



# wwPDB X-ray Structure Validation Summary Report

May 22, 2014 – 01:30 AM EDT

PDB ID : 4OSJ  
Title : Crystal structure of TAL effector reveals the recognition between asparagine and adenine  
Authors : Deng, D.; Wu, J.P.; Yan, C.Y.; Pan, X.J.; Yan, N.  
Deposited on : 2014-02-13  
Resolution : 2.79 Å(reported)

This is a wwPDB validation summary report for a publicly released PDB entry.

We welcome your comments at [validation@mail.wwpdb.org](mailto:validation@mail.wwpdb.org)

A user guide is available at <http://wwpdb.org/ValidationPDFNotes.html>

---

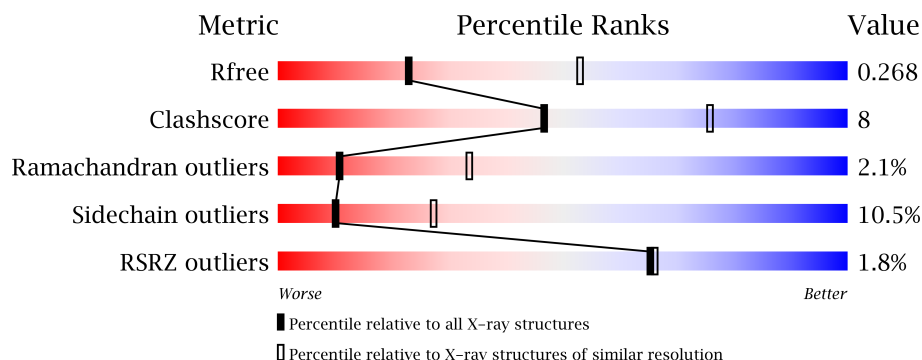
The following versions of software and data (see [references](#)) were used in the production of this report:

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 2.79 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	1799 (2.80-2.80)
Clashscore	79885	2295 (2.80-2.80)
Ramachandran outliers	78287	2252 (2.80-2.80)
Sidechain outliers	78261	2254 (2.80-2.80)
RSRZ outliers	66119	1802 (2.80-2.80)

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  $\geq 3$ , 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	

## 2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 8390 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	487	Total	C	N	O	S	15	0	0
			3514	2197	650	655	12			
1	B	489	Total	C	N	O	S	13	0	0
			3532	2206	658	656	12			

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	ASN	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	ASN	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\*AP\*TP\*CP\*TP\*CP\*TP\*CP\*T)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	I	16	Total	C	N	O	P	0	0	0
			316	153	44	103	16			
2	G	16	Total	C	N	O	P	0	0	0
			314	154	44	101	15			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	J	16	Total 336	C 159	N 75	O 87	P 15	0	0	0
3	H	16	Total 339	C 159	N 75	O 89	P 16	0	0	0

- Molecule 4 is water.

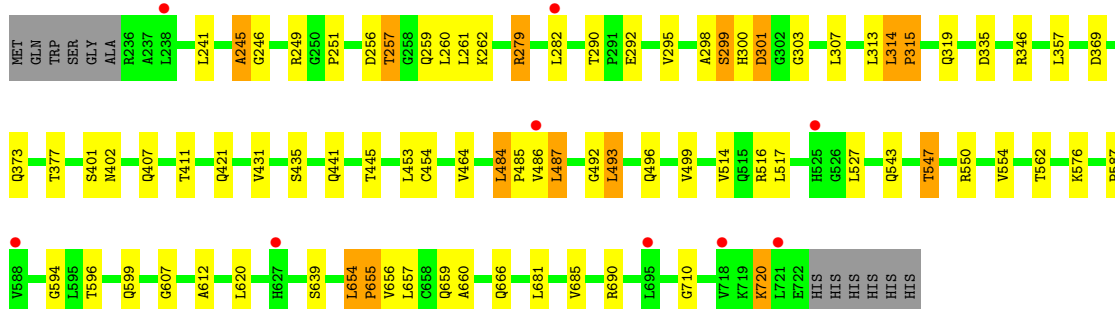
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	8	Total 8	O 8	0	0
4	I	4	Total 4	O 4	0	0
4	B	20	Total 20	O 20	0	0
4	G	7	Total 7	O 7	0	0

### 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.

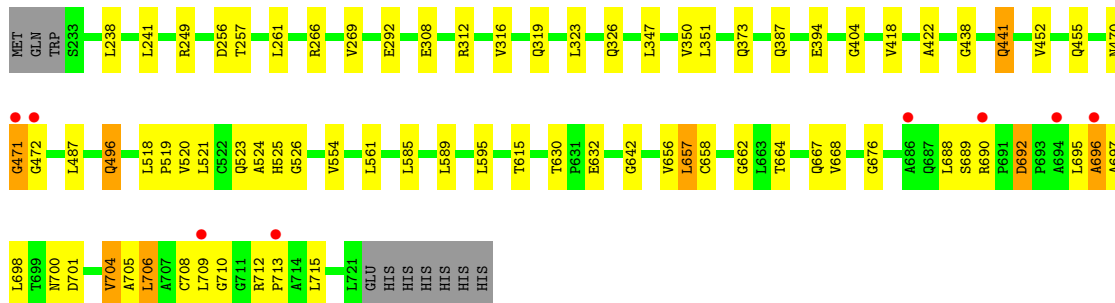
- Molecule 1: Hax3

Chain A: 



- Molecule 1: Hax3

Chain B: 



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

Chain I: 



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

Chain G: 



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

Chain J: A horizontal bar representing the sequence of Chain J. It is primarily green, with a yellow segment towards the right end and a grey segment at the far right.



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

Chain H: A horizontal bar representing the sequence of Chain H. It is primarily green, with a red segment at the beginning, a yellow segment towards the right end, and a grey segment at the far right.



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	84.41Å 80.81Å 88.82Å 90.00° 103.68° 90.00°	Depositor
Resolution (Å)	38.06 – 2.79 38.06 – 2.79	Depositor EDS
% Data completeness (in resolution range)	98.0 (38.06-2.79) 98.1 (38.06-2.79)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.51 (at 2.81Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.7.3_928)	Depositor
R, $R_{free}$	0.214 , 0.271 0.212 , 0.268	Depositor DCC
$R_{free}$ test set	1442 reflections (5.06%)	DCC
Wilson B-factor (Å <sup>2</sup> )	46.2	Xtriage
Anisotropy	0.578	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.30 , 19.2	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.32$	Xtriage
Outliers	0 of 28520 reflections	Xtriage
$F_o, F_c$ correlation	0.93	EDS
Total number of atoms	8390	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	54.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 7.41% of the height of the origin peak. No significant pseudotranslation is detected.*

<sup>1</sup>Intensities estimated from amplitudes.



## 5 Model quality

### 5.1 Standard geometry

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	0.40	0/3563	0.61	3/4868 (0.1%)
1	B	0.45	0/3581	0.66	3/4890 (0.1%)
2	G	0.78	0/347	1.56	7/532 (1.3%)
2	I	0.77	0/349	1.58	4/534 (0.7%)
3	H	0.73	0/384	1.48	2/592 (0.3%)
3	J	0.90	0/381	1.67	5/588 (0.9%)
All	All	0.51	0/8605	0.89	24/12004 (0.2%)

There are no bond length outliers.

The worst 5 of 24 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	698	LEU	O-C-N	15.47	147.44	122.70
1	A	245	ALA	N-CA-C	12.38	144.43	111.00
1	B	698	LEU	CA-C-N	-11.99	90.81	117.20
3	J	10	DG	O4'-C1'-N9	9.73	114.81	108.00
3	J	13	DA	O4'-C4'-C3'	-8.93	100.64	106.00

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	3514	0	0	34	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	B	3532	0	0	25	0
2	G	314	0	0	5	0
2	I	316	0	0	1	0
3	H	339	0	0	2	0
3	J	336	0	0	0	0
4	A	8	0	0	0	0
4	B	20	0	0	3	0
4	G	7	0	0	1	0
4	I	4	0	0	0	0
All	All	8390	0	0	62	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 8.

The worst 5 of 62 close contacts within the same asymmetric unit are listed below.

Atom-1	Atom-2	Distance(Å)	Clash(Å)
1:A:262:LYS:NZ	1:A:295:VAL:CG1	1.98	1.25
1:B:657:LEU:O	1:B:657:LEU:CD2	2.30	0.80
1:A:486:VAL:O	1:A:487:LEU:CD2	2.31	0.79
1:A:314:LEU:O	1:A:314:LEU:CD1	2.30	0.78
1:A:486:VAL:O	1:A:486:VAL:CG1	2.30	0.78

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	485/499 (97%)	423 (87%)	48 (10%)	14 (3%)	7	23
1	B	487/499 (98%)	448 (92%)	33 (7%)	6 (1%)	19	54
All	All	972/998 (97%)	871 (90%)	81 (8%)	20 (2%)	11	33

5 of 20 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	493	LEU
1	B	696	ALA
1	A	607	GLY
1	A	655	PRO
1	B	471	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	371/383 (97%)	340 (92%)	31 (8%)	16	41
1	B	373/383 (97%)	326 (87%)	47 (13%)	7	19
All	All	744/766 (97%)	666 (90%)	78 (10%)	10	27

5 of 78 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	257	THR
1	B	326	GLN
1	B	690	ARG
1	B	261	LEU
1	B	292	GLU

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 ⓘ

There are no ligands in this entry.

## 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, 95<sup>th</sup> 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 <sup>2</sup> )	Q<0.9
1	A	487/499 (97%)	0.03	9 (1%) 65 66	29, 56, 92, 126	14 (2%)
1	B	489/499 (97%)	-0.08	8 (1%) 68 69	22, 44, 86, 144	11 (2%)
2	G	16/17 (94%)	-0.31	1 (6%) 19 18	34, 38, 99, 110	0
2	I	16/17 (94%)	-0.46	0 100 100	34, 47, 87, 138	0
3	H	16/17 (94%)	0.09	1 (6%) 19 18	37, 54, 116, 138	0
3	J	16/17 (94%)	-0.33	0 100 100	49, 58, 97, 119	0
All	All	1040/1066 (97%)	-0.04	19 (1%) 65 66	22, 50, 92, 144	25 (2%)

The worst 5 of 19 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	721	LEU	4.2
3	H	15	DA	4.0
1	A	718	VAL	3.1
1	B	696	ALA	2.8
1	A	486	VAL	2.7

### 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 ⓘ

There are no ligands in this entry.

## 6.5 Other polymers ⓘ

There are no such residues in this entry.