wwPDB X-ray Structure Validation Summary Report (i)

Feb 1, 2016 – 04:28 PM GMT

PDB ID	:	4F35
Title	:	Crystal Structure of a bacterial dicarboxylate/sodium symporter
Authors	:	Mancusso, R.L.; Gregorio, G.G.; Liu, Q.; Wang, D.N.
Deposited on	:	2012-05-08
Resolution	:	3.20 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
http://wwpdb.org/validation/2016/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:

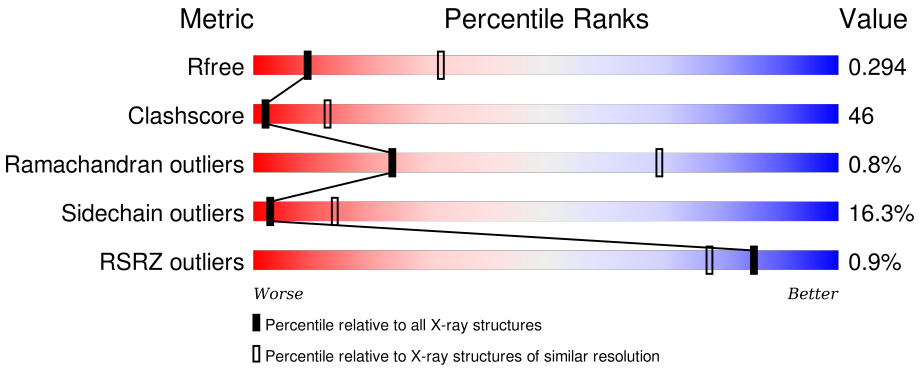
MolProbity	:	4.02b-467
Mogul	:	1.7 (RC4), CSD as536be (2015)
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20026688
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	trunk26865

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 3.20 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}	91344	1124 (3.24-3.16)
Clashscore	102246	1024 (3.22-3.18)
Ramachandran outliers	100387	1004 (3.22-3.18)
Sidechain outliers	100360	1003 (3.22-3.18)
RSRZ outliers	91569	1129 (3.24-3.16)

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. 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	449	% 42% 43% 11% ·
1	B	449	% 39% 44% 9% 8%
1	C	449	% 37% 49% 9% ·
1	D	449	% 39% 49% 7% · ·

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit crite-

ria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
2	NA	A	502	-	-	-	X
2	NA	B	502	-	-	-	X
2	NA	D	501	-	-	-	X
3	BNG	B	503	-	-	-	X
3	BNG	C	503	-	-	-	X
3	BNG	D	502	-	-	-	X
4	CIT	A	501	-	-	-	X
4	CIT	B	501	-	-	X	X

2 Entry composition (i)

There are 5 unique types of molecules in this entry. The entry contains 12390 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.

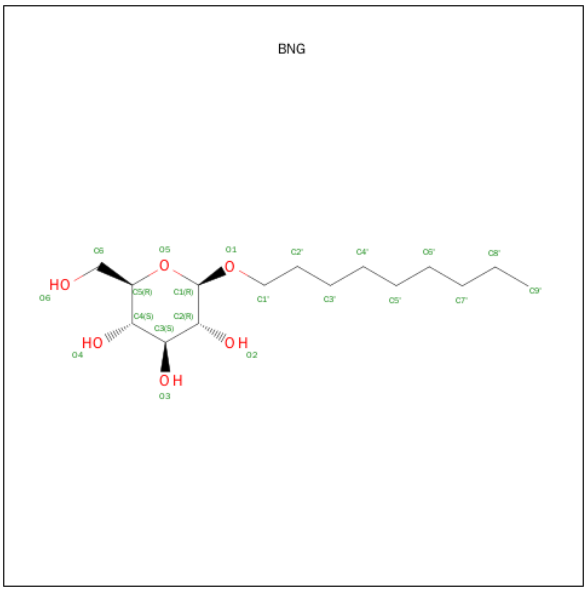
- Molecule 1 is a protein called Transporter, NadC family.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	D	430	Total	C	N	O	S	Se	3	0	0
			3076	2046	479	526	3	22			
1	B	414	Total	C	N	O	S	Se	0	0	0
			2980	1991	457	508	3	21			
1	A	434	Total	C	N	O	S	Se	0	0	0
			3110	2068	481	536	3	22			
1	C	430	Total	C	N	O	S	Se	0	0	0
			3113	2077	482	529	3	22			

- Molecule 2 is SODIUM ION (three-letter code: NA) (formula: Na).

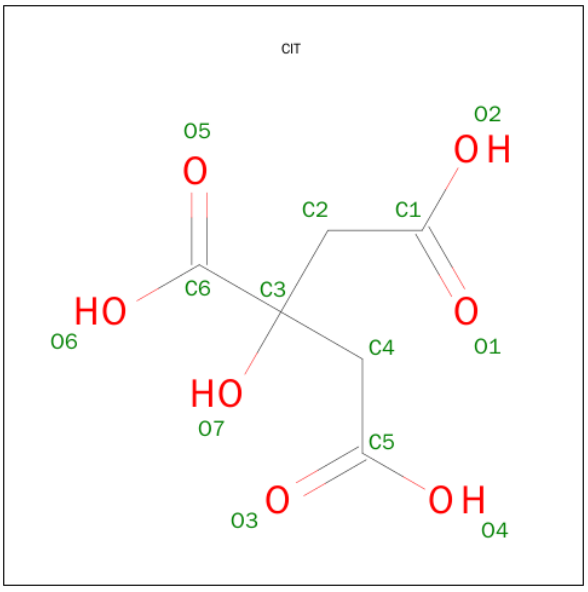
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	B	1	Total	Na	0	0
			1	1		
2	A	1	Total	Na	0	0
			1	1		
2	D	1	Total	Na	0	0
			1	1		
2	C	1	Total	Na	0	0
			1	1		

- Molecule 3 is SUGAR (B-NONYLGLUCOSIDE) (three-letter code: BNG) (formula: $C_{15}H_{30}O_6$).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	D	1	Total C O 21 15 6	0	0
3	B	1	Total C O 21 15 6	0	0
3	C	1	Total C O 21 15 6	0	0

- Molecule 4 is CITRIC ACID (three-letter code: CIT) (formula: $C_6H_8O_7$).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	B	1	Total C O 13 6 7	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	A	1	Total C O 13 6 7	0	0
4	C	1	Total C O 13 6 7	0	0

- Molecule 5 is water.

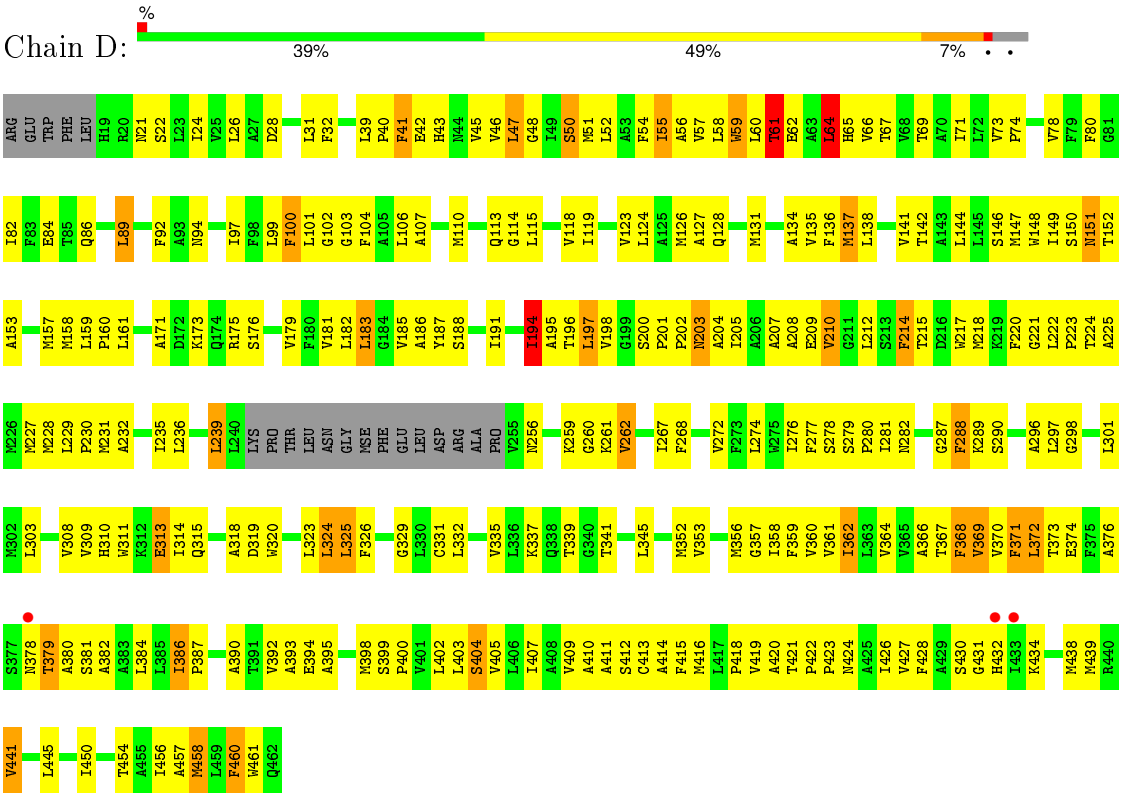
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	D	3	Total O 3 3	0	0
5	B	1	Total O 1 1	0	0
5	C	1	Total O 1 1	0	0

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 errors 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.

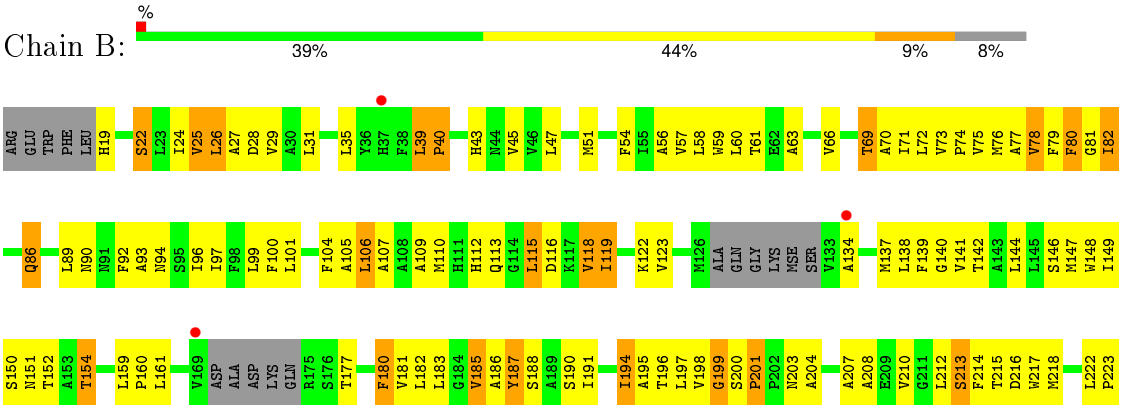
- Molecule 1: Transporter, NadC family

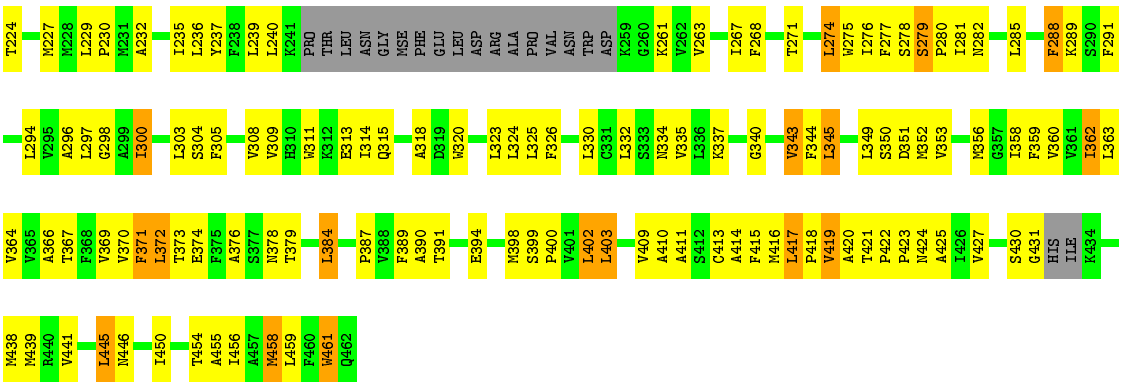
Chain D:



- Molecule 1: Transporter, NadC family

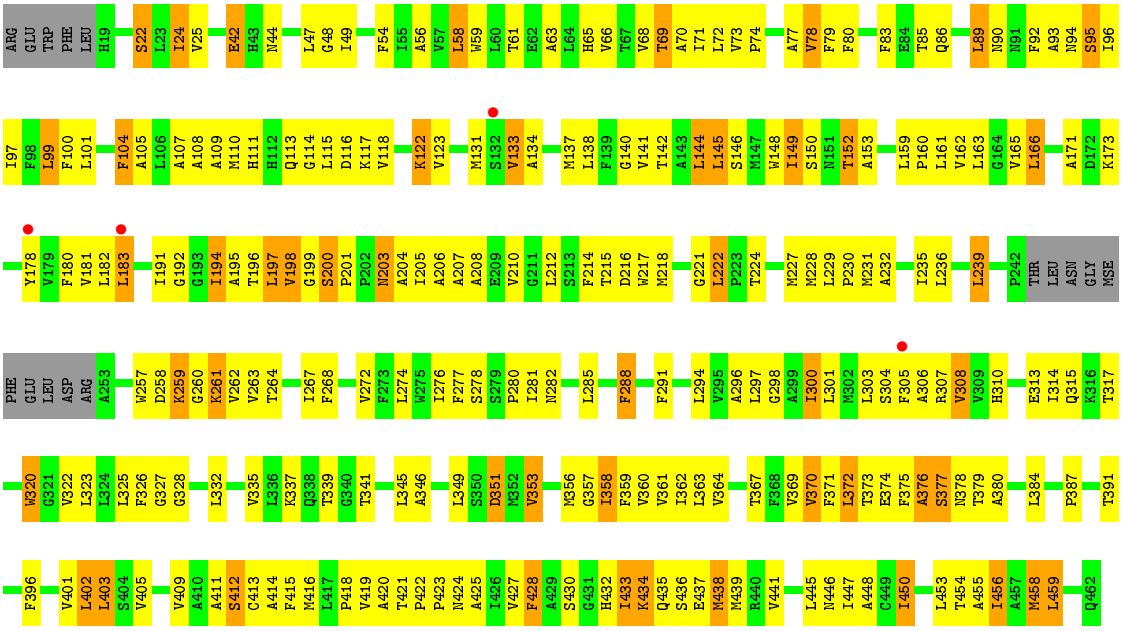
Chain B:





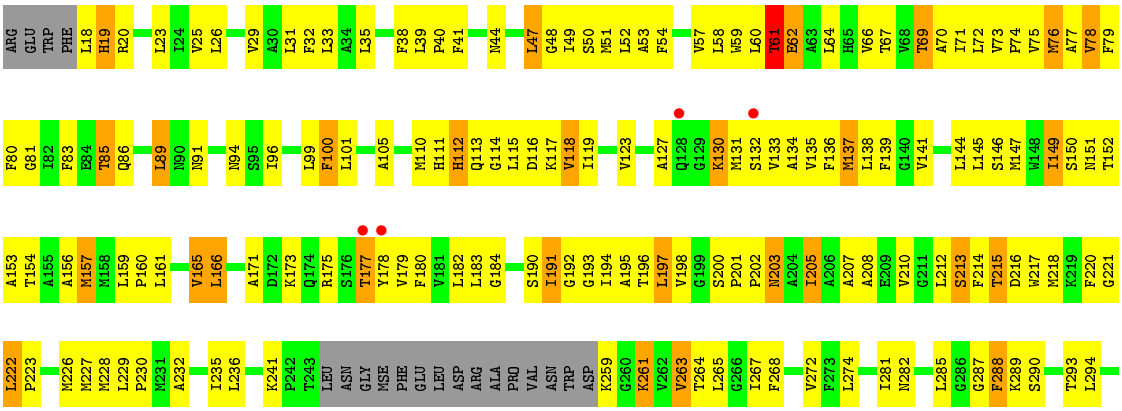
• Molecule 1: Transporter, NadC family

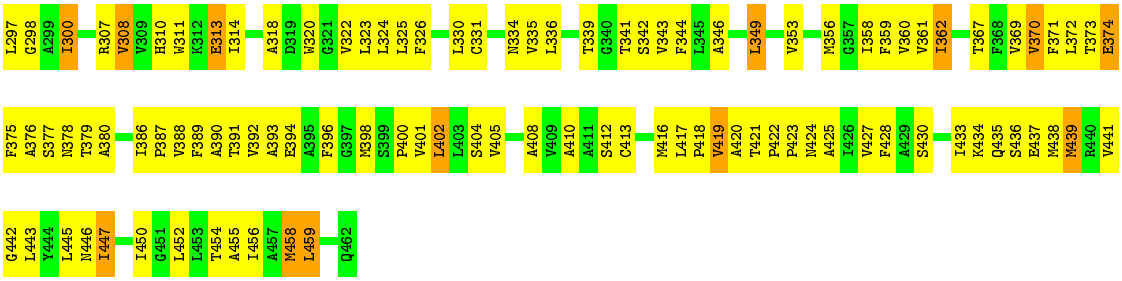
Chain A:



• Molecule 1: Transporter, NadC family

Chain C:




L297 G298 A299 I300 R307 V308 V309 H310 W311 K312 E313 I314 A318 D319 W320 G321 V322 L323 L324 L325 F326 L330 C331 N334 V335 L336 T339 G340 T341 S342 V343 F344 L345 A346 L349 V353 M356 G357 I358 F359 V360 V361 I362 T367 F368 V369 V370 F371 L372 T373 E374 F375 A376 S377 N378 T379 A380 I386 P387 V388 F389 A390 T391 V392 A393 E394 A395 F396 G397 M398 S399 P400 V401 L402 L403 S404 V405 A408 V409 A410 A411 S412 C413 M416 L417 P418 V419 A420 T421 P422 P423 N424 A425 I426 V427 F428 A429 S430 I433 K434 Q435 S436 E437 M438 M439 R440 V441 G442 L443 Y444 L445 N446 I447 I450 G451 L452 L453 T454 A455 I456 A457 M458 L459 Q462

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	101.33Å 100.87Å 164.57Å 90.00° 101.74° 90.00°	Depositor
Resolution (Å)	44.38 – 3.20 44.38 – 3.20	Depositor EDS
% Data completeness (in resolution range)	96.5 (44.38-3.20) 96.6 (44.38-3.20)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.53 (at 3.19Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.7.1_743)	Depositor
R, R_{free}	0.228 , 0.291 0.230 , 0.294	Depositor DCC
R_{free} test set	2026 reflections (4.04%)	DCC
Wilson B-factor (Å^2)	93.9	Xtriage
Anisotropy	0.263	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.28 , 100.4	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning[2]	$< \|L\| > = 0.48$, $< L^2 > = 0.31$	Xtriage
Outliers	1 of 52223 reflections (0.002%)	Xtriage
F_o,F_c correlation	0.92	EDS
Total number of atoms	12390	wwPDB-VP
Average B, all atoms (Å^2)	98.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.48% 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, 0.375 respectively for untwinned datasets, and 0.333, 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, BNG, CIT

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	0.47	0/3153	0.71	2/4278 (0.0%)
1	B	0.42	0/3024	0.67	2/4106 (0.0%)
1	C	0.50	0/3163	0.73	1/4292 (0.0%)
1	D	0.50	0/3119	0.74	2/4231 (0.0%)
All	All	0.47	0/12459	0.71	7/16907 (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	1
1	C	0	2
1	D	0	3
All	All	0	6

There are no bond length outliers.

The worst 5 of 7 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	B	199	GLY	N-CA-C	-6.01	98.08	113.10
1	D	325	LEU	CA-CB-CG	5.79	128.63	115.30
1	A	199	GLY	N-CA-C	-5.39	99.61	113.10
1	B	402	LEU	CA-CB-CG	-5.20	103.33	115.30
1	C	402	LEU	CA-CB-CG	-5.11	103.55	115.30

There are no chirality outliers.

5 of 6 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	433	ILE	Peptide
1	C	61	THR	Peptide
1	D	371	PHE	Peptide
1	D	61	THR	Peptide
1	D	64	LEU	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	3110	0	3137	280	0
1	B	2980	0	3005	296	0
1	C	3113	0	3139	284	0
1	D	3076	0	3109	293	0
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	21	0	30	5	0
3	C	21	0	30	1	0
3	D	21	0	30	3	0
4	A	13	0	5	3	0
4	B	13	0	5	6	0
4	C	13	0	5	5	0
5	B	1	0	0	0	0
5	C	1	0	0	0	0
5	D	3	0	0	1	0
All	All	12390	0	12495	1133	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 46.

The worst 5 of 1133 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:A:419:VAL:HG11	1:A:425:ALA:HB2	1.25	1.13
1:C:420:ALA:HB1	1:C:421:THR:HA	1.36	1.07

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:B:420:ALA:HB1	1:B:421:THR:HA	1.38	1.06
1:C:200:SER:HB2	1:C:201:PRO:HD2	1.37	1.04
1:A:377:SER:HB3	1:A:380:ALA:HB3	1.40	1.03

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	430/449 (96%)	355 (83%)	72 (17%)	3 (1%)	26	72
1	B	404/449 (90%)	360 (89%)	40 (10%)	4 (1%)	19	65
1	C	426/449 (95%)	369 (87%)	53 (12%)	4 (1%)	21	67
1	D	426/449 (95%)	356 (84%)	68 (16%)	2 (0%)	34	78
All	All	1686/1796 (94%)	1440 (85%)	233 (14%)	13 (1%)	24	69

5 of 13 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	78	VAL
1	B	78	VAL
1	C	78	VAL
1	B	40	PRO
1	A	376	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	307/334 (92%)	253 (82%)	54 (18%)	2 11
1	B	297/334 (89%)	251 (84%)	46 (16%)	3 15
1	C	310/334 (93%)	256 (83%)	54 (17%)	2 12
1	D	303/334 (91%)	259 (86%)	44 (14%)	4 19
All	All	1217/1336 (91%)	1019 (84%)	198 (16%)	3 14

5 of 198 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	461	TRP
1	A	200	SER
1	C	307	ARG
1	A	42	GLU
1	A	133	VAL

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (4) such sidechains are listed below:

Mol	Chain	Res	Type
1	B	90	ASN
1	B	282	ASN
1	A	94	ASN
1	C	151	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 carbohydrates in this entry.

5.6 Ligand geometry (i)

Of 10 ligands modelled in this entry, 4 are monoatomic - leaving 6 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	$\#\|Z\| > 2$
4	CIT	A	501	-	3,12,12	1.23	0	3,17,17	2.06	1 (33%)
4	CIT	B	501	-	3,12,12	1.38	0	3,17,17	2.28	2 (66%)
3	BNG	B	503	-	21,21,21	1.31	3 (14%)	26,26,26	1.86	5 (19%)
4	CIT	C	501	-	3,12,12	1.23	1 (33%)	3,17,17	1.99	1 (33%)
3	BNG	C	503	-	21,21,21	1.20	3 (14%)	26,26,26	1.25	3 (11%)
3	BNG	D	502	-	21,21,21	1.21	2 (9%)	26,26,26	1.89	9 (34%)

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
4	CIT	A	501	-	-	0/6/16/16	0/0/0/0
4	CIT	B	501	-	-	0/6/16/16	0/0/0/0
3	BNG	B	503	-	-	1/12/32/32	0/1/1/1
4	CIT	C	501	-	-	0/6/16/16	0/0/0/0
3	BNG	C	503	-	-	0/12/32/32	0/1/1/1
3	BNG	D	502	-	-	0/12/32/32	0/1/1/1

The worst 5 of 9 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	B	503	BNG	C3-C2	-4.06	1.41	1.52
3	D	502	BNG	C3-C2	-3.90	1.42	1.52
3	C	503	BNG	C3-C2	-3.56	1.43	1.52
3	B	503	BNG	C4-C3	-2.12	1.46	1.52
3	C	503	BNG	C4-C3	-2.01	1.47	1.52

The worst 5 of 21 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
3	D	502	BNG	O2-C2-C3	-3.50	102.45	110.34
4	A	501	CIT	C3-C4-C5	-3.45	109.44	114.96
4	B	501	CIT	C3-C4-C5	-3.40	109.52	114.96
4	C	501	CIT	C3-C4-C5	-3.35	109.60	114.96
3	B	503	BNG	O2-C2-C3	-2.92	103.77	110.34

There are no chirality outliers.

All (1) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	B	503	BNG	C1'-O1-C1-O5

There are no ring outliers.

6 monomers are involved in 23 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	A	501	CIT	3	0
4	B	501	CIT	6	0
3	B	503	BNG	5	0
4	C	501	CIT	5	0
3	C	503	BNG	1	0
3	D	502	BNG	3	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

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^{th} 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	412/449 (91%)	-0.31	4 (0%)	84	75	57, 93, 150, 196	0
1	B	393/449 (87%)	-0.32	3 (0%)	87	80	61, 103, 160, 188	0
1	C	408/449 (90%)	-0.43	4 (0%)	84	75	53, 87, 139, 187	0
1	D	408/449 (90%)	-0.41	3 (0%)	89	83	48, 89, 142, 229	1 (0%)
All	All	1621/1796 (90%)	-0.37	14 (0%)	85	78	48, 93, 150, 229	1 (0%)

The worst 5 of 14 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	177	THR	4.8
1	A	132	SER	3.8
1	D	432	HIS	3.5
1	C	132	SER	3.0
1	D	378	ASN	2.8

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^{th} 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	LLDF	B-factors(Å^2)	Q<0.9
2	NA	A	502	1/1	0.98	0.52	9.78	77,77,77,77	0
4	CIT	B	501	13/13	0.83	0.37	3.71	127,135,161,164	0
4	CIT	A	501	13/13	0.92	0.40	3.37	101,117,130,136	0
3	BNG	C	503	21/21	0.87	0.31	2.78	58,111,159,160	0
3	BNG	D	502	21/21	0.88	0.35	2.46	75,100,143,143	0
3	BNG	B	503	21/21	0.77	0.33	2.45	62,175,190,333	0
2	NA	B	502	1/1	0.97	0.40	2.26	71,71,71,71	0
2	NA	D	501	1/1	0.97	0.36	2.11	63,63,63,63	0
4	CIT	C	501	13/13	0.92	0.25	1.43	88,110,139,149	0
2	NA	C	502	1/1	0.97	0.19	-0.08	53,53,53,53	0

6.5 Other polymers (i)

There are no such residues in this entry.