WORLDWIDE PDB PROTEIN DATA BANK

wwPDB X-ray Structure Validation Summary Report ⓘ

Jun 19, 2024 – 06:04 AM EDT

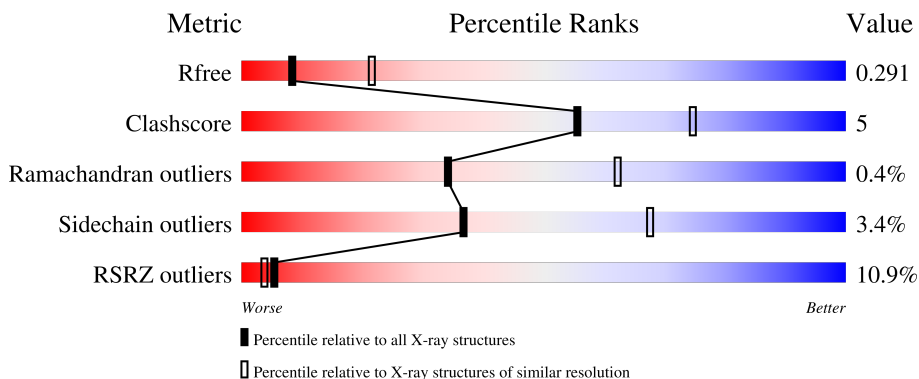|		
PDB ID	:	4EMY
Title	:	Crystal structure of aminotransferase from anaerococcus prevotii dsm 20548.
Authors	:	Chang, C.; Tesar, C.; Bearden, J.; Joachimiak, A.; Midwest Center for Structural Genomics (MCSG)
Deposited on	:	2012-04-12
Resolution	:	2.86 Å(reported)

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 2.86 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	Percentile Ranks	Value
Rfree		0.291
Clashscore		5
Ramachandran outliers		0.4%
Sidechain outliers		3.4%
RSRZ outliers		10.9%

Worse   *Better*

▮ Percentile relative to all X-ray structures

▯ Percentile relative to X-ray structures of similar resolution

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R$_{free}$	130704	3168 (2.90-2.82)
Clashscore	141614	3438 (2.90-2.82)
Ramachandran outliers	138981	3348 (2.90-2.82)
Sidechain outliers	138945	3351 (2.90-2.82)
RSRZ outliers	127900	3103 (2.90-2.82)

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 of 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% 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	413	82% 16%
1	B	413	85% 13%
1	C	413	27% 87% 11%
1	D	413	14% 81% 16%

2 Entry composition ⓘ

There are 3 unique types of molecules in this entry. The entry contains 13055 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 Aminotransferase class I and II.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	409	Total	C	N	O	S	Se	0	0	0
			3249	2080	538	621	4	6			
1	B	409	Total	C	N	O	S	Se	0	0	0
			3249	2080	538	621	4	6			
1	C	407	Total	C	N	O	S	Se	0	0	0
			3230	2065	536	619	4	6			
1	D	407	Total	C	N	O	S	Se	0	0	0
			3235	2073	536	616	4	6			

There are 12 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-2	SER	-	EXPRESSION TAG	UNP C7REB0
A	-1	ASN	-	EXPRESSION TAG	UNP C7REB0
A	0	ALA	-	EXPRESSION TAG	UNP C7REB0
B	-2	SER	-	EXPRESSION TAG	UNP C7REB0
B	-1	ASN	-	EXPRESSION TAG	UNP C7REB0
B	0	ALA	-	EXPRESSION TAG	UNP C7REB0
C	-2	SER	-	EXPRESSION TAG	UNP C7REB0
C	-1	ASN	-	EXPRESSION TAG	UNP C7REB0
C	0	ALA	-	EXPRESSION TAG	UNP C7REB0
D	-2	SER	-	EXPRESSION TAG	UNP C7REB0
D	-1	ASN	-	EXPRESSION TAG	UNP C7REB0
D	0	ALA	-	EXPRESSION TAG	UNP C7REB0

- Molecule 2 is PYRIDOXAL-5'-PHOSPHATE (three-letter code: PLP) (formula: $C_8H_{10}NO_6P$).

PLP

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf
2	A	1	Total	C	N	O	P		0	0
			16	8	1	6	1			
2	B	1	Total	C	N	O	P		0	0
			16	8	1	6	1			
2	C	1	Total	C	N	O	P		0	0
			16	8	1	6	1			
2	D	1	Total	C	N	O	P		0	0
			16	8	1	6	1			

- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	7	Total	O	0	0
			7	7		
3	B	10	Total	O	0	0
			10	10		
3	C	4	Total	O	0	0
			4	4		
3	D	7	Total	O	0	0
			7	7		

3   Residue-property plots ⓘ

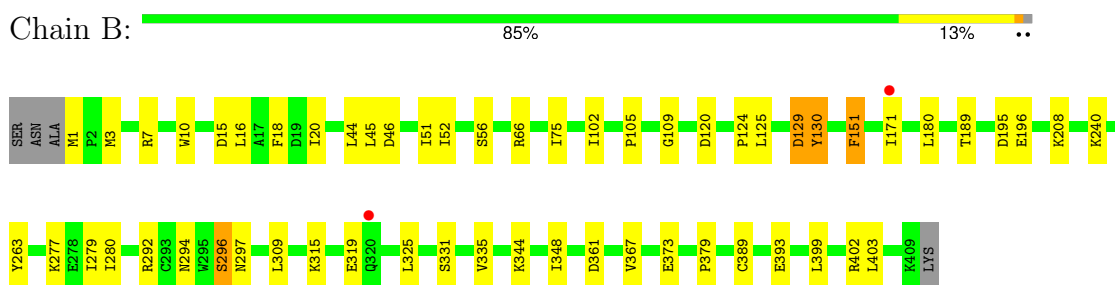These plots are drawn for all protein, RNA, DNA and oligosaccharide chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. 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.
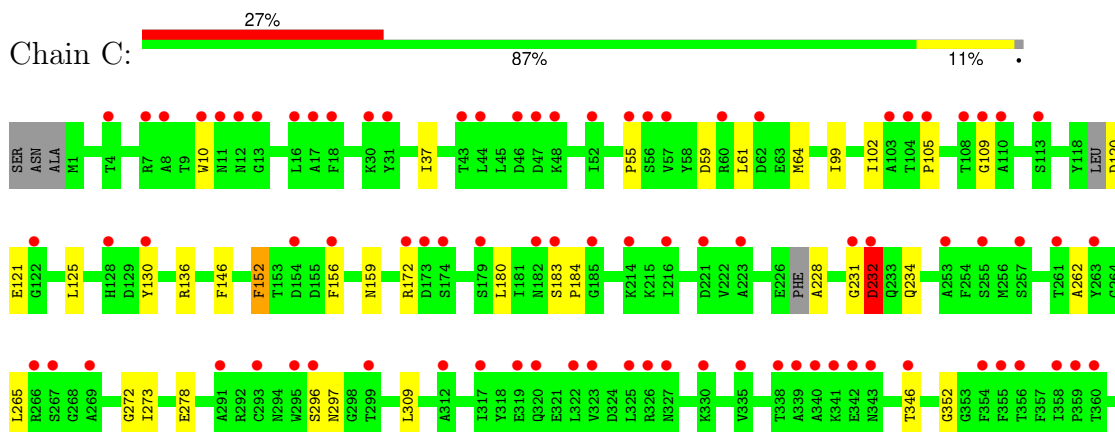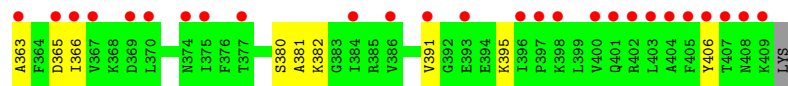
- Molecule 1: Aminotransferase class I and II

- Molecule 1: Aminotransferase class I and II

- Molecule 1: Aminotransferase class I and II

A363 F364 D365 I366 V367 K368 D369 L370 N374 I375 F376 T377 S380 A381 K382 G383 I384 R385 V386 V391 G392 E393 E394 K395 I396 P397 K398 L399 V400 Q401 R402 L403 A404 F405 Y406 T407 N408 K409 LYS

• Molecule 1: Aminotransferase class I and II

Chain D:

14%

81%   16%

SER ASN ALA M1 I5 K6 R7 R33 L44 L45 D46 D47 K48 I52 V57 R60 E63 M64 D65 R66 I69 A70 E76 G77 E78 R82 K83 D87 T88 L89 F90 G91 P92 Y93 K94 Y98 I99 I102 A103 T104 P105 G106 G109 R112

L119 ASP E121 G122 D123 P124 L125 I126 D129 Y130 Y131 W132 A133 P134 R144 T148 F151 F152 F156 A157 F158 A166 I167 D168 E169 G170 I171 R172 D175 R176 I177 A178 T189 E196 E197 W198 D199 L205 E210 D232 K240 F241 L248 F249 V250

K258 S267 G272 S276 K277 E278 E281 E282 F283 S290 A291 R292 C293 N294 W295 SER N297 G298 T299 H300 A301 I305 L309 E310 R311 A312 E313 N314 K315 K316 E319 V323 D324 L325 F336 K341 M347 I348 P349 G353 F354 P359 D365 I366 V367

E373 P379 K382 A387 I388 V391 G392 E393 E394 K395 I396 L399 V400 Q401 R402 L403 K409 LYS

4 Data and refinement statistics ⓘ

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	64.30Å 131.18Å 239.91Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	50.00 – 2.86 45.10 – 2.86	Depositor EDS
% Data completeness (in resolution range)	99.4 (50.00-2.86) 99.6 (45.10-2.86)	Depositor EDS
R_{merge}	0.13	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	4.04 (at 2.86Å)	Xtriage
Refinement program	REFMAC 5.6.0117	Depositor
R, R_{free}	0.245 , 0.290 0.246 , 0.291	Depositor DCC
R_{free} test set	2420 reflections (5.05%)	wwPDB-VP
Wilson B-factor (Å^2)	50.4	Xtriage
Anisotropy	0.960	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.29 , 36.8	EDS
L-test for twinning[2]	$< |L| > = 0.50, < L^2 > = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.91	EDS
Total number of atoms	13055	wwPDB-VP
Average B, all atoms (Å^2)	75.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.05% 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: PLP

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	0.42	0/3315	0.55	0/4470
1	B	0.40	0/3315	0.54	0/4470
1	C	0.41	1/3293 (0.0%)	0.51	0/4437
1	D	0.41	1/3299 (0.0%)	0.51	0/4445
All	All	0.41	2/13222 (0.0%)	0.53	0/17822

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	C	10	TRP	CD2-CE2	5.11	1.47	1.41
1	D	198	TRP	CD2-CE2	5.05	1.47	1.41

There are no bond angle outliers.

There are no chirality outliers.

There are no planarity outliers.

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	3249	0	3200	46	0
1	B	3249	0	3200	32	0
1	C	3230	0	3178	21	0
1	D	3235	0	3189	40	0
2	A	16	0	7	1	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	B	16	0	8	0	0
2	C	16	0	8	0	0
2	D	16	0	8	2	0
3	A	7	0	0	1	0
3	B	10	0	0	1	0
3	C	4	0	0	0	0
3	D	7	0	0	0	0
All	All	13055	0	12798	132	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 5.

The worst 5 of 132 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:C:228:ALA:HA	1:C:352:GLY:HA3	1.46	0.94
1:A:112:ARG:HD2	1:B:294:ASN:HA	1.64	0.78
1:D:394:GLU:CD	1:D:394:GLU:H	1.89	0.75
1:A:111:ILE:HD13	1:A:135:TYR:HE1	1.57	0.70
1:B:105:PRO:HD2	1:B:109:GLY:HA3	1.74	0.69

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	407/413 (98%)	386 (95%)	20 (5%)	1 (0%)	47	75
1	B	407/413 (98%)	391 (96%)	14 (3%)	2 (0%)	29	57
1	C	401/413 (97%)	370 (92%)	29 (7%)	2 (0%)	29	57
1	D	401/413 (97%)	375 (94%)	24 (6%)	2 (0%)	29	57

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles
All	All	1616/1652 (98%)	1522 (94%)	87 (5%)	7 (0%)	34 62

5 of 7 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	33	ARG
1	C	381	ALA
1	B	296	SER
1	D	172	ARG
1	D	387	ALA

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	344/341 (101%)	334 (97%)	10 (3%)	42 72
1	B	344/341 (101%)	332 (96%)	12 (4%)	36 67
1	C	342/341 (100%)	331 (97%)	11 (3%)	39 69
1	D	342/341 (100%)	329 (96%)	13 (4%)	33 64
All	All	1372/1364 (101%)	1326 (97%)	46 (3%)	37 67

5 of 46 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	278	GLU
1	D	151	PHE
1	C	296	SER
1	C	346	THR
1	D	196	GLU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

5.3.3 RNA ⓘ

There are no RNA molecules 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 monosaccharides in this entry.

5.6 Ligand geometry ⓘ

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	PLP	C	501	-	16,16,16	1.28	2 (12%)	20,23,23	1.24	3 (15%)
2	PLP	D	501	-	16,16,16	1.26	2 (12%)	20,23,23	1.39	2 (10%)
2	PLP	B	501	-	16,16,16	1.30	1 (6%)	20,23,23	1.44	4 (20%)
2	PLP	A	501	-	16,16,16	1.30	1 (6%)	20,23,23	1.29	3 (15%)

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	PLP	C	501	-	-	2/8/8/8	0/1/1/1
2	PLP	D	501	-	-	2/8/8/8	0/1/1/1
2	PLP	B	501	-	-	4/8/8/8	0/1/1/1
2	PLP	A	501	-	-	5/8/8/8	0/1/1/1

The worst 5 of 6 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	D	501	PLP	C2-N1	3.14	1.39	1.33
2	C	501	PLP	C2-N1	3.06	1.39	1.33
2	B	501	PLP	C2-N1	2.97	1.39	1.33
2	A	501	PLP	C2-N1	2.84	1.38	1.33
2	D	501	PLP	C6-N1	2.33	1.39	1.34

The worst 5 of 12 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	B	501	PLP	C5-C6-N1	-2.80	119.27	123.83
2	D	501	PLP	C3-C4-C5	2.80	120.52	118.28
2	D	501	PLP	C5-C6-N1	-2.76	119.34	123.83
2	A	501	PLP	C3-C4-C4A	-2.65	116.20	119.84
2	C	501	PLP	C3-C4-C5	2.43	120.23	118.28

There are no chirality outliers.

5 of 13 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	A	501	PLP	C3-C4-C4A-O4A
2	A	501	PLP	C5A-O4P-P-O1P
2	A	501	PLP	C5A-O4P-P-O2P
2	B	501	PLP	C3-C4-C4A-O4A
2	B	501	PLP	C5A-O4P-P-O3P

There are no ring outliers.

2 monomers are involved in 3 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	D	501	PLP	2	0
2	A	501	PLP	1	0

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	403/413 (97%)	-0.10	5 (1%)	79	78	34, 48, 64, 86	0
1	B	403/413 (97%)	-0.07	2 (0%)	91	90	33, 46, 70, 96	0
1	C	401/413 (97%)	1.41	111 (27%)	0	0	61, 108, 144, 233	0
1	D	401/413 (97%)	0.84	57 (14%)	2	2	50, 95, 129, 220	0
All	All	1608/1652 (97%)	0.52	175 (10%)	5	4	33, 67, 136, 233	0

The worst 5 of 175 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	384	ILE	8.4
1	C	363	ALA	7.8
1	C	11	ASN	7.1
1	C	408	ASN	7.0
1	C	400	VAL	6.8

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 monosaccharides 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. 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	RSCC	RSR	B-factors($Å^2$)	Q<0.9
2	PLP	C	501	16/16	0.84	0.35	59,68,74,79	0
2	PLP	D	501	16/16	0.91	0.17	38,46,57,57	0
2	PLP	A	501	16/16	0.96	0.18	23,28,33,42	0
2	PLP	B	501	16/16	0.97	0.20	30,37,43,50	0

6.5 Other polymers ⓘ

There are no such residues in this entry.