WORLDWIDE wwPDB PROTEIN DATA BANK

wwPDB X-ray Structure Validation Summary Report (i)

Feb 1, 2016 – 04:59 PM GMT

PDB ID   :   4GTU
Title   :   LIGAND-FREE HOMODIMERIC HUMAN GLUTATHIONE S-TRANSFERASE M4-4
Authors   :   Patskovsky, Y.V.; Patskovska, L.N.; Listowsky, I.
Deposited on :   1999-05-12
Resolution :   3.30 Å(reported)

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

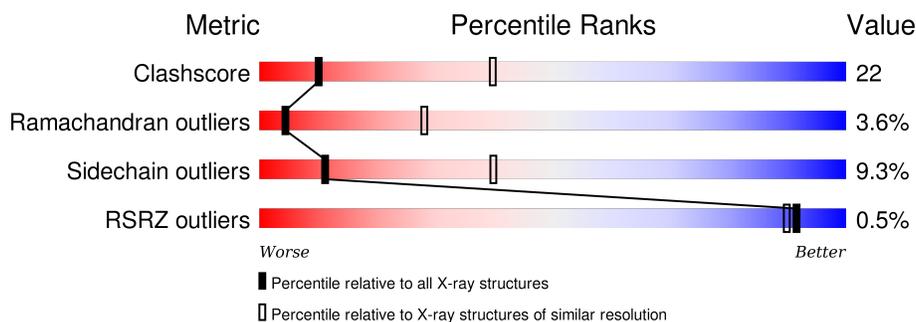|		
MolProbity	:	4.02b-467
Mogul	:	1.7 (RC4), CSD as536be (2015)
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20026688
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	trunk26865

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	Percentile Ranks	Value
Clashscore		22
Ramachandran outliers		3.6%
Sidechain outliers		9.3%
RSRZ outliers		0.5%

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))
Clashscore	102246	1058 (3.38-3.22)
Ramachandran outliers	100387	1038 (3.38-3.22)
Sidechain outliers	100360	1037 (3.38-3.22)
RSRZ outliers	91569	2070 (3.40-3.20)

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	217	51% 44% 5%
1	B	217	54% 40% 6%
1	C	217	47% 46% 7%
1	D	217	% 50% 47% •
1	E	217	53% 41% 6%
1	F	217	% 39% 53% 8%
1	G	217	52% 41% 6%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	H	217	55% 39% 6%

2 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 14352 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 GLUTATHIONE S-TRANSFERASE.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	217	Total	C	N	O	S	0	0	0
			1794	1160	295	329	10			
1	B	217	Total	C	N	O	S	0	0	0
			1794	1160	295	329	10			
1	C	217	Total	C	N	O	S	0	0	0
			1794	1160	295	329	10			
1	D	217	Total	C	N	O	S	0	0	0
			1794	1160	295	329	10			
1	E	217	Total	C	N	O	S	0	0	0
			1794	1160	295	329	10			
1	F	217	Total	C	N	O	S	0	0	0
			1794	1160	295	329	10			
1	G	217	Total	C	N	O	S	0	0	0
			1794	1160	295	329	10			
1	H	217	Total	C	N	O	S	0	0	0
			1794	1160	295	329	10			

3 Residue-property plots ⓘ

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: GLUTATHIONE S-TRANSFERASE

- Molecule 1: GLUTATHIONE S-TRANSFERASE

- Molecule 1: GLUTATHIONE S-TRANSFERASE

● Molecule 1: GLUTATHIONE S-TRANSFERASE

Chain D:

● Molecule 1: GLUTATHIONE S-TRANSFERASE

Chain E:

● Molecule 1: GLUTATHIONE S-TRANSFERASE

Chain F:

● Molecule 1: GLUTATHIONE S-TRANSFERASE

Chain G:

D175 A176 F177 L180 K181 D182 F183 I184 F187 E188 K192 I193 Y196 M197 R201 K205 P206 L207 Y208 T209 R210 V211 A212 V213 W214 G215 N216 K217

• Molecule 1: GLUTATHIONE S-TRANSFERASE

Chain H:

55%　　　　　　　　　　　　39%　　　　6%

S1 M2 G5 Y6 W7 R10 H14 R17 L18 L19 L20 E21 S25 K30 K31 Y32 T33 D36 A37 P38 D39 Y40 D41 R42 S43 Q44 W45 L46 N47 E48 K49 L52 P57 N58 L59 P60 Y61 L62 I63 K68 I69 T70 Q71 S72 N73 A74 I75 Y78 I79

K82 H83 E88 T89 E92 K93 I94 R95 V96 Q102 D105 V106 S107 N108 Q109 L110 A111 R112 V113 C114 F119 K123 P124 E125 Y126 L127 L130 M133 F140 L141 G142 K143 R144 P145 W146 F147 V148 K151 I152 T153 F154 V155 A159 Y160 D161 V162 L163

H166 R167 I168 L174 D175 A176 F177 P178 N179 L180 K181 I184 S185 R186 F187 E188 E191 Y196 L203 P204 L207 Y208 T209 R210 V213 W214 K217

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	51.00Å 215.70Å 94.56Å 90.00° 97.40° 90.00°	Depositor
Resolution (Å)	10.00 – 3.30 39.38 – 2.80	Depositor EDS
% Data completeness (in resolution range)	79.5 (10.00-3.30) 63.5 (39.38-2.80)	Depositor EDS
R_{merge}	0.17	Depositor
R_{sym}	0.16	Depositor
$< I/\sigma(I) >$ [1]	1.49 (at 2.81Å)	Xtriage
Refinement program	X-PLOR 3.851	Depositor
R, R_{free}	0.245 , 0.315 0.260 , (Not available)	Depositor DCC
R_{free} test set	No test flags present.	DCC
Wilson B-factor (Å^2)	39.1	Xtriage
Anisotropy	0.142	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.26 , 73.6	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning[2]	$< |L| > = 0.42, < L^2 > = 0.24$	Xtriage
Outliers	1 of 31560 reflections (0.003%)	Xtriage
F_o,F_c correlation	0.78	EDS
Total number of atoms	14352	wwPDB-VP
Average B, all atoms (Å^2)	18.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 17.09% 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, 0.375 respectively for untwinned datasets, and 0.333, 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	0.43	0/1842	0.61	0/2489
1	B	0.44	0/1842	0.62	0/2489
1	C	0.42	0/1842	0.61	0/2489
1	D	0.42	0/1842	0.57	0/2489
1	E	0.41	0/1842	0.60	0/2489
1	F	0.43	0/1842	0.58	0/2489
1	G	0.42	0/1842	0.63	0/2489
1	H	0.40	0/1842	0.59	0/2489
All	All	0.42	0/14736	0.60	0/19912

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 (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	1794	0	1762	83	0
1	B	1794	0	1762	83	0
1	C	1794	0	1762	84	0
1	D	1794	0	1762	74	0
1	E	1794	0	1762	75	0
1	F	1794	0	1762	107	0
1	G	1794	0	1762	75	0
1	H	1794	0	1762	70	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
All	All	14352	0	14096	623	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 22.

The worst 5 of 623 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:206:PRO:HG3	1:C:217:LYS:HB3	1.50	0.94
1:A:144:ARG:HD3	1:A:149:GLY:HA2	1.50	0.92
1:C:37:ALA:HB1	1:C:38:PRO:HD2	1.55	0.89
1:E:80:ALA:HB1	1:E:86:CYS:SG	2.12	0.89
1:A:114:CYS:SG	1:A:214:TRP:HB3	2.14	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	215/217 (99%)	184 (86%)	24 (11%)	7 (3%)	5	30
1	B	215/217 (99%)	176 (82%)	28 (13%)	11 (5%)	2	19
1	C	215/217 (99%)	180 (84%)	29 (14%)	6 (3%)	6	34
1	D	215/217 (99%)	183 (85%)	27 (13%)	5 (2%)	8	39
1	E	215/217 (99%)	188 (87%)	20 (9%)	7 (3%)	5	30
1	F	215/217 (99%)	183 (85%)	25 (12%)	7 (3%)	5	30
1	G	215/217 (99%)	182 (85%)	23 (11%)	10 (5%)	3	20
1	H	215/217 (99%)	184 (86%)	22 (10%)	9 (4%)	3	23
All	All	1720/1736 (99%)	1460 (85%)	198 (12%)	62 (4%)	4	28

5 of 62 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	49	LYS
1	B	147	PHE
1	B	204	PRO
1	C	14	HIS
1	C	71	GLN

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	195/195 (100%)	183 (94%)	12 (6%)	23	61
1	B	195/195 (100%)	180 (92%)	15 (8%)	16	51
1	C	195/195 (100%)	176 (90%)	19 (10%)	10	38
1	D	195/195 (100%)	177 (91%)	18 (9%)	11	40
1	E	195/195 (100%)	175 (90%)	20 (10%)	9	34
1	F	195/195 (100%)	172 (88%)	23 (12%)	6	27
1	G	195/195 (100%)	175 (90%)	20 (10%)	9	34
1	H	195/195 (100%)	177 (91%)	18 (9%)	11	40
All	All	1560/1560 (100%)	1415 (91%)	145 (9%)	11	40

5 of 145 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	E	2	MET
1	E	201	ARG
1	H	71	GLN
1	E	41	ASP
1	E	118	ASP

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 22 such sidechains are listed below:

Mol	Chain	Res	Type
1	D	14	HIS
1	D	216	ASN
1	H	84	ASN
1	D	108	ASN
1	D	179	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 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 (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	217/217 (100%)	-0.22	0	100	100	6, 18, 29, 35	0
1	B	217/217 (100%)	-0.21	1 (0%)	91	90	5, 17, 28, 35	0
1	C	217/217 (100%)	-0.24	1 (0%)	91	90	5, 18, 28, 34	0
1	D	217/217 (100%)	-0.24	2 (0%)	85	82	5, 17, 28, 35	0
1	E	217/217 (100%)	-0.15	1 (0%)	91	90	6, 17, 29, 34	0
1	F	217/217 (100%)	-0.22	2 (0%)	85	82	5, 18, 29, 34	0
1	G	217/217 (100%)	-0.22	1 (0%)	91	90	5, 17, 29, 34	0
1	H	217/217 (100%)	-0.31	0	100	100	4, 17, 28, 34	0
All	All	1736/1736 (100%)	-0.23	8 (0%)	91	90	4, 18, 29, 35	0

The worst 5 of 8 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	F	39	ASP	4.0
1	D	38	PRO	3.4
1	F	38	PRO	2.8
1	B	38	PRO	2.8
1	D	39	ASP	2.6

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 carbohydrates 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.