wwPDB X-ray Structure Validation Summary Report (i)

Jun 12, 2024 – 06:18 PM EDT

PDB ID	:	4BXR
Title	:	Structure of the wild-type TCP10 domain of Danio rerio CPAP in complex with a peptide of Danio rerio STIL
Authors	:	van Breugel, M.
Deposited on	:	2013-07-15
Resolution	:	2.20 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:

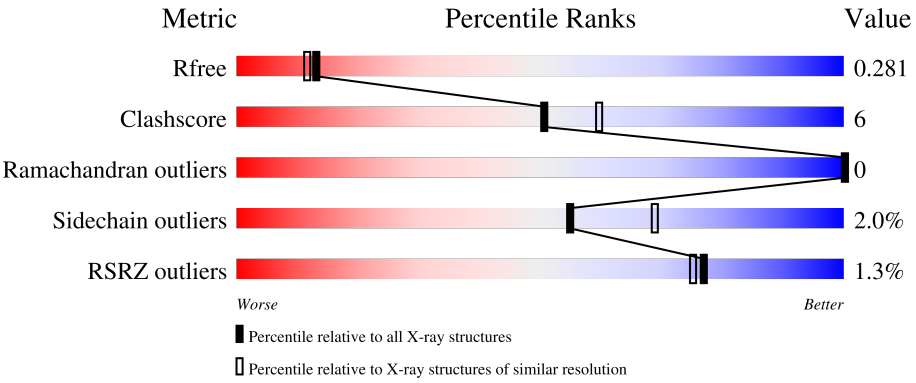
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.20.1
EDS	:	2.36.2
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36.2

1 Overall quality at a glance (i)

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

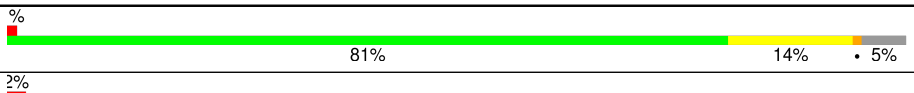
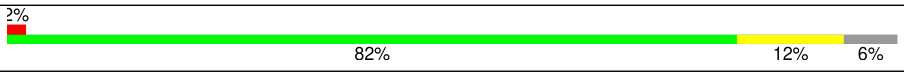
The reported resolution of this entry is 2.20 Å.

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	4898 (2.20-2.20)
Clashscore	141614	5594 (2.20-2.20)
Ramachandran outliers	138981	5503 (2.20-2.20)
Sidechain outliers	138945	5504 (2.20-2.20)
RSRZ outliers	127900	4800 (2.20-2.20)

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	192	
1	B	192	
2	C	28	
2	D	28	

2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 3176 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 CPAP.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	183	Total	C	N	O	S	0	1	0
			1475	920	260	290	5			
1	B	181	Total	C	N	O	S	0	0	0
			1450	907	256	282	5			

There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	933	GLY	-	expression tag	UNP E7FCY1
A	934	PRO	-	expression tag	UNP E7FCY1
A	935	HIS	-	expression tag	UNP E7FCY1
A	936	MET	-	expression tag	UNP E7FCY1
B	933	GLY	-	expression tag	UNP E7FCY1
B	934	PRO	-	expression tag	UNP E7FCY1
B	935	HIS	-	expression tag	UNP E7FCY1
B	936	MET	-	expression tag	UNP E7FCY1

- Molecule 2 is a protein called SCL-INTERRUPTING LOCUS PROTEIN HOMOLOG.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
2	C	12	Total	C	N	O	0	0	0
			92	57	18	17			
2	D	14	Total	C	N	O	0	0	0
			105	65	20	20			

There are 14 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	407	GLY	-	expression tag	UNP Q8JGS1
C	429	GLU	-	expression tag	UNP Q8JGS1
C	430	ASN	-	expression tag	UNP Q8JGS1
C	431	LEU	-	expression tag	UNP Q8JGS1

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
C	432	TYR	-	expression tag	UNP Q8JGS1
C	433	PHE	-	expression tag	UNP Q8JGS1
C	434	GLN	-	expression tag	UNP Q8JGS1
D	407	GLY	-	expression tag	UNP Q8JGS1
D	429	GLU	-	expression tag	UNP Q8JGS1
D	430	ASN	-	expression tag	UNP Q8JGS1
D	431	LEU	-	expression tag	UNP Q8JGS1
D	432	TYR	-	expression tag	UNP Q8JGS1
D	433	PHE	-	expression tag	UNP Q8JGS1
D	434	GLN	-	expression tag	UNP Q8JGS1

- Molecule 3 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	30	Total O 30 30	0	0
3	B	20	Total O 20 20	0	0
3	D	4	Total O 4 4	0	0

3 Residue-property plots (i)

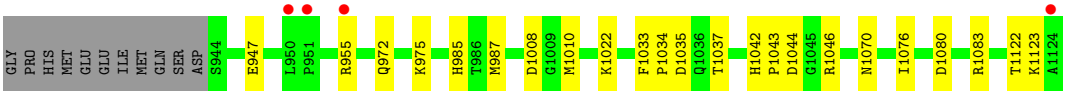
These plots are drawn for all protein, RNA, DNA and oligosaccharide chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. 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: CPAP

Chain A:

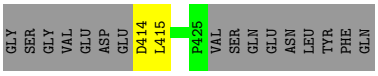
• Molecule 1: CPAP

Chain B:



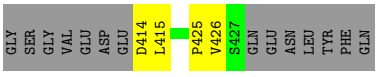
• Molecule 2: SCL-INTERRUPTING LOCUS PROTEIN HOMOLOG

Chain C:



• Molecule 2: SCL-INTERRUPTING LOCUS PROTEIN HOMOLOG

Chain D:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	60.25Å 67.47Å 61.65Å 90.00° 113.92° 90.00°	Depositor
Resolution (Å)	51.13 – 2.20 51.08 – 2.20	Depositor EDS
% Data completeness (in resolution range)	99.8 (51.13-2.20) 99.8 (51.08-2.20)	Depositor EDS
R_{merge}	0.10	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.75 (at 2.20Å)	Xtriage
Refinement program	REFMAC 5.7.0029	Depositor
R, R_{free}	0.234 , 0.277 0.237 , 0.281	Depositor DCC
R_{free} test set	1145 reflections (4.97%)	wwPDB-VP
Wilson B-factor (Å^2)	50.2	Xtriage
Anisotropy	0.307	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.33 , 44.6	EDS
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.32$	Xtriage
Estimated twinning fraction	0.036 for l,-k,h	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	3176	wwPDB-VP
Average B, all atoms (Å^2)	59.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.31% 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	0.79	0/1506	0.84	0/2036
1	B	0.73	0/1481	0.83	1/2002 (0.0%)
2	C	0.64	0/97	0.74	0/135
2	D	0.75	0/110	0.80	0/153
All	All	0.76	0/3194	0.83	1/4326 (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
2	D	0	1

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	B	1008	ASP	CB-CG-OD1	5.66	123.39	118.30

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
2	D	425	PRO	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.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	1475	0	1440	18	0
1	B	1450	0	1422	14	0
2	C	92	0	85	1	0
2	D	105	0	99	2	0
3	A	30	0	0	4	0
3	B	20	0	0	1	0
3	D	4	0	0	0	0
All	All	3176	0	3046	35	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 6.

The worst 5 of 35 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:1122:THR:HG22	1:A:1122:THR:O	1.63	0.95
2:D:414:ASP:OD1	2:D:415:LEU:N	2.18	0.73
1:B:947:GLU:OE2	1:B:955:ARG:HD3	1.87	0.73
1:A:990:GLN:O	1:A:990:GLN:NE2	2.25	0.70
1:A:1122:THR:O	1:A:1122:THR:CG2	2.38	0.70

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	182/192 (95%)	176 (97%)	6 (3%)	0	100	100
1	B	179/192 (93%)	174 (97%)	5 (3%)	0	100	100
2	C	10/28 (36%)	8 (80%)	2 (20%)	0	100	100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
2	D	12/28 (43%)	12 (100%)	0	0	100	100
All	All	383/440 (87%)	370 (97%)	13 (3%)	0	100	100

There are no Ramachandran outliers to report.

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	165/171 (96%)	160 (97%)	5 (3%)	41	53
1	B	161/171 (94%)	159 (99%)	2 (1%)	71	83
2	C	12/26 (46%)	12 (100%)	0	100	100
2	D	14/26 (54%)	14 (100%)	0	100	100
All	All	352/394 (89%)	345 (98%)	7 (2%)	55	69

5 of 7 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	1059	LEU
1	A	1080	ASP
1	B	1080	ASP
1	B	1070	ASN
1	A	990	GLN

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

Mol	Chain	Res	Type
1	A	941	GLN
1	A	1014	GLN
1	B	1000	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 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	A	183/192 (95%)	0.02	1 (0%) 91 90	35, 55, 85, 98	0
1	B	181/192 (94%)	0.17	4 (2%) 62 59	30, 61, 93, 112	0
2	C	12/28 (42%)	0.16	0 100 100	58, 71, 93, 96	0
2	D	14/28 (50%)	-0.04	0 100 100	41, 48, 80, 90	0
All	All	390/440 (88%)	0.09	5 (1%) 77 75	30, 58, 92, 112	0

All (5) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	1124	ALA	2.8
1	A	1043	PRO	2.3
1	B	950	LEU	2.2
1	B	951	PRO	2.1
1	B	955	ARG	2.0

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 (i)

There are no such residues in this entry.