W O R L D W I D E

wwPDB

PROTEIN DATA BANK

wwPDB X-ray Structure Validation Summary Report (i)

Jun 23, 2024 – 12:24 AM EDT

PDB ID	:	5HMQ
Title	:	xylose isomerase-like TIM barrel/4-hydroxyphenylpyruvate dioxygenase fusion protein
Authors	:	Peek, J.; Christendat, D.
Deposited on	:	2016-01-16
Resolution	:	2.37 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:

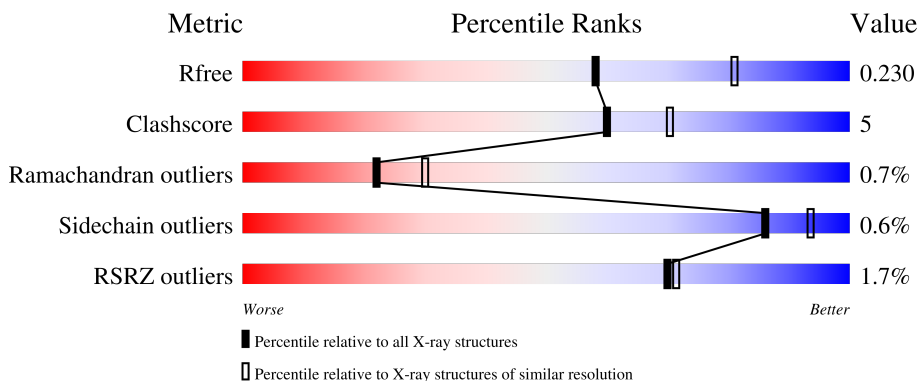|		
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.37.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.37.1

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 2.37 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		0.230
Clashscore		5
Ramachandran outliers		0.7%
Sidechain outliers		0.6%
RSRZ outliers		1.7%

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	5509 (2.40-2.36)
Clashscore	141614	6082 (2.40-2.36)
Ramachandran outliers	138981	5973 (2.40-2.36)
Sidechain outliers	138945	5975 (2.40-2.36)
RSRZ outliers	127900	5397 (2.40-2.36)

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	637	2% 84% 13% •
1	B	637	2% 90% 8% •
1	C	637	% 86% 11% •
1	D	637	2% 87% 10% ••
1	E	637	% 85% 12% •

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	F	637	

2 Entry composition ⓘ

There are 3 unique types of molecules in this entry. The entry contains 30198 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 4-hydroxyphenylpyruvate dioxygenase.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	621	Total	C	N	O	S	Se	0	0	0
			4850	3080	853	900	8	9			
1	B	621	Total	C	N	O	S	Se	0	0	0
			4854	3083	853	901	8	9			
1	C	622	Total	C	N	O	S	Se	0	0	0
			4824	3058	848	901	8	9			
1	D	624	Total	C	N	O	S	Se	0	0	0
			4869	3094	855	903	8	9			
1	E	621	Total	C	N	O	S	Se	0	0	0
			4845	3075	852	901	8	9			
1	F	623	Total	C	N	O	S	Se	0	0	0
			4860	3087	849	907	8	9			

There are 12 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-1	ALA	-	expression tag	UNP Q88JU3
A	0	LYS	-	expression tag	UNP Q88JU3
B	-1	ALA	-	expression tag	UNP Q88JU3
B	0	LYS	-	expression tag	UNP Q88JU3
C	-1	ALA	-	expression tag	UNP Q88JU3
C	0	LYS	-	expression tag	UNP Q88JU3
D	-1	ALA	-	expression tag	UNP Q88JU3
D	0	LYS	-	expression tag	UNP Q88JU3
E	-1	ALA	-	expression tag	UNP Q88JU3
E	0	LYS	-	expression tag	UNP Q88JU3
F	-1	ALA	-	expression tag	UNP Q88JU3
F	0	LYS	-	expression tag	UNP Q88JU3

- Molecule 2 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	2	Total 2	Mg 2	0	0
2	B	2	Total 2	Mg 2	0	0
2	C	2	Total 2	Mg 2	0	0
2	D	2	Total 2	Mg 2	0	0
2	E	2	Total 2	Mg 2	0	0
2	F	2	Total 2	Mg 2	0	0

- Molecule 3 is water.

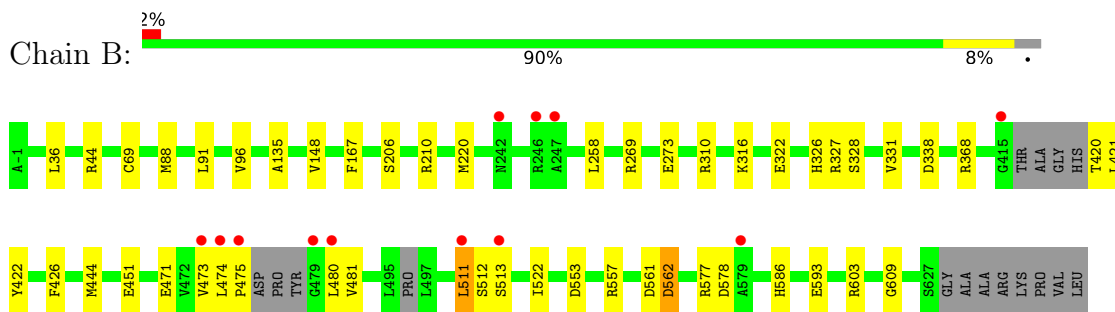| Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	140	Total 140	O 140	0	0
3	B	195	Total 195	O 195	0	0
3	C	128	Total 128	O 128	0	0
3	D	231	Total 231	O 231	0	0
3	E	236	Total 236	O 236	0	0
3	F	154	Total 154	O 154	0	0

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.
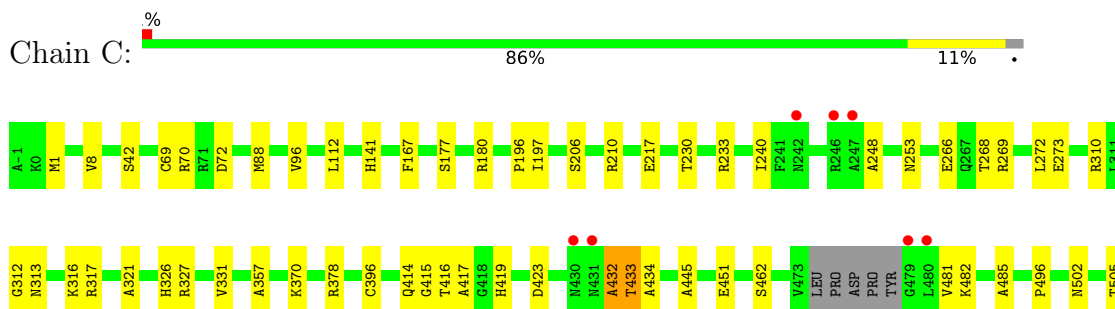
- Molecule 1: 4-hydroxyphenylpyruvate dioxygenase

- Molecule 1: 4-hydroxyphenylpyruvate dioxygenase

- Molecule 1: 4-hydroxyphenylpyruvate dioxygenase

A506 I507 A508 H509 A510 L511 S512 G516 S517 L543 E567 Y571 H586 T589 E593 E594 R595 F596 R603 G609 A625 ARG SER GLY ALA ALA ARG LYS PRO VAL LEU

• Molecule 1: 4-hydroxyphenylpyruvate dioxygenase

Chain D:

A-1 K0 L36 C69 N77 D86 L91 D94 K126 R127 G128 L129 W138 F167 M192 I197 V202 S206 E217 F223 G244 F245 R246 A247 A248 N253 R269 E273 N313 R317 H326 V331 Q332 L333 L334 R335

H358 G359 P360 R383 C396 G415 T416 ALA GLY H419 T420 L421 N431 M444 A450 D469 V472 V473 L474 P475 ASP PRO TYR G479 L480 V481 K482 A485 E501 H509 A510 L511 S512 S513 Y514 R515 G516 S517 I522 F560 D561 D562 E563 F564 L565 D576

Q580 G581 G582 H586 R595 F596 F597 A605 G606 Y607 M622 A623 R626 S627 GLY ALA ALA ARG LYS PRO VAL LEU

• Molecule 1: 4-hydroxyphenylpyruvate dioxygenase

Chain E:

A-1 R64 R70 R71 D72 K76 N77 R80 A104 A135 W138 G139 R140 V148 F167 S177 R180 L198 A199 M200 D201 V202 L203 E204 W205 S206 E217 M220 F245 ARG A247 L261 E266 R269 E273 L299 D304 V307

G319 H326 V331 D338 R368 R378 G389 L390 V391 C396 D405 Q414 G415 THR ALA GLY HIS T420 L421 Y422 D423 S427 L428 D429 A432 T435 G436 G437 M444 S462 V472 V473 LEU PRO ASP PRO TYR G479 L480 V481 K482 R487

S500 N504 I507 L511 S512 S513 Y514 R515 G516 V519 I529 E532 L538 Y550 Y551 R557 A569 Y575 D576 R577 E583 H586 R595 F596 F597 Q602 R603 K604 V617 M622 A629 ALA ARG LYS PRO VAL LEU
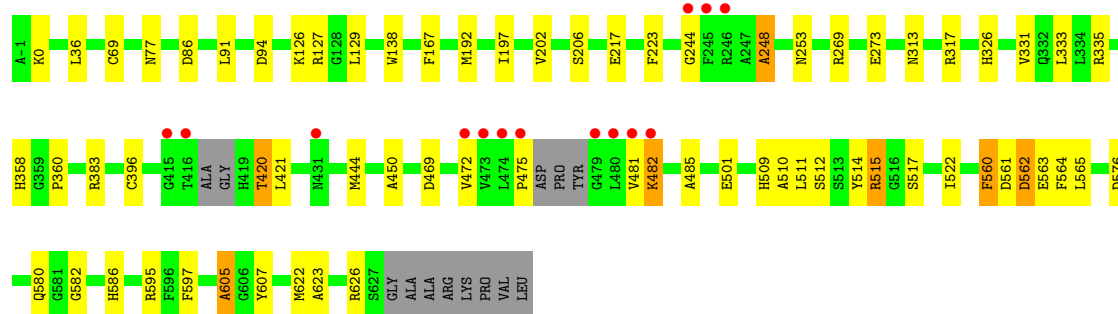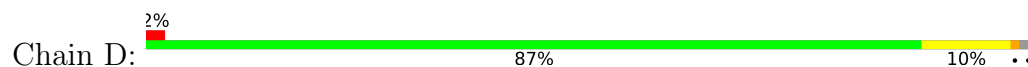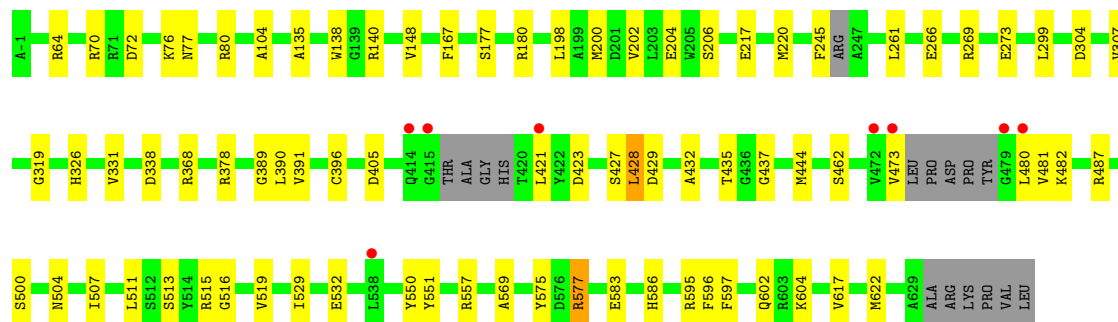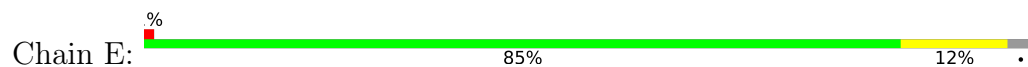
• Molecule 1: 4-hydroxyphenylpyruvate dioxygenase

Chain F:

A-1 S11 R44 R70 R71 D72 M88 Q89 E90 V96 K126 F167 M200 S206 E217 M218 D219 M220 R233 D243 R246 A247 A248 E266 R269 E273 I279 V307 R310 N313 K316 G319 E322 H326 V331

L334 R378 C396 G415 THR ALA GLY HIS T420 L421 Y422 M430 N431 A432 T433 A434 T435 H443 E451 S452 L463 V457 S462 V473 L474 P475 ASP PRO TYR G479 L480 V481 K482 L486 R487 S488 Q489 C490 R494 E501 T505 A508 H509 A510

R535 Y550 Y551 D562 E563 A569 H586 E593 A605 V617 G628 ALA ALA ARG LYS PRO VAL LEU

4 Data and refinement statistics ⓘ

Property	Value	Source
Space group	P 63	Depositor
Cell constants a, b, c, α, β, γ	262.50Å 262.50Å 139.64Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	40.44 – 2.37 88.15 – 2.37	Depositor EDS
% Data completeness (in resolution range)	100.0 (40.44-2.37) 95.6 (88.15-2.37)	Depositor EDS
R_{merge}	0.05	Depositor
R_{sym}	0.08	Depositor
$< I/\sigma(I) >$ [1]	1.40 (at 2.37Å)	Xtriage
Refinement program	PHENIX (1.10.1_2155: ???)	Depositor
R, R_{free}	0.204 , 0.229 0.204 , 0.230	Depositor DCC
R_{free} test set	10963 reflections (4.96%)	wwPDB-VP
Wilson B-factor (Å^2)	41.3	Xtriage
Anisotropy	0.035	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.30 , 19.6	EDS
L-test for twinning[2]	$< |L| > = 0.44, < L^2 > = 0.26$	Xtriage
Estimated twinning fraction	0.125 for h,-h-k,-l	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	30198	wwPDB-VP
Average B, all atoms (Å^2)	49.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.02% 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: MG

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	0.29	0/4946	0.51	2/6682 (0.0%)
1	B	0.30	0/4948	0.50	1/6684 (0.0%)
1	C	0.29	0/4918	0.49	1/6645 (0.0%)
1	D	0.30	0/4966	0.54	4/6713 (0.1%)
1	E	0.30	0/4939	0.51	3/6672 (0.0%)
1	F	0.29	0/4956	0.50	1/6699 (0.0%)
All	All	0.30	0/29673	0.51	12/40095 (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	B	0	1

There are no bond length outliers.

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

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	D	511	LEU	CB-CG-CD1	-10.27	93.54	111.00
1	B	511	LEU	CB-CG-CD2	8.96	126.22	111.00
1	D	515	ARG	NE-CZ-NH2	-7.80	116.40	120.30
1	D	472	VAL	CG1-CB-CG2	-6.82	99.99	110.90
1	F	481	VAL	CG1-CB-CG2	6.23	120.87	110.90

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	B	471	GLU	Sidechain

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	4850	0	4743	54	0
1	B	4854	0	4750	32	0
1	C	4824	0	4697	53	0
1	D	4869	0	4759	50	0
1	E	4845	0	4729	56	0
1	F	4860	0	4745	43	0
2	A	2	0	0	0	0
2	B	2	0	0	0	0
2	C	2	0	0	0	0
2	D	2	0	0	0	0
2	E	2	0	0	0	0
2	F	2	0	0	0	0
3	A	140	0	0	2	0
3	B	195	0	0	4	0
3	C	128	0	0	3	0
3	D	231	0	0	5	0
3	E	236	0	0	5	0
3	F	154	0	0	2	0
All	All	30198	0	28423	265	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 265 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:E:217:GLU:OE1	1:E:482:LYS:NZ	1.94	0.99
1:E:427:SER:HA	1:E:428:LEU:HB2	1.42	0.99
1:C:502:ASN:OD1	1:C:505:THR:N	2.05	0.89
1:B:511:LEU:HD23	1:B:513:SER:H	1.37	0.89
1:B:603:ARG:NH2	1:B:609:GLY:O	2.07	0.87

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	615/637 (96%)	592 (96%)	19 (3%)	4 (1%)	22	30
1	B	613/637 (96%)	597 (97%)	13 (2%)	3 (0%)	29	39
1	C	618/637 (97%)	593 (96%)	19 (3%)	6 (1%)	15	21
1	D	618/637 (97%)	593 (96%)	17 (3%)	8 (1%)	12	15
1	E	613/637 (96%)	601 (98%)	10 (2%)	2 (0%)	41	53
1	F	617/637 (97%)	597 (97%)	16 (3%)	4 (1%)	25	34
All	All	3694/3822 (97%)	3573 (97%)	94 (2%)	27 (1%)	22	30

5 of 27 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	243	ASP
1	A	432	ALA
1	A	433	THR
1	B	562	ASP
1	C	433	THR

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	495/501 (99%)	494 (100%)	1 (0%)	93	97

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	B	496/501 (99%)	491 (99%)	5 (1%)	76	87
1	C	490/501 (98%)	487 (99%)	3 (1%)	86	93
1	D	497/501 (99%)	493 (99%)	4 (1%)	81	91
1	E	494/501 (99%)	491 (99%)	3 (1%)	86	93
1	F	497/501 (99%)	495 (100%)	2 (0%)	91	96
All	All	2969/3006 (99%)	2951 (99%)	18 (1%)	86	93

5 of 18 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	E	577	ARG
1	F	586	HIS
1	F	44	ARG
1	C	586	HIS
1	E	500	SER

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (3) such sidechains are listed below:

Mol	Chain	Res	Type
1	C	373	GLN
1	E	373	GLN
1	F	580	GLN

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 ⓘ

Of 12 ligands modelled in this entry, 12 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.

No monomer is involved in short contacts.

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(Å²)	Q<0.9
1	A	612/637 (96%)	0.01	15 (2%)	57	59	38, 51, 70, 90	0
1	B	612/637 (96%)	-0.09	12 (1%)	65	66	34, 44, 70, 96	0
1	C	613/637 (96%)	-0.08	7 (1%)	80	81	37, 50, 76, 98	0
1	D	615/637 (96%)	0.02	14 (2%)	60	62	31, 44, 75, 100	0
1	E	612/637 (96%)	0.01	8 (1%)	77	78	30, 44, 71, 83	0
1	F	614/637 (96%)	-0.10	8 (1%)	77	78	35, 47, 67, 96	0
All	All	3678/3822 (96%)	-0.04	64 (1%)	70	71	30, 47, 72, 100	0

The worst 5 of 64 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	D	415	GLY	9.7
1	B	475	PRO	9.1
1	B	480	LEU	7.2
1	B	474	LEU	6.8
1	D	480	LEU	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 (i)

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	MG	D	702	1/1	0.92	0.08	47,47,47,47	0
2	MG	C	702	1/1	0.93	0.13	45,45,45,45	0
2	MG	A	701	1/1	0.93	0.18	45,45,45,45	0
2	MG	B	701	1/1	0.94	0.11	38,38,38,38	0
2	MG	E	702	1/1	0.97	0.11	42,42,42,42	0
2	MG	D	701	1/1	0.98	0.08	33,33,33,33	0
2	MG	B	702	1/1	0.98	0.16	38,38,38,38	0
2	MG	E	701	1/1	0.98	0.16	30,30,30,30	0
2	MG	A	702	1/1	0.98	0.13	44,44,44,44	0
2	MG	F	701	1/1	0.98	0.05	41,41,41,41	0
2	MG	F	702	1/1	0.98	0.14	34,34,34,34	0
2	MG	C	701	1/1	0.99	0.07	43,43,43,43	0

6.5   Other polymers (i)

There are no such residues in this entry.