wwPDB X-ray Structure Validation Summary Report (i)

Mar 28, 2016 – 07:50 PM EDT

PDB ID	:	2RGU
Title	:	Crystal structure of complex of human DPP4 and inhibitor
Authors	:	Nar, H.; Himmelsbach, F.; Eckhardt, M.
Deposited on	:	2007-10-05
Resolution	:	2.60 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 (i) symbol.

The following versions of software and data (see references (1)) were used in the production of this report:

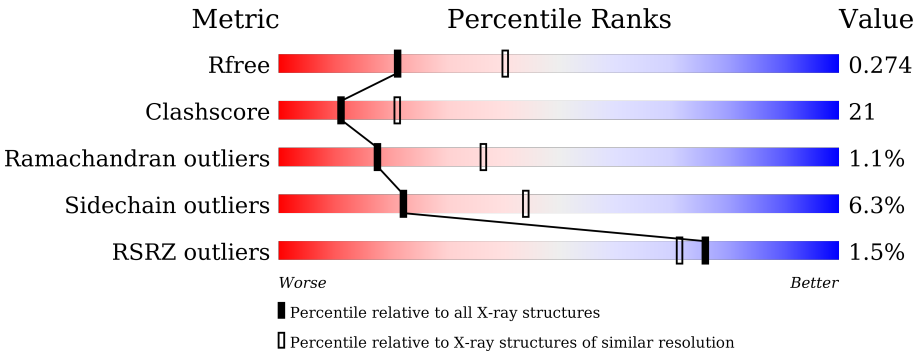
MolProbity	:	4.02b-467
Mogul	:	1.7.1 (RC1), CSD as537be (2016)
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20027107
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	rb-20027107

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 2.60 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}	91344	2328 (2.60-2.60)
Clashscore	102246	2679 (2.60-2.60)
Ramachandran outliers	100387	2635 (2.60-2.60)
Sidechain outliers	100360	2635 (2.60-2.60)
RSRZ outliers	91569	2334 (2.60-2.60)

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 <=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	734	2% 60% 35%
1	B	734	4% 64% 32%

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
2	NAG	A	795	X	-	-	-
2	NAG	A	796	X	-	-	-
2	NAG	B	793	X	-	-	-
2	NAG	B	794	X	-	-	-
2	NAG	B	796	X	-	-	-

2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 12323 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.

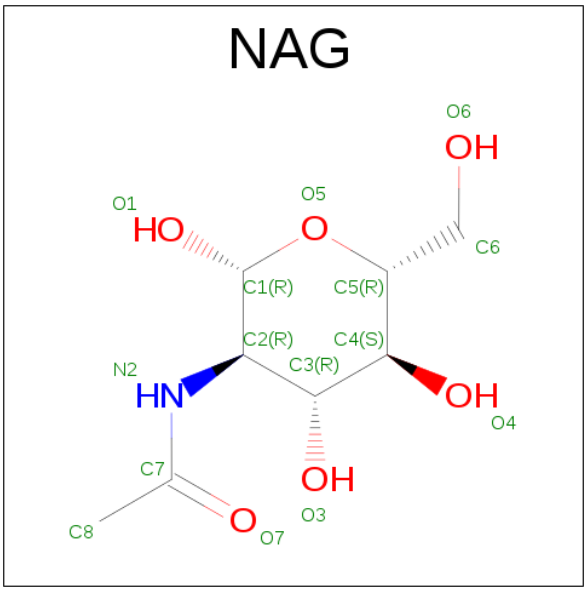
- Molecule 1 is a protein called Dipeptidyl peptidase 4.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	728	Total	C	N	O	S	0	0	0
			5963	3827	982	1128	26			
1	B	728	Total	C	N	O	S	0	0	0
			5963	3827	982	1128	26			

There are 12 discrepancies between the modelled and reference sequences:

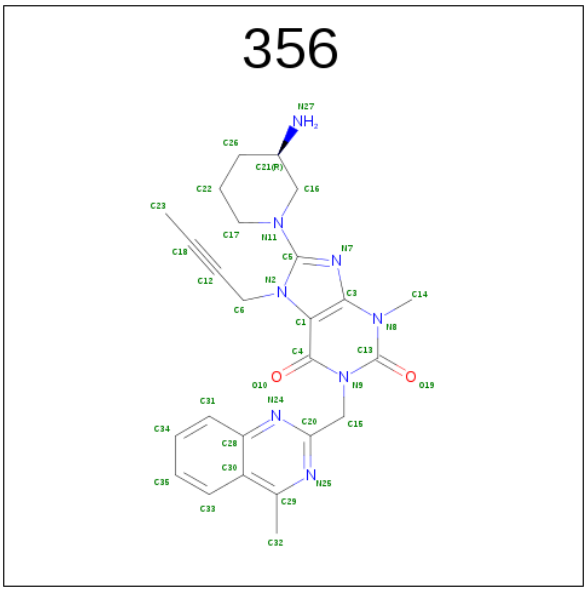
Chain	Residue	Modelled	Actual	Comment	Reference
A	767	HIS	-	EXPRESSION TAG	UNP P27487
A	768	HIS	-	EXPRESSION TAG	UNP P27487
A	769	HIS	-	EXPRESSION TAG	UNP P27487
A	770	HIS	-	EXPRESSION TAG	UNP P27487
A	771	HIS	-	EXPRESSION TAG	UNP P27487
A	772	HIS	-	EXPRESSION TAG	UNP P27487
B	767	HIS	-	EXPRESSION TAG	UNP P27487
B	768	HIS	-	EXPRESSION TAG	UNP P27487
B	769	HIS	-	EXPRESSION TAG	UNP P27487
B	770	HIS	-	EXPRESSION TAG	UNP P27487
B	771	HIS	-	EXPRESSION TAG	UNP P27487
B	772	HIS	-	EXPRESSION TAG	UNP P27487

- Molecule 2 is SUGAR (N-ACETYL-D-GLUCOSAMINE) (three-letter code: NAG) (formula: $C_8H_{15}NO_6$).



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

- Molecule 3 is 8-[(3R)-3-AMINOPIPERIDIN-1-YL]-7-BUT-2-YN-1-YL-3-METHYL-1-[(4-METHYLQUINAZOLIN-2-YL)METHYL]-3,7-DIHYDRO-1H-PURINE-2,6-DIONE (three-letter code: 356) (formula: $C_{25}H_{28}N_8O_2$).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	B	1	Total C N O 35 25 8 2	0	0
3	A	1	Total C N O 35 25 8 2	0	0

- Molecule 4 is water.

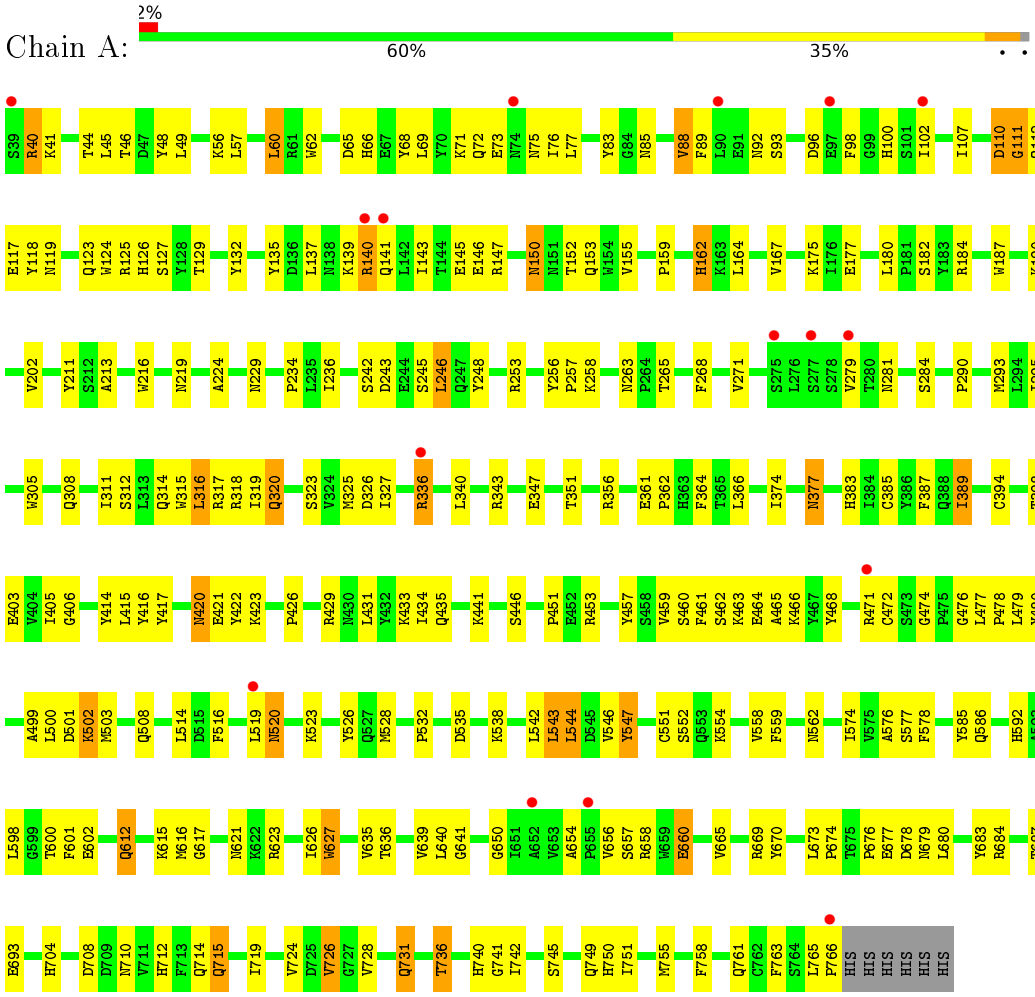
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	A	96	Total O 96 96	0	0
4	B	111	Total O 111 111	0	0

3 Residue-property plots (i)

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.

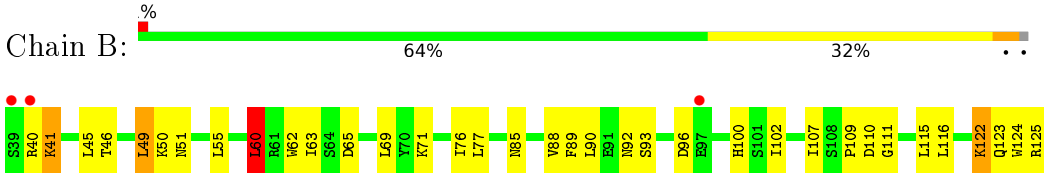
- Molecule 1: Dipeptidyl peptidase 4

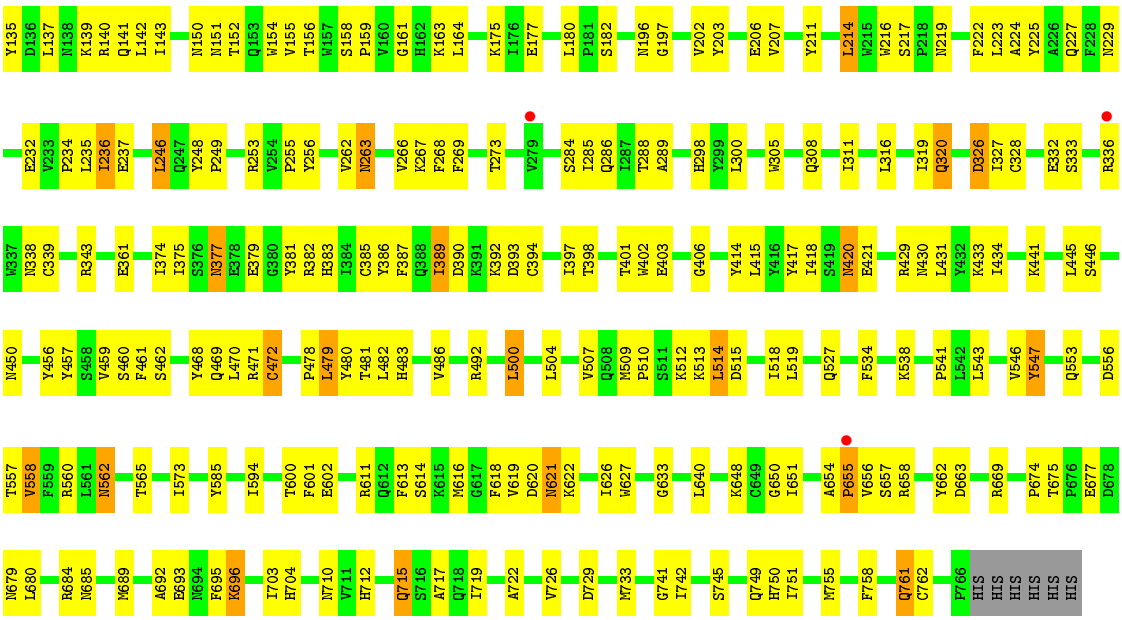
Chain A:



- Molecule 1: Dipeptidyl peptidase 4

Chain B:




Y135 D136 L137 N138 K139 R140 Q141 L142 I143 N150 N151 T152 Q153 W154 V155 T156 W157 S158 P159 Y160 G161 H162 K163 L164 K175 I176 E177 L180 P181 S182 N196 G197 Y202 Y203 E206 V207 Y211 L214 W215 W216 S217 P218 M219 F222 L223 A224 Y225 A226 Q227 F228 M229
E232 V233 P234 L235 I236 E237 L246 Q247 Y248 P249 R253 V254 P255 Y256 V262 M263 V266 K267 F268 F269 T273 V279 S284 I285 Q286 I287 T288 A289 H298 Y299 L300 W305 Q308 I311 L316 I319 Q320 D326 I327 C328 E332 S333 R336
W337 N338 C339 R343 E361 I374 I375 S376 N377 E378 E379 G380 Y381 R382 H383 I384 C385 Y386 F387 Q388 I389 D390 K391 K392 D393 C394 I397 T398 T401 W402 E403 G406 Y414 L415 Y416 Y417 I418 S419 N420 E421 R429 N430 L431 Y432 K433 I434 K441 L445 S446
N450 Y456 Y457 S458 V459 S460 F461 S462 Y468 Q469 L470 R471 C472 P478 L479 Y480 T481 L482 H483 V486 R492 L500 L504 V507 Q508 M509 P510 S511 K512 K513 L514 D515 I518 L519 Q527 F534 K538 P541 L542 L543 V546 Y547 Q553 D556
T557 V558 F559 R560 L561 N562 T565 I573 Y585 I594 T600 F601 E602 R611 Q612 F613 S614 K615 M616 G617 F618 V619 D620 N621 K622 I626 W627 G633 L640 K648 C649 G650 I651 A654 P655 V656 S657 R658 Y662 D663 R669 P674 T675 P676 E677 D678
N679 L680 R684 N685 M689 A692 E693 N694 F695 K696 I703 H704 M710 V711 H712 Q715 S716 A717 Q718 I719 A722 V726 D729 M733 G741 I742 S745 Q749 H750 I751 M755 F758 Q761 C762 P766 HIS HIS HIS HIS HIS HIS

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	65.30Å 67.10Å 419.90Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	40.00 – 2.60 33.18 – 2.60	Depositor EDS
% Data completeness (in resolution range)	97.5 (40.00-2.60) 97.6 (33.18-2.60)	Depositor EDS
R_{merge}	0.10	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.58 (at 2.61Å)	Xtriage
Refinement program	CNS	Depositor
R, R_{free}	0.217 , 0.276 0.217 , 0.274	Depositor DCC
R_{free} test set	2874 reflections (5.33%)	DCC
Wilson B-factor (Å^2)	53.5	Xtriage
Anisotropy	0.654	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.32 , 37.6	EDS
Estimated twinning fraction	0.034 for k,h,-l	Xtriage
L-test for twinning[2]	$< \|L\| > = 0.48$, $< L^2 > = 0.31$	Xtriage
Outliers	0 of 56805 reflections	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	12323	wwPDB-VP
Average B, all atoms (Å^2)	53.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.67% 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, 0.375 respectively for untwinned datasets, and 0.333, 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: 356, 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	A	0.39	0/6135	0.65	0/8344
1	B	0.41	0/6135	0.67	1/8344 (0.0%)
All	All	0.40	0/12270	0.66	1/16688 (0.0%)

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	B	60	LEU	CA-CB-CG	5.47	127.89	115.30

There are no chirality outliers.

There are no planarity outliers.

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	5963	0	5685	269	0
1	B	5963	0	5685	231	0
2	A	60	0	60	21	0
2	B	60	0	60	14	0
3	A	35	0	28	0	0
3	B	35	0	28	1	0
4	A	96	0	0	28	0
4	B	111	0	0	18	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
All	All	12323	0	11546	492	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 21.

The worst 5 of 492 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:B:85:ASN:ND2	2:B:794:NAG:H1	1.30	1.42
1:B:85:ASN:HD21	2:B:794:NAG:C1	1.43	1.31
1:A:229:ASN:HD21	2:A:796:NAG:C1	1.72	1.01
1:B:403:GLU:H	1:B:420:ASN:HD21	1.09	0.98
1:A:85:ASN:ND2	2:A:794:NAG:H1	1.79	0.96

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	726/734 (99%)	657 (90%)	59 (8%)	10 (1%)	14	28
1	B	726/734 (99%)	662 (91%)	58 (8%)	6 (1%)	24	46
All	All	1452/1468 (99%)	1319 (91%)	117 (8%)	16 (1%)	17	36

5 of 16 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	320	GLN
1	B	140	ARG
1	B	333	SER
1	B	393	ASP

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	A	111	GLY

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	A	653/659 (99%)	611 (94%)	42 (6%)	22	43
1	B	653/659 (99%)	613 (94%)	40 (6%)	23	46
All	All	1306/1318 (99%)	1224 (94%)	82 (6%)	22	44

5 of 82 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	710	ASN
1	B	51	ASN
1	B	655	PRO
1	A	715	GLN
1	A	736	THR

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 48 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	679	ASN
1	A	750	HIS
1	B	685	ASN
1	A	710	ASN
1	A	731	GLN

5.3.3 RNA

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)

10 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	$\#\|Z\| > 2$
2	NAG	A	793	-	15,15,15	0.29	0	17,21,21	0.56	0
2	NAG	A	794	-	15,15,15	0.43	0	17,21,21	0.68	0
2	NAG	A	795	-	15,15,15	0.41	0	17,21,21	0.54	0
2	NAG	A	796	-	15,15,15	0.45	0	17,21,21	0.58	0
3	356	A	901	-	27,39,39	1.63	7 (25%)	29,57,57	1.98	6 (20%)
2	NAG	B	793	-	15,15,15	0.46	0	17,21,21	0.64	0
2	NAG	B	794	-	15,15,15	0.34	0	17,21,21	0.55	0
2	NAG	B	796	-	15,15,15	0.32	0	17,21,21	0.73	0
2	NAG	B	797	-	15,15,15	0.45	0	17,21,21	0.60	0
3	356	B	902	-	27,39,39	1.89	7 (25%)	29,57,57	1.95	6 (20%)

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
2	NAG	A	793	-	-	0/6/26/26	0/1/1/1
2	NAG	A	794	-	-	0/6/26/26	0/1/1/1
2	NAG	A	795	-	1/1/6/7	0/6/26/26	0/1/1/1
2	NAG	A	796	-	1/1/6/7	0/6/26/26	0/1/1/1

Continued on next page...

Continued from previous page...

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
3	356	A	901	-	-	0/6/22/22	0/5/5/5
2	NAG	B	793	-	1/1/6/7	0/6/26/26	0/1/1/1
2	NAG	B	794	-	1/1/6/7	0/6/26/26	0/1/1/1
2	NAG	B	796	-	1/1/6/7	0/6/26/26	0/1/1/1
2	NAG	B	797	-	-	0/6/26/26	0/1/1/1
3	356	B	902	-	-	0/6/22/22	0/5/5/5

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

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	B	902	356	C3-N7	-4.95	1.28	1.33
3	B	902	356	C5-N7	-3.98	1.28	1.33
3	A	901	356	C3-N7	-3.65	1.30	1.33
3	A	901	356	C5-N7	-2.99	1.30	1.33
3	A	901	356	C17-N11	-2.29	1.43	1.46

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

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
3	B	902	356	C22-C17-N11	-4.58	102.38	111.03
3	A	901	356	C30-C28-N24	-4.44	118.82	122.89
3	B	902	356	C30-C28-N24	-4.28	118.96	122.89
3	A	901	356	C4-C1-C3	-3.78	117.23	119.93
3	A	901	356	C22-C17-N11	-3.74	103.95	111.03

All (5) chirality outliers are listed below:

Mol	Chain	Res	Type	Atom
2	A	796	NAG	C1
2	B	794	NAG	C1
2	B	793	NAG	C1
2	B	796	NAG	C1
2	A	795	NAG	C1

There are no torsion outliers.

There are no ring outliers.

9 monomers are involved in 36 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	A	793	NAG	3	0
2	A	794	NAG	6	0

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	A	795	NAG	6	0
2	A	796	NAG	6	0
2	B	793	NAG	2	0
2	B	794	NAG	6	0
2	B	796	NAG	5	0
2	B	797	NAG	1	0
3	B	902	356	1	0

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

6.1 Protein, DNA and RNA chains

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^{th} 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	728/734 (99%)	-0.10	16 (2%) 65 59	36, 55, 82, 98	0
1	B	728/734 (99%)	-0.23	6 (0%) 87 85	34, 48, 72, 86	0
All	All	1456/1468 (99%)	-0.17	22 (1%) 76 71	34, 51, 77, 98	0

The worst 5 of 22 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	39	SER	5.0
1	B	279	VAL	3.5
1	A	279	VAL	3.2
1	B	97	GLU	3.1
1	A	97	GLU	3.0

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

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

6.3 Carbohydrates

There are no carbohydrates in this entry.

6.4 Ligands

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, 95^{th} 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(Å^2)	Q<0.9
2	NAG	B	794	15/15	0.94	0.16	1.40	50,52,54,57	0
2	NAG	B	793	15/15	0.81	0.22	0.98	98,99,100,100	0
3	356	B	902	35/35	0.94	0.23	0.97	34,36,45,48	0
2	NAG	B	796	15/15	0.91	0.17	0.68	65,65,66,67	0
2	NAG	A	796	15/15	0.86	0.19	0.55	79,80,81,82	0
3	356	A	901	35/35	0.93	0.20	0.32	40,42,45,46	0
2	NAG	A	794	15/15	0.93	0.18	0.27	68,68,69,69	0
2	NAG	A	793	15/15	0.92	0.17	-0.25	95,96,96,97	0
2	NAG	B	797	15/15	0.88	0.14	-	92,93,93,93	0
2	NAG	A	795	15/15	0.82	0.32	-	99,99,100,100	0

6.5 Other polymers (i)

There are no such residues in this entry.