WORLDWIDE PDB
PROTEIN DATA BANK

wwPDB X-ray Structure Validation Summary Report (i)

May 22, 2020 – 02:30 am BST

PDB ID	:	5N7Q
Title	:	CRYSTAL STRUCTURE OF MATURE CATHEPSIN D FROM THE TICK IXODES RICINUS (IRCD1) IN COMPLEX WITH THE INHIBITOR PEP-STATIN A
Authors	:	Brynda, J.; Hanova, I.; Hobizalova, R.; Mares, M.
Deposited on	:	2017-02-21
Resolution	:	1.45 Å(reported)

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

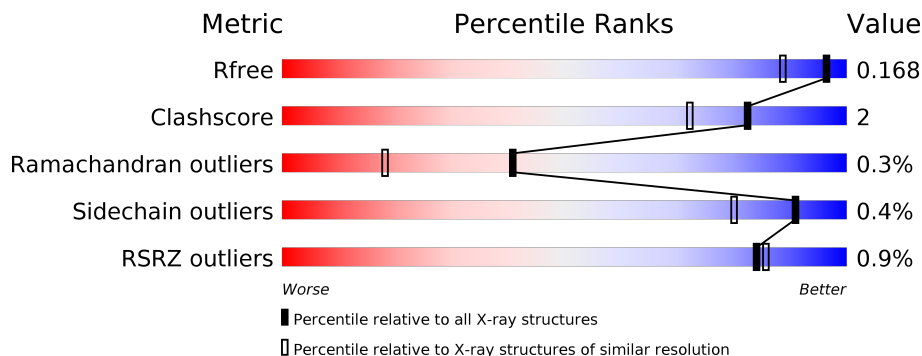|		
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.11

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 1.45 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		0.168
Clashscore		2
Ramachandran outliers		0.3%
Sidechain outliers		0.4%
RSRZ outliers		0.9%

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	1156 (1.46-1.46)
Clashscore	141614	1202 (1.46-1.46)
Ramachandran outliers	138981	1178 (1.46-1.46)
Sidechain outliers	138945	1178 (1.46-1.46)
RSRZ outliers	127900	1139 (1.46-1.46)

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 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	339	96%
1	B	339	96%
2	I	6	50% 50%
2	J	6	83% 17%

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 crite-

ria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
3	PEG	A	401	-	-	X	-

2 Entry composition (i)

There are 5 unique types of molecules in this entry. The entry contains 6647 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 Putative cathepsin d.

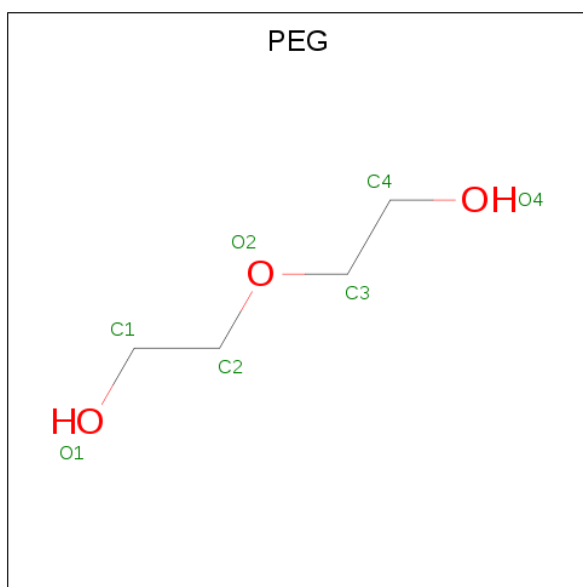| Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	337	Total 2614	C 1675	N 417	O 510	S 12	0	13	0
1	B	337	Total 2600	C 1663	N 414	O 511	S 12	0	7	0

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	17	VAL	GLU	conflict	UNP V5HCK7
A	160	LEU	ARG	conflict	UNP V5HCK7
B	17	VAL	GLU	conflict	UNP V5HCK7
B	160	LEU	ARG	conflict	UNP V5HCK7

- Molecule 2 is a protein called PEPSTATIN A.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
2	I	6	Total 61	C 44	N 6	O 11	0	2	0
2	J	6	Total 48	C 34	N 5	O 9	0	0	0

- Molecule 3 is DI(HYDROXYETHYL)ETHER (three-letter code: PEG) (formula: $C_4H_{10}O_3$).

PEG

C4

OH O4

O2

O

C3

C1

HO

O1

C2

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
3	A	1	Total	C	O	0	0
			7	4	3		

- Molecule 4 is GLYCEROL (three-letter code: GOL) (formula: $C_3H_8O_3$).

GOL

C3

OH O3

C1

C2

HO

OH

O1

O2
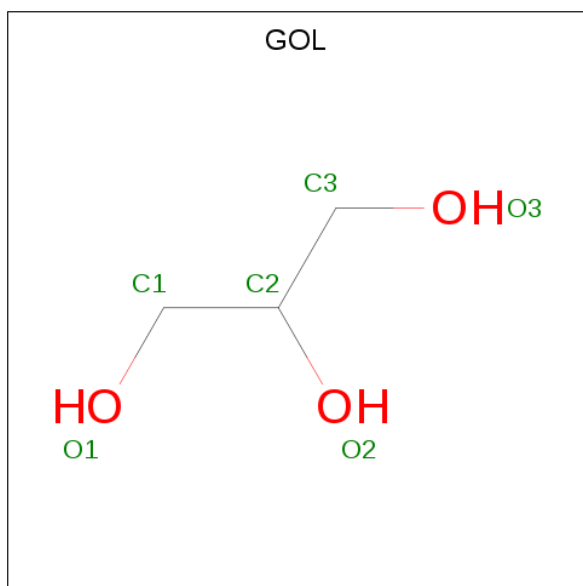
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
4	B	1	Total	C	O	0	0
			6	3	3		

- Molecule 5 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	A	636	Total 638 O 638	0	2
5	B	650	Total 651 O 651	0	1
5	I	12	Total 12 O 12	0	0
5	J	10	Total 10 O 10	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 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: Putative cathepsin d

Chain A:

• Molecule 1: Putative cathepsin d

Chain B:

• Molecule 2: PEPSTATIN A

Chain I:

• Molecule 2: PEPSTATIN A

Chain J:

4   Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	61.14Å 82.82Å 61.39Å 90.00° 91.13° 90.00°	Depositor
Resolution (Å)	61.38 – 1.45 27.21 – 1.46	Depositor EDS
% Data completeness (in resolution range)	99.9 (61.38-1.45) 99.9 (27.21-1.46)	Depositor EDS
R_{merge}	0.08	Depositor
R_{sym}	0.08	Depositor
$< I/\sigma(I) >$ [1]	2.75 (at 1.46Å)	Xtriage
Refinement program	REFMAC 5.8.0103	Depositor
R, R_{free}	0.145 , 0.166 0.146 , 0.168	Depositor DCC
R_{free} test set	5335 reflections (4.99%)	wwPDB-VP
Wilson B-factor (Å^2)	10.5	Xtriage
Anisotropy	0.121	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.30 , 40.3	EDS
L-test for twinning[2]	$< |L| > = 0.49, < L^2 > = 0.33$	Xtriage
Estimated twinning fraction	0.022 for l,k,-h 0.024 for h,-k,-l 0.018 for l,-k,h	Xtriage
F_o,F_c correlation	0.97	EDS
Total number of atoms	6647	wwPDB-VP
Average B, all atoms (Å^2)	15.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.66% 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)

Bond lengths and bond angles in the following residue types are not validated in this section: GOL, PEG, STA, IVA

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	0.54	0/2707	0.74	0/3680
1	B	0.53	0/2684	0.73	1/3648 (0.0%)
2	I	0.71	0/24	0.71	0/31
2	J	0.67	0/17	0.81	0/21
All	All	0.54	0/5432	0.74	1/7380 (0.0%)

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
2	I	0	2
2	J	0	2
All	All	0	4

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	B	282	ARG	NE-CZ-NH1	5.61	123.11	120.30

There are no chirality outliers.

All (4) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
2	I	504	STA	Mainchain,Peptide
2	J	504	STA	Mainchain,Peptide

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	2614	0	2549	10	0
1	B	2600	0	2531	8	0
2	I	61	0	77	5	0
2	J	48	0	60	0	0
3	A	7	0	10	5	0
4	B	6	0	8	2	0
5	A	638	0	0	3	1
5	B	651	0	0	5	1
5	I	12	0	0	0	0
5	J	10	0	0	0	0
All	All	6647	0	5235	26	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 2.

The worst 5 of 26 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:I:501[B]:IVA:HG22	2:I:502[B]:VAL:N	1.92	0.82
1:A:83[A]:ASN:OD1	5:A:501:HOH:O	2.00	0.78
2:I:501[B]:IVA:CG2	2:I:502[B]:VAL:N	2.56	0.67
2:I:501[B]:IVA:HG22	2:I:502[B]:VAL:H	1.60	0.65
1:A:83[A]:ASN:CG	3:A:401:PEG:H41	2.19	0.63

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)
5:A:792:HOH:O	5:B:851:HOH:O[1_655]	1.81	0.39

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	347/339 (102%)	342 (99%)	4 (1%)	1 (0%)	41	18
1	B	342/339 (101%)	337 (98%)	4 (1%)	1 (0%)	41	18
2	I	4/6 (67%)	4 (100%)	0	0	100	100
2	J	3/6 (50%)	3 (100%)	0	0	100	100
All	All	696/690 (101%)	686 (99%)	8 (1%)	2 (0%)	41	18

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	195	GLY
1	B	195	GLY

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	290/286 (101%)	289 (100%)	1 (0%)	92	82
1	B	290/286 (101%)	289 (100%)	1 (0%)	92	82
2	I	3/2 (150%)	3 (100%)	0	100	100
2	J	2/2 (100%)	2 (100%)	0	100	100
All	All	585/576 (102%)	583 (100%)	2 (0%)	91	82

All (2) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	150	LEU
1	B	105	LEU

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

Mol	Chain	Res	Type
1	A	292	GLN

5.3.3 RNA (i)

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains (i)

4 non-standard protein/DNA/RNA residues 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	STA	I	504	2	10,10,11	0.69	0	9,12,14	0.96	1 (11%)
2	STA	J	504	2	10,10,11	0.78	0	9,12,14	1.16	1 (11%)
2	STA	J	506	2	8,11,11	0.37	0	7,14,14	1.07	0
2	STA	I	506	2	8,11,11	0.79	0	7,14,14	0.47	0

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	STA	I	504	2	-	1/11/11/12	-
2	STA	J	504	2	-	2/11/11/12	-
2	STA	J	506	2	-	3/10/12/12	-
2	STA	I	506	2	-	0/10/12/12	-

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	J	504	STA	O-C-CM	-3.25	115.96	125.43
2	I	504	STA	O-C-CM	-2.36	118.56	125.43

There are no chirality outliers.

5 of 6 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	I	504	STA	O-C-CM-CH
2	J	506	STA	N-CA-CB-CG
2	J	506	STA	CH-CA-CB-CG
2	J	506	STA	CA-CB-CG-CD1
2	J	504	STA	O-C-CM-CH

There are no ring outliers.

No monomer is involved in short contacts.

5.5 Carbohydrates (i)

There are no carbohydrates in this entry.

5.6 Ligand geometry (i)

2 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	#$|Z| > 2$
4	GOL	B	401	-	5,5,5	0.45	0	5,5,5	1.03	0
3	PEG	A	401	-	6,6,6	0.38	0	5,5,5	0.59	0

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
4	GOL	B	401	-	-	4/4/4/4	-
3	PEG	A	401	-	-	3/4/4/4	-

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

5 of 7 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
4	B	401	GOL	C1-C2-C3-O3
4	B	401	GOL	O1-C1-C2-C3
4	B	401	GOL	O2-C2-C3-O3
3	A	401	PEG	O2-C3-C4-O4
3	A	401	PEG	C1-C2-O2-C3

There are no ring outliers.

2 monomers are involved in 7 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	B	401	GOL	2	0
3	A	401	PEG	5	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(Å²)	Q<0.9
1	A	337/339 (99%)	-0.27	5 (1%)	73	74	7, 11, 22, 35	0
1	B	337/339 (99%)	-0.31	1 (0%)	94	95	7, 11, 21, 31	0
2	I	3/6 (50%)	-0.32	0	100	100	9, 9, 10, 13	0
2	J	3/6 (50%)	-0.58	0	100	100	11, 11, 11, 11	0
All	All	680/690 (98%)	-0.29	6 (0%)	84	86	7, 11, 21, 35	0

The worst 5 of 6 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	252	THR	4.1
1	A	254	GLY	2.5
1	B	3	GLU	2.4
1	A	253	GLU	2.3
1	A	255	GLY	2.3

6.2 Non-standard residues in protein, DNA, RNA chains (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. 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(Å²)	Q<0.9
2	STA	J	506	12/12	0.74	0.18	18,26,33,34	0
2	STA	I	506	12/12	0.81	0.15	14,18,25,26	0
2	STA	J	504	11/12	0.97	0.08	8,9,10,10	0
2	STA	I	504	11/12	0.98	0.08	7,7,8,8	0

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. 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
3	PEG	A	401	7/7	0.70	0.27	23,24,26,26	0
4	GOL	B	401	6/6	0.85	0.20	13,13,14,14	6

6.5 Other polymers (i)

There are no such residues in this entry.