WORLDWIDE wwPDB PROTEIN DATA BANK

wwPDB X-ray Structure Validation Summary Report (i)

Jul 14, 2025 – 12:39 PM JST

PDB ID	:	9IS1 / pdb_00009is1
Title	:	Pasteurization-treated beta-conglycinin
Authors	:	Zhang, T.; Li, J.Y.
Deposited on	:	2024-07-16
Resolution	:	2.01 Å(reported)

The following versions of software and data (see references (i)) were used in the production of this report:

|		
MolProbity	:	4-5-2 with Phenix2.0rc1
Xtriage (Phenix)	:	2.0rc1
EDS	:	3.0
Percentile statistics	:	20231227.v01 (using entries in the PDB archive December 27th 2023)
CCP4	:	9.0.006 (Gargrove)
Density-Fitness	:	1.0.12
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.44

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 2.01 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		0.293
Clashscore		14
Ramachandran outliers		0.6%
Sidechain outliers		5.4%
RSRZ outliers		6.5%

Worse — Better

▌ Percentile relative to all X-ray structures

▯ Percentile relative to X-ray structures of similar resolution
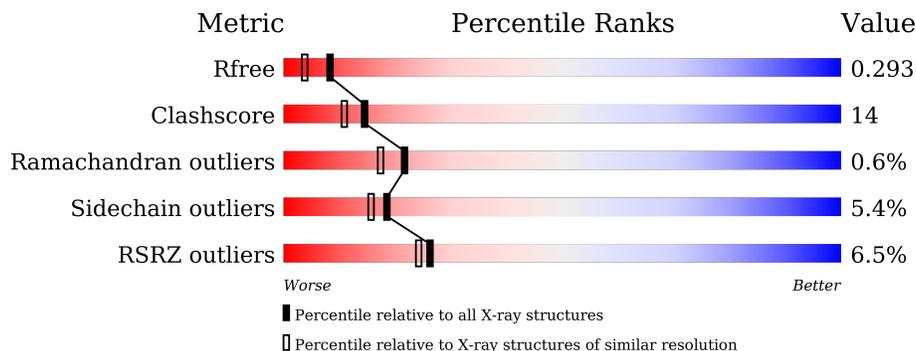
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R$_{free}$	164625	9409 (2.00-2.00)
Clashscore	180529	10737 (2.00-2.00)
Ramachandran outliers	177936	10628 (2.00-2.00)
Sidechain outliers	177891	10627 (2.00-2.00)
RSRZ outliers	164620	9409 (2.00-2.00)

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	385	7% / 66% / 26% / · ·
1	B	385	5% / 71% / 23% / · ·
1	C	385	6% / 66% / 28% / · ·

W O R L D W I D E
wwPDB
PROTEIN DATA BANK

2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 9461 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 Beta-conglycinin beta subunit 2.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
1	C	368	Total	C	N	O	0	0	0
			2995	1890	537	568			
1	A	368	Total	C	N	O	0	0	0
			2995	1890	537	568			
1	B	368	Total	C	N	O	0	0	0
			2995	1890	537	568			

- Molecule 2 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	C	163	Total	O	0	0
			163	163		
2	A	171	Total	O	0	0
			171	171		
2	B	142	Total	O	0	0
			142	142		

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 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: Beta-conglycinin beta subunit 2
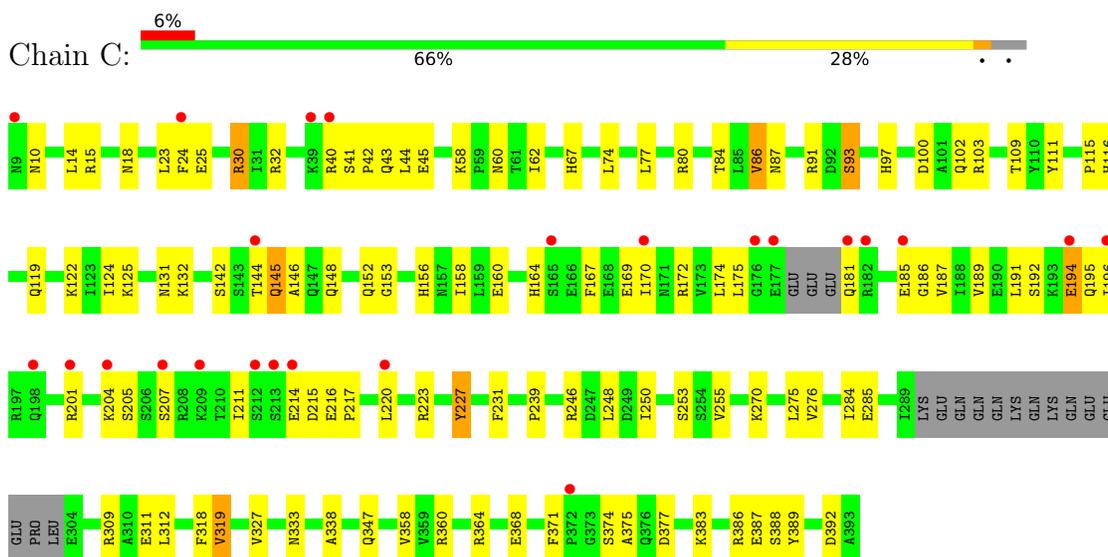
Chain C:

• Molecule 1: Beta-conglycinin beta subunit 2

Chain A:

• Molecule 1: Beta-conglycinin beta subunit 2

Chain B:

5%

71%

23%
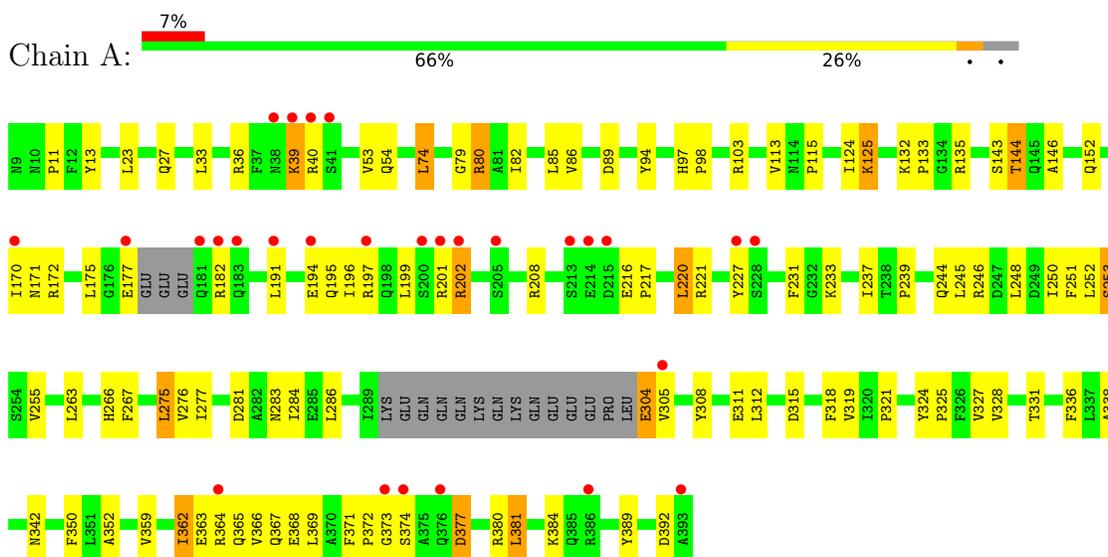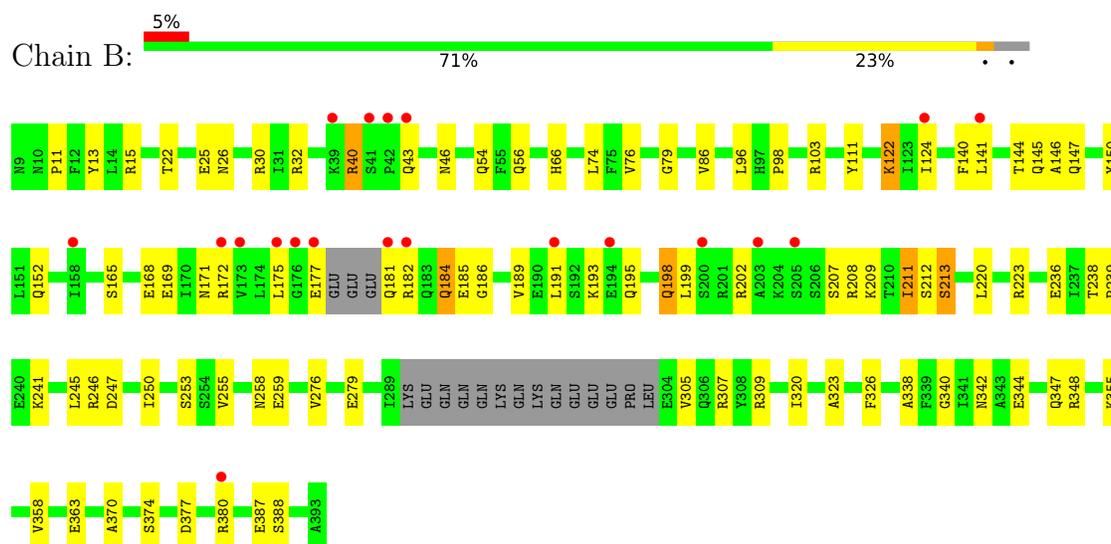
4 Data and refinement statistics ⓘ

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	85.11Å 83.46Å 86.92Å 90.00° 118.36° 90.00°	Depositor
Resolution (Å)	37.45 – 2.01 37.45 – 2.01	Depositor EDS
% Data completeness (in resolution range)	89.5 (37.45-2.01) 90.9 (37.45-2.01)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.79 (at 2.01Å)	Xtriage
Refinement program	PHENIX 1.20.1_4487	Depositor
R, R_{free}	0.244 , 0.294 0.244 , 0.293	Depositor DCC
R_{free} test set	69398 reflections (2.95%)	wwPDB-VP
Wilson B-factor (Å^2)	27.3	Xtriage
Anisotropy	0.188	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.35 , 41.3	EDS
L-test for twinning [2]	$< |L| > = 0.49$, $< L^2 > = 0.33$	Xtriage
Estimated twinning fraction	0.000 for -h-l,k,h 0.000 for l,k,-h-l 0.022 for h,-k,-h-l 0.017 for -h-l,-k,l 0.017 for l,-k,h	Xtriage
F_o,F_c correlation	0.92	EDS
Total number of atoms	9461	wwPDB-VP
Average B, all atoms (Å^2)	32.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.57% 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 (i)

5.1 Standard geometry (i)

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	0.39	0/3053	0.63	1/4129 (0.0%)
1	B	0.35	0/3053	0.58	1/4129 (0.0%)
1	C	0.41	0/3053	0.64	0/4129
All	All	0.38	0/9159	0.62	2/12387 (0.0%)

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	B	220	LEU	CA-CB-CG	6.43	138.80	116.30
1	A	74	LEU	CA-CB-CG	5.23	134.59	116.30

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	2995	0	2941	91	1
1	B	2995	0	2941	79	0
1	C	2995	0	2941	92	1
2	A	171	0	0	18	0
2	B	142	0	0	13	0
2	C	163	0	0	25	0
All	All	9461	0	8823	245	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 14.

The worst 5 of 245 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:ARG:NH2	2:A:401:HOH:O	1.85	1.09
1:A:36:ARG:HE	1:A:39:LYS:HG3	1.31	0.94
1:C:181:GLN:N	2:C:402:HOH:O	1.99	0.94
1:B:207:SER:OG	1:B:209:LYS:HD3	1.71	0.91
1:C:214:GLU:OE1	2:C:401:HOH:O	1.89	0.89

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:C:386:ARG:O	1:A:39:LYS:NZ[2_545]	2.17	0.03

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	362/385 (94%)	348 (96%)	12 (3%)	2 (1%)	22	17
1	B	362/385 (94%)	351 (97%)	10 (3%)	1 (0%)	37	35
1	C	362/385 (94%)	349 (96%)	10 (3%)	3 (1%)	16	12
All	All	1086/1155 (94%)	1048 (96%)	32 (3%)	6 (1%)	22	17

5 of 6 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	C	145	GLN
1	A	202	ARG
1	A	220	LEU

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	C	10	ASN
1	C	42	PRO

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	332/349 (95%)	310 (93%)	22 (7%)	14	10
1	B	332/349 (95%)	320 (96%)	12 (4%)	30	30
1	C	332/349 (95%)	312 (94%)	20 (6%)	16	13
All	All	996/1047 (95%)	942 (95%)	54 (5%)	18	16

5 of 54 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	201	ARG
1	A	304	GLU
1	B	211	ILE
1	A	216	GLU
1	A	255	VAL

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 22 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	54	GLN
1	B	156	HIS
1	B	147	GLN
1	B	164	HIS
1	A	87	ASN

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 oligosaccharides 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, 95th 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	368/385 (95%)	0.61	28 (7%)	21	20	17, 29, 50, 69	0
1	B	368/385 (95%)	0.67	20 (5%)	32	30	17, 33, 51, 70	0
1	C	368/385 (95%)	0.55	24 (6%)	26	24	17, 28, 50, 63	0
All	All	1104/1155 (95%)	0.61	72 (6%)	26	24	17, 30, 50, 70	0

The worst 5 of 72 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	201	ARG	5.2
1	C	181	GLN	5.0
1	C	212	SER	4.3
1	C	177	GLU	4.3
1	C	9	ASN	4.2

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 oligosaccharides in this entry.

6.4 Ligands ⓘ

There are no ligands in this entry.

WORLDWIDE
wwPDB
PROTEIN DATA BANK

6.5 Other polymers ⓘ

There are no such residues in this entry.