wwPDB X-ray Structure Validation Summary Report (i)

Oct 30, 2023 – 07:52 PM JST

PDB ID	:	4YTR
Title	:	Crystal structure of D-tagatose 3-epimerase C66S from Pseudomonas cichorii in complex with 1-deoxy L-tagatose
Authors	:	Yoshida, H.; Yoshihara, A.; Ishii, T.; Izumori, K.; Kamitori, S.
Deposited on	:	2015-03-18
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:

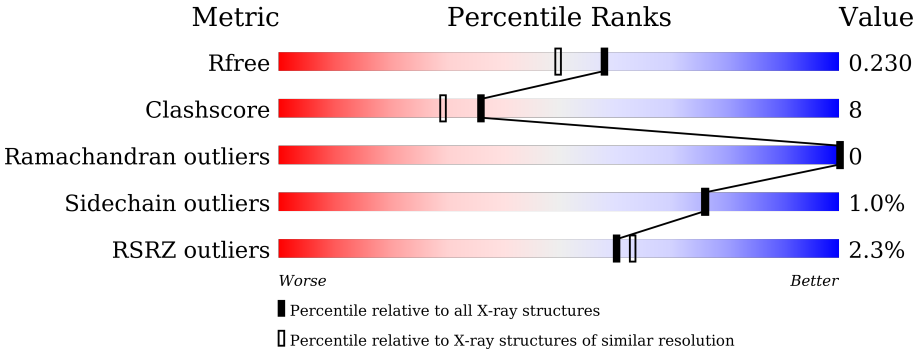
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, 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 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	300	3% 79% 18% • •
1	B	300	2% 82% 16% • •
1	C	300	3% 81% 18% • •
1	D	300	2% 81% 15% • •

2 Entry composition (i)

There are 5 unique types of molecules in this entry. The entry contains 10544 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 D-tagatose 3-epimerase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	294	2314	1465	399	432	18	0	0	0
1	B	296	2334	1477	405	434	18	0	0	0
1	C	297	2344	1483	408	435	18	0	0	0
1	D	290	2287	1451	392	426	18	0	0	0

There are 44 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	66	SER	CYS	engineered mutation	UNP O50580
A	291	GLY	-	expression tag	UNP O50580
A	292	SER	-	expression tag	UNP O50580
A	293	ARG	-	expression tag	UNP O50580
A	294	SER	-	expression tag	UNP O50580
A	295	HIS	-	expression tag	UNP O50580
A	296	HIS	-	expression tag	UNP O50580
A	297	HIS	-	expression tag	UNP O50580
A	298	HIS	-	expression tag	UNP O50580
A	299	HIS	-	expression tag	UNP O50580
A	300	HIS	-	expression tag	UNP O50580
B	66	SER	CYS	engineered mutation	UNP O50580
B	291	GLY	-	expression tag	UNP O50580
B	292	SER	-	expression tag	UNP O50580
B	293	ARG	-	expression tag	UNP O50580
B	294	SER	-	expression tag	UNP O50580
B	295	HIS	-	expression tag	UNP O50580
B	296	HIS	-	expression tag	UNP O50580
B	297	HIS	-	expression tag	UNP O50580
B	298	HIS	-	expression tag	UNP O50580
B	299	HIS	-	expression tag	UNP O50580

Continued on next page...

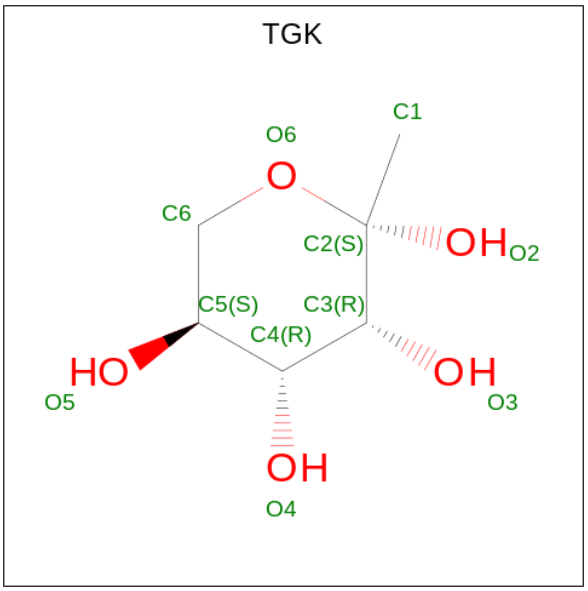
Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	300	HIS	-	expression tag	UNP O50580
C	66	SER	CYS	engineered mutation	UNP O50580
C	291	GLY	-	expression tag	UNP O50580
C	292	SER	-	expression tag	UNP O50580
C	293	ARG	-	expression tag	UNP O50580
C	294	SER	-	expression tag	UNP O50580
C	295	HIS	-	expression tag	UNP O50580
C	296	HIS	-	expression tag	UNP O50580
C	297	HIS	-	expression tag	UNP O50580
C	298	HIS	-	expression tag	UNP O50580
C	299	HIS	-	expression tag	UNP O50580
C	300	HIS	-	expression tag	UNP O50580
D	66	SER	CYS	engineered mutation	UNP O50580
D	291	GLY	-	expression tag	UNP O50580
D	292	SER	-	expression tag	UNP O50580
D	293	ARG	-	expression tag	UNP O50580
D	294	SER	-	expression tag	UNP O50580
D	295	HIS	-	expression tag	UNP O50580
D	296	HIS	-	expression tag	UNP O50580
D	297	HIS	-	expression tag	UNP O50580
D	298	HIS	-	expression tag	UNP O50580
D	299	HIS	-	expression tag	UNP O50580
D	300	HIS	-	expression tag	UNP O50580

- Molecule 2 is MANGANESE (II) ION (three-letter code: MN) (formula: Mn).

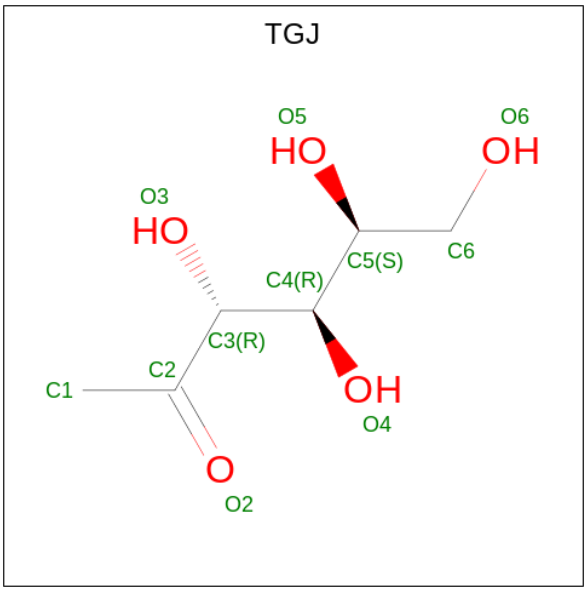
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	1	Total Mn 1 1	0	0
2	B	1	Total Mn 1 1	0	0
2	C	1	Total Mn 1 1	0	0
2	D	1	Total Mn 1 1	0	0

- Molecule 3 is 1-deoxy-beta-L-tagatopyranose (three-letter code: TGK) (formula: $C_6H_{12}O_5$).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
3	B	1	Total	C	O		0	0
			11	6	5			
3	B	1	Total	C	O		0	0
			11	6	5			
3	B	1	Total	C	O		0	0
			11	6	5			
3	C	1	Total	C	O		0	0
			11	6	5			
3	C	1	Total	C	O		0	0
			11	6	5			
3	C	1	Total	C	O		0	0
			11	6	5			
3	D	1	Total	C	O		0	0
			11	6	5			
3	D	1	Total	C	O		0	0
			11	6	5			
3	D	1	Total	C	O		0	0
			11	6	5			

- Molecule 4 is 1-deoxy-L-tagatose (three-letter code: TGJ) (formula: $C_6H_{12}O_5$).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
			Total	C	O		
4	C	1	11	6	5	0	0

- Molecule 5 is water.

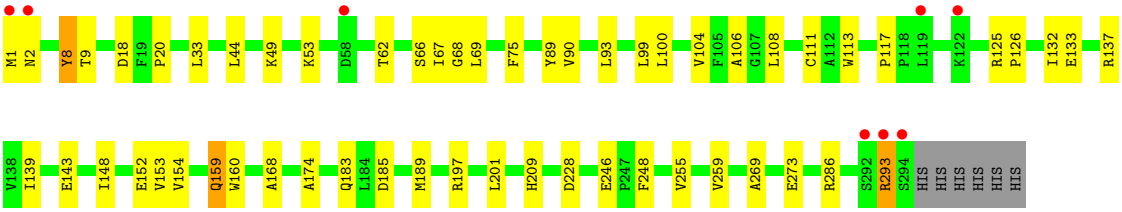
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
5	A	253	253	253	0	0
			Total	O		
5	B	309	309	309	0	0
			Total	O		
5	C	304	304	304	0	0
			Total	O		
5	D	285	285	285	0	0

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.

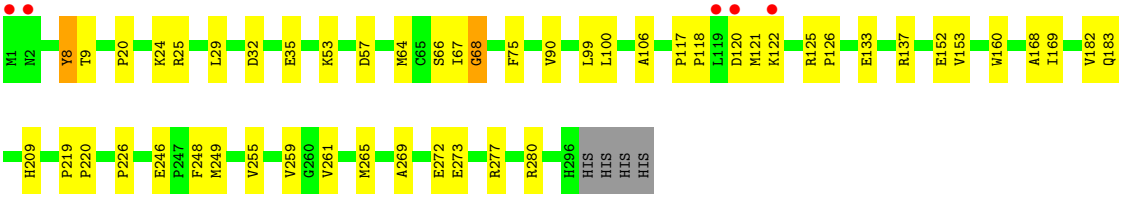
- Molecule 1: D-tagatose 3-epimerase

Chain A:



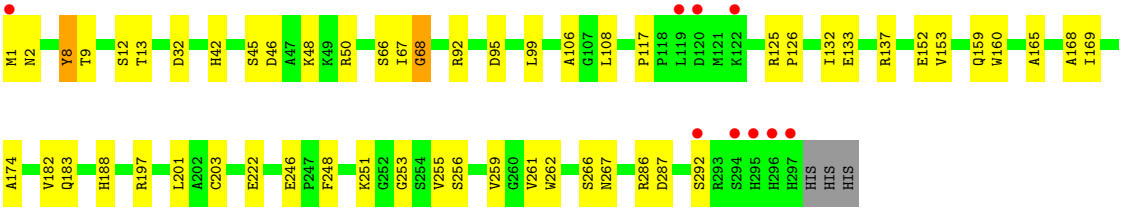
- Molecule 1: D-tagatose 3-epimerase

Chain B:



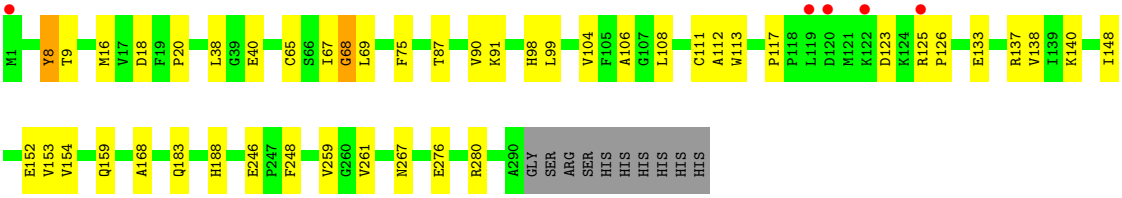
- Molecule 1: D-tagatose 3-epimerase

Chain C:



- Molecule 1: D-tagatose 3-epimerase

Chain D:


M1
Y8
T9
M16
V17
D18
F19
P20
L38
G39
E40
C65
S66
I67
G68
L69
F75
T87
V90
K91
H98
L99
V104
F105
A106
G107
L108
C111
A112
W113
P117
P118
L119
D120
M121
K122
D123
K124
R125
P126
E133
R137
V138
I139
K140
I148
E152
V153
V154
Q159
A168
Q183
H188
E246
P247
F248
V259
G260
V261
N267
E276
R280
A290
GLY
SER
ARG
SER
HIS
HIS
HIS
HIS
HIS
HIS

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	52.38Å 126.33Å 92.92Å 90.00° 98.70° 90.00°	Depositor
Resolution (Å)	21.71 – 1.90 21.71 – 1.90	Depositor EDS
% Data completeness (in resolution range)	98.0 (21.71-1.90) 98.2 (21.71-1.90)	Depositor EDS
R_{merge}	0.09	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.69 (at 1.90Å)	Xtriage
Refinement program	CNS 1.3	Depositor
R, R_{free}	0.188 , 0.235 0.184 , 0.230	Depositor DCC
R_{free} test set	9204 reflections (10.00%)	wwPDB-VP
Wilson B-factor (Å^2)	18.5	Xtriage
Anisotropy	0.189	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.38 , 59.4	EDS
L-test for twinning[2]	$< \|L\| > = 0.44$, $< L^2 > = 0.27$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	10544	wwPDB-VP
Average B, all atoms (Å^2)	20.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.29% 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: TGK, TGJ, MN

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	0.32	0/2365	0.57	0/3190
1	B	0.34	0/2387	0.59	1/3220 (0.0%)
1	C	0.33	0/2398	0.59	1/3235 (0.0%)
1	D	0.32	0/2338	0.58	1/3155 (0.0%)
All	All	0.33	0/9488	0.58	3/12800 (0.0%)

There are no bond length outliers.

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	D	68	GLY	N-CA-C	-5.58	99.14	113.10
1	B	68	GLY	N-CA-C	-5.52	99.30	113.10
1	C	68	GLY	N-CA-C	-5.06	100.45	113.10

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	2314	0	2271	36	0
1	B	2334	0	2285	34	0
1	C	2344	0	2292	44	0
1	D	2287	0	2245	36	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	A	1	0	0	0	0
2	B	1	0	0	0	0
2	C	1	0	0	0	0
2	D	1	0	0	0	0
3	B	33	0	0	0	0
3	C	33	0	0	1	0
3	D	33	0	0	0	0
4	C	11	0	0	2	0
5	A	253	0	0	2	0
5	B	309	0	0	2	0
5	C	304	0	0	5	0
5	D	285	0	0	3	0
All	All	10544	0	9093	148	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 8.

The worst 5 of 148 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:276:GLU:HG3	1:D:280:ARG:NH1	1.97	0.80
1:A:106:ALA:HB1	1:A:183:GLN:OE1	1.86	0.76
1:B:106:ALA:HB1	1:B:183:GLN:OE1	1.90	0.72
1:D:276:GLU:HG3	1:D:280:ARG:HH12	1.53	0.71
1:C:106:ALA:HB1	1:C:183:GLN:OE1	1.90	0.70

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	292/300 (97%)	285 (98%)	7 (2%)	0	100	100
1	B	294/300 (98%)	287 (98%)	7 (2%)	0	100	100
1	C	295/300 (98%)	289 (98%)	6 (2%)	0	100	100
1	D	288/300 (96%)	283 (98%)	5 (2%)	0	100	100
All	All	1169/1200 (97%)	1144 (98%)	25 (2%)	0	100	100

There are no Ramachandran outliers to report.

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	243/249 (98%)	240 (99%)	3 (1%)	71	70
1	B	245/249 (98%)	243 (99%)	2 (1%)	81	82
1	C	246/249 (99%)	244 (99%)	2 (1%)	81	82
1	D	240/249 (96%)	237 (99%)	3 (1%)	69	68
All	All	974/996 (98%)	964 (99%)	10 (1%)	76	76

5 of 10 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	D	8	TYR
1	D	159	GLN
1	D	267	ASN
1	B	8	TYR
1	B	32	ASP

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (5) such sidechains are listed below:

Mol	Chain	Res	Type
1	C	267	ASN
1	C	283	GLN
1	D	2	ASN

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	D	98	HIS
1	D	267	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)

Of 14 ligands modelled in this entry, 4 are monoatomic - leaving 10 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	$\#\|Z\| > 2$
3	TGK	C	404	-	11,11,11	0.86	1 (9%)	16,17,17	1.06	1 (6%)
3	TGK	C	405	-	11,11,11	0.80	1 (9%)	16,17,17	1.11	2 (12%)
3	TGK	D	404	-	11,11,11	0.80	0	16,17,17	1.07	1 (6%)
3	TGK	B	404	-	11,11,11	0.78	0	16,17,17	1.05	2 (12%)
3	TGK	C	403	-	11,11,11	0.83	1 (9%)	16,17,17	1.09	1 (6%)
3	TGK	B	403	-	11,11,11	0.81	1 (9%)	16,17,17	1.12	1 (6%)
3	TGK	D	402	-	11,11,11	0.81	0	16,17,17	1.16	2 (12%)
3	TGK	B	401	-	11,11,11	0.82	1 (9%)	16,17,17	1.23	1 (6%)
3	TGK	D	403	-	11,11,11	0.73	0	16,17,17	1.04	1 (6%)
4	TGJ	C	402	2	10,10,10	0.38	0	11,13,13	0.96	1 (9%)

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
3	TGK	C	404	-	-	-	0/1/1/1
3	TGK	C	405	-	-	-	0/1/1/1
3	TGK	D	404	-	-	-	0/1/1/1
3	TGK	B	404	-	-	-	0/1/1/1
3	TGK	C	403	-	-	-	0/1/1/1
3	TGK	B	403	-	-	-	0/1/1/1
3	TGK	D	402	-	-	-	0/1/1/1
3	TGK	B	401	-	-	-	0/1/1/1
3	TGK	D	403	-	-	-	0/1/1/1
4	TGJ	C	402	2	-	8/14/14/14	-

All (5) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	B	401	TGK	O6-C2	2.12	1.45	1.42
3	B	403	TGK	O6-C2	2.07	1.45	1.42
3	C	403	TGK	O6-C2	2.07	1.45	1.42
3	C	404	TGK	O6-C2	2.06	1.45	1.42
3	C	405	TGK	O6-C2	2.06	1.45	1.42

The worst 5 of 13 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
3	C	405	TGK	O6-C2-C3	2.94	110.41	105.96
3	C	403	TGK	O6-C2-C3	2.91	110.36	105.96
3	B	403	TGK	O6-C2-C3	2.89	110.34	105.96
3	C	404	TGK	O6-C2-C3	2.84	110.25	105.96
3	D	402	TGK	O6-C2-C3	2.81	110.21	105.96

There are no chirality outliers.

5 of 8 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
4	C	402	TGJ	C2-C3-C4-O4
4	C	402	TGJ	C2-C3-C4-C5
4	C	402	TGJ	O3-C3-C4-O4
4	C	402	TGJ	O3-C3-C4-C5
4	C	402	TGJ	O4-C4-C5-O5

There are no ring outliers.

2 monomers are involved in 3 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	C	404	TGK	1	0
4	C	402	TGJ	2	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	294/300 (98%)	-0.05	8 (2%)	54	57	11, 20, 34, 62	0
1	B	296/300 (98%)	-0.24	5 (1%)	70	72	11, 17, 29, 58	0
1	C	297/300 (99%)	-0.20	9 (3%)	50	53	11, 16, 28, 90	0
1	D	290/300 (96%)	-0.17	5 (1%)	70	72	11, 18, 30, 51	0
All	All	1177/1200 (98%)	-0.16	27 (2%)	60	63	11, 18, 31, 90	0

The worst 5 of 27 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	294	SER	9.4
1	A	1	MET	9.3
1	C	297	HIS	8.8
1	C	296	HIS	7.9
1	A	292	SER	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, 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
4	TGJ	C	402	11/11	0.76	0.21	40,45,49,49	0
3	TGK	B	403	11/11	0.78	0.24	44,46,47,48	0
3	TGK	C	404	11/11	0.83	0.19	27,30,33,33	0
3	TGK	D	403	11/11	0.83	0.26	45,48,48,49	0
3	TGK	C	403	11/11	0.83	0.13	35,37,37,38	0
3	TGK	B	401	11/11	0.85	0.16	36,37,38,38	0
3	TGK	B	404	11/11	0.86	0.17	34,38,40,42	0
3	TGK	C	405	11/11	0.87	0.25	41,42,43,43	0
3	TGK	D	404	11/11	0.90	0.14	33,34,35,35	0
3	TGK	D	402	11/11	0.92	0.14	26,29,31,31	0
2	MN	B	402	1/1	0.99	0.04	27,27,27,27	0
2	MN	C	401	1/1	0.99	0.03	25,25,25,25	0
2	MN	A	401	1/1	1.00	0.02	20,20,20,20	0
2	MN	D	401	1/1	1.00	0.02	21,21,21,21	0

6.5 Other polymers (i)

There are no such residues in this entry.