Full wwPDB X-ray Structure Validation Report ⓘ

Feb 24, 2025 – 03:12 pm GMT

PDB ID	:	8S5A
Title	:	The crystal structure of FAN1 Nuclease bound to 5' phosphorylated p(dG)/3'(dT-dT-dT-dT) double flap DNA
Authors	:	Costanzi, E.; Thomsen, M.
Deposited on	:	2024-02-23
Resolution	:	2.65 Å(reported)

This is a Full wwPDB X-ray Structure Validation 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 ⓘ 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 ⓘ) were used in the production of this report:

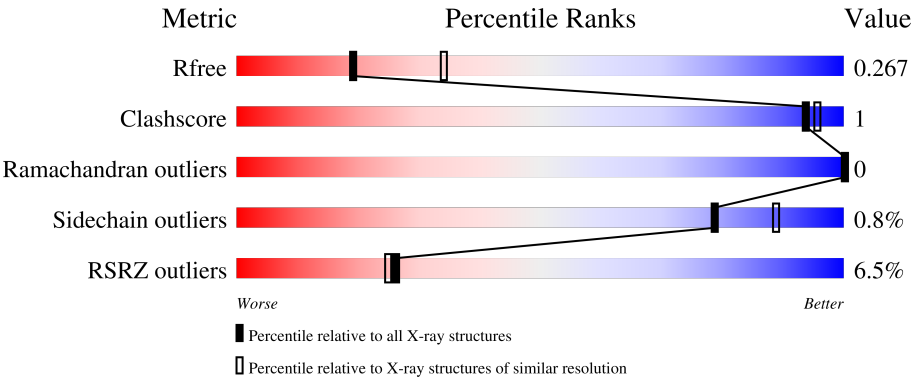
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	3.0
Percentile statistics	:	20231227.v01 (using entries in the PDB archive December 27th 2023)
CCP4	:	9.0.003 (Gargrove)
Density-Fitness	:	1.0.11
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.41

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 2.65 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}	164625	1851 (2.66-2.62)
Clashscore	180529	1953 (2.66-2.62)
Ramachandran outliers	177936	1929 (2.66-2.62)
Sidechain outliers	177891	1929 (2.66-2.62)
RSRZ outliers	164620	1850 (2.66-2.62)

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 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	654	6%, 86%, ., 9%
2	B	19	100%
3	C	8	88%, 12%
4	D	16	62%, 19%, 19%

2 Entry composition (i)

There are 7 unique types of molecules in this entry. The entry contains 5620 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 Fanconi-associated nuclease 1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	592	4785	3053	851	857	24	169	0	0

There are 20 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	355	GLY	-	expression tag	UNP Q9Y2M0
A	356	PRO	-	expression tag	UNP Q9Y2M0
A	357	MET	-	expression tag	UNP Q9Y2M0
A	358	GLY	-	expression tag	UNP Q9Y2M0
A	359	ALA	-	expression tag	UNP Q9Y2M0
A	360	HIS	-	expression tag	UNP Q9Y2M0
A	361	MET	-	expression tag	UNP Q9Y2M0
A	362	THR	-	expression tag	UNP Q9Y2M0
A	363	ARG	-	expression tag	UNP Q9Y2M0
A	507	HIS	ARG	conflict	UNP Q9Y2M0
A	?	-	CYS	deletion	UNP Q9Y2M0
A	?	-	THR	deletion	UNP Q9Y2M0
A	?	-	TRP	deletion	UNP Q9Y2M0
A	?	-	GLY	deletion	UNP Q9Y2M0
A	?	-	LYS	deletion	UNP Q9Y2M0
A	?	-	ASN	deletion	UNP Q9Y2M0
A	?	-	LYS	deletion	UNP Q9Y2M0
A	?	-	PRO	deletion	UNP Q9Y2M0
A	?	-	GLY	deletion	UNP Q9Y2M0
A	794	ALA	LYS	conflict	UNP Q9Y2M0

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	19	Total	C	N	O	P	0	0	0
			379	182	70	109	18			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	C	8	Total	C	N	O	P	0	0	0
			171	79	35	49	8			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	D	13	Total	C	N	O	P	0	0	0
			266	129	45	80	12			

- Molecule 5 is CALCIUM ION (three-letter code: CA) (formula: Ca).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	2	Total	Ca	0	0
			2	2		

- Molecule 6 is CHLORIDE ION (three-letter code: CL) (formula: Cl).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	C	1	Total	Cl	0	0
			1	1		

- Molecule 7 is water.

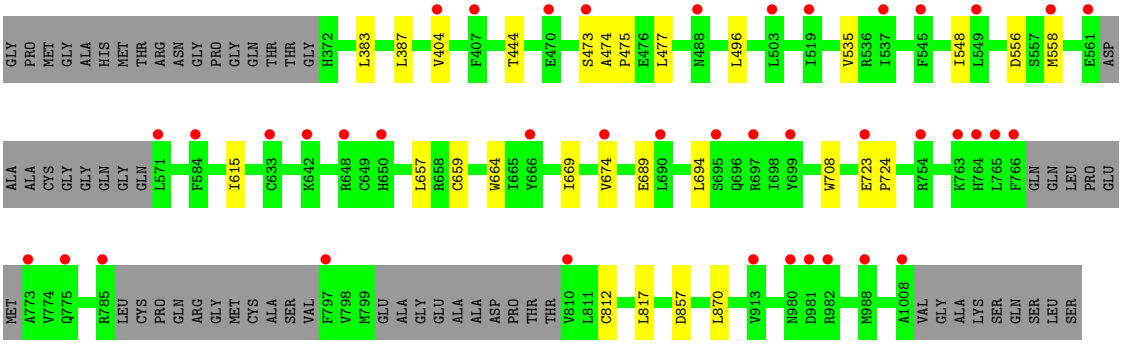
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
7	A	15	Total	O	0	0
			15	15		
7	B	1	Total	O	0	0
			1	1		

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 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.

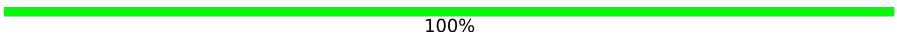
• Molecule 1: Fanconi-associated nuclease 1

Chain A:



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

Chain B:



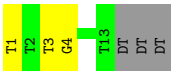
There are no outlier residues recorded for this chain.

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

Chain C:

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

Chain D:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	91.63Å 100.55Å 112.88Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	43.40 – 2.65 43.40 – 2.65	Depositor EDS
% Data completeness (in resolution range)	99.8 (43.40-2.65) 99.8 (43.40-2.65)	Depositor EDS
R_{merge}	0.07	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.96 (at 2.65Å)	Xtriage
Refinement program	REFMAC 5.8.0267	Depositor
R, R_{free}	0.234 , 0.273 0.236 , 0.267	Depositor DCC
R_{free} test set	826 reflections (2.65%)	wwPDB-VP
Wilson B-factor (Å^2)	85.5	Xtriage
Anisotropy	0.519	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.30 , 50.0	EDS
L-test for twinning[2]	$< \|L\| > = 0.50$, $< L^2 > = 0.34$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	5620	wwPDB-VP
Average B, all atoms (Å^2)	95.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.58% 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

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

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	0.68	0/4885	0.77	0/6602
2	B	0.31	0/424	0.85	0/650
3	C	0.77	1/192 (0.5%)	0.70	0/294
4	D	0.36	0/297	0.78	0/458
All	All	0.65	1/5798 (0.0%)	0.77	0/8004

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	C	1	DG	OP3-P	-10.08	1.49	1.61

There are no bond angle outliers.

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	4785	0	4819	12	0
2	B	379	0	214	0	0
3	C	171	0	90	0	0
4	D	266	0	151	2	0
5	A	2	0	0	0	0
6	C	1	0	0	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
7	A	15	0	0	0	0
7	B	1	0	0	0	0
All	All	5620	0	5274	14	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 1.

All (14) 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:387:LEU:HD11	1:A:404:VAL:HG11	1.86	0.57
4:D:1:DT:O2	4:D:1:DT:H2'	2.04	0.56
1:A:659:CYS:HA	1:A:664:TRP:CD2	2.46	0.51
1:A:694:LEU:HD22	1:A:708:TRP:CE2	2.49	0.47
1:A:474:ALA:N	1:A:475:PRO:HD2	2.32	0.45
4:D:3:DT:H2"	4:D:4:DG:C8	2.53	0.44
1:A:383:LEU:HD23	1:A:548:ILE:HG21	1.99	0.43
1:A:657:LEU:HD11	1:A:857:ASP:HB3	2.01	0.43
1:A:477:LEU:HB3	1:A:496:LEU:HD23	2.02	0.42
1:A:615:ILE:HD13	1:A:669:ILE:HG23	2.01	0.41
1:A:723:GLU:HB3	1:A:724:PRO:HD3	2.03	0.41
1:A:812:CYS:SG	1:A:817:LEU:HD13	2.61	0.41
1:A:558:MET:CE	1:A:870:LEU:HD23	2.50	0.41
1:A:674:VAL:HG12	1:A:689:GLU:HB2	2.03	0.40

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	582/654 (89%)	562 (97%)	20 (3%)	0	100 100

There are no Ramachandran outliers to report.

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	522/565 (92%)	518 (99%)	4 (1%)	79 89

All (4) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	444	THR
1	A	473	SER
1	A	535	VAL
1	A	556	ASP

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

Mol	Chain	Res	Type
1	A	492	GLN

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 oligosaccharides in this entry.

5.6 Ligand geometry

Of 3 ligands modelled in this entry, 3 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.

No monomer is involved in short contacts.

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	592/654 (90%)	0.53	41 (6%) 24 23	52, 90, 118, 141	56 (9%)
2	B	19/19 (100%)	-0.36	0 100 100	82, 90, 107, 113	0
3	C	8/8 (100%)	-0.35	0 100 100	98, 102, 107, 109	0
4	D	13/16 (81%)	-0.12	0 100 100	81, 96, 124, 164	0
All	All	632/697 (90%)	0.48	41 (6%) 26 25	52, 90, 118, 164	56 (8%)

All (41) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	561	GLU	8.0
1	A	571	LEU	4.2
1	A	699	TYR	4.0
1	A	775	GLN	3.9
1	A	797	PHE	3.7
1	A	764	HIS	3.7
1	A	674	VAL	3.2
1	A	785	ARG	3.0
1	A	723	GLU	3.0
1	A	765	LEU	3.0
1	A	695	SER	3.0
1	A	763	LYS	2.9
1	A	690	LEU	2.9
1	A	773	ALA	2.9
1	A	537	ILE	2.8
1	A	1008	ALA	2.7
1	A	766	PHE	2.6
1	A	404	VAL	2.5
1	A	981	ASP	2.4
1	A	913	VAL	2.4
1	A	980	ASN	2.4

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	A	754	ARG	2.3
1	A	503	LEU	2.3
1	A	642	LYS	2.3
1	A	982	ARG	2.3
1	A	407	PHE	2.3
1	A	584	PHE	2.3
1	A	558	MET	2.2
1	A	650	HIS	2.2
1	A	648	ARG	2.2
1	A	473	SER	2.2
1	A	633	CYS	2.2
1	A	988	MET	2.2
1	A	666	TYR	2.1
1	A	519	ILE	2.1
1	A	697	ARG	2.1
1	A	470	GLU	2.1
1	A	810	VAL	2.1
1	A	545	PHE	2.0
1	A	549	LEU	2.0
1	A	488	ASN	2.0

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

There are no non-standard protein/DNA/RNA residues in this entry.

6.3 Carbohydrates (i)

There are no monosaccharides in this entry.

6.4 Ligands (i)

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. The B-factors column lists the minimum, median, 95^{th} 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	RSCC	RSR	B-factors(Å^2)	Q<0.9
5	CA	A	1102	1/1	0.84	0.20	109,109,109,109	0
5	CA	A	1101	1/1	0.90	0.11	100,100,100,100	0
6	CL	C	101	1/1	0.91	0.18	116,116,116,116	0

6.5 Other polymers (i)

There are no such residues in this entry.