wwPDB X-ray Structure Validation Summary Report (i)

Oct 16, 2023 – 11:32 AM EDT

PDB ID	:	1UHV
Title	:	Crystal structure of beta-D-xylosidase from Thermoanaerobacterium saccharolyticum, a family 39 glycoside hydrolase
Authors	:	Yang, J.K.; Yoon, H.J.; Ahn, H.J.; Il Lee, B.; Pedelacq, J.D.; Liong, E.C.; Berendzen, J.; Laivenieks, M.; Vieille, C.; Zeikus, G.J.; Vocadlo, D.J.; Withers, S.G.; Suh, S.W.
Deposited on	:	2003-07-11
Resolution	:	2.10 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:

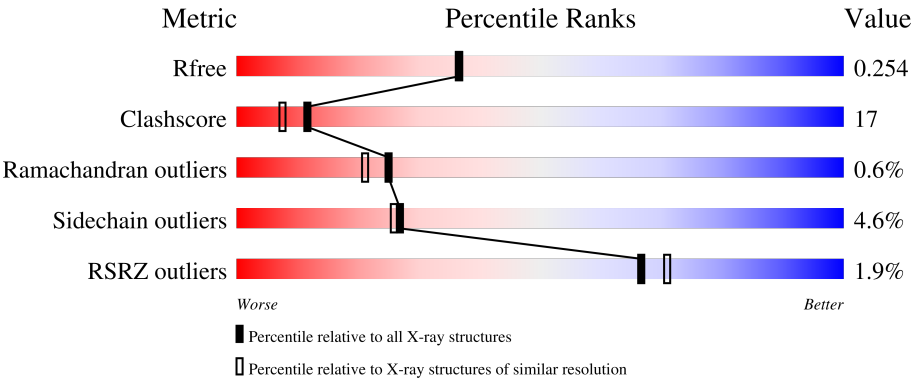
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36

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 2.10 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	5197 (2.10-2.10)
Clashscore	141614	5710 (2.10-2.10)
Ramachandran outliers	138981	5647 (2.10-2.10)
Sidechain outliers	138945	5648 (2.10-2.10)
RSRZ outliers	127900	5083 (2.10-2.10)

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	500	3% 62% 35% ·
1	B	500	2% 64% 32% ·
1	C	500	% 69% 28% ·
1	D	500	2% 64% 34% ·

2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 17868 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.

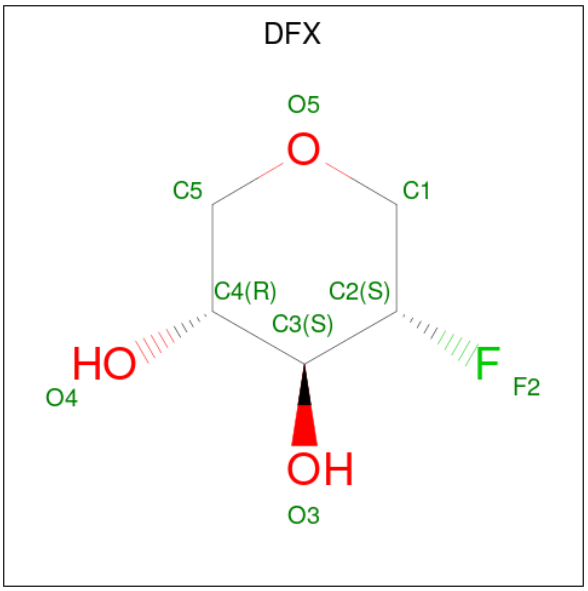
- Molecule 1 is a protein called Beta-xylosidase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	500	Total	C	N	O	S	0	0	0
			4153	2691	676	771	15			
1	B	500	Total	C	N	O	S	0	0	0
			4153	2691	676	771	15			
1	C	500	Total	C	N	O	S	0	0	0
			4153	2691	676	771	15			
1	D	500	Total	C	N	O	S	0	0	0
			4153	2691	676	771	15			

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	230	TYR	THR	SEE REMARK 999	UNP P36906
B	230	TYR	THR	SEE REMARK 999	UNP P36906
C	230	TYR	THR	SEE REMARK 999	UNP P36906
D	230	TYR	THR	SEE REMARK 999	UNP P36906

- Molecule 2 is 1,5-anhydro-2-deoxy-2-fluoro-D-xylitol (three-letter code: DFX) (formula: $C_5H_9FO_3$).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	1	Total C F O 9 5 1 3	0	0
2	B	1	Total C F O 9 5 1 3	0	0
2	C	1	Total C F O 9 5 1 3	0	0
2	D	1	Total C F O 9 5 1 3	0	0

- Molecule 3 is water.

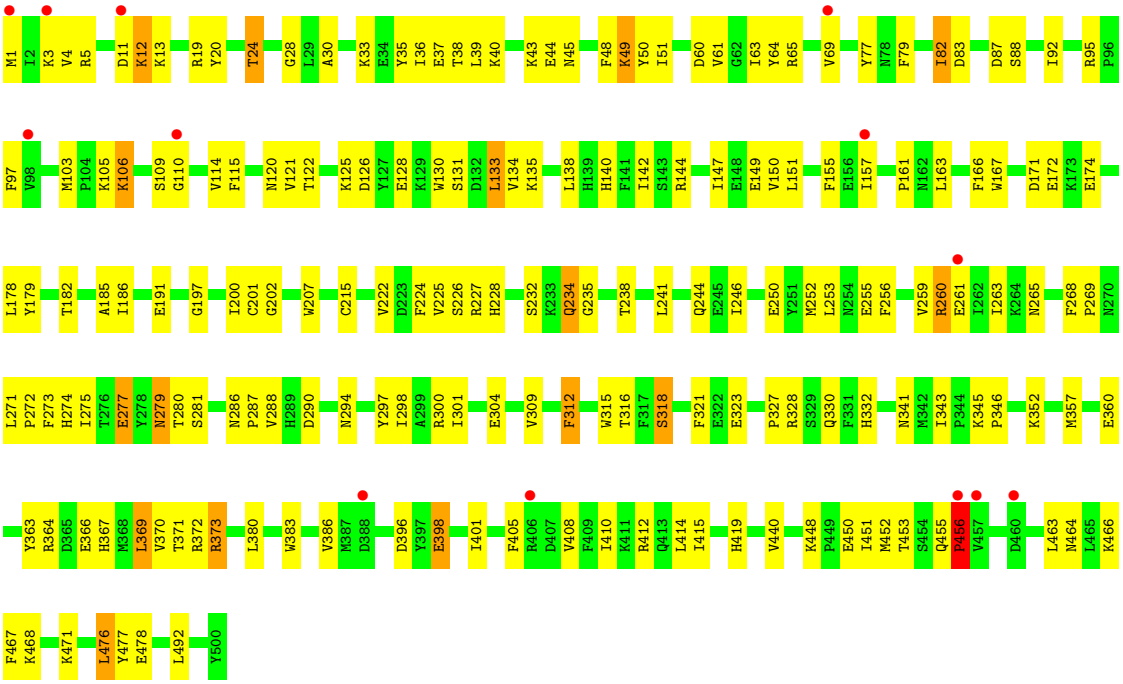
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	247	Total O 247 247	0	0
3	B	331	Total O 331 331	0	0
3	C	377	Total O 377 377	0	0
3	D	265	Total O 265 265	0	0

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.

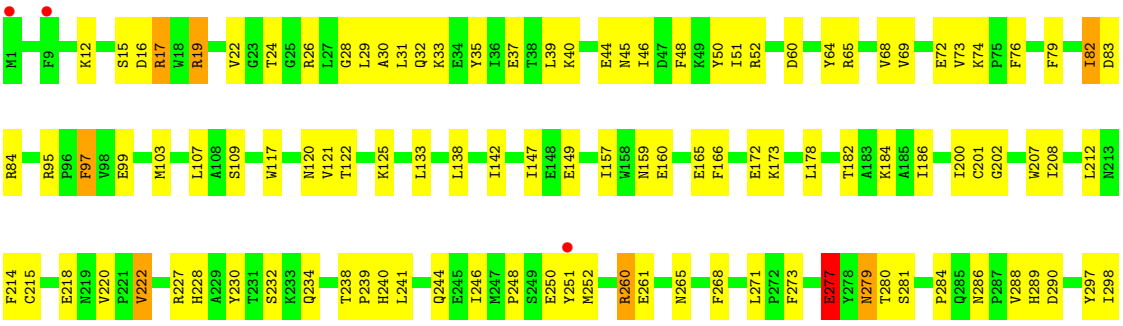
• Molecule 1: Beta-xylosidase

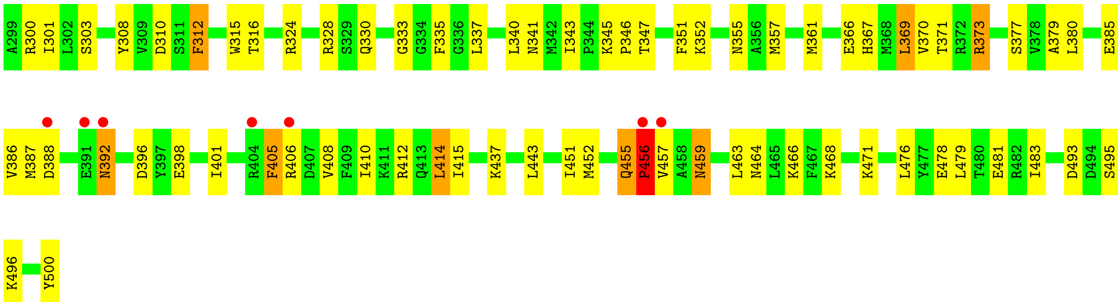
Chain A:



• Molecule 1: Beta-xylosidase

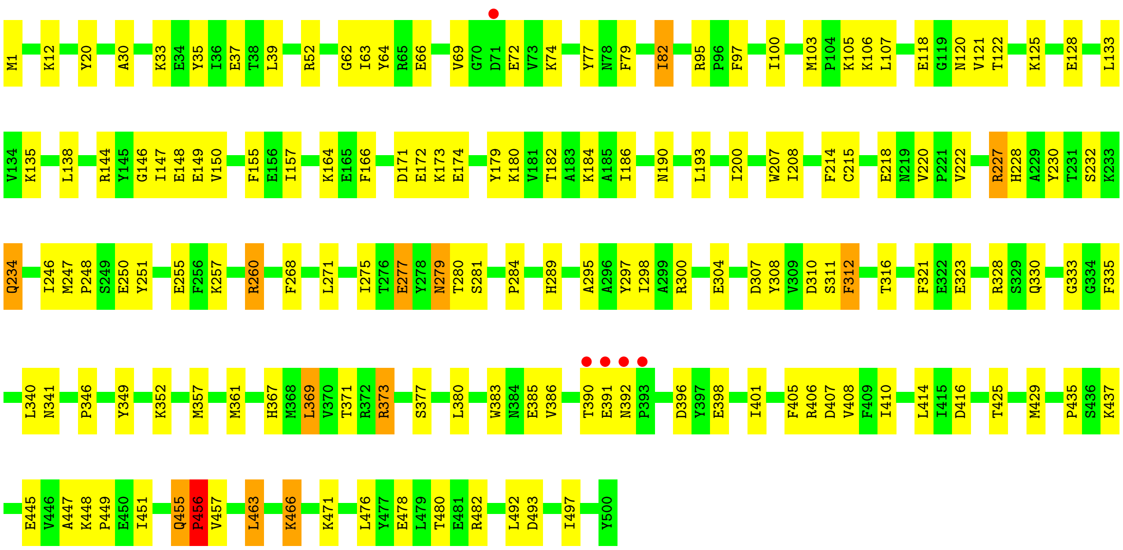
Chain B:





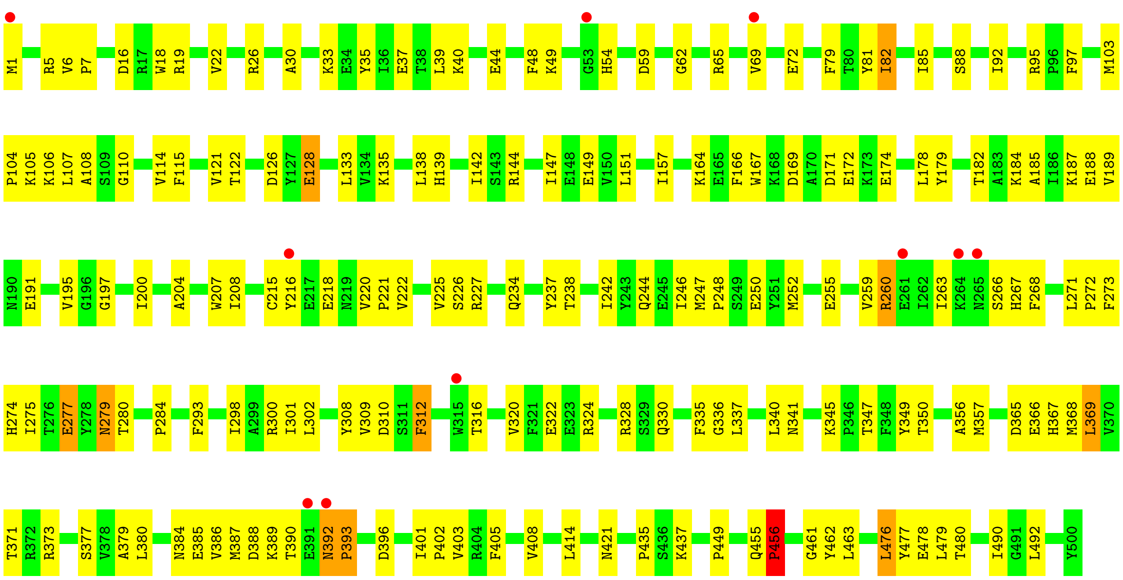
• Molecule 1: Beta-xylosidase

Chain C:



• Molecule 1: Beta-xylosidase

Chain D:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	95.33Å 152.28Å 159.68Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	32.33 – 2.10 33.95 – 2.10	Depositor EDS
% Data completeness (in resolution range)	88.8 (32.33-2.10) 92.2 (33.95-2.10)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.07	Depositor
$< I/\sigma(I) >$ [1]	4.03 (at 2.10Å)	Xtriage
Refinement program	CNS 1.1	Depositor
R, R_{free}	0.210 , 0.262 0.204 , 0.254	Depositor DCC
R_{free} test set	12504 reflections (9.98%)	wwPDB-VP
Wilson B-factor (Å^2)	20.3	Xtriage
Anisotropy	0.804	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.34 , 54.8	EDS
L-test for twinning[2]	$< \|L\| > = 0.45$, $< L^2 > = 0.27$	Xtriage
Estimated twinning fraction	0.000 for -h,l,k	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	17868	wwPDB-VP
Average B, all atoms (Å^2)	27.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 32.47 % of the origin peak, indicating pseudo-translational symmetry. The chance of finding a peak of this or larger height randomly in a structure without pseudo-translational symmetry is equal to 9.3530e-04. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

[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

5.1 Standard geometry

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

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	0.41	1/4277 (0.0%)	0.60	2/5799 (0.0%)
1	B	0.42	1/4277 (0.0%)	0.63	3/5799 (0.1%)
1	C	0.44	1/4277 (0.0%)	0.64	3/5799 (0.1%)
1	D	0.42	1/4277 (0.0%)	0.62	3/5799 (0.1%)
All	All	0.42	4/17108 (0.0%)	0.62	11/23196 (0.0%)

All (4) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	D	277	GLU	CD-OE1	14.67	1.41	1.25
1	C	277	GLU	CD-OE1	14.43	1.41	1.25
1	B	277	GLU	CD-OE1	13.82	1.40	1.25
1	A	277	GLU	CD-OE1	13.79	1.40	1.25

The worst 5 of 11 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	238	THR	N-CA-C	-6.55	93.33	111.00
1	A	97	PHE	N-CA-C	-6.09	94.55	111.00
1	D	238	THR	N-CA-C	-5.93	94.99	111.00
1	C	97	PHE	N-CA-C	-5.92	95.00	111.00
1	D	97	PHE	N-CA-C	-5.40	96.42	111.00

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	4153	0	3993	163	0
1	B	4153	0	3993	155	0
1	C	4153	0	3993	132	0
1	D	4153	0	3993	133	0
2	A	9	0	8	0	0
2	B	9	0	8	0	0
2	C	9	0	8	0	0
2	D	9	0	8	0	0
3	A	247	0	0	6	0
3	B	331	0	0	17	0
3	C	377	0	0	11	0
3	D	265	0	0	8	0
All	All	17868	0	16004	570	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 17.

The worst 5 of 570 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:D:330:GLN:HE22	1:D:341:ASN:HD22	1.15	0.94
1:C:330:GLN:HE22	1:C:341:ASN:HD22	1.11	0.92
1:B:330:GLN:HE22	1:B:341:ASN:HD22	1.18	0.91
1:C:63:ILE:HD11	1:C:82:ILE:HG12	1.54	0.88
1:A:410:ILE:HD11	1:A:456:PRO:HG3	1.55	0.88

There are no symmetry-related clashes.

5.3 Torsion angles

5.3.1 Protein backbone

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	498/500 (100%)	461 (93%)	35 (7%)	2 (0%)	34	32
1	B	498/500 (100%)	470 (94%)	26 (5%)	2 (0%)	34	32
1	C	498/500 (100%)	470 (94%)	25 (5%)	3 (1%)	25	21
1	D	498/500 (100%)	454 (91%)	39 (8%)	5 (1%)	15	11
All	All	1992/2000 (100%)	1855 (93%)	125 (6%)	12 (1%)	25	21

5 of 12 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	456	PRO
1	B	456	PRO
1	D	366	GLU
1	C	456	PRO
1	D	456	PRO

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	450/450 (100%)	428 (95%)	22 (5%)	25	23
1	B	450/450 (100%)	430 (96%)	20 (4%)	28	28
1	C	450/450 (100%)	431 (96%)	19 (4%)	30	30
1	D	450/450 (100%)	429 (95%)	21 (5%)	26	25
All	All	1800/1800 (100%)	1718 (95%)	82 (5%)	27	26

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

Mol	Chain	Res	Type
1	C	463	LEU
1	D	312	PHE
1	C	476	LEU

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	D	133	LEU
1	D	392	ASN

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 49 such sidechains are listed below:

Mol	Chain	Res	Type
1	C	265	ASN
1	C	384	ASN
1	C	267	HIS
1	C	294	ASN
1	C	455	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 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	DFX	C	3277	1	9,9,9	2.01	4 (44%)	7,12,12	1.09	0

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\#\|Z\|>2$	Counts	RMSZ	$\#\|Z\|>2$
2	DFX	D	3277	1	9,9,9	2.01	3 (33%)	7,12,12	1.03	1 (14%)
2	DFX	B	3277	1	9,9,9	1.83	3 (33%)	7,12,12	0.99	0
2	DFX	A	3277	1	9,9,9	1.93	4 (44%)	7,12,12	0.98	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	DFX	C	3277	1	-	-	0/1/1/1
2	DFX	D	3277	1	-	-	0/1/1/1
2	DFX	B	3277	1	-	-	0/1/1/1
2	DFX	A	3277	1	-	-	0/1/1/1

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

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	C	3277	DFX	C2-C3	3.68	1.56	1.51
2	D	3277	DFX	O5-C1	3.37	1.49	1.42
2	A	3277	DFX	C2-C3	3.22	1.55	1.51
2	D	3277	DFX	C2-C3	3.17	1.55	1.51
2	B	3277	DFX	O5-C1	3.03	1.48	1.42

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	D	3277	DFX	C5-O5-C1	2.04	114.66	111.52

There are no chirality outliers.

There are no torsion outliers.

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 (i)

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	500/500 (100%)	0.07	13 (2%)	56	61	14, 30, 51, 68	0
1	B	500/500 (100%)	-0.24	10 (2%)	65	69	13, 25, 45, 66	0
1	C	500/500 (100%)	-0.31	5 (1%)	82	85	12, 22, 37, 59	0
1	D	500/500 (100%)	-0.06	10 (2%)	65	69	14, 27, 47, 69	0
All	All	2000/2000 (100%)	-0.14	38 (1%)	66	71	12, 26, 46, 69	0

The worst 5 of 38 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	D	1	MET	4.5
1	A	110	GLY	4.1
1	B	1	MET	3.9
1	A	456	PRO	3.4
1	D	392	ASN	3.2

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 monosaccharides 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. 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	B-factors(Å^2)	Q<0.9
2	DFX	A	3277	9/9	0.83	0.14	32,35,37,39	0
2	DFX	B	3277	9/9	0.88	0.14	32,33,34,36	0
2	DFX	D	3277	9/9	0.91	0.17	37,37,38,39	0
2	DFX	C	3277	9/9	0.92	0.11	28,30,31,31	0

6.5 Other polymers (i)

There are no such residues in this entry.