Full wwPDB X-ray Structure Validation Report (i)

Sep 2, 2024 – 05:11 pm BST

PDB ID	:	9ESA
Title	:	Aurora-C with SER mutation in complex with INCENP peptide
Authors	:	Hillig, R.C.
Deposited on	:	2024-03-26
Resolution	:	2.80 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 (i)) were used in the production of this report:

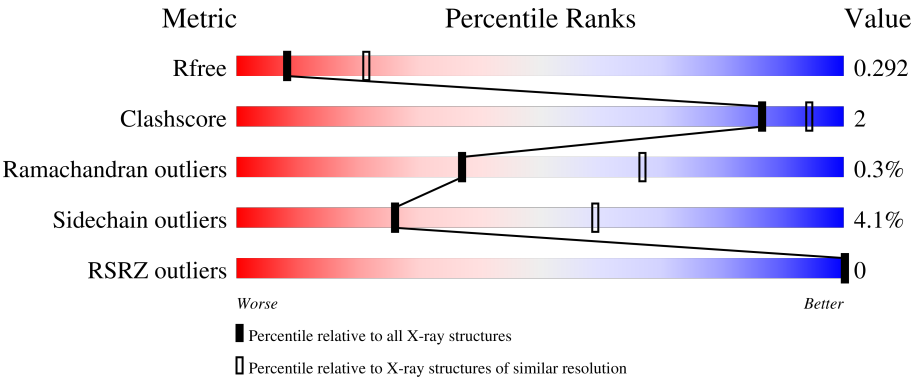
MolProbity	:	4.02b-467
Mogul	:	1.8.4, CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	3.0
Percentile statistics	:	20231227.v01 (using entries in the PDB archive December 27th 2023)
CCP4	:	9.0.002 (Gargrove)
Density-Fitness	:	1.0.11
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.38.2

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 2.80 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}	164625	3657 (2.80-2.80)
Clashscore	180529	4123 (2.80-2.80)
Ramachandran outliers	177936	4071 (2.80-2.80)
Sidechain outliers	177891	4073 (2.80-2.80)
RSRZ outliers	164620	3659 (2.80-2.80)

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	AAA	303	82% 9% • 9%
1	BBB	303	83% 8% 9%
2	CCC	58	71% • 26%
2	DDD	58	72% • 24%

2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 5297 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 Aurora kinase C.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	AAA	275	Total	C	N	O	S	0	0	0
			2252	1449	399	395	9			
1	BBB	275	Total	C	N	O	S	0	3	0
			2275	1463	404	399	9			

There are 18 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
AAA	7	HIS	-	expression tag	UNP Q9UQB9
AAA	8	HIS	-	expression tag	UNP Q9UQB9
AAA	9	HIS	-	expression tag	UNP Q9UQB9
AAA	10	HIS	-	expression tag	UNP Q9UQB9
AAA	11	HIS	-	expression tag	UNP Q9UQB9
AAA	12	HIS	-	expression tag	UNP Q9UQB9
AAA	195	ALA	ARG	engineered mutation	UNP Q9UQB9
AAA	196	ALA	ARG	engineered mutation	UNP Q9UQB9
AAA	197	ALA	LYS	engineered mutation	UNP Q9UQB9
BBB	7	HIS	-	expression tag	UNP Q9UQB9
BBB	8	HIS	-	expression tag	UNP Q9UQB9
BBB	9	HIS	-	expression tag	UNP Q9UQB9
BBB	10	HIS	-	expression tag	UNP Q9UQB9
BBB	11	HIS	-	expression tag	UNP Q9UQB9
BBB	12	HIS	-	expression tag	UNP Q9UQB9
BBB	195	ALA	ARG	engineered mutation	UNP Q9UQB9
BBB	196	ALA	ARG	engineered mutation	UNP Q9UQB9
BBB	197	ALA	LYS	engineered mutation	UNP Q9UQB9

- Molecule 2 is a protein called Inner centromere protein.

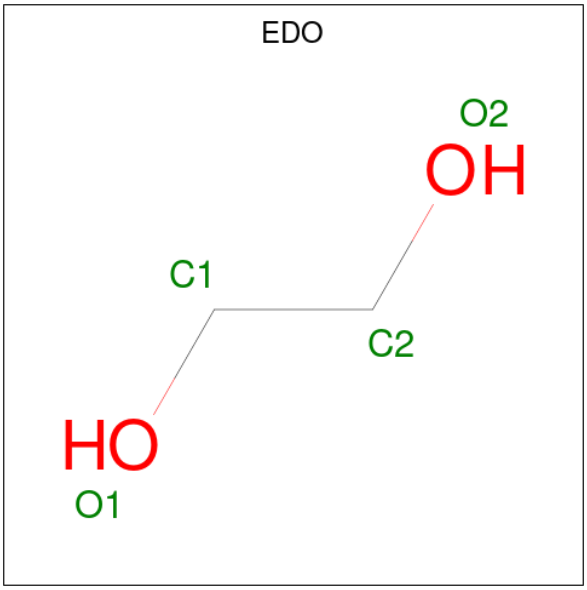
Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
2	CCC	43	Total	C	N	O	0	0	0
			352	236	56	60			

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
2	DDD	44	Total	C	N	O	0	0	0
			358	239	57	62			

- Molecule 3 is 1,2-ETHANEDIOL (three-letter code: EDO) (formula: $C_2H_6O_2$).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
3	BBB	1	Total	C	O	0	0
			4	2	2		

- Molecule 4 is water.

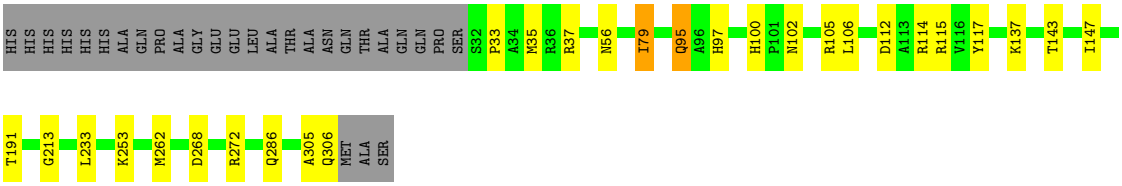
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	AAA	28	Total	O	0	0
			28	28		
4	BBB	25	Total	O	0	0
			25	25		
4	CCC	1	Total	O	0	0
			1	1		
4	DDD	2	Total	O	0	0
			2	2		

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: Aurora kinase C

Chain AAA:

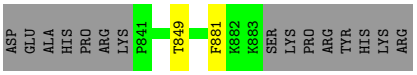


- Molecule 1: Aurora kinase C

Chain BBB:

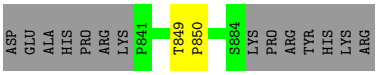
- Molecule 2: Inner centromere protein

Chain CCC:



- Molecule 2: Inner centromere protein

Chain DDD:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 2 2 21	Depositor
Cell constants a, b, c, α, β, γ	79.28Å 79.49Å 265.24Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	19.87 – 2.80 19.87 – 2.80	Depositor EDS
% Data completeness (in resolution range)	99.3 (19.87-2.80) 98.9 (19.87-2.80)	Depositor EDS
R_{merge}	0.13	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.33 (at 2.79Å)	Xtriage
Refinement program	REFMAC 5.8.0267	Depositor
R, R_{free}	0.228 , 0.294 0.230 , 0.292	Depositor DCC
R_{free} test set	1075 reflections (5.07%)	wwPDB-VP
Wilson B-factor (Å^2)	65.1	Xtriage
Anisotropy	0.047	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.29 , 40.4	EDS
L-test for twinning[2]	$< \|L\| > = 0.53$, $< L^2 > = 0.37$	Xtriage
Estimated twinning fraction	0.429 for -k,-h,-l	Xtriage
F_o,F_c correlation	0.96	EDS
Total number of atoms	5297	wwPDB-VP
Average B, all atoms (Å^2)	74.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.81% 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

Bond lengths and bond angles in the following residue types are not validated in this section: EDO

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	AAA	0.65	0/2307	0.75	0/3122
1	BBB	0.65	0/2336	0.76	0/3160
2	CCC	0.63	0/365	0.68	0/500
2	DDD	0.64	0/371	0.68	0/508
All	All	0.65	0/5379	0.74	0/7290

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	AAA	0	1

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	AAA	213	GLY	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	AAA	2252	0	2291	9	0
1	BBB	2275	0	2317	10	0
2	CCC	352	0	357	1	0
2	DDD	358	0	362	2	0
3	BBB	4	0	6	0	0
4	AAA	28	0	0	0	0
4	BBB	25	0	0	0	0
4	CCC	1	0	0	0	0
4	DDD	2	0	0	0	0
All	All	5297	0	5333	21	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 2.

All (21) 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:AAA:95:GLN:HG2	1:AAA:106:LEU:HB3	1.78	0.66
1:BBB:90:ARG:CZ	1:BBB:94:ILE:HD11	2.27	0.65
1:BBB:90:ARG:NH2	1:BBB:94:ILE:HD11	2.19	0.58
1:AAA:268:ASP:O	1:AAA:272:ARG:HG3	2.09	0.53
1:BBB:277:GLN:HB3	1:BBB:280:GLU:HG2	1.92	0.52
1:AAA:79:ILE:HD11	2:CCC:881:PHE:CZ	2.44	0.51
1:AAA:56:ASN:HD22	1:AAA:56:ASN:N	2.12	0.46
1:AAA:100:HIS:CD2	1:AAA:102:ASN:H	2.34	0.45
1:BBB:202:THR:O	1:BBB:202:THR:CG2	2.65	0.45
1:BBB:100:HIS:CD2	1:BBB:102:ASN:H	2.35	0.44
1:AAA:143:THR:HG21	1:AAA:233:LEU:HG	2.00	0.44
2:DDD:849:THR:N	2:DDD:850:PRO:CD	2.81	0.43
1:AAA:143:THR:O	1:AAA:147:ILE:HG12	2.19	0.43
1:BBB:158:HIS:NE2	1:BBB:218[A]:GLU:OE1	2.50	0.43
1:AAA:305:ALA:O	1:AAA:306:GLN:HG2	2.20	0.42
1:BBB:94:ILE:HD12	1:BBB:162:VAL:HG11	2.01	0.42
1:BBB:143:THR:O	1:BBB:147:ILE:HG12	2.20	0.41
1:BBB:143:THR:HG21	1:BBB:233:LEU:HG	2.02	0.41
2:DDD:849:THR:OG1	2:DDD:850:PRO:HD3	2.21	0.41
1:BBB:290:HIS:CD2	1:BBB:291:PRO:HD2	2.56	0.41
1:AAA:115:ARG:HB3	1:AAA:117:TYR:CE1	2.56	0.40

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	AAA	273/303 (90%)	254 (93%)	18 (7%)	1 (0%)	30 61
1	BBB	276/303 (91%)	260 (94%)	15 (5%)	1 (0%)	30 61
2	CCC	41/58 (71%)	41 (100%)	0	0	100 100
2	DDD	42/58 (72%)	41 (98%)	1 (2%)	0	100 100
All	All	632/722 (88%)	596 (94%)	34 (5%)	2 (0%)	37 67

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	AAA	33	PRO
1	BBB	33	PRO

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	AAA	243/264 (92%)	230 (95%)	13 (5%)	19 49
1	BBB	246/264 (93%)	237 (96%)	9 (4%)	29 63
2	CCC	39/53 (74%)	38 (97%)	1 (3%)	41 75
2	DDD	40/53 (76%)	40 (100%)	0	100 100
All	All	568/634 (90%)	545 (96%)	23 (4%)	26 60

All (23) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	AAA	35	MET
1	AAA	37	ARG
1	AAA	79	ILE
1	AAA	95	GLN
1	AAA	97	HIS
1	AAA	105	ARG
1	AAA	112	ASP
1	AAA	114	ARG
1	AAA	137	LYS
1	AAA	191	THR
1	AAA	253	LYS
1	AAA	262	MET
1	AAA	286	GLN
1	BBB	35	MET
1	BBB	51	LYS
1	BBB	88	LEU
1	BBB	89	ARG
1	BBB	97	HIS
1	BBB	112	ASP
1	BBB	137	LYS
1	BBB	202	THR
1	BBB	267	ARG
2	CCC	849	THR

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

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)

1 ligand is modelled in this entry.

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		
					Counts	RMSZ	$\#\|Z\| > 2$	Counts	RMSZ	$\#\|Z\| > 2$
3	EDO	BBB	400	-	3,3,3	0.07	0	2,2,2	0.21	0

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
3	EDO	BBB	400	-	-	1/1/1/1	-

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (1) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	BBB	400	EDO	O1-C1-C2-O2

There are no ring outliers.

No monomer is involved in short contacts.

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

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	AAA	275/303 (90%)	-1.52	0	100	100	41, 71, 107, 152	0
1	BBB	275/303 (90%)	-1.45	0	100	100	26, 71, 106, 149	3 (1%)
2	CCC	43/58 (74%)	-1.39	0	100	100	51, 74, 106, 112	0
2	DDD	44/58 (75%)	-1.38	0	100	100	54, 74, 116, 127	0
All	All	637/722 (88%)	-1.47	0	100	100	26, 72, 108, 152	3 (0%)

There are no RSRZ outliers to report.

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

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	B-factors(Å^2)	Q<0.9
3	EDO	BBB	400	4/4	0.96	0.08	77,84,89,90	0

6.5 Other polymers (i)

There are no such residues in this entry.