



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 15, 2017 – 12:43 am GMT

PDB ID : 1RUY  
Title : 1930 Swine 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 : 2.70 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](mailto: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) : **NOT EXECUTED**  
EDS : **NOT EXECUTED**  
Percentile statistics : 20161228.v01 (using entries in the PDB archive December 28th 2016)  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : recalc28949

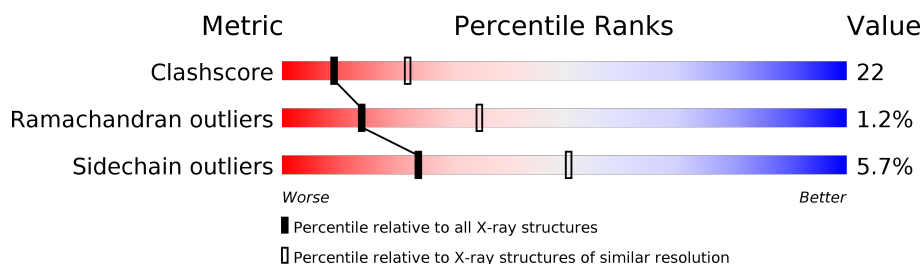
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.70 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	112137	2590 (2.70-2.70)
Ramachandran outliers	110173	2550 (2.70-2.70)
Sidechain outliers	110143	2550 (2.70-2.70)

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  $\geq 3$ , 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.

Note EDS was not executed.

Mol	Chain	Length	Quality of chain
1	H	328	
1	J	328	
1	L	328	
2	I	160	
2	K	160	
2	M	160	

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 crite-

ria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
3	NDG	H	640	-	-	X	-
3	NDG	J	643	-	-	X	-
3	NDG	J	644	-	-	X	-
3	NDG	L	646	-	-	X	-
4	NAG	J	645	-	-	X	-

## 2 Entry composition

There are 5 unique types of molecules in this entry. The entry contains 11966 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.

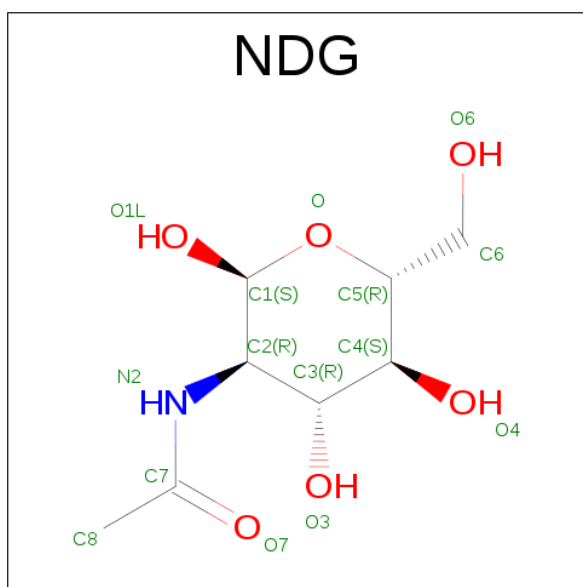
- Molecule 1 is a protein called hemagglutinin.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	H	324	Total	C	N	O	S	0	0	0
			2515	1584	442	478	11			
1	J	324	Total	C	N	O	S	0	0	0
			2515	1584	442	478	11			
1	L	324	Total	C	N	O	S	0	0	0
			2515	1584	442	478	11			

- Molecule 2 is a protein called hemagglutinin.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	I	160	Total	C	N	O	S	0	0	0
			1284	805	219	255	5			
2	K	160	Total	C	N	O	S	0	0	0
			1284	805	219	255	5			
2	M	160	Total	C	N	O	S	0	0	0
			1284	805	219	255	5			

- 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	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	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		

- 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	C	N	O	0	0
			15	8	1	6		
4	H	1	Total	C	N	O	0	0
			15	8	1	6		
4	J	1	Total	C	N	O	0	0
			15	8	1	6		
4	L	1	Total	C	N	O	0	0
			15	8	1	6		

- Molecule 5 is water.

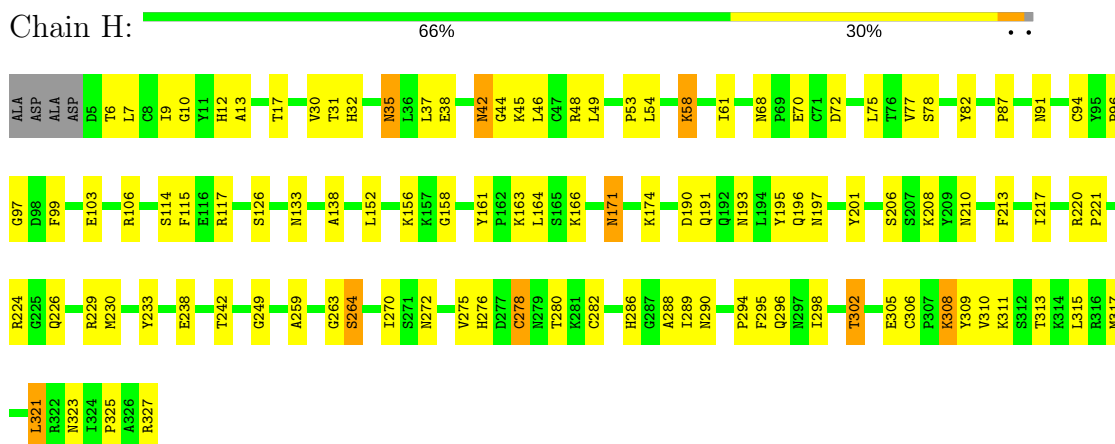
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	H	110	Total	O	0	0
			110	110		
5	I	26	Total	O	0	0
			26	26		
5	J	106	Total	O	0	0
			106	106		
5	K	39	Total	O	0	0
			39	39		
5	L	119	Total	O	0	0
			119	119		
5	M	34	Total	O	0	0
			34	34		

### 3 Residue-property plots

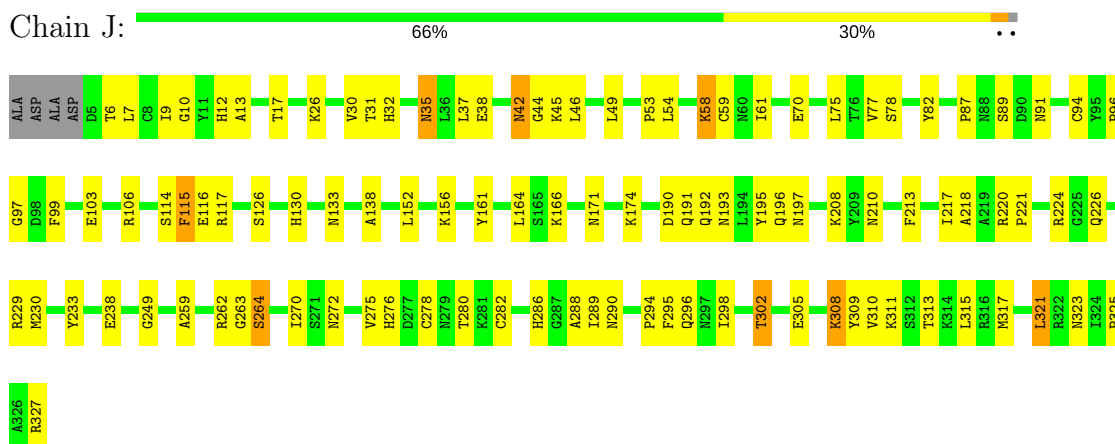
These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are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

Note EDS was not executed.

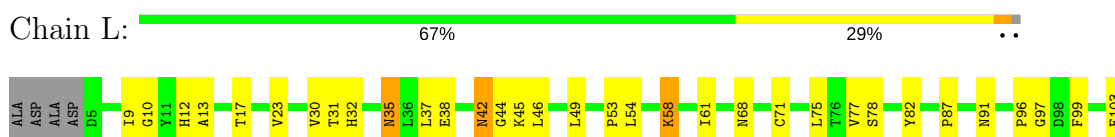
- Molecule 1: hemagglutinin

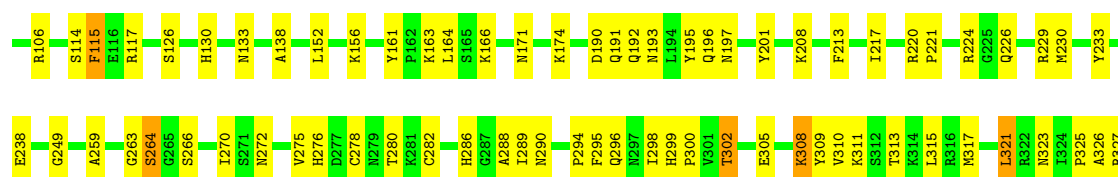


- Molecule 1: hemagglutinin



- Molecule 1: hemagglutinin





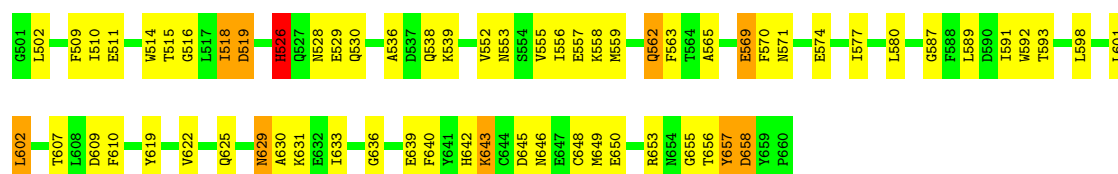
• Molecule 2: hemagglutinin

Chain I: 58% 36% 5% •



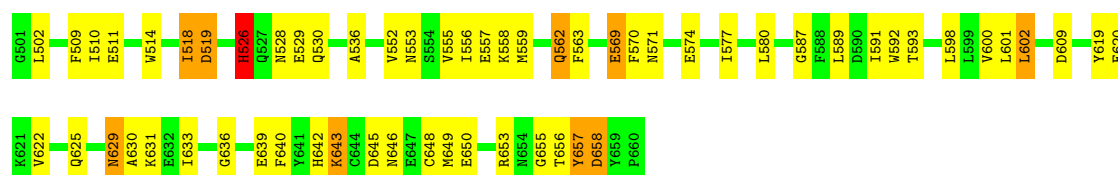
• Molecule 2: hemagglutinin

Chain K: 59% 34% 6% •



• Molecule 2: hemagglutinin

Chain M: 63% 31% 6% •





## 4 Data and refinement statistics

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

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	203.11Å 83.28Å 178.23Å 90.00° 106.67° 90.00°	Depositor
Resolution (Å)	20.00 – 2.70	Depositor
% Data completeness (in resolution range)	(Not available) (20.00-2.70)	Depositor
$R_{merge}$	0.08	Depositor
$R_{sym}$	(Not available)	Depositor
Refinement program	CNS 1.0	Depositor
R, $R_{free}$	0.248 , 0.265	Depositor
Estimated twinning fraction	No twinning to report.	Xtriage
Total number of atoms	11966	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	44.0	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	0.36	0/2579	0.64	0/3515
1	J	0.37	0/2579	0.64	0/3515
1	L	0.39	0/2579	0.65	0/3515
2	I	0.33	0/1310	0.55	0/1765
2	K	0.35	0/1310	0.56	0/1765
2	M	0.35	0/1310	0.55	0/1765
All	All	0.36	0/11667	0.61	0/15840

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	2515	0	2449	112	0
1	J	2515	0	2449	114	1
1	L	2515	0	2449	114	0
2	I	1284	0	1203	76	0
2	K	1284	0	1203	69	0
2	M	1284	0	1203	67	0
3	H	15	0	15	13	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	17	0
3	L	30	0	30	15	0
4	H	30	0	30	8	0
4	J	15	0	15	8	0
4	L	15	0	15	2	0
5	H	110	0	0	4	0
5	I	26	0	0	4	0
5	J	106	0	0	5	0
5	K	39	0	0	4	0
5	L	119	0	0	7	0
5	M	34	0	0	1	0
All	All	11966	0	11091	500	1

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 22.

The worst 5 of 500 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:J:91:ASN:ND2	3:J:643:NDG:H1	1.38	1.38
1:J:87:PRO:HG2	1:J:272:ASN:HD21	1.09	1.17
1:J:290:ASN:HD21	4:J:645:NAG:H1	1.08	1.15
1:H:87:PRO:HG2	1:H:272:ASN:HD21	1.10	1.13
1:H:58:LYS:H	1:H:58:LYS:HD3	0.95	1.11

All (1) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:J:114:SER:OG	1:J:116:GLU:OE2[2_655]	2.11	0.09

## 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%)	303 (94%)	18 (6%)	1 (0%)	44	73
1	J	322/328 (98%)	302 (94%)	19 (6%)	1 (0%)	44	73
1	L	322/328 (98%)	302 (94%)	19 (6%)	1 (0%)	44	73
2	I	158/160 (99%)	136 (86%)	17 (11%)	5 (3%)	5	11
2	K	158/160 (99%)	136 (86%)	17 (11%)	5 (3%)	5	11
2	M	158/160 (99%)	136 (86%)	17 (11%)	5 (3%)	5	11
All	All	1440/1464 (98%)	1315 (91%)	107 (7%)	18 (1%)	14	35

5 of 18 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	H	264	SER
2	I	643	LYS
2	I	657	TYR
2	I	658	ASP
1	J	264	SER

### 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%)	264 (95%)	15 (5%)	26	54
1	J	279/281 (99%)	264 (95%)	15 (5%)	26	54
1	L	279/281 (99%)	264 (95%)	15 (5%)	26	54
2	I	136/136 (100%)	128 (94%)	8 (6%)	23	49
2	K	136/136 (100%)	127 (93%)	9 (7%)	19	43
2	M	136/136 (100%)	127 (93%)	9 (7%)	19	43
All	All	1245/1251 (100%)	1174 (94%)	71 (6%)	24	51

5 of 71 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	J	278	CYS
2	K	519	ASP
2	M	562	GLN
1	J	302	THR
1	J	321	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 54 such sidechains are listed below:

Mol	Chain	Res	Type
1	J	191	GLN
2	K	528	ASN
2	M	530	GLN
1	J	193	ASN
1	J	250	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 carbohydrates in this entry.

## 5.6 Ligand geometry [i](#)

9 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	0.37	0	21,21,21	0.66	0
4	NAG	H	641	-	15,15,15	0.54	0	21,21,21	1.09	3 (14%)
4	NAG	H	642	-	15,15,15	0.50	0	21,21,21	0.55	0
3	NDG	J	643	-	15,15,15	0.55	0	21,21,21	0.79	0
3	NDG	J	644	-	15,15,15	0.42	0	21,21,21	0.55	0
4	NAG	J	645	-	15,15,15	0.50	0	21,21,21	0.71	0
3	NDG	L	646	-	15,15,15	0.34	0	21,21,21	0.71	0
4	NAG	L	647	-	15,15,15	0.48	0	21,21,21	0.59	0
3	NDG	L	648	-	15,15,15	0.44	0	21,21,21	0.57	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	-	-	1/6/26/26	0/1/1/1
4	NAG	H	642	-	-	0/6/26/26	0/1/1/1
3	NDG	J	643	-	-	0/6/26/26	0/1/1/1
3	NDG	J	644	-	-	0/6/26/26	0/1/1/1
4	NAG	J	645	-	-	0/6/26/26	0/1/1/1
3	NDG	L	646	-	-	0/6/26/26	0/1/1/1
4	NAG	L	647	-	-	0/6/26/26	0/1/1/1
3	NDG	L	648	-	-	0/6/26/26	0/1/1/1

There are no bond length outliers.

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
4	H	641	NAG	C1-C2-N2	-2.01	108.40	110.73
4	H	641	NAG	C1-C2-C3	2.19	113.53	110.54
4	H	641	NAG	O5-C1-C2	2.43	111.95	109.52

There are no chirality outliers.

All (1) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
4	H	641	NAG	O7-C7-N2-C2

There are no ring outliers.

9 monomers are involved in 59 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	H	640	NDG	13	0
4	H	641	NAG	3	0
4	H	642	NAG	5	0
3	J	643	NDG	17	0
3	J	644	NDG	7	0
4	J	645	NAG	8	0
3	L	646	NDG	13	0
4	L	647	NAG	2	0
3	L	648	NDG	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

### 6.1 Protein, DNA and RNA chains

EDS was not executed - this section is therefore empty.

### 6.2 Non-standard residues in protein, DNA, RNA chains

EDS was not executed - this section is therefore empty.

### 6.3 Carbohydrates

EDS was not executed - this section is therefore empty.

### 6.4 Ligands

EDS was not executed - this section is therefore empty.

### 6.5 Other polymers

EDS was not executed - this section is therefore empty.