WORLDWIDE PDB PROTEIN DATA BANK

wwPDB X-ray Structure Validation Summary Report (i)

Jun 16, 2024 – 05:51 PM EDT

PDB ID	:	3FZ3
Title	:	Crystal Structure of almond Pru1 protein
Authors	:	Jin, T.C.; Zhang, Y.Z.
Deposited on	:	2009-01-23
Resolution	:	2.40 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 (i)) were used in the production of this report:

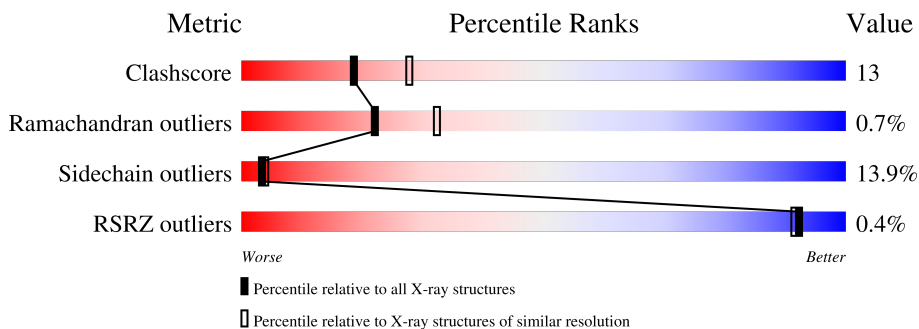|		
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.20.1
EDS	:	2.37.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.37.1

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 2.40 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
Clashscore		13
Ramachandran outliers		0.7%
Sidechain outliers		13.9%
RSRZ outliers		0.4%

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))
Clashscore	141614	4398 (2.40-2.40)
Ramachandran outliers	138981	4318 (2.40-2.40)
Sidechain outliers	138945	4319 (2.40-2.40)
RSRZ outliers	127900	3811 (2.40-2.40)

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	531	48% 20% 5% 27%
1	B	531	% 51% 21% 5% 23%
1	C	531	% 53% 19% • 23%
1	D	531	52% 18% • 27%
1	E	531	50% 18% 5% 27%
1	F	531	52% 20% 5% 23%

WORLDWIDE **wwPDB** PROTEIN DATA BANK

2 Entry composition ⓘ

There are 4 unique types of molecules in this entry. The entry contains 19957 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 Prunin.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	388	Total 3104	C 1926	N 583	O 591	S 4	0	2	0
1	B	410	Total 3296	C 2034	N 626	O 632	S 4	0	1	0
1	C	410	Total 3298	C 2035	N 628	O 631	S 4	0	1	0
1	D	388	Total 3102	C 1924	N 583	O 591	S 4	0	1	0
1	E	388	Total 3097	C 1922	N 580	O 591	S 4	0	1	0
1	F	410	Total 3302	C 2038	N 627	O 633	S 4	0	2	0

- Molecule 2 is CALCIUM ION (three-letter code: CA) (formula: Ca).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	1	Total 1	Ca 1	0	0
2	B	1	Total 1	Ca 1	0	0
2	C	1	Total 1	Ca 1	0	0
2	D	1	Total 1	Ca 1	0	0
2	E	1	Total 1	Ca 1	0	0
2	F	1	Total 1	Ca 1	0	0

- Molecule 3 is SODIUM ION (three-letter code: NA) (formula: Na).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	2	Total 2	Na 2	0	0
3	B	1	Total 1	Na 1	0	0
3	C	1	Total 1	Na 1	0	0
3	D	1	Total 1	Na 1	0	0
3	E	2	Total 2	Na 2	0	0
3	F	2	Total 2	Na 2	0	0

- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	90	Total 90	O 90	0	0
4	B	162	Total 162	O 162	0	0
4	C	123	Total 123	O 123	0	0
4	D	136	Total 136	O 136	0	0
4	E	102	Total 102	O 102	0	0
4	F	130	Total 130	O 130	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 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: Prunin

- Molecule 1: Prunin

• Molecule 1: Prunin

Chain C:

• Molecule 1: Prunin

Chain D:

E370 E371 R377 L378 K379 E380 N381 I382 G383 N384 P385 E386 R387 A388 S392 P393 R394 S399 I408 R413 L414 R418 G419 F420 R423 N424 G425 I426 Y427 H430 H436 R443 V448 N454 I458 L468 N474 Q479 E495 N500 T501 L502

T506 S507 F508 L509 R510 P513 V516 R525 R529 N534 R535 Q536 L541 S542 S543 SER GLN GLN ARG ARG ALA VAL VAL

- Molecule 1: Prunin

Chain E:

50%   18% 5%   27%

ALA ARG GLN SER GLN LEU S27 P28 Q29 N30 L37 Q38 A39 R40 E41 P42 R45 I46 Q47 Q52 N57 Q60 G61 D62 F63 V68 T74 R77 L80 N87 Y93 I94 V95 V100 V104 F105 S106 E110 T111 F112 E113 E114 S115 Q116 Q117 SER

SER GLN GLN GLY ARG GLN GLN GLU GLN GLU GLN GLU GLU ARG GLN GLN GLN GLU GLY GLU GLN GLY ARG GLN GLN GLN GLU GLN GLN GLU ARG GLN GLY ARG GLN GLN GLU GLU GLY ARG GLN GLU GLN GLU GLN GLN GLY GLN GLN GLY ARG PRO

GLN GLN GLN GLN GLN PHE R185 Q186 T193 R194 R195 E198 V202 P205 S212 Y213 G216 D217 V221 A222 V223 L235 K241 N247 N250 N253 Q254 GLN GLY GLN SER GLN PRO ARG GLN GLN GLY GLU GLN GLY ARG PRO GLY HIS GLN GLN PRO

PHE GLY ARG PRO ARG GLN GLN GLU GLN GLN GLY N287 G288 N289 N290 V291 F292 S293 L299 Q302 A303 L304 N305 V306 N307 E308 R312 Q317 R321 R322 Q323 I324 I325 Q326 V327 R328 G329 N330 L331 V334 P337 ARG GLY ARG GLN GLU ARG GLU HIS GLU GLU ARG GLN

GLN GLU GLN LEU GLN GLN GLU ARG GLN GLN GLN GLY GLU GLU LEU MET ALA ASN G368 L369 E370 E371 T372 F373 L378 K379 I382 G383 N384 P385 E386 T400 H404 I408 R413 L414 E417 R418 G419 F420 I426 Y427 S428 P429 V438 R443 R447 V448

N454 E462 V463 Q464 Q465 L468 F469 V477 I478 Q479 Q484 E487 A490 N500 T501 L502 A503 G504 R505 L517 Y521 R525 E526 R529 K532 E537 L541 S542 S543 SER GLN ARG ARG ALA VAL VAL

- Molecule 1: Prunin

Chain F:

52%   20% 5%   23%

ALA ARG GLN SER GLN LEU S27 P28 Q29 N30 Q31 C32 Q33 L34 L37 I46 Q47 A48 E49 I53 E54 N59 Q60 Q64 R72 I73 T74 I75 Q76 R77 L80 N87 V95 G102 A103 V104 F105 S106 F112 Q117 SER SER GLN GLN GLY ARG GLN GLN

GLU GLN GLU GLU GLU ARG GLN GLN GLN GLN GLN GLY GLU GLN GLY ARG GLN GLN GLY GLN GLN GLU GLN GLN GLU ARG GLN GLN GLN GLU GLU GLY ARG GLN GLY ARG GLN GLN GLU GLN GLN GLY GLN GLN GLY ARG PRO GLN GLN GLN GLN PHE R185

Q186 L187 D188 K192 T193 R194 R195 I196 R197 P205 Y210 W211 S212 Y213 N214 D215 G216 D217 A222 V223 N224 L225 L235 L244 N250 E251 F252 N253 Q254 GLN GLY GLN SER GLN PRO ARG GLN GLN GLY GLU GLN GLY ARG PRO GLY GLN HIS GLN GLN PRO PHE GLY ARG

PRO ARG GLN GLN GLU GLN GLN GLY N287 G288 N289 N296 T297 Q298 L299 L300 A301 N305 V306 N307 E308 R321 R328 G339 R340 Q341 E344 E347 R348 Q349 Q352 L353 Q354 Q355 E356 Q359 GLN GLY GLU GLN LEU MET ALA ASN G368 L369 E370 E371 C374

L378 K379 I382 P385 E386 R387 I390 F391 S392 P393 R394 A395 T400 I408 L412 R413 L414 R415 R418 G419 F420 R423 N424 G425 I426 Y427 S428 P429 V439 R447 V448 Q449 N454 L459 L468 Q473 N474 V477 I478 Q479 Q480 A481 G482 N483

Y488 K492 F498 L502 A503 L512 P513 D514 E515 V516 L517 A518 I523 S524 R525 R529 Q530 L531 Q536 E537 T538 I539 S543 SER GLN GLN ARG ARG ALA VAL VAL

4 Data and refinement statistics ⓘ

Property	Value	Source
Space group	P 41	Depositor
Cell constants a, b, c, α, β, γ	151.02Å 151.02Å 165.37Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	36.08 – 2.40 36.09 – 2.40	Depositor EDS
% Data completeness (in resolution range)	88.7 (36.08-2.40) 88.0 (36.09-2.40)	Depositor EDS
R_{merge}	0.13	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.67 (at 2.39Å)	Xtriage
Refinement program	REFMAC 5.2.0019	Depositor
R, R_{free}	0.172 , 0.229 0.175 , (Not available)	Depositor DCC
R_{free} test set	No test flags present.	wwPDB-VP
Wilson B-factor (Å^2)	37.3	Xtriage
Anisotropy	0.575	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.38 , 38.1	EDS
L-test for twinning[2]	$< |L| > = 0.49, < L^2 > = 0.33$	Xtriage
Estimated twinning fraction	0.478 for h,-k,-l	Xtriage
F_o,F_c correlation	0.96	EDS
Total number of atoms	19957	wwPDB-VP
Average B, all atoms (Å^2)	43.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.84% 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: CA, NA

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	0.69	0/3168	0.90	1/4291 (0.0%)
1	B	0.78	0/3358	1.01	6/4542 (0.1%)
1	C	0.74	0/3360	1.00	9/4544 (0.2%)
1	D	0.75	0/3163	0.98	3/4284 (0.1%)
1	E	0.72	0/3158	0.94	1/4278 (0.0%)
1	F	0.80	0/3367	1.02	3/4554 (0.1%)
All	All	0.75	0/19574	0.97	23/26493 (0.1%)

There are no bond length outliers.

The worst 5 of 23 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	B	321	ARG	NE-CZ-NH1	-7.18	116.71	120.30
1	D	195	ARG	NE-CZ-NH1	-6.43	117.09	120.30
1	C	189	ARG	NE-CZ-NH1	5.96	123.28	120.30
1	A	194	ARG	NE-CZ-NH1	5.93	123.27	120.30
1	C	236	ASP	CB-CG-OD2	5.87	123.58	118.30

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.

WORLDWIDE
wwPDB
PROTEIN DATA BANK

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	3104	0	2974	95	0
1	B	3296	0	3143	97	0
1	C	3298	0	3148	88	0
1	D	3102	0	2972	66	0
1	E	3097	0	2963	97	0
1	F	3302	0	3151	111	0
2	A	1	0	0	0	0
2	B	1	0	0	0	0
2	C	1	0	0	0	0
2	D	1	0	0	0	0
2	E	1	0	0	0	0
2	F	1	0	0	0	0
3	A	2	0	0	0	0
3	B	1	0	0	0	0
3	C	1	0	0	0	0
3	D	1	0	0	0	0
3	E	2	0	0	0	0
3	F	2	0	0	0	0
4	A	90	0	0	7	0
4	B	162	0	0	11	0
4	C	123	0	0	5	0
4	D	136	0	0	4	0
4	E	102	0	0	5	0
4	F	130	0	0	4	0
All	All	19957	0	18351	494	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 13.

The worst 5 of 494 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:B:250:ASN:HB3	1:B:253:ASN:ND2	1.43	1.30
1:B:250:ASN:CB	1:B:253:ASN:HD21	1.49	1.23
1:A:195:ARG:HH21	1:A:195:ARG:HG3	1.05	1.11
1:F:195:ARG:HH21	1:F:195:ARG:CG	1.67	1.08
1:F:195:ARG:NH2	1:F:195:ARG:HG3	1.53	1.05

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	382/531 (72%)	342 (90%)	36 (9%)	4 (1%)	15	23
1	B	403/531 (76%)	371 (92%)	32 (8%)	0	100	100
1	C	403/531 (76%)	375 (93%)	24 (6%)	4 (1%)	15	23
1	D	381/531 (72%)	356 (93%)	23 (6%)	2 (0%)	29	41
1	E	381/531 (72%)	354 (93%)	22 (6%)	5 (1%)	12	17
1	F	404/531 (76%)	372 (92%)	30 (7%)	2 (0%)	29	41
All	All	2354/3186 (74%)	2170 (92%)	167 (7%)	17 (1%)	22	32

5 of 17 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	328[A]	ARG
1	A	328[B]	ARG
1	C	340	ARG
1	E	328	ARG
1	E	330	ASN

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	330/452 (73%)	272 (82%)	58 (18%)	2	2
1	B	349/452 (77%)	293 (84%)	56 (16%)	2	3
1	C	349/452 (77%)	312 (89%)	37 (11%)	6	9

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles		
1	D	330/452 (73%)	289 (88%)	41 (12%)	4	5	
1	E	329/452 (73%)	281 (85%)	48 (15%)	3	3	
1	F	350/452 (77%)	304 (87%)	46 (13%)	4	4	
All	All	2037/2712 (75%)	1751 (86%)	286 (14%)	3	4	

5 of 286 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	E	420	PHE
1	E	525	ARG
1	F	299	LEU
1	B	369	LEU
1	B	351	GLU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 89 such sidechains are listed below:

Mol	Chain	Res	Type
1	D	480	GLN
1	E	522	GLN
1	D	530	GLN
1	E	247	ASN
1	F	76	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 ⓘ

Of 15 ligands modelled in this entry, 15 are monoatomic - leaving 0 for Mogul analysis.

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 ⓘ

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	388/531 (73%)	-0.06	1 (0%)	94	93	27, 45, 66, 84	0
1	B	410/531 (77%)	-0.05	5 (1%)	79	77	21, 39, 81, 96	0
1	C	410/531 (77%)	-0.09	3 (0%)	87	86	22, 38, 80, 103	0
1	D	388/531 (73%)	-0.11	0	100	100	21, 37, 61, 87	0
1	E	388/531 (73%)	-0.06	1 (0%)	94	93	26, 45, 67, 80	0
1	F	410/531 (77%)	-0.11	0	100	100	21, 38, 75, 86	0
All	All	2394/3186 (75%)	-0.08	10 (0%)	92	91	21, 41, 68, 103	0

The worst 5 of 10 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	359	GLN	5.0
1	B	27	SER	2.9
1	B	357	ARG	2.8
1	C	353	LEU	2.7
1	B	353	LEU	2.5

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.

WORLDWIDE
wwPDB
PROTEIN DATA BANK

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, 95th 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
2	CA	A	1	1/1	0.92	0.21	52,52,52,52	0
3	NA	D	552	1/1	0.94	0.35	58,58,58,58	0
3	NA	E	552	1/1	0.94	0.25	67,67,67,67	0
3	NA	A	552	1/1	0.96	0.26	53,53,53,53	0
3	NA	C	552	1/1	0.96	0.31	51,51,51,51	0
2	CA	B	2	1/1	0.97	0.15	53,53,53,53	0
3	NA	F	552	1/1	0.97	0.23	43,43,43,43	0
3	NA	F	7	1/1	0.97	0.23	53,53,53,53	0
3	NA	A	9	1/1	0.98	0.18	50,50,50,50	0
2	CA	F	6	1/1	0.98	0.14	51,51,51,51	0
2	CA	D	4	1/1	0.99	0.17	45,45,45,45	0
2	CA	E	5	1/1	0.99	0.18	51,51,51,51	0
3	NA	E	8	1/1	0.99	0.20	44,44,44,44	0
3	NA	B	552	1/1	0.99	0.29	40,40,40,40	0
2	CA	C	3	1/1	0.99	0.18	45,45,45,45	0

6.5 Other polymers ⓘ

There are no such residues in this entry.