Full wwPDB X-ray Structure Validation Report ⓘ

May 29, 2020 – 08:43 am BST

PDB ID	:	5WCH
Title	:	Crystal structure of the catalytic domain of human USP9X
Authors	:	Dong, A.; Zhang, Q.; Walker, J.R.; Bountra, C.; Arrowsmith, C.H.; Edwards, A.M.; Tong, Y.; Structural Genomics Consortium (SGC)
Deposited on	:	2017-06-30
Resolution	:	2.50 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:

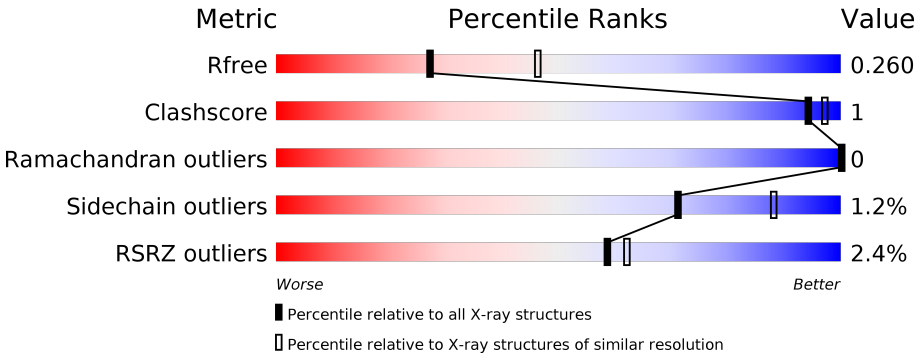
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.11

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 2.50 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 on 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	421	% 80% · 17%
1	B	421	3% 76% · 21%
1	C	421	% 76% 6% 19%
1	D	421	3% 76% · 21%

2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 10802 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 Probable ubiquitin carboxyl-terminal hydrolase FAF-X.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	350	Total	C	N	O	S	0	1	0
			2801	1787	483	511	20			
1	B	331	Total	C	N	O	S	0	1	0
			2626	1688	442	477	19			
1	C	342	Total	C	N	O	S	0	1	0
			2717	1744	460	496	17			
1	D	331	Total	C	N	O	S	0	1	1
			2613	1680	443	470	20			

There are 12 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1550	GLY	-	expression tag	UNP Q93008
A	1637	ALA	LYS	engineered mutation	UNP Q93008
A	1638	ALA	GLU	engineered mutation	UNP Q93008
B	1550	GLY	-	expression tag	UNP Q93008
B	1637	ALA	LYS	engineered mutation	UNP Q93008
B	1638	ALA	GLU	engineered mutation	UNP Q93008
C	1550	GLY	-	expression tag	UNP Q93008
C	1637	ALA	LYS	engineered mutation	UNP Q93008
C	1638	ALA	GLU	engineered mutation	UNP Q93008
D	1550	GLY	-	expression tag	UNP Q93008
D	1637	ALA	LYS	engineered mutation	UNP Q93008
D	1638	ALA	GLU	engineered mutation	UNP Q93008

- Molecule 2 is ZINC ION (three-letter code: ZN) (formula: Zn).

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

Continued on next page...

Continued from previous page...

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

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

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

- Molecule 4 is water.

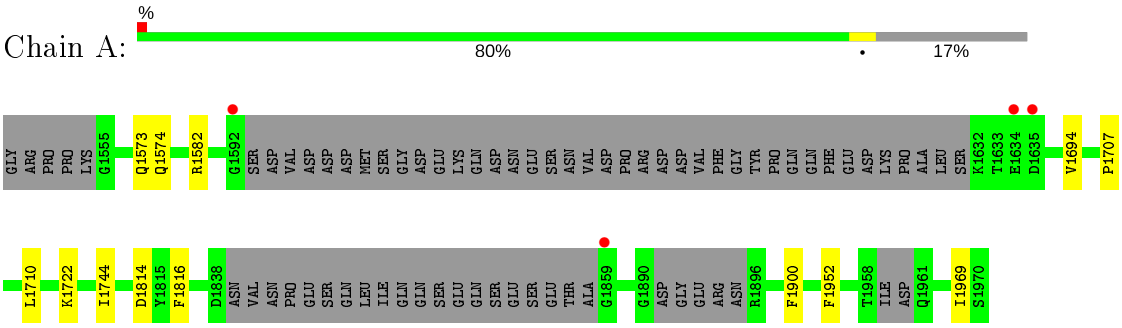
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	A	11	Total O 11 11	0	0
4	B	7	Total O 7 7	0	0
4	C	16	Total O 16 16	0	0
4	D	6	Total O 6 6	0	0

3 Residue-property plots (i)

These plots are drawn for all protein, RNA and DNA 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.

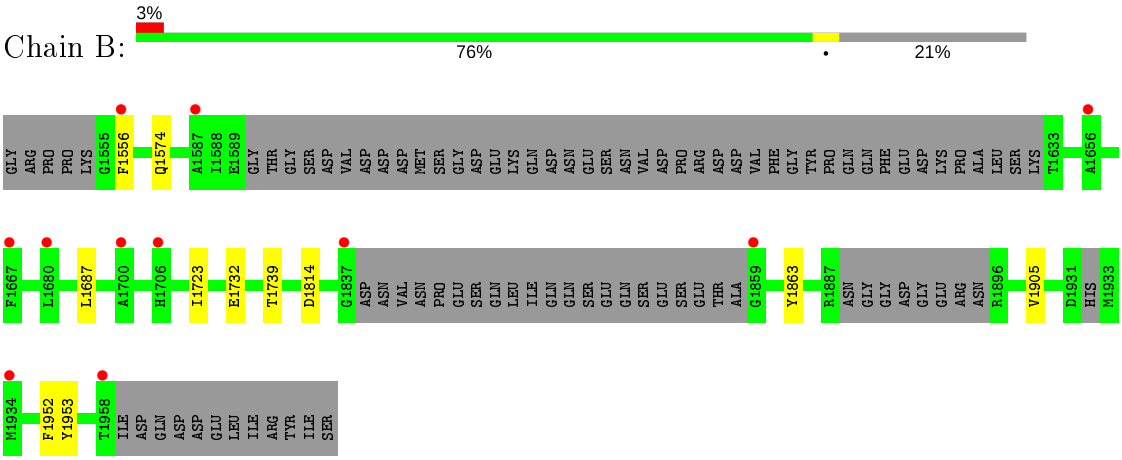
• Molecule 1: Probable ubiquitin carboxyl-terminal hydrolase FAF-X

Chain A:



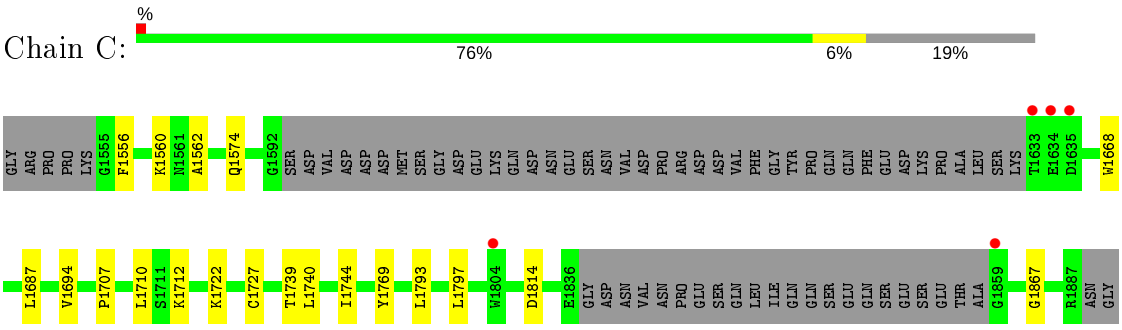
• Molecule 1: Probable ubiquitin carboxyl-terminal hydrolase FAF-X

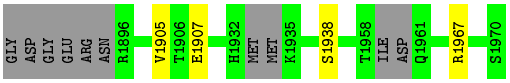
Chain B:



• Molecule 1: Probable ubiquitin carboxyl-terminal hydrolase FAF-X

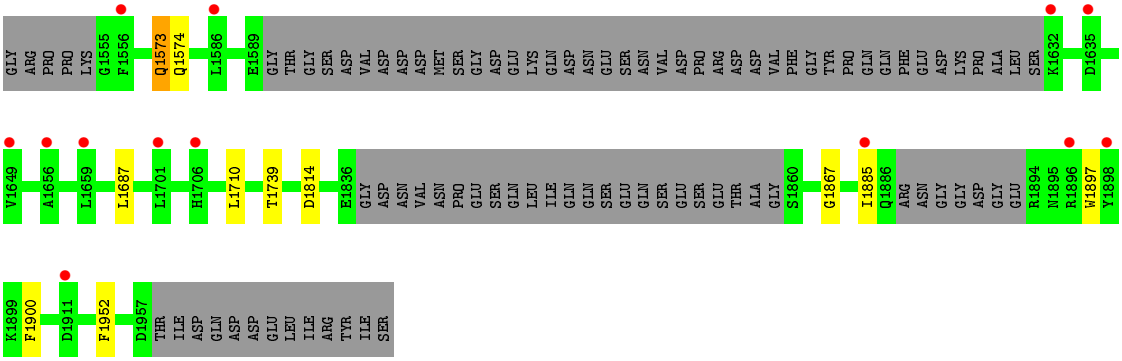
Chain C:





• Molecule 1: Probable ubiquitin carboxyl-terminal hydrolase FAF-X

Chain D:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	79.37Å 79.05Å 131.93Å 90.00° 103.76° 90.00°	Depositor
Resolution (Å)	50.01 – 2.50 49.78 – 2.50	Depositor EDS
% Data completeness (in resolution range)	96.1 (50.01-2.50) 96.1 (49.78-2.50)	Depositor EDS
R_{merge}	0.12	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.55 (at 2.51Å)	Xtriage
Refinement program	REFMAC 5.8.0124	Depositor
R, R_{free}	0.207 , 0.261 0.209 , 0.260	Depositor DCC
R_{free} test set	1087 reflections (2.06%)	wwPDB-VP
Wilson B-factor (Å^2)	42.6	Xtriage
Anisotropy	0.413	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.33 , 35.2	EDS
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	10802	wwPDB-VP
Average B, all atoms (Å^2)	46.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.65% 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, 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	0.45	0/2864	0.67	1/3858 (0.0%)
1	B	0.44	0/2688	0.63	0/3630
1	C	0.46	0/2780	0.66	0/3753
1	D	0.44	0/2677	0.62	0/3619
All	All	0.45	0/11009	0.64	1/14860 (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	A	1582	ARG	NE-CZ-NH1	5.17	122.89	120.30

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	2801	0	2663	6	0
1	B	2626	0	2461	6	0
1	C	2717	0	2556	11	0
1	D	2613	0	2430	5	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	C	1	0	0	0	0
4	A	11	0	0	0	0
4	B	7	0	0	0	0
4	C	16	0	0	0	0
4	D	6	0	0	0	0
All	All	10802	0	10110	28	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 1.

All (28) 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:1687:LEU:HD11	1:C:1739:THR:HG21	1.77	0.66
1:D:1687:LEU:HD11	1:D:1739:THR:HG21	1.77	0.66
1:A:1574:GLN:HG2	1:A:1952:PHE:CE1	2.36	0.61
1:D:1574:GLN:HG2	1:D:1952:PHE:CE1	2.36	0.59
1:C:1740:LEU:HD12	1:C:1793:LEU:HD11	1.86	0.57
1:B:1574:GLN:HG2	1:B:1952:PHE:CE1	2.40	0.56
1:D:1574:GLN:HE22	1:D:1867:GLY:HA3	1.78	0.49
1:C:1707:PRO:HD3	1:C:1967:ARG:O	2.15	0.47
1:B:1556:PHE:HB2	1:B:1905:VAL:HG11	1.98	0.46
1:C:1556:PHE:HB2	1:C:1905:VAL:HG11	1.97	0.46
1:C:1722:LYS:HD3	1:C:1769:TYR:CE1	2.51	0.45
1:A:1573:GLN:HG2	1:A:1900:PHE:CD2	2.52	0.45
1:B:1574:GLN:HG2	1:B:1952:PHE:CZ	2.52	0.45
1:C:1694:VAL:HG12	1:C:1710:LEU:CD2	2.47	0.45
1:B:1723:ILE:N	1:B:1723:ILE:HD12	2.33	0.44
1:B:1687:LEU:HD11	1:B:1739:THR:HG21	1.99	0.44
1:C:1562:ALA:HB2	1:C:1668:TRP:CZ2	2.53	0.44
1:D:1573:GLN:HG2	1:D:1900:PHE:CD2	2.53	0.43
1:C:1574:GLN:NE2	1:C:1867:GLY:HA3	2.34	0.43
1:C:1694:VAL:HG12	1:C:1710:LEU:HD22	2.01	0.42
1:A:1744:ILE:HD12	1:A:1816:PHE:CG	2.54	0.42
1:A:1694:VAL:HG12	1:A:1710:LEU:CD2	2.50	0.42
1:A:1707:PRO:HB3	1:A:1969:ILE:HD12	2.02	0.41
1:D:1885:ILE:HD11	1:D:1897:TRP:NE1	2.35	0.41
1:B:1863:TYR:HB3	1:B:1953:TYR:HB3	2.02	0.41
1:C:1744:ILE:HD13	1:C:1744:ILE:HA	1.89	0.41

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:C:1744:ILE:HD11	1:C:1797:LEU:HG	2.00	0.41
1:A:1694:VAL:HG12	1:A:1710:LEU:HD22	2.02	0.41

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	341/421 (81%)	331 (97%)	10 (3%)	0	100	100
1	B	322/421 (76%)	314 (98%)	8 (2%)	0	100	100
1	C	331/421 (79%)	322 (97%)	9 (3%)	0	100	100
1	D	324/421 (77%)	314 (97%)	10 (3%)	0	100	100
All	All	1318/1684 (78%)	1281 (97%)	37 (3%)	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	285/366 (78%)	283 (99%)	2 (1%)	84	94
1	B	262/366 (72%)	260 (99%)	2 (1%)	81	93
1	C	273/366 (75%)	267 (98%)	6 (2%)	52	77

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
1	D	256/366 (70%)	253 (99%)	3 (1%)	71 88
All	All	1076/1464 (74%)	1063 (99%)	13 (1%)	71 88

All (13) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	1722	LYS
1	A	1814	ASP
1	B	1732	GLU
1	B	1814	ASP
1	C	1560	LYS
1	C	1712	LYS
1	C	1727	CYS
1	C	1814	ASP
1	C	1907	GLU
1	C	1938	SER
1	D	1573	GLN
1	D	1710	LEU
1	D	1814	ASP

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (4) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	1660	GLN
1	A	1871	HIS
1	C	1574	GLN
1	D	1574	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 carbohydrates in this entry.

5.6 Ligand geometry (i)

Of 5 ligands modelled in this entry, 1 is unknown and 4 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

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	350/421 (83%)	0.01	4 (1%) 80 82	26, 43, 67, 97	0
1	B	331/421 (78%)	0.23	11 (3%) 46 50	26, 47, 70, 111	0
1	C	342/421 (81%)	0.02	5 (1%) 73 75	24, 42, 70, 101	0
1	D	331/421 (78%)	0.23	13 (3%) 39 42	28, 49, 76, 101	0
All	All	1354/1684 (80%)	0.12	33 (2%) 59 62	24, 45, 72, 111	0

All (33) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	D	1556	PHE	4.5
1	C	1634	GLU	4.1
1	B	1859	GLY	4.1
1	D	1898	TYR	3.8
1	A	1859	GLY	3.6
1	C	1635	ASP	3.5
1	C	1804	TRP	3.4
1	A	1634	GLU	3.1
1	B	1556	PHE	3.0
1	D	1656	ALA	2.9
1	B	1656	ALA	2.8
1	D	1896	ARG	2.8
1	B	1700	ALA	2.7
1	B	1706	HIS	2.6
1	C	1633	THR	2.6
1	D	1659	LEU	2.6
1	B	1587	ALA	2.5
1	B	1958	THR	2.5
1	A	1592	GLY	2.5
1	A	1635	ASP	2.5
1	D	1911	ASP	2.3

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	D	1649	VAL	2.2
1	C	1859	GLY	2.2
1	B	1667	PHE	2.2
1	D	1701	LEU	2.1
1	D	1706	HIS	2.1
1	D	1632	LYS	2.1
1	D	1885	ILE	2.1
1	B	1837	GLY	2.1
1	B	1680	LEU	2.0
1	D	1635	ASP	2.0
1	D	1586	LEU	2.0
1	B	1934	MET	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 carbohydrates 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	C	2002	1/1	0.80	0.38	30,30,30,30	0
2	ZN	D	2000	1/1	1.00	0.12	29,29,29,29	0
2	ZN	C	2001	1/1	1.00	0.11	34,34,34,34	0
2	ZN	B	2000	1/1	1.00	0.13	31,31,31,31	0
2	ZN	A	2000	1/1	1.00	0.11	34,34,34,34	0

6.5 Other polymers (i)

There are no such residues in this entry.