



wwPDB X-ray Structure Validation Summary Report

Feb 28, 2014 – 07:00 PM GMT

PDB ID : 2O18
Title : Crystal structure of a Thiamine biosynthesis lipoprotein apbE, NorthEast Structural Genomics target ER559
Authors : Seetharaman, J.; Su, M.; Wang, D.; Fang, Y.; Cunningham, K.; Ma, L.; Xiao, R.; Liu, J.; Baran, M.C.; Acton, T.B.; Rost, B.; Montelione, G.T.; Hunt, J.F.; Tong, L.; Northeast Structural Genomics Consortium (NESG)
Deposited on : 2006-11-28
Resolution : 2.20 Å(reported)

This is a wwPDB 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/ValidationPDFNotes.html>

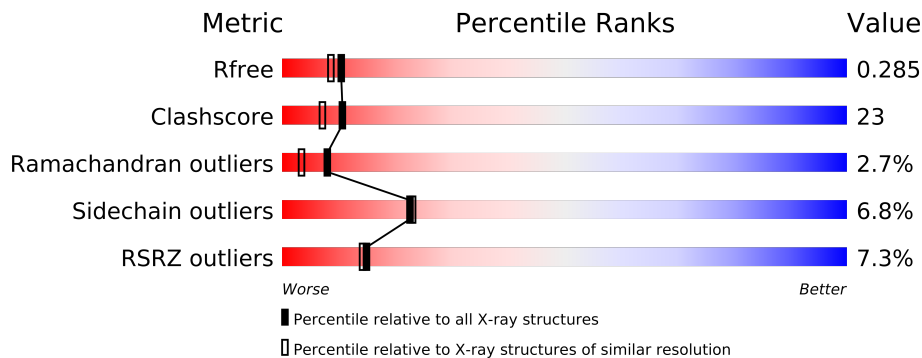
The following versions of software and data (see [references](#)) were used in the production of this report:

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 2.20 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}	66092	2938 (2.20-2.20)
Clashscore	79885	3751 (2.20-2.20)
Ramachandran outliers	78287	3681 (2.20-2.20)
Sidechain outliers	78261	3682 (2.20-2.20)
RSRZ outliers	66119	2939 (2.20-2.20)

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. 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	340	
1	B	340	
1	C	340	
1	D	340	

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

Mol	Type	Chain	Res	Geometry	Electron density
2	CA	B	341	-	X

2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 9555 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 Thiamine biosynthesis lipoprotein apbE.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	307	Total	C	N	O	Se	0	0	0
			2351	1476	406	459	10			
1	B	300	Total	C	N	O	Se	0	0	0
			2307	1449	396	452	10			
1	C	313	Total	C	N	O	Se	0	0	0
			2374	1491	410	463	10			
1	D	306	Total	C	N	O	Se	0	0	0
			2316	1457	398	451	10			

There are 76 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	MSE	-	INITIATING METHIONINE	UNP P0AB85
A	23	MSE	MET	MODIFIED RESIDUE	UNP P0AB85
A	67	MSE	MET	MODIFIED RESIDUE	UNP P0AB85
A	84	MSE	MET	MODIFIED RESIDUE	UNP P0AB85
A	102	MSE	MET	MODIFIED RESIDUE	UNP P0AB85
A	133	MSE	MET	MODIFIED RESIDUE	UNP P0AB85
A	180	MSE	MET	MODIFIED RESIDUE	UNP P0AB85
A	201	MSE	MET	MODIFIED RESIDUE	UNP P0AB85
A	287	MSE	MET	MODIFIED RESIDUE	UNP P0AB85
A	307	MSE	MET	MODIFIED RESIDUE	UNP P0AB85
A	319	MSE	MET	MODIFIED RESIDUE	UNP P0AB85
A	333	LEU	-	CLONING ARTIFACT	UNP P0AB85
A	334	GLU	-	CLONING ARTIFACT	UNP P0AB85
A	335	HIS	-	EXPRESSION TAG	UNP P0AB85
A	336	HIS	-	EXPRESSION TAG	UNP P0AB85
A	337	HIS	-	EXPRESSION TAG	UNP P0AB85
A	338	HIS	-	EXPRESSION TAG	UNP P0AB85
A	339	HIS	-	EXPRESSION TAG	UNP P0AB85
A	340	HIS	-	EXPRESSION TAG	UNP P0AB85
B	1	MSE	-	INITIATING METHIONINE	UNP P0AB85
B	23	MSE	MET	MODIFIED RESIDUE	UNP P0AB85

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	67	MSE	MET	MODIFIED RESIDUE	UNP P0AB85
B	84	MSE	MET	MODIFIED RESIDUE	UNP P0AB85
B	102	MSE	MET	MODIFIED RESIDUE	UNP P0AB85
B	133	MSE	MET	MODIFIED RESIDUE	UNP P0AB85
B	180	MSE	MET	MODIFIED RESIDUE	UNP P0AB85
B	201	MSE	MET	MODIFIED RESIDUE	UNP P0AB85
B	287	MSE	MET	MODIFIED RESIDUE	UNP P0AB85
B	307	MSE	MET	MODIFIED RESIDUE	UNP P0AB85
B	319	MSE	MET	MODIFIED RESIDUE	UNP P0AB85
B	333	LEU	-	CLONING ARTIFACT	UNP P0AB85
B	334	GLU	-	CLONING ARTIFACT	UNP P0AB85
B	335	HIS	-	EXPRESSION TAG	UNP P0AB85
B	336	HIS	-	EXPRESSION TAG	UNP P0AB85
B	337	HIS	-	EXPRESSION TAG	UNP P0AB85
B	338	HIS	-	EXPRESSION TAG	UNP P0AB85
B	339	HIS	-	EXPRESSION TAG	UNP P0AB85
B	340	HIS	-	EXPRESSION TAG	UNP P0AB85
C	1	MSE	-	INITIATING METHIONINE	UNP P0AB85
C	23	MSE	MET	MODIFIED RESIDUE	UNP P0AB85
C	67	MSE	MET	MODIFIED RESIDUE	UNP P0AB85
C	84	MSE	MET	MODIFIED RESIDUE	UNP P0AB85
C	102	MSE	MET	MODIFIED RESIDUE	UNP P0AB85
C	133	MSE	MET	MODIFIED RESIDUE	UNP P0AB85
C	180	MSE	MET	MODIFIED RESIDUE	UNP P0AB85
C	201	MSE	MET	MODIFIED RESIDUE	UNP P0AB85
C	287	MSE	MET	MODIFIED RESIDUE	UNP P0AB85
C	307	MSE	MET	MODIFIED RESIDUE	UNP P0AB85
C	319	MSE	MET	MODIFIED RESIDUE	UNP P0AB85
C	333	LEU	-	CLONING ARTIFACT	UNP P0AB85
C	334	GLU	-	CLONING ARTIFACT	UNP P0AB85
C	335	HIS	-	EXPRESSION TAG	UNP P0AB85
C	336	HIS	-	EXPRESSION TAG	UNP P0AB85
C	337	HIS	-	EXPRESSION TAG	UNP P0AB85
C	338	HIS	-	EXPRESSION TAG	UNP P0AB85
C	339	HIS	-	EXPRESSION TAG	UNP P0AB85
C	340	HIS	-	EXPRESSION TAG	UNP P0AB85
D	1	MSE	-	INITIATING METHIONINE	UNP P0AB85
D	23	MSE	MET	MODIFIED RESIDUE	UNP P0AB85
D	67	MSE	MET	MODIFIED RESIDUE	UNP P0AB85
D	84	MSE	MET	MODIFIED RESIDUE	UNP P0AB85
D	102	MSE	MET	MODIFIED RESIDUE	UNP P0AB85
D	133	MSE	MET	MODIFIED RESIDUE	UNP P0AB85

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
D	180	MSE	MET	MODIFIED RESIDUE	UNP P0AB85
D	201	MSE	MET	MODIFIED RESIDUE	UNP P0AB85
D	287	MSE	MET	MODIFIED RESIDUE	UNP P0AB85
D	307	MSE	MET	MODIFIED RESIDUE	UNP P0AB85
D	319	MSE	MET	MODIFIED RESIDUE	UNP P0AB85
D	333	LEU	-	CLONING ARTIFACT	UNP P0AB85
D	334	GLU	-	CLONING ARTIFACT	UNP P0AB85
D	335	HIS	-	EXPRESSION TAG	UNP P0AB85
D	336	HIS	-	EXPRESSION TAG	UNP P0AB85
D	337	HIS	-	EXPRESSION TAG	UNP P0AB85
D	338	HIS	-	EXPRESSION TAG	UNP P0AB85
D	339	HIS	-	EXPRESSION TAG	UNP P0AB85
D	340	HIS	-	EXPRESSION TAG	UNP P0AB85

- Molecule 2 is CALCIUM ION (three-letter code: CA) (formula: Ca).

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

- Molecule 3 is water.

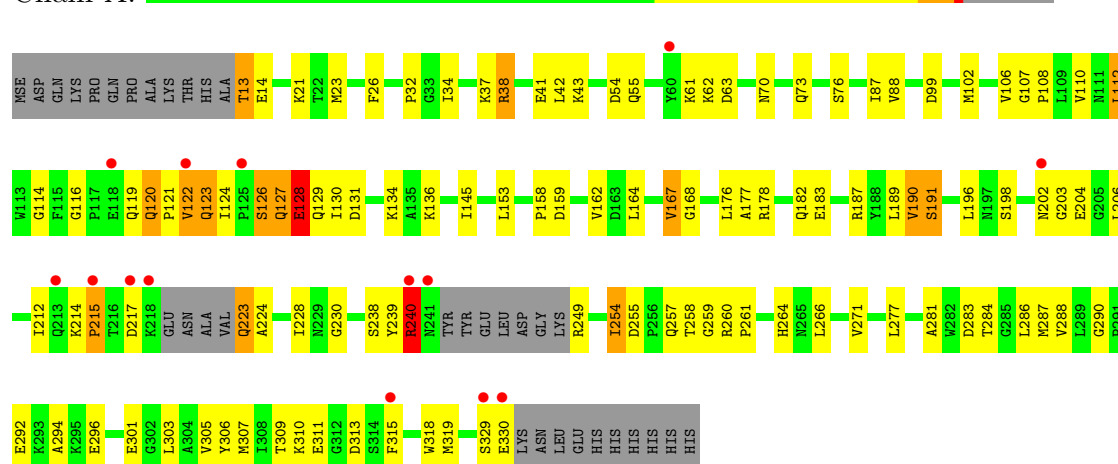
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	52	Total O 52 52	0	0
3	B	47	Total O 47 47	0	0
3	C	48	Total O 48 48	0	0
3	D	56	Total O 56 56	0	0

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 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.

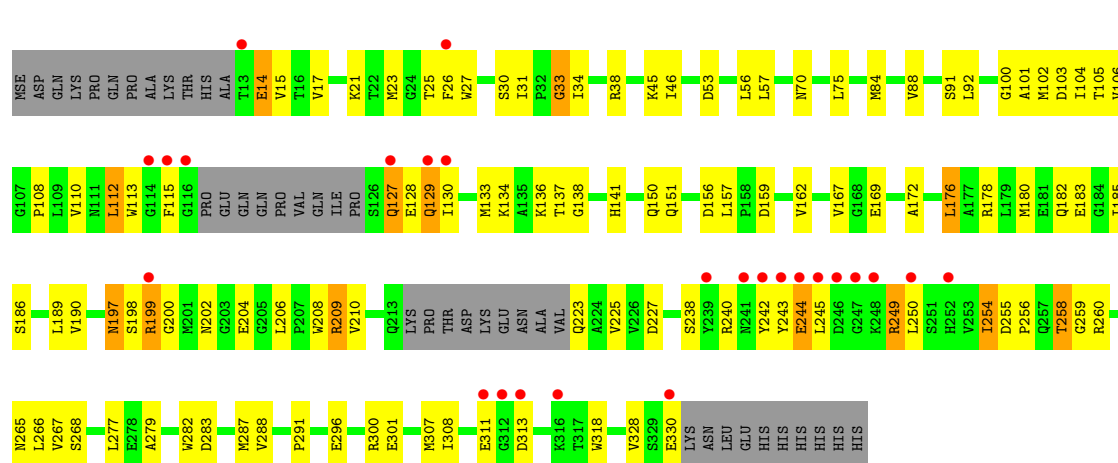
- Molecule 1: Thiamine biosynthesis lipoprotein apbE

Chain A:



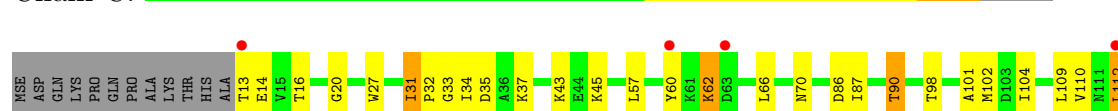
- Molecule 1: Thiamine biosynthesis lipoprotein apbE

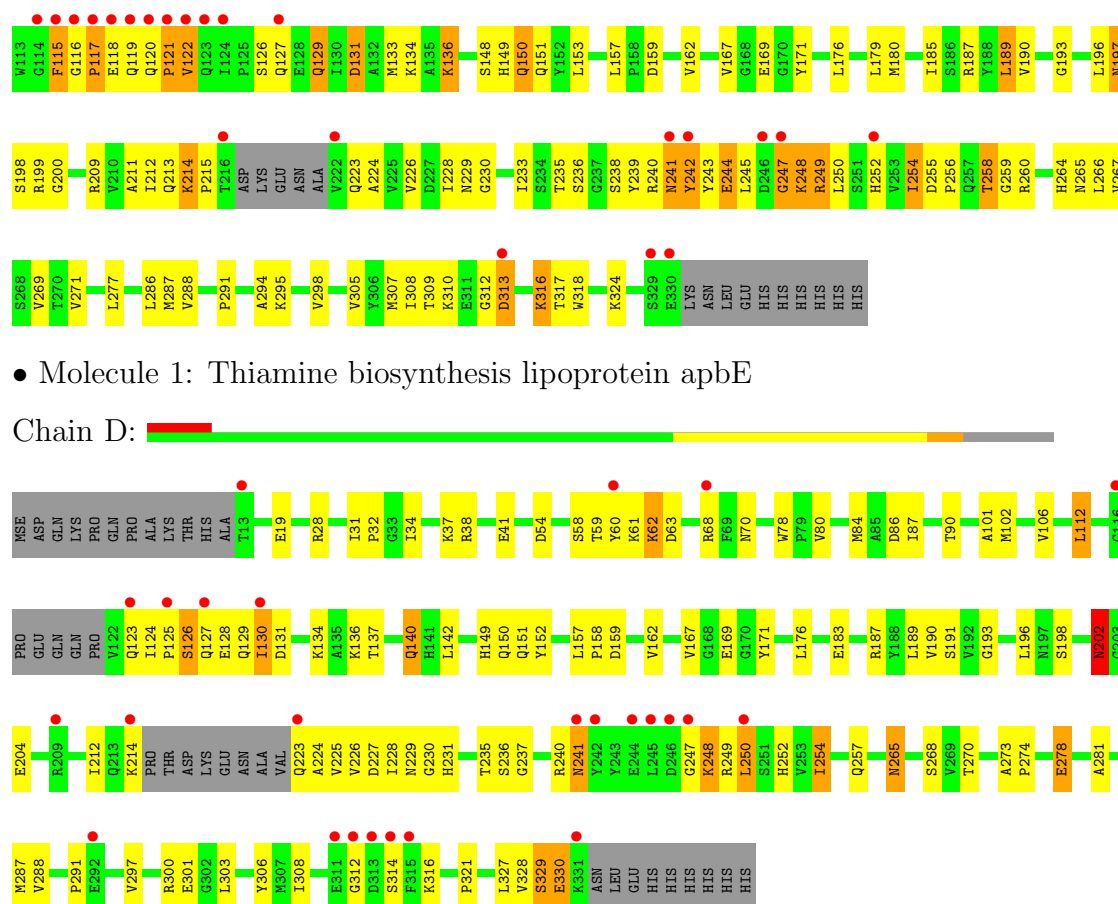
Chain B:



- Molecule 1: Thiamine biosynthesis lipoprotein apbE

Chain C:





4 Data and refinement statistics

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, α , β , γ	57.32Å 70.65Å 86.64Å 75.76° 71.67° 69.46°	Depositor
Resolution (Å)	49.74 – 2.20 49.74 – 2.20	Depositor EDS
% Data completeness (in resolution range)	78.0 (49.74-2.20) 94.5 (49.74-2.20)	Depositor EDS
R_{merge}	0.06	Depositor
R_{sym}	0.04	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.86 (at 2.20Å)	Xtriage
Refinement program	CNS 1.1	Depositor
R, R_{free}	0.233 , 0.284 0.236 , 0.285	Depositor DCC
R_{free} test set	2486 reflections (5.10%)	DCC
Wilson B-factor (Å ²)	20.9	Xtriage
Anisotropy	0.358	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.34 , 19.3	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning	$\langle L \rangle = 0.50$, $\langle L^2 \rangle = 0.33$	Xtriage
Outliers	0 of 114432 reflections	Xtriage
F_o, F_c correlation	0.89	EDS
Total number of atoms	9555	wwPDB-VP
Average B, all atoms (Å ²)	25.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.59% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities 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: CA

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	0.35	0/2383	0.65	1/3215 (0.0%)
1	B	0.34	0/2338	0.63	1/3149 (0.0%)
1	C	0.36	0/2407	0.63	1/3250 (0.0%)
1	D	0.36	0/2346	0.62	0/3165
All	All	0.35	0/9474	0.64	3/12779 (0.0%)

There are no bond length outliers.

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	C	121	PRO	N-CA-CB	5.60	110.03	103.30
1	B	243	TYR	N-CA-C	5.51	125.89	111.00
1	A	215	PRO	N-CA-CB	5.22	109.57	103.30

There are no chirality outliers.

There are no planarity outliers.

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	2351	0	2321	106	0
1	B	2307	0	2275	113	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	C	2374	0	2333	129	0
1	D	2316	0	2277	89	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	A	52	0	0	7	0
3	B	47	0	0	9	0
3	C	48	0	0	5	0
3	D	56	0	0	4	0
All	All	9555	0	9206	427	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 23.

The worst 5 of 427 close contacts within the same asymmetric unit are listed below.

Atom-1	Atom-2	Distance(Å)	Clash(Å)
1:A:202:ASN:HB3	3:A:343:HOH:O	1.71	0.89
1:B:189:LEU:HD12	1:B:208:TRP:HB3	1.53	0.89
1:B:46:ILE:HG23	1:B:176:LEU:HD21	1.56	0.86
1:A:112:LEU:HD22	1:A:126:SER:HB3	1.56	0.86
1:A:120:GLN:HB2	1:A:121:PRO:C	1.95	0.86

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	301/340 (88%)	275 (91%)	18 (6%)	8 (3%)	8 3
1	B	294/340 (86%)	268 (91%)	21 (7%)	5 (2%)	14 8
1	C	309/340 (91%)	273 (88%)	24 (8%)	12 (4%)	5 2
1	D	300/340 (88%)	273 (91%)	20 (7%)	7 (2%)	10 5

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles
All	All	1204/1360 (88%)	1089 (90%)	83 (7%)	32 (3%)	8 3

5 of 32 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	123	GLN
1	C	35	ASP
1	C	117	PRO
1	C	241	ASN
1	C	242	TYR

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	254/277 (92%)	238 (94%)	16 (6%)	25 27
1	B	248/277 (90%)	232 (94%)	16 (6%)	24 25
1	C	253/277 (91%)	234 (92%)	19 (8%)	19 19
1	D	246/277 (89%)	229 (93%)	17 (7%)	22 22
All	All	1001/1108 (90%)	933 (93%)	68 (7%)	22 23

5 of 68 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	313	ASP
1	C	129	GLN
1	D	227	ASP
1	B	330	GLU
1	C	62	LYS

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 29 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	197	ASN
1	C	70	ASN

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	D	202	ASN
1	B	241	ASN
1	C	147	GLN

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 4 ligands modelled in this entry, 4 are monoatomic - leaving 0 for Mogul analysis.

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, 95th 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	307/340 (90%)	0.23	14 (4%) 31 31	7, 21, 51, 61	0
1	B	300/340 (88%)	0.51	25 (8%) 11 11	9, 25, 52, 73	0
1	C	313/340 (92%)	0.52	26 (8%) 11 11	9, 23, 56, 72	0
1	D	306/340 (90%)	0.44	25 (8%) 12 11	6, 22, 53, 67	0
All	All	1226/1360 (90%)	0.43	90 (7%) 15 14	6, 23, 53, 73	0

The worst 5 of 90 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	D	246	ASP	11.1
1	C	122	VAL	9.0
1	C	116	GLY	8.4
1	C	246	ASP	8.2
1	C	117	PRO	6.6

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, 95th 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(\AA^2)	Q<0.9
2	CA	B	341	1/1	0.22	2.42	24,24,24,24	0
2	CA	A	341	1/1	0.14	1.57	23,23,23,23	0
2	CA	C	341	1/1	0.15	0.95	13,13,13,13	0
2	CA	D	341	1/1	0.10	-1.13	20,20,20,20	0

6.5 Other polymers ⓘ

There are no such residues in this entry.