Full wwPDB X-ray Structure Validation Report (i)

Nov 13, 2021 – 05:02 am GMT

PDB ID	:	7OOL
Title	:	Crystal structure of a Candidatus photodesmus katoptron thioredoxin chimera containing an ancestral loop
Authors	:	Gavira, J.A.; Ibarra-Molero, B.; Gamiz-Arco, G.; Risso, V.; Sanchez-Ruiz, J.M.
Deposited on	:	2021-05-28
Resolution	:	2.85 Å(reported)

This is a Full wwPDB X-ray Structure Validation 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 following versions of software and data (see references (i)) were used in the production of this report:

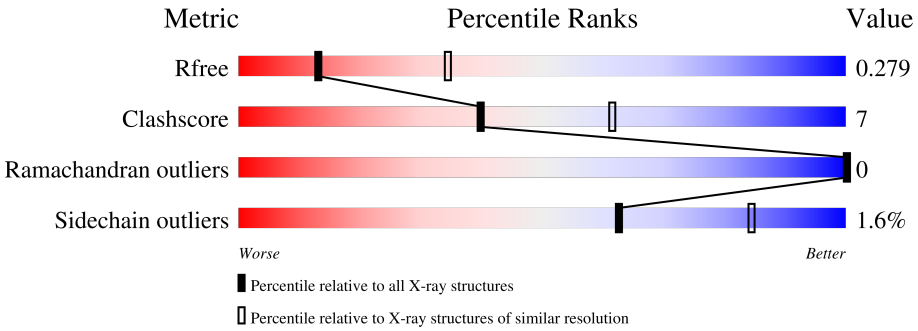
MolProbity	:	4.02b-467
Mogul	:	1.8.4 (270009), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.23.2
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0267
CCP4	:	7.1.010 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.23.2

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 2.85 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	3168 (2.90-2.82)
Clashscore	141614	3438 (2.90-2.82)
Ramachandran outliers	138981	3348 (2.90-2.82)
Sidechain outliers	138945	3351 (2.90-2.82)

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%

Mol	Chain	Length	Quality of chain
1	A	109	77% 20% .
1	B	109	80% 17% .

2 Entry composition (i)

There are 8 unique types of molecules in this entry. The entry contains 1749 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.

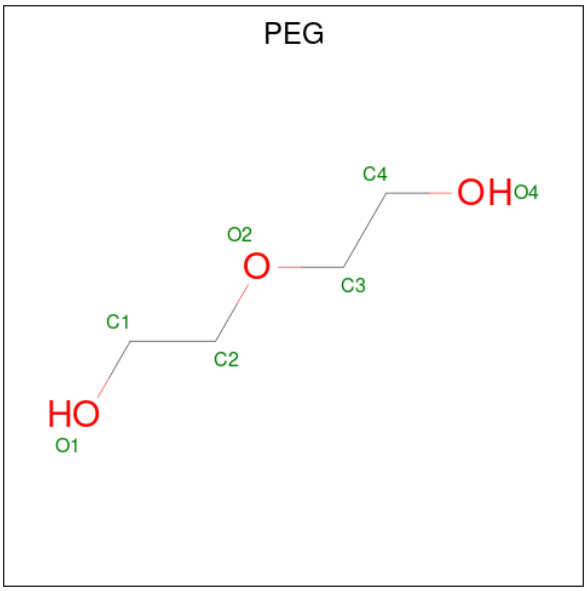
- Molecule 1 is a protein called Thioredoxin.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	106	Total	C	N	O	S	0	2	0
			844	548	134	160	2			
1	B	106	Total	C	N	O	S	0	1	0
			830	541	129	158	2			

There are 8 discrepancies between the modelled and reference sequences:

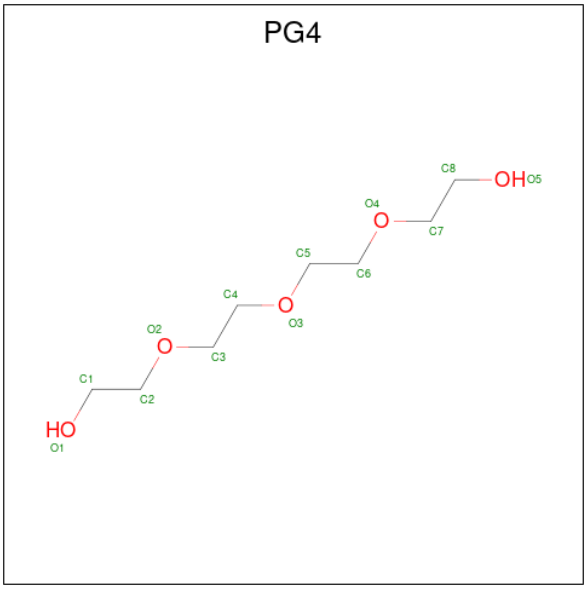
Chain	Residue	Modelled	Actual	Comment	Reference
A	69	TYR	PHE	conflict	UNP S3DGC4
A	73	SER	GLY	conflict	UNP S3DGC4
A	74	ILE	VAL	conflict	UNP S3DGC4
A	76	THR	SER	conflict	UNP S3DGC4
B	69	TYR	PHE	conflict	UNP S3DGC4
B	73	SER	GLY	conflict	UNP S3DGC4
B	74	ILE	VAL	conflict	UNP S3DGC4
B	76	THR	SER	conflict	UNP S3DGC4

- Molecule 2 is DI(HYDROXYETHYL)ETHER (three-letter code: PEG) (formula: $C_4H_{10}O_3$).



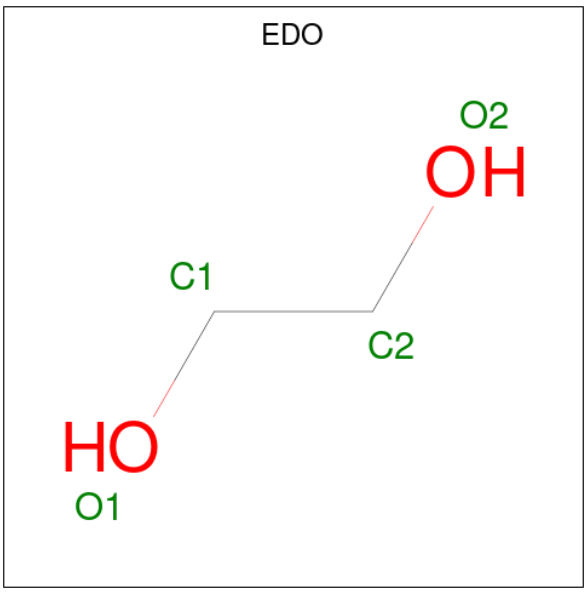
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	1	Total C O 7 4 3	0	0
2	A	1	Total C O 7 4 3	0	0
2	A	1	Total C O 7 4 3	0	0

- Molecule 3 is TETRAETHYLENE GLYCOL (three-letter code: PG4) (formula: $C_8H_{18}O_5$).



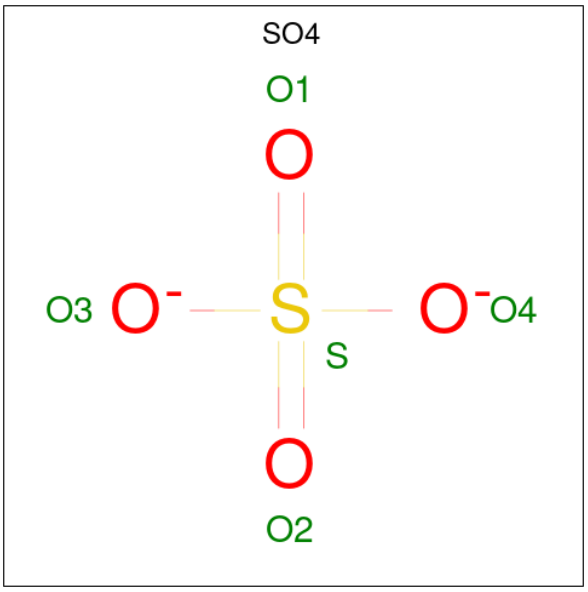
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	1	Total C O 13 8 5	0	0

- Molecule 4 is 1,2-ETHANEDIOL (three-letter code: EDO) (formula: $C_2H_6O_2$).



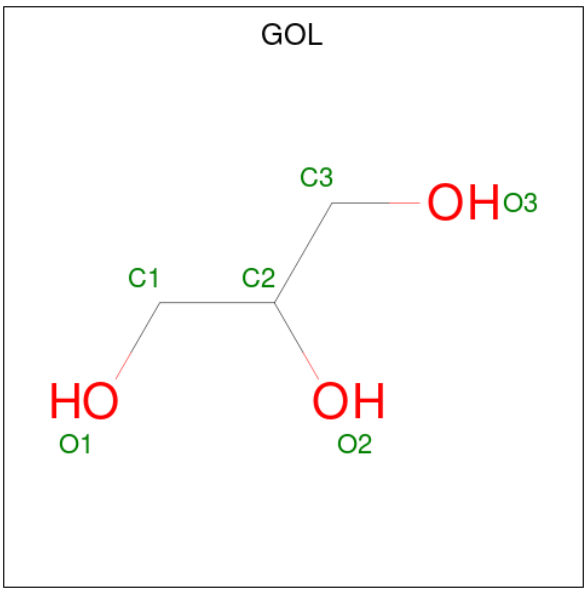
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	A	1	Total C O 4 2 2	0	0
4	B	1	Total C O 4 2 2	0	0

- Molecule 5 is SULFATE ION (three-letter code: SO4) (formula: O_4S).



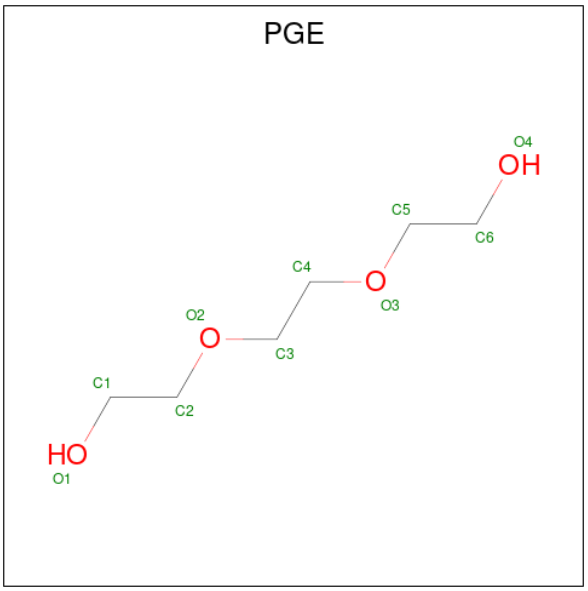
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	A	1	Total O S 5 4 1	0	0

- Molecule 6 is GLYCEROL (three-letter code: GOL) (formula: $C_3H_8O_3$).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
6	B	1	Total C O 6 3 3	0	0

- Molecule 7 is TRIETHYLENE GLYCOL (three-letter code: PGE) (formula: $C_6H_{14}O_4$).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
7	B	1	Total C O 10 6 4	0	0

- Molecule 8 is water.

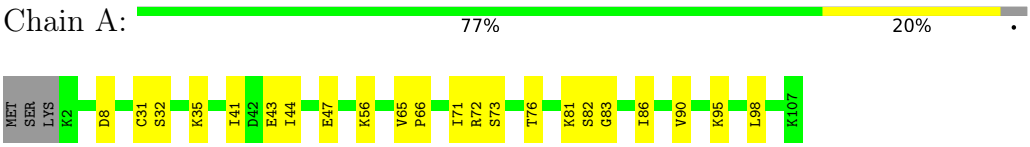
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
8	A	10	Total O 10 10	0	0
8	B	2	Total O 2 2	0	0

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. 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. 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.

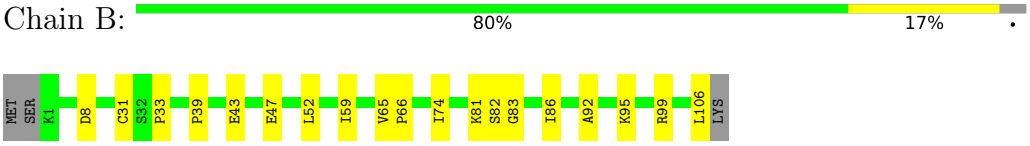
• Molecule 1: Thioredoxin

Chain A:



• Molecule 1: Thioredoxin

Chain B:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	I 21 3	Depositor
Cell constants a, b, c, α, β, γ	140.08Å 140.08Å 140.08Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	44.30 – 2.85 44.30 – 2.65	Depositor EDS
% Data completeness (in resolution range)	93.4 (44.30-2.85) 99.5 (44.30-2.65)	Depositor EDS
R_{merge}	0.12	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.25 (at 2.65Å)	Xtriage
Refinement program	PHENIX 1.19	Depositor
R, R_{free}	0.261 , 0.287 0.253 , 0.279	Depositor DCC
R_{free} test set	672 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å^2)	87.2	Xtriage
Anisotropy	0.000	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	(Not available) , (Not available)	EDS
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.32$	Xtriage
Estimated twinning fraction	0.026 for -l,-k,-h	Xtriage
F_o,F_c correlation	0.90	EDS
Total number of atoms	1749	wwPDB-VP
Average B, all atoms (Å^2)	71.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.58% 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: PEG, PGE, SO4, PG4, EDO, GOL

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	0.26	0/859	0.48	0/1166
1	B	0.27	0/848	0.51	0/1152
All	All	0.27	0/1707	0.50	0/2318

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	844	0	871	13	0
1	B	830	0	863	13	0
2	A	21	0	30	0	0
3	A	13	0	18	0	0
4	A	4	0	6	0	0
4	B	4	0	6	0	0
5	A	5	0	0	0	0
6	B	6	0	8	0	0
7	B	10	0	14	0	0
8	A	10	0	0	0	0
8	B	2	0	0	0	0
All	All	1749	0	1816	26	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.

All (26) 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:A:66:PRO:HA	1:A:71:ILE:HD12	1.65	0.77
1:A:81:LYS:HB2	1:A:86:ILE:HD13	1.70	0.73
1:B:59:ILE:HD11	1:B:74:ILE:HD13	1.72	0.71
1:B:52:LEU:HD23	1:B:106:LEU:HD11	1.80	0.62
1:B:81:LYS:HB2	1:B:86:ILE:HD13	1.85	0.58
1:B:82:SER:OG	1:B:83:GLY:N	2.41	0.53
1:A:47:GLU:OE1	1:A:95:LYS:NZ	2.41	0.53
1:A:41:ILE:HG21	1:A:56:LYS:HE3	1.92	0.52
1:A:82:SER:OG	1:A:83:GLY:N	2.43	0.52
1:A:8:ASP:OD1	1:A:65:VAL:HG23	2.09	0.51
1:B:8:ASP:OD1	1:B:65:VAL:HG23	2.11	0.50
1:B:59:ILE:HG22	1:B:66:PRO:HG3	1.93	0.50
1:A:76:THR:HG23	1:A:90:VAL:HG22	1.92	0.50
1:B:33:PRO:HB3	1:B:92:ALA:HB2	1.95	0.49
1:B:39:PRO:O	1:B:43:GLU:HG3	2.13	0.49
1:B:95:LYS:HE2	1:B:99:ARG:NH2	2.28	0.48
1:B:59:ILE:HG22	1:B:66:PRO:CG	2.45	0.46
1:A:44:ILE:HG13	1:A:98:LEU:HD23	1.97	0.46
1:B:95:LYS:O	1:B:99:ARG:HG3	2.16	0.46
1:A:72[B]:ARG:CZ	1:A:72[B]:ARG:HB2	2.45	0.44
1:A:72[B]:ARG:HH22	1:A:90:VAL:HG11	1.83	0.44
1:A:43:GLU:OE1	1:A:95:LYS:HE3	2.18	0.43
1:A:73:SER:O	1:A:76:THR:OG1	2.36	0.43
1:B:47:GLU:OE2	1:B:95:LYS:NZ	2.31	0.43
1:B:43:GLU:O	1:B:47:GLU:HG3	2.19	0.42
1:A:35:LYS:HD2	1:A:35:LYS:HA	1.80	0.41

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	106/109 (97%)	105 (99%)	1 (1%)	0	100	100
1	B	105/109 (96%)	104 (99%)	1 (1%)	0	100	100
All	All	211/218 (97%)	209 (99%)	2 (1%)	0	100	100

There are no Ramachandran outliers to report.

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	96/97 (99%)	94 (98%)	2 (2%)	53	79
1	B	95/97 (98%)	94 (99%)	1 (1%)	73	90
All	All	191/194 (98%)	188 (98%)	3 (2%)	62	84

All (3) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	31	CYS
1	A	32	SER
1	B	31	CYS

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

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)

9 ligands are modelled in this entry.

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

Unable to reproduce the depositors R factor - this section is therefore empty.

6.2 Non-standard residues in protein, DNA, RNA chains

Unable to reproduce the depositors R factor - this section is therefore empty.

6.3 Carbohydrates

Unable to reproduce the depositors R factor - this section is therefore empty.

6.4 Ligands

Unable to reproduce the depositors R factor - this section is therefore empty.

6.5 Other polymers

Unable to reproduce the depositors R factor - this section is therefore empty.