wwPDB X-ray Structure Validation Summary Report (i)

Sep 24, 2023 – 04:12 AM EDT

PDB ID	:	5SY7
Title	:	Crystal Structure of the Heterodimeric NPAS3-ARNT Complex with HRE DNA
Authors	:	Wu, D.; Su, X.; Potluri, N.; Kim, Y.; Rastinejad, F.
Deposited on	:	2016-08-10
Resolution	:	4.20 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 types of validation reports are described at
http://www.wwpdb.org/validation/2017/FAQs#types.

The following versions of software and data (see references (1)) were used in the production of this report:

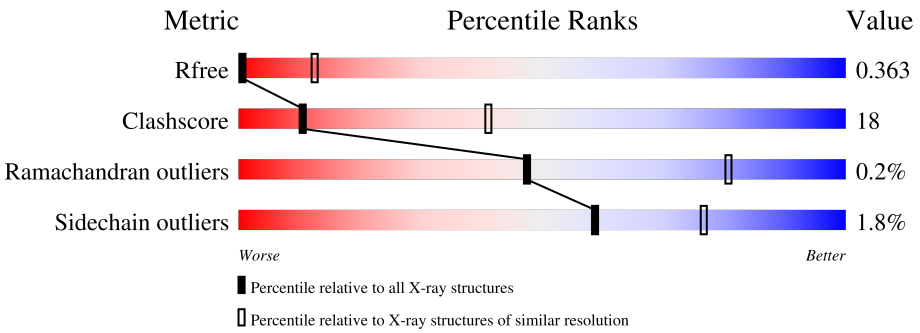
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.35.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.35.1

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 4.20 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	1005 (4.62-3.78)
Clashscore	141614	1044 (4.60-3.80)
Ramachandran outliers	138981	1000 (4.60-3.80)
Sidechain outliers	138945	1007 (4.62-3.78)

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	384	41% 30% • 28%
2	B	410	37% 30% • 33%
3	C	21	33% 67%
4	D	21	48% 48% 5%

2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 5313 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 Aryl hydrocarbon receptor nuclear translocator.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	276	2225	1402	394	414	15	0	0	0

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	81	MET	-	initiating methionine	UNP P53762

- Molecule 2 is a protein called Neuronal PAS domain-containing protein 3.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	B	276	2233	1432	385	403	13	0	0	0

There are 11 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	54	MET	-	initiating methionine	UNP Q9QZQ0
B	55	ILE	-	expression tag	UNP Q9QZQ0
B	302	SER	PRO	conflict	UNP Q9QZQ0
B	456	VAL	-	expression tag	UNP Q9QZQ0
B	457	GLU	-	expression tag	UNP Q9QZQ0
B	458	HIS	-	expression tag	UNP Q9QZQ0
B	459	HIS	-	expression tag	UNP Q9QZQ0
B	460	HIS	-	expression tag	UNP Q9QZQ0
B	461	HIS	-	expression tag	UNP Q9QZQ0
B	462	HIS	-	expression tag	UNP Q9QZQ0
B	463	HIS	-	expression tag	UNP Q9QZQ0

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
3	C	21	433	205	80	128	20	0	0	0

- Molecule 4 is a DNA chain called DNA (5'-D(*CP*AP*CP*GP*AP*CP*CP*CP*GP*CP*AP*CP*GP*TP*AP*CP*GP*CP*AP*GP*C)-3').

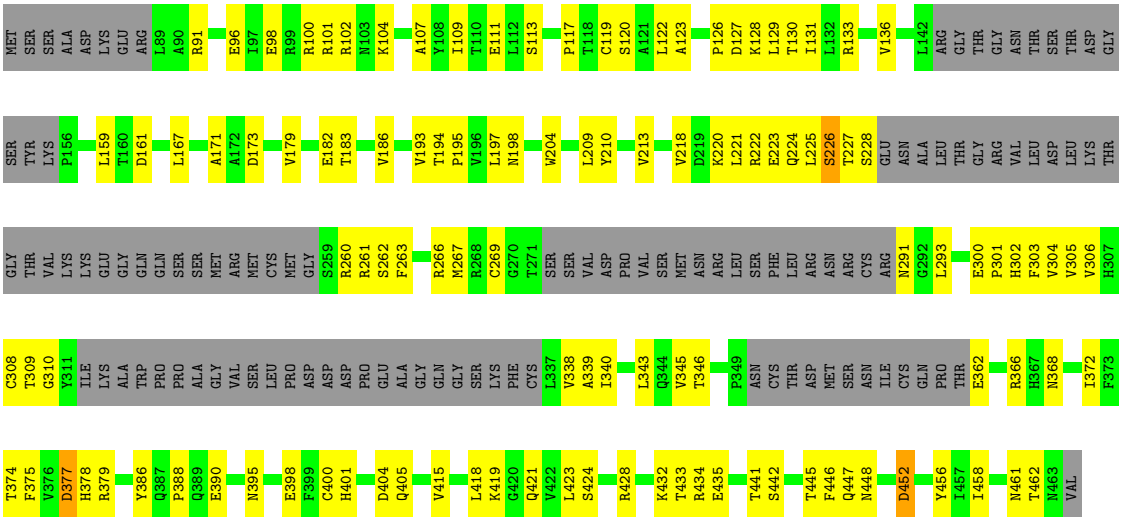
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
4	D	21	422	200	82	120	20	0	0	0

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.

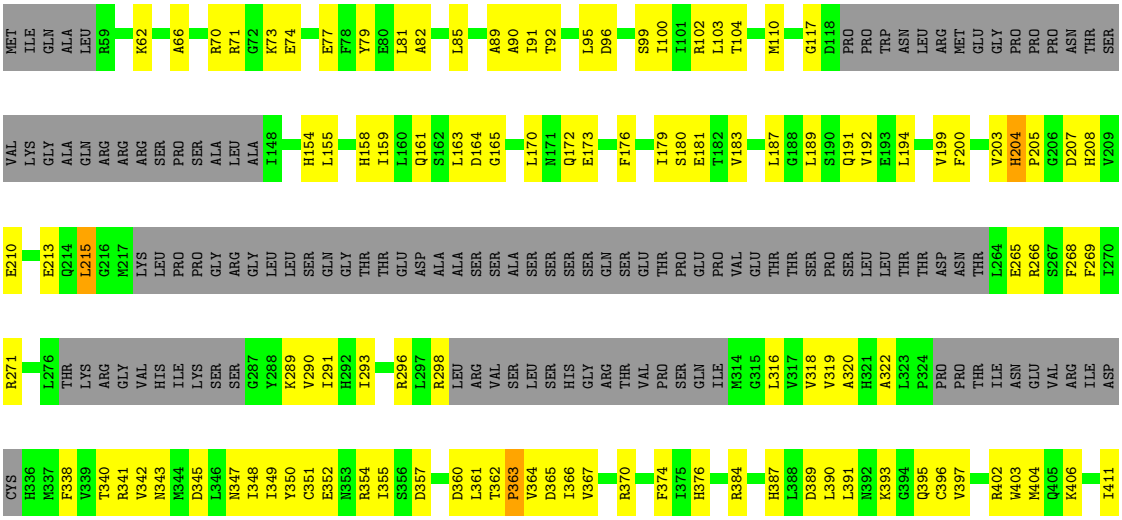
- Molecule 1: Aryl hydrocarbon receptor nuclear translocator

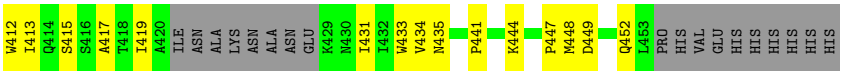
Chain A: 41% 30% • 28%



- Molecule 2: Neuronal PAS domain-containing protein 3

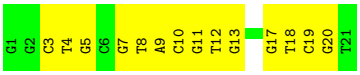
Chain B: 37% 30% • 33%





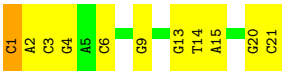
• Molecule 3: DNA (5'-D(*GP*GP*CP*TP*GP*CP*GP*TP*AP*CP*GP*TP*GP*CP*GP*GP *GP*TP*CP*GP*T)-3')

Chain C:



• Molecule 4: DNA (5'-D(*CP*AP*CP*GP*AP*CP*CP*CP*GP*CP*AP*CP*GP*TP*AP*CP *GP*CP*AP*GP*C)-3')

Chain D:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 43	Depositor
Cell constants a, b, c, α, β, γ	64.83Å 64.83Å 249.10Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	45.84 – 4.20 45.84 – 4.19	Depositor EDS
% Data completeness (in resolution range)	77.7 (45.84-4.20) 77.9 (45.84-4.19)	Depositor EDS
R_{merge}	0.06	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	6.51 (at 4.14Å)	Xtriage
Refinement program	PHENIX 1.10.1_2155	Depositor
R, R_{free}	0.272 , 0.361 0.273 , 0.363	Depositor DCC
R_{free} test set	277 reflections (4.71%)	wwPDB-VP
Wilson B-factor (Å^2)	37.5	Xtriage
Anisotropy	0.806	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.32 , 93.8	EDS
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.31$	Xtriage
Estimated twinning fraction	0.099 for h,-k,-l	Xtriage
F_o,F_c correlation	0.76	EDS
Total number of atoms	5313	wwPDB-VP
Average B, all atoms (Å^2)	66.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.86% 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 (i)

5.1 Standard geometry (i)

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	0.28	0/2267	0.49	2/3057 (0.1%)
2	B	0.29	0/2276	0.49	0/3063
3	C	0.63	0/485	0.94	0/749
4	D	0.67	0/473	0.94	1/726 (0.1%)
All	All	0.38	0/5501	0.60	3/7595 (0.0%)

There are no bond length outliers.

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	226	SER	N-CA-CB	-6.48	100.78	110.50
1	A	226	SER	N-CA-C	5.60	126.11	111.00
4	D	1	DC	O4'-C1'-N1	5.02	111.52	108.00

There are no chirality outliers.

There are no planarity outliers.

5.2 Too-close contacts (i)

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	2225	0	2201	92	0
2	B	2233	0	2230	91	0
3	C	433	0	238	16	0
4	D	422	0	234	12	0
All	All	5313	0	4903	189	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 18.

The worst 5 of 189 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:220:LYS:O	1:A:223:GLU:HG2	1.61	1.00
2:B:419:ILE:HD13	2:B:431:ILE:HG12	1.42	0.99
2:B:342:VAL:HG12	2:B:348:ILE:HA	1.48	0.95
1:A:227:THR:HG22	1:A:228:SER:H	1.35	0.91
1:A:267:MET:O	1:A:303:PHE:HA	1.71	0.90

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	264/384 (69%)	236 (89%)	28 (11%)	0	100	100
2	B	262/410 (64%)	232 (88%)	29 (11%)	1 (0%)	34	72
All	All	526/794 (66%)	468 (89%)	57 (11%)	1 (0%)	47	80

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	B	363	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	254/347 (73%)	251 (99%)	3 (1%)	71	83
2	B	243/361 (67%)	237 (98%)	6 (2%)	47	68
All	All	497/708 (70%)	488 (98%)	9 (2%)	59	76

5 of 9 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	B	204	HIS
2	B	215	LEU
2	B	74	GLU
2	B	79	TYR
2	B	154	HIS

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (3) such sidechains are listed below:

Mol	Chain	Res	Type
2	B	158	HIS
2	B	292	HIS
2	B	387	HIS

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)

There are no ligands in this entry.

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 (i)

6.1 Protein, DNA and RNA chains (i)

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

6.2 Non-standard residues in protein, DNA, RNA chains (i)

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

6.3 Carbohydrates (i)

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

6.4 Ligands (i)

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

6.5 Other polymers (i)

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