wwPDB X-ray Structure Validation Summary Report (i)

Jun 13, 2024 – 12:34 PM EDT

PDB ID	:	1PFQ
Title	:	crystal structure of human apo dipeptidyl peptidase IV / CD26
Authors	:	Oefner, C.; D'Arcy, A.; Mac Sweeney, A.; Pierau, S.; Gardiner, R.; Dale, G.E.
Deposited on	:	2003-05-27
Resolution	:	1.90 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
https://www.wwpdb.org/validation/2017/XrayValidationReportHelp
with specific help available everywhere you see the (i) symbol.

The types of validation reports are described at
http://www.wwpdb.org/validation/2017/FAQs#types.

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

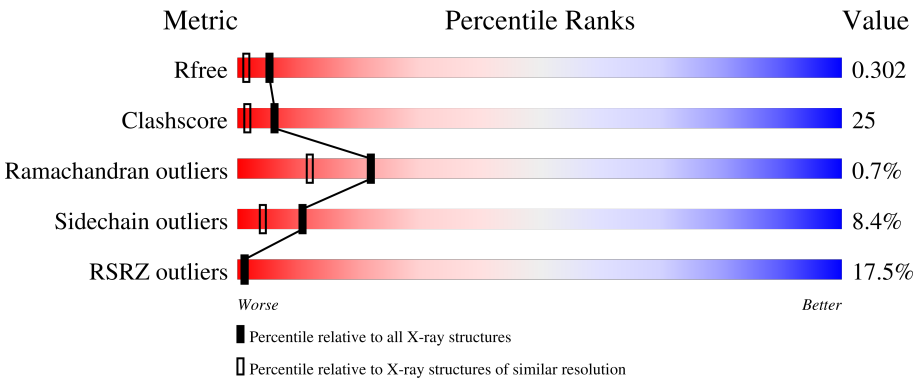
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.36.2
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36.2

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 1.90 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	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	6207 (1.90-1.90)
Clashscore	141614	6847 (1.90-1.90)
Ramachandran outliers	138981	6760 (1.90-1.90)
Sidechain outliers	138945	6760 (1.90-1.90)
RSRZ outliers	127900	6082 (1.90-1.90)

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 of the lower bar indicate the fraction of residues that contain outliers for >=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 <=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	731	16% 64% 30% 5%
1	B	731	19% 59% 34% 6% ••

2 Entry composition (i)

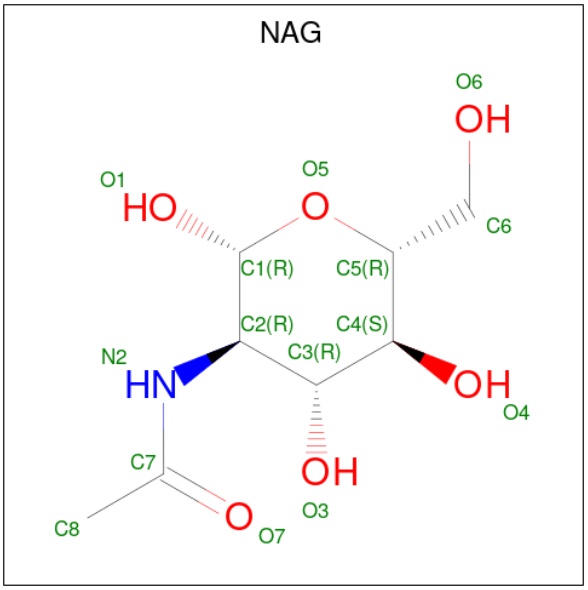
There are 3 unique types of molecules in this entry. The entry contains 12418 atoms, of which 0 are hydrogens and 0 are deuteriums.

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 IV soluble form.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	728	5964	3827	982	1129	26	0	0	0
1	B	725	5926	3801	977	1122	26	0	0	0

- Molecule 2 is 2-acetamido-2-deoxy-beta-D-glucopyranose (three-letter code: NAG) (formula: $C_8H_{15}NO_6$).



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

- Molecule 3 is water.

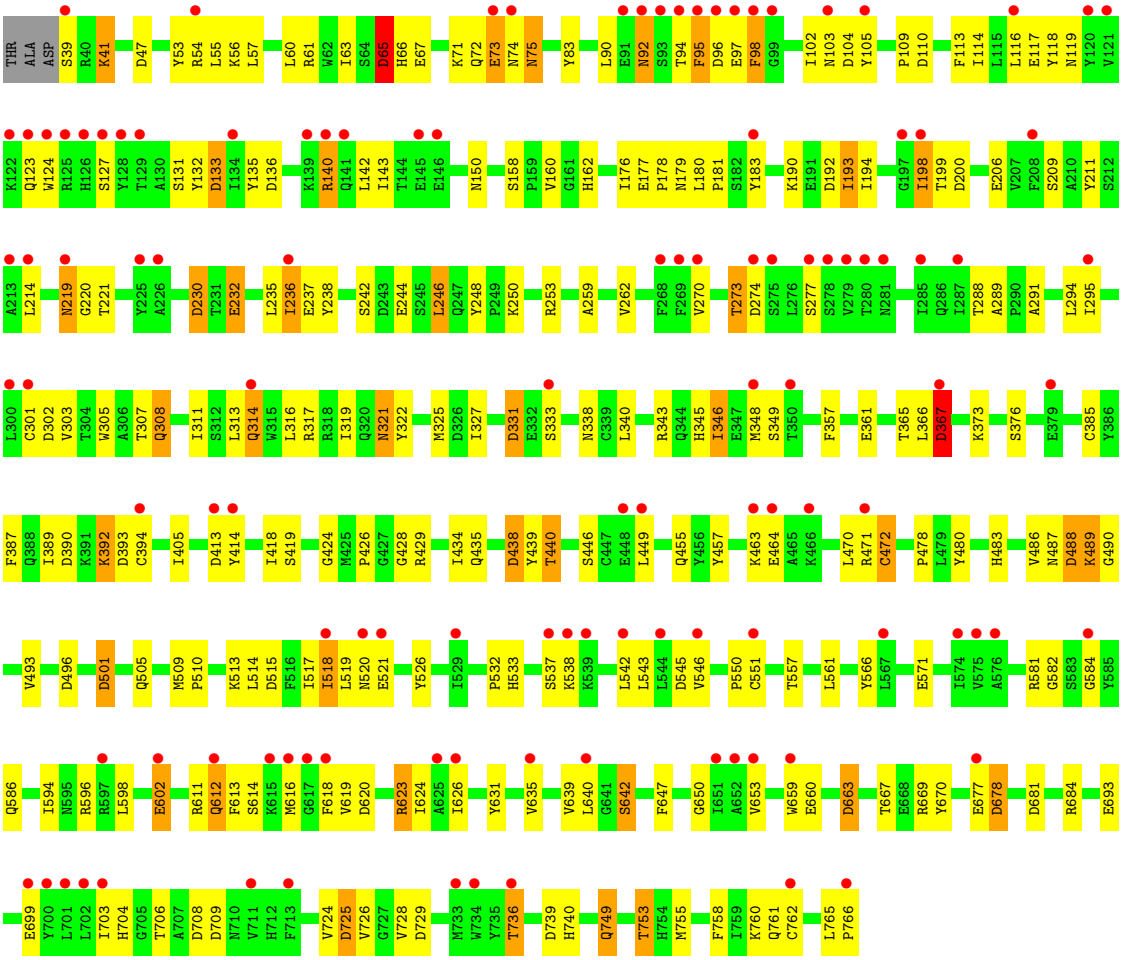
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
3	A	268	268	268	0	0
			Total	O		
3	B	204	204	204	0	0

3 Residue-property plots (i)

These plots are drawn for all protein, RNA, DNA and oligosaccharide 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.

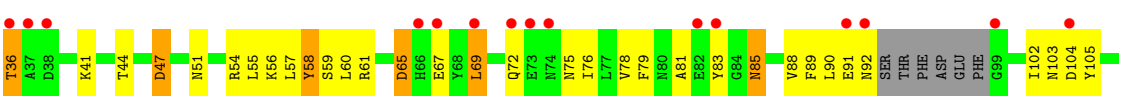
• Molecule 1: Dipeptidyl peptidase IV soluble form

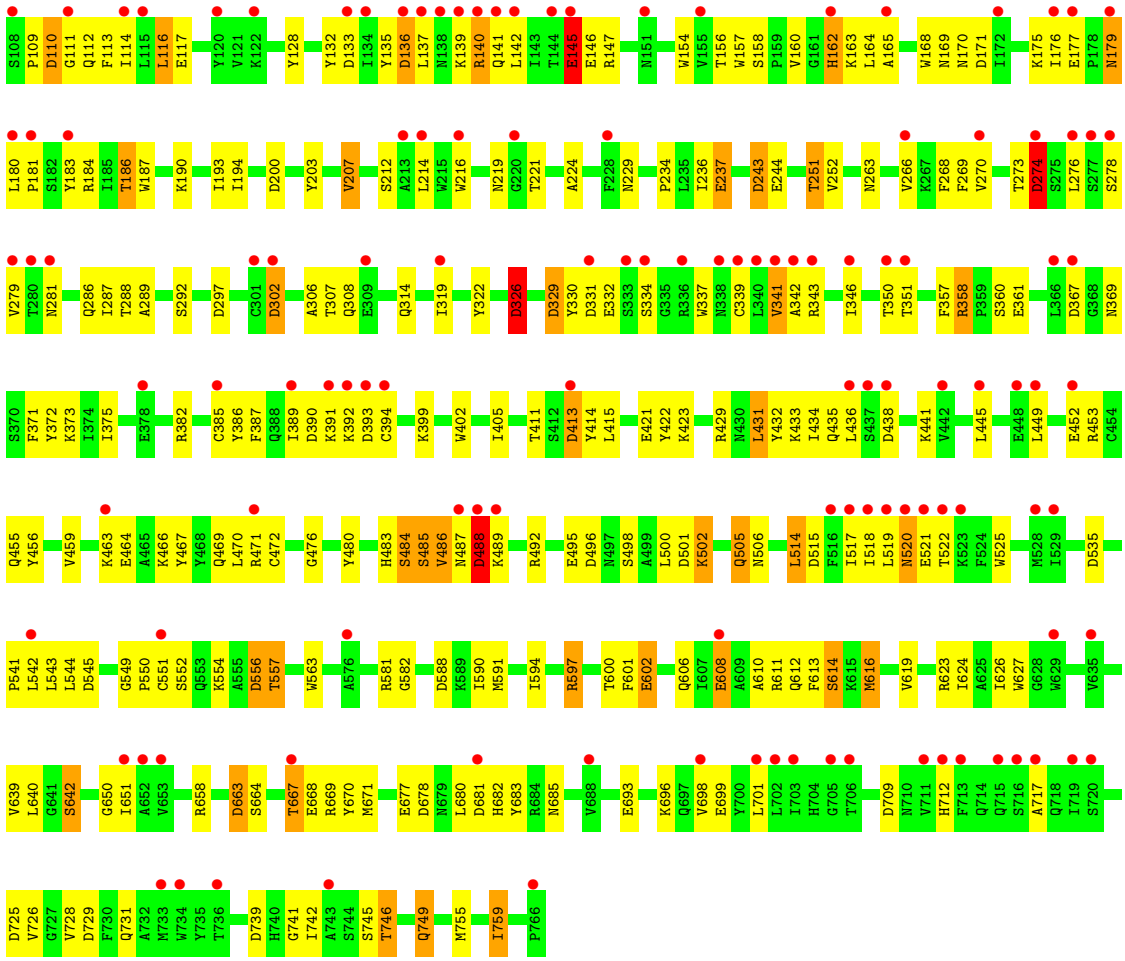
Chain A:



• Molecule 1: Dipeptidyl peptidase IV soluble form

Chain B:




S108 P109 D110 G111 Q112 F113 I114 L115 L116 E117 Y120 V121 K122 Y128 Y132 D133 I134 Y135 D136 L137 N138 K139 R140 Q141 L142 I143 T144 E145 E146 R147 N151 W154 V155 T156 W157 S158 P159 V160 G161 H162 K163 L164 A165 W168 N169 N170 D171 I172 K175 I176 E177 P178 N179
L180 P181 S182 Y183 R184 I185 T186 W187 K190 I193 I194 D200 Y203 V207 S212 A213 L214 W215 W216 N219 G220 T221 A224 F228 N229 P234 L235 I236 E237 D243 E244 T251 V252 N263 V266 K267 F268 F269 V270 T273 D274 S275 L276 S277 S278
V279 T280 N281 Q286 I287 T288 A289 S292 D297 C301 D302 A306 T307 Q308 E309 Q314 I319 Y322 D326 D329 Y330 D331 E332 S333 S334 G335 R336 W337 N338 C339 L340 V341 A342 R343 I346 T350 T351 F357 R358 P359 S360 E361 L366 D367 G368 N369
S370 F371 Y372 K373 I374 I375 E378 R382 C385 Y386 F387 Q388 I389 D390 K391 K392 D393 C394 K399 W402 I405 T411 S412 D413 Y414 L415 E421 Y422 K423 R429 N430 L431 Y432 K433 I434 Q435 L436 S437 D438 K441 V442 L445 E448 L449 E452 R453 G454
Q455 Y456 V459 K463 E464 A465 K466 Y467 Y468 Q469 L470 R471 C472 G476 Y480 H483 S484 S485 V486 N487 D488 K489 R492 E495 D496 N497 S498 A499 L500 D501 K502 Q505 N506 L514 D515 F516 I517 I518 L519 N520 E521 T522 K523 F524 W525 M528 I529 D535
P541 L542 L543 L544 D545 G549 P550 C551 S552 Q553 K554 A555 D556 T557 W563 A576 R581 G582 D588 K589 I590 M591 I594 R597 T600 F601 E602 Q606 I607 E608 A609 A610 R611 Q612 F613 S614 K615 M616 V619 R623 I624 A625 I626 W627 G628 W629 V635
V639 L640 G641 S642 G650 I651 A652 V653 R658 D663 S664 T667 E668 R669 Y670 M671 E677 D678 N679 L680 D681 H682 Y683 R684 N685 V688 E693 K696 Q697 V698 E699 Y700 L701 L702 I703 H704 G705 T706 D709 N710 V711 H712 F713 Q714 Q715 S716 A717 Q718 I719 S720
D725 V726 G727 V728 D729 F730 Q731 A732 M733 W734 Y735 T736 D739 H740 G741 I742 A743 S744 S745 T746 Q749 M755 I759 P766

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	71.03Å 118.14Å 184.58Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	20.00 – 1.90 29.77 – 1.90	Depositor EDS
% Data completeness (in resolution range)	98.7 (20.00-1.90) 98.7 (29.77-1.90)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.13	Depositor
$< I/\sigma(I) >$ [1]	1.64 (at 1.91Å)	Xtriage
Refinement program	REFMAC 5.0	Depositor
R, R_{free}	0.254 , 0.298 0.259 , 0.302	Depositor DCC
R_{free} test set	11796 reflections (5.06%)	wwPDB-VP
Wilson B-factor (Å^2)	37.5	Xtriage
Anisotropy	0.196	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.35 , 61.9	EDS
L-test for twinning[2]	$< \|L\| > = 0.48$, $< L^2 > = 0.31$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.93	EDS
Total number of atoms	12418	wwPDB-VP
Average B, all atoms (Å^2)	42.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.76% 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, 0.333 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: 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	0.27	0/6136	0.60	27/8344 (0.3%)
1	B	0.26	0/6094	0.60	33/8287 (0.4%)
All	All	0.26	0/12230	0.60	60/16631 (0.4%)

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
1	B	0	1

There are no bond length outliers.

The worst 5 of 60 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	B	545	ASP	CB-CG-OD2	6.25	123.93	118.30
1	A	545	ASP	CB-CG-OD2	5.91	123.62	118.30
1	B	725	ASP	CB-CG-OD2	5.82	123.54	118.30
1	A	200	ASP	CB-CG-OD2	5.79	123.51	118.30
1	A	725	ASP	CB-CG-OD2	5.79	123.51	118.30

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	B	581	ARG	Peptide

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	5964	0	5683	271	3
1	B	5926	0	5650	314	3
2	A	28	0	26	1	0
2	B	28	0	26	1	0
3	A	268	0	0	26	0
3	B	204	0	0	16	2
All	All	12418	0	11385	573	5

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 25.

The worst 5 of 573 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:A:517:ILE:HG21	1:A:612:GLN:NE2	1.28	1.45
1:A:517:ILE:CG2	1:A:612:GLN:HE22	1.37	1.36
1:A:198:ILE:HD11	1:A:211:TYR:CE2	1.73	1.22
1:B:544:LEU:HD21	1:B:606:GLN:NE2	1.56	1.20
1:A:150:ASN:ND2	3:A:1095:HOH:O	1.71	1.17

All (5) 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)
3:B:1043:HOH:O	3:B:1054:HOH:O[3_555]	1.53	0.67
1:A:98:PHE:CZ	1:B:168:TRP:CZ3[1_455]	1.81	0.39
1:A:83:TYR:CD1	1:A:677:GLU:OE1[1_455]	1.89	0.31
1:B:642:SER:O	3:B:1017:HOH:O[1_455]	1.91	0.29
1:A:98:PHE:CZ	1:B:168:TRP:CH2[1_455]	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	A	726/731 (99%)	678 (93%)	45 (6%)	3 (0%)	34 24
1	B	721/731 (99%)	667 (92%)	47 (6%)	7 (1%)	15 6
All	All	1447/1462 (99%)	1345 (93%)	92 (6%)	10 (1%)	22 12

5 of 10 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	140	ARG
1	B	140	ARG
1	B	341	VAL
1	A	98	PHE
1	B	274	ASP

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	653/655 (100%)	596 (91%)	57 (9%)	10 4
1	B	647/655 (99%)	595 (92%)	52 (8%)	12 5
All	All	1300/1310 (99%)	1191 (92%)	109 (8%)	11 4

5 of 109 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	36	THR

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	B	251	THR
1	B	608	GLU
1	B	54	ARG
1	B	145	GLU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 33 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	505	GLN
1	B	606	GLN
1	B	749	GLN
1	A	505	GLN
1	A	487	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 monosaccharides in this entry.

5.6 Ligand geometry (i)

4 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	B	852	1	14,14,15	0.56	0	17,19,21	0.81	1 (5%)
2	NAG	A	853	1	14,14,15	0.55	0	17,19,21	0.95	1 (5%)
2	NAG	A	851	1	14,14,15	0.61	0	17,19,21	0.83	1 (5%)
2	NAG	B	855	1	14,14,15	0.57	0	17,19,21	0.72	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
2	NAG	B	852	1	-	0/6/23/26	0/1/1/1
2	NAG	A	853	1	-	2/6/23/26	0/1/1/1
2	NAG	A	851	1	-	0/6/23/26	0/1/1/1
2	NAG	B	855	1	-	2/6/23/26	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(^{o})	Ideal(^{o})
2	A	853	NAG	O5-C1-C2	-2.45	107.42	111.29
2	A	851	NAG	O5-C5-C6	2.19	110.64	107.20
2	B	852	NAG	O5-C5-C6	2.15	110.58	107.20

There are no chirality outliers.

All (4) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	B	855	NAG	C4-C5-C6-O6
2	B	855	NAG	O5-C5-C6-O6
2	A	853	NAG	C4-C5-C6-O6
2	A	853	NAG	O5-C5-C6-O6

There are no ring outliers.

2 monomers are involved in 2 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	A	853	NAG	1	0
2	B	855	NAG	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	728/731 (99%)	1.04	116 (15%)	1	2	14, 40, 55, 60	0
1	B	725/731 (99%)	1.26	138 (19%)	1	1	17, 43, 64, 72	0
All	All	1453/1462 (99%)	1.15	254 (17%)	1	1	14, 41, 59, 72	0

The worst 5 of 254 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	96	ASP	11.2
1	A	95	PHE	10.8
1	B	340	LEU	10.7
1	B	279	VAL	9.1
1	A	766	PRO	8.9

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, 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	B-factors(Å^2)	Q<0.9
2	NAG	B	855	14/15	0.63	0.25	68,69,69,69	0
2	NAG	B	852	14/15	0.87	0.14	42,42,44,45	0
2	NAG	A	853	14/15	0.89	0.15	42,43,43,44	0
2	NAG	A	851	14/15	0.90	0.10	48,48,49,49	0

6.5 Other polymers (i)

There are no such residues in this entry.