



wwPDB X-ray Structure Validation Summary Report

May 22, 2014 – 01:17 AM EDT

PDB ID : 4OSL
Title : Crystal structure of TAL effector reveals the recognition between histidine and guanine
Authors : Deng, D.; Wu, J.P.; Yan, C.Y.; Pan, X.J.; Yan, N.
Deposited on : 2014-02-13
Resolution : 2.45 Å(reported)

This is a wwPDB 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/ValidationPDFNotes.html>

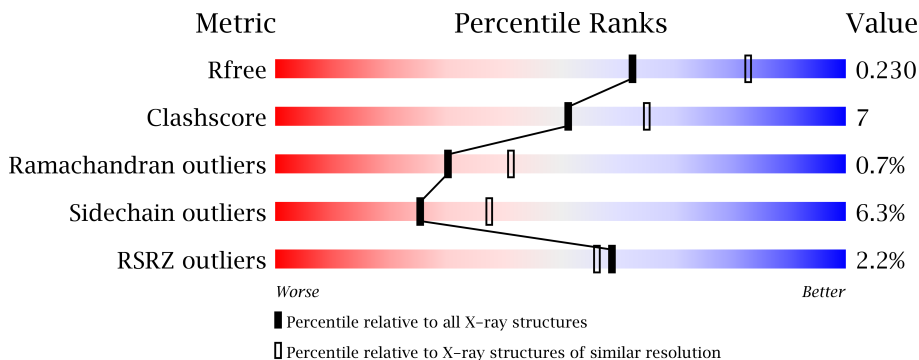
The following versions of software and data (see [references](#)) were used in the production of this report:

MolProbity : 4.02b-467
Mogul : 1.16 November 2013
Xtriage (Phenix) : dev-1439
EDS : stable22978
Percentile statistics : 21963
Refmac : 5.8.0049
CCP4 : 6.3.0 (Settle)
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et. al. (1996)
Validation Pipeline (wwPDB-VP) : stable22978

1 Overall quality at a glance

The reported resolution of this entry is 2.45 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}	66092	2989 (2.48-2.40)
Clashscore	79885	3698 (2.48-2.40)
Ramachandran outliers	78287	3639 (2.48-2.40)
Sidechain outliers	78261	3640 (2.48-2.40)
RSRZ outliers	66119	2993 (2.48-2.40)

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. The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density.

Mol	Chain	Length	Quality of chain
1	A	499	
1	B	499	
2	G	17	
2	I	17	
3	H	17	
3	J	17	

The following table lists non-polymeric compounds that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Geometry	Electron density
4	MG	B	801	-	X

2 Entry composition

There are 5 unique types of molecules in this entry. The entry contains 8937 atoms, of which 0 are hydrogen and 0 are deuterium.

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 Hax3.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	492	Total	C	N	O	S	0	0	0
			3566	2229	664	661	12			
1	B	488	Total	C	N	O	S	0	7	0
			3585	2239	667	666	13			

There are 54 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	230	MET	-	EXPRESSION TAG	UNP Q3ZD72
A	300	HIS	ASN	ENGINEERED MUTATION	UNP Q3ZD72
A	301	ASP	ILE	ENGINEERED MUTATION	UNP Q3ZD72
A	368	HIS	ASN	ENGINEERED MUTATION	UNP Q3ZD72
A	369	ASP	ILE	ENGINEERED MUTATION	UNP Q3ZD72
A	402	ASN	HIS	ENGINEERED MUTATION	UNP Q3ZD72
A	403	GLY	ASP	ENGINEERED MUTATION	UNP Q3ZD72
A	436	ASN	HIS	ENGINEERED MUTATION	UNP Q3ZD72
A	437	GLY	ASP	ENGINEERED MUTATION	UNP Q3ZD72
A	470	ASN	HIS	ENGINEERED MUTATION	UNP Q3ZD72
A	471	GLY	ASP	ENGINEERED MUTATION	UNP Q3ZD72
A	505	HIS	SER	ENGINEERED MUTATION	UNP Q3ZD72
A	539	GLY	SER	ENGINEERED MUTATION	UNP Q3ZD72
A	572	HIS	ASN	ENGINEERED MUTATION	UNP Q3ZD72
A	573	ASP	SER	ENGINEERED MUTATION	UNP Q3ZD72
A	606	ASN	HIS	ENGINEERED MUTATION	UNP Q3ZD72
A	607	GLY	ASP	ENGINEERED MUTATION	UNP Q3ZD72
A	640	HIS	ASN	ENGINEERED MUTATION	UNP Q3ZD72
A	641	ASP	ILE	ENGINEERED MUTATION	UNP Q3ZD72
A	721	LEU	-	EXPRESSION TAG	UNP Q3ZD72
A	722	GLU	-	EXPRESSION TAG	UNP Q3ZD72
A	723	HIS	-	EXPRESSION TAG	UNP Q3ZD72
A	724	HIS	-	EXPRESSION TAG	UNP Q3ZD72
A	725	HIS	-	EXPRESSION TAG	UNP Q3ZD72
A	726	HIS	-	EXPRESSION TAG	UNP Q3ZD72

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
A	727	HIS	-	EXPRESSION TAG	UNP Q3ZD72
A	728	HIS	-	EXPRESSION TAG	UNP Q3ZD72
B	230	MET	-	EXPRESSION TAG	UNP Q3ZD72
B	300	HIS	ASN	ENGINEERED MUTATION	UNP Q3ZD72
B	301	ASP	ILE	ENGINEERED MUTATION	UNP Q3ZD72
B	368	HIS	ASN	ENGINEERED MUTATION	UNP Q3ZD72
B	369	ASP	ILE	ENGINEERED MUTATION	UNP Q3ZD72
B	402	ASN	HIS	ENGINEERED MUTATION	UNP Q3ZD72
B	403	GLY	ASP	ENGINEERED MUTATION	UNP Q3ZD72
B	436	ASN	HIS	ENGINEERED MUTATION	UNP Q3ZD72
B	437	GLY	ASP	ENGINEERED MUTATION	UNP Q3ZD72
B	470	ASN	HIS	ENGINEERED MUTATION	UNP Q3ZD72
B	471	GLY	ASP	ENGINEERED MUTATION	UNP Q3ZD72
B	505	HIS	SER	ENGINEERED MUTATION	UNP Q3ZD72
B	539	GLY	SER	ENGINEERED MUTATION	UNP Q3ZD72
B	572	HIS	ASN	ENGINEERED MUTATION	UNP Q3ZD72
B	573	ASP	SER	ENGINEERED MUTATION	UNP Q3ZD72
B	606	ASN	HIS	ENGINEERED MUTATION	UNP Q3ZD72
B	607	GLY	ASP	ENGINEERED MUTATION	UNP Q3ZD72
B	640	HIS	ASN	ENGINEERED MUTATION	UNP Q3ZD72
B	641	ASP	ILE	ENGINEERED MUTATION	UNP Q3ZD72
B	721	LEU	-	EXPRESSION TAG	UNP Q3ZD72
B	722	GLU	-	EXPRESSION TAG	UNP Q3ZD72
B	723	HIS	-	EXPRESSION TAG	UNP Q3ZD72
B	724	HIS	-	EXPRESSION TAG	UNP Q3ZD72
B	725	HIS	-	EXPRESSION TAG	UNP Q3ZD72
B	726	HIS	-	EXPRESSION TAG	UNP Q3ZD72
B	727	HIS	-	EXPRESSION TAG	UNP Q3ZD72
B	728	HIS	-	EXPRESSION TAG	UNP Q3ZD72

- Molecule 2 is a DNA chain called DNA (5'-D(*TP*GP*TP*CP*CP*CP*TP*TP*TP*GP*TP*CP*TP*CP*TP*CP*T)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	G	17	Total	C	N	O	P	0	0	0
			334	163	46	109	16			
2	I	17	Total	C	N	O	P	0	0	0
			335	164	46	109	16			

- Molecule 3 is a DNA chain called DNA (5'-D(*AP*GP*AP*GP*AP*GP*AP*CP*AP*AP*AP*GP*GP*GP*AP*CP*A)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	H	17	Total	C	N	O	P	0	0	0
			356	168	81	91	16			
3	J	17	Total	C	N	O	P	0	0	0
			356	168	81	91	16			

- Molecule 4 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	B	1	Total	Mg	0	0
			1	1		
4	J	1	Total	Mg	0	0
			1	1		

- Molecule 5 is water.

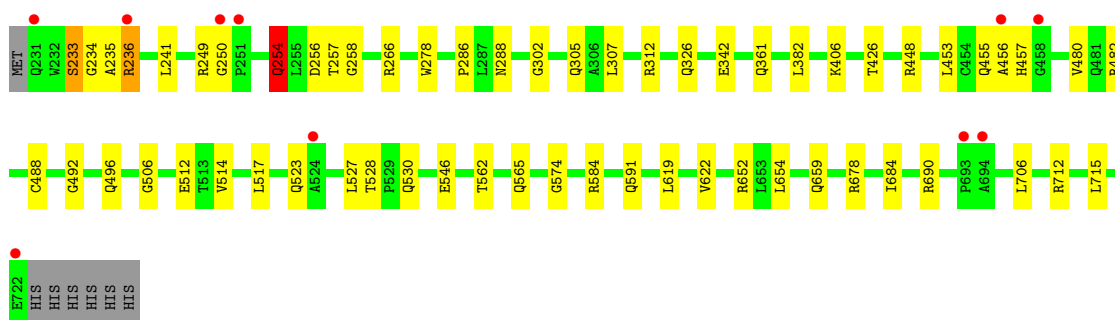
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	144	Total	O	0	0
			144	144		
5	B	169	Total	O	0	0
			169	169		
5	G	31	Total	O	0	0
			31	31		
5	H	15	Total	O	0	0
			15	15		
5	I	26	Total	O	0	0
			26	26		
5	J	18	Total	O	0	0
			18	18		

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.

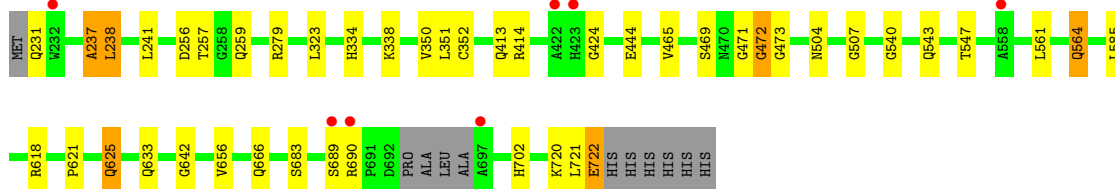
- Molecule 1: Hax3

Chain A: 



- Molecule 1: Hax3

Chain B: 



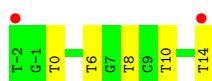
- Molecule 2: DNA (5'-D(*TP*GP*TP*CP*CP*CP*TP*TP*TP*GP*TP*CP*TP*CP*TP*CP*T)-3')

Chain G: 



- Molecule 2: DNA (5'-D(*TP*GP*TP*CP*CP*CP*TP*TP*TP*GP*TP*CP*TP*CP*TP*CP*T)-3')

Chain I: 



- Molecule 3: DNA (5'-D(*AP*GP*AP*GP*AP*GP*AP*CP*AP*AP*AP*GP*GP*GP*AP*CP*A)-3')

Chain H: 



- Molecule 3: DNA (5'-D(*AP*GP*AP*GP*AP*GP*AP*CP*AP*AP*AP*GP*GP*GP*AP*CP*A)-3')

Chain J: 



4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α , β , γ	81.53Å 87.61Å 87.84Å 90.00° 102.72° 90.00°	Depositor
Resolution (Å)	26.55 – 2.45 26.55 – 2.45	Depositor EDS
% Data completeness (in resolution range)	95.9 (26.55-2.45) 95.9 (26.55-2.45)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	4.39 (at 2.44Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.7.3_928)	Depositor
R, R_{free}	0.225 , 0.242 0.207 , 0.230	Depositor DCC
R_{free} test set	2161 reflections (5.05%)	DCC
Wilson B-factor (Å ²)	35.2	Xtriage
Anisotropy	0.172	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.34 , 22.8	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning	$\langle L \rangle = 0.47$, $\langle L^2 \rangle = 0.31$	Xtriage
Outliers	0 of 42801 reflections	Xtriage
F_o, F_c correlation	0.93	EDS
Total number of atoms	8937	wwPDB-VP
Average B, all atoms (Å ²)	37.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.55% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

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	0.41	0/3618	0.58	0/4941
1	B	0.39	0/3636	0.57	1/4962 (0.0%)
2	G	1.07	1/369 (0.3%)	1.31	1/566 (0.2%)
2	I	0.87	0/370	1.53	6/568 (1.1%)
3	H	0.87	0/404	1.39	4/623 (0.6%)
3	J	0.81	0/404	1.39	2/623 (0.3%)
All	All	0.53	1/8801 (0.0%)	0.81	14/12283 (0.1%)

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	G	13	DC	O3'-P	12.98	1.76	1.61

The worst 5 of 14 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	G	13	DC	O4'-C1'-N1	7.97	113.58	108.00
2	I	6	DT	C5-C4-O4	-6.82	120.13	124.90
3	J	-2	DG	O4'-C1'-N9	6.74	112.72	108.00
3	H	-7	DC	O4'-C1'-N1	6.15	112.31	108.00
2	I	8	DT	O4'-C1'-N1	-5.84	103.91	108.00

There are no chirality outliers.

There are no planarity outliers.

5.2 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 hydrogens

added by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, and the number in parentheses is this value normalized per 1000 atoms of the molecule in the chain. The Symm-Clashes column gives symmetry related clashes, in the same way as for the Clashes column.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	3566	0	0	28	0
1	B	3585	0	0	23	0
2	G	334	0	0	2	0
2	I	335	0	0	1	0
3	H	356	0	0	3	0
3	J	356	0	0	1	0
4	B	1	0	0	0	0
4	J	1	0	0	0	0
5	A	144	0	0	15	0
5	B	169	0	0	15	0
5	G	31	0	0	0	0
5	H	15	0	0	0	0
5	I	26	0	0	0	0
5	J	18	0	0	1	0
All	All	8937	0	0	55	0

Clashscore is defined as the number of clashes calculated for the entry per 1000 atoms (including hydrogens) of the entry. The overall clashscore for this entry is 7.

The worst 5 of 55 close contacts within the same asymmetric unit are listed below.

Atom-1	Atom-2	Distance(Å)	Clash(Å)
1:B:722:GLU:O	5:B:957:HOH:O	1.74	1.06
1:B:472:GLY:O	5:B:1054:HOH:O	1.78	1.02
1:A:278:TRP:NE1	5:A:926:HOH:O	2.00	0.93
1:A:236:ARG:CD	1:A:236:ARG:N	2.35	0.89
1:B:352:CYS:O	5:B:992:HOH:O	1.93	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	490/499 (98%)	461 (94%)	25 (5%)	4 (1%)	27	37
1	B	491/499 (98%)	461 (94%)	26 (5%)	4 (1%)	27	37
All	All	981/998 (98%)	922 (94%)	51 (5%)	8 (1%)	30	37

5 of 8 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	234	GLY
1	B	237	ALA
1	B	472	GLY
1	B	424[A]	GLY
1	B	424[B]	GLY

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	376/383 (98%)	349 (93%)	27 (7%)	21	30
1	B	378/383 (99%)	358 (95%)	20 (5%)	32	47
All	All	754/766 (98%)	707 (94%)	47 (6%)	25	38

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

Mol	Chain	Res	Type
1	A	659	GLN
1	A	715	LEU
1	B	702	HIS
1	A	684	ILE
1	B	231	GLN

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

5.3.3 RNA ⓘ

There are no RNA chains 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 carbohydrates in this entry.

5.6 Ligand geometry ⓘ

Of 2 ligands modelled in this entry, 2 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.

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(Å ²)	Q<0.9
1	A	492/499 (98%)	-0.05	10 (2%) 62 59	14, 34, 71, 112	6 (1%)
1	B	488/499 (97%)	-0.08	7 (1%) 72 71	12, 33, 61, 85	7 (1%)
2	G	17/17 (100%)	-0.27	1 (5%) 22 19	18, 23, 57, 115	0
2	I	17/17 (100%)	-0.33	2 (11%) 5 4	18, 23, 69, 78	0
3	H	17/17 (100%)	0.56	2 (11%) 5 4	30, 38, 71, 81	0
3	J	17/17 (100%)	0.41	1 (5%) 22 19	29, 43, 65, 71	0
All	All	1048/1066 (98%)	-0.05	23 (2%) 59 56	12, 33, 67, 115	13 (1%)

The worst 5 of 23 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	456	ALA	4.8
2	G	14	DT	4.5
3	H	-14	DA	4.2
1	B	423[A]	HIS	3.2
1	A	693	PRO	3.0

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, 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	RSR	LLDF	B-factors(Å ²)	Q<0.9
4	MG	B	801	1/1	0.21	2.56	65,65,65,65	0
4	MG	J	101	1/1	0.17	0.32	81,81,81,81	0

6.5 Other polymers

There are no such residues in this entry.