wwPDB X-ray Structure Validation Summary Report (i)

Aug 19, 2023 – 05:19 PM EDT

PDB ID	:	2FO1
Title	:	Crystal Structure of the CSL-Notch-Mastermind ternary complex bound to DNA
Authors	:	Wilson, J.J.; Kovall, R.A.
Deposited on	:	2006-01-12
Resolution	:	3.12 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 (i)) were used in the production of this report:

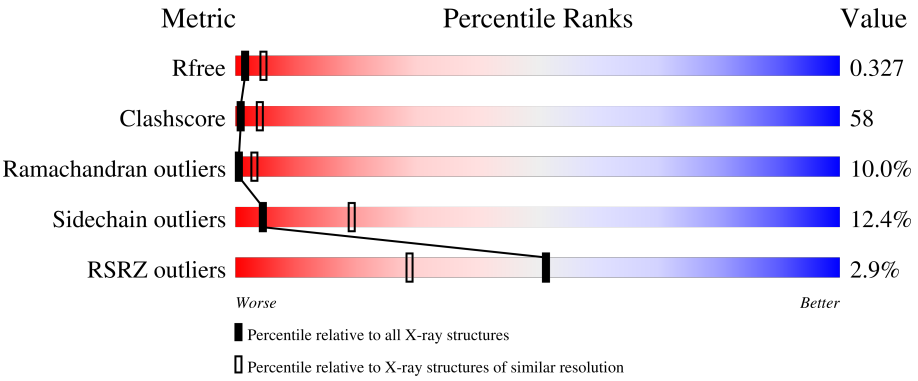
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.35
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.35

1 Overall quality at a glance (i)

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

The reported resolution of this entry is 3.12 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	1292 (3.14-3.10)
Clashscore	141614	1389 (3.14-3.10)
Ramachandran outliers	138981	1337 (3.14-3.10)
Sidechain outliers	138945	1337 (3.14-3.10)
RSRZ outliers	127900	1260 (3.14-3.10)

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	B	15	33% 60% 7%
2	C	15	33% 60% 7%
3	A	477	4% 34% 48% 10% 8%
4	D	85	21% 44% 8% 26%
5	E	373	2% 16% 46% 16% 20%

2 Entry composition (i)

There are 5 unique types of molecules in this entry. The entry contains 6939 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 DNA chain called 5'-D(*TP*TP*AP*CP*TP*GP*TP*GP*GP*GP*AP*AP*AP*GP*A)-3'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
1	B	15	311	149	61	87	14	0	0	0

- Molecule 2 is a DNA chain called 5'-D(*AP*AP*TP*CP*TP*TP*TP*CP*CP*CP*AP*CP*AP*GP*T)-3'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
2	C	15	298	145	50	89	14	0	0	0

- Molecule 3 is a protein called Lin-12 and glp-1 phenotype protein 1, isoform b.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
3	A	439	3497	2218	608	654	17	0	0	0

There are 5 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	187	GLY	-	cloning artifact	GB 22532887
A	188	PRO	-	cloning artifact	GB 22532887
A	189	LEU	-	cloning artifact	GB 22532887
A	190	GLY	-	cloning artifact	GB 22532887
A	191	SER	-	cloning artifact	GB 22532887

- Molecule 4 is a protein called Protein lag-3.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
4	D	63	524	316	102	105	1	0	0	0

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
D	48	SER	-	cloning artifact	UNP Q09260

- Molecule 5 is a protein called Lin-12 protein.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	N	O	S	Se			
5	E	297	2309	1420	427	445	3	14	0	0	0

There are 16 discrepancies between the modelled and reference sequences:

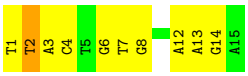
Chain	Residue	Modelled	Actual	Comment	Reference
E	929	GLY	-	cloning artifact	UNP P14585
E	930	SER	-	cloning artifact	UNP P14585
E	939	MSE	MET	modified residue	UNP P14585
E	946	MSE	MET	modified residue	UNP P14585
E	949	MSE	MET	modified residue	UNP P14585
E	1090	MSE	MET	modified residue	UNP P14585
E	1099	MSE	MET	modified residue	UNP P14585
E	1114	MSE	MET	modified residue	UNP P14585
E	1142	MSE	MET	modified residue	UNP P14585
E	1143	MSE	MET	modified residue	UNP P14585
E	1146	MSE	MET	modified residue	UNP P14585
E	1164	MSE	MET	modified residue	UNP P14585
E	1168	MSE	MET	modified residue	UNP P14585
E	1220	MSE	MET	modified residue	UNP P14585
E	1246	MSE	MET	modified residue	UNP P14585
E	1258	MSE	MET	modified residue	UNP P14585

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.

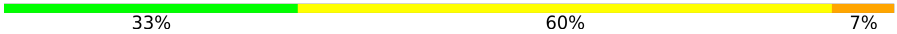
• Molecule 1: 5'-D(*TP*TP*AP*CP*TP*GP*TP*GP*GP*GP*AP*AP*AP*GP*A)-3'

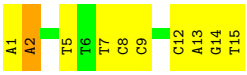
Chain B:



• Molecule 2: 5'-D(*AP*AP*TP*CP*TP*TP*TP*TP*CP*CP*CP*AP*CP*AP*GP*T)-3'

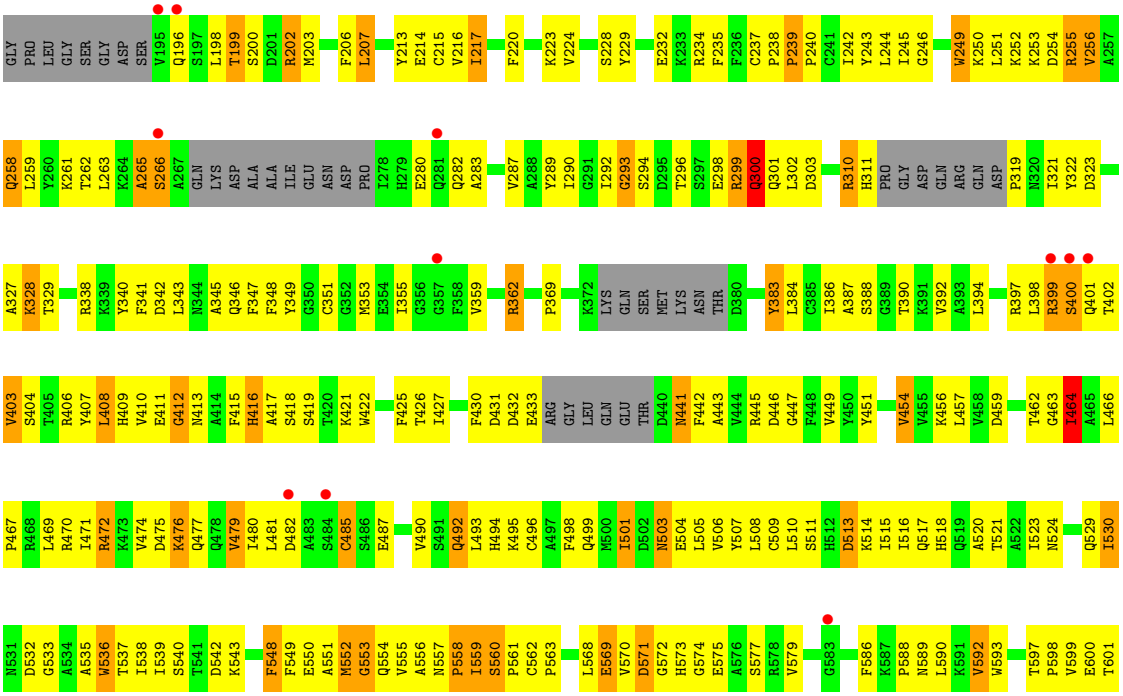
Chain C:

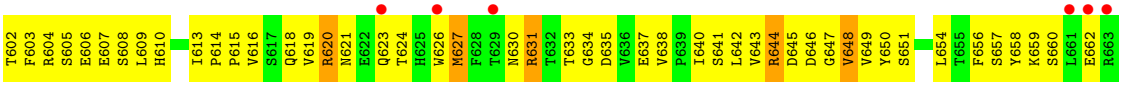




• Molecule 3: Lin-12 and glp-1 phenotype protein 1, isoform b

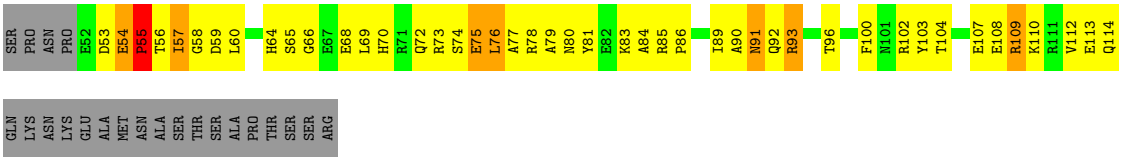
Chain A:





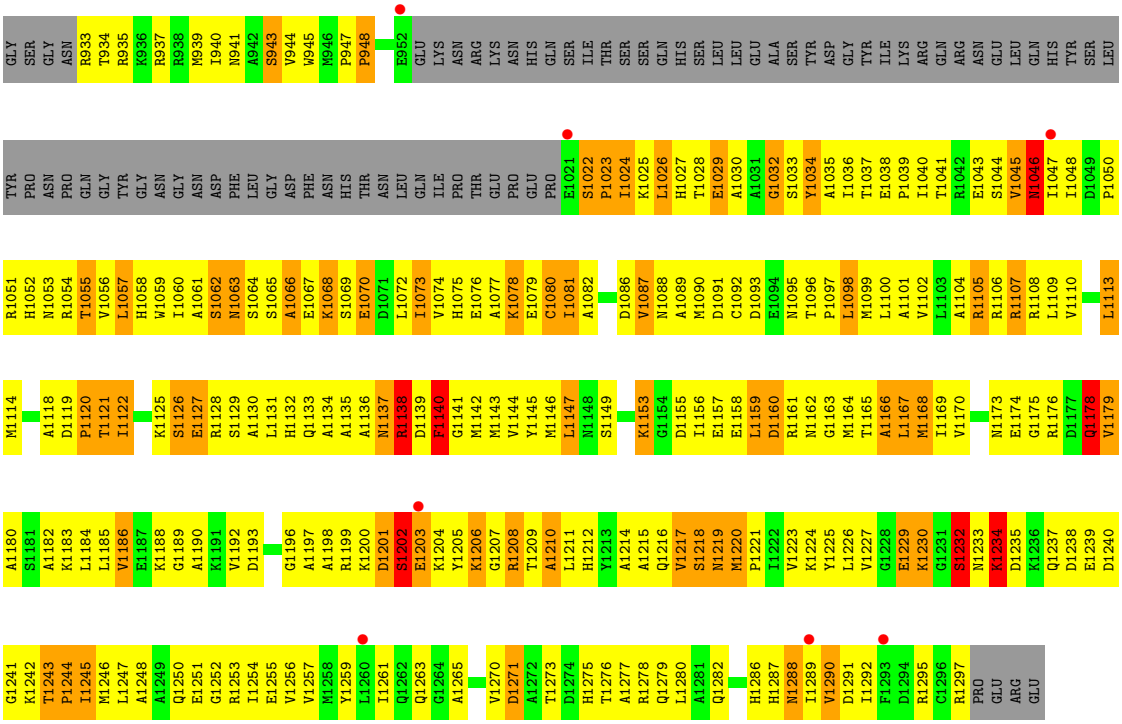
• Molecule 4: Protein lag-3

Chain D:



• Molecule 5: Lin-12 protein

Chain E:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	66.09Å 96.78Å 243.54Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	43.51 – 3.12 43.51 – 3.12	Depositor EDS
% Data completeness (in resolution range)	88.6 (43.51-3.12) 93.0 (43.51-3.12)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.09	Depositor
$< I/\sigma(I) >$ [1]	2.67 (at 3.12Å)	Xtriage
Refinement program	CNS 1.1	Depositor
R, R_{free}	0.273 , 0.340 0.276 , 0.327	Depositor DCC
R_{free} test set	4706 reflections (9.04%)	wwPDB-VP
Wilson B-factor (Å^2)	94.2	Xtriage
Anisotropy	0.612	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.23 , 72.0	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	0.92	EDS
Total number of atoms	6939	wwPDB-VP
Average B, all atoms (Å^2)	103.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.03% 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	B	0.66	0/350	0.80	0/540
2	C	0.64	0/332	0.81	0/509
3	A	0.45	0/3573	0.70	0/4822
4	D	0.37	0/532	0.69	1/714 (0.1%)
5	E	0.42	0/2327	0.76	0/3113
All	All	0.46	0/7114	0.73	1/9698 (0.0%)

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	B	0	2
2	C	0	2
All	All	0	4

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
4	D	54	GLU	C-N-CD	-5.81	107.83	120.60

There are no chirality outliers.

All (4) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	B	12	DA	Sidechain
1	B	2	DT	Sidechain
2	C	2	DA	Sidechain
2	C	5	DT	Sidechain

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	B	311	0	171	20	0
2	C	298	0	172	17	0
3	A	3497	0	3441	343	0
4	D	524	0	491	51	0
5	E	2309	0	2301	392	0
All	All	6939	0	6576	785	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 58.

The worst 5 of 785 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)
5:E:1022:SER:H	5:E:1023:PRO:HD2	1.17	1.08
5:E:1167:LEU:H	5:E:1167:LEU:HD12	1.14	1.07
3:A:283:ALA:HB1	3:A:351:CYS:HB3	1.33	1.07
5:E:935:ARG:NH1	5:E:1158:GLU:HG2	1.69	1.06
2:C:8:DC:H2”	2:C:9:DC:H5’	1.33	1.04

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
3	A	429/477 (90%)	324 (76%)	77 (18%)	28 (6%)	1 7

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles
4	D	61/85 (72%)	44 (72%)	14 (23%)	3 (5%)	2 13
5	E	293/373 (79%)	170 (58%)	76 (26%)	47 (16%)	0 0
All	All	783/935 (84%)	538 (69%)	167 (21%)	78 (10%)	0 3

5 of 78 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
3	A	299	ARG
3	A	399	ARG
3	A	400	SER
3	A	403	VAL
3	A	504	GLU

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
3	A	385/416 (92%)	341 (89%)	44 (11%)	5 23
4	D	54/73 (74%)	47 (87%)	7 (13%)	4 18
5	E	244/298 (82%)	210 (86%)	34 (14%)	3 15
All	All	683/787 (87%)	598 (88%)	85 (12%)	4 19

5 of 85 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
5	E	1073	ILE
5	E	1159	LEU
5	E	1080	CYS
5	E	1137	ASN
5	E	1178	GLN

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 35 such sidechains are listed below:

Mol	Chain	Res	Type
5	E	1162	ASN
5	E	1172	HIS
5	E	1275	HIS
3	A	557	ASN
3	A	528	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, 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	B	15/15 (100%)	-0.34	0 100 100	72, 88, 113, 115	0
2	C	15/15 (100%)	-0.38	0 100 100	66, 85, 113, 113	0
3	A	439/477 (92%)	0.03	17 (3%) 39 20	37, 97, 165, 197	0
4	D	63/85 (74%)	-0.50	0 100 100	68, 123, 174, 200	0
5	E	283/373 (75%)	-0.22	7 (2%) 57 35	52, 94, 165, 202	0
All	All	815/965 (84%)	-0.11	24 (2%) 51 29	37, 98, 166, 202	0

The worst 5 of 24 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
3	A	400	SER	8.0
3	A	484	SER	5.3
5	E	1203	GLU	4.2
3	A	626	TRP	3.6
3	A	661	LEU	3.6

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 (i)

There are no such residues in this entry.