wwPDB X-ray Structure Validation Summary Report (i)

Jun 16, 2024 – 06:34 PM EDT

PDB ID	:	3D0I
Title	:	Crystal structure of spike protein receptor-binding domain from the 2005-2006 SARS coronavirus civet strain complexed with human-civet chimeric receptor ACE2
Authors	:	Li, F.
Deposited on	:	2008-05-01
Resolution	:	2.90 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 (i)) were used in the production of this report:

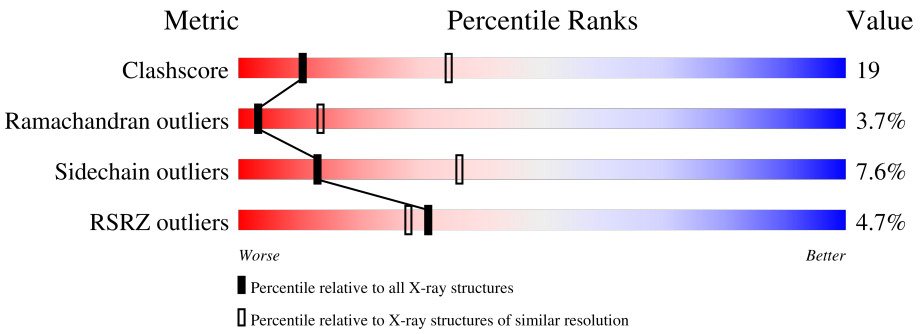
MolProbity	:	4.02b-467
Mogul	:	2022.3.0, CSD as543be (2022)
Xtriage (Phenix)	:	1.20.1
EDS	:	2.37.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.37.1

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 2.90 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))
Clashscore	141614	2172 (2.90-2.90)
Ramachandran outliers	138981	2115 (2.90-2.90)
Sidechain outliers	138945	2117 (2.90-2.90)
RSRZ outliers	127900	1906 (2.90-2.90)

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	597	6%, 56%, 40%, ·
1	B	597	4%, 59%, 37%, ·
2	E	179	4%, 64%, 28%, ·, ·
2	F	179	4%, 62%, 28%, 6%, ·

2 Entry composition (i)

There are 7 unique types of molecules in this entry. The entry contains 12639 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 Angiotensin-converting enzyme 2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	597	4864	3110	803	922	29	0	0	0
1	B	597	4864	3110	803	922	29	0	0	0

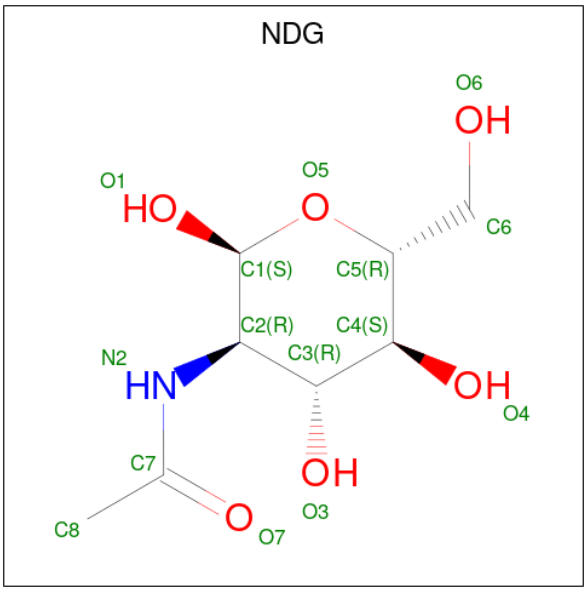
- Molecule 2 is a protein called Spike glycoprotein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	E	173	1395	905	229	255	6	0	0	0
2	F	173	1395	905	229	255	6	0	0	0

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
E	479	ARG	ASN	conflict	UNP P59594
E	480	GLY	ASP	conflict	UNP P59594
E	487	SER	THR	conflict	UNP P59594
F	479	ARG	ASN	conflict	UNP P59594
F	480	GLY	ASP	conflict	UNP P59594
F	487	SER	THR	conflict	UNP P59594

- Molecule 3 is 2-acetamido-2-deoxy-alpha-D-glucopyranose (three-letter code: NDG) (formula: $C_8H_{15}NO_6$).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
			Total	C	N	O		
3	A	1	15	8	1	6	0	0
3	B	1	15	8	1	6	0	0
3	B	1	15	8	1	6	0	0
3	B	1	15	8	1	6	0	0
3	B	1	15	8	1	6	0	0

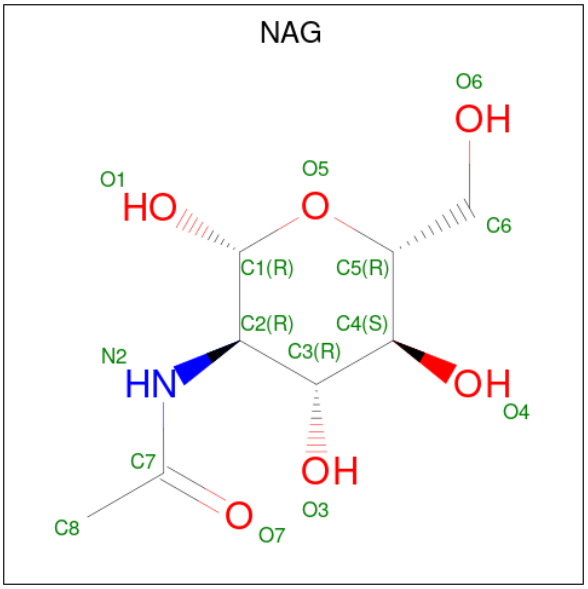
- Molecule 4 is ZINC ION (three-letter code: ZN) (formula: Zn).

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

- Molecule 5 is CHLORIDE ION (three-letter code: CL) (formula: Cl).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	Cl		
5	A	1	1	1	0	0
5	B	1	1	1	0	0

- Molecule 6 is 2-acetamido-2-deoxy-beta-D-glucopyranose (three-letter code: NAG) (formula: $C_8H_{15}NO_6$).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
			Total	C	N	O		
6	E	1	15	8	1	6	0	0

- Molecule 7 is water.

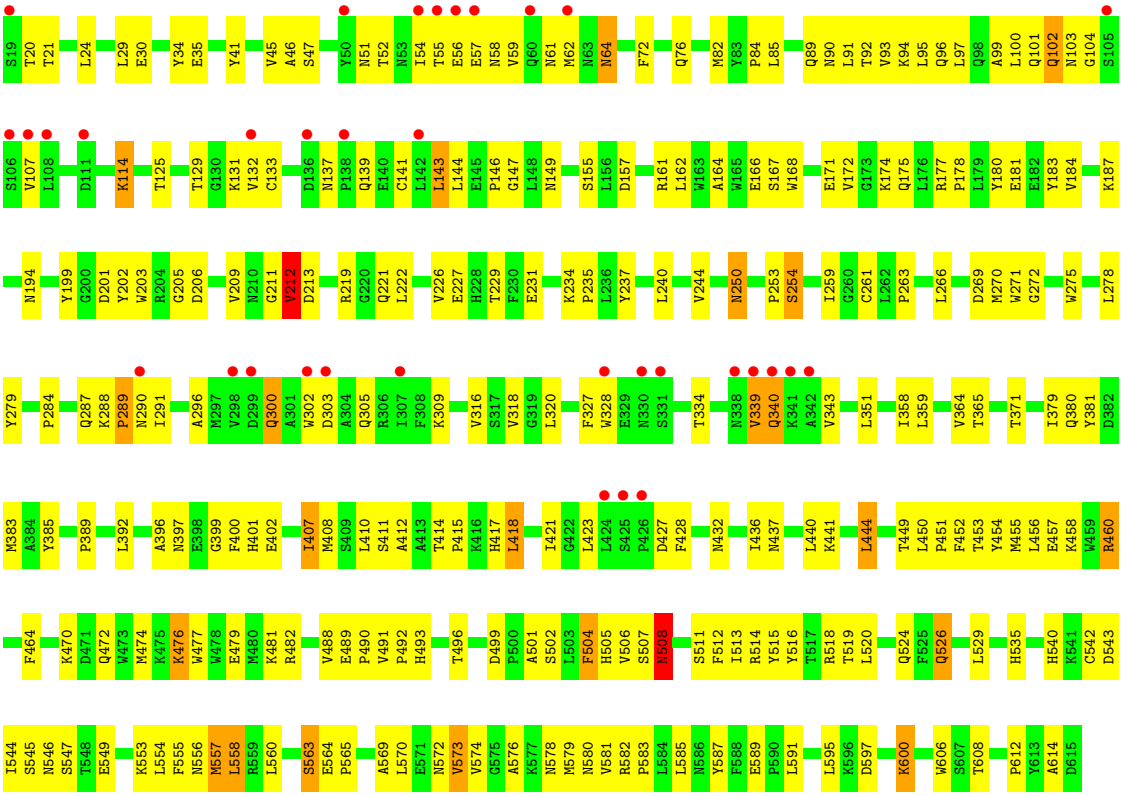
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
7	B	21	21	21	0	0
			Total	O		
7	F	6	6	6	0	0

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.

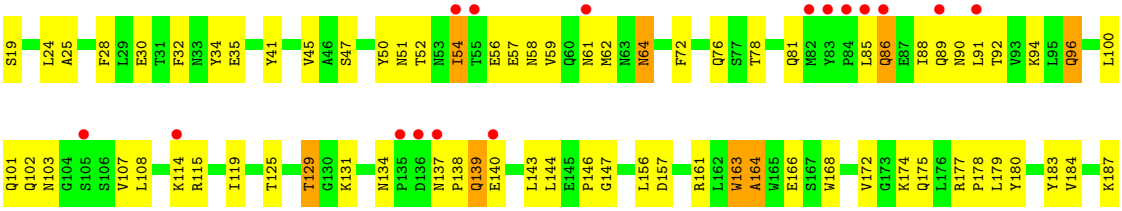
• Molecule 1: Angiotensin-converting enzyme 2

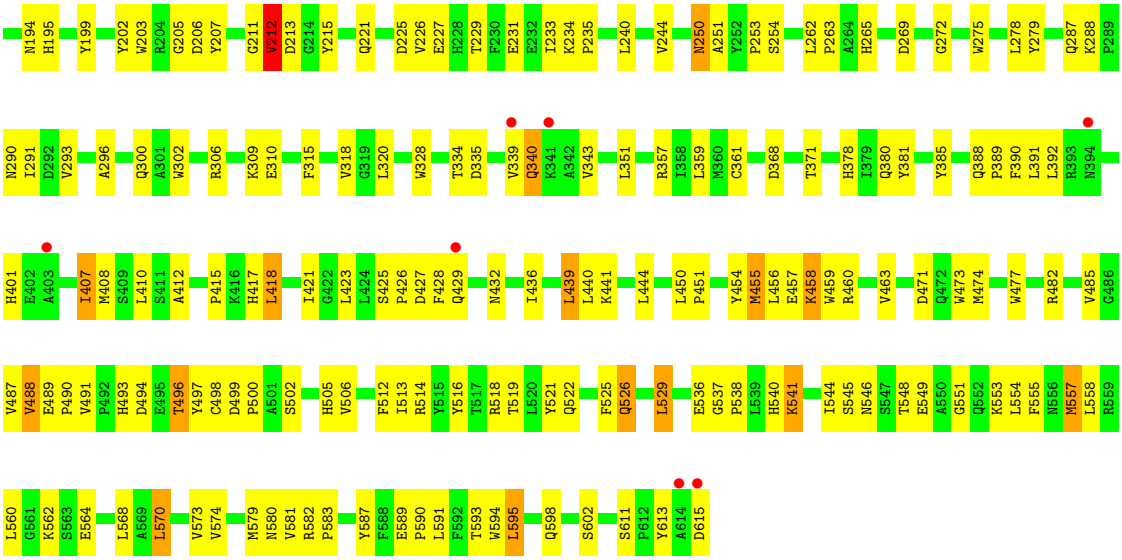
Chain A:



• Molecule 1: Angiotensin-converting enzyme 2

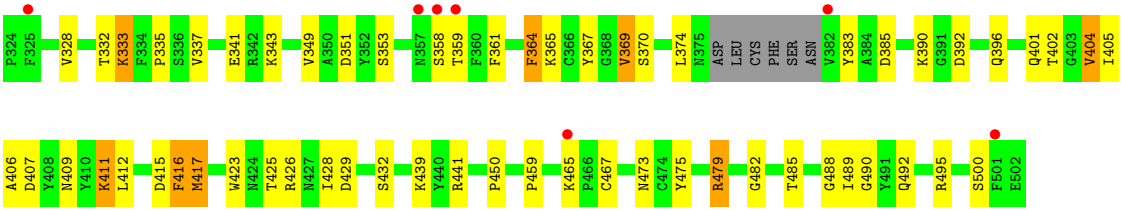
Chain B:





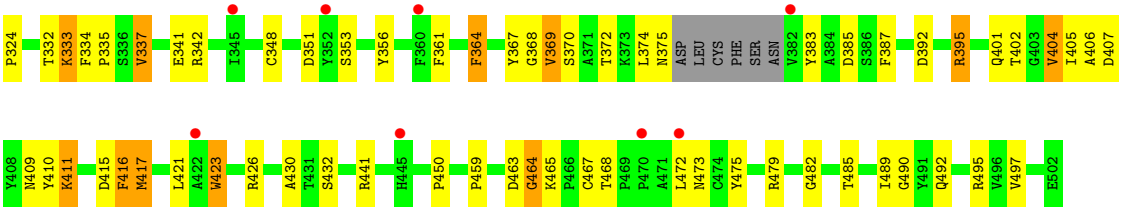
- Molecule 2: Spike glycoprotein

Chain E:



- Molecule 2: Spike glycoprotein

Chain F:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	80.42Å 119.82Å 109.77Å 90.00° 95.50° 90.00°	Depositor
Resolution (Å)	36.47 – 2.90 36.47 – 2.89	Depositor EDS
% Data completeness (in resolution range)	89.9 (36.47-2.90) 89.3 (36.47-2.89)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.85 (at 2.90Å)	Xtriage
Refinement program	REFMAC 5.2.0019	Depositor
R, R_{free}	0.224 , 0.278 0.243 , (Not available)	Depositor DCC
R_{free} test set	No test flags present.	wwPDB-VP
Wilson B-factor (Å^2)	72.6	Xtriage
Anisotropy	0.175	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.30 , 44.9	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	0.92	EDS
Total number of atoms	12639	wwPDB-VP
Average B, all atoms (Å^2)	84.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.91% 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, NAG, CL, NDG

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	0.53	1/5000 (0.0%)	0.60	0/6796
1	B	0.53	0/5000	0.62	0/6796
2	E	0.55	0/1440	0.61	0/1958
2	F	0.54	0/1440	0.61	0/1958
All	All	0.53	1/12880 (0.0%)	0.61	0/17508

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	300	GLN	CD-OE1	6.23	1.37	1.24

There are no bond angle outliers.

There are no chirality outliers.

There are no planarity outliers.

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	4864	0	4638	187	0
1	B	4864	0	4638	196	0
2	E	1395	0	1329	34	0
2	F	1395	0	1329	48	0
3	A	15	0	12	4	0
3	B	60	0	48	4	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
4	A	1	0	0	0	0
4	B	1	0	0	0	0
5	A	1	0	0	0	0
5	B	1	0	0	1	0
6	E	15	0	15	2	0
7	B	21	0	0	1	0
7	F	6	0	0	1	0
All	All	12639	0	12009	456	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 19.

The worst 5 of 456 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)
2:E:409:ASN:HD21	2:E:441:ARG:H	1.13	0.96
1:A:574:VAL:HG23	1:A:576:ALA:H	1.35	0.89
1:A:47:SER:HA	1:A:62:MET:HG3	1.56	0.88
1:A:474:MET:HE2	1:A:499:ASP:H	1.38	0.88
1:A:526:GLN:HE21	1:A:526:GLN:HA	1.39	0.85

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	595/597 (100%)	524 (88%)	52 (9%)	19 (3%)	4 16
1	B	595/597 (100%)	528 (89%)	50 (8%)	17 (3%)	4 18
2	E	169/179 (94%)	138 (82%)	21 (12%)	10 (6%)	1 5
2	F	169/179 (94%)	141 (83%)	17 (10%)	11 (6%)	1 3

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles
All	All	1528/1552 (98%)	1331 (87%)	140 (9%)	57 (4%)	3 13

5 of 57 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	146	PRO
1	A	212	VAL
1	A	289	PRO
1	A	340	GLN
1	A	614	ALA

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	527/527 (100%)	491 (93%)	36 (7%)	16 42
1	B	527/527 (100%)	484 (92%)	43 (8%)	11 32
2	E	150/156 (96%)	137 (91%)	13 (9%)	10 30
2	F	150/156 (96%)	139 (93%)	11 (7%)	14 38
All	All	1354/1366 (99%)	1251 (92%)	103 (8%)	13 36

5 of 103 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	418	LEU
1	B	541	LYS
2	F	417	MET
1	B	439	LEU
1	B	473	TRP

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 47 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	149	ASN
1	B	505	HIS
1	B	194	ASN
1	B	300	GLN
1	B	546	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 10 ligands modelled in this entry, 4 are monoatomic - leaving 6 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	NDG	A	616	-	15,15,15	0.49	0	21,21,21	0.64	0
3	NDG	B	618	-	15,15,15	0.75	0	21,21,21	1.07	1 (4%)
3	NDG	B	619	-	15,15,15	0.66	0	21,21,21	1.10	1 (4%)
6	NAG	E	91	-	15,15,15	0.77	0	21,21,21	1.30	2 (9%)
3	NDG	B	617	-	15,15,15	0.57	0	21,21,21	1.27	2 (9%)
3	NDG	B	616	-	15,15,15	0.63	0	21,21,21	0.59	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
3	NDG	A	616	-	-	5/6/26/26	0/1/1/1
3	NDG	B	618	-	-	3/6/26/26	0/1/1/1
3	NDG	B	619	-	-	2/6/26/26	0/1/1/1
6	NAG	E	91	-	-	4/6/26/26	0/1/1/1
3	NDG	B	617	-	-	3/6/26/26	0/1/1/1
3	NDG	B	616	-	-	4/6/26/26	0/1/1/1

There are no bond length outliers.

The worst 5 of 6 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
3	B	617	NDG	C1-C2-N2	3.37	114.63	110.73
6	E	91	NAG	O5-C1-C2	3.07	112.60	109.52
6	E	91	NAG	C1-O5-C5	3.01	119.47	113.65
3	B	618	NDG	O5-C1-C2	2.87	112.40	109.52
3	B	617	NDG	C4-C3-C2	2.75	114.41	110.40

There are no chirality outliers.

5 of 21 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	A	616	NDG	C1-C2-N2-C7
3	A	616	NDG	C8-C7-N2-C2
3	A	616	NDG	O7-C7-N2-C2
3	B	616	NDG	C8-C7-N2-C2
3	B	616	NDG	O7-C7-N2-C2

There are no ring outliers.

3 monomers are involved in 8 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	A	616	NDG	4	0
3	B	619	NDG	4	0
6	E	91	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	597/597 (100%)	0.30	34 (5%)	23	19	47, 80, 130, 154	0
1	B	597/597 (100%)	0.18	23 (3%)	39	35	38, 74, 130, 156	0
2	E	173/179 (96%)	0.30	7 (4%)	38	33	56, 89, 132, 134	0
2	F	173/179 (96%)	0.28	8 (4%)	32	29	58, 90, 131, 134	0
All	All	1540/1552 (99%)	0.25	72 (4%)	31	28	38, 80, 131, 156	0

The worst 5 of 72 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	615	ASP	6.9
1	A	290	ASN	6.7
1	A	105	SER	5.4
1	A	106	SER	4.7
1	A	302	TRP	4.6

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
6	NAG	E	91	15/15	0.66	0.28	137,140,141,142	0
3	NDG	B	618	15/15	0.75	0.23	111,113,114,114	0
3	NDG	B	616	15/15	0.79	0.35	123,125,125,126	0
3	NDG	B	619	15/15	0.82	0.18	72,76,81,81	0
3	NDG	B	617	15/15	0.82	0.30	126,128,129,129	0
4	ZN	A	901	1/1	0.85	0.43	123,123,123,123	0
3	NDG	A	616	15/15	0.92	0.12	109,113,114,114	0
5	CL	A	902	1/1	0.93	0.20	84,84,84,84	0
4	ZN	B	901	1/1	0.93	0.29	131,131,131,131	0
5	CL	B	902	1/1	0.95	0.17	73,73,73,73	0

6.5 Other polymers (i)

There are no such residues in this entry.