wwPDB X-ray Structure Validation Summary Report (i)

May 23, 2020 – 08:22 am BST

PDB ID	:	4GG4
Title	:	Crystal structure of the TAL effector dHax3 bound to specific DNA-RNA hybrid
Authors	:	Yin, P.; Deng, D.; Yan, C.Y.; Pan, X.J.; Yan, N.; Shi, Y.G.
Deposited on	:	2012-08-05
Resolution	:	2.50 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
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 (1)) were used in the production of this report:

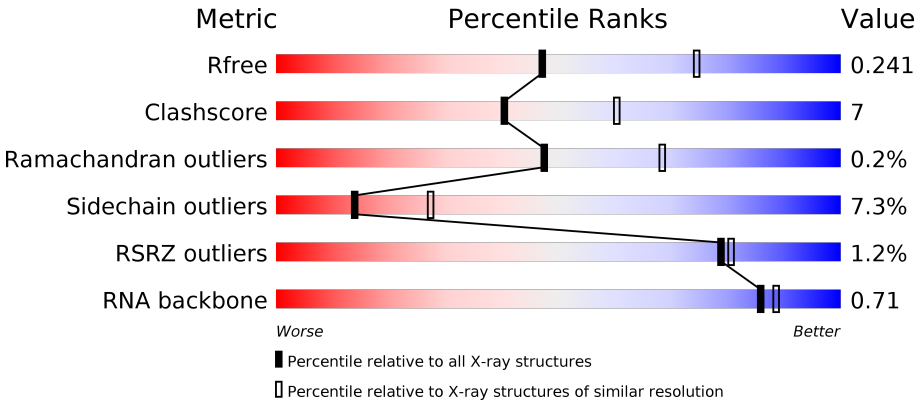
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.11

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 2.50 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	4661 (2.50-2.50)
Clashscore	141614	5346 (2.50-2.50)
Ramachandran outliers	138981	5231 (2.50-2.50)
Sidechain outliers	138945	5233 (2.50-2.50)
RSRZ outliers	127900	4559 (2.50-2.50)
RNA backbone	3102	1008 (2.84-2.16)

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 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% 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	499	% 82% 14%
2	G	17	6% 65% 35%
3	H	17	6% 71% 24% 6%

2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 4265 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 Hax3.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	488	3522	2203	656	651	12	17	0	0

There are 27 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	ILE	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
A	727	HIS	-	EXPRESSION TAG	UNP Q3ZD72
A	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	G	17	Total	C	N	O	P	0	0	0
			333	163	46	108	16			

- Molecule 3 is a RNA chain called RNA (5'-R(*AP*GP*AP*GP*AP*GP*AP*UP*AP*AP*AP*GP*GP*GP*AP*CP*A)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	H	16	Total	C	N	O	P	0	0	0
			354	158	75	105	16			

- Molecule 4 is water.

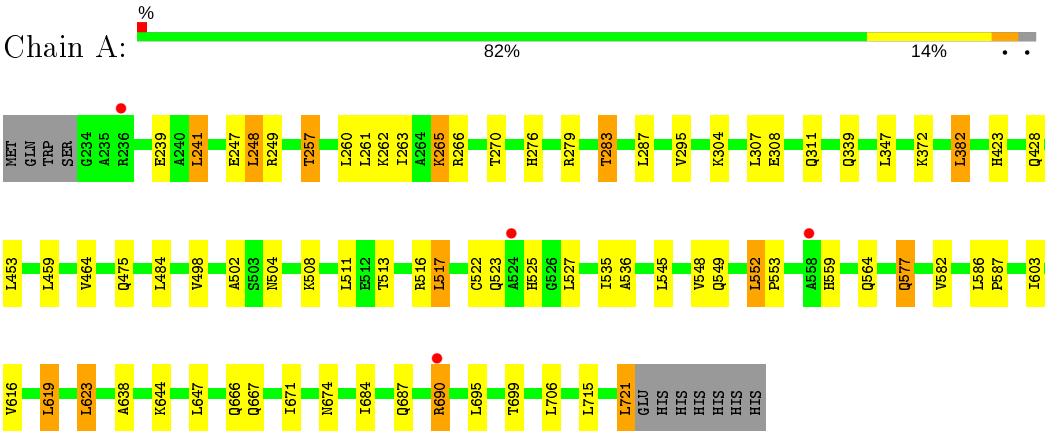
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	35	Total	O	0	0
			35	35		
4	G	21	Total	O	0	0
			21	21		

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 the various outlier classes 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.

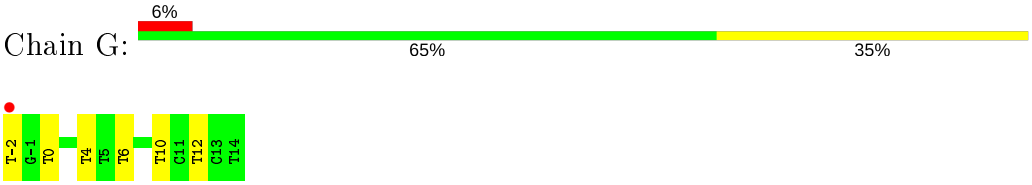
• Molecule 1: Hax3

Chain A:



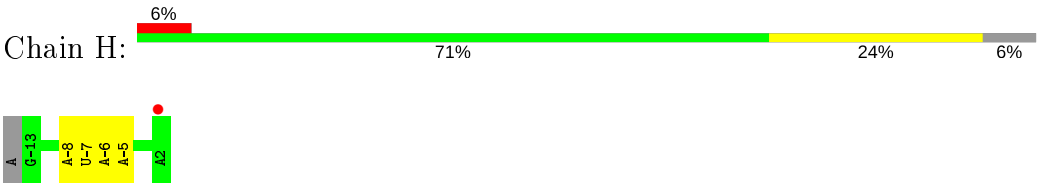
• Molecule 2: DNA (5'-D(*TP*GP*TP*CP*CP*CP*TP*TP*TP*AP*TP*CP*TP*CP*TP*CP*T)-3')

Chain G:



• Molecule 3: RNA (5'-R(*AP*GP*AP*GP*AP*GP*AP*UP*AP*AP*AP*GP*GP*GP*AP*CP*A)-3')

Chain H:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 61	Depositor
Cell constants a, b, c, α, β, γ	99.73Å 99.73Å 134.49Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	31.73 – 2.50 31.72 – 2.50	Depositor EDS
% Data completeness (in resolution range)	98.4 (31.73-2.50) 98.4 (31.72-2.50)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.74 (at 2.51Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.7.3_928)	Depositor
R, R_{free}	0.193 , 0.243 0.194 , 0.241	Depositor DCC
R_{free} test set	1328 reflections (5.15%)	wwPDB-VP
Wilson B-factor (Å^2)	47.4	Xtriage
Anisotropy	0.373	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.34 , 35.4	EDS
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.33$	Xtriage
Estimated twinning fraction	0.059 for h,-h-k,-l	Xtriage
F_o,F_c correlation	0.96	EDS
Total number of atoms	4265	wwPDB-VP
Average B, all atoms (Å^2)	50.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.69% 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

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	0.37	0/3571	0.56	1/4877 (0.0%)
2	G	0.86	0/368	1.64	8/564 (1.4%)
3	H	0.64	0/399	1.00	0/622
All	All	0.46	0/4338	0.78	9/6063 (0.1%)

There are no bond length outliers.

The worst 5 of 9 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	G	12	DT	N3-C4-O4	7.35	124.31	119.90
1	A	382	LEU	CA-CB-CG	6.41	130.03	115.30
2	G	4	DT	C6-C5-C7	-5.67	119.50	122.90
2	G	12	DT	C5-C4-O4	-5.64	120.95	124.90
2	G	6	DT	N3-C4-O4	5.30	123.08	119.90

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	3522	0	3686	55	0
2	G	333	0	195	0	0
3	H	354	0	176	5	0
4	A	35	0	0	2	0
4	G	21	0	0	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
All	All	4265	0	4057	58	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.

The worst 5 of 58 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:611:GLN:HB3	1:A:644:LYS:HD2	1.60	0.83
1:A:270:THR:HG22	1:A:304:LYS:HB2	1.67	0.76
1:A:674:ASN:ND2	4:A:835:HOH:O	2.17	0.72
1:A:265:LYS:HD3	1:A:265:LYS:O	1.94	0.67
1:A:523:GLN:OE1	1:A:523:GLN:N	2.30	0.65

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	486/499 (97%)	460 (95%)	25 (5%)	1 (0%)	47 68

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	553	PRO

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	371/383 (97%)	344 (93%)	27 (7%)	14 27

5 of 27 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	459	LEU
1	A	527	LEU
1	A	690	ARG
1	A	516	ARG
1	A	261	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (5) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	311	GLN
1	A	321	HIS
1	A	538	ASN
1	A	577	GLN
1	A	625	GLN

5.3.3 RNA

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
3	H	15/17 (88%)	0	0

There are no RNA backbone outliers to report.

There are no RNA pucker outliers to report.

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^{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	488/499 (97%)	-0.30	4 (0%)	86	87	27, 48, 79, 117	13 (2%)
2	G	17/17 (100%)	-0.38	1 (5%)	22	23	29, 33, 84, 117	0
3	H	16/17 (94%)	-0.76	1 (6%)	20	21	34, 37, 66, 104	0
All	All	521/533 (97%)	-0.32	6 (1%)	79	80	27, 47, 79, 117	13 (2%)

The worst 5 of 6 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	236	ARG	2.9
1	A	690	ARG	2.9
1	A	558	ALA	2.6
2	G	-2	DT	2.6
1	A	524	ALA	2.3

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 (i)

There are no such residues in this entry.