wwPDB X-ray Structure Validation Summary Report (i)

Jan 31, 2016 – 10:04 PM GMT

|		
PDB ID	:	1RUZ
Title	:	1918 H1 Hemagglutinin
Authors	:	Skehel, J.J.; Gamblin, S.J.; Haire, L.F.; Russell, R.J.; Stevens, D.J.; Xiao, B.; Ha, Y.; Vasisht, N.; Steinhauer, D.A.; Daniels, R.S.
Deposited on	:	2003-12-12
Resolution	:	2.90 Å(reported)

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	Percentile Ranks	Value
Clashscore		24
Ramachandran outliers		2.9%
Sidechain outliers		6.4%

Worse   *Better*

▌ Percentile relative to all X-ray structures

▯ Percentile relative to X-ray structures of similar resolution
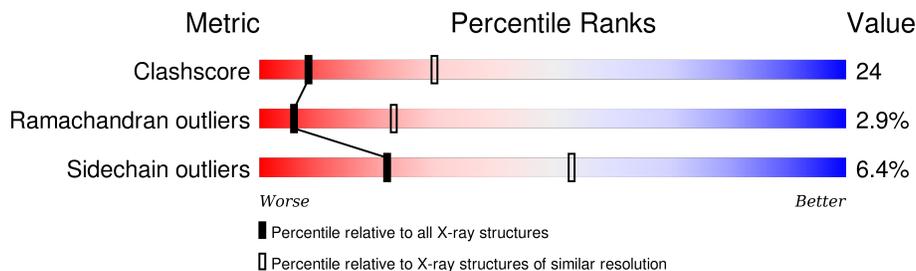
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
Clashscore	102246	1668 (2.90-2.90)
Ramachandran outliers	100387	1630 (2.90-2.90)
Sidechain outliers	100360	1632 (2.90-2.90)

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 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 <=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.

Note EDS was not executed.

Mol	Chain	Length	Quality of chain
1	H	328	64% 30% • •
1	J	328	65% 30% • •
1	L	328	63% 33% • •
2	I	160	65% 25% 8% •
2	K	160	63% 28% 8% •
2	M	160	63% 26% 9% •

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	NDG	H	640	-	-	X	-
3	NDG	J	650	-	-	X	-
3	NDG	L	660	-	-	X	-

2 Entry composition (i)

There are 5 unique types of molecules in this entry. The entry contains 12108 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 hemagglutinin.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	H	324	Total 2516	C 1586	N 434	O 485	S 11	0	0	0
1	J	324	Total 2516	C 1586	N 434	O 485	S 11	0	0	0
1	L	324	Total 2516	C 1586	N 434	O 485	S 11	0	0	0

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
H	326	ALA	SER	CONFLICT	UNP Q9WFZ1
H	327	ARG	ILE	CONFLICT	UNP Q9WFZ1
J	326	ALA	SER	CONFLICT	UNP Q9WFZ1
J	327	ARG	ILE	CONFLICT	UNP Q9WFZ1
L	326	ALA	SER	CONFLICT	UNP Q9WFZ1
L	327	ARG	ILE	CONFLICT	UNP Q9WFZ1

- Molecule 2 is a protein called hemagglutinin.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	I	160	Total 1282	C 802	N 220	O 254	S 6	0	0	0
2	K	160	Total 1282	C 802	N 220	O 254	S 6	0	0	0
2	M	160	Total 1282	C 802	N 220	O 254	S 6	0	0	0

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

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
3	H	1	Total	C	N	O	0	0
			15	8	1	6		
3	H	1	Total	C	N	O	0	0
			15	8	1	6		
3	H	1	Total	C	N	O	0	0
			15	8	1	6		
3	J	1	Total	C	N	O	0	0
			15	8	1	6		
3	J	1	Total	C	N	O	0	0
			15	8	1	6		
3	K	1	Total	C	N	O	0	0
			15	8	1	6		
3	L	1	Total	C	N	O	0	0
			15	8	1	6		
3	L	1	Total	C	N	O	0	0
			15	8	1	6		
3	M	1	Total	C	N	O	0	0
			15	8	1	6		

- Molecule 4 is SUGAR (N-ACETYL-D-GLUCOSAMINE) (three-letter code: NAG) (formula: $C_8H_{15}NO_6$).
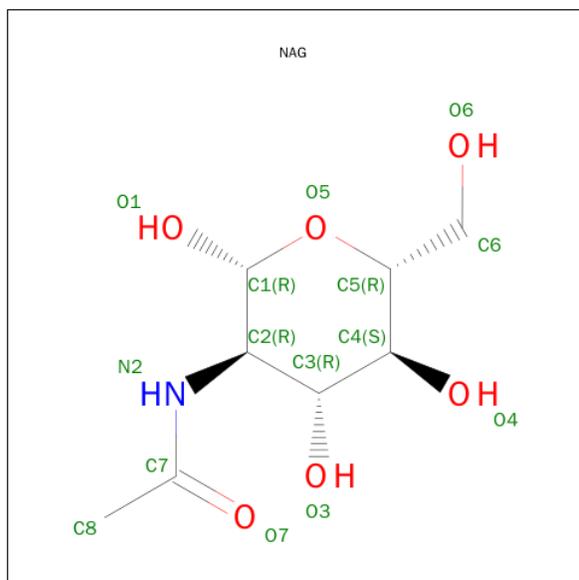
Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
4	H	1	Total 15	C 8	N 1	O 6	0	0
4	H	1	Total 15	C 8	N 1	O 6	0	0

- Molecule 5 is water.

| Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	H	95	Total 95	O 95	0	0
5	I	75	Total 75	O 75	0	0
5	J	100	Total 100	O 100	0	0
5	K	78	Total 78	O 78	0	0
5	L	128	Total 128	O 128	0	0
5	M	73	Total 73	O 73	0	0

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 errors 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.

Note EDS was not executed.

- Molecule 1: hemagglutinin
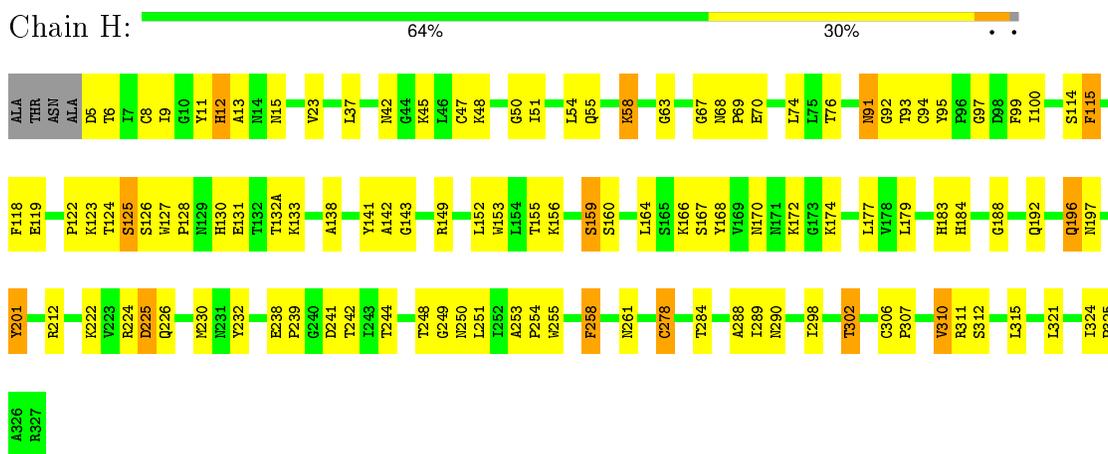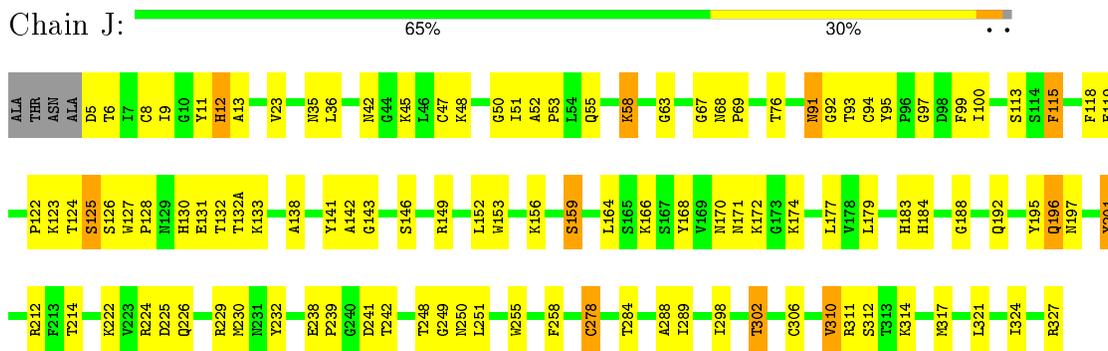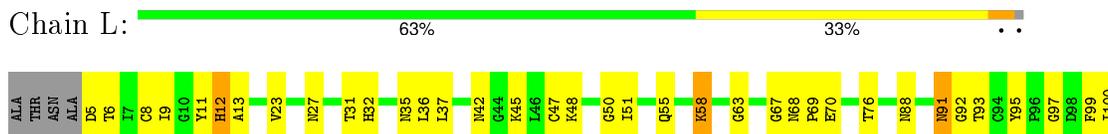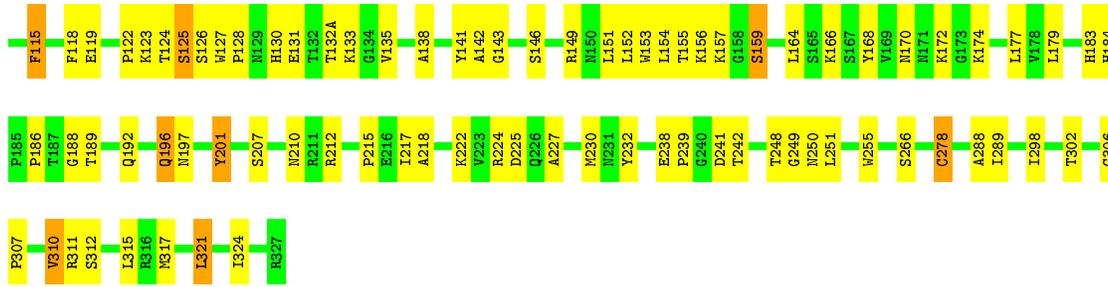
Chain H:

- Molecule 1: hemagglutinin

Chain J:

- Molecule 1: hemagglutinin

Chain L:

- Molecule 2: hemagglutinin

Chain I:

- Molecule 2: hemagglutinin

Chain K:

- Molecule 2: hemagglutinin

Chain M:
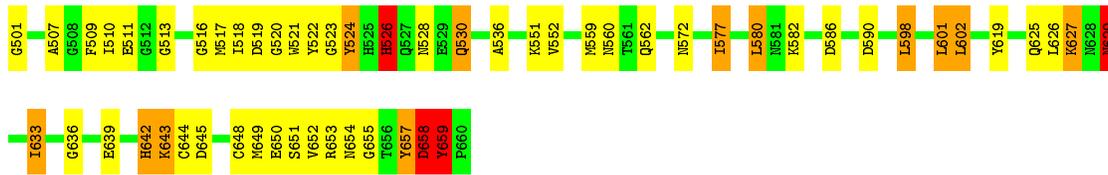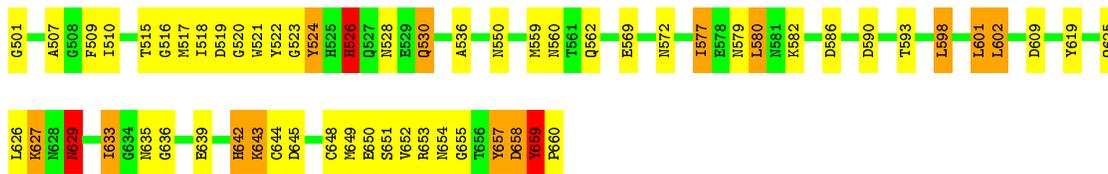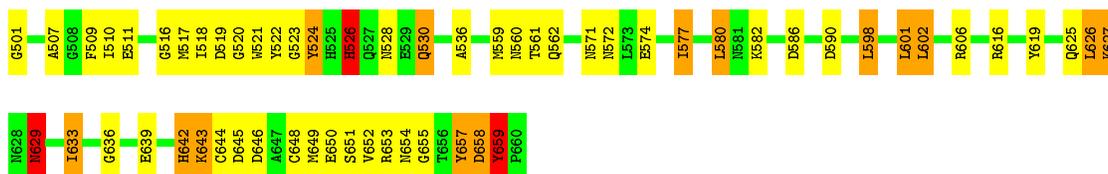
4 Data and refinement statistics (i)

Xtriage (Phenix) and EDS were not executed - this section will therefore be incomplete.

Property	Value	Source
Space group	P 41 21 2	Depositor
Cell constants a, b, c, α, β, γ	171.46Å 171.46Å 153.44Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	20.00 – 2.90	Depositor
% Data completeness (in resolution range)	(Not available) (20.00-2.90)	Depositor
R_{merge}	0.09	Depositor
R_{sym}	(Not available)	Depositor
Refinement program	CNS 1.0	Depositor
R, R_{free}	0.248 , 0.289	Depositor
Estimated twinning fraction	No twinning to report.	Xtriage
Total number of atoms	12108	wwPDB-VP
Average B, all atoms (Å^2)	73.0	wwPDB-VP

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: NAG, 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	H	0.39	0/2580	0.62	0/3514
1	J	0.40	0/2580	0.62	0/3514
1	L	0.40	0/2580	0.62	0/3514
2	I	0.39	0/1308	0.57	0/1762
2	K	0.39	0/1308	0.57	0/1762
2	M	0.40	0/1308	0.57	0/1762
All	All	0.39	0/11664	0.60	0/15828

There are no bond length outliers.

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	H	2516	0	2442	120	0
1	J	2516	0	2442	117	0
1	L	2516	0	2442	133	0
2	I	1282	0	1200	66	0
2	K	1282	0	1200	73	0
2	M	1282	0	1200	65	0
3	H	45	0	45	18	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
3	J	30	0	30	16	0
3	K	15	0	15	4	0
3	L	30	0	30	11	0
3	M	15	0	15	1	0
4	H	30	0	30	8	0
5	H	95	0	0	14	0
5	I	75	0	0	14	0
5	J	100	0	0	20	0
5	K	78	0	0	14	0
5	L	128	0	0	32	0
5	M	73	0	0	12	0
All	All	12108	0	11091	548	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 24.

The worst 5 of 548 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:H:13:ALA:HB1	5:H:699:HOH:O	1.68	0.94
1:H:290:ASN:CG	4:H:644:NAG:HN2	1.72	0.92
2:K:633:ILE:HD11	2:K:639:GLU:HB2	1.54	0.90
2:I:629:ASN:HD22	2:I:629:ASN:N	1.69	0.89
2:M:633:ILE:HD11	2:M:639:GLU:HB2	1.52	0.89

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	H	322/328 (98%)	282 (88%)	36 (11%)	4 (1%)	16	48

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	J	322/328 (98%)	282 (88%)	36 (11%)	4 (1%)	16	48
1	L	322/328 (98%)	281 (87%)	37 (12%)	4 (1%)	16	48
2	I	158/160 (99%)	131 (83%)	17 (11%)	10 (6%)	2	5
2	K	158/160 (99%)	133 (84%)	16 (10%)	9 (6%)	2	6
2	M	158/160 (99%)	131 (83%)	16 (10%)	11 (7%)	1	4
All	All	1440/1464 (98%)	1240 (86%)	158 (11%)	42 (3%)	6	23

5 of 42 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	H	125	SER
2	I	643	LYS
2	I	659	TYR
1	J	125	SER
2	K	643	LYS

5.3.2 Protein sidechains ⓘ

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	H	279/281 (99%)	265 (95%)	14 (5%)	30	65
1	J	279/281 (99%)	266 (95%)	13 (5%)	32	68
1	L	279/281 (99%)	268 (96%)	11 (4%)	39	75
2	I	135/135 (100%)	121 (90%)	14 (10%)	9	26
2	K	135/135 (100%)	121 (90%)	14 (10%)	9	26
2	M	135/135 (100%)	121 (90%)	14 (10%)	9	26
All	All	1242/1248 (100%)	1162 (94%)	80 (6%)	22	53

5 of 80 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	J	278	CYS

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
2	K	572	ASN
2	M	601	LEU
1	J	284	THR
1	J	321	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 44 such sidechains are listed below:

Mol	Chain	Res	Type
1	J	191	GLN
2	K	562	GLN
2	M	572	ASN
1	J	196	GLN
1	J	276	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 carbohydrates in this entry.

5.6   Ligand geometry (i)

11 ligands are modelled in this entry.

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	H	640	-	15,15,15	0.40	0	17,21,21	0.55	0
4	NAG	H	641	-	15,15,15	0.70	0	17,21,21	1.04	1 (5%)
4	NAG	H	644	-	15,15,15	0.59	0	17,21,21	0.53	0
3	NDG	H	645	-	15,15,15	0.61	0	17,21,21	0.63	0
3	NDG	H	646	-	15,15,15	0.79	0	17,21,21	0.56	0
3	NDG	J	650	-	15,15,15	0.38	0	17,21,21	0.56	0
3	NDG	J	651	-	15,15,15	0.47	0	17,21,21	0.60	0
3	NDG	K	661	-	15,15,15	0.93	1 (6%)	17,21,21	0.61	0
3	NDG	L	660	-	15,15,15	0.41	0	17,21,21	0.53	0
3	NDG	L	664	-	15,15,15	0.73	0	17,21,21	0.57	0
3	NDG	M	665	-	15,15,15	0.83	0	17,21,21	0.63	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	H	640	-	-	0/6/26/26	0/1/1/1
4	NAG	H	641	-	-	0/6/26/26	0/1/1/1
4	NAG	H	644	-	-	0/6/26/26	0/1/1/1
3	NDG	H	645	-	-	0/6/26/26	0/1/1/1
3	NDG	H	646	-	-	0/6/26/26	0/1/1/1
3	NDG	J	650	-	-	0/6/26/26	0/1/1/1
3	NDG	J	651	-	-	0/6/26/26	0/1/1/1
3	NDG	K	661	-	-	0/6/26/26	0/1/1/1
3	NDG	L	660	-	-	0/6/26/26	0/1/1/1
3	NDG	L	664	-	-	0/6/26/26	0/1/1/1
3	NDG	M	665	-	-	0/6/26/26	0/1/1/1

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	K	661	NDG	C1-C2	2.92	1.56	1.53

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
4	H	641	NAG	C4-C3-C2	2.16	113.42	110.43

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

11 monomers are involved in 54 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	H	640	NDG	14	0
4	H	641	NAG	5	0
4	H	644	NAG	3	0
3	H	645	NDG	3	0
3	H	646	NDG	1	0
3	J	650	NDG	16	0
3	J	651	NDG	4	0
3	K	661	NDG	4	0
3	L	660	NDG	10	0
3	L	664	NDG	1	0
3	M	665	NDG	1	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)

EDS was not executed - this section will therefore be empty.

6.2 Non-standard residues in protein, DNA, RNA chains (i)

EDS was not executed - this section will therefore be empty.

6.3 Carbohydrates (i)

EDS was not executed - this section will therefore be empty.

6.4 Ligands (i)

EDS was not executed - this section will therefore be empty.

6.5 Other polymers (i)

EDS was not executed - this section will therefore be empty.