



wwPDB X-ray Structure Validation Summary Report ⓘ

Oct 15, 2017 – 08:09 AM EDT

PDB ID : 2RGU
Title : Crystal structure of complex of human DPP4 and inhibitor
Authors : Nar, H.; Himmelsbach, F.; Eckhardt, M.
Deposited on : unknown
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 ⓘ 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) : 1.9-1692
EDS : rb-20030345
Percentile statistics : 20161228.v01 (using entries in the PDB archive December 28th 2016)
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-20030345

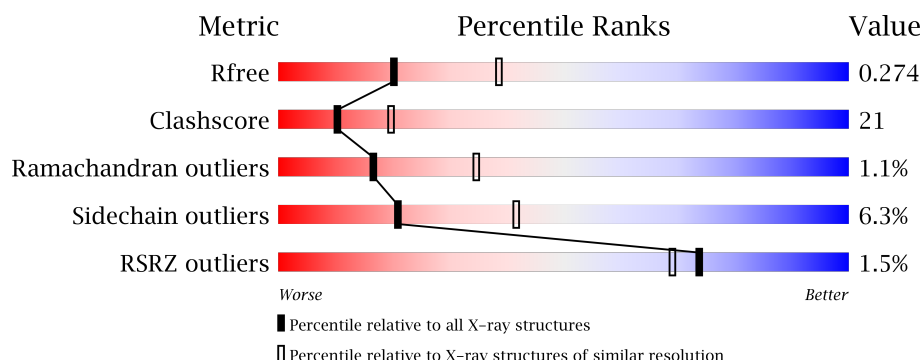
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.60 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}	100719	2542 (2.60-2.60)
Clashscore	112137	2895 (2.60-2.60)
Ramachandran outliers	110173	2848 (2.60-2.60)
Sidechain outliers	110143	2848 (2.60-2.60)
RSRZ outliers	101464	2550 (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 $\leq 5\%$. 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	A	734	<div> <div>2%</div> <div>60%</div> <div>35%</div> <div>..</div> </div>
1	B	734	<div> <div>%</div> <div>64%</div> <div>32%</div> <div>..</div> </div>

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

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:

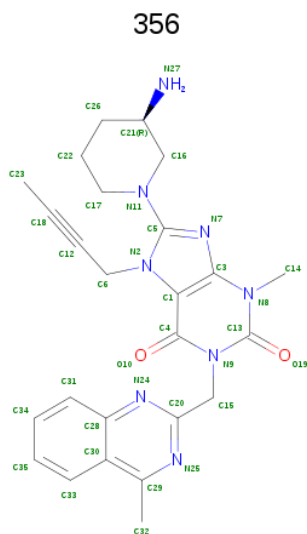
Chain	Residue	Modelled	Actual	Comment	Reference
A	767	HIS	-	EXPRESSION TAG	UNP P27487
A	768	HIS	-	EXPRESSION TAG	UNP P27487
A	769	HIS	-	EXPRESSION TAG	UNP P27487
A	770	HIS	-	EXPRESSION TAG	UNP P27487
A	771	HIS	-	EXPRESSION TAG	UNP P27487
A	772	HIS	-	EXPRESSION TAG	UNP P27487
B	767	HIS	-	EXPRESSION TAG	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 N-ACETYL-D-GLUCOSAMINE (three-letter code: NAG) (formula: C₈H₁₅NO₆).



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

- 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: C₂₅H₂₈N₈O₂).



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

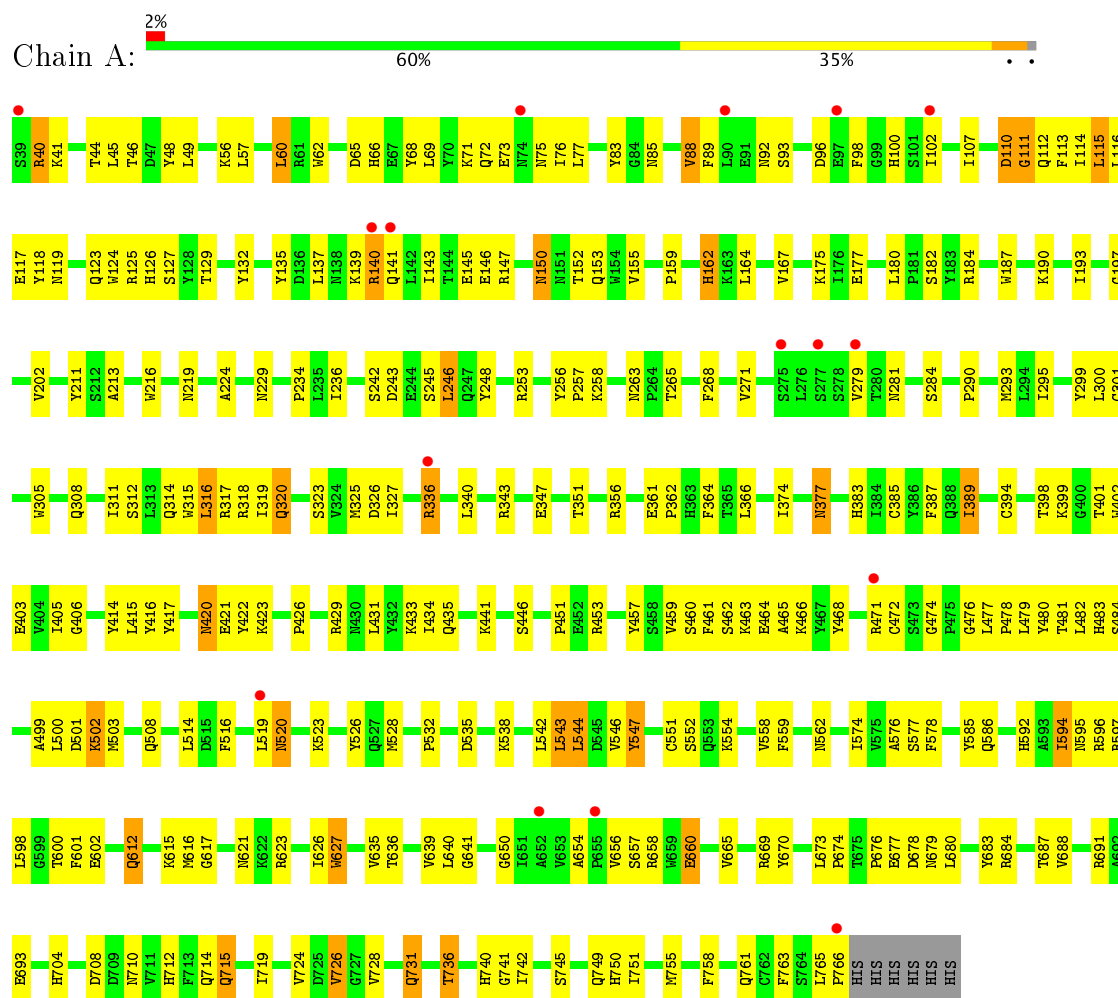
- Molecule 4 is water.

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

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 ($\text{RSRZ} > 2$). 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



N679	T557	N450	G337	E232	Y135
N680	V558	Y456	N338	V233	D136
N684	F559	Y457	C339	P234	L137
N685	R560	Y458	R343	I235	N138
N689	L561	V459	R343	I236	K139
	N562	S460	E361	E237	R140
	T565	F461			Q141
A692		S462	I374	L246	L142
E593	I573	Y468	I375	Q247	I143
T695	Y585	Q469	S376	Y248	N150
K696		L470	D377	P249	N151
	I594	R471	E378		T152
I703		O472	E379	R253	I153
H704	T600	P478	G380	V254	N154
	F601	L479	R381	Y255	V155
N710	E602	Y480	R382	Y256	T156
V711		Y481	R383		I157
H712	R611	L482	I384	V262	S158
	Q612	H483	C385	M263	P159
Q715	F613		Y386	V266	V160
S716	S614	V486	F387	K267	G161
A717	K615		Q388	F268	H162
Q718	M616	R492	I389	F269	K163
I719	G617	L500	D390		L164
	F618	L504	K391	T273	
A722	V619		K392	I176	K175
V726	D620		D393	E177	
	N621		C394	V279	
	K622	V507	I397	S284	L180
D729		Q508	T398	I285	P181
	I626	M509	T401	K286	S182
M733	W627	P510	W402	I287	
G741	G633	S511	E403	A289	N196
I742		K512			G197
	L640	K513	G406	H298	V202
S745	K648	L514	Y414	Y299	T203
Q749	C649	D515	L415	L300	
H750	G650	I518	Y416	W305	E206
I751	I651	L519	Y417		V207
			I418	Q308	Y211
M755	A654	Q527	S419		
	P655		M420	I311	D214
F758	V656	F534	E421		W215
	S657			L316	W216
Q761	R658	K538	R429	I319	S217
C762			M430	Q320	P218
	Y662	P541	L431		N219
F766	D663	L542	Y432	D326	F222
HIS		L543	K433	I327	L223
HIS	R669	V546	I434	C328	A224
HIS	P674	Y547	K441		Y225
HIS	T675			E332	A226
HIS	P676	Q553	L445	S333	Q227
HIS	E677	D556	S446	R336	F228
	D678				N229

4 Data and refinement statistics

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
$\langle I/\sigma(I) \rangle$ 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
L-test for twinning ²	$\langle L \rangle = 0.48$, $\langle L^2 \rangle = 0.31$	Xtriage
Estimated twinning fraction	0.034 for k,h,-l	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.*

¹Intensities estimated from amplitudes.

²Theoretical 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: 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(°)	Ideal(°)
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 [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	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%)	13	26
1	B	726/734 (99%)	662 (91%)	58 (8%)	6 (1%)	22	44
All	All	1452/1468 (99%)	1319 (91%)	117 (8%)	16 (1%)	17	35

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 [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	653/659 (99%)	611 (94%)	42 (6%)	20	40
1	B	653/659 (99%)	613 (94%)	40 (6%)	22	43
All	All	1306/1318 (99%)	1224 (94%)	82 (6%)	21	42

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 [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](#)

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	21,21,21	0.63	0
2	NAG	A	794	-	15,15,15	0.44	0	21,21,21	0.83	1 (4%)
2	NAG	A	795	-	15,15,15	0.42	0	21,21,21	0.55	0
2	NAG	A	796	-	15,15,15	0.43	0	21,21,21	0.71	1 (4%)
3	356	A	901	-	28,39,39	1.66	9 (32%)	33,57,57	1.82	9 (27%)
2	NAG	B	793	-	15,15,15	0.47	0	21,21,21	0.62	0
2	NAG	B	794	-	15,15,15	0.34	0	21,21,21	0.65	0
2	NAG	B	796	-	15,15,15	0.32	0	21,21,21	0.91	1 (4%)
2	NAG	B	797	-	15,15,15	0.46	0	21,21,21	0.70	0
3	356	B	902	-	28,39,39	1.90	8 (28%)	33,57,57	1.82	7 (21%)

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 17 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	-3.69	1.28	1.34
3	A	901	356	C3-N7	-3.65	1.30	1.33
3	A	901	356	C5-N7	-2.81	1.30	1.34
3	A	901	356	C17-N11	-2.25	1.43	1.46

The worst 5 of 19 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	B	902	356	C22-C17-N11	-4.43	102.38	111.09
3	A	901	356	C4-C1-C3	-3.90	117.23	119.92
3	A	901	356	C30-C28-N24	-3.89	118.82	122.84
3	B	902	356	C4-C1-C3	-3.84	117.27	119.92
3	B	902	356	C30-C28-N24	-3.76	118.96	122.84

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 [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	728/734 (99%)	-0.10	16 (2%) 62 56	36, 55, 82, 98	0
1	B	728/734 (99%)	-0.23	6 (0%) 86 83	34, 48, 72, 86	0
All	All	1456/1468 (99%)	-0.17	22 (1%) 74 69	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 [i](#)

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

6.3 Carbohydrates [i](#)

There are no carbohydrates 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. 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, 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	LLDF	B-factors(\AA^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.