



wwPDB X-ray Structure Validation Summary Report ⓘ

Oct 16, 2017 – 12:18 PM EDT

PDB ID : 3CDZ
Title : Crystal structure of human factor VIII
Authors : Ngo, J.C.; Huang, M.; Roth, D.A.; Furie, B.C.; Furie, B.
Deposited on : unknown
Resolution : 3.98 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

<http://wwpdb.org/validation/2016/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 : 1.7.2 (RC1), CSD as538be (2017)
Xtriage (Phenix) : 1.9-1692
EDS : rb-20030345
Percentile statistics : 20161228.v01 (using entries in the PDB archive December 28th 2016)
Refmac : 5.8.0135
CCP4 : 6.5.0
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP) : rb-20030345

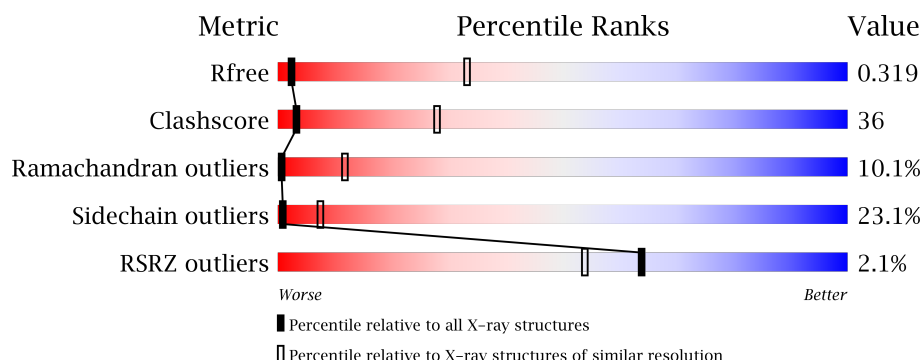
1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 3.98 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}	100719	1066 (4.36-3.60)
Clashscore	112137	1163 (4.36-3.60)
Ramachandran outliers	110173	1119 (4.36-3.60)
Sidechain outliers	110143	1108 (4.36-3.60)
RSRZ outliers	101464	1078 (4.36-3.60)

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. 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 $\leq 5\%$. 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	754	<div> <div>2%</div> <div>27%</div> <div>38%</div> <div>15%</div> <div>•</div> <div>16%</div> </div>
2	B	684	<div> <div>2%</div> <div>36%</div> <div>41%</div> <div>14%</div> <div>•</div> <div>8%</div> </div>

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
3	NAG	B	2333	X	-	-	-
3	NAG	B	2334	X	-	-	-
6	MAN	B	2336	X	-	-	-
6	MAN	B	2338	X	-	-	-

2 Entry composition

There are 6 unique types of molecules in this entry. The entry contains 10317 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 Coagulation factor VIII heavy chain.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	630	Total	C	N	O	S	0	0	0
			5086	3278	855	928	25			

There are 14 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	741	SER	-	SEE REMARK 999	UNP P00451
A	742	PHE	-	SEE REMARK 999	UNP P00451
A	743	SER	-	SEE REMARK 999	UNP P00451
A	744	GLN	-	SEE REMARK 999	UNP P00451
A	745	ASN	-	SEE REMARK 999	UNP P00451
A	746	PRO	-	SEE REMARK 999	UNP P00451
A	747	PRO	-	SEE REMARK 999	UNP P00451
A	748	VAL	-	SEE REMARK 999	UNP P00451
A	749	LEU	-	SEE REMARK 999	UNP P00451
A	750	LYS	-	SEE REMARK 999	UNP P00451
A	751	ARG	-	SEE REMARK 999	UNP P00451
A	752	HIS	-	SEE REMARK 999	UNP P00451
A	753	GLN	-	SEE REMARK 999	UNP P00451
A	754	ARG	-	SEE REMARK 999	UNP P00451

- Molecule 2 is a protein called Coagulation factor VIII light chain.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	631	Total	C	N	O	S	0	0	0
			5125	3288	881	924	32			

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	1880	LEU	PHE	VARIANT	UNP P00451

- Molecule 3 is N-ACETYL-D-GLUCOSAMINE (three-letter code: NAG) (formula: $C_8H_{15}NO_6$).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
3	A	1	Total	C	N	O	0	0
			14	8	1	5		
3	A	1	Total	C	N	O	0	0
			14	8	1	5		
3	B	1	Total	C	N	O	0	0
			14	8	1	5		
3	B	1	Total	C	N	O	0	0
			14	8	1	5		
3	B	1	Total	C	N	O	0	0
			14	8	1	5		

- Molecule 4 is COPPER (II) ION (three-letter code: CU) (formula: Cu).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	B	1	Total	Cu	0	0
			1	1		
4	A	1	Total	Cu	0	0
			1	1		

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

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	1	Total	Ca	0	0
			1	1		

- Molecule 6 is ALPHA-D-MANNOSE (three-letter code: MAN) (formula: $C_6H_{12}O_6$).

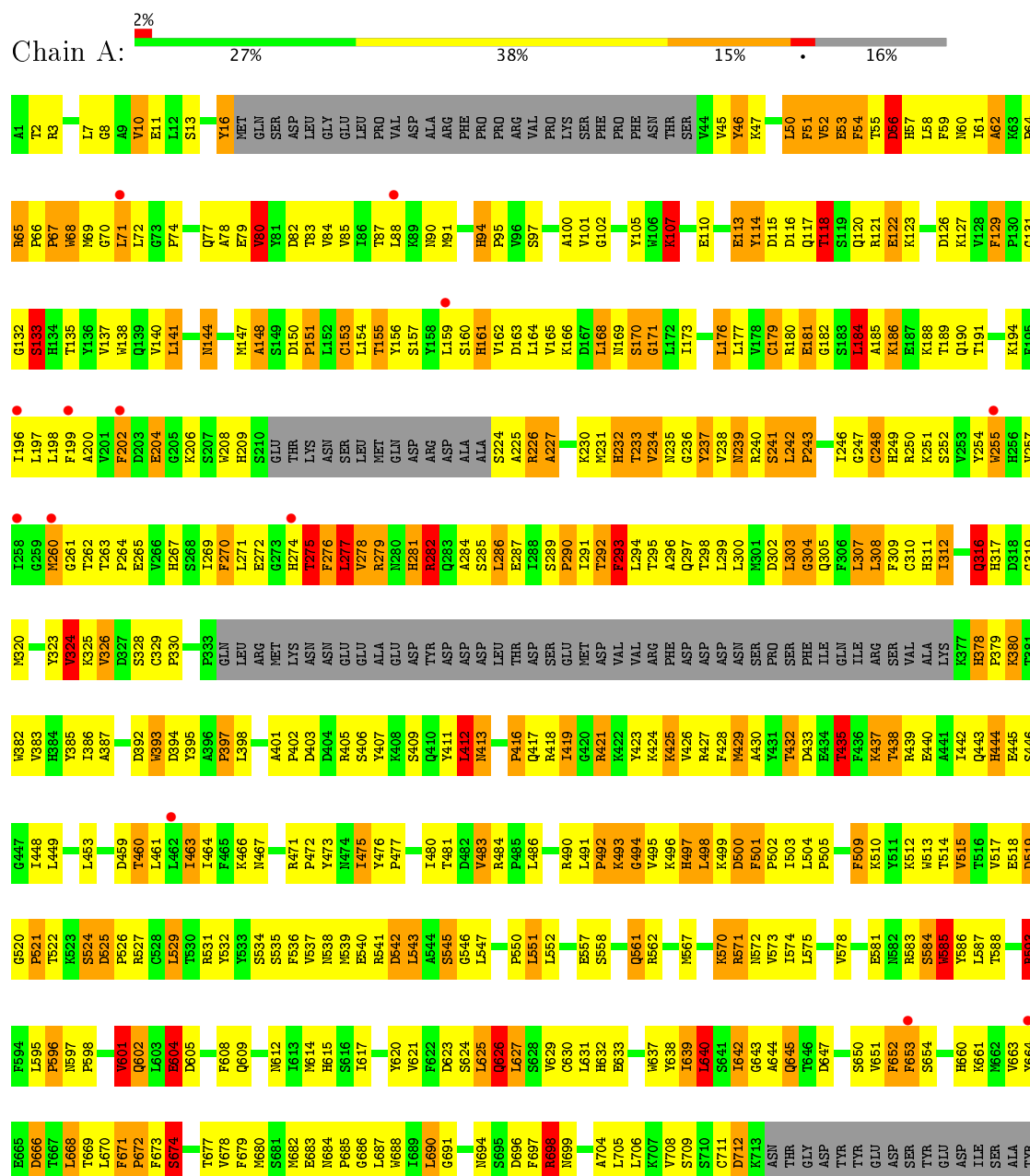


Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
6	B	1	Total	C	O	0	0
			11	6	5		
6	B	1	Total	C	O	0	0
			11	6	5		
6	B	1	Total	C	O	0	0
			11	6	5		

3 Residue-property plots

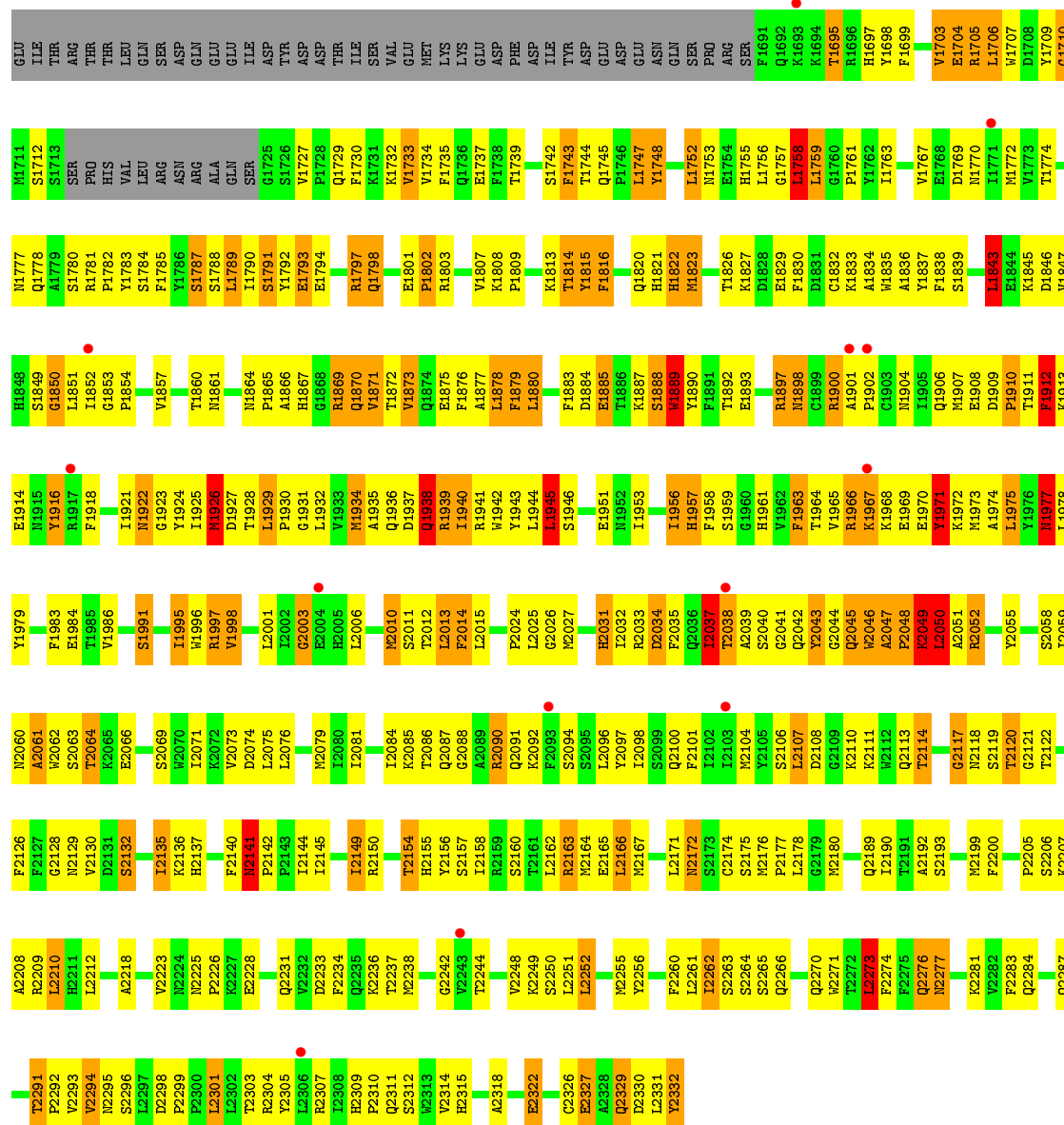
These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are 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: Coagulation factor VIII heavy chain



LEU
LEU
SER
LYS
ASN
ASN
ALA
ILE
GLU
GLU
PRO
ARG
SER
PHE
SER
GLN
ASN
PRO
PRO
VAL
LEU
LYS
ARG
HIS
GLN
ARG

• Molecule 2: Coagulation factor VIII light chain



4 Data and refinement statistics

Property	Value	Source
Space group	P 41 21 2	Depositor
Cell constants a, b, c, α , β , γ	134.11Å 134.11Å 349.76Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	50.00 – 3.98 49.66 – 3.98	Depositor EDS
% Data completeness (in resolution range)	98.5 (50.00-3.98) 98.5 (49.66-3.98)	Depositor EDS
R_{merge}	0.13	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	3.92 (at 4.00Å)	Xtriage
Refinement program	REFMAC 5.3.0037	Depositor
R, R_{free}	0.256 , 0.327 0.251 , 0.319	Depositor DCC
R_{free} test set	1399 reflections (5.29%)	DCC
Wilson B-factor (Å ²)	151.1	Xtriage
Anisotropy	0.480	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	(Not available) , (Not available)	EDS
L-test for twinning ²	$\langle L \rangle = 0.49$, $\langle L^2 \rangle = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.93	EDS
Total number of atoms	10317	wwPDB-VP
Average B, all atoms (Å ²)	199.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.22% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

²Theoretical values of $\langle |L| \rangle$, $\langle L^2 \rangle$ 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: CA, NAG, CU, MAN

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	0.74	8/5230 (0.2%)	0.94	8/7098 (0.1%)
2	B	0.62	2/5270 (0.0%)	0.77	6/7136 (0.1%)
All	All	0.68	10/10500 (0.1%)	0.86	14/14234 (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	7
2	B	0	6
All	All	0	13

The worst 5 of 10 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	121	ARG	CZ-NH1	13.13	1.50	1.33
2	B	1904	ASN	CG-OD1	9.72	1.45	1.24
1	A	121	ARG	NE-CZ	8.65	1.44	1.33
2	B	1904	ASN	CG-ND2	8.33	1.53	1.32
1	A	107	LYS	CG-CD	6.53	1.74	1.52

The worst 5 of 14 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	121	ARG	NE-CZ-NH2	-24.01	108.29	120.30
1	A	121	ARG	NE-CZ-NH1	23.51	132.06	120.30
2	B	1945	LEU	CA-CB-CG	7.79	133.22	115.30
1	A	277	LEU	CA-CB-CG	7.46	132.46	115.30

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	B	2050	LEU	CA-CB-CG	6.82	130.98	115.30

There are no chirality outliers.

5 of 13 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	237	TYR	Peptide
1	A	278	VAL	Peptide
1	A	281	HIS	Peptide
1	A	282	ARG	Peptide
1	A	56	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	5086	0	4971	402	0
2	B	5125	0	4994	354	0
3	A	28	0	25	1	0
3	B	42	0	38	2	0
4	A	1	0	0	1	0
4	B	1	0	0	0	0
5	A	1	0	0	0	0
6	B	33	0	28	0	0
All	All	10317	0	10056	737	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 36.

The worst 5 of 737 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:107:LYS:CG	1:A:107:LYS:CD	1.74	1.58
1:A:378:HIS:HB3	1:A:379:PRO:CD	1.71	1.19
2:B:1963:PHE:CD2	2:B:1986:VAL:HG21	1.77	1.18

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:412:LEU:CB	1:A:421:ARG:HG2	1.76	1.15
1:A:572:ASN:HB3	1:A:637:TRP:CZ3	1.82	1.15

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	622/754 (82%)	430 (69%)	130 (21%)	62 (10%)	1	12
2	B	627/684 (92%)	437 (70%)	126 (20%)	64 (10%)	1	11
All	All	1249/1438 (87%)	867 (69%)	256 (20%)	126 (10%)	1	12

5 of 126 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	118	THR
1	A	133	SER
1	A	181	GLU
1	A	227	ALA
1	A	232	HIS

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	559/674 (83%)	414 (74%)	145 (26%)	0	5
2	B	560/612 (92%)	446 (80%)	114 (20%)	1	11
All	All	1119/1286 (87%)	860 (77%)	259 (23%)	1	7

5 of 259 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	585	TRP
1	A	712	ASP
2	B	2164	MET
1	A	597	ASN
1	A	640	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 27 such sidechains are listed below:

Mol	Chain	Res	Type
2	B	1778	GLN
2	B	1822	HIS
2	B	2266	GLN
2	B	1805	ASN
1	A	281	HIS

5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains ⓘ

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

5.5 Carbohydrates ⓘ

There are no carbohydrates in this entry.

5.6 Ligand geometry ⓘ

Of 11 ligands modelled in this entry, 3 are monoatomic - leaving 8 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	$\# Z > 2$
3	NAG	A	755	1,3	14,14,15	0.84	0	15,19,21	1.13	1 (6%)
3	NAG	A	756	3	14,14,15	2.79	3 (21%)	15,19,21	1.79	2 (13%)
3	NAG	B	2333	2	14,14,15	0.75	0	15,19,21	2.06	4 (26%)
3	NAG	B	2334	2	14,14,15	0.81	0	15,19,21	1.31	2 (13%)
3	NAG	B	2335	6	14,14,15	0.93	0	15,19,21	1.60	5 (33%)
6	MAN	B	2336	3,6	11,11,12	0.65	0	13,15,17	1.67	2 (15%)
6	MAN	B	2337	6	11,11,12	0.70	0	13,15,17	2.06	2 (15%)
6	MAN	B	2338	6	11,11,12	0.82	0	13,15,17	1.77	3 (23%)

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
3	NAG	A	755	1,3	-	0/6/23/26	0/1/1/1
3	NAG	A	756	3	-	0/6/23/26	0/1/1/1
3	NAG	B	2333	2	1/1/5/7	0/6/23/26	0/1/1/1
3	NAG	B	2334	2	1/1/5/7	0/6/23/26	0/1/1/1
3	NAG	B	2335	6	-	0/6/23/26	0/1/1/1
6	MAN	B	2336	3,6	1/1/4/5	0/2/19/22	0/1/1/1
6	MAN	B	2337	6	-	0/2/19/22	0/1/1/1
6	MAN	B	2338	6	1/1/4/5	0/2/19/22	0/1/1/1

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	A	756	NAG	C8-C7	-5.77	1.38	1.50
3	A	756	NAG	C1-C2	2.84	1.56	1.52
3	A	756	NAG	O7-C7	7.70	1.41	1.23

The worst 5 of 21 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	B	2335	NAG	C1-O5-C5	-2.19	109.15	112.17
3	B	2335	NAG	O7-C7-C8	-2.15	118.14	122.06
3	B	2333	NAG	C4-C3-C2	2.17	114.20	111.02
3	B	2335	NAG	C2-N2-C7	2.36	126.39	122.94
3	B	2334	NAG	C2-N2-C7	2.39	126.42	122.94

All (4) chirality outliers are listed below:

Mol	Chain	Res	Type	Atom
3	B	2334	NAG	C1
6	B	2338	MAN	C1
6	B	2336	MAN	C1
3	B	2333	NAG	C1

There are no torsion outliers.

There are no ring outliers.

2 monomers are involved in 3 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	A	755	NAG	1	0
3	B	2335	NAG	2	0

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	630/754 (83%)	0.17	13 (2%) 64 54	182, 196, 215, 239	0
2	B	631/684 (92%)	0.08	13 (2%) 64 54	181, 198, 214, 231	1 (0%)
All	All	1261/1438 (87%)	0.12	26 (2%) 64 54	181, 197, 215, 239	1 (0%)

The worst 5 of 26 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	260	MET	5.1
2	B	1693	LYS	3.8
1	A	258	ILE	3.7
2	B	2243	VAL	3.2
2	B	1901	ALA	3.1

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. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. 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	LLDF	B-factors(\AA^2)	Q<0.9
4	CU	A	757	1/1	0.99	0.26	0.02	171,171,171,171	0
3	NAG	B	2333	14/15	0.91	0.22	-0.02	210,213,214,214	0
4	CU	B	1	1/1	0.97	0.24	-0.45	174,174,174,174	0
5	CA	A	758	1/1	0.94	0.22	-0.52	172,172,172,172	0
6	MAN	B	2338	11/12	0.66	0.28	-	254,255,256,256	0
3	NAG	B	2335	14/15	0.83	0.18	-	257,257,257,258	0
3	NAG	A	756	14/15	0.61	0.27	-	227,229,230,231	0
6	MAN	B	2336	11/12	0.55	0.16	-	257,259,260,261	0
3	NAG	A	755	14/15	0.76	0.20	-	210,216,218,222	0
3	NAG	B	2334	14/15	0.90	0.18	-	191,194,196,197	0
6	MAN	B	2337	11/12	0.69	0.24	-	258,261,261,261	0

6.5 Other polymers [i](#)

There are no such residues in this entry.