wwPDB X-ray Structure Validation Summary Report (i)

Aug 8, 2020 – 10:57 AM BST

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 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 following versions of software and data (see references (1)) were used in the production of this report:

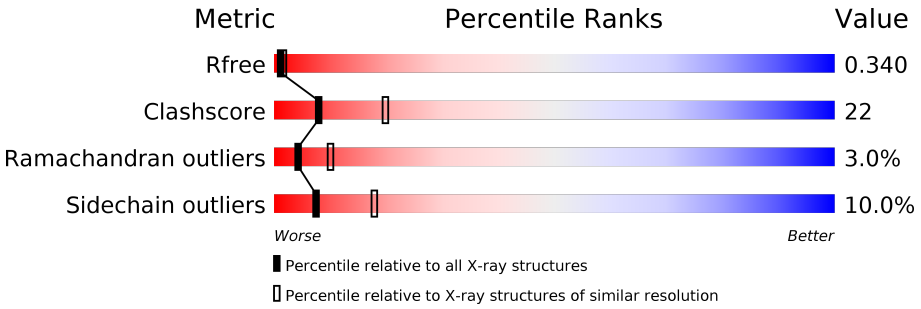
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.13.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.13.1

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 2.70 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}	130704	2808 (2.70-2.70)
Clashscore	141614	3122 (2.70-2.70)
Ramachandran outliers	138981	3069 (2.70-2.70)
Sidechain outliers	138945	3069 (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 >=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%

Mol	Chain	Length	Quality of chain
1	A	530	57% 34% 5% •
1	B	530	58% 32% 6% •
1	C	530	58% 32% 6% •
1	D	530	58% 33% 5% •

2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 15472 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 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...

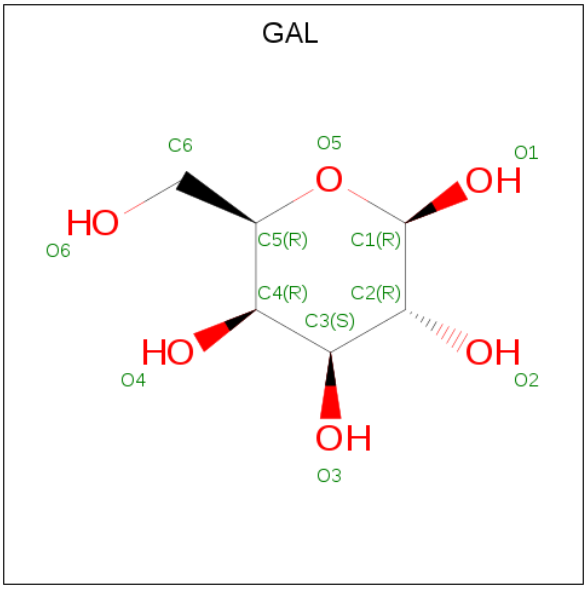
Chain	Residue	Modelled	Actual	Comment	Reference
B	549	GLY	-	expression tag	UNP P96169
B	550	GLY	-	expression tag	UNP P96169
B	551	SER	-	expression tag	UNP P96169
B	552	PRO	-	expression tag	UNP P96169
B	553	GLY	-	expression tag	UNP P96169
B	554	HIS	-	expression tag	UNP P96169
B	555	HIS	-	expression tag	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 beta-D-galactopyranose (three-letter code: GAL) (formula: $C_6H_{12}O_6$).



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

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

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

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

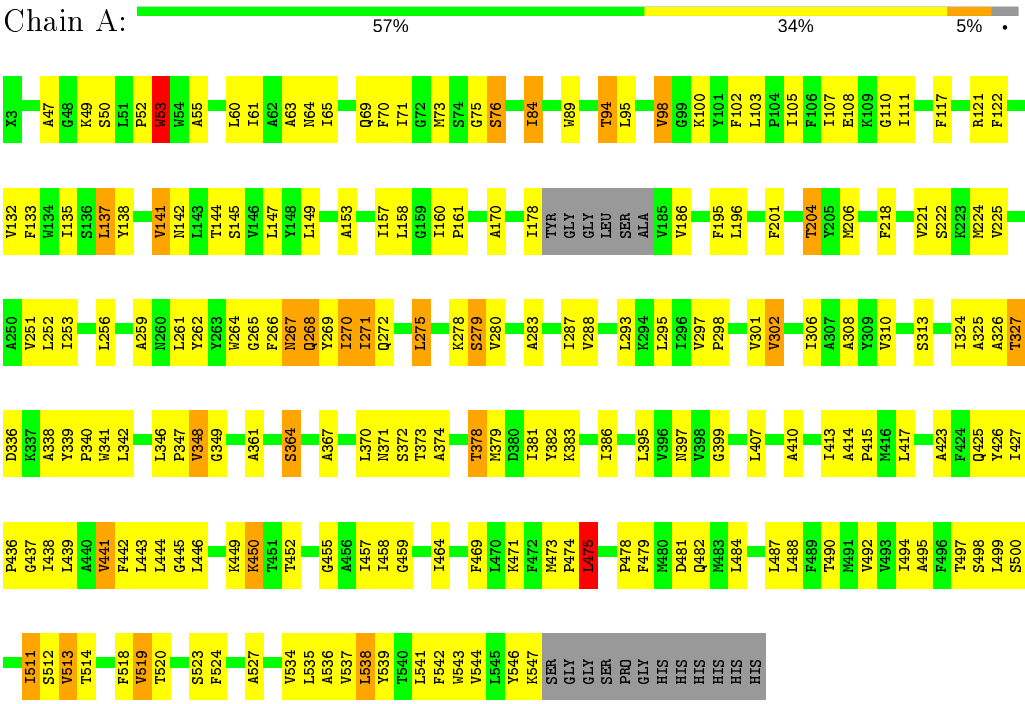
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	B	1	Total Na 1 1	0	0
4	A	1	Total Na 1 1	0	0
4	D	1	Total Na 1 1	0	0
4	C	1	Total Na 1 1	0	0

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. 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. 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.

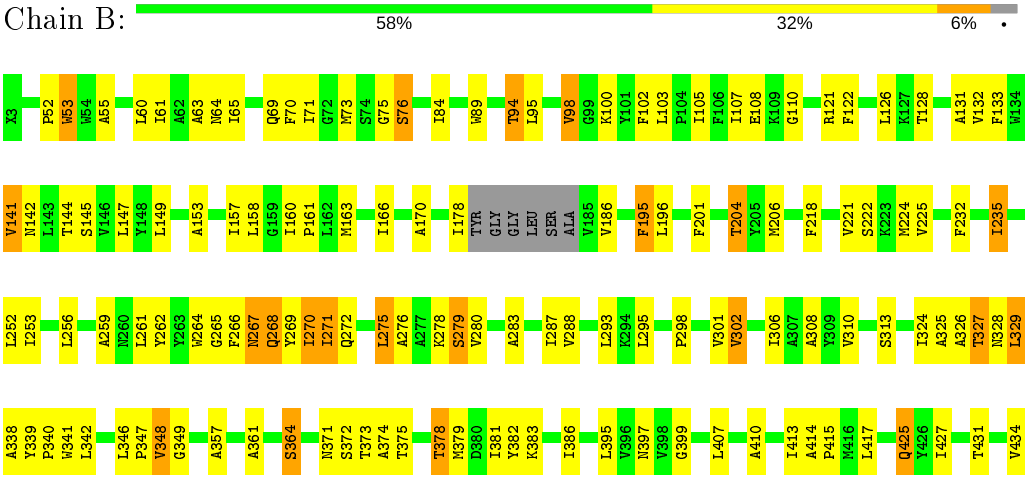
- Molecule 1: Sodium/glucose cotransporter

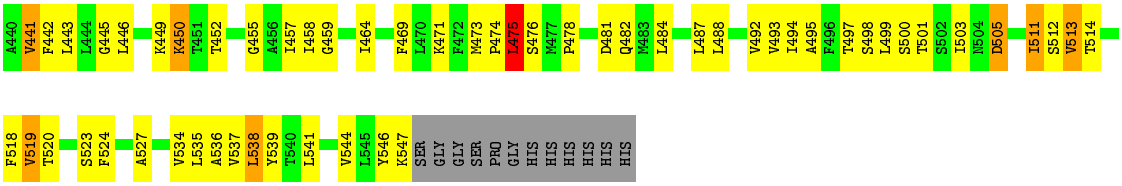
Chain A:



- Molecule 1: Sodium/glucose cotransporter

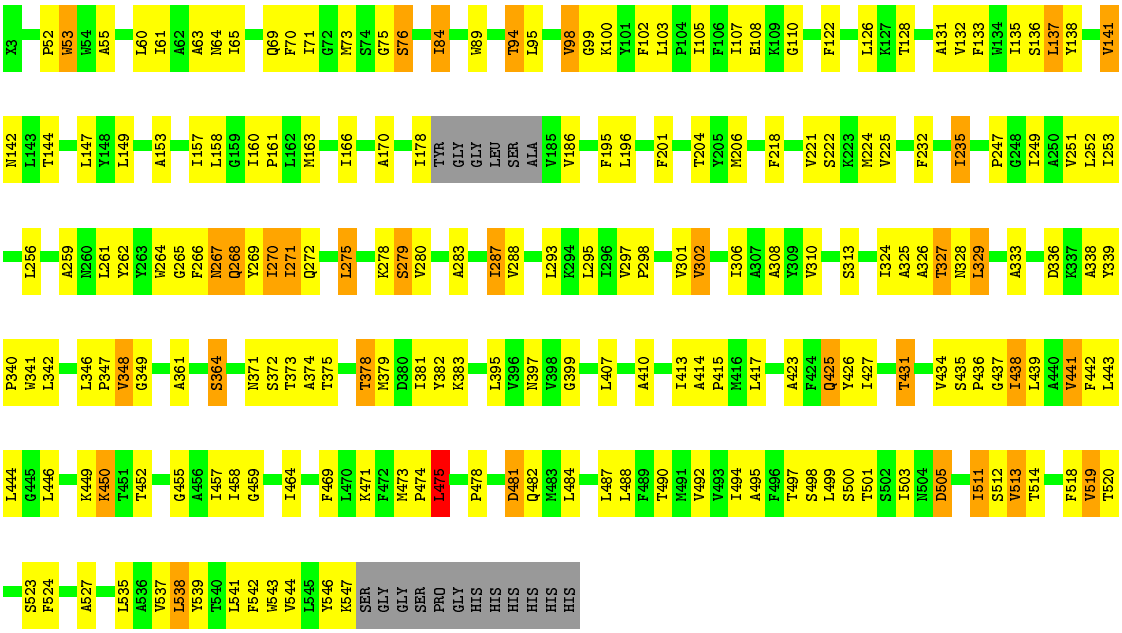
Chain B:





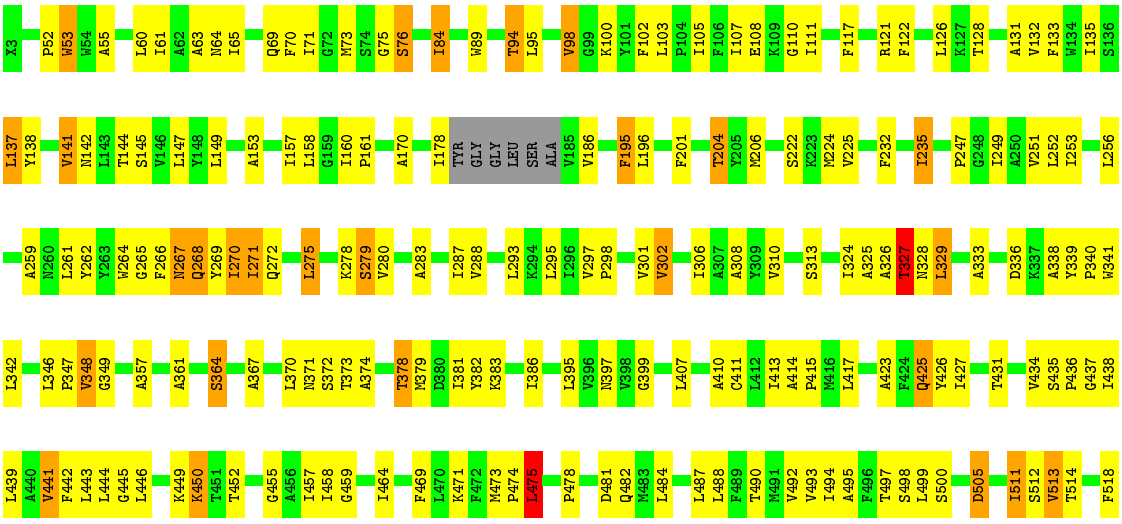
• Molecule 1: Sodium/glucose cotransporter

Chain C:



• Molecule 1: Sodium/glucose cotransporter

Chain D:




V519
T520
S523
F524
N525
V534
L535
A536
V537
L538
Y539
T540
L541
F542
W543
V544
L545
Y546
K547
SER
GLY
GLY
SER
PRO
GLY
HIS
HIS
HIS
HIS
HIS
HIS

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, α, β, γ	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
$< I/\sigma(I) >$ [1]	1.45 (at 2.68Å)	Xtriage
Refinement program	REFMAC	Depositor
R, R_{free}	0.270 , 0.287 0.330 , 0.340	Depositor DCC
R_{free} test set	5170 reflections (5.02%)	wwPDB-VP
Wilson B-factor (Å^2)	63.0	Xtriage
Anisotropy	0.303	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.17 , 32.5	EDS
L-test for twinning[2]	$< \|L\| > = 0.43$, $< L^2 > = 0.25$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.91	EDS
Total number of atoms	15472	wwPDB-VP
Average B, all atoms (Å^2)	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.*

[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: 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 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	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

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 22.

The worst 5 of 668 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: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 (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	491/530 (93%)	424 (86%)	53 (11%)	14 (3%)	4 10
1	B	491/530 (93%)	417 (85%)	59 (12%)	15 (3%)	4 9
1	C	491/530 (93%)	418 (85%)	59 (12%)	14 (3%)	4 10
1	D	491/530 (93%)	421 (86%)	54 (11%)	16 (3%)	4 8
All	All	1964/2120 (93%)	1680 (86%)	225 (12%)	59 (3%)	4 10

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 (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	396/413 (96%)	357 (90%)	39 (10%)	8 18
1	B	396/413 (96%)	357 (90%)	39 (10%)	8 18
1	C	396/413 (96%)	356 (90%)	40 (10%)	7 17

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
1	D	396/413 (96%)	356 (90%)	40 (10%)	7 17
All	All	1584/1652 (96%)	1426 (90%)	158 (10%)	7 18

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 (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 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	D	701	-	12,12,12	0.55	0	17,17,17	0.61	0
2	GAL	C	701	-	12,12,12	0.55	0	17,17,17	0.63	0
2	GAL	B	701	-	12,12,12	0.55	0	17,17,17	0.58	0
2	GAL	A	701	-	12,12,12	0.55	0	17,17,17	0.60	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	D	701	-	-	0/2/22/22	0/1/1/1
2	GAL	C	701	-	-	0/2/22/22	0/1/1/1
2	GAL	B	701	-	-	0/2/22/22	0/1/1/1
2	GAL	A	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.

No monomer is involved in short contacts.

5.7 Other polymers (i)

There are no such residues in this entry.

5.8 Polymer linkage issues (i)

The following chains have linkage breaks:

Mol	Chain	Number of breaks
1	B	1

Continued on next page...

Continued from previous page...

Mol	Chain	Number of breaks
1	A	1
1	D	1
1	C	1

All chain breaks are listed below:

Model	Chain	Residue-1	Atom-1	Residue-2	Atom-2	Distance (Å)
1	B	19:UNK	C	47:ALA	N	17.80
1	C	19:UNK	C	47:ALA	N	17.80
1	A	19:UNK	C	47:ALA	N	17.79
1	D	19:UNK	C	47:ALA	N	17.78

6 Fit of model and data

6.1 Protein, DNA and RNA chains

Unable to reproduce the depositors R factor - this section is therefore empty.

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

Unable to reproduce the depositors R factor - this section is therefore empty.

6.3 Carbohydrates

Unable to reproduce the depositors R factor - this section is therefore empty.

6.4 Ligands

Unable to reproduce the depositors R factor - this section is therefore empty.

6.5 Other polymers

Unable to reproduce the depositors R factor - this section is therefore empty.