wwPDB X-ray Structure Validation Summary Report (i)

Apr 28, 2024 – 10:19 am BST

PDB ID	:	2VHI
Title	:	Crystal structure of a pyrimidine degrading enzyme from Drosophila melanogaster
Authors	:	Lundgren, S.; Lohkamp, B.; Andersen, B.; Piskur, J.; Dobritzsch, D.
Deposited on	:	2007-11-21
Resolution	:	3.30 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 types of validation reports are described at
http://www.wwpdb.org/validation/2017/FAQs#types.

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

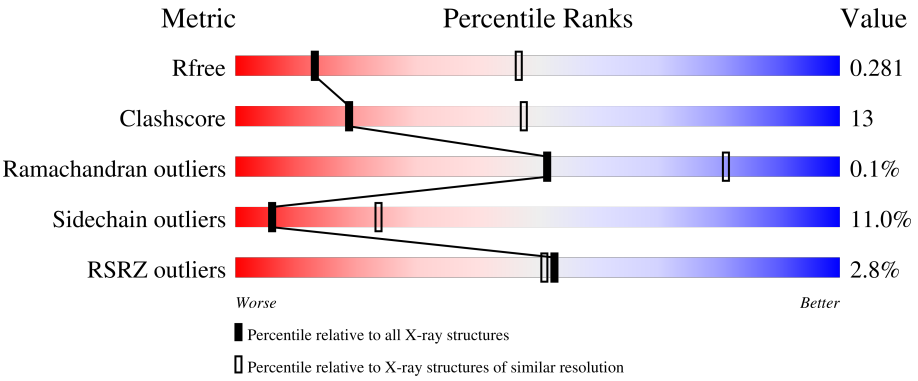
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.36.2
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36.2

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 3.30 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.



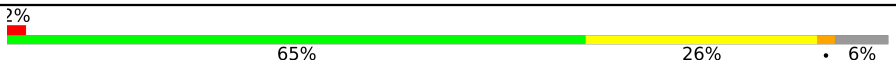
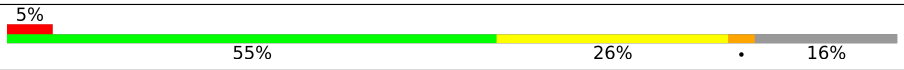
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	1149 (3.34-3.26)
Clashscore	141614	1205 (3.34-3.26)
Ramachandran outliers	138981	1183 (3.34-3.26)
Sidechain outliers	138945	1182 (3.34-3.26)
RSRZ outliers	127900	1115 (3.34-3.26)

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	405	4% 54% 28% · 14%
1	B	405	3% 62% 28% · 6%
1	C	405	2% 61% 30% · 6%
1	D	405	% 62% 27% · 6%
1	E	405	2% 63% 27% · 6%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	F	405	
1	G	405	
1	H	405	

2 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 23689 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 CG3027-PA.

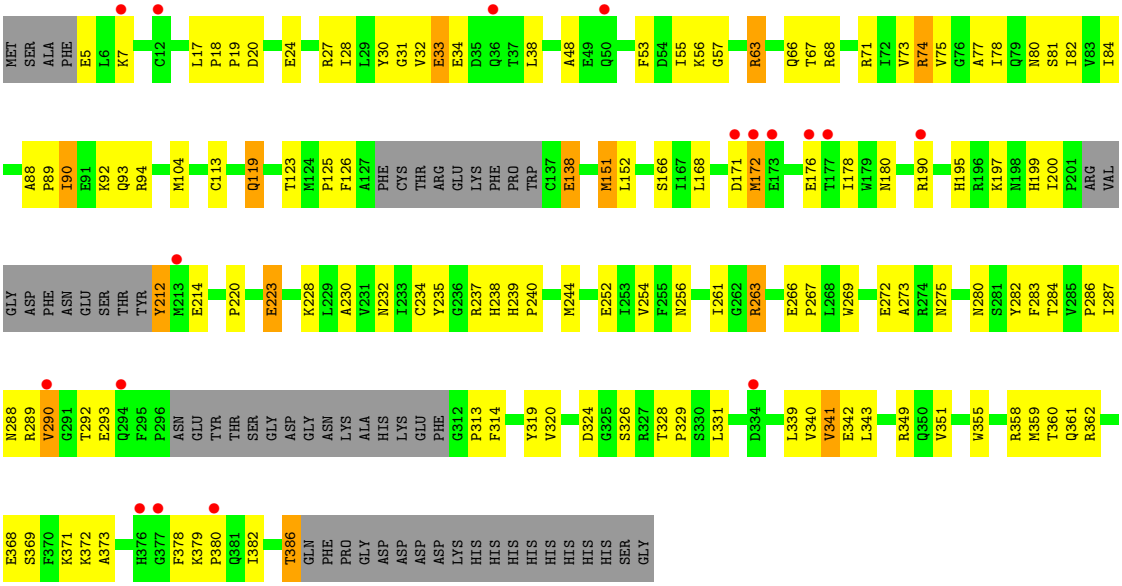
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	348	Total	C	N	O	S	0	1	0
			2774	1756	487	516	15			
1	B	379	Total	C	N	O	S	0	0	0
			3027	1916	530	565	16			
1	C	381	Total	C	N	O	S	0	0	0
			3044	1928	533	567	16			
1	D	379	Total	C	N	O	S	0	0	0
			3027	1916	530	565	16			
1	E	380	Total	C	N	O	S	0	0	0
			3036	1922	532	566	16			
1	F	379	Total	C	N	O	S	0	0	0
			3027	1916	530	565	16			
1	G	381	Total	C	N	O	S	0	0	0
			3044	1928	533	567	16			
1	H	342	Total	C	N	O	S	0	0	0
			2710	1712	478	505	15			

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 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.

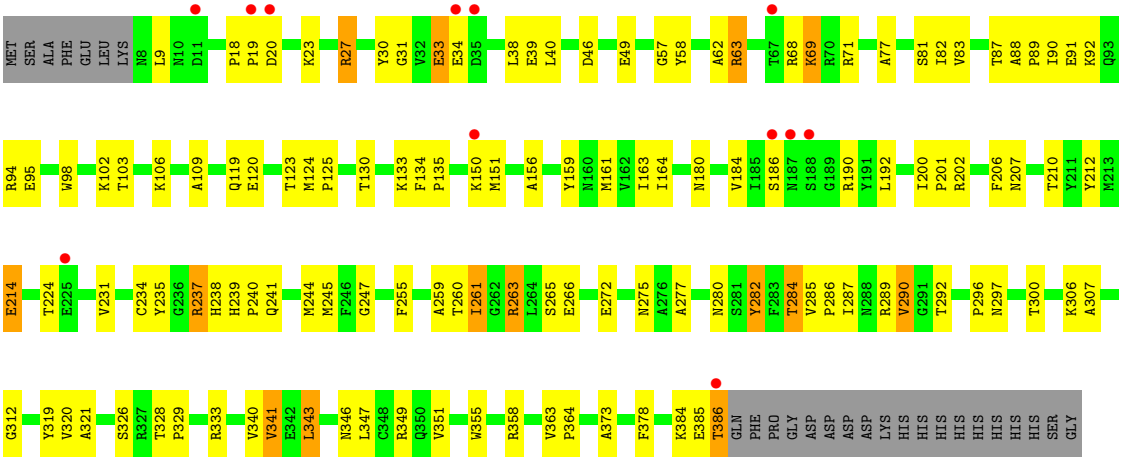
- Molecule 1: CG3027-PA

Chain A:



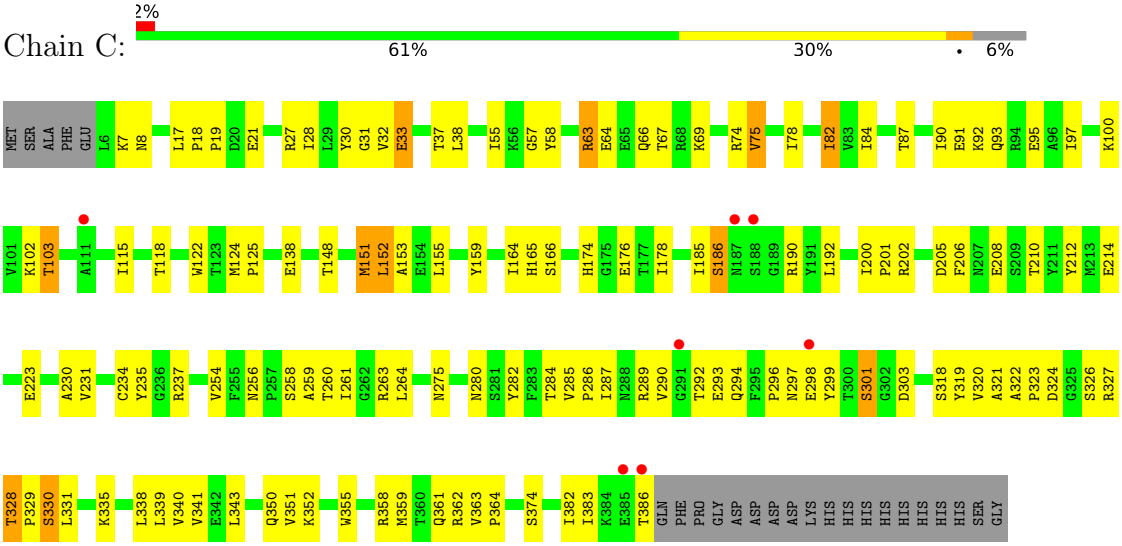
- Molecule 1: CG3027-PA

Chain B:



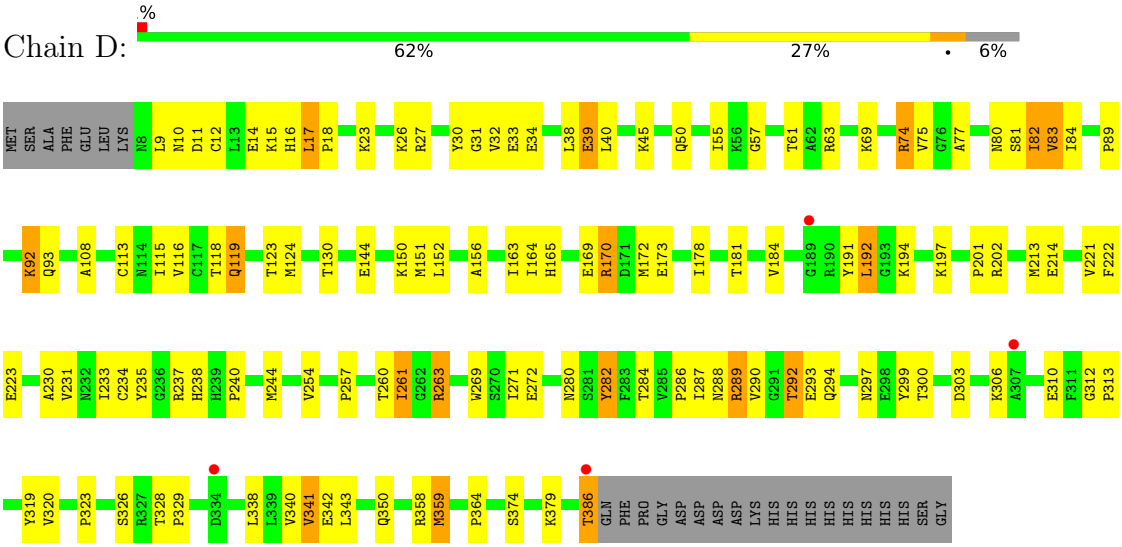
• Molecule 1: CG3027-PA

Chain C:



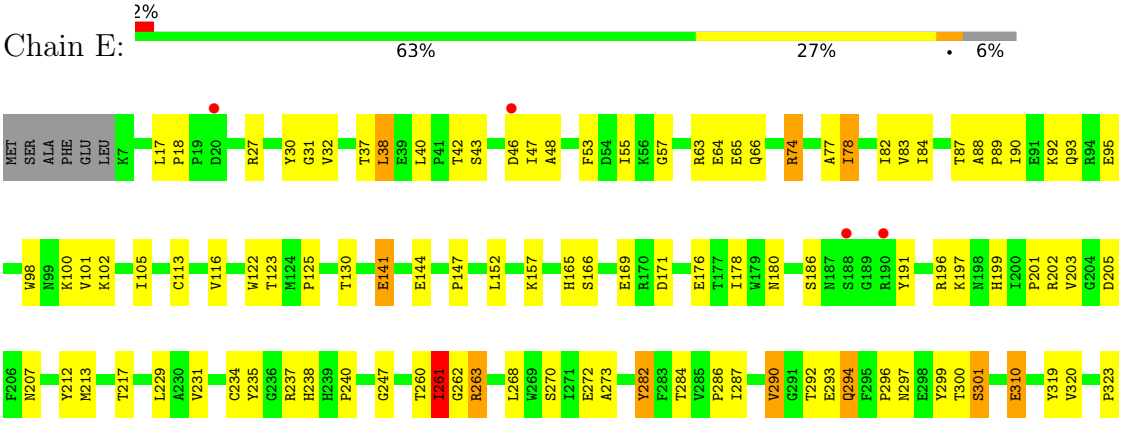
• Molecule 1: CG3027-PA

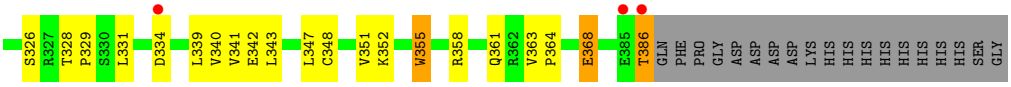
Chain D:



• Molecule 1: CG3027-PA

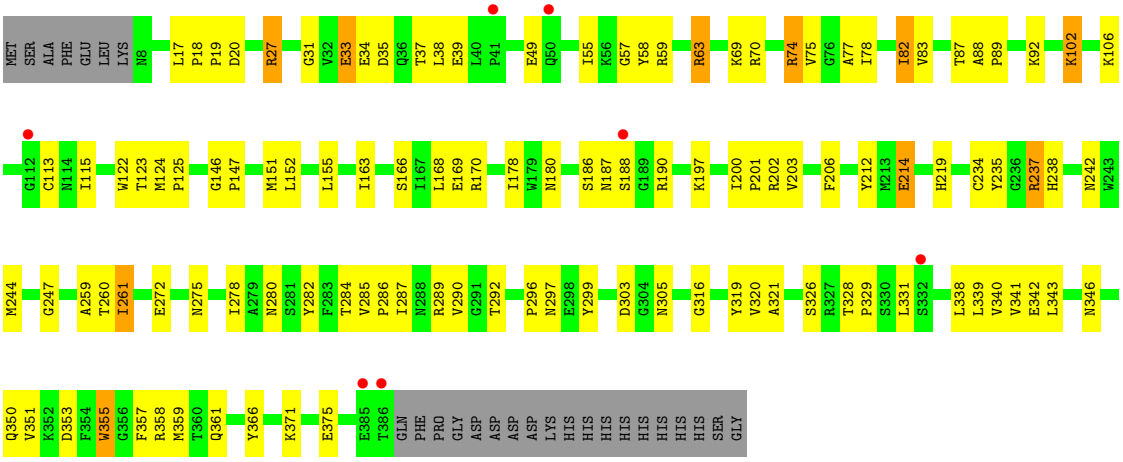
Chain E:





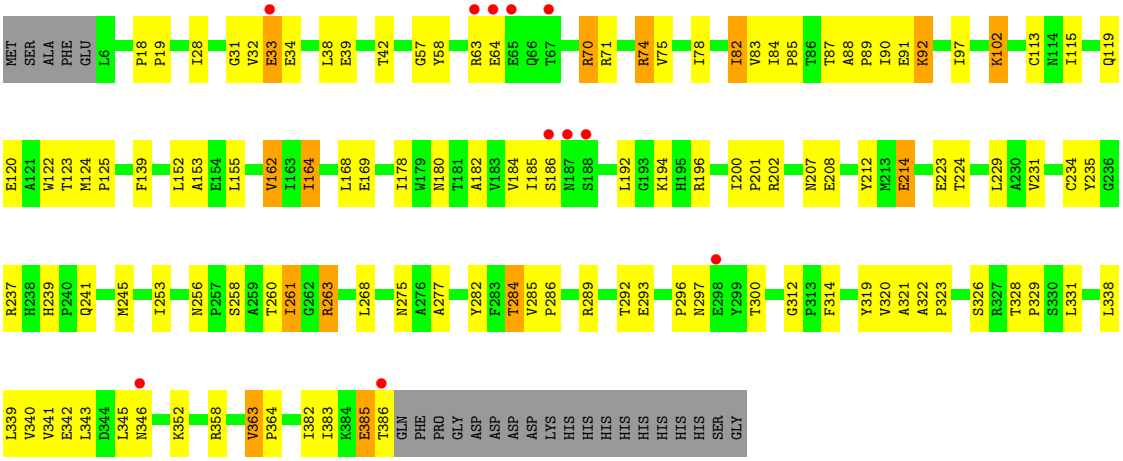
• Molecule 1: CG3027-PA

Chain F:



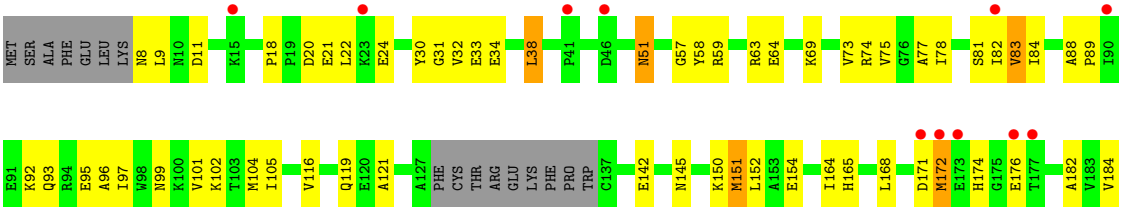
• Molecule 1: CG3027-PA

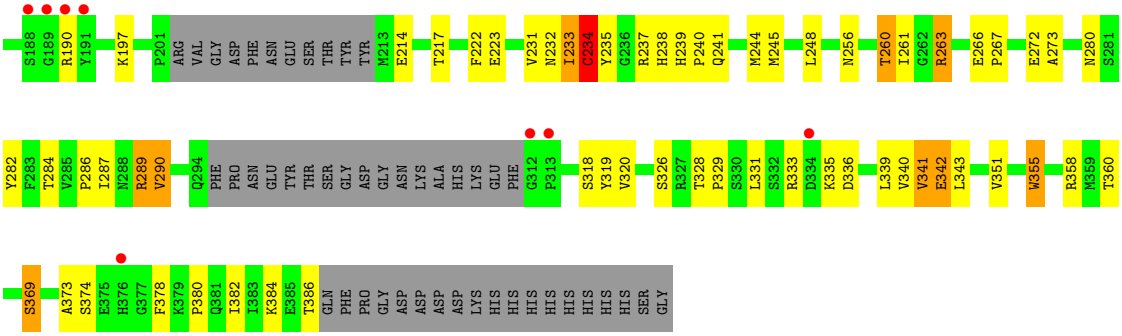
Chain G:



• Molecule 1: CG3027-PA

Chain H:




S188 G189 R190 Y191 K197 P201 ARG VAL GLY ASP PHE ASN GLU SER THR TYR TYR M213 E214 T217 F222 E223 V231 N232 I233 C234 Y235 G236 R237 H238 H239 P240 Q241 M244 M245 L248 N256 T260 I261 G262 R263 E266 P267 E272 A273 N280 S281
Y282 F283 T284 V285 P286 I287 N288 R289 V290 Q294 PHE PRO ASN GLU TYR THR SER GLY ASP GLY ASN LYS ALA HIS LYS GLU PHE G312 P313 S318 Y319 V320 S326 R327 T328 P329 S330 L331 S332 R333 D334 K335 D336 L339 V340 V341 E342 L343 V351 W355 R358 M359 T360
S369 A373 S374 E375 H376 G377 F378 K379 P380 Q381 I382 I383 K384 E385 T386 GLN PHE PRO GLY ASP ASP ASP ASP LYS HIS HIS HIS HIS HIS HIS HIS HIS HIS SER GLY

4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α, β, γ	278.86Å 95.05Å 199.31Å 90.00° 125.82° 90.00°	Depositor
Resolution (Å)	30.00 – 3.30 49.20 – 3.30	Depositor EDS
% Data completeness (in resolution range)	85.6 (30.00-3.30) 85.6 (49.20-3.30)	Depositor EDS
R_{merge}	0.17	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.17 (at 3.33Å)	Xtriage
Refinement program	REFMAC 5.2.0019	Depositor
R, R_{free}	0.226 , 0.282 0.229 , 0.281	Depositor DCC
R_{free} test set	2766 reflections (5.05%)	wwPDB-VP
Wilson B-factor (Å^2)	45.8	Xtriage
Anisotropy	0.162	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.35 , 34.0	EDS
L-test for twinning[2]	$< \|L\| > = 0.42$, $< L^2 > = 0.24$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.86	EDS
Total number of atoms	23689	wwPDB-VP
Average B, all atoms (Å^2)	41.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 15.82% 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	0.49	0/2835	0.63	0/3833
1	B	0.47	0/3101	0.61	0/4197
1	C	0.47	0/3118	0.67	0/4219
1	D	0.47	0/3101	0.61	1/4197 (0.0%)
1	E	0.48	0/3110	0.65	0/4208
1	F	0.46	0/3101	0.62	0/4197
1	G	0.51	0/3118	0.65	0/4219
1	H	0.50	0/2768	0.62	0/3742
All	All	0.48	0/24252	0.63	1/32812 (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	B	0	1
1	C	0	2
1	D	0	2
1	E	0	1
1	F	0	1
1	G	0	2
1	H	0	1
All	All	0	11

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	D	17	LEU	CA-CB-CG	5.32	127.54	115.30

There are no chirality outliers.

5 of 11 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	261	ILE	Peptide
1	B	261	ILE	Peptide
1	C	261	ILE	Peptide
1	C	82	ILE	Peptide
1	D	82	ILE	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	2774	0	2738	85	0
1	B	3027	0	2958	90	0
1	C	3044	0	2982	94	0
1	D	3027	0	2958	86	0
1	E	3036	0	2971	84	0
1	F	3027	0	2958	70	0
1	G	3044	0	2982	80	0
1	H	2710	0	2678	89	0
All	All	23689	0	23225	614	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 13.

The worst 5 of 614 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:63:ARG:HG3	1:C:350:GLN:OE1	1.58	1.04
1:B:63:ARG:HB2	1:B:63:ARG:HH11	1.17	1.04
1:H:197:LYS:HA	1:H:233:ILE:HD13	1.42	1.01
1:G:87:THR:HG22	1:G:296:PRO:HG3	1.43	0.97
1:E:66:GLN:HE22	1:G:90:ILE:H	0.98	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	341/405 (84%)	317 (93%)	24 (7%)	0	100 100
1	B	377/405 (93%)	363 (96%)	14 (4%)	0	100 100
1	C	379/405 (94%)	362 (96%)	17 (4%)	0	100 100
1	D	377/405 (93%)	362 (96%)	14 (4%)	1 (0%)	41 71
1	E	378/405 (93%)	360 (95%)	18 (5%)	0	100 100
1	F	377/405 (93%)	360 (96%)	17 (4%)	0	100 100
1	G	379/405 (94%)	362 (96%)	16 (4%)	1 (0%)	41 71
1	H	334/405 (82%)	313 (94%)	20 (6%)	1 (0%)	41 71
All	All	2942/3240 (91%)	2799 (95%)	140 (5%)	3 (0%)	51 81

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	H	234	CYS
1	D	83	VAL
1	G	83	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	297/346 (86%)	263 (89%)	34 (11%)	5 22
1	B	323/346 (93%)	289 (90%)	34 (10%)	7 25

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
1	C	325/346 (94%)	289 (89%)	36 (11%)	6 23
1	D	323/346 (93%)	290 (90%)	33 (10%)	7 27
1	E	324/346 (94%)	288 (89%)	36 (11%)	6 23
1	F	323/346 (93%)	291 (90%)	32 (10%)	8 28
1	G	325/346 (94%)	292 (90%)	33 (10%)	7 27
1	H	290/346 (84%)	251 (87%)	39 (13%)	4 16
All	All	2530/2768 (91%)	2253 (89%)	277 (11%)	6 24

5 of 277 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	G	282	TYR
1	H	20	ASP
1	H	233	ILE
1	C	301	SER
1	C	292	THR

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 37 such sidechains are listed below:

Mol	Chain	Res	Type
1	F	239	HIS
1	H	242	ASN
1	G	239	HIS
1	H	8	ASN
1	B	297	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)

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 (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	348/405 (85%)	0.15	17 (4%)	29	27	23, 40, 65, 70	0
1	B	379/405 (93%)	0.03	12 (3%)	47	46	23, 40, 62, 70	0
1	C	381/405 (94%)	0.03	7 (1%)	68	67	23, 40, 63, 70	0
1	D	379/405 (93%)	-0.02	4 (1%)	80	81	23, 40, 62, 70	0
1	E	380/405 (93%)	-0.01	7 (1%)	68	67	23, 40, 63, 70	0
1	F	379/405 (93%)	0.03	7 (1%)	68	67	23, 40, 62, 70	0
1	G	381/405 (94%)	0.02	11 (2%)	51	50	23, 40, 63, 70	0
1	H	342/405 (84%)	0.24	19 (5%)	24	23	23, 39, 63, 70	0
All	All	2969/3240 (91%)	0.06	84 (2%)	53	51	23, 40, 64, 70	0

The worst 5 of 84 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	H	376	HIS	4.1
1	G	67	THR	3.9
1	E	334	ASP	3.8
1	F	386	THR	3.8
1	G	187	ASN	3.7

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

There are no ligands in this entry.

6.5 Other polymers

There are no such residues in this entry.