wwPDB X-ray Structure Validation Summary Report (i)

Nov 2, 2023 – 04:41 PM EDT

PDB ID	:	3USV
Title	:	Structure of the precursor of a thermostable variant of papain at 3.8 A resolution from a crystal soaked at pH 4
Authors	:	Roy, S.; Choudhury, D.; Biswas, S.; Dattagupta, J.K.
Deposited on	:	2011-11-24
Resolution	:	3.80 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:

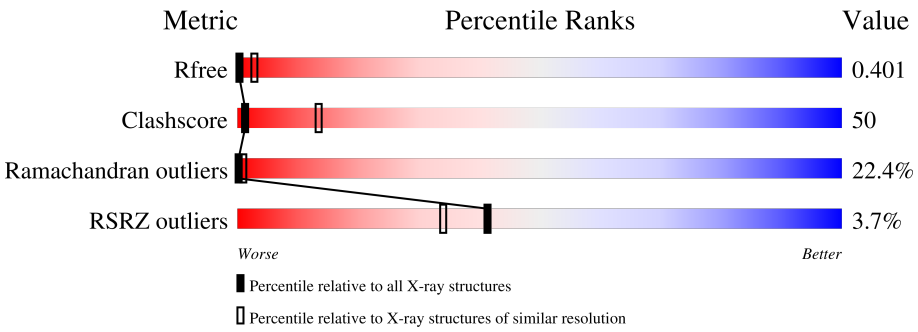
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation 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 3.80 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	1212 (4.00-3.60)
Clashscore	141614	1288 (4.00-3.60)
Ramachandran outliers	138981	1243 (4.00-3.60)
RSRZ outliers	127900	1121 (4.00-3.60)

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	363	3% 43% 31% 10% • 15%
1	C	363	3% 42% 33% 9% • 15%

2 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 3020 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 Papain.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
			Total	C	N	O			
1	A	308	1510	893	308	309	0	0	0
1	C	308	1510	893	308	309	0	0	0

There are 96 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-43	MET	-	expression tag	UNP P00784
A	-42	HIS	-	expression tag	UNP P00784
A	-41	HIS	-	expression tag	UNP P00784
A	-40	HIS	-	expression tag	UNP P00784
A	-39	HIS	-	expression tag	UNP P00784
A	-38	HIS	-	expression tag	UNP P00784
A	-37	HIS	-	expression tag	UNP P00784
A	-36	SER	-	expression tag	UNP P00784
A	-35	SER	-	expression tag	UNP P00784
A	-34	GLY	-	expression tag	UNP P00784
A	-33	LEU	-	expression tag	UNP P00784
A	-32	VAL	-	expression tag	UNP P00784
A	-31	PRO	-	expression tag	UNP P00784
A	-30	ARG	-	expression tag	UNP P00784
A	-29	GLY	-	expression tag	UNP P00784
A	-28	SER	-	expression tag	UNP P00784
A	-27	GLY	-	expression tag	UNP P00784
A	-26	MET	-	expression tag	UNP P00784
A	-25	LYS	-	expression tag	UNP P00784
A	-24	GLU	-	expression tag	UNP P00784
A	-23	THR	-	expression tag	UNP P00784
A	-22	ALA	-	expression tag	UNP P00784
A	-21	ALA	-	expression tag	UNP P00784
A	-20	ALA	-	expression tag	UNP P00784
A	-19	LYS	-	expression tag	UNP P00784

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
A	-18	PHE	-	expression tag	UNP P00784
A	-17	GLU	-	expression tag	UNP P00784
A	-16	ARG	-	expression tag	UNP P00784
A	-15	GLN	-	expression tag	UNP P00784
A	-14	HIS	-	expression tag	UNP P00784
A	-13	MET	-	expression tag	UNP P00784
A	-12	ASP	-	expression tag	UNP P00784
A	-11	SER	-	expression tag	UNP P00784
A	-10	PRO	-	expression tag	UNP P00784
A	-9	ASP	-	expression tag	UNP P00784
A	-8	LEU	-	expression tag	UNP P00784
A	-7	GLY	-	expression tag	UNP P00784
A	-6	THR	-	expression tag	UNP P00784
A	-5	ASP	-	expression tag	UNP P00784
A	-4	ASP	-	expression tag	UNP P00784
A	-3	ASP	-	expression tag	UNP P00784
A	-2	ASP	-	expression tag	UNP P00784
A	-1	LYS	-	expression tag	UNP P00784
A	0	MET	-	expression tag	UNP P00784
A	132	ALA	CYS	engineered mutation	UNP P00784
A	139	SER	VAL	engineered mutation	UNP P00784
A	143	SER	GLY	engineered mutation	UNP P00784
A	281	ARG	LYS	engineered mutation	UNP P00784
C	-43	MET	-	expression tag	UNP P00784
C	-42	HIS	-	expression tag	UNP P00784
C	-41	HIS	-	expression tag	UNP P00784
C	-40	HIS	-	expression tag	UNP P00784
C	-39	HIS	-	expression tag	UNP P00784
C	-38	HIS	-	expression tag	UNP P00784
C	-37	HIS	-	expression tag	UNP P00784
C	-36	SER	-	expression tag	UNP P00784
C	-35	SER	-	expression tag	UNP P00784
C	-34	GLY	-	expression tag	UNP P00784
C	-33	LEU	-	expression tag	UNP P00784
C	-32	VAL	-	expression tag	UNP P00784
C	-31	PRO	-	expression tag	UNP P00784
C	-30	ARG	-	expression tag	UNP P00784
C	-29	GLY	-	expression tag	UNP P00784
C	-28	SER	-	expression tag	UNP P00784
C	-27	GLY	-	expression tag	UNP P00784
C	-26	MET	-	expression tag	UNP P00784
C	-25	LYS	-	expression tag	UNP P00784

Continued on next page...

Continued from previous page...

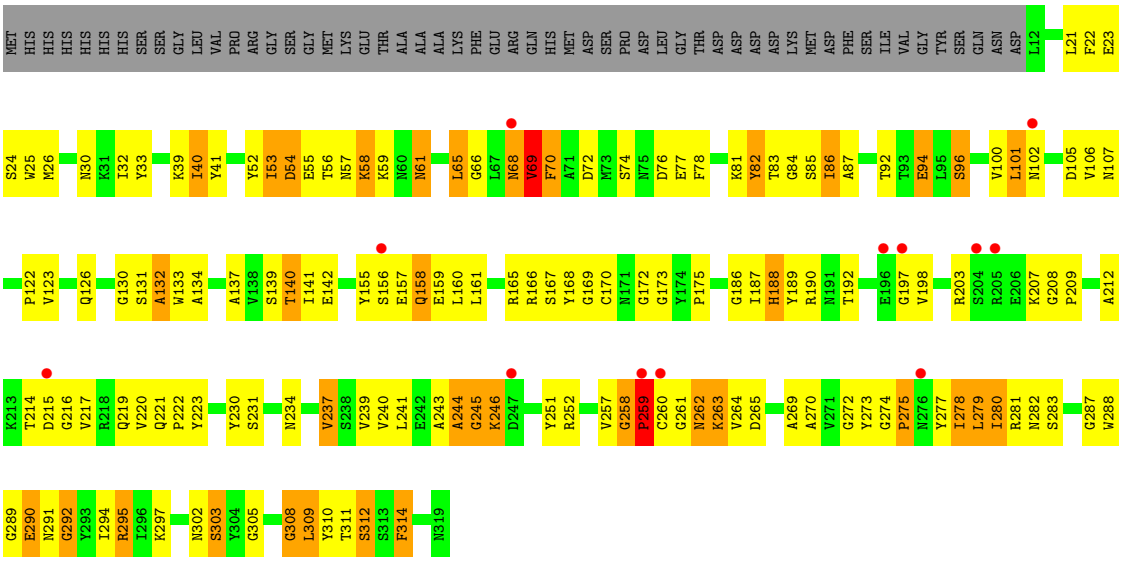
Chain	Residue	Modelled	Actual	Comment	Reference
C	-24	GLU	-	expression tag	UNP P00784
C	-23	THR	-	expression tag	UNP P00784
C	-22	ALA	-	expression tag	UNP P00784
C	-21	ALA	-	expression tag	UNP P00784
C	-20	ALA	-	expression tag	UNP P00784
C	-19	LYS	-	expression tag	UNP P00784
C	-18	PHE	-	expression tag	UNP P00784
C	-17	GLU	-	expression tag	UNP P00784
C	-16	ARG	-	expression tag	UNP P00784
C	-15	GLN	-	expression tag	UNP P00784
C	-14	HIS	-	expression tag	UNP P00784
C	-13	MET	-	expression tag	UNP P00784
C	-12	ASP	-	expression tag	UNP P00784
C	-11	SER	-	expression tag	UNP P00784
C	-10	PRO	-	expression tag	UNP P00784
C	-9	ASP	-	expression tag	UNP P00784
C	-8	LEU	-	expression tag	UNP P00784
C	-7	GLY	-	expression tag	UNP P00784
C	-6	THR	-	expression tag	UNP P00784
C	-5	ASP	-	expression tag	UNP P00784
C	-4	ASP	-	expression tag	UNP P00784
C	-3	ASP	-	expression tag	UNP P00784
C	-2	ASP	-	expression tag	UNP P00784
C	-1	LYS	-	expression tag	UNP P00784
C	0	MET	-	expression tag	UNP P00784
C	132	ALA	CYS	engineered mutation	UNP P00784
C	139	SER	VAL	engineered mutation	UNP P00784
C	143	SER	GLY	engineered mutation	UNP P00784
C	281	ARG	LYS	engineered mutation	UNP P00784

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.

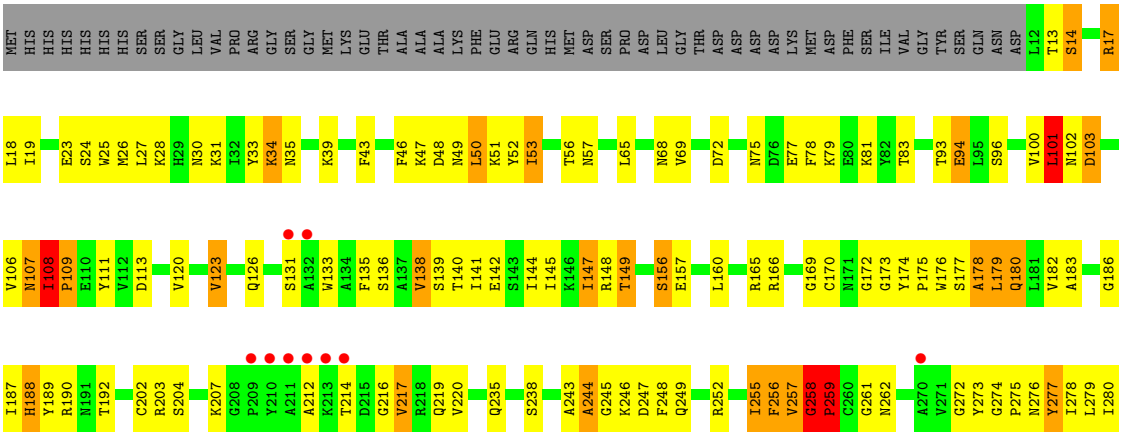
- Molecule 1: Papain

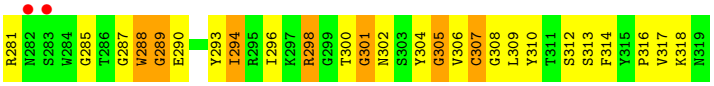
Chain A:



- Molecule 1: Papain

Chain C:




R281
N282
S283
W284
G285
T286
G287
W288
G289
E290
Y293
I294
R295
I296
K297
R298
G299
T300
G301
N302
S303
Y304
G305
V306
C307
G308
L309
Y310
T311
S312
S313
F314
Y315
P316
V317
K318
N319

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	42.85Å 75.97Å 109.71Å 90.00° 97.71° 90.00°	Depositor
Resolution (Å)	29.60 – 3.80 29.60 – 3.80	Depositor EDS
% Data completeness (in resolution range)	(Not available) (29.60-3.80) 70.5 (29.60-3.80)	Depositor EDS
R_{merge}	0.12	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.35 (at 3.75Å)	Xtriage
Refinement program	CNS	Depositor
R, R_{free}	0.358 , 0.408 0.359 , 0.401	Depositor DCC
R_{free} test set	579 reflections (9.95%)	wwPDB-VP
Wilson B-factor (Å^2)	103.7	Xtriage
Anisotropy	0.540	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	(Not available) , (Not available)	EDS
L-test for twinning[2]	$< \|L\| > = 0.22$, $< L^2 > = 0.09$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.77	EDS
Total number of atoms	3020	wwPDB-VP
Average B, all atoms (Å^2)	44.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 9.13% 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	0.56	1/1509 (0.1%)	1.08	10/2091 (0.5%)
1	C	0.56	0/1509	1.06	9/2091 (0.4%)
All	All	0.56	1/3018 (0.0%)	1.07	19/4182 (0.5%)

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	C	0	1

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	258	GLY	C-O	-5.12	1.15	1.23

The worst 5 of 19 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	C	259	PRO	N-CA-C	-13.93	75.88	112.10
1	A	258	GLY	N-CA-C	10.08	138.30	113.10
1	A	259	PRO	N-CA-C	-7.51	92.56	112.10
1	C	94	GLU	N-CA-C	7.03	129.98	111.00
1	A	100	VAL	N-CA-C	6.99	129.87	111.00

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	C	259	PRO	Mainchain

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	1510	0	687	123	0
1	C	1510	0	687	97	2
All	All	3020	0	1374	220	2

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 50.

The worst 5 of 220 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:C:217:VAL:HA	1:C:316:PRO:HA	1.05	1.05
1:A:168:TYR:O	1:A:173:GLY:HA2	1.62	0.98
1:C:217:VAL:CA	1:C:316:PRO:HA	1.96	0.96
1:C:219:GLN:HA	1:C:314:PHE:HA	1.48	0.94
1:A:170:CYS:C	1:A:172:GLY:H	1.75	0.89

All (2) 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:102:ASN:O	1:C:192:THR:O[1_655]	2.03	0.17
1:C:103:ASP:O	1:C:192:THR:O[1_655]	2.17	0.03

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	A	306/363 (84%)	164 (54%)	73 (24%)	69 (22%)	0 1
1	C	306/363 (84%)	165 (54%)	73 (24%)	68 (22%)	0 1
All	All	612/726 (84%)	329 (54%)	146 (24%)	137 (22%)	0 1

5 of 137 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	53	ILE
1	A	54	ASP
1	A	70	PHE
1	A	74	SER
1	A	82	TYR

5.3.2 Protein sidechains (i)

There are no protein residues with a non-rotameric sidechain to report in this entry.

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 (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, 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	308/363 (84%)	-0.17	12 (3%)	39	32	19, 45, 57, 59	0
1	C	308/363 (84%)	-0.22	11 (3%)	42	35	19, 45, 58, 60	0
All	All	616/726 (84%)	-0.20	23 (3%)	41	34	19, 45, 58, 60	0

The worst 5 of 23 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	214	THR	4.4
1	C	209	PRO	4.4
1	C	211	ALA	3.9
1	A	102	ASN	3.7
1	A	156	SER	3.5

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)

There are no ligands in this entry.

6.5 Other polymers (i)

There are no such residues in this entry.