W O R L D W I D E

wwPDB

P R O T E I N   D A T A   B A N K

wwPDB X-ray Structure Validation Summary Report (i)

Jan 20, 2024 – 10:47 pm GMT

PDB ID	:	7A0F
Title	:	The Crystal Structure of Bovine Thrombin in complex with Hirudin (C22U/C39U) at 2.7 Angstroms Resolution
Authors	:	Hidmi, T.; Mousa, R.; Pomyalov, S.; Lansky, S.; Khouri, L.; Metanis, N.; Shoham, G.
Deposited on	:	2020-08-07
Resolution	:	2.70 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	:	1.8.4, CSD as541be (2020)
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 ⓘ

The following experimental techniques were used to determine the structure:
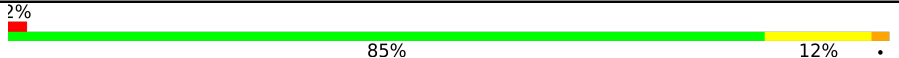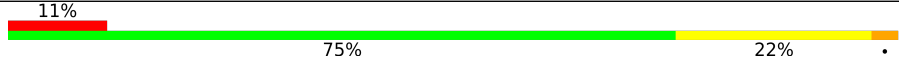*X-RAY DIFFRACTION*

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.
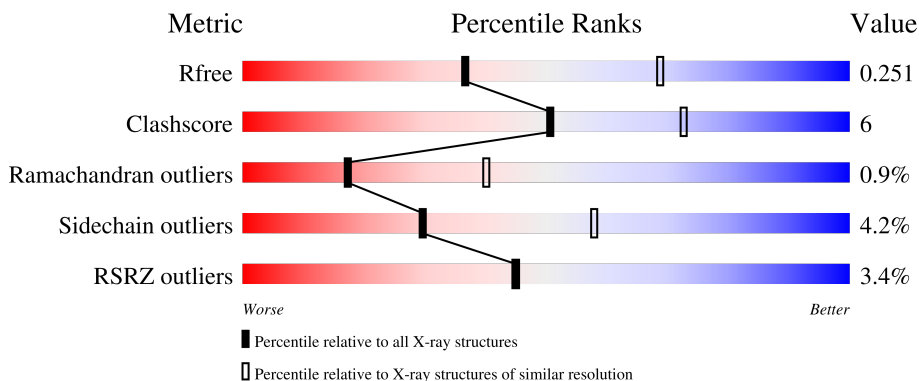
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	2808 (2.70-2.70)
Clashscore	141614	3122 (2.70-2.70)
Ramachandran outliers	138981	3069 (2.70-2.70)
Sidechain outliers	138945	3069 (2.70-2.70)
RSRZ outliers	127900	2737 (2.70-2.70)

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 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	HHH	259	2% / 85% / 12% / •
2	III	65	11% / 75% / 22% / •
3	LLL	49	2% / 55% / 16% / 29%

2 Entry composition ⓘ

There are 6 unique types of molecules in this entry. The entry contains 2936 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 Prothrombin.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	HHH	259	Total	C	N	O	S	0	0	0
			2083	1331	372	368	12			

- Molecule 2 is a protein called Hirudin variant-1.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
2	III	65	Total	C	N	O	S	Se	0	0	0
			482	287	80	109	4	2			

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
III	22	SEC	CYS	modified residue	UNP P01050
III	39	SEC	CYS	modified residue	UNP P01050

- Molecule 3 is a protein called Prothrombin.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	LLL	35	Total	C	N	O	S	0	0	0
			283	179	45	58	1			

- Molecule 4 is 2-acetamido-2-deoxy-beta-D-glucopyranose (three-letter code: NAG) (formula: $C_8H_{15}NO_6$).

NAG

O6
OH

O5
O1
HO
O

C6

C1(R) C5(R)

C2(R) C4(S)
N2 C3(R)
HN
OH
O4

C7
OH

C8
O
O3
O7

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
4	HHH	1	Total	C	N	O	0	0
			14	8	1	5		

- Molecule 5 is SODIUM ION (three-letter code: NA) (formula: Na).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	HHH	1	Total	Na	0	0
			1	1		

- Molecule 6 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	HHH	56	Total	O	0	0
			56	56		
6	III	13	Total	O	0	0
			13	13		
6	LLL	4	Total	O	0	0
			4	4		

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.
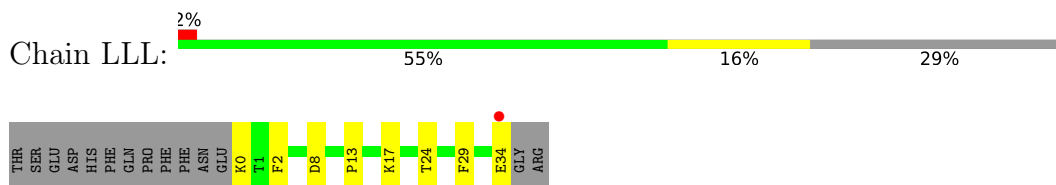
• Molecule 1: Prothrombin

Chain HHH:
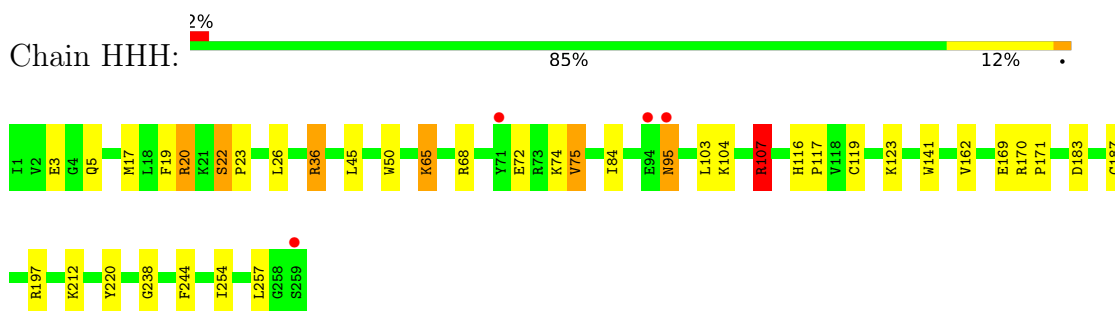
• Molecule 2: Hirudin variant-1

Chain III:
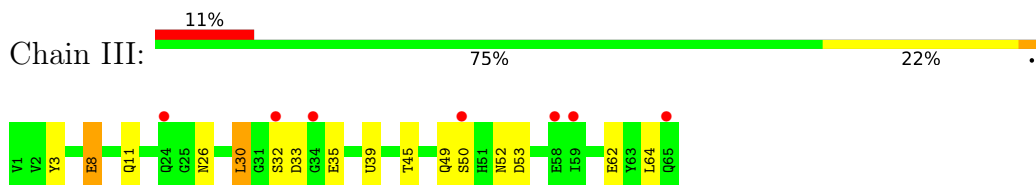
• Molecule 3: Prothrombin

Chain LLL:

4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 2 2 21	Depositor
Cell constants a, b, c, α, β, γ	58.07Å 101.50Å 142.07Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	47.79 – 2.70 47.79 – 2.70	Depositor EDS
% Data completeness (in resolution range)	99.9 (47.79-2.70) 100.0 (47.79-2.70)	Depositor EDS
R_{merge}	0.29	Depositor
R_{sym}	0.29	Depositor
$< I/\sigma(I) >$ [1]	1.69 (at 2.69Å)	Xtriage
Refinement program	REFMAC 5.8.0258	Depositor
R, R_{free}	0.194 , 0.253 0.195 , 0.251	Depositor DCC
R_{free} test set	586 reflections (4.92%)	wwPDB-VP
Wilson B-factor (Å^2)	44.3	Xtriage
Anisotropy	0.062	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.34 , 36.2	EDS
L-test for twinning[2]	$< |L| > = 0.48, < L^2 > = 0.32$	Xtriage
Estimated twinning fraction	0.027 for 1/2*h-1/2*k,-3/2*h-1/2*k,-l 0.045 for 1/2*h+1/2*k,3/2*h-1/2*k,-l	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	2936	wwPDB-VP
Average B, all atoms (Å^2)	45.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 7.02% 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: NAG, SEC, 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	HHH	0.71	2/2137 (0.1%)	0.94	7/2894 (0.2%)
2	III	0.47	0/474	1.01	2/637 (0.3%)
3	LLL	0.37	0/287	0.78	0/382
All	All	0.65	2/2898 (0.1%)	0.94	9/3913 (0.2%)

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	HHH	0	3
2	III	0	1
All	All	0	4

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	HHH	107	ARG	CZ-NH1	25.93	1.66	1.33
1	HHH	107	ARG	CZ-NH2	6.61	1.41	1.33

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

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	HHH	107	ARG	NH1-CZ-NH2	-12.53	105.62	119.40
1	HHH	20	ARG	NE-CZ-NH2	12.22	126.41	120.30
1	HHH	107	ARG	NE-CZ-NH1	11.73	126.17	120.30
1	HHH	20	ARG	NE-CZ-NH1	-9.13	115.73	120.30
1	HHH	20	ARG	CD-NE-CZ	6.38	132.53	123.60

There are no chirality outliers.

All (4) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	HHH	107	ARG	Sidechain
1	HHH	22	SER	Peptide
1	HHH	95	ASN	Peptide
2	III	52	ASN	Peptide

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.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	HHH	2083	0	2077	26	0
2	III	482	0	425	7	0
3	LLL	283	0	273	6	0
4	HHH	14	0	13	0	0
5	HHH	1	0	0	0	0
6	HHH	56	0	0	0	0
6	III	13	0	0	1	0
6	LLL	4	0	0	0	0
All	All	2936	0	2788	32	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 6.

The worst 5 of 32 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:HHH:17:MET:HG3	1:HHH:26:LEU:HD12	1.62	0.80
1:HHH:45:LEU:HD11	1:HHH:103:LEU:HD21	1.66	0.77
1:HHH:170:ARG:HB2	1:HHH:171:PRO:HD3	1.72	0.70
1:HHH:17:MET:HG3	1:HHH:26:LEU:CD1	2.22	0.69
1:HHH:45:LEU:HD13	1:HHH:84:ILE:HG23	1.74	0.68

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	HHH	257/259 (99%)	242 (94%)	13 (5%)	2 (1%)	19	43
2	III	61/65 (94%)	54 (88%)	6 (10%)	1 (2%)	9	24
3	LLL	33/49 (67%)	32 (97%)	1 (3%)	0	100	100
All	All	351/373 (94%)	328 (93%)	20 (6%)	3 (1%)	17	40

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	HHH	23	PRO
2	III	33	ASP
1	HHH	75	VAL

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	HHH	224/226 (99%)	218 (97%)	6 (3%)	44	74
2	III	54/54 (100%)	49 (91%)	5 (9%)	9	21
3	LLL	30/43 (70%)	28 (93%)	2 (7%)	16	37
All	All	308/323 (95%)	295 (96%)	13 (4%)	30	58

5 of 13 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	III	30	LEU
2	III	50	SER
3	LLL	34	GLU
2	III	64	LEU
3	LLL	8	ASP

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

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 2 ligands modelled in this entry, 1 is monoatomic - leaving 1 for Mogul analysis.

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	NAG	HHH	301	1	14,14,15	0.86	0	17,19,21	2.86	4 (23%)

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	NAG	HHH	301	1	-	1/6/23/26	0/1/1/1

There are no bond length outliers.

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
4	HHH	301	NAG	C1-O5-C5	8.66	123.93	112.19
4	HHH	301	NAG	C4-C3-C2	-5.46	103.01	111.02
4	HHH	301	NAG	C1-C2-N2	3.00	115.62	110.49
4	HHH	301	NAG	O5-C5-C4	2.65	117.27	110.83

There are no chirality outliers.

All (1) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
4	HHH	301	NAG	C1-C2-N2-C7

There are no ring outliers.

No monomer is involved in short contacts.

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 ⓘ

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	HHH	259/259 (100%)	0.09	4 (1%)	73	76	24, 38, 67, 113	0
2	III	63/65 (96%)	0.55	7 (11%)	5	4	32, 47, 91, 102	0
3	LLL	35/49 (71%)	0.04	1 (2%)	51	52	27, 44, 94, 102	0
All	All	357/373 (95%)	0.16	12 (3%)	45	45	24, 40, 82, 113	0

The worst 5 of 12 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	HHH	94	GLU	6.6
2	III	65	GLN	4.8
2	III	58	GLU	3.3
1	HHH	259	SER	3.2
2	III	32	SER	2.8

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.

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
4	NAG	HHH	301	14/15	0.81	0.17	57,81,85,86	0
5	NA	HHH	302	1/1	0.96	0.11	28,28,28,28	0

6.5   Other polymers ⓘ

There are no such residues in this entry.