



# wwPDB X-ray Structure Validation Summary Report ⓘ

Mar 9, 2018 – 01:02 am GMT

PDB ID : 4E5Y  
Title : Structure of human FX protein, the key enzyme in the biosynthesis of GDP-L-fucose  
Authors : Zhou, H.; He, J.H.  
Deposited on : 2012-03-15  
Resolution : 2.37 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

<https://www.wwpdb.org/validation/2017/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 : 1.7.3 (157068), CSD as539be (2018)  
Xtriage (Phenix) : 1.13  
EDS : trunk30967  
Percentile statistics : 20171227.v01 (using entries in the PDB archive December 27th 2017)  
Refmac : 5.8.0158  
CCP4 : 7.0 (Gargrove)  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : trunk30967

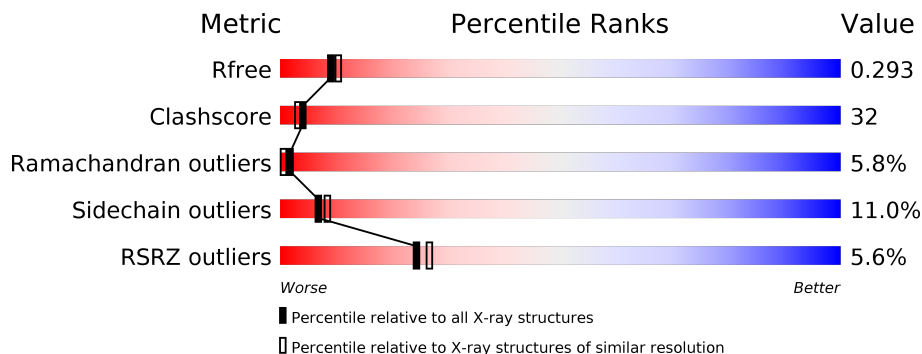
# 1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 2.37 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))
$R_{free}$	111664	4850 (2.40-2.36)
Clashscore	122126	5405 (2.40-2.36)
Ramachandran outliers	120053	5324 (2.40-2.36)
Sidechain outliers	120020	5326 (2.40-2.36)
RSRZ outliers	108989	4741 (2.40-2.36)

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.

Mol	Chain	Length	Quality of chain
1	A	321	<div> <div>4%</div> <div> <div>59%</div> <div>26%</div> <div>8%</div> <div>• •</div> </div> </div>
1	B	321	<div> <div>2%</div> <div> <div>56%</div> <div>30%</div> <div>9%</div> <div>• •</div> </div> </div>
1	C	321	<div> <div>10%</div> <div> <div>50%</div> <div>27%</div> <div>11%</div> <div>• 10%</div> </div> </div>
1	D	321	<div> <div>4%</div> <div> <div>63%</div> <div>29%</div> <div>7%</div> <div>• •</div> </div> </div>

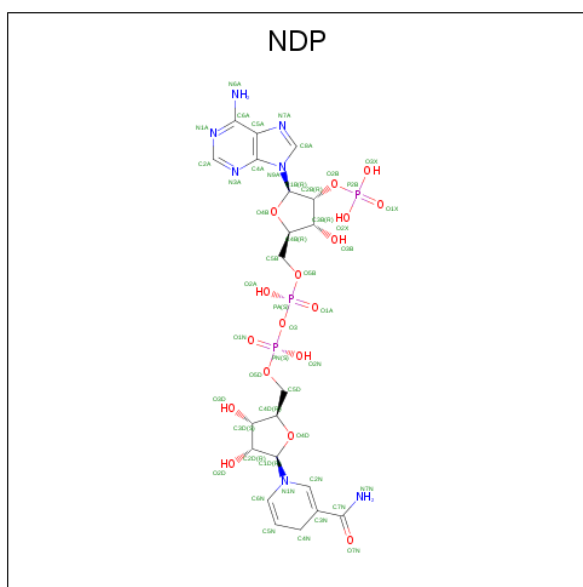


In the tables below, the ZeroOcc column contains the number of atoms modelled with zero occupancy, the AltConf column contains the number of residues with at least one atom in alternate conformation and the Trace column contains the number of residues modelled with at most 2 atoms.

- Molecule 1 is a protein called GDP-L-fucose synthase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	307	Total 2433	C 1555	N 418	O 450	S 10	0	0	0
1	B	309	Total 2452	C 1566	N 421	O 455	S 10	0	0	0
1	C	290	Total 2295	C 1462	N 392	O 431	S 10	0	0	0
1	D	318	Total 2509	C 1601	N 431	O 467	S 10	0	0	0

- Molecule 2 is NADPH DIHYDRO-NICOTINAMIDE-ADENINE-DINUCLEOTIDE PHOSPHATE (three-letter code: NDP) (formula:  $C_{21}H_{30}N_7O_{17}P_3$ ).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
2	A	1	Total 48	C 21	N 7	O 17	P 3	0	0
2	B	1	Total 48	C 21	N 7	O 17	P 3	0	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
2	C	1	Total	C	N	O	P	0	0
			48	21	7	17	3		
2	D	1	Total	C	N	O	P	0	0
			48	21	7	17	3		

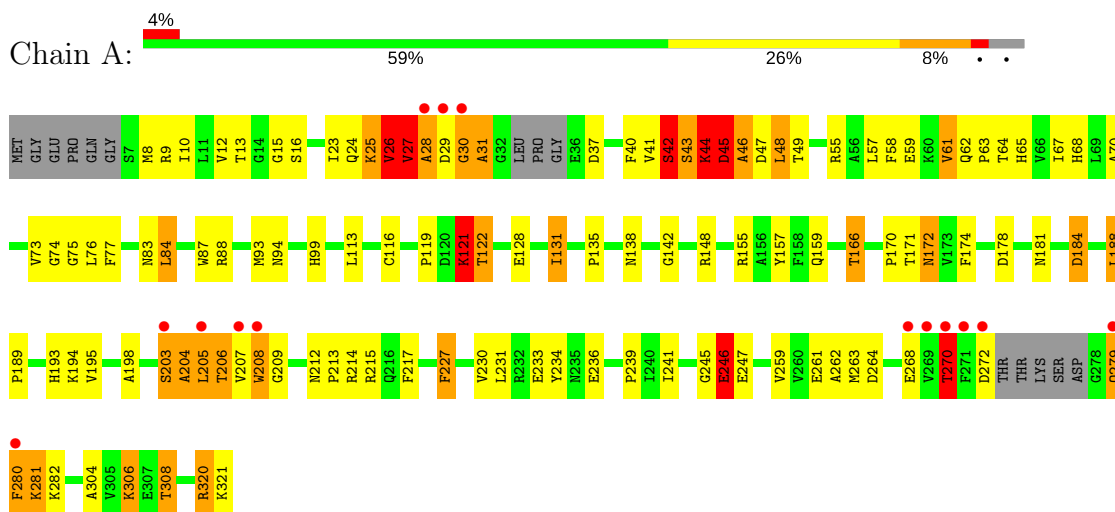
- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	3	Total	O	0	0
			3	3		
3	B	5	Total	O	0	0
			5	5		
3	C	6	Total	O	0	0
			6	6		
3	D	8	Total	O	0	0
			8	8		

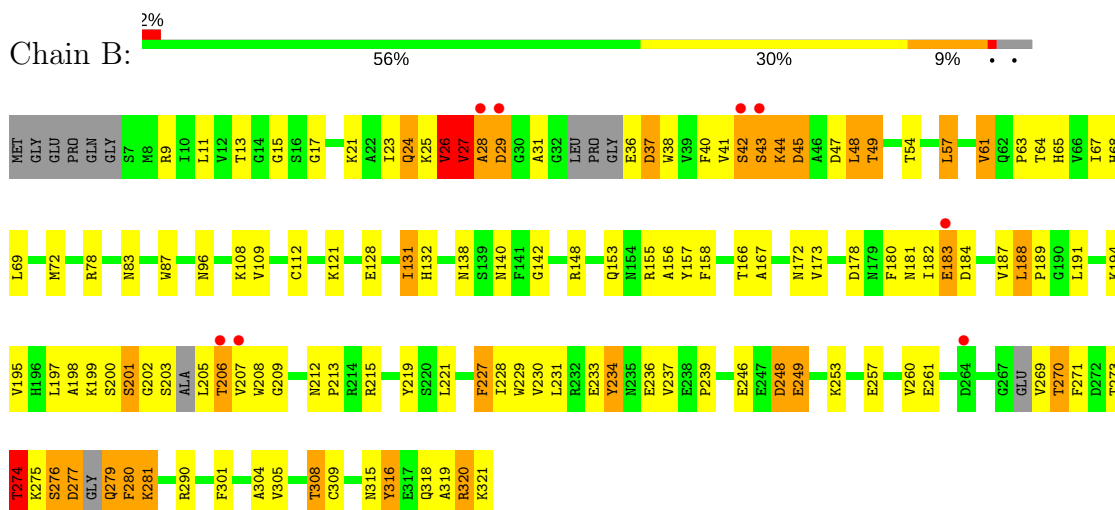
### 3 Residue-property plots

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 ( $RSRZ > 2$ ). 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: GDP-L-fucose synthase

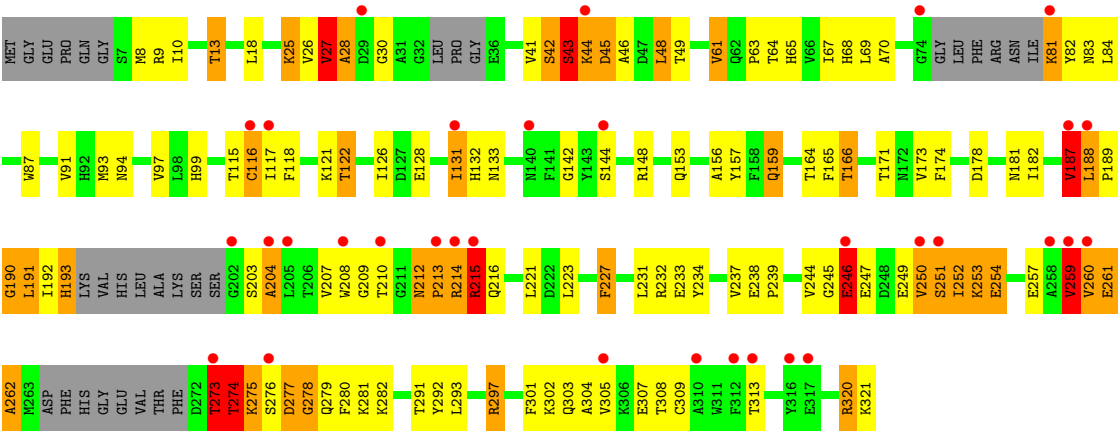


#### • Molecule 1: GDP-L-fucose synthase

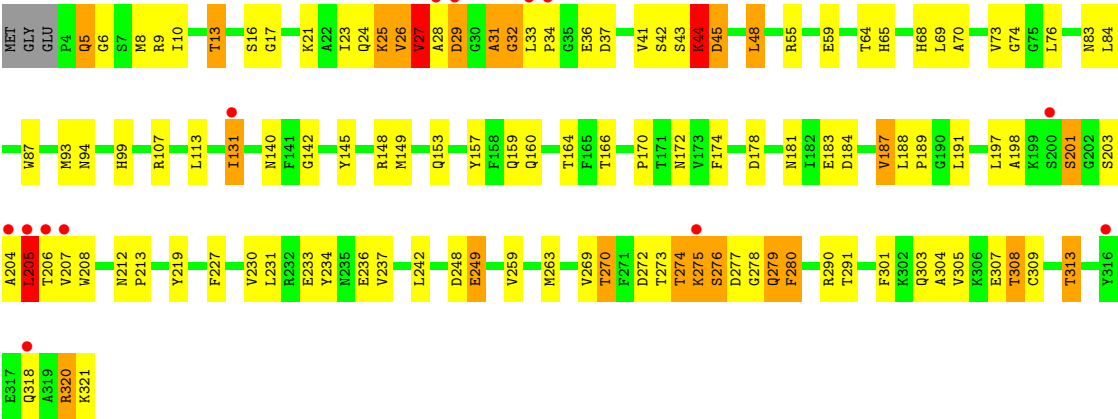


#### • Molecule 1: GDP-L-fucose synthase





• Molecule 1: GDP-L-fucose synthase



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	85.77Å 136.44Å 139.85Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	39.27 – 2.37 39.27 – 2.37	Depositor EDS
% Data completeness (in resolution range)	93.6 (39.27-2.37) 92.6 (39.27-2.37)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	3.34 (at 2.37Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.6.2_432)	Depositor
R, $R_{free}$	0.232 , 0.291 0.240 , 0.293	Depositor DCC
$R_{free}$ test set	2000 reflections (2.98%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	33.9	Xtriage
Anisotropy	0.569	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.34 , 38.0	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.33$	Xtriage
Estimated twinning fraction	0.012 for -h,l,k	Xtriage
$F_o, F_c$ correlation	0.92	EDS
Total number of atoms	9903	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	40.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.14% of the height of the origin peak. No significant pseudotranslation is detected.*

<sup>1</sup>Intensities estimated from amplitudes.

<sup>2</sup>Theoretical 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: NDP

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	0.45	0/2493	0.64	1/3379 (0.0%)
1	B	0.44	0/2510	0.58	0/3399
1	C	0.48	0/2348	0.69	2/3182 (0.1%)
1	D	0.47	0/2573	0.63	0/3491
All	All	0.46	0/9924	0.64	3/13451 (0.0%)

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	A	0	3
1	C	0	3
1	D	0	1
All	All	0	7

There are no bond length outliers.

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	C	274	THR	N-CA-C	5.50	125.86	111.00
1	A	75	GLY	N-CA-C	-5.42	99.56	113.10
1	C	273	THR	N-CA-C	5.05	124.64	111.00

There are no chirality outliers.

5 of 7 planarity outliers are listed below:



Mol	Chain	Res	Type	Group
1	A	184	ASP	Peptide
1	A	42	SER	Peptide
1	A	44	LYS	Peptide
1	C	209	GLY	Peptide
1	C	43	SER	Peptide

## 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	2433	0	2364	140	0
1	B	2452	0	2384	159	0
1	C	2295	0	2228	210	0
1	D	2509	0	2442	139	0
2	A	48	0	26	2	0
2	B	48	0	26	5	0
2	C	48	0	26	3	0
2	D	48	0	26	1	0
3	A	3	0	0	0	0
3	B	5	0	0	0	0
3	C	6	0	0	0	0
3	D	8	0	0	0	0
All	All	9903	0	9522	619	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 32.

The worst 5 of 619 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:42:SER:HA	1:B:43:SER:O	1.39	1.20
1:A:64:THR:HG23	1:A:65:HIS:HD2	1.00	1.14
1:C:214:ARG:HG2	1:C:214:ARG:HH11	1.13	1.13
1:A:64:THR:HG23	1:A:65:HIS:CD2	1.85	1.11
1:C:64:THR:HG23	1:C:65:HIS:HD2	1.05	1.09

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	301/321 (94%)	265 (88%)	17 (6%)	19 (6%)	1	0
1	B	299/321 (93%)	252 (84%)	31 (10%)	16 (5%)	2	1
1	C	280/321 (87%)	221 (79%)	37 (13%)	22 (8%)	1	0
1	D	316/321 (98%)	270 (85%)	34 (11%)	12 (4%)	3	2
All	All	1196/1284 (93%)	1008 (84%)	119 (10%)	69 (6%)	2	0

5 of 69 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	24	GLN
1	A	26	VAL
1	A	27	VAL
1	A	28	ALA
1	A	44	LYS

### 5.3.2 Protein sidechains ⓘ

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	261/273 (96%)	234 (90%)	27 (10%)	8	10
1	B	265/273 (97%)	236 (89%)	29 (11%)	7	8
1	C	248/273 (91%)	216 (87%)	32 (13%)	5	5
1	D	270/273 (99%)	243 (90%)	27 (10%)	8	11
All	All	1044/1092 (96%)	929 (89%)	115 (11%)	7	8

5 of 115 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	281	LYS
1	C	81	LYS
1	D	227	PHE
1	B	308	THR
1	C	27	VAL

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 34 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	153	GLN
1	C	65	HIS
1	D	159	GLN
1	B	181	ASN
1	A	159	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](#)

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	NDP	A	401	-	45,52,52	1.76	14 (31%)	54,80,80	1.43	7 (12%)
2	NDP	B	401	-	45,52,52	1.49	7 (15%)	54,80,80	1.54	7 (12%)
2	NDP	C	401	-	45,52,52	1.85	16 (35%)	54,80,80	1.55	10 (18%)
2	NDP	D	401	-	45,52,52	1.76	12 (26%)	54,80,80	1.45	2 (3%)

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	NDP	A	401	-	-	0/30/77/77	0/5/5/5
2	NDP	B	401	-	-	0/30/77/77	0/5/5/5
2	NDP	C	401	-	-	0/30/77/77	0/5/5/5
2	NDP	D	401	-	-	0/30/77/77	0/5/5/5

The worst 5 of 49 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	C	401	NDP	C8A-N9A	-5.31	1.30	1.36
2	D	401	NDP	C8A-N9A	-4.72	1.31	1.36
2	A	401	NDP	C8A-N9A	-4.47	1.31	1.36
2	B	401	NDP	C8A-N9A	-3.94	1.32	1.36
2	D	401	NDP	P2B-O2X	-3.24	1.41	1.54

The worst 5 of 26 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	B	401	NDP	N3A-C2A-N1A	-7.71	122.27	128.86
2	D	401	NDP	N3A-C2A-N1A	-6.41	123.38	128.86
2	C	401	NDP	N3A-C2A-N1A	-5.85	123.85	128.86
2	A	401	NDP	N3A-C2A-N1A	-5.37	124.27	128.86
2	A	401	NDP	C4A-C5A-N7A	-3.50	106.03	109.41

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

4 monomers are involved in 11 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	A	401	NDP	2	0
2	B	401	NDP	5	0
2	C	401	NDP	3	0
2	D	401	NDP	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, 95<sup>th</sup> 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 <sup>2</sup> )	Q<0.9
1	A	307/321 (95%)	0.17	14 (4%) 32 34	23, 34, 58, 79	0
1	B	309/321 (96%)	0.27	8 (2%) 56 57	24, 44, 66, 81	0
1	C	290/321 (90%)	0.60	33 (11%) 5 5	24, 39, 69, 84	0
1	D	318/321 (99%)	0.17	13 (4%) 37 40	23, 35, 59, 73	0
All	All	1224/1284 (95%)	0.30	68 (5%) 24 26	23, 38, 64, 84	0

The worst 5 of 68 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	207	VAL	10.5
1	D	204	ALA	7.5
1	A	208	TRP	6.7
1	C	208	TRP	5.9
1	C	250	VAL	4.9

### 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 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. The B-factors column lists the minimum, median, 95<sup>th</sup> 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( $\text{\AA}^2$ )	Q<0.9
2	NDP	D	401	48/48	0.95	0.10	20,20,20,20	0
2	NDP	B	401	48/48	0.95	0.12	20,20,20,20	0
2	NDP	C	401	48/48	0.96	0.09	20,20,20,20	0
2	NDP	A	401	48/48	0.96	0.09	20,20,20,20	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.