W O R L D W I D E

wwPDB

P R O T E I N   D A T A   B A N K

wwPDB X-ray Structure Validation Summary Report ⓘ

Dec 2, 2023 – 03:21 pm GMT

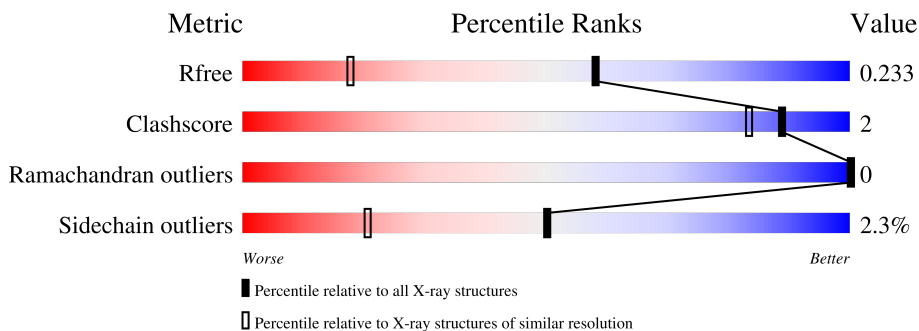|		
PDB ID	:	2VHS
Title	:	Cathsilicatein, a chimera
Authors	:	Fairhead, M.; Kowatz, T.; McMahon, S.A.; Carter, L.G.; Oke, M.; Johnson, K.A.; Liu, H.; Naismith, J.H.; Wal, C.F.V.D.
Deposited on	:	2007-11-24
Resolution	:	1.50 Å(reported)

1 Overall quality at a glance ⓘ

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

The reported resolution of this entry is 1.50 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	Percentile Ranks	Value
Rfree		0.233
Clashscore		2
Ramachandran outliers		0
Sidechain outliers		2.3%

Worse ▬ *Better*

▮ Percentile relative to all X-ray structures

▯ Percentile relative to X-ray structures of similar resolution

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R$_{free}$	130704	2936 (1.50-1.50)
Clashscore	141614	3144 (1.50-1.50)
Ramachandran outliers	138981	3066 (1.50-1.50)
Sidechain outliers	138945	3064 (1.50-1.50)

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%

Mol	Chain	Length	Quality of chain
1	A	217	90% 9% •
1	B	217	94% 5% •
1	C	217	94% 5% •
1	D	217	90% 9% •

WORLDWIDE
wwPDB
PROTEIN DATA BANK

2   Entry composition ⓘ

There are 3 unique types of molecules in this entry. The entry contains 7848 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 CATHSILICATEIN.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	215	Total	C	N	O	S	0	3	0
			1661	1047	275	326	13			
1	B	215	Total	C	N	O	S	0	2	0
			1657	1043	275	326	13			
1	C	215	Total	C	N	O	S	0	3	0
			1661	1047	275	326	13			
1	D	215	Total	C	N	O	S	0	2	0
			1657	1043	275	326	13			

There are 48 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	110	VAL	THR	conflict	UNP P07711
B	110	VAL	THR	conflict	UNP P07711
C	110	VAL	THR	conflict	UNP P07711
D	110	VAL	THR	conflict	UNP P07711
A	24	ALA	SER	engineered mutation	UNP P07711
A	25	SER	CYS	engineered mutation	UNP P07711
A	26	TYR	TRP	engineered mutation	UNP P07711
A	153	SER	GLU	engineered mutation	UNP P07711
A	154	SER	PRO	engineered mutation	UNP P07711
A	159	SER	GLU	engineered mutation	UNP P07711
A	160	SER	ASP	engineered mutation	UNP P07711
A	161	LEU	MET	engineered mutation	UNP P07711
A	162	ASN	ASP	engineered mutation	UNP P07711
A	164	ALA	GLY	engineered mutation	UNP P07711
A	165	MET	VAL	engineered mutation	UNP P07711
B	24	ALA	SER	engineered mutation	UNP P07711
B	25	SER	CYS	engineered mutation	UNP P07711
B	26	TYR	TRP	engineered mutation	UNP P07711
B	153	SER	GLU	engineered mutation	UNP P07711
B	154	SER	PRO	engineered mutation	UNP P07711
B	159	SER	GLU	engineered mutation	UNP P07711

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	160	SER	ASP	engineered mutation	UNP P07711
B	161	LEU	MET	engineered mutation	UNP P07711
B	162	ASN	ASP	engineered mutation	UNP P07711
B	164	ALA	GLY	engineered mutation	UNP P07711
B	165	MET	VAL	engineered mutation	UNP P07711
C	24	ALA	SER	engineered mutation	UNP P07711
C	25	SER	CYS	engineered mutation	UNP P07711
C	26	TYR	TRP	engineered mutation	UNP P07711
C	153	SER	GLU	engineered mutation	UNP P07711
C	154	SER	PRO	engineered mutation	UNP P07711
C	159	SER	GLU	engineered mutation	UNP P07711
C	160	SER	ASP	engineered mutation	UNP P07711
C	161	LEU	MET	engineered mutation	UNP P07711
C	162	ASN	ASP	engineered mutation	UNP P07711
C	164	ALA	GLY	engineered mutation	UNP P07711
C	165	MET	VAL	engineered mutation	UNP P07711
D	24	ALA	SER	engineered mutation	UNP P07711
D	25	SER	CYS	engineered mutation	UNP P07711
D	26	TYR	TRP	engineered mutation	UNP P07711
D	153	SER	GLU	engineered mutation	UNP P07711
D	154	SER	PRO	engineered mutation	UNP P07711
D	159	SER	GLU	engineered mutation	UNP P07711
D	160	SER	ASP	engineered mutation	UNP P07711
D	161	LEU	MET	engineered mutation	UNP P07711
D	162	ASN	ASP	engineered mutation	UNP P07711
D	164	ALA	GLY	engineered mutation	UNP P07711
D	165	MET	VAL	engineered mutation	UNP P07711

- Molecule 2 is SULFATE ION (three-letter code: SO4) (formula: O_4S).

SO4

O1

O

O3 O⁻ — S — O⁻ O4

S

O

O2

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
2	A	1	Total 5	O 4	S 1	0	0
2	A	1	Total 5	O 4	S 1	0	0
2	B	1	Total 5	O 4	S 1	0	0
2	B	1	Total 5	O 4	S 1	0	0
2	C	1	Total 5	O 4	S 1	0	0
2	C	1	Total 5	O 4	S 1	0	0
2	D	1	Total 5	O 4	S 1	0	0
2	D	1	Total 5	O 4	S 1	0	0

- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	303	Total 303	O 303	0	0
3	B	264	Total 264	O 264	0	0
3	C	307	Total 307	O 307	0	0
3	D	298	Total 298	O 298	0	0

3 Residue-property plots ⓘ

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. 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. 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.
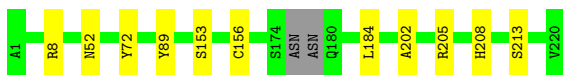
- Molecule 1: CATHSILICATEIN

Chain A:

90%   9%   ·

A1 P2 R3 · R8 · N52 · Y72 · Y89 D109 V110 · S133 · I150 Y151 F152 · C156 · I173 S174 ASN ASN Q180 · Y198 · A202 · R205 · S213 A214 A215 · V220

- Molecule 1: CATHSILICATEIN

Chain B:

94%   5% ·

A1 · R8 · N52 · Y72 · Y89 · S153 · C156 · S174 ASN ASN Q180 · L184 · A202 · R205 · H208 · S213 V220

- Molecule 1: CATHSILICATEIN

Chain C:

94%   5% ·

A1 · R8 E9 K10 · N52 · Y89 · S153 · C156 · S159 · S174 ASN ASN Q180 · Y198 · A202 · R205 · H208 · V220

- Molecule 1: CATHSILICATEIN

Chain D:

90%   9% ·

A1 · W7 · N52 · G61 · Y89 · D109 · K120 · K124 · G130 · H140 · L144 · I150 Y151 F152 · C156 · I173 S174 ASN ASN Q180 · A202 · R205 · H208 · V220

4 Data and refinement statistics ⓘ

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, α, β, γ	56.83Å 58.10Å 70.21Å 105.67° 104.97° 105.05°	Depositor
Resolution (Å)	63.25 – 1.50 27.40 – 1.50	Depositor EDS
% Data completeness (in resolution range)	91.7 (63.25-1.50) 80.1 (27.40-1.50)	Depositor EDS
R_{merge}	0.04	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.23 (at 1.50Å)	Xtriage
Refinement program	REFMAC 5.2.0019	Depositor
R, R_{free}	0.188 , 0.220 0.196 , 0.233	Depositor DCC
R_{free} test set	5042 reflections (5.04%)	wwPDB-VP
Wilson B-factor (Å^2)	16.5	Xtriage
Anisotropy	0.097	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.41 , 46.0	EDS
L-test for twinning[2]	$< |L| > = 0.51, < L^2 > = 0.34$	Xtriage
Estimated twinning fraction	0.036 for -k,-h,-l	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	7848	wwPDB-VP
Average B, all atoms (Å^2)	26.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 85.58 % of the origin peak, indicating pseudo-translational symmetry. The chance of finding a peak of this or larger height randomly in a structure without pseudo-translational symmetry is equal to 1.2083e-07. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

---

[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: SO4

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	0.33	0/1709	0.50	0/2306
1	B	0.32	0/1702	0.49	0/2296
1	C	0.34	0/1709	0.51	0/2306
1	D	0.34	0/1702	0.50	0/2296
All	All	0.33	0/6822	0.50	0/9204

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	1661	0	1576	9	0
1	B	1657	0	1567	5	0
1	C	1661	0	1576	5	0
1	D	1657	0	1567	9	0
2	A	10	0	0	0	0
2	B	10	0	0	0	0
2	C	10	0	0	0	0
2	D	10	0	0	0	0
3	A	303	0	0	2	0

Continued on next page...

WORLDWIDE
wwPDB
PROTEIN DATA BANK

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
3	B	264	0	0	0	0
3	C	307	0	0	1	0
3	D	298	0	0	1	0
All	All	7848	0	6286	25	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.

The worst 5 of 25 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:202:ALA:HB1	1:A:205:ARG:HD3	1.78	0.66
1:A:72:TYR:OH	1:B:213:SER:OG	2.10	0.62
1:C:202:ALA:HB1	1:C:205:ARG:HD3	1.85	0.57
1:D:180:GLN:N	3:D:2254:HOH:O	2.40	0.54
1:A:213:SER:OG	1:B:72:TYR:OH	2.13	0.52

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	214/217 (99%)	211 (99%)	3 (1%)	0	100	100
1	B	213/217 (98%)	209 (98%)	4 (2%)	0	100	100
1	C	214/217 (99%)	211 (99%)	3 (1%)	0	100	100
1	D	213/217 (98%)	209 (98%)	4 (2%)	0	100	100
All	All	854/868 (98%)	840 (98%)	14 (2%)	0	100	100

There are no Ramachandran outliers to report.

5.3.2 Protein sidechains ⓘ

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	176/175 (101%)	170 (97%)	6 (3%)	37	9
1	B	175/175 (100%)	172 (98%)	3 (2%)	60	33
1	C	176/175 (101%)	173 (98%)	3 (2%)	60	33
1	D	175/175 (100%)	171 (98%)	4 (2%)	50	20
All	All	702/700 (100%)	686 (98%)	16 (2%)	50	20

5 of 16 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	D	109	ASP
1	D	89	TYR
1	B	156	CYS
1	D	52	ASN
1	B	89	TYR

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
1	C	180	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 ⓘ

8 ligands are 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	$\#|Z| > 2$
2	SO4	B	1221	-	4,4,4	0.18	0	6,6,6	0.18	0
2	SO4	B	1222	-	4,4,4	0.13	0	6,6,6	0.12	0
2	SO4	D	1221	-	4,4,4	0.21	0	6,6,6	0.28	0
2	SO4	A	1222	-	4,4,4	0.14	0	6,6,6	0.16	0
2	SO4	C	1222	-	4,4,4	0.14	0	6,6,6	0.15	0
2	SO4	C	1221	-	4,4,4	0.16	0	6,6,6	0.27	0
2	SO4	A	1221	-	4,4,4	0.19	0	6,6,6	0.25	0
2	SO4	D	1222	-	4,4,4	0.17	0	6,6,6	0.14	0

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

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 ⓘ

Unable to reproduce the depositors R factor - this section is therefore empty.

6.2　Non-standard residues in protein, DNA, RNA chains ⓘ

Unable to reproduce the depositors R factor - this section is therefore empty.

6.3　Carbohydrates ⓘ

Unable to reproduce the depositors R factor - this section is therefore empty.

6.4　Ligands ⓘ

Unable to reproduce the depositors R factor - this section is therefore empty.

6.5　Other polymers ⓘ

Unable to reproduce the depositors R factor - this section is therefore empty.