wwPDB X-ray Structure Validation Summary Report (i)

Feb 1, 2016 – 09:45 PM GMT

PDB ID	:	4W5B
Title	:	Crystal structure analysis of cruzain with Fragment 1 (N-(1H-benzimidazol-2 -yl)-1,3-dimethyl-pyrazole-4-carboxamide)
Authors	:	Tochowicz, A.; McKerrow, J.H.; Craik, C.S.
Deposited on	:	2014-08-17
Resolution	:	2.70 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
http://wwpdb.org/validation/2016/XrayValidationReportHelp
with specific help available everywhere you see the (i) symbol.

The following versions of software and data (see references (1)) were used in the production of this report:

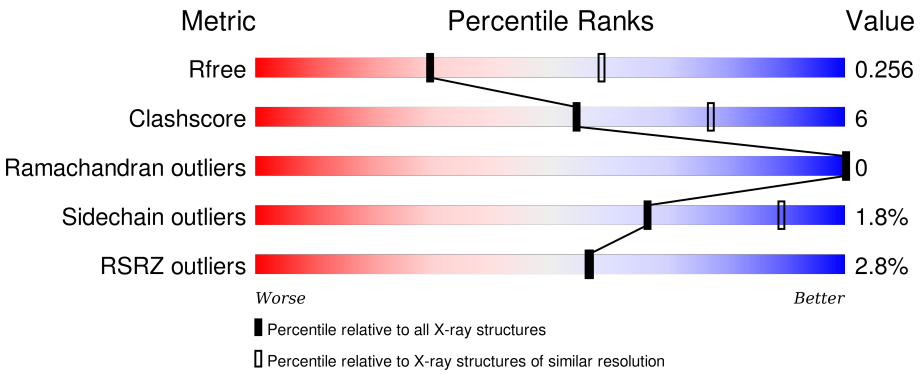
MolProbity	:	4.02b-467
Mogul	:	1.7 (RC4), CSD as536be (2015)
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20026688
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	trunk26865

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 2.70 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}	91344	2103 (2.70-2.70)
Clashscore	102246	2422 (2.70-2.70)
Ramachandran outliers	100387	2382 (2.70-2.70)
Sidechain outliers	100360	2382 (2.70-2.70)
RSRZ outliers	91569	2107 (2.70-2.70)

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 on the lower bar indicate the fraction of residues that contain outliers for >=3, 2, 1 and 0 types of geometric quality criteria. 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	215	2% 87% 13%
1	B	215	% 85% 15%
1	C	215	5% 86% 13% •

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
2	3H5	A	301	-	-	-	X
2	3H5	B	301	-	-	-	X
2	3H5	C	301	-	-	-	X

2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 4927 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.

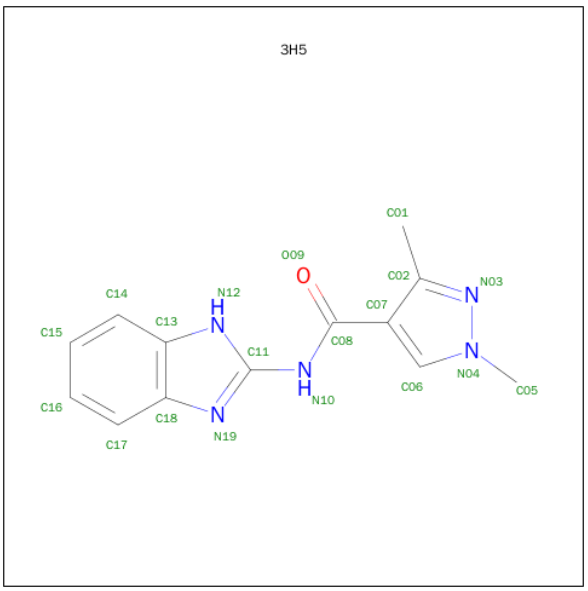
- Molecule 1 is a protein called Cruzipain.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	215	1598	999	266	322	11	0	1	0
1	B	215	1592	995	266	320	11	0	0	0
1	C	215	1592	995	266	320	11	0	0	0

There are 3 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	25	SER	CYS	engineered mutation	UNP P25779
B	25	SER	CYS	engineered mutation	UNP P25779
C	25	SER	CYS	engineered mutation	UNP P25779

- Molecule 2 is N-(1H-benzimidazol-2-yl)-1,3-dimethyl-1H-pyrazole-4-carboxamide (three-letter code: 3H5) (formula: $C_{13}H_{13}N_5O$).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
			Total	C	N	O		
2	A	1	19	13	5	1	0	0
2	B	1	19	13	5	1	0	0
2	C	1	19	13	5	1	0	0

- Molecule 3 is water.

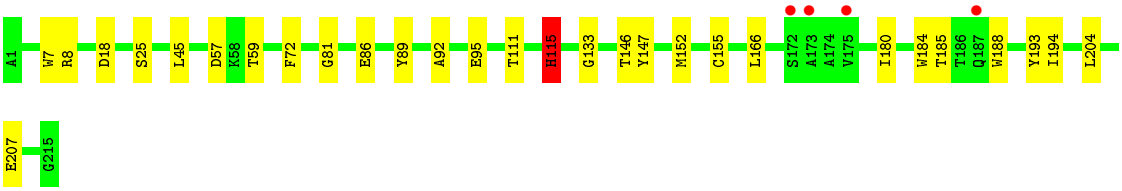
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
3	A	39	39	39	0	0
3	B	29	29	29	0	0
3	C	20	20	20	0	0

3 Residue-property plots (i)

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors 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.

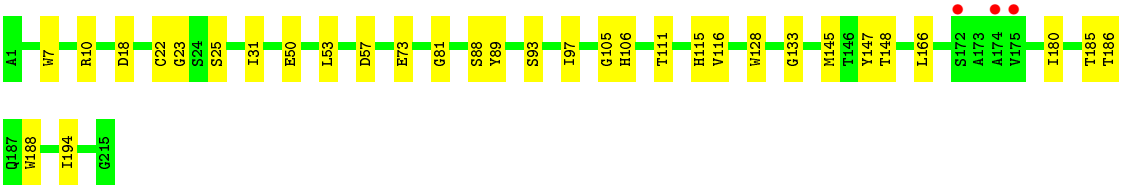
- Molecule 1: Cruzipain

Chain A:



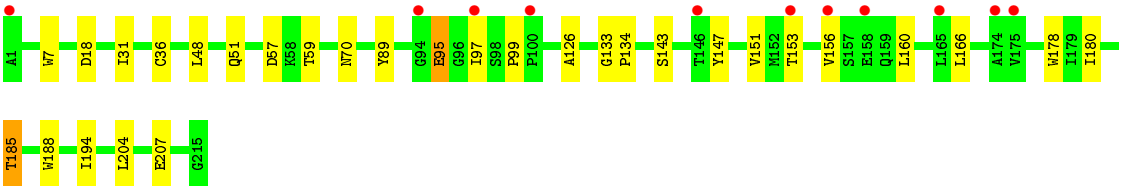
- Molecule 1: Cruzipain

Chain B:



- Molecule 1: Cruzipain

Chain C:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 43 21 2	Depositor
Cell constants a, b, c, α, β, γ	137.10Å 137.10Å 109.80Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	29.53 – 2.70 29.53 – 2.70	Depositor EDS
% Data completeness (in resolution range)	99.9 (29.53-2.70) 99.9 (29.53-2.70)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.84 (at 2.72Å)	Xtriage
Refinement program	PHENIX (phenix.refine: dev_1738)	Depositor
R, R_{free}	0.212 , 0.254 0.223 , 0.256	Depositor DCC
R_{free} test set	1170 reflections (4.17%)	DCC
Wilson B-factor (Å^2)	29.3	Xtriage
Anisotropy	0.981	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.34 , 40.4	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning[2]	$< \|L\| > = 0.48$, $< L^2 > = 0.31$	Xtriage
Outliers	0 of 29249 reflections	Xtriage
F_o,F_c correlation	0.91	EDS
Total number of atoms	4927	wwPDB-VP
Average B, all atoms (Å^2)	40.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 11.38% 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, 0.375 respectively for untwinned datasets, and 0.333, 0.2 for perfectly twinned datasets.

5 Model quality (i)

5.1 Standard geometry (i)

Bond lengths and bond angles in the following residue types are not validated in this section: 3H5

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	0.50	0/1640	0.61	0/2244
1	B	0.48	0/1631	0.62	0/2232
1	C	0.41	0/1631	0.55	0/2232
All	All	0.46	0/4902	0.60	0/6708

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	A	0	1

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	115	HIS	Sidechain

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	1598	0	1496	24	0
1	B	1592	0	1490	16	2
1	C	1592	0	1490	18	0
2	A	19	0	13	2	0
2	B	19	0	13	2	0
2	C	19	0	13	3	0
3	A	39	0	0	2	4
3	B	29	0	0	0	2
3	C	20	0	0	1	0
All	All	4927	0	4515	60	4

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 6.

The worst 5 of 60 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)
2:C:301:3H5:H4	2:C:301:3H5:H11	1.45	0.81
1:B:73:GLU:HG3	1:B:115:HIS:HE1	1.56	0.70
1:A:95:GLU:OE2	3:A:401:HOH:O	2.10	0.69
1:A:72:PHE:HB2	1:A:115:HIS:HD2	1.58	0.68
1:A:72:PHE:HB2	1:A:115:HIS:CD2	2.28	0.67

All (4) 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:B:105:GLY:O	3:A:401:HOH:O[3_444]	2.00	0.20
1:B:97:ILE:O	3:A:402:HOH:O[3_444]	2.01	0.19
3:A:411:HOH:O	3:B:408:HOH:O[2_545]	2.07	0.13
3:A:402:HOH:O	3:B:401:HOH:O[4_545]	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	214/215 (100%)	206 (96%)	8 (4%)	0	100	100
1	B	213/215 (99%)	205 (96%)	8 (4%)	0	100	100
1	C	213/215 (99%)	205 (96%)	8 (4%)	0	100	100
All	All	640/645 (99%)	616 (96%)	24 (4%)	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	168/167 (101%)	165 (98%)	3 (2%)	66	89
1	B	167/167 (100%)	164 (98%)	3 (2%)	66	89
1	C	167/167 (100%)	164 (98%)	3 (2%)	66	89
All	All	502/501 (100%)	493 (98%)	9 (2%)	66	89

5 of 9 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	145	MET
1	C	185	THR
1	C	89	TYR
1	A	115	HIS
1	B	186	THR

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

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 carbohydrates in this entry.

5.6 Ligand geometry (i)

3 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	3H5	A	301	-	19,21,21	2.16	3 (15%)	14,30,30	1.71	2 (14%)
2	3H5	B	301	-	19,21,21	2.03	3 (15%)	14,30,30	1.72	1 (7%)
2	3H5	C	301	-	19,21,21	1.97	3 (15%)	14,30,30	1.85	5 (35%)

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the chemical component dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
2	3H5	A	301	-	-	0/4/8/8	0/3/3/3
2	3H5	B	301	-	-	0/4/8/8	0/3/3/3
2	3H5	C	301	-	-	0/4/8/8	0/3/3/3

The worst 5 of 9 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	B	301	3H5	C15-C14	2.24	1.41	1.36
2	A	301	3H5	C15-C14	2.36	1.42	1.36
2	C	301	3H5	N03-N04	3.09	1.40	1.37
2	C	301	3H5	C08-N10	3.85	1.45	1.35

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	B	301	3H5	C08-N10	3.88	1.45	1.35

The worst 5 of 8 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	C	301	3H5	C11-N10-C08	-2.85	122.83	128.18
2	A	301	3H5	C11-N19-C18	2.11	109.61	106.64
2	C	301	3H5	C01-C02-N03	2.11	124.51	119.87
2	C	301	3H5	C07-C08-N10	2.14	120.48	116.13
2	C	301	3H5	C11-N12-C13	2.43	110.07	106.64

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

3 monomers are involved in 7 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	A	301	3H5	2	0
2	B	301	3H5	2	0
2	C	301	3H5	3	0

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	215/215 (100%)	-0.32	4 (1%)	70	70	20, 32, 52, 66	0
1	B	215/215 (100%)	-0.35	3 (1%)	78	77	18, 34, 53, 64	0
1	C	215/215 (100%)	0.06	11 (5%)	32	30	36, 49, 65, 73	0
All	All	645/645 (100%)	-0.20	18 (2%)	56	57	18, 40, 59, 73	0

The worst 5 of 18 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	174	ALA	3.8
1	C	175	VAL	3.6
1	A	175	VAL	2.8
1	B	175	VAL	2.7
1	C	1	ALA	2.5

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 carbohydrates in this entry.

6.4 Ligands (i)

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. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. 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	LLDF	B-factors(Å^2)	Q<0.9
2	3H5	C	301	19/19	0.79	0.38	10.76	50,67,74,75	19
2	3H5	B	301	19/19	0.70	0.42	10.25	37,56,69,69	19
2	3H5	A	301	19/19	0.70	0.38	7.55	58,71,76,78	19

6.5 Other polymers (i)

There are no such residues in this entry.