W O R L D W I D E
wwPDB
P R O T E I N   D A T A   B A N K

wwPDB X-ray Structure Validation Summary Report (i)

Sep 3, 2023 – 11:21 AM EDT

PDB ID	:	3S4W
Title	:	Structure of the FANCI-FANCD2 complex
Authors	:	Pavletich, N.P.
Deposited on	:	2011-05-20
Resolution	:	3.41 Å(reported)

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

MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.35
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 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 3.41 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.
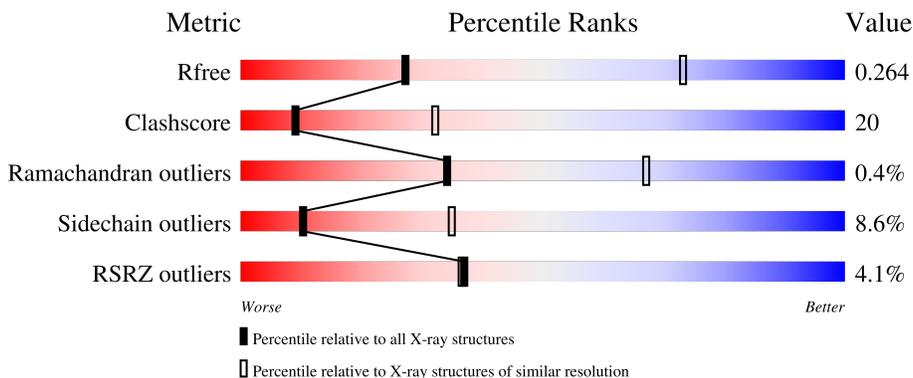
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	1026 (3.48-3.32)
Clashscore	141614	1055 (3.48-3.32)
Ramachandran outliers	138981	1038 (3.48-3.32)
Sidechain outliers	138945	1038 (3.48-3.32)
RSRZ outliers	127900	2173 (3.50-3.30)

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	1308	4% / 54% / 35% / • / 8%
2	B	1323	3% / 50% / 33% / • / 13%

2 Entry composition ⓘ

There are 2 unique types of molecules in this entry. The entry contains 18671 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 anemia group I protein homolog.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	1206	Total 9506	C 6093	N 1579	O 1778	S 56	0	0	0

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1303	HIS	-	expression tag	UNP Q8K368
A	1304	HIS	-	expression tag	UNP Q8K368
A	1305	HIS	-	expression tag	UNP Q8K368
A	1306	HIS	-	expression tag	UNP Q8K368
A	1307	HIS	-	expression tag	UNP Q8K368
A	1308	HIS	-	expression tag	UNP Q8K368

- Molecule 2 is a protein called Fanconi anemia group D2 protein homolog.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	1146	Total 9165	C 5905	N 1543	O 1664	S 53	0	0	0

There are 60 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	?	-	SER	deletion	UNP Q80V62
B	?	-	SER	deletion	UNP Q80V62
B	?	-	SER	deletion	UNP Q80V62
B	?	-	ALA	deletion	UNP Q80V62
B	?	-	VAL	deletion	UNP Q80V62
B	?	-	ALA	deletion	UNP Q80V62
B	?	-	ALA	deletion	UNP Q80V62
B	?	-	LYS	deletion	UNP Q80V62
B	?	-	ASN	deletion	UNP Q80V62
B	?	-	ARG	deletion	UNP Q80V62

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	?	-	ASN	deletion	UNP Q80V62
B	?	-	LYS	deletion	UNP Q80V62
B	?	-	GLY	deletion	UNP Q80V62
B	?	-	LYS	deletion	UNP Q80V62
B	?	-	THR	deletion	UNP Q80V62
B	?	-	GLY	deletion	UNP Q80V62
B	?	-	GLY	deletion	UNP Q80V62
B	?	-	LYS	deletion	UNP Q80V62
B	?	-	LYS	deletion	UNP Q80V62
B	?	-	GLN	deletion	UNP Q80V62
B	?	-	LYS	deletion	UNP Q80V62
B	?	-	ALA	deletion	UNP Q80V62
B	?	-	ASP	deletion	UNP Q80V62
B	?	-	SER	deletion	UNP Q80V62
B	?	-	ASN	deletion	UNP Q80V62
B	?	-	LYS	deletion	UNP Q80V62
B	?	-	ALA	deletion	UNP Q80V62
B	?	-	SER	deletion	UNP Q80V62
B	?	-	CYS	deletion	UNP Q80V62
B	?	-	SER	deletion	UNP Q80V62
B	?	-	ASP	deletion	UNP Q80V62
B	?	-	THR	deletion	UNP Q80V62
B	?	-	LEU	deletion	UNP Q80V62
B	?	-	LEU	deletion	UNP Q80V62
B	?	-	THR	deletion	UNP Q80V62
B	?	-	GLU	deletion	UNP Q80V62
B	?	-	ASP	deletion	UNP Q80V62
B	?	-	THR	deletion	UNP Q80V62
B	?	-	SER	deletion	UNP Q80V62
B	?	-	GLU	deletion	UNP Q80V62
B	?	-	CYS	deletion	UNP Q80V62
B	?	-	ASP	deletion	UNP Q80V62
B	?	-	MET	deletion	UNP Q80V62
B	?	-	ALA	deletion	UNP Q80V62
B	?	-	PRO	deletion	UNP Q80V62
B	?	-	SER	deletion	UNP Q80V62
B	?	-	GLY	deletion	UNP Q80V62
B	?	-	ARG	deletion	UNP Q80V62
B	?	-	SER	deletion	UNP Q80V62
B	?	-	HIS	deletion	UNP Q80V62
B	?	-	VAL	deletion	UNP Q80V62
B	?	-	ASP	deletion	UNP Q80V62

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	?	-	LYS	deletion	UNP Q80V62
B	?	-	GLU	deletion	UNP Q80V62
B	?	-	SER	deletion	UNP Q80V62
B	?	-	THR	deletion	UNP Q80V62
B	?	-	GLY	deletion	UNP Q80V62
B	?	-	LYS	deletion	UNP Q80V62
B	?	-	GLU	deletion	UNP Q80V62
B	?	-	GLY	deletion	UNP Q80V62

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: Fanconi anemia group I protein homolog

• Molecule 2: Fanconi anemia group D2 protein homolog
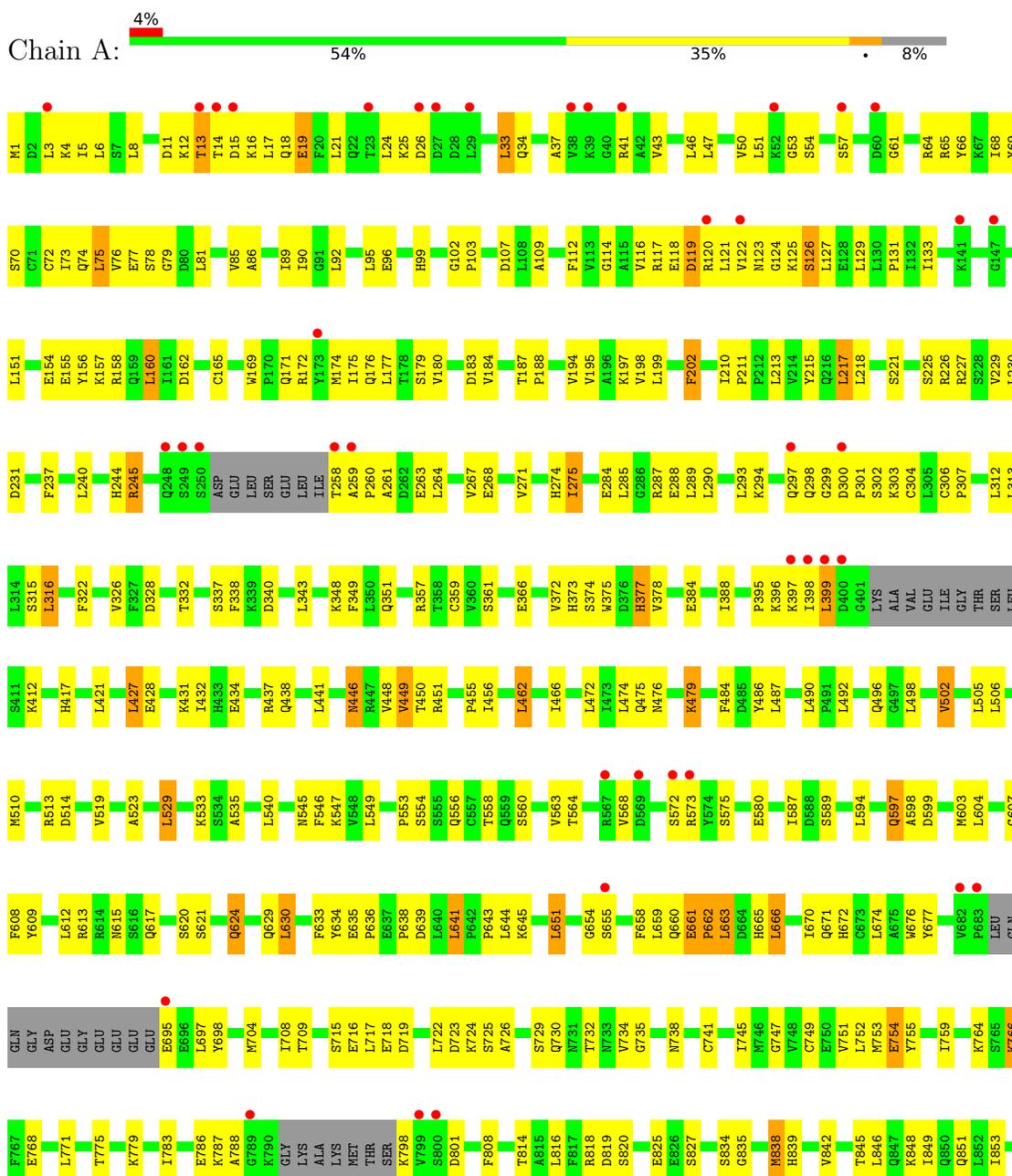
4 Data and refinement statistics ⓘ

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	83.80Å 110.40Å 350.50Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	19.86 – 3.41 20.02 – 3.41	Depositor EDS
% Data completeness (in resolution range)	90.0 (19.86-3.41) 88.9 (20.02-3.41)	Depositor EDS
R_{merge}	0.09	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.72 (at 3.44Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.7_650)	Depositor
R, R_{free}	0.240 , 0.272 0.232 , 0.264	Depositor DCC
R_{free} test set	1738 reflections (4.05%)	wwPDB-VP
Wilson B-factor (Å^2)	90.3	Xtriage
Anisotropy	0.261	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.26 , 53.5	EDS
L-test for twinning[2]	$< |L| > = 0.47, < L^2 > = 0.30$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.93	EDS
Total number of atoms	18671	wwPDB-VP
Average B, all atoms (Å^2)	122.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.68% 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	0.28	0/9655	0.47	0/13038
2	B	0.26	0/9336	0.46	0/12621
All	All	0.27	0/18991	0.46	0/25659

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
1	A	0	1
2	B	0	3
All	All	0	4

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (4) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	1062	ILE	Peptide
2	B	1094	ASN	Peptide
2	B	1284	ASP	Peptide
2	B	290	THR	Peptide

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.

WORLDWIDE
wwPDB
PROTEIN DATA BANK

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	9506	0	9792	373	0
2	B	9165	0	9297	394	0
All	All	18671	0	19089	759	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 20.

The worst 5 of 759 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)
2:B:698:LEU:HD22	2:B:760:PRO:HD3	1.38	1.04
1:A:764:LYS:HG2	1:A:827:SER:HB3	1.40	1.02
2:B:1123:VAL:HG13	2:B:1128:CYS:HB2	1.48	0.95
1:A:1071:VAL:HG13	1:A:1076:ALA:HB2	1.49	0.95
2:B:426:LEU:HD23	2:B:429:ILE:HD12	1.51	0.93

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	1190/1308 (91%)	1108 (93%)	75 (6%)	7 (1%)	25	57
2	B	1120/1323 (85%)	1043 (93%)	75 (7%)	2 (0%)	47	78
All	All	2310/2631 (88%)	2151 (93%)	150 (6%)	9 (0%)	34	67

5 of 9 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	122	VAL
1	A	568	VAL
2	B	1285	SER

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	A	25	LYS
1	A	717	LEU

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	1093/1188 (92%)	992 (91%)	101 (9%)	9	31
2	B	1030/1193 (86%)	949 (92%)	81 (8%)	12	39
All	All	2123/2381 (89%)	1941 (91%)	182 (9%)	10	35

5 of 182 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	B	441	PHE
2	B	800	GLN
2	B	463	PHE
2	B	650	VAL
2	B	928	ILE

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 15 such sidechains are listed below:

Mol	Chain	Res	Type
2	B	800	GLN
2	B	1342	HIS
2	B	1003	HIS
2	B	1354	HIS
2	B	1086	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	1206/1308 (92%)	-0.13	55 (4%)	32	32	48, 100, 204, 311	0
2	B	1146/1323 (86%)	-0.01	41 (3%)	42	42	74, 127, 192, 236	0
All	All	2352/2631 (89%)	-0.07	96 (4%)	37	36	48, 116, 198, 311	0

The worst 5 of 96 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	399	LEU	7.1
2	B	543	PRO	6.9
1	A	14	THR	6.8
1	A	397	LYS	6.0
1	A	258	THR	5.6

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.