



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 14, 2017 – 10:18 pm GMT

PDB ID : 1MS1  
Title : Monoclinic form of Trypanosoma cruzi trans-sialidase, in complex with 3-deoxy-2,3-dehydro-N-acetylneuraminic acid (DANA)  
Authors : Buschiazzo, A.; Amaya, M.F.; Cremona, M.L.; Frasch, A.C.; Alzari, P.M.  
Deposited on : 2002-09-19  
Resolution : 1.80 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](mailto: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)  
Xtrriage (Phenix) : **NOT EXECUTED**  
EDS : **NOT EXECUTED**  
Percentile statistics : 20161228.v01 (using entries in the PDB archive December 28th 2016)  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : recalc28949

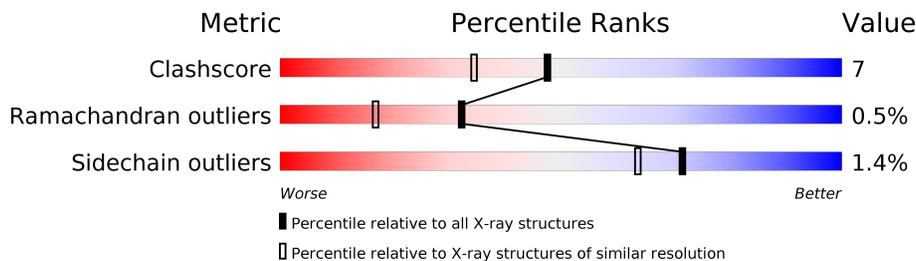
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 1.80 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))
Clashscore	112137	5742 (1.80-1.80)
Ramachandran outliers	110173	5676 (1.80-1.80)
Sidechain outliers	110143	5675 (1.80-1.80)

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  $\geq 3$ , 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.

Note EDS was not executed.

Mol	Chain	Length	Quality of chain
1	A	648	
1	B	648	

## 2 Entry composition [i](#)

There are 4 unique types of molecules in this entry. The entry contains 10915 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 trans-sialidase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	622	4871	3087	847	923	14	0	4	0
1	B	622	4878	3091	848	925	14	0	5	0

There are 50 discrepancies between the modelled and reference sequences:

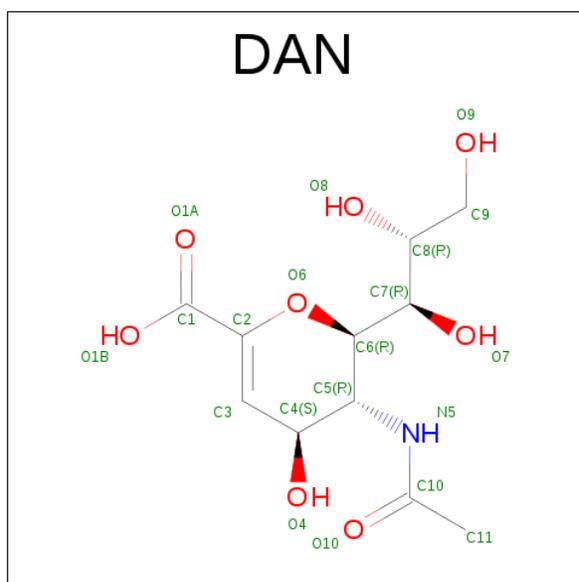
Chain	Residue	Modelled	Actual	Comment	Reference
A	-13	MET	-	EXPRESSION TAG	UNP Q26964
A	-12	GLY	-	EXPRESSION TAG	UNP Q26964
A	-11	GLY	-	EXPRESSION TAG	UNP Q26964
A	-10	SER	-	EXPRESSION TAG	UNP Q26964
A	-9	HIS	-	EXPRESSION TAG	UNP Q26964
A	-8	HIS	-	EXPRESSION TAG	UNP Q26964
A	-7	HIS	-	EXPRESSION TAG	UNP Q26964
A	-6	HIS	-	EXPRESSION TAG	UNP Q26964
A	-5	HIS	-	EXPRESSION TAG	UNP Q26964
A	-4	HIS	-	EXPRESSION TAG	UNP Q26964
A	-3	GLY	-	EXPRESSION TAG	UNP Q26964
A	-2	MET	-	EXPRESSION TAG	UNP Q26964
A	-1	ALA	-	EXPRESSION TAG	UNP Q26964
A	0	SER	-	EXPRESSION TAG	UNP Q26964
A	58	PHE	ASN	ENGINEERED	UNP Q26964
A	262	THR	SER	SEE REMARK 999	UNP Q26964
A	476	HIS	ARG	SEE REMARK 999	UNP Q26964
A	484	LEU	VAL	SEE REMARK 999	UNP Q26964
A	495	LYS	SER	ENGINEERED	UNP Q26964
A	496	GLY	VAL	ENGINEERED	UNP Q26964
A	520	LYS	GLU	ENGINEERED	UNP Q26964
A	558	VAL	GLU	SEE REMARK 999	UNP Q26964
A	593	GLY	ASP	ENGINEERED	UNP Q26964
A	597	ASP	ILE	ENGINEERED	UNP Q26964
A	599	ARG	HIS	ENGINEERED	UNP Q26964

*Continued on next page...*

*Continued from previous page...*

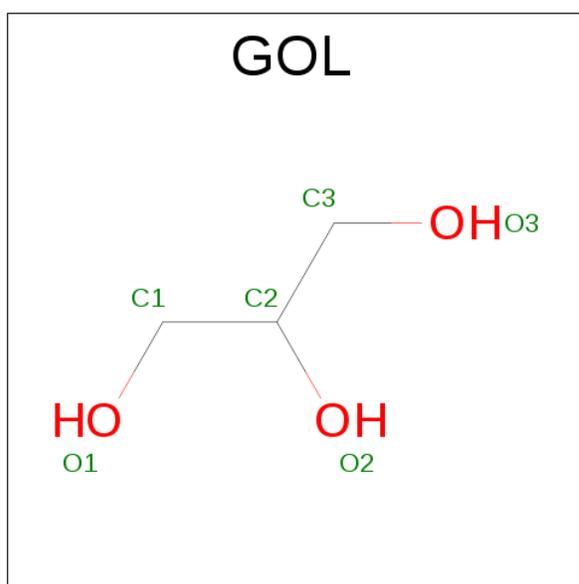
Chain	Residue	Modelled	Actual	Comment	Reference
B	-13	MET	-	EXPRESSION TAG	UNP Q26964
B	-12	GLY	-	EXPRESSION TAG	UNP Q26964
B	-11	GLY	-	EXPRESSION TAG	UNP Q26964
B	-10	SER	-	EXPRESSION TAG	UNP Q26964
B	-9	HIS	-	EXPRESSION TAG	UNP Q26964
B	-8	HIS	-	EXPRESSION TAG	UNP Q26964
B	-7	HIS	-	EXPRESSION TAG	UNP Q26964
B	-6	HIS	-	EXPRESSION TAG	UNP Q26964
B	-5	HIS	-	EXPRESSION TAG	UNP Q26964
B	-4	HIS	-	EXPRESSION TAG	UNP Q26964
B	-3	GLY	-	EXPRESSION TAG	UNP Q26964
B	-2	MET	-	EXPRESSION TAG	UNP Q26964
B	-1	ALA	-	EXPRESSION TAG	UNP Q26964
B	0	SER	-	EXPRESSION TAG	UNP Q26964
B	58	PHE	ASN	ENGINEERED	UNP Q26964
B	262	THR	SER	SEE REMARK 999	UNP Q26964
B	476	HIS	ARG	SEE REMARK 999	UNP Q26964
B	484	LEU	VAL	SEE REMARK 999	UNP Q26964
B	495	LYS	SER	ENGINEERED	UNP Q26964
B	496	GLY	VAL	ENGINEERED	UNP Q26964
B	520	LYS	GLU	ENGINEERED	UNP Q26964
B	558	VAL	GLU	SEE REMARK 999	UNP Q26964
B	593	GLY	ASP	ENGINEERED	UNP Q26964
B	597	ASP	ILE	ENGINEERED	UNP Q26964
B	599	ARG	HIS	ENGINEERED	UNP Q26964

- Molecule 2 is 2-DEOXY-2,3-DEHYDRO-N-ACETYL-NEURAMINIC ACID (three-letter code: DAN) (formula: C<sub>11</sub>H<sub>17</sub>NO<sub>8</sub>).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf	
2	A	1	Total	C	N	O	0	1
			26	14	1	11		
2	B	1	Total	C	N	O	0	1
			26	14	1	11		

- Molecule 3 is GLYCEROL (three-letter code: GOL) (formula:  $C_3H_8O_3$ ).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
3	A	1	Total	C	O	0	0
			6	3	3		
3	A	1	Total	C	O	0	0
			6	3	3		

*Continued on next page...*

*Continued from previous page...*

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

- Molecule 4 is water.

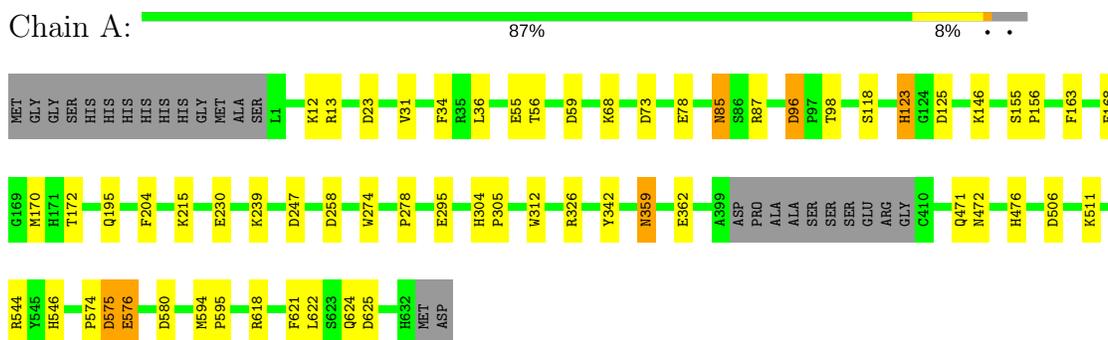
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	577	Total	O	0	0
			577	577		
4	B	519	Total	O	0	0
			519	519		

### 3 Residue-property plots [i](#)

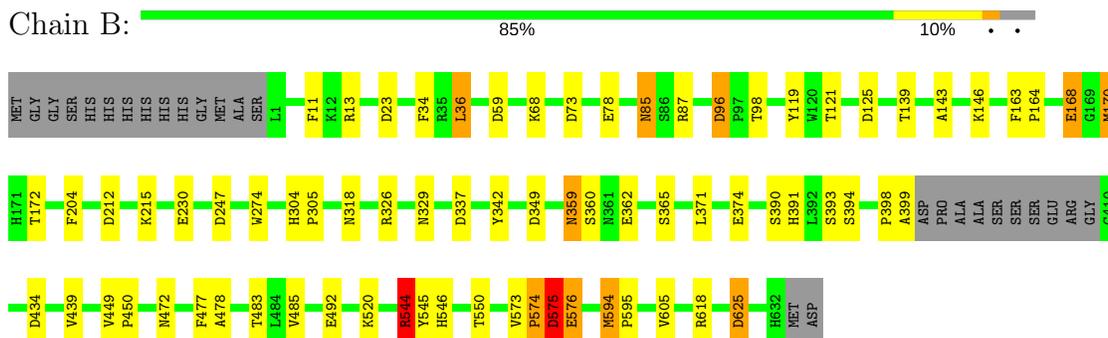
These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are 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.

Note EDS was not executed.

- Molecule 1: trans-sialidase



- Molecule 1: trans-sialidase



## 4 Data and refinement statistics

Xtrriage (Phenix) and EDS were not executed - this section is therefore incomplete.

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	63.72Å 129.79Å 88.04Å 90.00° 90.67° 90.00°	Depositor
Resolution (Å)	15.00 – 1.80	Depositor
% Data completeness (in resolution range)	99.3 (15.00-1.80)	Depositor
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	0.09	Depositor
Refinement program	REFMAC 5	Depositor
R, $R_{free}$	0.175 , 0.199	Depositor
Estimated twinning fraction	No twinning to report.	Xtrriage
Total number of atoms	10915	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	19.0	wwPDB-VP

## 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: GOL, DAN

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	0/4986	0.76	12/6773 (0.2%)
1	B	0.39	0/4993	0.77	15/6783 (0.2%)
All	All	0.40	0/9979	0.77	27/13556 (0.2%)

There are no bond length outliers.

The worst 5 of 27 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	544	ARG	CD-NE-CZ	7.30	133.83	123.60
1	B	544	ARG	CA-CB-CG	-6.89	98.23	113.40
1	B	544	ARG	CG-CD-NE	-6.89	97.34	111.80
1	B	575	ASP	CB-CG-OD2	6.78	124.41	118.30
1	B	23	ASP	CB-CG-OD2	6.31	123.98	118.30

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	4871	0	4789	61	0
1	B	4878	0	4795	75	0
2	A	26	0	16	6	0
2	B	26	0	16	9	0

*Continued on next page...*

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
3	A	12	0	16	4	0
3	B	6	0	8	0	0
4	A	577	0	0	11	1
4	B	519	0	0	12	1
All	All	10915	0	9640	135	1

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 7.

The worst 5 of 135 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:230:GLU:OE2	2:B:2701[B]:DAN:H92	1.61	1.00
1:A:511:LYS:HE2	1:A:511:LYS:HA	1.44	0.99
1:A:230:GLU:OE2	2:A:2700[A]:DAN:H92	1.72	0.87
1:A:295:GLU:HG3	4:A:3362:HOH:O	1.76	0.84
1:B:85:ASN:HD22	1:B:87:ARG:H	1.25	0.84

All (1) 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)
4:A:3316:HOH:O	4:B:3054:HOH:O[1_455]	2.14	0.06

## 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	622/648 (96%)	599 (96%)	21 (3%)	2 (0%)	44 29
1	B	623/648 (96%)	597 (96%)	22 (4%)	4 (1%)	28 13

Continued on next page...

*Continued from previous page...*

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles
All	All	1245/1296 (96%)	1196 (96%)	43 (4%)	6 (0%)	32 17

5 of 6 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	576	GLU
1	B	575	ASP
1	B	576	GLU
1	A	575	ASP
1	B	143	ALA

### 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	530/547 (97%)	525 (99%)	5 (1%)	82 78
1	B	531/547 (97%)	520 (98%)	11 (2%)	59 46
All	All	1061/1094 (97%)	1045 (98%)	16 (2%)	71 61

5 of 16 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	85	ASN
1	B	146	LYS
1	B	359	ASN
1	B	36[B]	LEU
1	B	544	ARG

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 22 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	604	ASN
1	B	85	ASN
1	B	604	ASN

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	A	611	GLN
1	A	624	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](#)

7 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	DAN	A	2700[A]	-	17,20,20	4.49	3 (17%)	18,28,28	1.18	1 (5%)
2	DAN	A	2700[B]	-	17,20,20	4.49	3 (17%)	18,28,28	1.48	3 (16%)
3	GOL	A	2800	-	5,5,5	0.32	0	5,5,5	0.40	0
3	GOL	A	2801	-	5,5,5	0.39	0	5,5,5	0.23	0
2	DAN	B	2701[A]	-	17,20,20	4.43	4 (23%)	18,28,28	1.31	1 (5%)
2	DAN	B	2701[B]	-	17,20,20	4.43	4 (23%)	18,28,28	1.29	1 (5%)
3	GOL	B	2802	-	5,5,5	0.30	0	5,5,5	0.31	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	DAN	A	2700[A]	-	-	0/14/34/34	0/1/1/1
2	DAN	A	2700[B]	-	-	0/14/34/34	0/1/1/1
3	GOL	A	2800	-	-	0/4/4/4	0/0/0/0
3	GOL	A	2801	-	-	0/4/4/4	0/0/0/0
2	DAN	B	2701[A]	-	-	0/14/34/34	0/1/1/1
2	DAN	B	2701[B]	-	-	0/14/34/34	0/1/1/1
3	GOL	B	2802	-	-	0/4/4/4	0/0/0/0

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

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	A	2700[A]	DAN	O6-C2	-6.64	1.26	1.37
2	A	2700[B]	DAN	O6-C2	-6.64	1.26	1.37
2	B	2701[A]	DAN	O6-C2	-6.62	1.26	1.37
2	B	2701[B]	DAN	O6-C2	-6.62	1.26	1.37
2	B	2701[A]	DAN	O6-C6	-2.50	1.41	1.46

The worst 5 of 6 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	B	2701[A]	DAN	C4-C3-C2	-4.61	113.80	121.60
2	B	2701[B]	DAN	C4-C3-C2	-4.61	113.80	121.60
2	A	2700[A]	DAN	C4-C3-C2	-4.37	114.20	121.60
2	A	2700[B]	DAN	C4-C3-C2	-4.37	114.20	121.60
2	A	2700[B]	DAN	C8-C7-C6	-2.03	109.09	113.04

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

5 monomers are involved in 19 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	A	2700[A]	DAN	6	0
3	A	2800	GOL	1	0
3	A	2801	GOL	3	0
2	B	2701[A]	DAN	3	0
2	B	2701[B]	DAN	6	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

### 6.1 Protein, DNA and RNA chains

EDS was not executed - this section is therefore empty.

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

EDS was not executed - this section is therefore empty.

### 6.3 Carbohydrates

EDS was not executed - this section is therefore empty.

### 6.4 Ligands

EDS was not executed - this section is therefore empty.

### 6.5 Other polymers

EDS was not executed - this section is therefore empty.