wwPDB X-ray Structure Validation Summary Report (i)

May 18, 2020 – 07:07 am BST

PDB ID	:	4JPB
Title	:	The structure of a ternary complex between CheA domains P4 and P5 with CheW and with an unzipped fragment of TM14, a chemoreceptor analog from Thermotoga maritima.
Authors	:	Li, X.; Bayas, C.; Bilwes, A.M.; Crane, B.R.
Deposited on	:	2013-03-19
Resolution	:	3.19 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 following versions of software and data (see references (1)) were used in the production of this report:

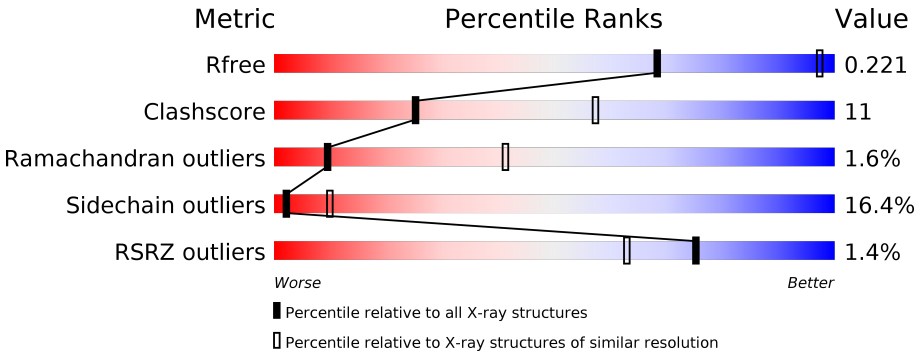
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.11

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 3.19 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	1467 (3.20-3.16)
Clashscore	141614	1599 (3.20-3.16)
Ramachandran outliers	138981	1574 (3.20-3.16)
Sidechain outliers	138945	1573 (3.20-3.16)
RSRZ outliers	127900	1423 (3.20-3.16)

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 on 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	319	24% 13% • 59%
2	W	151	% 60% 28% 5% 7%
3	B	90	68% 22% • 7%
3	C	90	4% 64% 24% 6% 6%

2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 3468 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 Chemotaxis protein CheA.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	131	Total	C	N	O	S	0	1	0
			1027	661	172	193	1			

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	353	GLY	-	EXPRESSION TAG	UNP Q56310
A	354	SER	-	EXPRESSION TAG	UNP Q56310
A	355	HIS	-	EXPRESSION TAG	UNP Q56310
A	?	-	PHE	DELETION	UNP Q56310

- Molecule 2 is a protein called Chemotaxis protein CheW.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	W	140	Total	C	N	O	S	0	0	0
			1113	716	184	211	2			

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
W	-3	GLY	-	EXPRESSION TAG	UNP Q56311
W	-2	SER	-	EXPRESSION TAG	UNP Q56311
W	-1	HIS	-	EXPRESSION TAG	UNP Q56311
W	0	MET	-	EXPRESSION TAG	UNP Q56311

- Molecule 3 is a protein called Methyl-accepting chemotaxis protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	B	84	Total	C	N	O	S	0	2	0
			667	407	125	133	2			
3	C	85	Total	C	N	O	S	0	0	0
			661	403	121	135	2			

There are 8 discrepancies between the modelled and reference sequences:

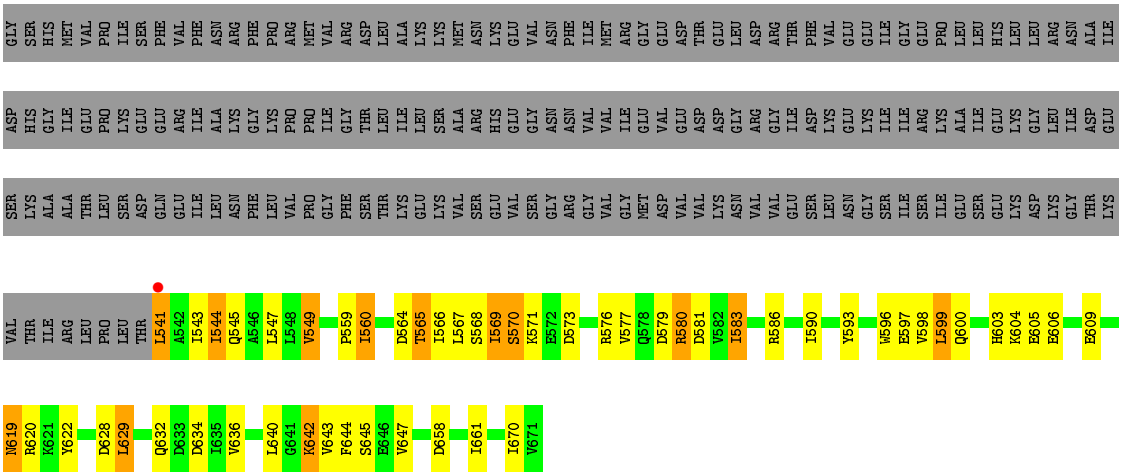
Chain	Residue	Modelled	Actual	Comment	Reference
B	103	GLY	-	EXPRESSION TAG	UNP Q7DFA3
B	104	SER	-	EXPRESSION TAG	UNP Q7DFA3
B	105	HIS	-	EXPRESSION TAG	UNP Q7DFA3
B	106	MET	-	EXPRESSION TAG	UNP Q7DFA3
C	103	GLY	-	EXPRESSION TAG	UNP Q7DFA3
C	104	SER	-	EXPRESSION TAG	UNP Q7DFA3
C	105	HIS	-	EXPRESSION TAG	UNP Q7DFA3
C	106	MET	-	EXPRESSION TAG	UNP Q7DFA3

3 Residue-property plots (i)

These plots are drawn for all protein, RNA and DNA 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.

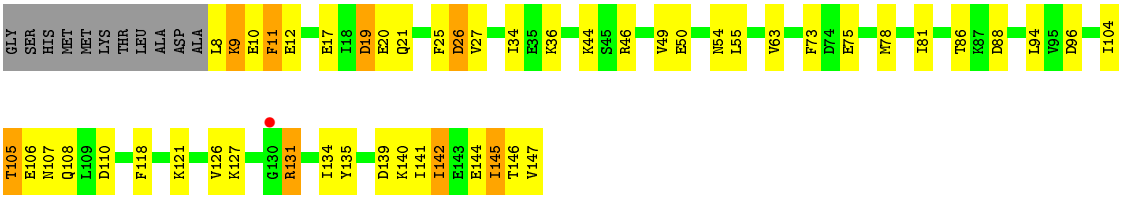
- Molecule 1: Chemotaxis protein CheA

Chain A:



- Molecule 2: Chemotaxis protein CheW

Chain W:



- Molecule 3: Methyl-accepting chemotaxis protein

Chain B:

- Molecule 3: Methyl-accepting chemotaxis protein

Chain C:

GLY
SER
HIS
MET
SER
Q108
T112
L113
E114
R117
S118
I119
I123
Q124
N125
I126
M127
R128
I129
L136
A137
L138
T141
V157
E160
V161
L164
E167
T172
V176
E177
L183
E184
S185
S186
Q187
R188
S189
L190
E191
N192

4 Data and refinement statistics (i)

Property	Value	Source
Space group	H 3 2	Depositor
Cell constants a, b, c, α, β, γ	213.59Å 213.59Å 208.84Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	46.16 – 3.19 46.16 – 3.19	Depositor EDS
% Data completeness (in resolution range)	99.4 (46.16-3.19) 93.3 (46.16-3.19)	Depositor EDS
R_{merge}	0.10	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.49 (at 3.19Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.8.1_1168)	Depositor
R, R_{free}	0.196 , 0.220 0.200 , 0.221	Depositor DCC
R_{free} test set	2000 reflections (6.55%)	wwPDB-VP
Wilson B-factor (Å^2)	91.6	Xtriage
Anisotropy	0.367	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.30 , 23.7	EDS
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	3468	wwPDB-VP
Average B, all atoms (Å^2)	35.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.96% 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	0.60	0/1040	0.76	0/1407
2	W	0.51	0/1124	0.66	0/1512
3	B	0.47	0/675	0.60	0/904
3	C	0.51	0/661	0.63	0/887
All	All	0.53	0/3500	0.67	0/4710

There are no bond length outliers.

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	1027	0	1088	29	1
2	W	1113	0	1177	26	1
3	B	667	0	691	19	0
3	C	661	0	679	15	0
All	All	3468	0	3635	76	1

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 11.

The worst 5 of 76 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)
3:B:157:VAL:HG23	3:C:136:LEU:HD12	1.65	0.78
2:W:110:ASP:HB2	2:W:127:LYS:HE2	1.67	0.77
1:A:560:ILE:HD12	1:A:632:GLN:HE21	1.57	0.69
3:C:114:GLU:OE1	3:C:117:ARG:NH2	2.26	0.68
1:A:569:ILE:HD13	1:A:583:ILE:HD13	1.77	0.67

All (1) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:577:VAL:O	2:W:21:GLN:NE2[3_555]	2.11	0.09

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	130/319 (41%)	109 (84%)	18 (14%)	3 (2%)	6 32
2	W	138/151 (91%)	125 (91%)	9 (6%)	4 (3%)	4 26
3	B	84/90 (93%)	82 (98%)	2 (2%)	0	100 100
3	C	83/90 (92%)	76 (92%)	7 (8%)	0	100 100
All	All	435/650 (67%)	392 (90%)	36 (8%)	7 (2%)	9 41

5 of 7 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	605	GLU
2	W	106	GLU
1	A	645	SER
2	W	19	ASP
2	W	78	MET

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	115/279 (41%)	88 (76%)	27 (24%)	1 3
2	W	128/136 (94%)	113 (88%)	15 (12%)	5 22
3	B	73/76 (96%)	66 (90%)	7 (10%)	8 30
3	C	72/76 (95%)	56 (78%)	16 (22%)	1 4
All	All	388/567 (68%)	323 (83%)	65 (17%)	2 9

5 of 65 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	W	12	GLU
2	W	126	VAL
3	C	164	LEU
2	W	25	PHE
2	W	50	GLU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (4) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	545	GLN
1	A	600	GLN
1	A	632	GLN
3	B	108	GLN

5.3.3 RNA (i)

There are no RNA molecules in this entry.

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

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

5.5 Carbohydrates

There are no carbohydrates 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	131/319 (41%)	-0.17	1 (0%) 86 77	6, 25, 64, 83	0
2	W	140/151 (92%)	-0.17	1 (0%) 87 81	8, 31, 78, 93	0
3	B	84/90 (93%)	-0.09	0 100 100	14, 30, 65, 89	0
3	C	85/90 (94%)	0.19	4 (4%) 31 18	12, 35, 106, 132	0
All	All	440/650 (67%)	-0.09	6 (1%) 75 63	6, 30, 78, 132	0

The worst 5 of 6 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
3	C	192	ASN	6.2
1	A	541	LEU	5.6
3	C	190	LEU	5.6
3	C	189	SER	5.1
2	W	130	GLY	3.3

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 carbohydrates 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.