



# wwPDB X-ray Structure Validation Summary Report ⓘ

Aug 7, 2020 – 08:30 PM BST

PDB ID : 5NHU  
Title : HUMAN ALPHA THROMBIN COMPLEXED WITH ANOPHELES GAM-BIAE cE5 ANTICOAGULANT  
Authors : Ripoll-Rozada, J.; Pereira, P.J.B.  
Deposited on : 2017-03-22  
Resolution : 1.45 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

<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  
Mogul : 1.8.5 (274361), CSD as541be (2020)  
Xtriage (Phenix) : 1.13  
EDS : 2.13.1  
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) : 2.13.1

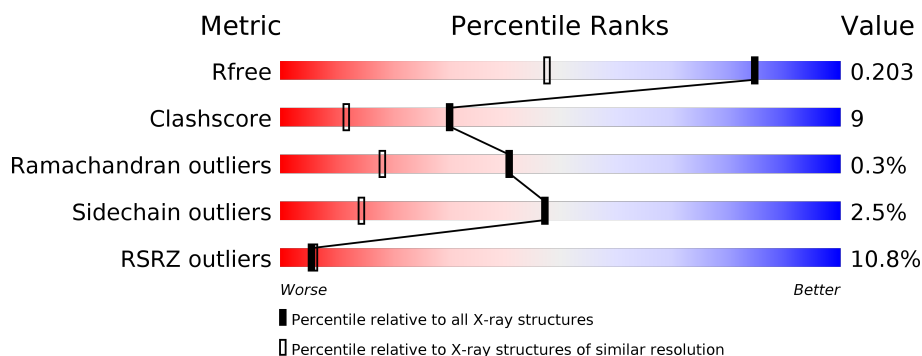
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 1.45 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	1156 (1.46-1.46)
Clashscore	141614	1202 (1.46-1.46)
Ramachandran outliers	138981	1178 (1.46-1.46)
Sidechain outliers	138945	1178 (1.46-1.46)
RSRZ outliers	127900	1139 (1.46-1.46)

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 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  $\leq 5\%$ . 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	B	36	<div> <div>6%</div> <div>86%</div> <div>6%</div> <div>8%</div> </div>
1	D	36	<div> <div>61%</div> <div>14%</div> <div>•</div> <div>•</div> <div>19%</div> </div>
1	L	36	<div> <div>14%</div> <div>92%</div> <div>6%</div> <div>•</div> </div>
2	A	259	<div> <div>2%</div> <div>88%</div> <div>8%</div> <div>•</div> </div>
2	C	259	<div> <div>26%</div> <div>54%</div> <div>39%</div> <div>•</div> <div>•</div> </div>
2	H	259	<div> <div>2%</div> <div>86%</div> <div>10%</div> <div>•</div> </div>

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
3	I	82	
3	J	82	
3	K	82	

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
4	NAG	C	601	X	-	-	-

## 2 Entry composition

There are 6 unique types of molecules in this entry. The entry contains 9088 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 Prothrombin.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	L	36	Total	C	N	O	S	0	0	0
			287	177	48	61	1			
1	B	33	Total	C	N	O	S	0	0	0
			265	162	45	57	1			
1	D	29	Total	C	N	O	S	0	1	0
			241	150	39	50	2			

- Molecule 2 is a protein called Prothrombin.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	H	252	Total	C	N	O	S	0	7	0
			2095	1335	373	373	14			
2	A	251	Total	C	N	O	S	0	8	0
			2090	1331	369	376	14			
2	C	249	Total	C	N	O	S	0	99	0
			2828	1814	498	495	21			

- Molecule 3 is a protein called AGAP008004-PA.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
3	I	24	Total	C	N	O	0	0	0
			189	118	34	37			
3	J	29	Total	C	N	O	0	0	0
			234	143	42	49			
3	K	10	Total	C	N	O	0	0	0
			84	52	17	15			

- Molecule 4 is 2-acetamido-2-deoxy-beta-D-glucopyranose (three-letter code: NAG) (formula: C<sub>8</sub>H<sub>15</sub>NO<sub>6</sub>).



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

- Molecule 5 is SODIUM ION (three-letter code: NA) (formula: Na).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	H	1	Total	Na	0	0
			1	1		
5	A	1	Total	Na	0	0
			1	1		
5	C	1	Total	Na	0	0
			1	1		

- Molecule 6 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	L	50	Total	O	0	0
			50	50		
6	H	214	Total	O	0	0
			214	214		
6	B	38	Total	O	0	0
			38	38		

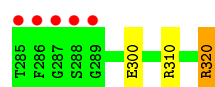
*Continued on next page...*


*Continued from previous page...*

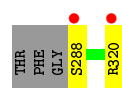
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	A	242	Total 242	O 242	0	0
6	D	23	Total 23	O 23	0	0
6	C	126	Total 128	O 128	0	6
6	I	12	Total 12	O 12	0	0
6	J	20	Total 20	O 20	0	0
6	K	3	Total 3	O 3	0	0



- Molecule 1: Prothrombin




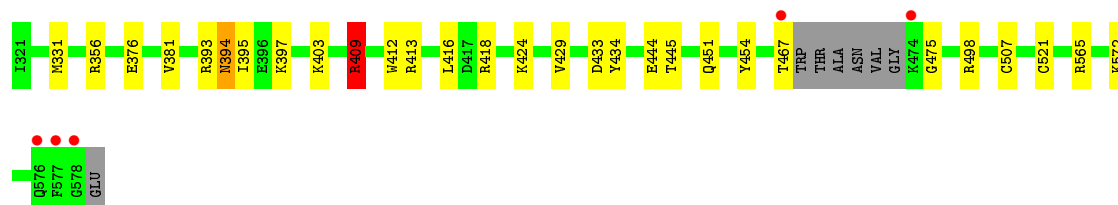
Chain B:  6% 86% 6% 8%

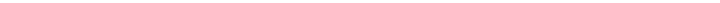


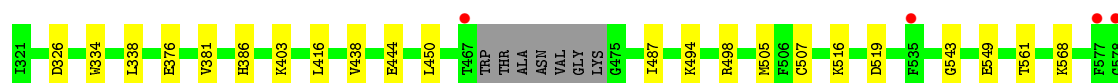
Chain D:  61% 14% . . 19%



Chain H:  29% 86% 10%



Chain A:  29% 88% 8%







## 4 Data and refinement statistics

Property	Value	Source
Space group	I 1 2 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	110.71Å 78.64Å 136.11Å 90.00° 101.84° 90.00°	Depositor
Resolution (Å)	54.18 – 1.45 54.18 – 1.45	Depositor EDS
% Data completeness (in resolution range)	98.3 (54.18-1.45) 98.3 (54.18-1.45)	Depositor EDS
$R_{merge}$	0.09	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.42 (at 1.45Å)	Xtriage
Refinement program	PHENIX (1.11.1 _2575: ???)	Depositor
R, $R_{free}$	0.179 , 0.203 0.179 , 0.203	Depositor DCC
$R_{free}$ test set	9836 reflections (4.95%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	18.2	Xtriage
Anisotropy	0.417	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.35 , 50.3	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.48$ , $\langle L^2 \rangle = 0.31$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.97	EDS
Total number of atoms	9088	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	32.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.20% of the height of the origin peak. No significant pseudotranslation is detected.*

<sup>1</sup>Intensities estimated from amplitudes.

<sup>2</sup>Theoretical 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 [i](#)

### 5.1 Standard geometry [i](#)

Bond lengths and bond angles in the following residue types are not validated in this section: NA, NAG

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	B	0.97	0/267	0.99	0/353
1	D	0.96	0/243	1.14	2/323 (0.6%)
1	L	1.00	1/290 (0.3%)	0.95	1/384 (0.3%)
2	A	0.98	0/2142	1.01	3/2893 (0.1%)
2	C	0.66	1/2902 (0.0%)	0.81	0/3920
2	H	0.92	4/2147 (0.2%)	1.01	5/2898 (0.2%)
3	I	0.83	0/193	0.98	1/261 (0.4%)
3	J	0.74	0/240	0.85	0/326
3	K	0.48	0/86	0.72	0/115
All	All	0.85	6/8510 (0.1%)	0.94	12/11473 (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
2	C	0	1

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

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	H	444	GLU	CG-CD	6.76	1.62	1.51
2	H	381	VAL	CB-CG2	-6.44	1.39	1.52
1	L	300	GLU	CD-OE2	-5.65	1.19	1.25
2	H	444	GLU	CB-CG	5.40	1.62	1.52
2	H	521	CYS	CB-SG	5.28	1.91	1.82

The worst 5 of 12 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	A	326	ASP	CB-CG-OD1	6.63	124.27	118.30
2	H	454	TYR	CB-CG-CD2	-6.19	117.28	121.00
2	A	519	ASP	CB-CG-OD1	5.86	123.58	118.30
2	H	433	ASP	CB-CG-OD2	-5.74	113.13	118.30
1	L	320	ARG	NE-CZ-NH2	-5.59	117.50	120.30

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
2	C	557	TYR	Mainchain

## 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	B	265	0	256	0	0
1	D	241	0	234	10	0
1	L	287	0	275	2	1
2	A	2090	0	2060	17	0
2	C	2828	0	2805	115	1
2	H	2095	0	2076	13	0
3	I	189	0	176	2	0
3	J	234	0	208	7	0
3	K	84	0	78	9	0
4	A	14	0	13	0	0
4	C	14	0	13	0	0
4	H	14	0	13	0	0
5	A	1	0	0	0	0
5	C	1	0	0	0	0
5	H	1	0	0	0	0
6	A	242	0	0	3	0
6	B	38	0	0	0	0
6	C	128	0	0	10	0
6	D	23	0	0	0	0
6	H	214	0	0	3	0
6	I	12	0	0	0	0
6	J	20	0	0	2	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
6	K	3	0	0	1	0
6	L	50	0	0	0	0
All	All	9088	0	8207	154	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 9.

The worst 5 of 154 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:C:401:LEU:HD22	2:C:423[B]:MET:HB3	1.51	0.92
2:C:365[B]:LEU:HD21	2:C:423[B]:MET:HE1	1.52	0.90
2:C:334:TRP:O	2:C:351[B]:SER:HA	1.75	0.87
1:L:320:ARG:NH2	2:H:451[A]:GLN:OE1	2.10	0.85
2:C:370[B]:TRP:O	2:C:372[B]:LYS:N	2.14	0.80

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:L:320:ARG:OXT	2:C:565[A]:ARG:NH2[4_544]	2.12	0.08

## 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	B	31/36 (86%)	30 (97%)	1 (3%)	0	100	100
1	D	28/36 (78%)	27 (96%)	1 (4%)	0	100	100
1	L	34/36 (94%)	33 (97%)	1 (3%)	0	100	100
2	A	255/259 (98%)	247 (97%)	8 (3%)	0	100	100

*Continued on next page...*

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
2	C	343/259 (132%)	321 (94%)	18 (5%)	4 (1%)	13	2
2	H	255/259 (98%)	247 (97%)	7 (3%)	1 (0%)	34	13
3	I	20/82 (24%)	20 (100%)	0	0	100	100
3	J	27/82 (33%)	25 (93%)	2 (7%)	0	100	100
3	K	8/82 (10%)	8 (100%)	0	0	100	100
All	All	1001/1131 (88%)	958 (96%)	38 (4%)	5 (0%)	41	9

All (5) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	C	371[A]	ASP
2	C	371[B]	ASP
2	C	418[A]	ARG
2	C	418[B]	ARG
2	H	394	ASN

### 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	B	29/31 (94%)	27 (93%)	2 (7%)	15	1
1	D	27/31 (87%)	22 (82%)	5 (18%)	1	0
1	L	31/31 (100%)	31 (100%)	0	100	100
2	A	227/225 (101%)	226 (100%)	1 (0%)	91	80
2	C	306/225 (136%)	295 (96%)	11 (4%)	35	6
2	H	227/225 (101%)	222 (98%)	5 (2%)	52	18
3	I	20/70 (29%)	20 (100%)	0	100	100
3	J	25/70 (36%)	25 (100%)	0	100	100
3	K	9/70 (13%)	9 (100%)	0	100	100
All	All	901/978 (92%)	877 (97%)	24 (3%)	47	12

5 of 24 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	D	293[B]	CYS
2	C	356[A]	ARG
2	C	565[A]	ARG
1	D	307	LYS
1	D	317	ILE

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

### 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 monosaccharides in this entry.

## 5.6 Ligand geometry ⓘ

Of 6 ligands modelled in this entry, 3 are monoatomic - leaving 3 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	$\# Z  > 2$
4	NAG	A	601	2	14,14,15	0.48	0	17,19,21	1.11	1 (5%)
4	NAG	C	601	2	14,14,15	2.60	2 (14%)	17,19,21	0.86	1 (5%)
4	NAG	H	601	2	14,14,15	0.38	0	17,19,21	0.89	1 (5%)

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
4	NAG	A	601	2	-	0/6/23/26	0/1/1/1
4	NAG	C	601	2	1/1/5/7	4/6/23/26	0/1/1/1
4	NAG	H	601	2	-	0/6/23/26	0/1/1/1

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
4	C	601	NAG	O5-C1	-9.16	1.29	1.43
4	C	601	NAG	C1-C2	-3.07	1.47	1.52

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
4	A	601	NAG	C1-O5-C5	3.97	117.56	112.19
4	C	601	NAG	C3-C4-C5	2.73	115.10	110.24
4	H	601	NAG	C1-O5-C5	2.66	115.80	112.19

All (1) chirality outliers are listed below:

Mol	Chain	Res	Type	Atom
4	C	601	NAG	C1

All (4) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
4	C	601	NAG	C8-C7-N2-C2
4	C	601	NAG	O7-C7-N2-C2
4	C	601	NAG	C4-C5-C6-O6
4	C	601	NAG	O5-C5-C6-O6

There are no ring outliers.

No monomer is involved in short contacts.

## 5.7 Other polymers ⓘ

There are no such residues in this entry.

## 5.8 Polymer linkage issues ⓘ

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, 95<sup>th</sup> 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 <sup>2</sup> )	Q<0.9
1	B	33/36 (91%)	0.17	2 (6%) 21 23	16, 24, 63, 76	0
1	D	29/36 (80%)	0.09	0 100 100	21, 30, 51, 60	0
1	L	36/36 (100%)	0.65	5 (13%) 2 3	17, 27, 63, 75	0
2	A	251/259 (96%)	-0.34	4 (1%) 72 72	13, 21, 40, 56	0
2	C	249/259 (96%)	1.22	68 (27%) 0 0	18, 38, 76, 90	0
2	H	252/259 (97%)	-0.21	5 (1%) 65 67	15, 23, 43, 63	0
3	I	24/82 (29%)	0.92	4 (16%) 1 2	19, 35, 66, 100	0
3	J	29/82 (35%)	0.96	5 (17%) 1 2	18, 32, 84, 94	0
3	K	10/82 (12%)	2.90	6 (60%) 0 0	63, 69, 75, 92	0
All	All	913/1131 (80%)	0.30	99 (10%) 5 6	13, 26, 67, 100	0

The worst 5 of 99 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
3	J	45	HIS	9.1
3	K	61	LEU	8.6
2	C	378[A]	ASP	8.4
2	C	374[A]	PHE	7.6
3	I	46	ALA	7.3

### 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, 95<sup>th</sup> 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 <sup>2</sup> )	Q<0.9
4	NAG	C	601	14/15	0.60	0.25	43,63,71,72	14
4	NAG	H	601	14/15	0.84	0.17	45,56,62,64	0
4	NAG	A	601	14/15	0.89	0.19	35,48,58,61	0
5	NA	C	602	1/1	0.97	0.15	32,32,32,32	0
5	NA	A	602	1/1	0.99	0.05	17,17,17,17	0
5	NA	H	602	1/1	0.99	0.05	20,20,20,20	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.