wwPDB X-ray Structure Validation Summary Report (i)

Dec 9, 2016 – 11:53 AM EST

PDB ID	:	5F75
Title	:	Thiocyanate dehydrogenase from Thioalkalivibrio paradoxus
Authors	:	Tsallagov, S.I.; Polyakov, K.M.; Tikhonova, T.V.; Trofimov, A.A.; Shabalin, I.G.; Popov, A.N.; Popov, V.O.
Deposited on	:	2015-12-07
Resolution	:	2.00 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
http://wwpdb.org/validation/2016/XrayValidationReportHelp
with specific help available everywhere you see the (i) symbol.

The following versions of software and data (see references (1)) were used in the production of this report:

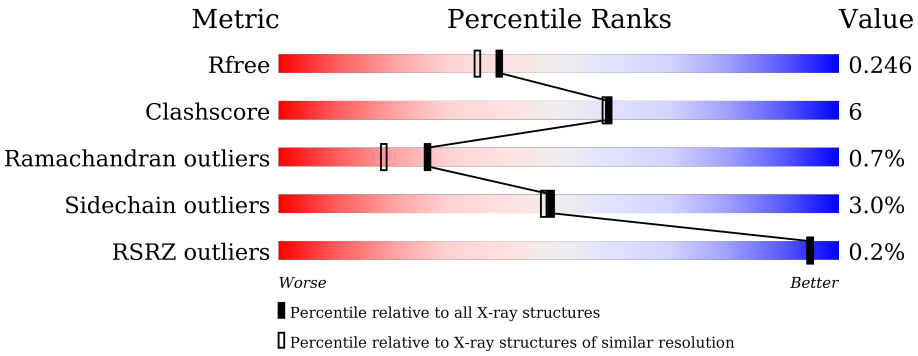
MolProbity	:	4.02b-467
Mogul	:	unknown
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20028442
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	rb-20028442

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 2.00 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}	91344	6249 (2.00-2.00)
Clashscore	102246	7340 (2.00-2.00)
Ramachandran outliers	100387	7248 (2.00-2.00)
Sidechain outliers	100360	7247 (2.00-2.00)
RSRZ outliers	91569	6262 (2.00-2.00)

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	548	72% 14% • 14%
1	B	548	72% 13% 15%
1	C	548	73% 13% • 13%
1	D	548	% 71% 13% • 15%

2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 15038 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 Thiocyanate dehydrogenase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	474	Total	C	N	O	S	0	1	0
			3670	2345	614	692	19			
1	B	466	Total	C	N	O	S	0	0	0
			3586	2292	600	676	18			
1	C	475	Total	C	N	O	S	0	3	0
			3673	2346	608	699	20			
1	D	466	Total	C	N	O	S	0	1	0
			3588	2293	594	683	18			

- Molecule 2 is COPPER (II) ION (three-letter code: CU) (formula: Cu).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	B	2	Total	Cu	0	0
			2	2		
2	A	2	Total	Cu	0	0
			2	2		
2	D	2	Total	Cu	0	0
			2	2		
2	C	2	Total	Cu	0	0
			2	2		

- Molecule 3 is water.

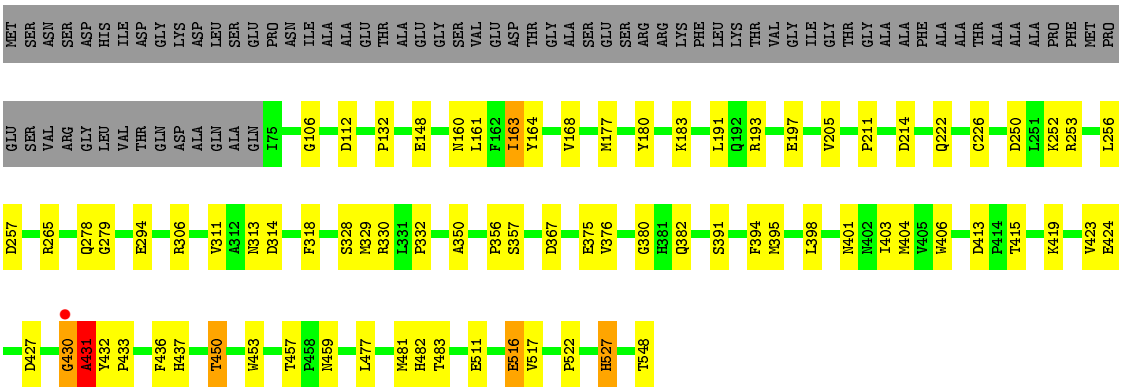
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	156	Total	O	0	0
			156	156		
3	B	149	Total	O	0	0
			149	149		
3	C	120	Total	O	0	0
			120	120		
3	D	88	Total	O	0	0
			88	88		

3 Residue-property plots (i)

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.

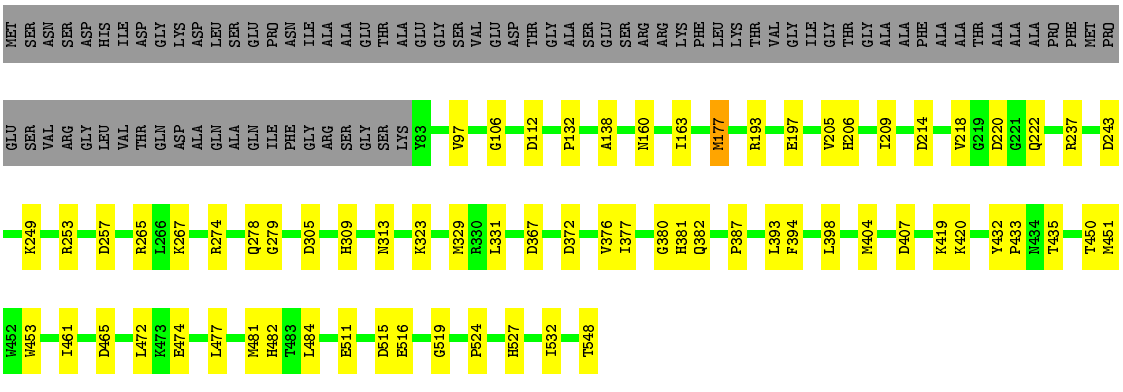
- Molecule 1: Thiocyanate dehydrogenase

Chain A:



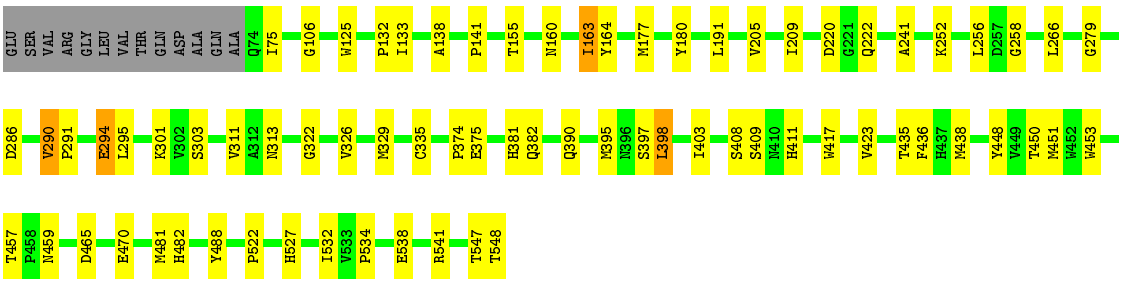
- Molecule 1: Thiocyanate dehydrogenase

Chain B:

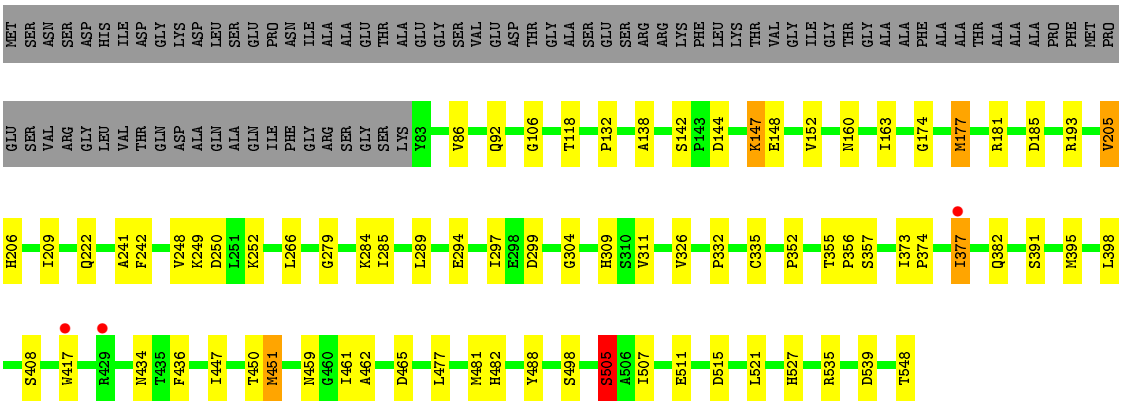


- Molecule 1: Thiocyanate dehydrogenase

Chain C:



• Molecule 1: Thiocyanate dehydrogenase



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	98.78Å 95.42Å 107.00Å 90.00° 100.66° 90.00°	Depositor
Resolution (Å)	49.00 – 2.00 48.54 – 2.00	Depositor EDS
% Data completeness (in resolution range)	94.3 (49.00-2.00) 94.3 (48.54-2.00)	Depositor EDS
R_{merge}	0.07	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.88 (at 2.00Å)	Xtriage
Refinement program	REFMAC 5.8.0073	Depositor
R, R_{free}	0.186 , 0.244 0.188 , 0.246	Depositor DCC
R_{free} test set	6063 reflections (5.13%)	DCC
Wilson B-factor (Å^2)	35.7	Xtriage
Anisotropy	0.088	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.35 , 46.0	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	0.97	EDS
Total number of atoms	15038	wwPDB-VP
Average B, all atoms (Å^2)	39.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.80% 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: CU

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	0.81	0/3779	0.91	8/5157 (0.2%)
1	B	0.74	0/3689	0.87	7/5043 (0.1%)
1	C	0.75	0/3792	0.87	2/5177 (0.0%)
1	D	0.69	1/3695 (0.0%)	0.83	1/5052 (0.0%)
All	All	0.75	1/14955 (0.0%)	0.87	18/20429 (0.1%)

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

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	D	505	SER	CB-OG	-5.12	1.35	1.42

The worst 5 of 18 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	306	ARG	NE-CZ-NH1	6.82	123.71	120.30
1	A	367	ASP	CB-CG-OD1	6.49	124.14	118.30
1	A	214	ASP	CB-CG-OD1	6.13	123.82	118.30
1	B	214	ASP	CB-CG-OD1	6.10	123.79	118.30
1	A	330	ARG	NE-CZ-NH1	5.96	123.28	120.30

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	430	GLY	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	3670	0	3534	57	0
1	B	3586	0	3426	32	0
1	C	3673	0	3514	44	0
1	D	3588	0	3417	39	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
3	A	156	0	0	1	0
3	B	149	0	0	2	0
3	C	120	0	0	1	0
3	D	88	0	0	0	0
All	All	15038	0	13891	166	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 6.

The worst 5 of 166 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:D:352:PRO:HB2	1:D:377:ILE:HD12	1.40	1.03
1:A:430:GLY:HA3	1:A:431:ALA:CB	1.92	0.99
1:A:430:GLY:HA3	1:A:431:ALA:HB3	1.49	0.92
1:C:160:ASN:HD21	1:C:222:GLN:HE22	1.18	0.91
1:A:148:GLU:OE1	1:A:183:LYS:NZ	2.07	0.87

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	473/548 (86%)	443 (94%)	27 (6%)	3 (1%)	30	22
1	B	464/548 (85%)	425 (92%)	35 (8%)	4 (1%)	21	13
1	C	476/548 (87%)	452 (95%)	22 (5%)	2 (0%)	39	33
1	D	465/548 (85%)	435 (94%)	26 (6%)	4 (1%)	21	13
All	All	1878/2192 (86%)	1755 (94%)	110 (6%)	13 (1%)	26	19

5 of 13 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	431	ALA
1	D	356	PRO
1	D	174	GLY
1	D	205	VAL
1	B	205	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	391/451 (87%)	381 (97%)	10 (3%)	54	54
1	B	378/451 (84%)	369 (98%)	9 (2%)	57	58
1	C	390/451 (86%)	381 (98%)	9 (2%)	58	60
1	D	380/451 (84%)	361 (95%)	19 (5%)	30	24
All	All	1539/1804 (85%)	1492 (97%)	47 (3%)	48	46

5 of 47 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	290	VAL
1	C	470	GLU
1	D	447	ILE
1	C	301	LYS
1	C	527	HIS

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 25 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	390	GLN
1	B	459	ASN
1	D	382	GLN
1	B	411	HIS
1	C	222	GLN

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)

Of 8 ligands modelled in this entry, 8 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 (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	474/548 (86%)	-0.44	1 (0%) 95 95	22, 35, 50, 61	0
1	B	466/548 (85%)	-0.38	0 100 100	23, 38, 61, 77	0
1	C	475/548 (86%)	-0.40	0 100 100	23, 37, 53, 66	0
1	D	466/548 (85%)	-0.22	3 (0%) 90 90	26, 44, 62, 74	0
All	All	1881/2192 (85%)	-0.36	4 (0%) 95 95	22, 38, 59, 77	0

All (4) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	D	377	ILE	2.7
1	A	430	GLY	2.3
1	D	429	ARG	2.3
1	D	417	TRP	2.1

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 carbohydrates 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. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. 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	LLDF	B-factors(Å^2)	Q<0.9
2	CU	C	601	1/1	1.00	0.11	1.53	32,32,32,32	0
2	CU	B	602	1/1	0.99	0.10	0.68	34,34,34,34	1
2	CU	B	601	1/1	0.99	0.09	0.10	37,37,37,37	0
2	CU	D	601	1/1	1.00	0.11	0.06	42,42,42,42	0
2	CU	A	602	1/1	0.99	0.09	-0.10	32,32,32,32	1
2	CU	D	602	1/1	0.99	0.10	-1.13	44,44,44,44	1
2	CU	A	601	1/1	0.99	0.09	-1.13	33,33,33,33	0
2	CU	C	602	1/1	1.00	0.10	-1.25	35,35,35,35	1

6.5 Other polymers

There are no such residues in this entry.