18, 33, 55	0

All (82) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	1505	SER	12.8
1	A	1503	ILE	9.4
1	A	1333	TYR	9.1
1	A	1506	GLY	8.0
1	A	1439	ALA	7.8
1	A	1504	SER	7.4
1	B	1445	VAL	7.1
1	A	1398	TYR	6.8
1	A	1438	SER	6.3
1	B	1446	ALA	5.9
1	A	1440	SER	5.9
1	A	1330	GLU	5.5
1	B	1527	GLY	5.3
1	A	1576	LEU	5.2
1	B	1332	LEU	4.7
1	B	1416	ARG	4.6
1	B	1333	TYR	4.4
1	B	1398	TYR	4.1
1	A	1332	LEU	4.1
1	B	1448	PRO	3.9
1	B	1449	SER	3.7
1	B	1415	ILE	3.7
1	A	1523	ASP	3.5
1	B	1525	THR	3.5

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	B	1437	CYS	3.3
1	A	1441	THR	3.3
1	B	1433	LEU	3.2
1	A	1639[A]	ILE	3.2
1	A	1527	GLY	3.1
1	B	1417	ASP	3.1
1	A	1462	THR	3.1
1	B	1410	ILE	3.1
1	B	1447	ILE	3.1
1	A	1416	ARG	3.0
1	A	1380[A]	ILE	2.9
1	A	1410	ILE	2.9
1	A	1529	LEU	2.8
1	A	1417	ASP	2.8
1	A	1331	ASN	2.7
1	B	1414	ASP	2.7
1	A	1575	LEU	2.7
1	B	1523[A]	ASP	2.7
1	B	1435	ASP	2.6
1	B	1587	VAL	2.6
1	A	1507	GLN	2.6
1	A	1642	ALA	2.6
1	B	1477	TRP	2.5
1	B	1419	ALA	2.5
1	B	1436	ALA	2.5
1	B	1418	SER	2.5
1	A	1525	THR	2.5
1	A	1518	TYR	2.5
1	B	1519	ILE	2.5
1	B	1546	ILE	2.5
1	B	1626	ASP	2.4
1	A	1415	ILE	2.4
1	B	1400	SER	2.3
1	A	1393[A]	VAL	2.3
1	B	1393[A]	VAL	2.3
1	A	1626[A]	ASP	2.3
1	A	1654	ALA	2.3
1	B	1345	GLY	2.3
1	A	1397	ASN	2.3
1	A	1390	VAL	2.3
1	B	1491[A]	HIS	2.2
1	A	1535	THR	2.2

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	A	1554	ASP	2.2
1	A	1402	VAL	2.2
1	A	1412	VAL	2.2
1	A	1501	ASP	2.1
1	A	1502	GLN	2.1
1	A	1572	GLN	2.1
1	A	1461	PRO	2.1
1	A	1537	ASN	2.0
1	A	1587	VAL	2.0
1	B	1528	ALA	2.0
1	A	1437	CYS	2.0
1	B	1334	PHE	2.0
1	B	1450	GLY	2.0
1	B	1629	ILE	2.0
1	B	1615	THR	2.0
1	B	1612	ASN	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 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, 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	UNX	A	1702	1/1	0.75	0.25	35,35,35,35	0
2	EDO	A	1701	4/4	0.85	0.13	37,37,40,41	0
3	UNX	A	1705	1/1	0.85	0.11	23,23,23,23	0
3	UNX	A	1704	1/1	0.90	0.14	25,25,25,25	0
3	UNX	A	1703	1/1	0.96	0.17	18,18,18,18	0
2	EDO	B	1701	4/4	0.97	0.09	25,27,28,29	0
3	UNX	B	1702	1/1	0.97	0.20	23,23,23,23	0

6.5 Other polymers (i)

There are no such residues in this entry.