



# wwPDB X-ray Structure Validation Summary Report

Feb 28, 2014 – 05:53 PM GMT

PDB ID : 3DH4  
Title : Crystal Structure of Sodium/Sugar symporter with bound Galactose from vibrio parahaemolyticus  
Authors : Abramson, J.; Faham, S.; Cascio, D.  
Deposited on : 2008-06-16  
Resolution : 2.70 Å(reported)

This is a wwPDB 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/ValidationPDFNotes.html>

---

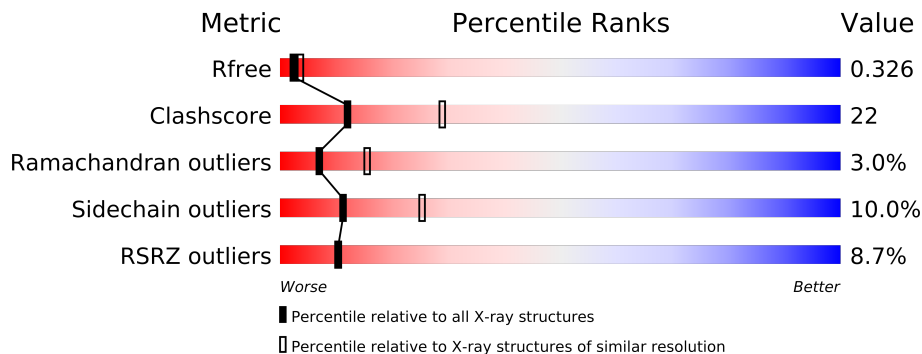
The following versions of software and data (see [references](#)) were used in the production of this report:

MolProbity : 4.02b-467  
Mogul : 1.15 2013  
Xtriage (Phenix) : dev-1323  
EDS : stable22639  
Percentile statistics : 21963  
Refmac : 5.8.0049  
CCP4 : 6.3.0 (Settle)  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et. al. (1996)  
Validation Pipeline (wwPDB-VP) : stable22683

# 1 Overall quality at a glance

The reported resolution of this entry is 2.70 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))
$R_{free}$	66092	1557 (2.70-2.70)
Clashscore	79885	1939 (2.70-2.70)
Ramachandran outliers	78287	1905 (2.70-2.70)
Sidechain outliers	78261	1905 (2.70-2.70)
RSRZ outliers	66119	1559 (2.70-2.70)

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. The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density.

Mol	Chain	Length	Quality of chain
1	A	530	
1	B	530	
1	C	530	
1	D	530	

The following table lists non-polymeric compounds that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Geometry	Electron density
4	NA	C	801	-	X

## 2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 15472 atoms, of which 0 are hydrogen and 0 are deuterium.

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 Sodium/glucose cotransporter.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	512	Total	C	N	O	S	0	0	0
			3854	2579	580	675	20			
1	B	512	Total	C	N	O	S	0	0	0
			3854	2579	580	675	20			
1	C	512	Total	C	N	O	S	0	0	0
			3854	2579	580	675	20			
1	D	512	Total	C	N	O	S	0	0	0
			3854	2579	580	675	20			

There are 64 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	544	VAL	-	EXPRESSION TAG	UNP P96169
A	545	LEU	-	EXPRESSION TAG	UNP P96169
A	546	TYR	-	EXPRESSION TAG	UNP P96169
A	547	LYS	-	EXPRESSION TAG	UNP P96169
A	548	SER	-	EXPRESSION TAG	UNP P96169
A	549	GLY	-	EXPRESSION TAG	UNP P96169
A	550	GLY	-	EXPRESSION TAG	UNP P96169
A	551	SER	-	EXPRESSION TAG	UNP P96169
A	552	PRO	-	EXPRESSION TAG	UNP P96169
A	553	GLY	-	EXPRESSION TAG	UNP P96169
A	554	HIS	-	EXPRESSION TAG	UNP P96169
A	555	HIS	-	EXPRESSION TAG	UNP P96169
A	556	HIS	-	EXPRESSION TAG	UNP P96169
A	557	HIS	-	EXPRESSION TAG	UNP P96169
A	558	HIS	-	EXPRESSION TAG	UNP P96169
A	559	HIS	-	EXPRESSION TAG	UNP P96169
B	544	VAL	-	EXPRESSION TAG	UNP P96169
B	545	LEU	-	EXPRESSION TAG	UNP P96169
B	546	TYR	-	EXPRESSION TAG	UNP P96169
B	547	LYS	-	EXPRESSION TAG	UNP P96169
B	548	SER	-	EXPRESSION TAG	UNP P96169

*Continued on next page...*

*Continued from previous page...*

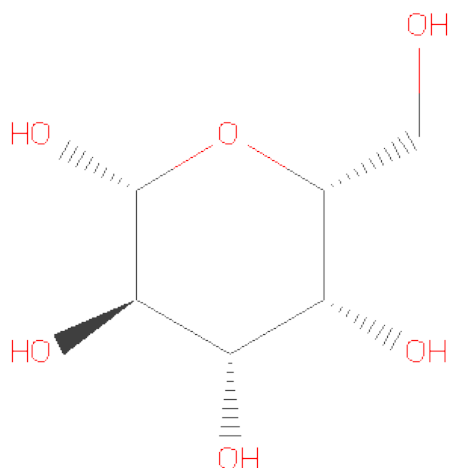
Chain	Residue	Modelled	Actual	Comment	Reference
B	549	GLY	-	EXPRESSION TAG	UNP P96169
B	550	GLY	-	EXPRESSION TAG	UNP P96169
B	551	SER	-	EXPRESSION TAG	UNP P96169
B	552	PRO	-	EXPRESSION TAG	UNP P96169
B	553	GLY	-	EXPRESSION TAG	UNP P96169
B	554	HIS	-	EXPRESSION TAG	UNP P96169
B	555	HIS	-	EXPRESSION TAG	UNP P96169
B	556	HIS	-	EXPRESSION TAG	UNP P96169
B	557	HIS	-	EXPRESSION TAG	UNP P96169
B	558	HIS	-	EXPRESSION TAG	UNP P96169
B	559	HIS	-	EXPRESSION TAG	UNP P96169
C	544	VAL	-	EXPRESSION TAG	UNP P96169
C	545	LEU	-	EXPRESSION TAG	UNP P96169
C	546	TYR	-	EXPRESSION TAG	UNP P96169
C	547	LYS	-	EXPRESSION TAG	UNP P96169
C	548	SER	-	EXPRESSION TAG	UNP P96169
C	549	GLY	-	EXPRESSION TAG	UNP P96169
C	550	GLY	-	EXPRESSION TAG	UNP P96169
C	551	SER	-	EXPRESSION TAG	UNP P96169
C	552	PRO	-	EXPRESSION TAG	UNP P96169
C	553	GLY	-	EXPRESSION TAG	UNP P96169
C	554	HIS	-	EXPRESSION TAG	UNP P96169
C	555	HIS	-	EXPRESSION TAG	UNP P96169
C	556	HIS	-	EXPRESSION TAG	UNP P96169
C	557	HIS	-	EXPRESSION TAG	UNP P96169
C	558	HIS	-	EXPRESSION TAG	UNP P96169
C	559	HIS	-	EXPRESSION TAG	UNP P96169
D	544	VAL	-	EXPRESSION TAG	UNP P96169
D	545	LEU	-	EXPRESSION TAG	UNP P96169
D	546	TYR	-	EXPRESSION TAG	UNP P96169
D	547	LYS	-	EXPRESSION TAG	UNP P96169
D	548	SER	-	EXPRESSION TAG	UNP P96169
D	549	GLY	-	EXPRESSION TAG	UNP P96169
D	550	GLY	-	EXPRESSION TAG	UNP P96169
D	551	SER	-	EXPRESSION TAG	UNP P96169
D	552	PRO	-	EXPRESSION TAG	UNP P96169
D	553	GLY	-	EXPRESSION TAG	UNP P96169
D	554	HIS	-	EXPRESSION TAG	UNP P96169
D	555	HIS	-	EXPRESSION TAG	UNP P96169
D	556	HIS	-	EXPRESSION TAG	UNP P96169
D	557	HIS	-	EXPRESSION TAG	UNP P96169
D	558	HIS	-	EXPRESSION TAG	UNP P96169

*Continued on next page...*

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
D	559	HIS	-	EXPRESSION TAG	UNP P96169

- Molecule 2 is SUGAR (BETA-D-GALACTOSE) (three-letter code: GAL) (formula:  $C_6H_{12}O_6$ ).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
2	A	1	Total	C	O	0	0
			12	6	6		
2	B	1	Total	C	O	0	0
			12	6	6		
2	C	1	Total	C	O	0	0
			12	6	6		
2	D	1	Total	C	O	0	0
			12	6	6		

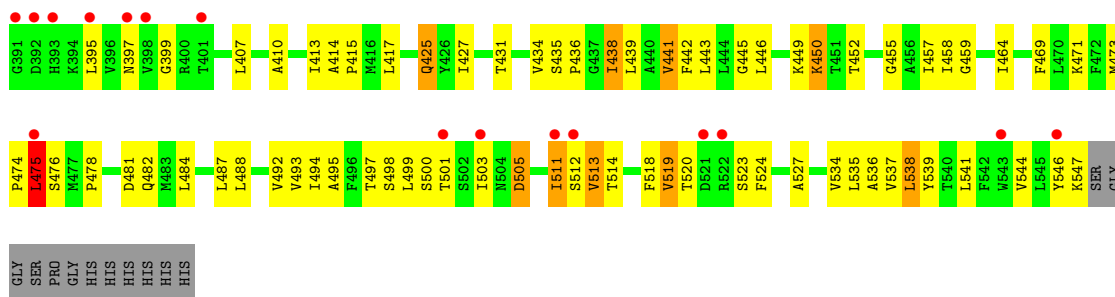
- Molecule 3 is ERBIUM (III) ION (three-letter code: ER3) (formula: Er).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	2	Total	Er	0	0
			2	2		
3	D	2	Total	Er	0	0
			2	2		

- Molecule 4 is SODIUM ION (three-letter code: NA) (formula: Na).

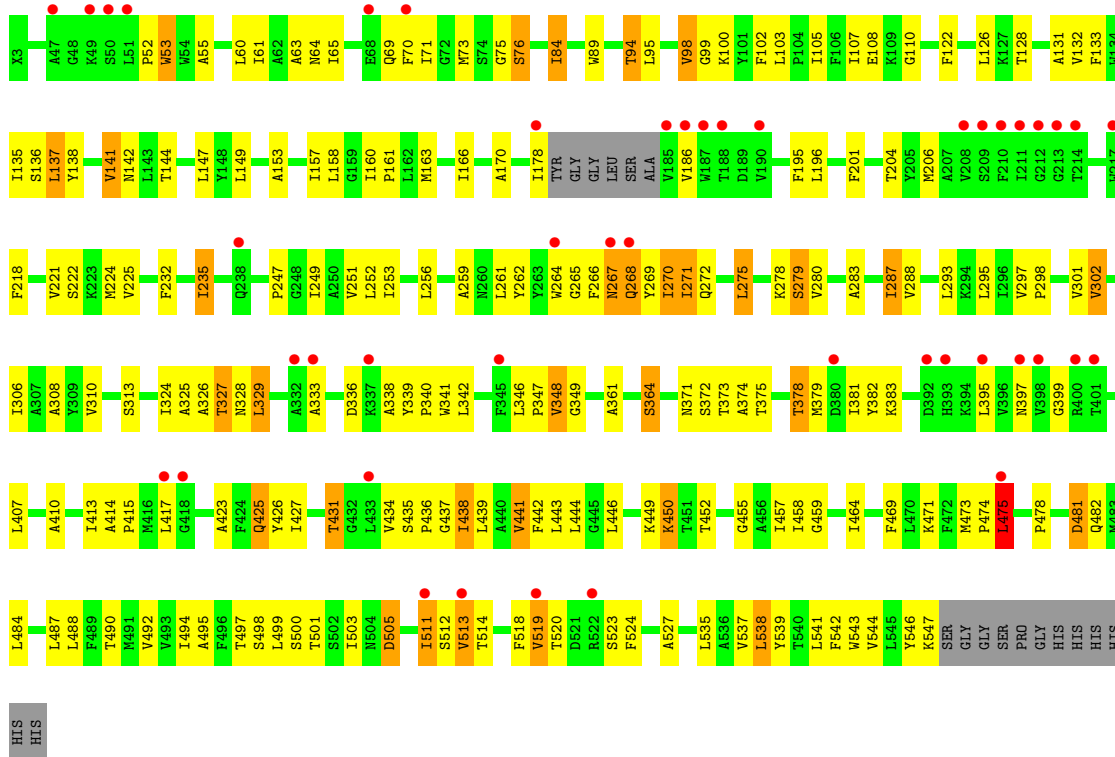
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	B	1	Total 1	Na 1	0	0
4	A	1	Total 1	Na 1	0	0
4	D	1	Total 1	Na 1	0	0
4	C	1	Total 1	Na 1	0	0





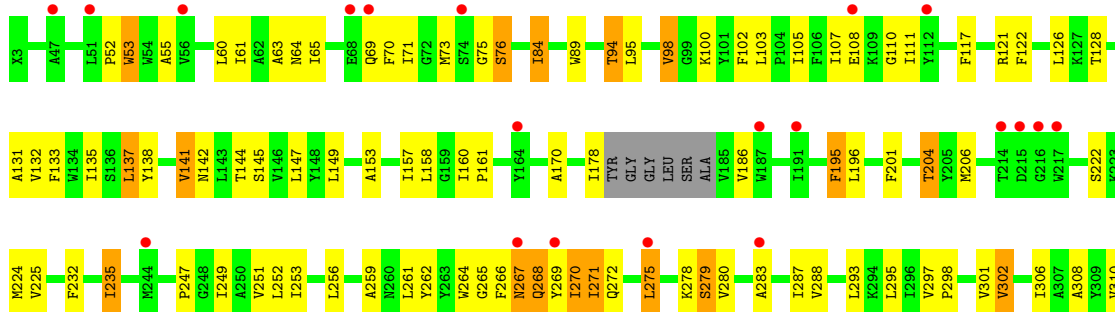
• Molecule 1: Sodium/glucose cotransporter

Chain C:



• Molecule 1: Sodium/glucose cotransporter

Chain D:





## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	83.28Å 109.21Å 127.58Å 109.70° 92.02° 102.11°	Depositor
Resolution (Å)	30.00 – 2.70 29.50 – 2.70	Depositor EDS
% Data completeness (in resolution range)	53.3 (30.00-2.70) 91.3 (29.50-2.70)	Depositor EDS
$R_{merge}$	0.11	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.45 (at 2.68Å)	Xtriage
Refinement program	REFMAC 5.4.0061	Depositor
R, $R_{free}$	0.270 , 0.287 0.310 , 0.326	Depositor DCC
$R_{free}$ test set	5166 reflections (5.02%)	DCC
Wilson B-factor (Å <sup>2</sup> )	63.0	Xtriage
Anisotropy	0.303	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.17 , 22.6	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning	$\langle  L  \rangle = 0.43$ , $\langle L^2 \rangle = 0.25$	Xtriage
Outliers	0 of 103009 reflections	Xtriage
$F_o, F_c$ correlation	0.91	EDS
Total number of atoms	15472	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	46.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.46% of the height of the origin peak. No significant pseudotranslation is detected.*

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

## 5 Model quality

### 5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: NA, GAL, ER3

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	0.48	0/3867	0.57	0/5279
1	B	0.48	0/3867	0.57	0/5279
1	C	0.48	1/3867 (0.0%)	0.57	0/5279
1	D	0.47	0/3867	0.57	0/5279
All	All	0.48	1/15468 (0.0%)	0.57	0/21116

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	1
1	B	0	1
1	C	0	1
1	D	0	1
All	All	0	4

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	C	481	ASP	CB-CG	5.38	1.63	1.51

There are no bond angle outliers.

There are no chirality outliers.

All (4) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	122	PHE	Peptide

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Group
1	B	122	PHE	Peptide
1	C	122	PHE	Peptide
1	D	122	PHE	Peptide

## 5.2 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 hydrogens added by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, and the number in parentheses is this value normalized per 1000 atoms of the molecule in the chain. The Symm-Clashes column gives symmetry related clashes, in the same way as for the Clashes column.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	3854	0	3887	180	0
1	B	3854	0	3887	164	0
1	C	3854	0	3887	164	0
1	D	3854	0	3887	188	0
2	A	12	0	12	0	0
2	B	12	0	12	0	0
2	C	12	0	12	0	0
2	D	12	0	12	0	0
3	A	2	0	0	0	0
3	D	2	0	0	0	0
4	A	1	0	0	0	0
4	B	1	0	0	0	0
4	C	1	0	0	0	0
4	D	1	0	0	0	0
All	All	15472	0	15596	668	0

Clashscore is defined as the number of clashes calculated for the entry per 1000 atoms (including hydrogens) of the entry. The overall clashscore for this entry is 22.

The worst 5 of 668 close contacts within the same asymmetric unit are listed below.

Atom-1	Atom-2	Distance(Å)	Clash(Å)
1:B:103:LEU:HD11	1:B:275:LEU:HD13	1.36	1.06
1:A:103:LEU:HD11	1:A:275:LEU:HD13	1.42	1.01
1:C:103:LEU:HD11	1:C:275:LEU:HD13	1.39	1.01
1:D:103:LEU:HD11	1:D:275:LEU:HD13	1.42	1.00
1:C:505:ASP:HB2	1:D:513:VAL:HB	1.47	0.96

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	491/530 (93%)	424 (86%)	53 (11%)	14 (3%)	7	16
1	B	491/530 (93%)	417 (85%)	59 (12%)	15 (3%)	7	14
1	C	491/530 (93%)	418 (85%)	59 (12%)	14 (3%)	7	16
1	D	491/530 (93%)	421 (86%)	54 (11%)	16 (3%)	6	13
All	All	1964/2120 (93%)	1680 (86%)	225 (12%)	59 (3%)	7	15

5 of 59 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	53	TRP
1	A	110	GLY
1	A	268	GLN
1	A	348	VAL
1	A	513	VAL

### 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	396/413 (96%)	357 (90%)	39 (10%)	12	26
1	B	396/413 (96%)	357 (90%)	39 (10%)	12	26
1	C	396/413 (96%)	356 (90%)	40 (10%)	11	24
1	D	396/413 (96%)	356 (90%)	40 (10%)	11	24
All	All	1584/1652 (96%)	1426 (90%)	158 (10%)	11	25

5 of 158 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	475	LEU
1	C	186	VAL
1	D	438	ILE
1	B	487	LEU
1	C	64	ASN

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 31 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	425	GLN
1	C	245	ASN
1	D	397	ASN
1	B	525	ASN
1	C	260	ASN

### 5.3.3 RNA ⓘ

There are no RNA chains in this entry.

## 5.4 Non-standard residues in protein, DNA, RNA chains ⓘ

There are no non-standard protein/DNA/RNA residues in this entry.

## 5.5 Carbohydrates ⓘ

There are no carbohydrates in this entry.

## 5.6 Ligand geometry ⓘ

Of 12 ligands modelled in this entry, 8 are monoatomic - leaving 4 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
2	GAL	A	701	-	12,12,12	0.56	0	17,17,17	0.58	0
2	GAL	B	701	-	12,12,12	0.56	0	17,17,17	0.55	0
2	GAL	C	701	-	12,12,12	0.56	0	17,17,17	0.60	0
2	GAL	D	701	-	12,12,12	0.55	0	17,17,17	0.59	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	GAL	A	701	-	-	0/2/22/22	0/1/1/1
2	GAL	B	701	-	-	0/2/22/22	0/1/1/1
2	GAL	C	701	-	-	0/2/22/22	0/1/1/1
2	GAL	D	701	-	-	0/2/22/22	0/1/1/1

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

## 5.7 Other polymers ⓘ

There are no such residues in this entry.

## 5.8 Polymer linkage issues

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<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	495/530 (93%)	0.28	32 (6%) 18 21	10, 47, 74, 91	0
1	B	495/530 (93%)	0.51	54 (10%) 6 6	10, 47, 75, 92	0
1	C	495/530 (93%)	0.36	44 (8%) 10 10	10, 47, 75, 92	0
1	D	495/530 (93%)	0.31	41 (8%) 11 12	10, 47, 74, 92	0
All	All	1980/2120 (93%)	0.37	171 (8%) 10 11	10, 47, 75, 92	0

The worst 5 of 171 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	513	VAL	8.5
1	B	175	VAL	8.2
1	B	392	ASP	8.1
1	C	185	VAL	8.0
1	C	186	VAL	8.0

### 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 carbohydrates 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. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. 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	RSR	LLDF	B-factors( $\text{\AA}^2$ )	Q<0.9
4	NA	C	801	1/1	0.84	9.33	43,43,43,43	0
4	NA	B	801	1/1	0.27	0.38	41,41,41,41	0
4	NA	D	801	1/1	0.20	0.17	42,42,42,42	0
2	GAL	B	701	12/12	0.15	-0.87	24,25,27,27	0
2	GAL	C	701	12/12	0.16	-0.90	24,25,26,27	0
2	GAL	A	701	12/12	0.15	-0.95	24,25,26,27	0
2	GAL	D	701	12/12	0.16	-1.22	24,25,26,27	0
3	ER3	A	704	1/1	0.06	-1.44	93,93,93,93	0
4	NA	A	801	1/1	0.14	-1.58	42,42,42,42	0
3	ER3	D	705	1/1	0.10	-2.12	99,99,99,99	0
3	ER3	A	702	1/1	0.03	-59.67	54,54,54,54	0
3	ER3	D	703	1/1	0.04	-60.33	53,53,53,53	0

## 6.5 Other polymers ⓘ

There are no such residues in this entry.