



wwPDB X-ray Structure Validation Summary Report i

Feb 27, 2014 – 07:57 AM GMT

PDB ID : 4HC7
Title : Crystal structure of the full DNA binding domain of GATA3-complex 2
Authors : Chen, Y.; Bates, D.L.; Dey, R.; Chen, L.
Deposited on : 2012-09-28
Resolution : 2.65 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>

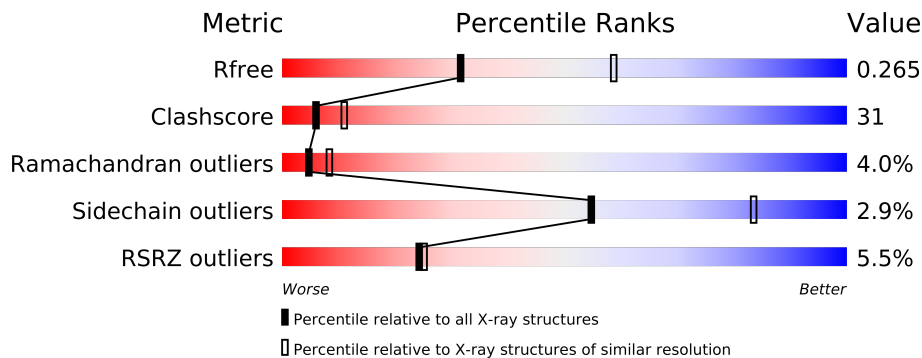
The following versions of software and data (see [references](#)) were used in the production of this report:

MolProbity	:	4.02b-467
Mogul	:	1.15 2013
Xtriage (Phenix)	:	dev-1323
EDS	:	stable22639
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)	:	stable22683

1 Overall quality at a glance

The reported resolution of this entry is 2.65 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	2232 (2.70-2.62)
Clashscore	79885	2700 (2.70-2.62)
Ramachandran outliers	78287	2657 (2.70-2.62)
Sidechain outliers	78261	2657 (2.70-2.62)
RSRZ outliers	66119	2234 (2.70-2.62)

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	115	
1	B	115	
2	W	20	
2	Y	20	
3	X	20	
3	Z	20	

2 Entry composition

There are 5 unique types of molecules in this entry. The entry contains 3442 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 Trans-acting T-cell-specific transcription factor GATA-3.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	108	Total	C	N	O	S	0	0	0
			843	510	174	148	11			
1	B	97	Total	C	N	O	S	0	0	0
			754	457	150	136	11			

There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	257	GLY	-	EXPRESSION TAG	UNP P23771
A	258	SER	-	EXPRESSION TAG	UNP P23771
A	259	HIS	-	EXPRESSION TAG	UNP P23771
A	260	MET	-	EXPRESSION TAG	UNP P23771
B	257	GLY	-	EXPRESSION TAG	UNP P23771
B	258	SER	-	EXPRESSION TAG	UNP P23771
B	259	HIS	-	EXPRESSION TAG	UNP P23771
B	260	MET	-	EXPRESSION TAG	UNP P23771

- Molecule 2 is a DNA chain called DNA (5'-D(*TP*TP*CP*CP*TP*AP*AP*AP*TP*CP*AP*GP*AP*GP*AP*TP*AP*AP*CP*C)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	W	20	Total	C	N	O	P	0	0	0
			404	195	75	115	19			
2	Y	20	Total	C	N	O	P	0	0	0
			404	195	75	115	19			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	X	20	Total	C	N	O	P	0	0	0
			410	198	72	121	19			

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	Z	20	Total	C	N	O	P	0	0	0
			410	198	72	121	19			

- Molecule 4 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	B	2	Total	Zn	0	0
			2	2		
4	A	2	Total	Zn	0	0
			2	2		

- Molecule 5 is water.

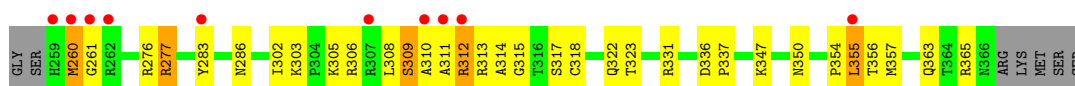
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	54	Total	O	0	0
			54	54		
5	W	23	Total	O	0	0
			23	23		
5	X	25	Total	O	0	0
			25	25		
5	B	46	Total	O	0	0
			46	46		
5	Y	30	Total	O	0	0
			30	30		
5	Z	35	Total	O	0	0
			35	35		

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.

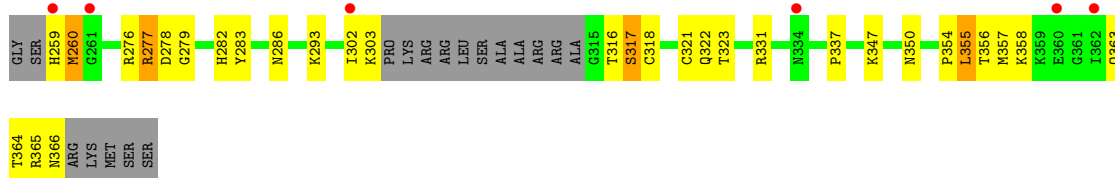
- Molecule 1: Trans-acting T-cell-specific transcription factor GATA-3

Chain A: 



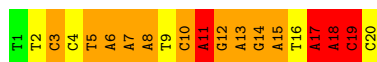
- Molecule 1: Trans-acting T-cell-specific transcription factor GATA-3

Chain B: 



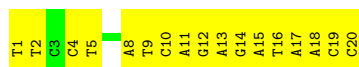
- Molecule 2: DNA (5'-D(*TP*TP*CP*CP*TP*AP*AP*AP*TP*CP*AP*GP*AP*GP*AP*TP*AP*AP*CP*C)-3')

Chain W: 



- Molecule 2: DNA (5'-D(*TP*TP*CP*CP*TP*AP*AP*AP*TP*CP*AP*GP*AP*GP*AP*TP*AP*AP*CP*C)-3')

Chain Y: 



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

Chain X: 



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

Chain Z: 

A1	A2	G3	G4	A7	T8	G9	T10	C11	T12	G13	A14	T15	T16	T17	A18	G19	G20
----	----	----	----	----	----	----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----

4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α , β , γ	65.77Å 65.37Å 72.92Å 90.00° 96.75° 90.00°	Depositor
Resolution (Å)	30.00 – 2.65 31.17 – 2.58	Depositor EDS
% Data completeness (in resolution range)	87.4 (30.00-2.65) 95.1 (31.17-2.58)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.74 (at 2.57Å)	Xtriage
Refinement program	CNS	Depositor
R, R_{free}	0.226 , 0.264 0.227 , 0.265	Depositor DCC
R_{free} test set	1817 reflections (9.83%)	DCC
Wilson B-factor (Å ²)	59.5	Xtriage
Anisotropy	0.471	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.30 , 41.7	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning	$\langle L \rangle = 0.50$, $\langle L^2 \rangle = 0.33$	Xtriage
Outliers	4 of 19436 reflections (0.021%)	Xtriage
F_o, F_c correlation	0.94	EDS
Total number of atoms	3442	wwPDB-VP
Average B, all atoms (Å ²)	60.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 8.52% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

5 Model quality i

5.1 Standard geometry i

Bond lengths and bond angles in the following residue types are not validated in this section: ZN

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	0.30	0/859	0.57	0/1158
1	B	0.31	0/768	0.58	0/1036
2	W	0.52	0/453	1.36	6/696 (0.9%)
2	Y	0.43	0/453	0.87	1/696 (0.1%)
3	X	0.43	0/459	0.82	0/708
3	Z	0.43	0/459	0.84	0/708
All	All	0.39	0/3451	0.84	7/5002 (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
2	W	0	12

There are no bond length outliers.

The worst 5 of 7 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed($^{\circ}$)	Ideal($^{\circ}$)
2	W	18	DA	N9-C1'-C2'	6.50	124.95	112.60
2	W	17	DA	N9-C1'-C2'	6.23	124.44	112.60
2	W	7	DA	N9-C1'-C2'	5.38	122.81	112.60
2	W	19	DC	N1-C1'-C2'	5.31	122.69	112.60
2	W	11	DA	N9-C1'-C2'	5.27	122.62	112.60

There are no chirality outliers.

5 of 12 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
2	W	10	DC	Sidechain
2	W	3	DC	Sidechain
2	W	5	DT	Sidechain
2	W	6	DA	Sidechain
2	W	8	DA	Sidechain

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	843	0	833	39	0
1	B	754	0	729	34	0
2	W	404	0	227	32	0
2	Y	404	0	227	24	0
3	X	410	0	230	30	0
3	Z	410	0	230	24	0
4	A	2	0	0	0	0
4	B	2	0	0	0	0
5	A	54	0	0	3	0
5	B	46	0	0	4	0
5	W	23	0	0	2	0
5	X	25	0	0	3	0
5	Y	30	0	0	3	0
5	Z	35	0	0	1	0
All	All	3442	0	2476	177	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 31.

The worst 5 of 177 close contacts within the same asymmetric unit are listed below.

Atom-1	Atom-2	Distance(Å)	Clash(Å)
3:X:18:DA:H2''	3:X:19:DG:H5''	1.37	1.02
3:X:10:DT:H2''	3:X:11:DC:H5''	1.46	0.94
1:A:308:LEU:HD12	1:A:312:ARG:HG3	1.55	0.85
3:X:18:DA:C2'	3:X:19:DG:H5''	2.08	0.83
3:Z:11:DC:H2'	3:Z:12:DT:H71	1.64	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	106/115 (92%)	92 (87%)	9 (8%)	5 (5%)	4	6
1	B	93/115 (81%)	86 (92%)	4 (4%)	3 (3%)	6	12
All	All	199/230 (86%)	178 (89%)	13 (6%)	8 (4%)	5	8

5 of 8 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	260	MET
1	B	317	SER
1	A	309	SER
1	A	310	ALA
1	A	312	ARG

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	90/96 (94%)	87 (97%)	3 (3%)	50	80
1	B	82/96 (85%)	80 (98%)	2 (2%)	61	88
All	All	172/192 (90%)	167 (97%)	5 (3%)	55	83

All (5) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	260	MET

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	A	277	ARG
1	A	350	ASN
1	B	277	ARG
1	B	350	ASN

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 12 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	350	ASN
1	B	282	HIS
1	B	298	ASN
1	A	349	HIS
1	B	292	HIS

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 4 ligands modelled in this entry, 4 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	108/115 (93%)	0.53	10 (9%) 9 8	36, 65, 96, 140	0
1	B	97/115 (84%)	0.36	6 (6%) 20 21	38, 64, 106, 119	0
2	W	20/20 (100%)	-0.19	0 100 100	42, 52, 71, 79	0
2	Y	20/20 (100%)	-0.35	0 100 100	41, 47, 71, 73	0
3	X	20/20 (100%)	-0.11	0 100 100	36, 50, 72, 77	0
3	Z	20/20 (100%)	-0.29	0 100 100	36, 50, 67, 69	0
All	All	285/310 (91%)	0.26	16 (5%) 24 24	36, 60, 96, 140	0

The worst 5 of 16 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	260	MET	5.6
1	A	259	HIS	4.1
1	A	355	LEU	3.6
1	A	283	TYR	3.3
1	B	360	GLU	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	ZN	A	401	1/1	0.16	-0.07	46,46,46,46	0
4	ZN	B	402	1/1	0.15	-0.30	53,53,53,53	0
4	ZN	A	402	1/1	0.13	-0.84	55,55,55,55	0
4	ZN	B	401	1/1	0.11	-0.99	70,70,70,70	0

6.5 Other polymers

There are no such residues in this entry.