WORLDWIDE wwPDB PROTEIN DATA BANK

wwPDB X-ray Structure Validation Summary Report ⓘ

Nov 22, 2023 – 09:59 PM JST

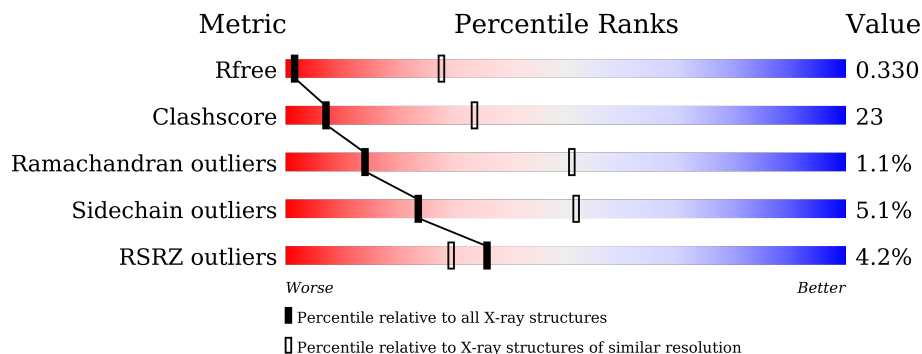|		
PDB ID	:	7XHV
Title	:	Crystal Structure of the NPAS4-ARNT heterodimer in complex with DNA
Authors	:	Sun, X.N.; Jing, L.Q.; Li, F.W.; Wu, D.L.
Deposited on	:	2022-04-10
Resolution	:	4.00 Å(reported)

1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

The reported resolution of this entry is 4.00 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	Percentile Ranks	Value
Rfree		0.330
Clashscore		23
Ramachandran outliers		1.1%
Sidechain outliers		5.1%
RSRZ outliers		4.2%

Worse | | Better

▌Percentile relative to all X-ray structures

▯Percentile relative to X-ray structures of similar resolution

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	1087 (4.30-3.70)
Clashscore	141614	1148 (4.30-3.70)
Ramachandran outliers	138981	1108 (4.30-3.70)
Sidechain outliers	138945	1099 (4.30-3.70)
RSRZ outliers	127900	1028 (4.34-3.66)

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% 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	279	4% 48% 15% 37%
2	B	211	3% 58% 29% • 9%
3	C	16	56% 44%
4	D	16	38% 62%

WORLDWIDE
wwPDB
PROTEIN DATA BANK

2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 3540 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
1	A	177	Total	C	N	O	S	0	0	0
			1400	877	254	258	11			

- Molecule 2 is a protein called Neuronal PAS domain-containing protein 4.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	192	Total	C	N	O	S	0	0	0
			1484	946	265	267	6			

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	207	HIS	-	expression tag	UNP Q8BGD7
B	208	HIS	-	expression tag	UNP Q8BGD7
B	209	HIS	-	expression tag	UNP Q8BGD7
B	210	HIS	-	expression tag	UNP Q8BGD7
B	211	HIS	-	expression tag	UNP Q8BGD7
B	212	HIS	-	expression tag	UNP Q8BGD7

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

| Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	C	16	Total	C	N	O	P	0	0	0
			338	159	66	97	16			

- Molecule 4 is a DNA chain called DNA (5'-D(P*CP*CP*AP*TP*CP*AP*CP*TP*CP*AP *CP*GP*AP*CP*CP*T)-3').

| Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	D	16	Total	C	N	O	P	0	0	0
			318	152	55	95	16			

3   Residue-property plots ⓘ

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 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: Aryl hydrocarbon receptor nuclear translocator
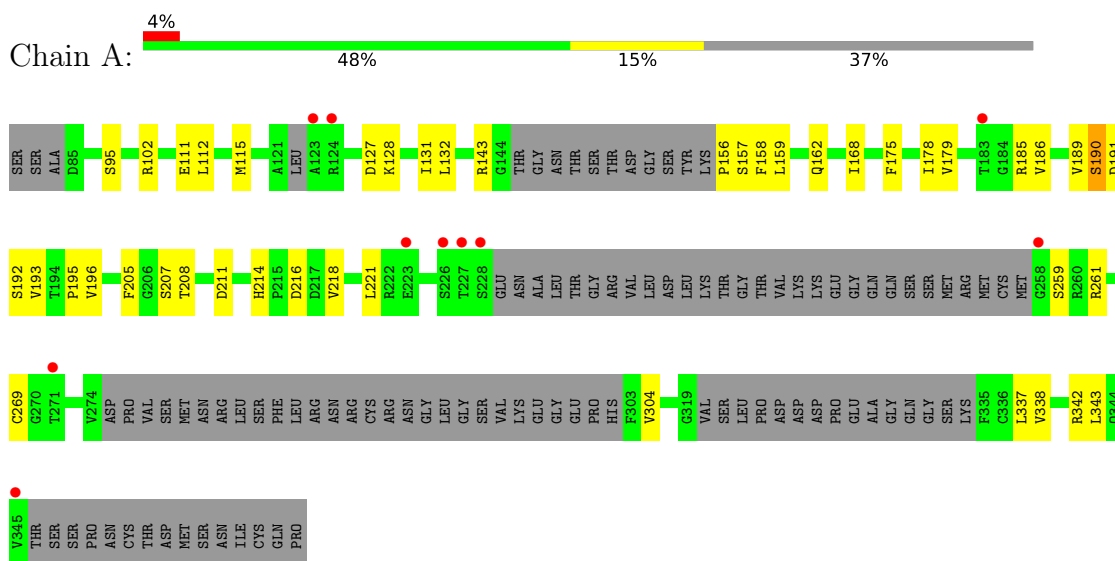
Chain A:

• Molecule 2: Neuronal PAS domain-containing protein 4
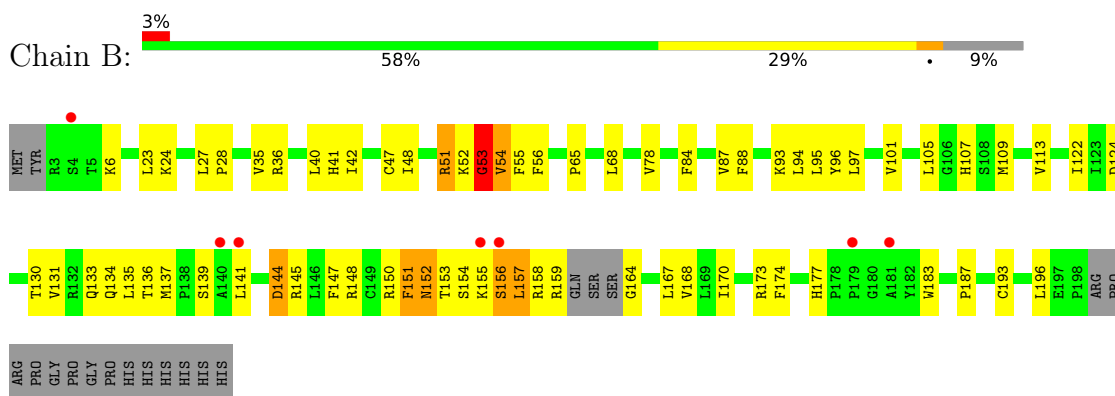
Chain B:

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

Chain C:

• Molecule 4: DNA (5'-D(P*CP*CP*AP*TP*CP*AP*CP*TP*CP*AP*CP*GP*AP*CP*CP*T)-3')

Chain D:

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 43 21 2	Depositor
Cell constants a, b, c, α, β, γ	72.85Å 72.85Å 272.74Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	31.07 – 4.00 32.35 – 4.00	Depositor EDS
% Data completeness (in resolution range)	99.7 (31.07-4.00) 99.7 (32.35-4.00)	Depositor EDS
R_{merge}	0.11	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	3.94 (at 3.99Å)	Xtriage
Refinement program	PHENIX 1.13_2998	Depositor
R, R_{free}	0.273 , 0.331 0.276 , 0.330	Depositor DCC
R_{free} test set	305 reflections (4.50%)	wwPDB-VP
Wilson B-factor (Å^2)	15.0	Xtriage
Anisotropy	0.518	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.19 , 15.2	EDS
L-test for twinning[2]	$< |L| > = 0.48$, $< L^2 > = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.79	EDS
Total number of atoms	3540	wwPDB-VP
Average B, all atoms (Å^2)	57.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.34% 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	0.26	0/1422	0.44	0/1912
2	B	0.43	0/1515	0.57	2/2051 (0.1%)
3	C	0.55	0/380	0.95	0/587
4	D	0.59	0/354	0.92	0/541
All	All	0.41	0/3671	0.63	2/5091 (0.0%)

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	B	53	GLY	N-CA-C	-6.58	96.65	113.10
2	B	23	LEU	CA-CB-CG	5.32	127.54	115.30

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	1400	0	1405	45	0
2	B	1484	0	1506	120	0
3	C	338	0	181	6	0
4	D	318	0	180	11	0
All	All	3540	0	3272	154	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 23.

The worst 5 of 154 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:159:LEU:CD1	2:B:52:LYS:HZ2	1.44	1.30
1:A:158:PHE:CE2	2:B:52:LYS:HE2	1.64	1.29
1:A:159:LEU:CD1	2:B:52:LYS:NZ	1.96	1.27
1:A:159:LEU:HD12	2:B:52:LYS:NZ	1.49	1.27
2:B:54:VAL:HG12	2:B:183:TRP:CH2	1.78	1.19

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	165/279 (59%)	158 (96%)	7 (4%)	0	100	100
2	B	188/211 (89%)	168 (89%)	16 (8%)	4 (2%)	7	39
All	All	353/490 (72%)	326 (92%)	23 (6%)	4 (1%)	14	51

All (4) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	B	151	PHE
2	B	54	VAL
2	B	156	SER
2	B	53	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	158/247 (64%)	150 (95%)	8 (5%)	24	52
2	B	157/174 (90%)	149 (95%)	8 (5%)	24	52
All	All	315/421 (75%)	299 (95%)	16 (5%)	24	52

5 of 16 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	B	152	ASN
2	B	148	ARG
2	B	51	ARG
2	B	144	ASP
1	A	342	ARG

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
2	B	107	HIS

5.3.3   RNA ⓘ

There are no RNA molecules 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 monosaccharides 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, 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	177/279 (63%)	0.38	10 (5%)	24	21	35, 66, 95, 105	0
2	B	192/211 (90%)	0.03	7 (3%)	42	34	30, 43, 64, 78	0
3	C	16/16 (100%)	0.49	0	100	100	62, 70, 81, 83	0
4	D	16/16 (100%)	0.50	0	100	100	54, 73, 88, 91	0
All	All	401/522 (76%)	0.22	17 (4%)	36	29	30, 51, 90, 105	0

The worst 5 of 17 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	271	THR	4.7
1	A	227	THR	3.9
1	A	258	GLY	3.7
1	A	183	THR	3.0
1	A	123	ALA	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 monosaccharides 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.