wwPDB X-ray Structure Validation Summary Report (i)

Sep 28, 2024 – 05:11 pm BST

PDB ID	:	5G1V
Title	:	Linalool Dehydratase Isomerase: Selenomethionine Derivative
Authors	:	Chambers, S.; Hau, A.; Man, H.; Omar, M.; Turkenburg, J.P.; Grogan, G.
Deposited on	:	2016-03-30
Resolution	:	2.68 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:

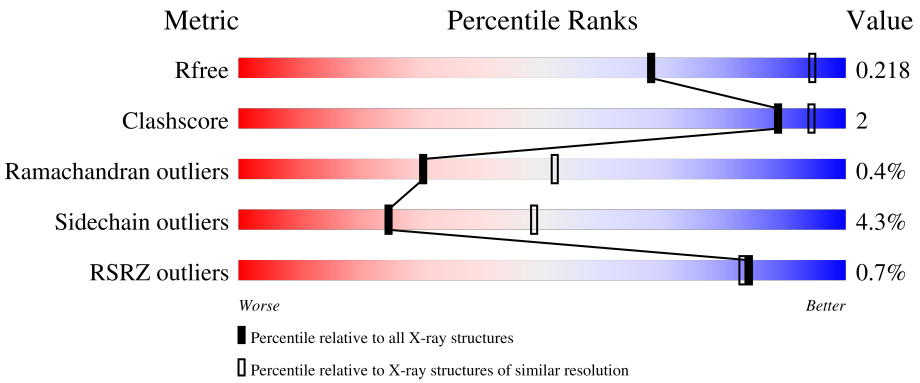
MolProbity	:	4.02b-467
Mogul	:	1.8.4, CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	3.0
Percentile statistics	:	20231227.v01 (using entries in the PDB archive December 27th 2023)
CCP4	:	9.0.003 (Gargrove)
Density-Fitness	:	1.0.11
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.39

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

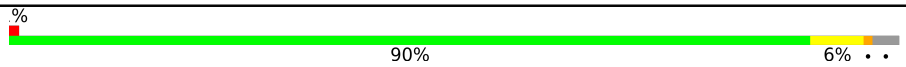
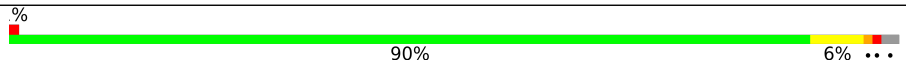
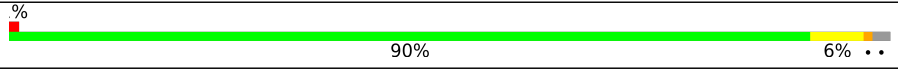
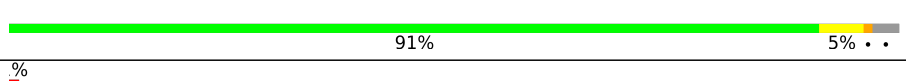
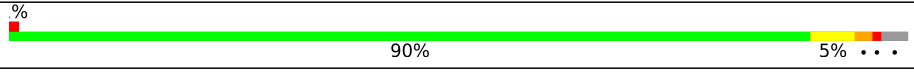
The reported resolution of this entry is 2.68 Å.

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}	164625	4708 (2.70-2.66)
Clashscore	180529	5138 (2.70-2.66)
Ramachandran outliers	177936	5071 (2.70-2.66)
Sidechain outliers	177891	5071 (2.70-2.66)
RSRZ outliers	164620	4708 (2.70-2.66)

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	372	
1	B	372	
1	C	372	
1	D	372	
1	E	372	

2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 15049 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 LINALOOL DEHYDRATASE ISOMERASE.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	N	O	S	Se			
1	A	362	2875	1859	480	523	4	9	0	0	0
1	B	363	2877	1860	479	525	4	9	0	0	0
1	C	363	2891	1868	482	528	4	9	0	0	0
1	D	362	2866	1855	475	523	4	9	0	0	0
1	E	362	2885	1864	485	523	4	9	0	0	0

There are 5 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	MSE	-	expression tag	UNP E1XUJ2
B	1	MSE	-	expression tag	UNP E1XUJ2
C	1	MSE	-	expression tag	UNP E1XUJ2
D	1	MSE	-	expression tag	UNP E1XUJ2
E	1	MSE	-	expression tag	UNP E1XUJ2

- Molecule 2 is water.

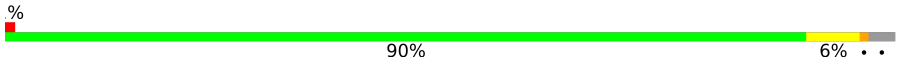
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
2	A	123	123	123	0	0
2	B	131	131	131	0	0
2	C	147	147	147	0	0
2	D	123	123	123	0	0
2	E	131	131	131	0	0

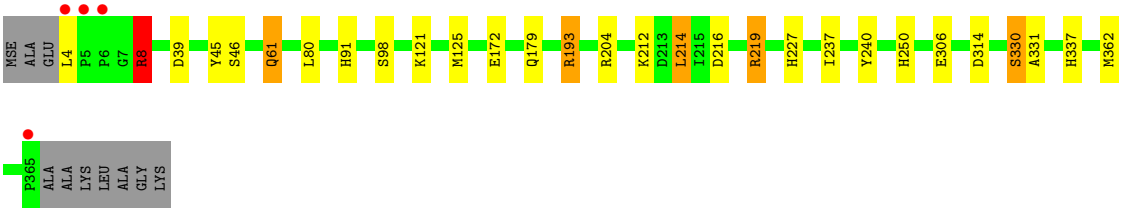
3 Residue-property plots (i)

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: LINALOOL DEHYDRATASE ISOMERASE

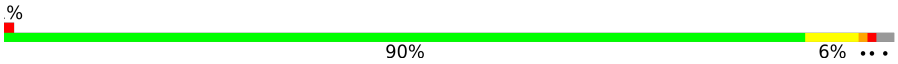
Chain A:





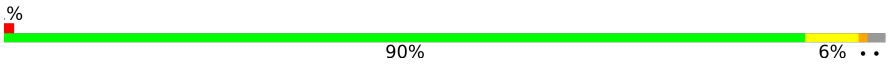
• Molecule 1: LINALOOL DEHYDRATASE ISOMERASE

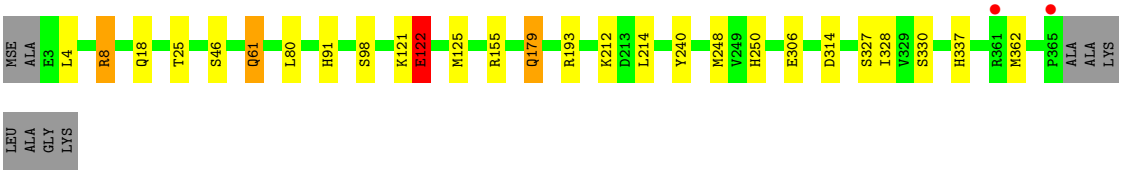
Chain B:



• Molecule 1: LINALOOL DEHYDRATASE ISOMERASE

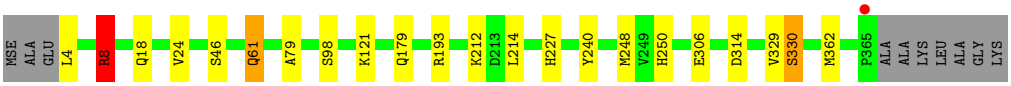
Chain C:





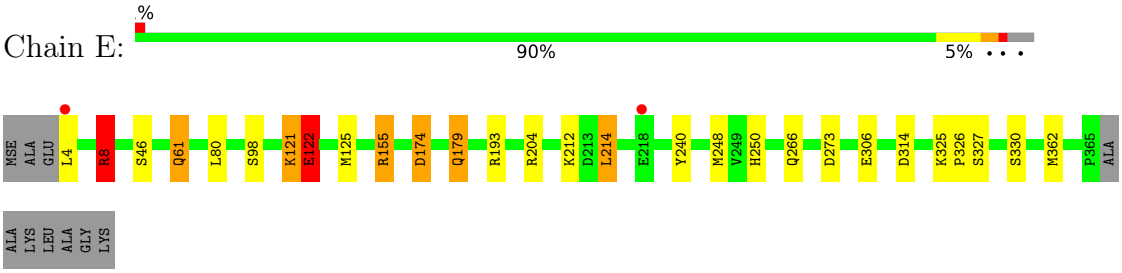
• Molecule 1: LINALOOL DEHYDRATASE ISOMERASE

Chain D:



• Molecule 1: LINALOOL DEHYDRATASE ISOMERASE

Chain E:


%
90%
5%
MSE
ALA
GLU
L4
R8
S46
Q61
L80
S98
K121
E122
M125
R155
D174
Q179
R193
R204
K212
D213
L214
E218
Y240
M248
V249
H250
Q266
D273
E306
D314
K325
P326
S327
S330
M362
P365
ALA
ALA
LYS
LEU
ALA
GLY
LYS

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	100.51Å 106.72Å 222.81Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	111.41 – 2.68 111.41 – 2.68	Depositor EDS
% Data completeness (in resolution range)	99.8 (111.41-2.68) 99.8 (111.41-2.68)	Depositor EDS
R_{merge}	0.24	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	3.58 (at 2.69Å)	Xtriage
Refinement program	REFMAC 5.8.0131	Depositor
R, R_{free}	0.168 , 0.203 0.180 , 0.218	Depositor DCC
R_{free} test set	3292 reflections (4.84%)	wwPDB-VP
Wilson B-factor (Å^2)	25.9	Xtriage
Anisotropy	1.012	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.33 , 24.8	EDS
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	15049	wwPDB-VP
Average B, all atoms (Å^2)	27.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.65% 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 (i)

5.1 Standard geometry (i)

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	0.72	0/2952	0.87	7/4003 (0.2%)
1	B	0.72	1/2954 (0.0%)	0.89	10/4007 (0.2%)
1	C	0.73	0/2968	0.85	4/4023 (0.1%)
1	D	0.72	0/2943	0.85	5/3993 (0.1%)
1	E	0.72	0/2962	0.87	9/4015 (0.2%)
All	All	0.72	1/14779 (0.0%)	0.87	35/20041 (0.2%)

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	122	GLU	CD-OE2	-5.25	1.19	1.25

The worst 5 of 35 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	E	8	ARG	NE-CZ-NH2	-8.86	115.87	120.30
1	A	8	ARG	NE-CZ-NH2	-8.79	115.91	120.30
1	D	8	ARG	NE-CZ-NH2	-8.55	116.02	120.30
1	B	8	ARG	NE-CZ-NH2	-7.86	116.37	120.30
1	A	8	ARG	NE-CZ-NH1	7.73	124.16	120.30

There are no chirality outliers.

There are no planarity outliers.

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	2875	0	2767	11	0
1	B	2877	0	2758	8	0
1	C	2891	0	2786	14	0
1	D	2866	0	2744	9	0
1	E	2885	0	2782	13	0
2	A	123	0	0	3	0
2	B	131	0	0	2	0
2	C	147	0	0	6	0
2	D	123	0	0	1	0
2	E	131	0	0	5	0
All	All	15049	0	13837	52	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 2.

The worst 5 of 52 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:121:LYS:HG2	1:E:122:GLU:HG2	1.35	1.06
1:A:172:GLU:OE2	2:A:2034:HOH:O	1.82	0.95
1:D:329:VAL:O	2:D:2116:HOH:O	1.83	0.94
1:E:121:LYS:HG2	1:E:122:GLU:CG	2.12	0.77
1:E:121:LYS:C	1:E:122:GLU:HG2	2.05	0.77

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	360/372 (97%)	350 (97%)	8 (2%)	2 (1%)	22 42
1	B	361/372 (97%)	350 (97%)	11 (3%)	0	100 100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	C	361/372 (97%)	352 (98%)	7 (2%)	2 (1%)	22	42
1	D	360/372 (97%)	351 (98%)	8 (2%)	1 (0%)	37	59
1	E	360/372 (97%)	351 (98%)	6 (2%)	3 (1%)	16	35
All	All	1802/1860 (97%)	1754 (97%)	40 (2%)	8 (0%)	30	52

5 of 8 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	E	174	ASP
1	A	212	LYS
1	C	122	GLU
1	D	212	LYS
1	E	212	LYS

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	293/293 (100%)	280 (96%)	13 (4%)	24	48
1	B	292/293 (100%)	278 (95%)	14 (5%)	21	44
1	C	296/293 (101%)	284 (96%)	12 (4%)	26	50
1	D	290/293 (99%)	280 (97%)	10 (3%)	32	58
1	E	294/293 (100%)	280 (95%)	14 (5%)	21	44
All	All	1465/1465 (100%)	1402 (96%)	63 (4%)	25	49

5 of 63 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	8	ARG
1	E	121	LYS
1	C	214	LEU
1	E	98	SER
1	E	204	ARG

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 29 such sidechains are listed below:

Mol	Chain	Res	Type
1	C	211	GLN
1	E	250	HIS
1	D	18	GLN
1	E	58	HIS
1	C	250	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 oligosaccharides 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 (i)

6.1 Protein, DNA and RNA chains (i)

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	353/372 (94%)	-0.49	4 (1%)	77	77	18, 26, 44, 64	0
1	B	354/372 (95%)	-0.47	4 (1%)	77	77	17, 25, 44, 70	0
1	C	354/372 (95%)	-0.54	2 (0%)	85	84	16, 25, 42, 66	0
1	D	353/372 (94%)	-0.53	1 (0%)	90	89	16, 26, 40, 54	0
1	E	353/372 (94%)	-0.51	2 (0%)	85	84	16, 27, 43, 67	0
All	All	1767/1860 (95%)	-0.51	13 (0%)	84	83	16, 26, 43, 70	0

The worst 5 of 13 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	3	GLU	5.4
1	E	4	LEU	4.2
1	B	212	LYS	3.4
1	E	218	GLU	3.2
1	C	365	PRO	3.1

6.2 Non-standard residues in protein, DNA, RNA chains (i)

There are no non-standard protein/DNA/RNA residues in this entry.

6.3 Carbohydrates (i)

There are no monosaccharides in this entry.

6.4 Ligands (i)

There are no ligands in this entry.

6.5 Other polymers (i)

There are no such residues in this entry.