wwPDB X-ray Structure Validation Summary Report (i)

Oct 27, 2016 – 05:08 PM EDT

PDB ID	:	5ERB
Title	:	Ketosynthase from module 5 of the bacillaene synthase from Bacillus amyloliquefaciens FZB42
Authors	:	Wagner, D.T.; Gay, D.C.; Keatinge-Clay, A.T.; Meinke, J.
Deposited on	:	2015-11-13
Resolution	:	4.20 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:

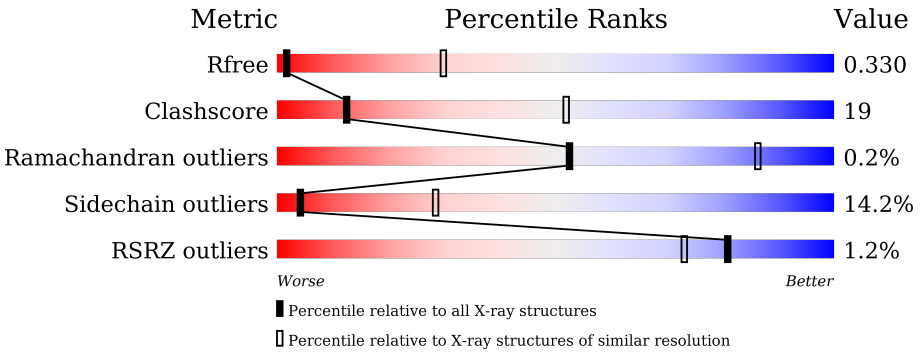
MolProbity	:	4.02b-467
Mogul	:	unknown
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20027939
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	rb-20027939

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 4.20 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	1039 (4.80-3.60)
Clashscore	102246	1140 (4.80-3.60)
Ramachandran outliers	100387	1083 (4.80-3.60)
Sidechain outliers	100360	1067 (4.80-3.60)
RSRZ outliers	91569	1042 (4.80-3.60)

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	640	61% 24% 5% • 9%
1	B	640	60% 25% 7% • 7%
1	C	640	60% 23% 7% • 8%
1	D	640	61% 24% 7% • 8%

2 Entry composition

There is only 1 type of molecule in this entry. The entry contains 18315 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 Polyketide synthase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	584	Total	C	N	O	S	0	0	0
			4543	2859	790	873	21			
1	B	593	Total	C	N	O	S	0	0	0
			4618	2907	798	892	21			
1	C	586	Total	C	N	O	S	0	0	0
			4562	2872	792	878	20			
1	D	590	Total	C	N	O	S	0	0	0
			4592	2886	797	888	21			

There are 80 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-19	MET	-	initiating methionine	UNP Q1RS72
A	-18	GLY	-	expression tag	UNP Q1RS72
A	-17	SER	-	expression tag	UNP Q1RS72
A	-16	SER	-	expression tag	UNP Q1RS72
A	-15	HIS	-	expression tag	UNP Q1RS72
A	-14	HIS	-	expression tag	UNP Q1RS72
A	-13	HIS	-	expression tag	UNP Q1RS72
A	-12	HIS	-	expression tag	UNP Q1RS72
A	-11	HIS	-	expression tag	UNP Q1RS72
A	-10	HIS	-	expression tag	UNP Q1RS72
A	-9	SER	-	expression tag	UNP Q1RS72
A	-8	SER	-	expression tag	UNP Q1RS72
A	-7	GLY	-	expression tag	UNP Q1RS72
A	-6	LEU	-	expression tag	UNP Q1RS72
A	-5	VAL	-	expression tag	UNP Q1RS72
A	-4	PRO	-	expression tag	UNP Q1RS72
A	-3	ARG	-	expression tag	UNP Q1RS72
A	-2	GLY	-	expression tag	UNP Q1RS72
A	-1	SER	-	expression tag	UNP Q1RS72
A	0	SER	-	expression tag	UNP Q1RS72
B	-19	MET	-	initiating methionine	UNP Q1RS72

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	-18	GLY	-	expression tag	UNP Q1RS72
B	-17	SER	-	expression tag	UNP Q1RS72
B	-16	SER	-	expression tag	UNP Q1RS72
B	-15	HIS	-	expression tag	UNP Q1RS72
B	-14	HIS	-	expression tag	UNP Q1RS72
B	-13	HIS	-	expression tag	UNP Q1RS72
B	-12	HIS	-	expression tag	UNP Q1RS72
B	-11	HIS	-	expression tag	UNP Q1RS72
B	-10	HIS	-	expression tag	UNP Q1RS72
B	-9	SER	-	expression tag	UNP Q1RS72
B	-8	SER	-	expression tag	UNP Q1RS72
B	-7	GLY	-	expression tag	UNP Q1RS72
B	-6	LEU	-	expression tag	UNP Q1RS72
B	-5	VAL	-	expression tag	UNP Q1RS72
B	-4	PRO	-	expression tag	UNP Q1RS72
B	-3	ARG	-	expression tag	UNP Q1RS72
B	-2	GLY	-	expression tag	UNP Q1RS72
B	-1	SER	-	expression tag	UNP Q1RS72
B	0	SER	-	expression tag	UNP Q1RS72
C	-19	MET	-	initiating methionine	UNP Q1RS72
C	-18	GLY	-	expression tag	UNP Q1RS72
C	-17	SER	-	expression tag	UNP Q1RS72
C	-16	SER	-	expression tag	UNP Q1RS72
C	-15	HIS	-	expression tag	UNP Q1RS72
C	-14	HIS	-	expression tag	UNP Q1RS72
C	-13	HIS	-	expression tag	UNP Q1RS72
C	-12	HIS	-	expression tag	UNP Q1RS72
C	-11	HIS	-	expression tag	UNP Q1RS72
C	-10	HIS	-	expression tag	UNP Q1RS72
C	-9	SER	-	expression tag	UNP Q1RS72
C	-8	SER	-	expression tag	UNP Q1RS72
C	-7	GLY	-	expression tag	UNP Q1RS72
C	-6	LEU	-	expression tag	UNP Q1RS72
C	-5	VAL	-	expression tag	UNP Q1RS72
C	-4	PRO	-	expression tag	UNP Q1RS72
C	-3	ARG	-	expression tag	UNP Q1RS72
C	-2	GLY	-	expression tag	UNP Q1RS72
C	-1	SER	-	expression tag	UNP Q1RS72
C	0	SER	-	expression tag	UNP Q1RS72
D	-19	MET	-	initiating methionine	UNP Q1RS72
D	-18	GLY	-	expression tag	UNP Q1RS72
D	-17	SER	-	expression tag	UNP Q1RS72

Continued on next page...

Continued from previous page...

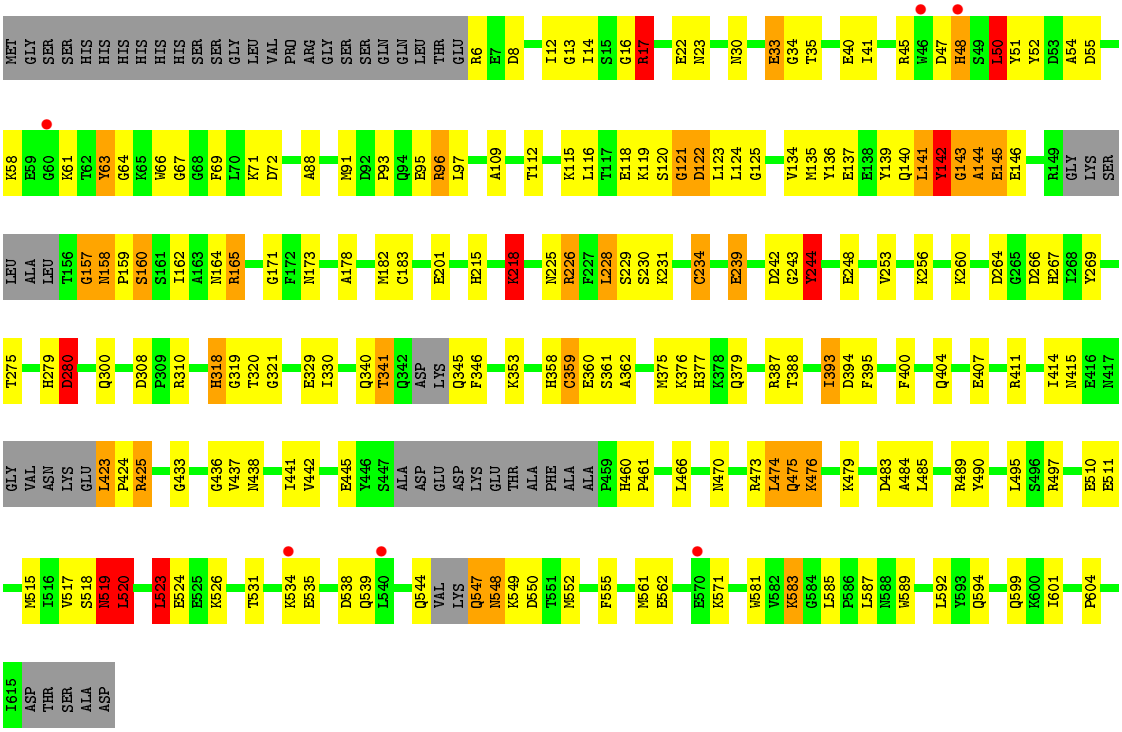
Chain	Residue	Modelled	Actual	Comment	Reference
D	-16	SER	-	expression tag	UNP Q1RS72
D	-15	HIS	-	expression tag	UNP Q1RS72
D	-14	HIS	-	expression tag	UNP Q1RS72
D	-13	HIS	-	expression tag	UNP Q1RS72
D	-12	HIS	-	expression tag	UNP Q1RS72
D	-11	HIS	-	expression tag	UNP Q1RS72
D	-10	HIS	-	expression tag	UNP Q1RS72
D	-9	SER	-	expression tag	UNP Q1RS72
D	-8	SER	-	expression tag	UNP Q1RS72
D	-7	GLY	-	expression tag	UNP Q1RS72
D	-6	LEU	-	expression tag	UNP Q1RS72
D	-5	VAL	-	expression tag	UNP Q1RS72
D	-4	PRO	-	expression tag	UNP Q1RS72
D	-3	ARG	-	expression tag	UNP Q1RS72
D	-2	GLY	-	expression tag	UNP Q1RS72
D	-1	SER	-	expression tag	UNP Q1RS72
D	0	SER	-	expression tag	UNP Q1RS72

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.

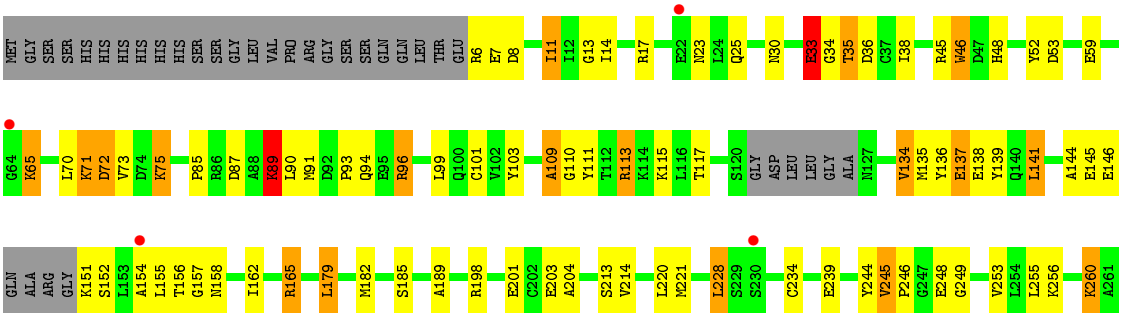
- Molecule 1: Polyketide synthase

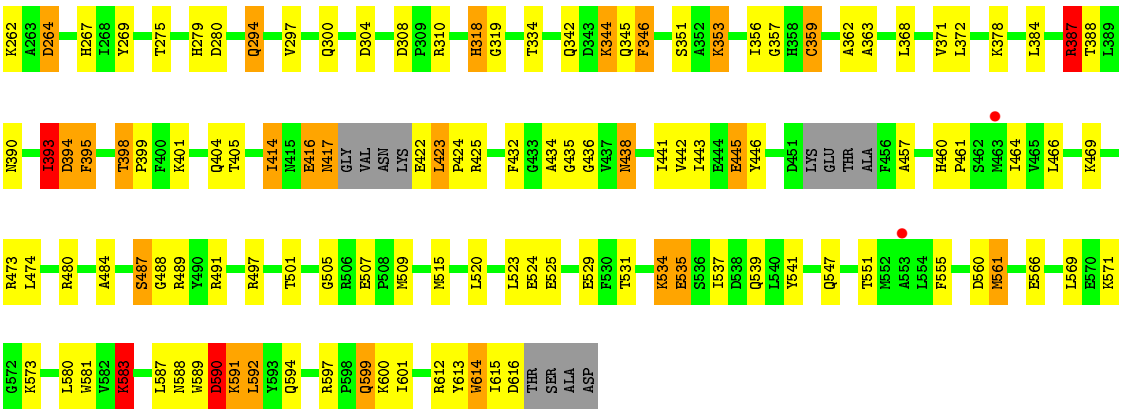
Chain A:



- Molecule 1: Polyketide synthase

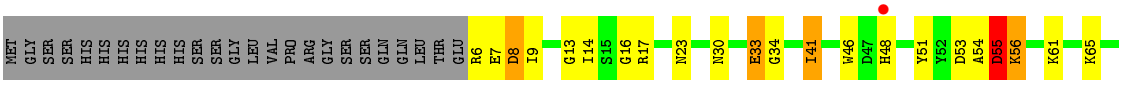
Chain B:

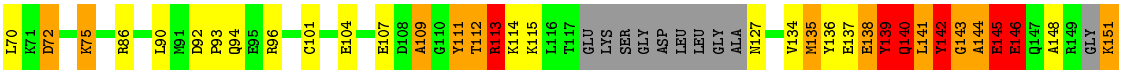


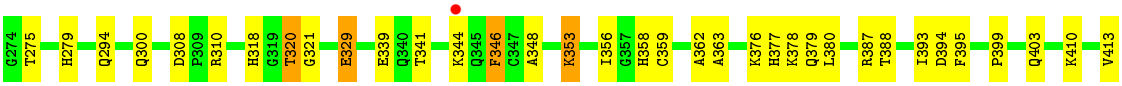


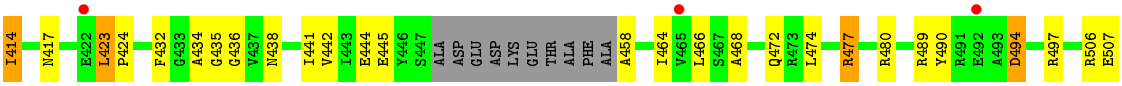
• Molecule 1: Polyketide synthase

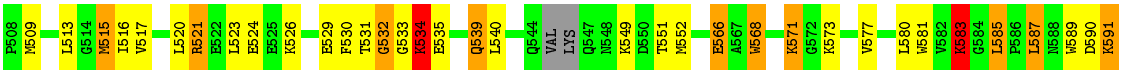
Chain C:





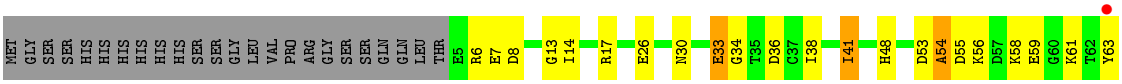


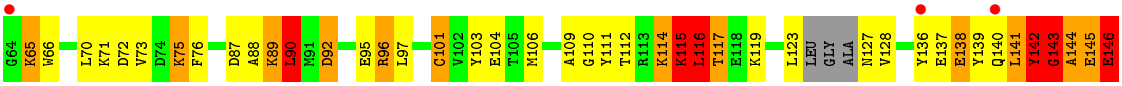


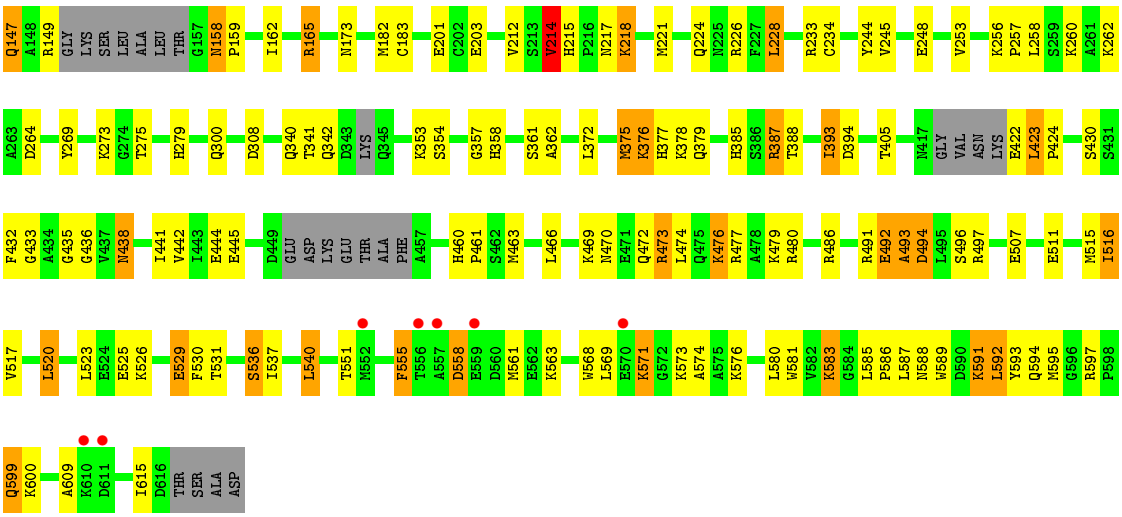


• Molecule 1: Polyketide synthase

Chain D:






Q147 A148 R149 GLY LYS SER LEU ALA LEU THR G157 M158 P159 I162 R165 N173 M182 C183 E201 C202 E203 V212 S213 Y214 H215 P216 N217 K218 M221 Q224 M225 R226 F227 L228 R233 C234 Y244 V245 E248 V253 K256 P257 L258 S259 K260 A261 K262
A263 D264 Y269 K273 G274 T275 H279 Q300 D308 Q340 T341 Q342 D343 LYS Q345 K353 S354 G357 H358 S361 A362 L372 M375 K376 H377 K378 Q379 H385 S386 R387 T388 I393 D394 T405 N417 GLY VAL ASN LYS E422 L423 P424 S430 S431
F432 G433 A434 G435 G436 V437 N438 I441 V442 I443 E444 E445 D449 GLU ASP LYS GLU THR ALA PHE A457 H460 P461 S462 M463 L466 K469 N470 E471 Q472 R473 L474 Q475 K476 R477 A478 K479 R480 R486 R491 E492 A493 D494 L495 S496 R497 E507 E511 M515 I516
V517 L520 L523 E524 E525 K526 E529 F530 T531 S536 I537 L540 T551 M552 F555 T556 A557 D558 E559 D560 M561 E562 K563 W568 L569 E570 K571 G572 K573 A574 A575 K576 L580 W581 V582 K583 G584 L585 P586 L587 N588 W589 D590 K591 L592 Y593 Q594 M595 G596 R597 P598
Q599 K600 A609 K610 D611 I615 D616 THR SER ALA ASP

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	70.76Å 319.13Å 103.47Å 90.00° 110.26° 90.00°	Depositor
Resolution (Å)	97.07 – 4.20 47.98 – 4.20	Depositor EDS
% Data completeness (in resolution range)	79.9 (97.07-4.20) 78.9 (47.98-4.20)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.19 (at 4.14Å)	Xtriage
Refinement program	REFMAC 5.8.0107	Depositor
R, R_{free}	0.288 , 0.341 0.281 , 0.330	Depositor DCC
R_{free} test set	1279 reflections (5.39%)	DCC
Wilson B-factor (Å^2)	168.6	Xtriage
Anisotropy	0.045	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.26 , 149.5	EDS
L-test for twinning[2]	$< \|L\| > = 0.44$, $< L^2 > = 0.26$	Xtriage
Estimated twinning fraction	0.115 for h,-k,-h-l	Xtriage
F_o,F_c correlation	0.90	EDS
Total number of atoms	18315	wwPDB-VP
Average B, all atoms (Å^2)	198.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.12% 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

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	0.62	0/4632	1.20	61/6245 (1.0%)
1	B	0.61	0/4709	1.28	56/6351 (0.9%)
1	C	0.62	0/4651	1.30	59/6272 (0.9%)
1	D	0.62	0/4681	1.22	50/6312 (0.8%)
All	All	0.62	0/18673	1.25	226/25180 (0.9%)

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	8
1	B	0	3
1	C	0	7
1	D	0	2
All	All	0	20

There are no bond length outliers.

The worst 5 of 226 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	D	54	ALA	CB-CA-C	24.75	147.22	110.10
1	C	152	SER	N-CA-C	17.36	157.87	111.00
1	C	152	SER	N-CA-CB	-17.30	84.55	110.50
1	B	144	ALA	N-CA-C	17.01	156.92	111.00
1	B	144	ALA	CB-CA-C	-15.76	86.46	110.10

There are no chirality outliers.

5 of 20 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	121	GLY	Mainchain
1	A	157	GLY	Mainchain
1	A	244	TYR	Mainchain
1	A	475	GLN	Peptide
1	A	72	ASP	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	4543	0	4470	151	0
1	B	4618	0	4540	145	1
1	C	4562	0	4491	258	0
1	D	4592	0	4505	211	1
All	All	18315	0	18006	708	1

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 19.

The worst 5 of 708 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:141:LEU:HD21	1:A:215:HIS:CE1	1.26	1.63
1:D:142:TYR:CE2	1:D:615:ILE:HG12	1.29	1.56
1:D:142:TYR:CE2	1:D:615:ILE:CG1	1.77	1.56
1:C:135:MET:CG	1:D:159:PRO:HD2	1.18	1.52
1:C:458:ALA:N	1:C:594:GLN:HB3	1.30	1.46

All (1) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:B:387:ARG:O	1:D:405:THR:OG1[2_646]	2.01	0.19

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	572/640 (89%)	529 (92%)	40 (7%)	3 (0%)	34 77
1	B	583/640 (91%)	543 (93%)	38 (6%)	2 (0%)	46 83
1	C	574/640 (90%)	533 (93%)	41 (7%)	0	100 100
1	D	578/640 (90%)	543 (94%)	35 (6%)	0	100 100
All	All	2307/2560 (90%)	2148 (93%)	154 (7%)	5 (0%)	52 86

All (5) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	538	ASP
1	A	141	LEU
1	A	142	TYR
1	B	113	ARG
1	B	246	PRO

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	479/524 (91%)	420 (88%)	59 (12%)	6 33
1	B	488/524 (93%)	411 (84%)	77 (16%)	3 23
1	C	482/524 (92%)	414 (86%)	68 (14%)	4 28
1	D	484/524 (92%)	414 (86%)	70 (14%)	4 27
All	All	1933/2096 (92%)	1659 (86%)	274 (14%)	4 28

5 of 274 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	547	GLN
1	C	155	LEU
1	D	480	ARG
1	B	573	LYS
1	C	46	TRP

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 28 such sidechains are listed below:

Mol	Chain	Res	Type
1	C	217	ASN
1	C	318	HIS
1	D	292	ASN
1	C	225	ASN
1	C	267	HIS

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)

There are no ligands in this entry.

5.7 Other polymers (i)

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^{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	584/640 (91%)	-0.23	6 (1%)	84	77	23, 191, 295, 377	0
1	B	593/640 (92%)	-0.20	6 (1%)	84	77	30, 189, 285, 364	0
1	C	586/640 (91%)	-0.18	6 (1%)	84	77	23, 193, 285, 402	0
1	D	590/640 (92%)	-0.19	11 (1%)	70	60	23, 191, 285, 399	0
All	All	2353/2560 (91%)	-0.20	29 (1%)	81	73	23, 191, 290, 402	0

The worst 5 of 29 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	D	557	ALA	4.8
1	A	60	GLY	3.5
1	C	344	LYS	3.5
1	C	422	GLU	3.4
1	B	154	ALA	3.3

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

There are no ligands in this entry.

6.5 Other polymers (i)

There are no such residues in this entry.