wwPDB X-ray Structure Validation Summary Report (i)

Feb 6, 2024 – 12:01 PM EST

PDB ID	:	2A10
Title	:	carboxysome shell protein ccmK4
Authors	:	Kerfeld, C.A.; Sawaya, M.R.; Tanaka, S.; Nguyen, C.V.; Phillips, M.; Beeby, M.; Yeates, T.O.
Deposited on	:	2005-06-17
Resolution	:	1.80 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:

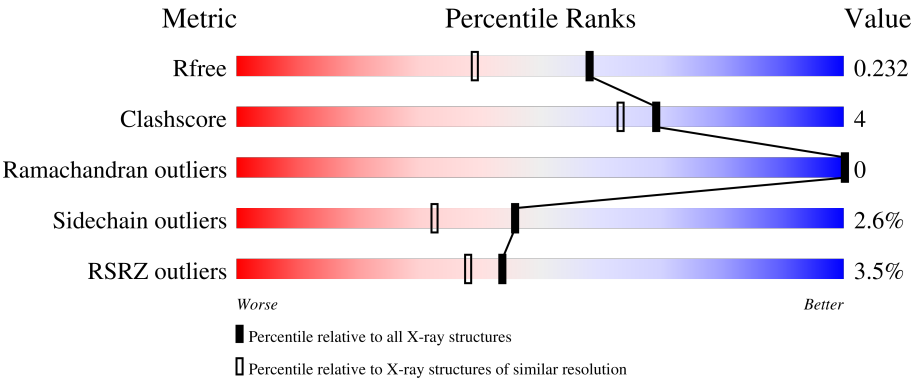
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36

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 1.80 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}	130704	5950 (1.80-1.80)
Clashscore	141614	6793 (1.80-1.80)
Ramachandran outliers	138981	6697 (1.80-1.80)
Sidechain outliers	138945	6696 (1.80-1.80)
RSRZ outliers	127900	5850 (1.80-1.80)

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	A	125	4% 78% • • 17%
1	B	125	5% 78% 6% • 15%
1	C	125	7% 81% • • 14%
1	D	125	73% 6% • 18%
1	E	125	% 74% 6% • 18%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	F	125	

2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 4962 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 Carbon dioxide concentrating mechanism protein ccmK homolog 4.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	104	Total	C	N	O	S	0	1	0
			771	490	136	143	2			
1	B	106	Total	C	N	O	S	0	1	0
			782	496	139	145	2			
1	C	107	Total	C	N	O	S	0	1	0
			792	503	140	147	2			
1	D	102	Total	C	N	O	S	0	2	0
			767	486	135	144	2			
1	E	102	Total	C	N	O	S	0	1	0
			764	485	135	142	2			
1	F	105	Total	C	N	O	S	0	2	0
			780	494	138	146	2			

There are 84 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	MET	-	initiating methionine	UNP P73407
A	113	GLY	-	expression tag	UNP P73407
A	114	VAL	-	expression tag	UNP P73407
A	115	PRO	-	expression tag	UNP P73407
A	116	ARG	-	expression tag	UNP P73407
A	117	GLY	-	expression tag	UNP P73407
A	118	LEU	-	expression tag	UNP P73407
A	119	GLU	-	expression tag	UNP P73407
A	120	HIS	-	expression tag	UNP P73407
A	121	HIS	-	expression tag	UNP P73407
A	122	HIS	-	expression tag	UNP P73407
A	123	HIS	-	expression tag	UNP P73407
A	124	HIS	-	expression tag	UNP P73407
A	125	HIS	-	expression tag	UNP P73407
B	1	MET	-	initiating methionine	UNP P73407
B	113	GLY	-	expression tag	UNP P73407

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	114	VAL	-	expression tag	UNP P73407
B	115	PRO	-	expression tag	UNP P73407
B	116	ARG	-	expression tag	UNP P73407
B	117	GLY	-	expression tag	UNP P73407
B	118	LEU	-	expression tag	UNP P73407
B	119	GLU	-	expression tag	UNP P73407
B	120	HIS	-	expression tag	UNP P73407
B	121	HIS	-	expression tag	UNP P73407
B	122	HIS	-	expression tag	UNP P73407
B	123	HIS	-	expression tag	UNP P73407
B	124	HIS	-	expression tag	UNP P73407
B	125	HIS	-	expression tag	UNP P73407
C	1	MET	-	initiating methionine	UNP P73407
C	113	GLY	-	expression tag	UNP P73407
C	114	VAL	-	expression tag	UNP P73407
C	115	PRO	-	expression tag	UNP P73407
C	116	ARG	-	expression tag	UNP P73407
C	117	GLY	-	expression tag	UNP P73407
C	118	LEU	-	expression tag	UNP P73407
C	119	GLU	-	expression tag	UNP P73407
C	120	HIS	-	expression tag	UNP P73407
C	121	HIS	-	expression tag	UNP P73407
C	122	HIS	-	expression tag	UNP P73407
C	123	HIS	-	expression tag	UNP P73407
C	124	HIS	-	expression tag	UNP P73407
C	125	HIS	-	expression tag	UNP P73407
D	1	MET	-	initiating methionine	UNP P73407
D	113	GLY	-	expression tag	UNP P73407
D	114	VAL	-	expression tag	UNP P73407
D	115	PRO	-	expression tag	UNP P73407
D	116	ARG	-	expression tag	UNP P73407
D	117	GLY	-	expression tag	UNP P73407
D	118	LEU	-	expression tag	UNP P73407
D	119	GLU	-	expression tag	UNP P73407
D	120	HIS	-	expression tag	UNP P73407
D	121	HIS	-	expression tag	UNP P73407
D	122	HIS	-	expression tag	UNP P73407
D	123	HIS	-	expression tag	UNP P73407
D	124	HIS	-	expression tag	UNP P73407
D	125	HIS	-	expression tag	UNP P73407
E	1	MET	-	initiating methionine	UNP P73407
E	113	GLY	-	expression tag	UNP P73407

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
E	114	VAL	-	expression tag	UNP P73407
E	115	PRO	-	expression tag	UNP P73407
E	116	ARG	-	expression tag	UNP P73407
E	117	GLY	-	expression tag	UNP P73407
E	118	LEU	-	expression tag	UNP P73407
E	119	GLU	-	expression tag	UNP P73407
E	120	HIS	-	expression tag	UNP P73407
E	121	HIS	-	expression tag	UNP P73407
E	122	HIS	-	expression tag	UNP P73407
E	123	HIS	-	expression tag	UNP P73407
E	124	HIS	-	expression tag	UNP P73407
E	125	HIS	-	expression tag	UNP P73407
F	1	MET	-	initiating methionine	UNP P73407
F	113	GLY	-	expression tag	UNP P73407
F	114	VAL	-	expression tag	UNP P73407
F	115	PRO	-	expression tag	UNP P73407
F	116	ARG	-	expression tag	UNP P73407
F	117	GLY	-	expression tag	UNP P73407
F	118	LEU	-	expression tag	UNP P73407
F	119	GLU	-	expression tag	UNP P73407
F	120	HIS	-	expression tag	UNP P73407
F	121	HIS	-	expression tag	UNP P73407
F	122	HIS	-	expression tag	UNP P73407
F	123	HIS	-	expression tag	UNP P73407
F	124	HIS	-	expression tag	UNP P73407
F	125	HIS	-	expression tag	UNP P73407

- Molecule 2 is water.

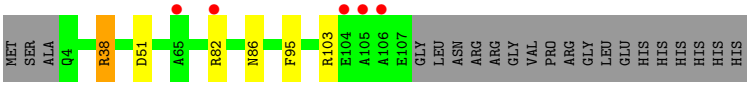
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	46	Total O 46 46	0	0
2	B	57	Total O 57 57	0	0
2	C	53	Total O 53 53	0	0
2	D	45	Total O 45 45	0	0
2	E	58	Total O 58 58	0	0
2	F	47	Total O 47 47	0	0

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.

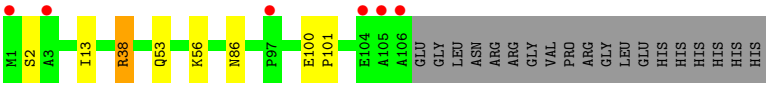
• Molecule 1: Carbon dioxide concentrating mechanism protein ccmK homolog 4

Chain A:



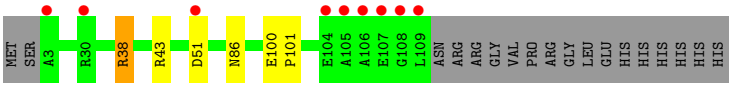
• Molecule 1: Carbon dioxide concentrating mechanism protein ccmK homolog 4

Chain B:



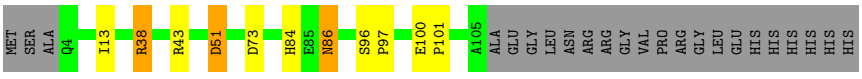
• Molecule 1: Carbon dioxide concentrating mechanism protein ccmK homolog 4

Chain C:



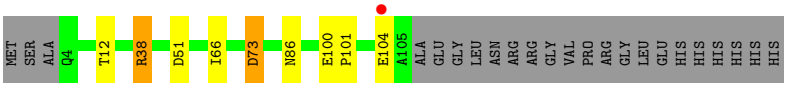
• Molecule 1: Carbon dioxide concentrating mechanism protein ccmK homolog 4

Chain D:



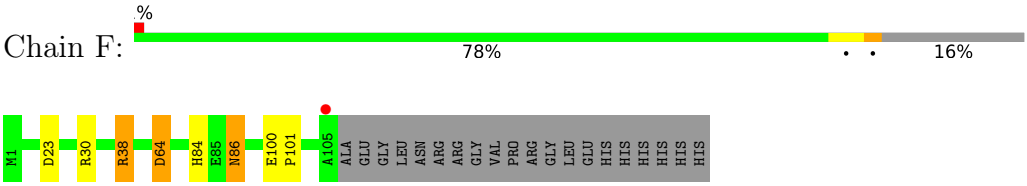
• Molecule 1: Carbon dioxide concentrating mechanism protein ccmK homolog 4

Chain E:



• Molecule 1: Carbon dioxide concentrating mechanism protein ccmK homolog 4

Chain F:


%
78%
16%
M1
D23
R30
R38
D64
H84
E85
N86
E100
P101
A105
ALA
GLU
GLY
LEU
ASN
ARG
ARG
GLY
VAL
PRO
ARG
GLY
LEU
GLU
HIS
HIS
HIS
HIS
HIS
HIS

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	84.76Å 87.00Å 96.51Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	64.55 – 1.80 39.66 – 1.80	Depositor EDS
% Data completeness (in resolution range)	98.2 (64.55-1.80) 98.2 (39.66-1.80)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.08	Depositor
$< I/\sigma(I) >$ [1]	4.41 (at 1.81Å)	Xtriage
Refinement program	REFMAC 5.2	Depositor
R, R_{free}	0.193 , 0.227 0.204 , 0.232	Depositor DCC
R_{free} test set	3340 reflections (5.12%)	wwPDB-VP
Wilson B-factor (Å^2)	22.1	Xtriage
Anisotropy	0.219	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.37 , 50.8	EDS
L-test for twinning[2]	$< \|L\| > = 0.51$, $< L^2 > = 0.35$	Xtriage
Estimated twinning fraction	0.007 for k,h,-l	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	4962	wwPDB-VP
Average B, all atoms (Å^2)	25.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.70% 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	2.18	2/792 (0.3%)	1.48	7/1077 (0.6%)
1	B	5.10	2/803 (0.2%)	1.72	6/1092 (0.5%)
1	C	6.76	2/813 (0.2%)	1.87	8/1105 (0.7%)
1	D	4.30	2/793 (0.3%)	1.65	9/1078 (0.8%)
1	E	1.28	2/785 (0.3%)	7.39	8/1067 (0.7%)
1	F	4.62	2/806 (0.2%)	1.58	9/1096 (0.8%)
All	All	4.45	12/4792 (0.3%)	3.36	47/6515 (0.7%)

The worst 5 of 12 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	C	38[A]	ARG	NE-CZ	135.82	3.09	1.33
1	C	38[B]	ARG	NE-CZ	135.82	3.09	1.33
1	B	38[A]	ARG	NE-CZ	101.55	2.65	1.33
1	B	38[B]	ARG	NE-CZ	101.55	2.65	1.33
1	F	38[A]	ARG	NE-CZ	92.18	2.52	1.33

The worst 5 of 47 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	E	38[A]	ARG	NE-CZ-NH1	-131.31	54.65	120.30
1	E	38[B]	ARG	NE-CZ-NH1	-131.31	54.65	120.30
1	E	38[A]	ARG	NE-CZ-NH2	107.04	173.82	120.30
1	E	38[B]	ARG	NE-CZ-NH2	107.04	173.82	120.30
1	C	38[A]	ARG	CD-NE-CZ	-37.91	70.52	123.60

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	771	0	783	4	0
1	B	782	0	798	4	0
1	C	792	0	808	3	0
1	D	767	0	780	11	0
1	E	764	0	780	7	0
1	F	780	0	793	6	0
2	A	46	0	0	2	0
2	B	57	0	0	1	0
2	C	53	0	0	3	0
2	D	45	0	0	3	0
2	E	58	0	0	8	0
2	F	47	0	0	3	0
All	All	4962	0	4742	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 4.

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:38[B]:ARG:NH1	2:A:171:HOH:O	1.92	1.01
1:D:38[B]:ARG:CZ	2:E:182:HOH:O	2.23	0.86
1:F:38[B]:ARG:NH1	2:F:172:HOH:O	2.07	0.85
1:E:38[B]:ARG:NH1	2:E:182:HOH:O	2.17	0.77
1:D:38[B]:ARG:CZ	2:D:170:HOH:O	2.34	0.75

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	
1	A	103/125 (82%)	103 (100%)	0	0	100	100
1	B	105/125 (84%)	105 (100%)	0	0	100	100
1	C	106/125 (85%)	106 (100%)	0	0	100	100
1	D	102/125 (82%)	102 (100%)	0	0	100	100
1	E	101/125 (81%)	101 (100%)	0	0	100	100
1	F	105/125 (84%)	105 (100%)	0	0	100	100
All	All	622/750 (83%)	622 (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	78/96 (81%)	76 (97%)	2 (3%)	46	32
1	B	79/96 (82%)	77 (98%)	2 (2%)	47	34
1	C	80/96 (83%)	79 (99%)	1 (1%)	69	62
1	D	80/96 (83%)	77 (96%)	3 (4%)	33	18
1	E	79/96 (82%)	77 (98%)	2 (2%)	47	34
1	F	80/96 (83%)	76 (95%)	4 (5%)	24	10
All	All	476/576 (83%)	462 (97%)	14 (3%)	46	29

5 of 14 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	D	86	ASN
1	E	73	ASP
1	F	86	ASN
1	F	64[A]	ASP
1	F	64[B]	ASP

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 7 such sidechains are listed below:

Mol	Chain	Res	Type
1	D	84	HIS
1	E	67	ASN
1	F	84	HIS
1	F	47	ASN
1	D	67	ASN

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 monosaccharides in this entry.

5.6 Ligand geometry (i)

There are no ligands in this entry.

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	A	104/125 (83%)	0.25	5 (4%) 30 25	20, 24, 35, 43	0
1	B	106/125 (84%)	0.20	6 (5%) 23 19	19, 22, 34, 44	0
1	C	107/125 (85%)	0.42	9 (8%) 11 8	19, 22, 37, 49	0
1	D	102/125 (81%)	-0.01	0 100 100	20, 23, 34, 40	0
1	E	102/125 (81%)	0.02	1 (0%) 82 80	20, 24, 35, 44	0
1	F	105/125 (84%)	-0.05	1 (0%) 82 80	19, 24, 37, 45	0
All	All	626/750 (83%)	0.14	22 (3%) 44 38	19, 23, 37, 49	0

The worst 5 of 22 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	108	GLY	6.6
1	C	109	LEU	6.4
1	C	106	ALA	5.6
1	C	107	GLU	5.1
1	C	105	ALA	4.7

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.