wwPDB X-ray Structure Validation Summary Report (i)

Sep 25, 2023 – 05:23 AM EDT

PDB ID	:	5VYJ
Title	:	Crystal structure of the photosynthetic phosphoenolpyruvate carboxylase isoenzyme from maize in complex with Gly
Authors	:	Gonzalez-Segura, L.; Guemez-Toro, R.; Munoz-Clares, R.A.
Deposited on	:	2017-05-25
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:

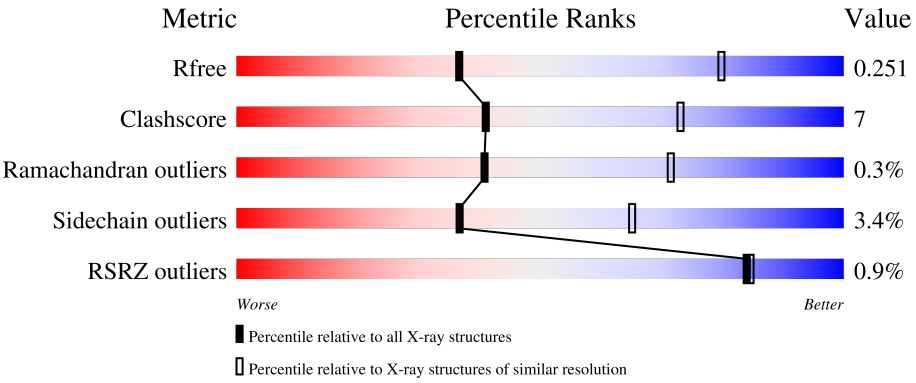
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.35.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.35.1

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	970	73% 20% • 5%
1	B	970	74% 20% • 5%
1	C	970	75% 19% • 5%
1	D	970	76% 18% • 5%

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 criteria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
2	GLY	B	1001	-	X	-	-

2 Entry composition (i)

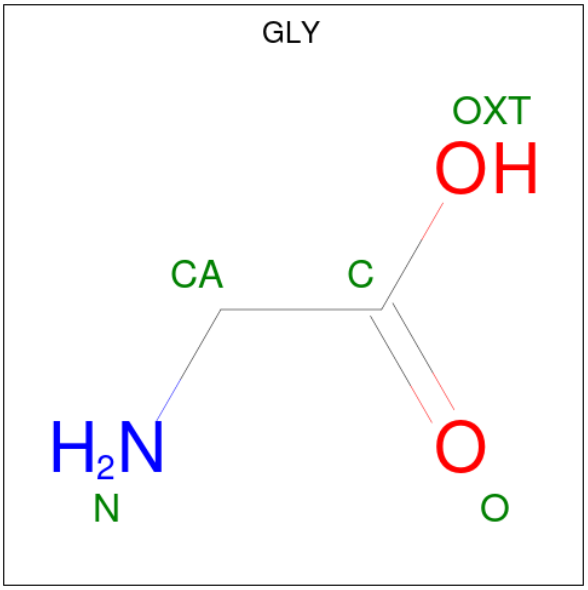
There are 3 unique types of molecules in this entry. The entry contains 29339 atoms, of which 0 are hydrogens and 0 are deuteriums.

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 Phosphoenolpyruvate carboxylase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	918	Total	C	N	O	S	0	0	0
			7318	4646	1275	1367	30			
1	B	917	Total	C	N	O	S	0	0	0
			7309	4640	1274	1365	30			
1	C	918	Total	C	N	O	S	0	0	0
			7318	4645	1275	1368	30			
1	D	918	Total	C	N	O	S	0	0	0
			7317	4646	1275	1366	30			

- Molecule 2 is GLYCINE (three-letter code: GLY) (formula: $C_2H_5NO_2$).



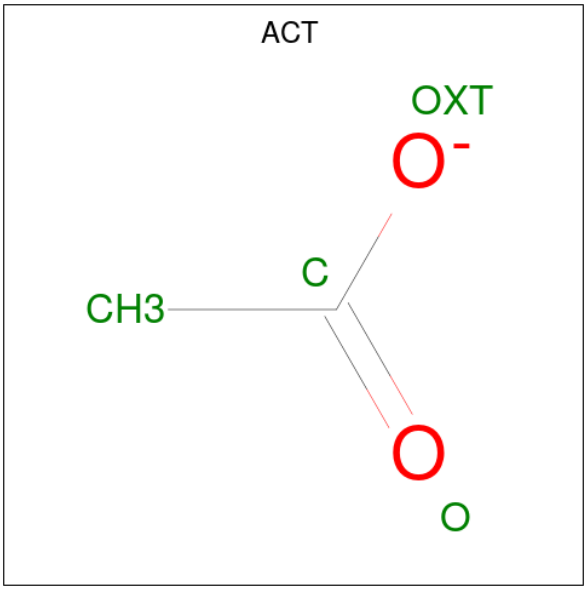
Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
2	A	1	Total	C	N	O	0	0
			5	2	1	2		
2	B	1	Total	C	N	O	0	0
			5	2	1	2		

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	B	1	Total C N O 5 2 1 2	0	0
2	C	1	Total C N O 5 2 1 2	0	0
2	D	1	Total C N O 5 2 1 2	0	0

- Molecule 3 is ACETATE ION (three-letter code: ACT) (formula: $C_2H_3O_2$).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	1	Total C O 4 2 2	0	0
3	A	1	Total C O 4 2 2	0	0
3	A	1	Total C O 4 2 2	0	0
3	B	1	Total C O 4 2 2	0	0
3	B	1	Total C O 4 2 2	0	0
3	B	1	Total C O 4 2 2	0	0
3	C	1	Total C O 4 2 2	0	0
3	C	1	Total C O 4 2 2	0	0

Continued on next page...

Continued from previous page...

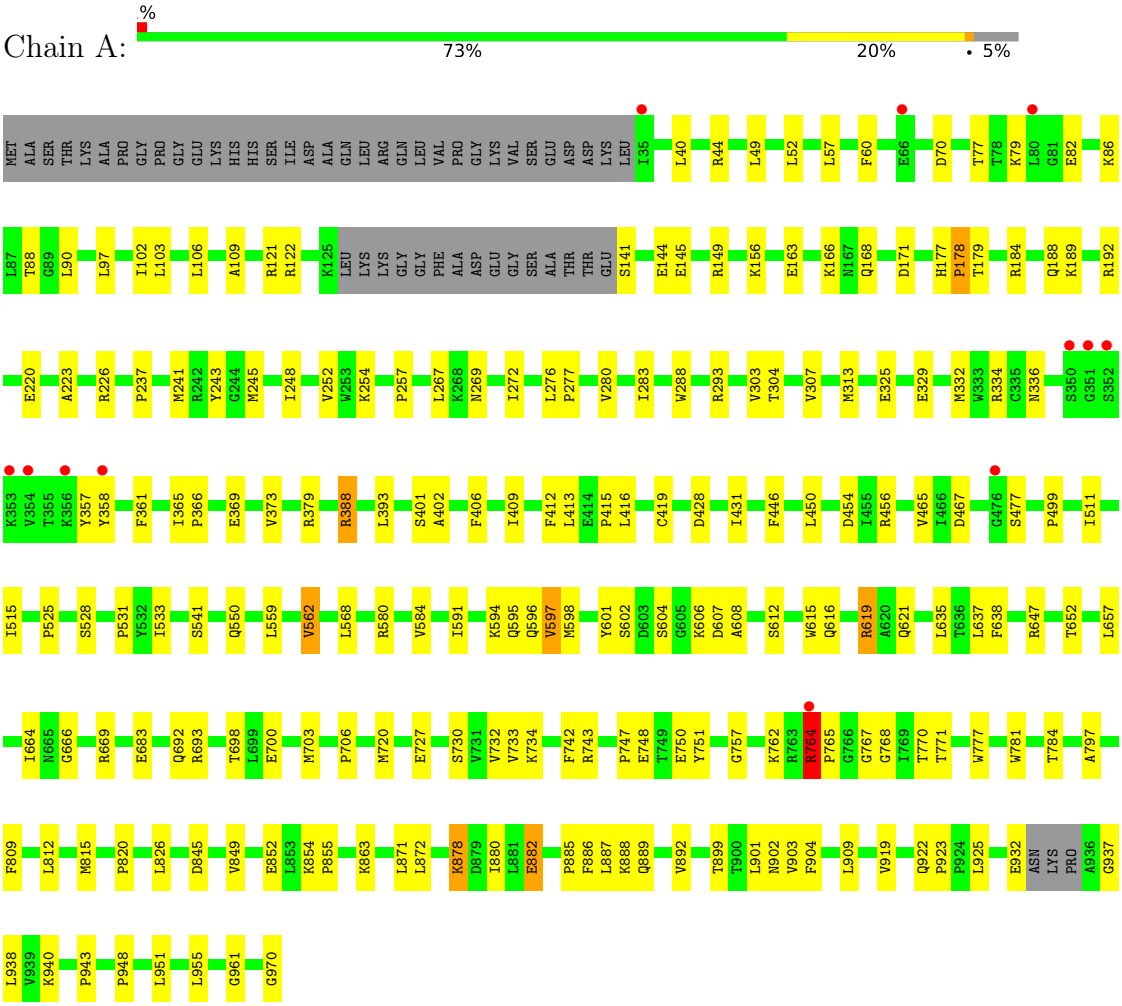
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
3	C	1	Total	C	O	0	0
			4	2	2		
3	C	1	Total	C	O	0	0
			4	2	2		
3	D	1	Total	C	O	0	0
			4	2	2		
3	D	1	Total	C	O	0	0
			4	2	2		
3	D	1	Total	C	O	0	0
			4	2	2		

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.

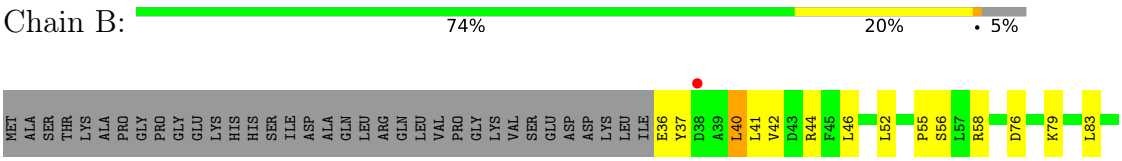
• Molecule 1: Phosphoenolpyruvate carboxylase

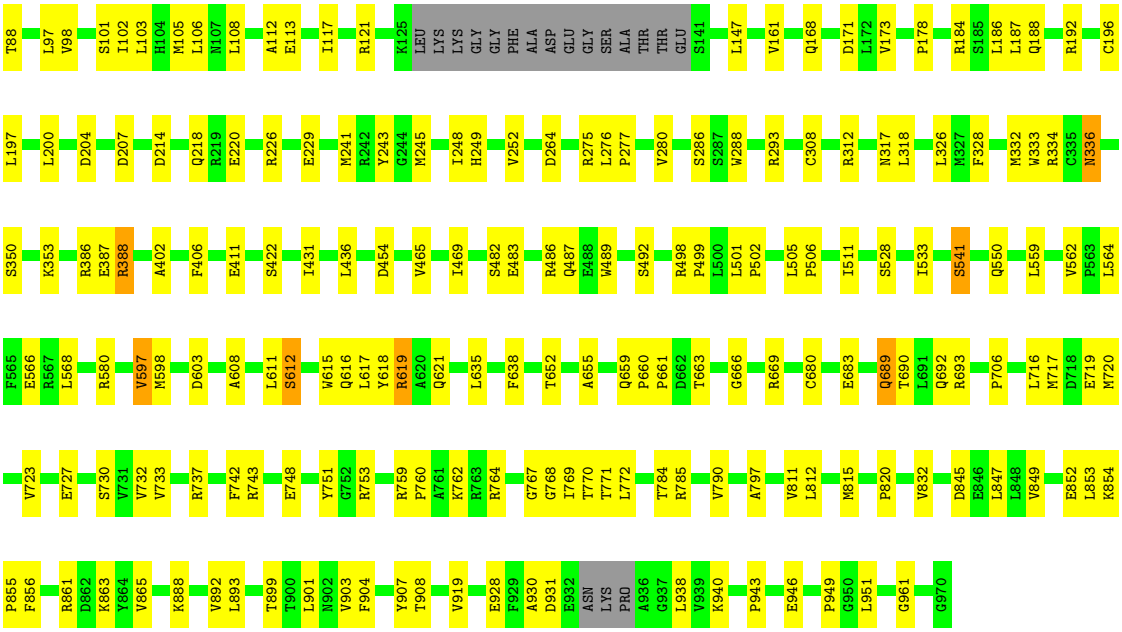
Chain A:



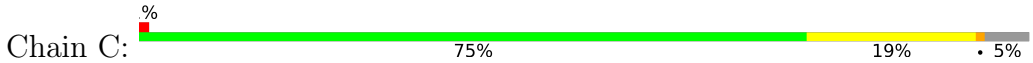
• Molecule 1: Phosphoenolpyruvate carboxylase

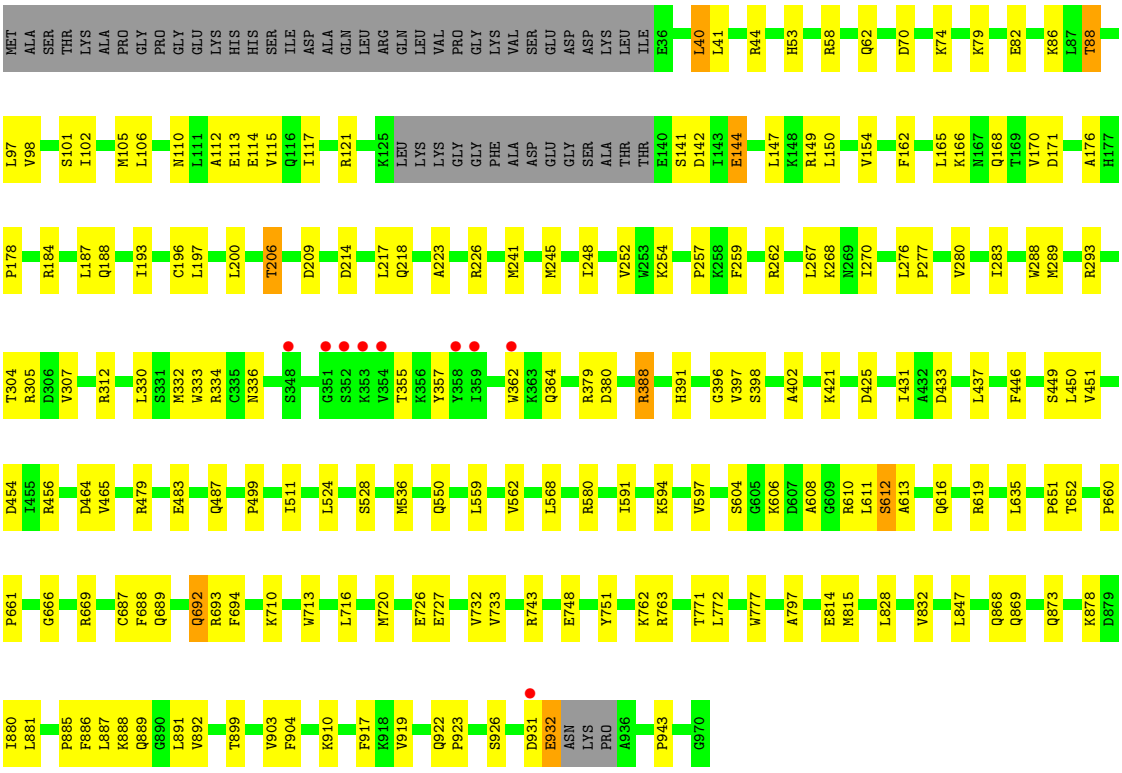
Chain B:



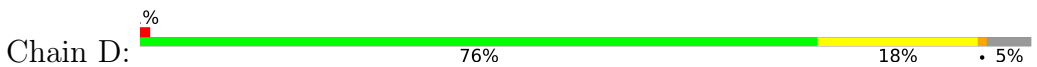


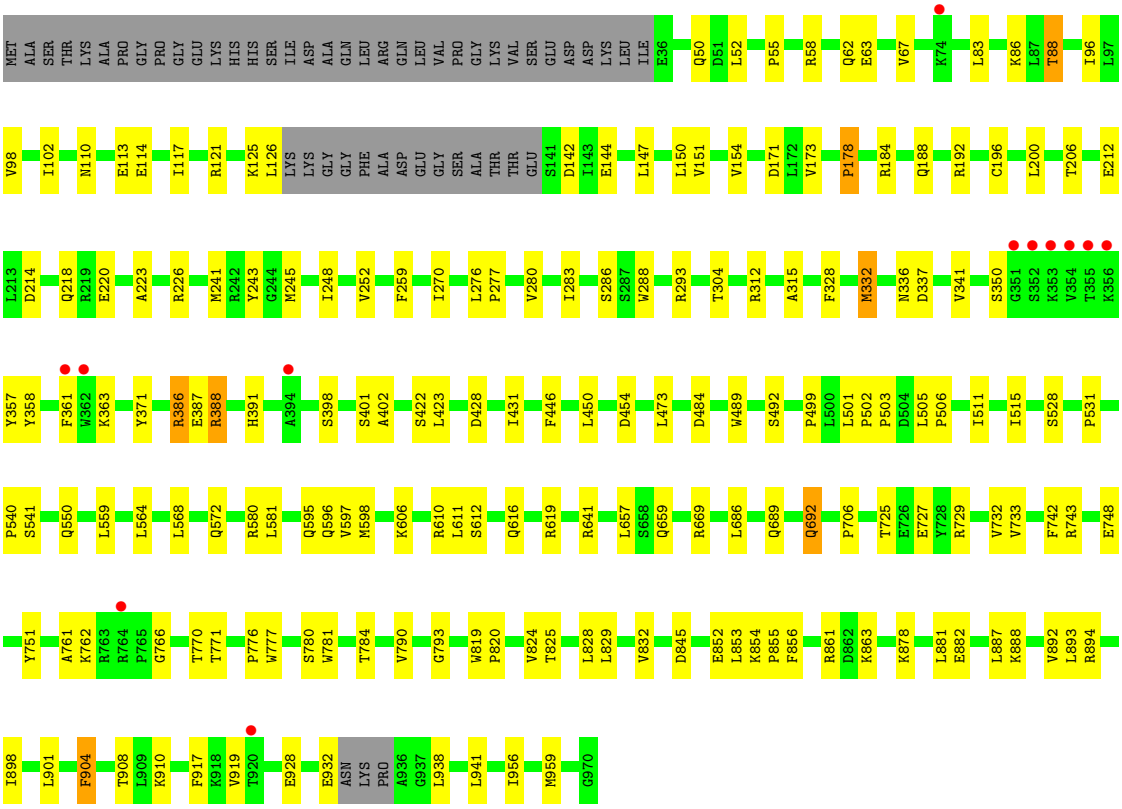
• Molecule 1: Phosphoenolpyruvate carboxylase





• Molecule 1: Phosphoenolpyruvate carboxylase




MET ALA SER THR LYS ALA PRO GLY PRO GLY GLU LYS HIS SER ILE ASP ALA GLN LEU ARG GLN LEU VAL PRO GLY LYS VAL SER GLU ASP ASP LYS LEU ILE E36 Q50 D51 L52 P55 R58 Q62 E63 V67 K74 L83 K86 L87 T88 I96 L97
V98 I102 N110 E113 E114 I117 R121 K125 L126 LYS LYS GLY GLY PHE ALA ASP GLU GLY GLY SER ALA THR THR GLU S141 D142 I143 E144 L147 L150 V151 V154 D171 L172 V173 P178 R184 Q188 R192 C196 L200 T206 E212
L213 D214 Q218 R219 E220 A223 R226 M241 R242 Y243 G244 M245 I248 V252 F259 I270 L276 P277 V280 I283 S286 S287 W288 R293 T304 R312 A315 F328 M332 N336 D337 V341 S350 G351 S352 K353 V354 T355 K356
Y357 Y358 F361 W362 K363 Y371 R386 E387 R388 H391 A394 S398 S401 A402 S422 L423 D428 I431 F446 L450 D464 L473 D484 W489 S492 P499 L500 L501 P502 P503 D504 L505 P506 I511 I515 S528 P531
P540 S541 Q550 L559 L564 L568 Q572 R580 L581 Q595 Q596 V597 M598 K606 R610 L611 S612 Q616 R619 R641 L657 S658 Q659 R669 L686 Q689 Q692 P706 T725 E726 E727 Y728 R729 V732 V733 F742 R743 E748
Y751 A761 K762 R763 R764 P765 G766 T770 T771 P776 W777 S780 W781 T784 V790 G793 W819 P820 V824 T825 L828 L829 V832 D845 E852 L853 K854 P855 F856 R861 D862 K863 K878 L881 E882 L887 K888 V892 L893 R894
I898 L901 F904 T908 L909 K910 F917 K918 V919 T920 E928 E932 ASN LYS PRO A936 G937 L938 L941 I956 M959 G970

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	156.95Å 167.24Å 242.90Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	57.08 – 3.30 57.08 – 3.30	Depositor EDS
% Data completeness (in resolution range)	93.2 (57.08-3.30) 93.3 (57.08-3.30)	Depositor EDS
R_{merge}	0.12	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.87 (at 3.33Å)	Xtriage
Refinement program	PHENIX 1.8.1_1168	Depositor
R, R_{free}	0.210 , 0.253 0.212 , 0.251	Depositor DCC
R_{free} test set	4518 reflections (5.01%)	wwPDB-VP
Wilson B-factor (Å^2)	45.0	Xtriage
Anisotropy	0.079	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.34 , 29.3	EDS
L-test for twinning[2]	$< \|L\| > = 0.46$, $< L^2 > = 0.29$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.91	EDS
Total number of atoms	29339	wwPDB-VP
Average B, all atoms (Å^2)	27.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 38.98 % of the origin peak, indicating pseudo-translational symmetry. The chance of finding a peak of this or larger height randomly in a structure without pseudo-translational symmetry is equal to 3.4006e-04. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

[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: ACT

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	0.34	0/7473	0.51	0/10119
1	B	0.35	0/7464	0.53	0/10108
1	C	0.33	0/7473	0.50	0/10120
1	D	0.33	0/7472	0.50	0/10119
All	All	0.34	0/29882	0.51	0/40466

There are no bond length outliers.

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	7318	0	7314	114	0
1	B	7309	0	7303	114	0
1	C	7318	0	7309	100	0
1	D	7317	0	7314	98	0
2	A	5	0	2	1	0
2	B	10	0	4	1	0
2	C	5	0	2	1	0
2	D	5	0	2	0	0
3	A	12	0	9	2	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
3	B	12	0	9	0	0
3	C	16	0	12	0	0
3	D	12	0	9	1	0
All	All	29339	0	29289	414	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 7.

The worst 5 of 414 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:B:388:ARG:HH12	1:B:402:ALA:HA	1.39	0.88
1:D:743:ARG:HH22	1:D:762:LYS:HE3	1.43	0.82
1:B:386:ARG:NH1	1:B:387:GLU:OE1	2.17	0.77
1:C:362:TRP:NE1	1:C:380:ASP:OD2	2.19	0.75
1:B:568:LEU:HD22	1:B:608:ALA:HB2	1.67	0.74

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	912/970 (94%)	855 (94%)	53 (6%)	4 (0%)	34	66
1	B	911/970 (94%)	863 (95%)	44 (5%)	4 (0%)	34	66
1	C	912/970 (94%)	867 (95%)	42 (5%)	3 (0%)	41	71
1	D	912/970 (94%)	857 (94%)	54 (6%)	1 (0%)	51	81
All	All	3647/3880 (94%)	3442 (94%)	193 (5%)	12 (0%)	41	71

5 of 12 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	C	943	PRO
1	B	943	PRO
1	C	330	LEU
1	A	764	ARG
1	A	943	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	791/832 (95%)	768 (97%)	23 (3%)	42	69
1	B	790/832 (95%)	762 (96%)	28 (4%)	36	64
1	C	791/832 (95%)	761 (96%)	30 (4%)	33	62
1	D	791/832 (95%)	763 (96%)	28 (4%)	36	64
All	All	3163/3328 (95%)	3054 (97%)	109 (3%)	37	65

5 of 109 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	144	GLU
1	C	687	CYS
1	D	612	SER
1	C	283	ILE
1	C	397	VAL

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 8 such sidechains are listed below:

Mol	Chain	Res	Type
1	D	692	GLN
1	D	659	GLN
1	D	53	HIS
1	C	211	GLN
1	D	336	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)

18 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	GLY	C	1001	-	4,4,4	1.02	0	3,4,4	1.53	0
3	ACT	C	1002	-	3,3,3	0.80	0	3,3,3	1.30	0
2	GLY	B	1001	-	4,4,4	1.11	1 (25%)	3,4,4	2.14	2 (66%)
3	ACT	B	1004	-	3,3,3	0.78	0	3,3,3	1.35	0
3	ACT	C	1003	-	3,3,3	0.77	0	3,3,3	1.38	0
3	ACT	A	1004	-	3,3,3	0.78	0	3,3,3	1.63	1 (33%)
3	ACT	D	1003	-	3,3,3	0.79	0	3,3,3	1.41	0
3	ACT	B	1005	-	3,3,3	0.73	0	3,3,3	1.60	1 (33%)
3	ACT	D	1002	-	3,3,3	0.79	0	3,3,3	1.62	1 (33%)
3	ACT	C	1005	-	3,3,3	0.82	0	3,3,3	1.36	0
3	ACT	A	1002	-	3,3,3	0.76	0	3,3,3	1.38	0
2	GLY	D	1001	-	4,4,4	0.91	0	3,4,4	1.70	1 (33%)
3	ACT	C	1004	-	3,3,3	0.76	0	3,3,3	1.53	0
2	GLY	A	1001	-	4,4,4	1.13	1 (25%)	3,4,4	1.63	1 (33%)
3	ACT	B	1003	-	3,3,3	0.77	0	3,3,3	1.58	0
3	ACT	D	1004	-	3,3,3	0.76	0	3,3,3	1.17	0

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\#\|Z\|>2$	Counts	RMSZ	$\#\|Z\|>2$
3	ACT	A	1003	-	3,3,3	0.72	0	3,3,3	1.45	0
2	GLY	B	1002	-	4,4,4	1.06	0	3,4,4	1.45	0

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	GLY	C	1001	-	-	0/2/2/2	-
2	GLY	B	1001	-	-	2/2/2/2	-
2	GLY	D	1001	-	-	0/2/2/2	-
2	GLY	A	1001	-	-	0/2/2/2	-
2	GLY	B	1002	-	-	2/2/2/2	-

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	B	1001	GLY	OXT-C	-2.11	1.23	1.30
2	A	1001	GLY	OXT-C	-2.06	1.23	1.30

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

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	B	1001	GLY	OXT-C-O	-3.01	115.79	123.30
2	A	1001	GLY	OXT-C-O	-2.48	117.11	123.30
2	D	1001	GLY	OXT-C-CA	2.25	122.38	113.45
2	B	1001	GLY	OXT-C-CA	2.17	122.07	113.45
3	A	1004	ACT	OXT-C-O	-2.11	114.28	122.05

There are no chirality outliers.

All (4) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	B	1001	GLY	O-C-CA-N
2	B	1001	GLY	OXT-C-CA-N
2	B	1002	GLY	OXT-C-CA-N
2	B	1002	GLY	O-C-CA-N

There are no ring outliers.

6 monomers are involved in 6 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	C	1001	GLY	1	0
3	A	1004	ACT	1	0
3	A	1002	ACT	1	0
2	A	1001	GLY	1	0
3	D	1004	ACT	1	0
2	B	1002	GLY	1	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	918/970 (94%)	-0.10	12 (1%)	77	77	3, 24, 57, 85	0
1	B	917/970 (94%)	-0.23	1 (0%)	95	97	0, 17, 50, 88	0
1	C	918/970 (94%)	-0.13	9 (0%)	82	82	8, 24, 60, 100	0
1	D	918/970 (94%)	-0.04	12 (1%)	77	77	2, 30, 74, 97	0
All	All	3671/3880 (94%)	-0.12	34 (0%)	84	84	0, 23, 63, 100	0

The worst 5 of 34 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	351	GLY	4.5
1	D	352	SER	4.3
1	C	352	SER	3.9
1	D	351	GLY	3.7
1	D	361	PHE	3.5

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, 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	B-factors(Å^2)	Q<0.9
3	ACT	C	1003	4/4	0.83	0.22	30,30,30,30	0
3	ACT	C	1005	4/4	0.84	0.27	37,37,37,37	0
3	ACT	D	1002	4/4	0.86	0.21	21,21,21,21	0
3	ACT	A	1003	4/4	0.89	0.24	40,40,40,40	0
2	GLY	B	1001	5/5	0.89	0.23	22,22,22,22	0
3	ACT	D	1003	4/4	0.90	0.28	39,39,39,39	0
3	ACT	D	1004	4/4	0.91	0.24	22,22,22,22	0
2	GLY	B	1002	5/5	0.92	0.19	8,8,8,8	0
3	ACT	B	1003	4/4	0.92	0.23	11,11,11,11	0
3	ACT	B	1005	4/4	0.92	0.21	11,11,11,11	0
2	GLY	C	1001	5/5	0.92	0.22	18,18,18,18	0
2	GLY	D	1001	5/5	0.93	0.18	8,8,8,8	0
3	ACT	C	1002	4/4	0.94	0.28	17,17,17,17	0
3	ACT	A	1002	4/4	0.94	0.18	16,16,16,16	0
2	GLY	A	1001	5/5	0.94	0.20	9,9,9,9	0
3	ACT	A	1004	4/4	0.95	0.14	20,20,20,20	0
3	ACT	C	1004	4/4	0.95	0.22	17,17,17,17	0
3	ACT	B	1004	4/4	0.97	0.16	10,10,10,10	0

6.5 Other polymers (i)

There are no such residues in this entry.