

Full wwPDB X-ray Structure Validation Report (i)

Sep 28, 2024 – 10:05 pm BST

PDB ID : 2J80

Title : Structure of Citrate-bound Periplasmic Domain of Sensor Histidine Kinase

CitA

Authors: Sevvana, M.; Vijayan, V.; Zweckstetter, M.; Reinelt, S.; Madden, D.R.;

Sheldrick, G.M.; Bott, M.; Griesinger, C.; Becker, S.

Deposited on : 2006-10-18

Resolution : 1.60 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 (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:

 $Mol Probity \quad : \quad 4.02b\text{-}467$

Mogul : 1.8.4, CSD as541be (2020)

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)

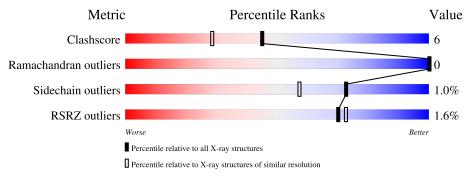
Validation Pipeline (wwPDB-VP) : 2.39

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 1.60 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	Similar resolution
Metric	$(\# \mathrm{Entries})$	$(\# ext{Entries}, ext{ resolution range}(ext{Å}))$
Clashscore	180529	4682 (1.60-1.60)
Ramachandran outliers	177936	4583 (1.60-1.60)
Sidechain outliers	177891	4582 (1.60-1.60)
RSRZ outliers	164620	4272 (1.60-1.60)

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	135	91%			
1	В	135	84%	7% • 8%		

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
5	GOL	В	1132	_	-	X	-



2 Entry composition (i)

There are 6 unique types of molecules in this entry. The entry contains 2136 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.

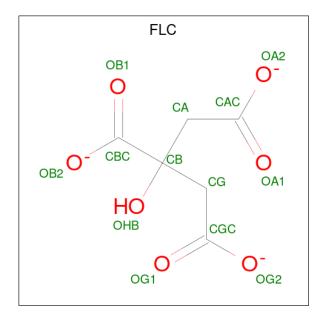
• Molecule 1 is a protein called SENSOR KINASE CITA.

\mathbf{Mol}	Chain	Residues	Atoms			ZeroOcc	AltConf	Trace			
1	A	129	Total 950		N 163	O 189	Se 4	0	0	0	
1	В	124	Total 927		N 164	O 181	Se 4	0	1	0	

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-1	GLY	-	expression tag	UNP P52687
A	0	HIS	-	expression tag	UNP P52687
A	1	MSE	-	expression tag	UNP P52687
В	-1	GLY	-	expression tag	UNP P52687
В	0	HIS	-	expression tag	UNP P52687
В	1	MSE	-	expression tag	UNP P52687

• Molecule 2 is CITRATE ANION (three-letter code: FLC) (formula: $C_6H_5O_7$).



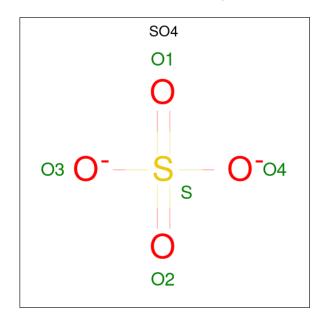


Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	1	Total C O 13 6 7	0	0
2	В	1	Total C O 13 6 7	0	0

• Molecule 3 is SODIUM ION (three-letter code: NA) (formula: Na).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	1	Total Na 1 1	0	0
3	В	1	Total Na 1 1	0	0

 \bullet Molecule 4 is SULFATE ION (three-letter code: SO4) (formula: $\mathrm{O_4S}).$



Mol	Chain	Residues	Ato	oms		ZeroOcc	AltConf
4	A	1	Total 5	O 4	S 1	0	0

 \bullet Molecule 5 is GLYCEROL (three-letter code: GOL) (formula: $\mathrm{C_3H_8O_3}).$





Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	A	1	Total C O 6 3 3	0	0
5	В	1	Total C O 6 3 3	0	0

• Molecule 6 is water.

\mathbf{Mol}	Chain	Residues	Atoms	ZeroOcc	AltConf
6	A	110	Total O 110 110	0	0
6	В	104	Total O 104 104	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: SENSOR KINASE CITA

Chain A:

91%

• Molecule 1: SENSOR KINASE CITA

Chain B:

84%

7% • 8%



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants	53.21Å 54.79Å 116.33Å	Donositon
a, b, c, α , β , γ	90.00° 90.00° 90.00°	Depositor
Resolution (Å)	30.00 - 1.60	Depositor
Resolution (A)	30.00 - 1.60	EDS
% Data completeness	96.6 (30.00-1.60)	Depositor
(in resolution range)	88.1 (30.00-1.60)	EDS
R_{merge}	0.05	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) > 1$	2.10 (at 1.48Å)	Xtriage
Refinement program	REFMAC 5.2.0005	Depositor
D D.	0.193 , 0.217	Depositor
R, R_{free}	0.186 , (Not available)	DCC
R_{free} test set	No test flags present.	wwPDB-VP
Wilson B-factor (Å ²)	19.2	Xtriage
Anisotropy	0.196	Xtriage
Bulk solvent $k_{sol}(e/Å^3)$, $B_{sol}(Å^2)$	0.39, 55.0	EDS
L-test for twinning ²	$< L >=0.49, < L^2>=0.32$	Xtriage
Estimated twinning fraction	0.016 for k,h,-l	Xtriage
F_o, F_c correlation	0.96	EDS
Total number of atoms	2136	wwPDB-VP
Average B, all atoms (Å ²)	21.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: The largest off-origin peak in the Patterson function is 14.96% of the height of the origin peak. No significant pseudotranslation is detected.

²Theoretical 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.



¹Intensities estimated from amplitudes.

5 Model quality (i)

5.1 Standard geometry (i)

Bond lengths and bond angles in the following residue types are not validated in this section: NA, GOL, SO4, FLC

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	Boı	nd lengths	Bond angles		
IVIOI	Mol Chain		# Z > 5	RMSZ	# Z > 5	
1	A	1.01	0/957	0.75	0/1287	
1	В	1.00	1/934~(0.1%)	0.75	0/1255	
All	All	1.00	1/1891 (0.1%)	0.75	0/2542	

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	$Ideal(\AA)$
1	В	26	MSE	SE-CE	-7.18	1.53	1.95

There are no bond angle outliers.

There are no chirality outliers.

There are no planarity outliers.

5.2 Too-close contacts (i)

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	950	0	932	13	0
1	В	927	0	922	18	0
2	A	13	0	5	0	0
2	В	13	0	5	0	0
3	A	1	0	0	0	0
3	В	1	0	0	0	0
4	A	5	0	0	0	0
5	A	6	0	6	0	0

Continued on next page...



Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
5	В	6	0	6	4	0
6	A	110	0	0	4	0
6	В	104	0	0	2	0
All	All	2136	0	1876	22	0

The all-atom clash score is defined as the number of clashes found per 1000 atoms (including hydrogen atoms). The all-atom clash score for this structure is 6.

All (22) 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:26:MSE:CE	1:A:26:MSE:SE	2.20	1.40
1:A:26:MSE:HE2	1:B:26:MSE:CE	1.53	1.38
1:A:26:MSE:HE2	1:B:26:MSE:HE3	1.33	1.06
1:B:65:GLN:HB3	6:B:2056:HOH:O	1.55	1.04
1:A:26:MSE:HE2	1:B:26:MSE:HE2	1.43	0.97
1:A:26:MSE:CE	1:B:26:MSE:CE	2.44	0.96
1:A:26:MSE:CE	1:B:26:MSE:HE2	2.06	0.81
1:B:21:MSE:HE3	5:B:1132:GOL:H32	1.68	0.76
1:A:86:GLU:HA	6:A:2077:HOH:O	1.85	0.76
1:B:21:MSE:HE3	5:B:1132:GOL:C3	2.22	0.69
1:A:26:MSE:SE	1:B:26:MSE:HE2	2.46	0.66
1:A:71:ASN:HB3	6:A:2064:HOH:O	2.01	0.61
6:A:2021:HOH:O	1:B:26:MSE:SE	2.72	0.57
1:B:21:MSE:CE	5:B:1132:GOL:H31	2.38	0.54
1:B:118:LYS:HE3	6:B:2095:HOH:O	2.08	0.54
1:A:18:ILE:HD12	1:B:15:ARG:HD2	1.94	0.49
1:A:26:MSE:CE	1:B:26:MSE:HE3	2.24	0.47
1:A:26:MSE:HE2	1:B:26:MSE:HE1	1.79	0.45
1:A:44:LEU:HG	6:A:2034:HOH:O	2.16	0.45
1:B:56:TYR:CZ	1:B:126:GLY:HA3	2.52	0.45
1:B:49:ARG:HD3	1:B:68:TYR:OH	2.20	0.42
1:B:21:MSE:CE	5:B:1132:GOL:C3	2.92	0.41

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	Perce	entiles	
1	A	127/135~(94%)	127 (100%)	0	0	100	100	
1	В	123/135~(91%)	123 (100%)	0	0	100	100	
All	All	250/270~(93%)	250 (100%)	0	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	97/106 (92%)	96 (99%)	1 (1%)	73 57		
1	В	96/106 (91%)	95 (99%)	1 (1%)	73 57		
All	All	193/212 (91%)	191 (99%)	2 (1%)	73 57		

All (2) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	107	ARG
1	В	107	ARG

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	65	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 (i)

There are no oligosaccharides in this entry.

5.6 Ligand geometry (i)

Of 7 ligands modelled in this entry, 2 are monoatomic - leaving 5 for Mogul analysis.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the Chemical Component Dictionary. The Link column lists molecule types, if any, to which the group is linked. 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| > 2 is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths				Bond angles		
MIOI	Type	Chain	nes	LIIIK	Counts	RMSZ	# Z > 2	Counts	RMSZ	# Z > 2	
2	FLC	В	1130	-	12,12,12	1.41	1 (8%)	17,17,17	1.75	3 (17%)	
5	GOL	В	1132	3	5,5,5	0.40	0	5,5,5	0.72	0	
4	SO4	A	1135	-	4,4,4	0.17	0	6,6,6	0.43	0	
5	GOL	A	1136	3	5,5,5	0.33	0	5,5,5	0.70	0	
2	FLC	A	1133	-	12,12,12	1.44	2 (16%)	17,17,17	1.60	2 (11%)	

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the Chemical Component Dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
5	GOL	В	1132	3	-	0/4/4/4	-
5	GOL	A	1136	3	-	0/4/4/4	-
2	FLC	В	1130	-	-	2/16/16/16	-
2	FLC	A	1133	-	-	2/16/16/16	-



All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	$\operatorname{Observed}(\text{\AA})$	$\operatorname{Ideal}(ext{\AA})$
2	A	1133	FLC	CB-CBC	2.48	1.56	1.53
2	A	1133	FLC	OHB-CB	2.36	1.47	1.43
2	В	1130	FLC	CB-CBC	2.19	1.55	1.53

All (5) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	\mathbf{Z}	$\mathbf{Observed}(^o)$	$\operatorname{Ideal}({}^{o})$
2	В	1130	FLC	OB2-CBC-CB	4.56	120.97	113.05
2	A	1133	FLC	OB2-CBC-CB	3.88	119.78	113.05
2	В	1130	FLC	OHB-CB-CBC	2.64	112.57	108.86
2	A	1133	FLC	OA2-CAC-CA	2.41	122.08	114.35
2	В	1130	FLC	OA2-CAC-CA	2.24	121.56	114.35

There are no chirality outliers.

All (4) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	A	1133	FLC	CB-CG-CGC-OG2
2	В	1130	FLC	CB-CA-CAC-OA2
2	A	1133	FLC	CB-CG-CGC-OG1
2	В	1130	FLC	CB-CA-CAC-OA1

There are no ring outliers.

1 monomer is involved in 4 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
5	В	1132	GOL	4	0

5.7 Other polymers (i)

There are no such residues in this entry.

5.8 Polymer linkage issues (i)

There are no chain breaks in this entry.



6 Fit of model and data (i)

6.1 Protein, DNA and RNA chains (i)

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 $>$	$\# \mathrm{RSRZ}{>}2$	$OWAB(A^2)$	Q < 0.9
1	A	125/135~(92%)	-0.18	1 (0%) 82 85	13, 19, 27, 31	0
1	В	120/135 (88%)	-0.05	3 (2%) 58 60	11, 20, 29, 38	1 (0%)
All	All	245/270 (90%)	-0.12	4 (1%) 70 73	11, 19, 28, 38	1 (0%)

All (4) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	В	6	GLU	3.7
1	В	7	ARG	2.3
1	A	4	THR	2.2
1	В	10	TYR	2.1

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	$ \mathbf{B}\text{-factors}(\mathbf{A}^2) $	Q<0.9
4	SO4	A	1135	5/5	0.89	0.17	24,24,26,26	5

Continued on next page...



 $Continued\ from\ previous\ page...$

Mol	Type	Chain	Res	Atoms	RSCC	RSR	$\mathbf{B} ext{-}\mathbf{factors}(\mathbf{\mathring{A}}^2)$	Q<0.9
5	GOL	В	1132	6/6	0.89	0.15	19,27,31,34	0
5	GOL	A	1136	6/6	0.92	0.12	16,20,26,30	0
2	FLC	В	1130	13/13	0.96	0.06	14,15,19,19	0
2	FLC	A	1133	13/13	0.96	0.06	15,16,19,21	0
3	NA	В	1131	1/1	0.99	0.02	16,16,16,16	0
3	NA	A	1134	1/1	0.99	0.02	15,15,15,15	0

6.5 Other polymers (i)

There are no such residues in this entry.

