wwPDB X-ray Structure Validation Summary Report (i)

May 17, 2020 – 10:21 pm BST

PDB ID	:	2UUF
Title	:	Thrombin-hirugen binary complex at 1.26A resolution
Authors	:	Ahmed, H.U.; Blakeley, M.P.; Cianci, M.; Cruickshank, D.W.J.; Hubbard, J.A.; Helliwell, J.R.
Deposited on	:	2007-03-02
Resolution	:	1.26 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 following versions of software and data (see references (1)) were used in the production of this report:

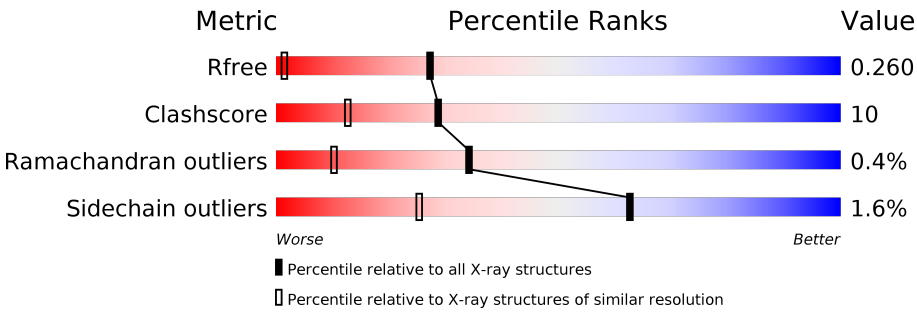
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, 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 1.26 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	1023 (1.28-1.24)
Clashscore	141614	1060 (1.28-1.24)
Ramachandran outliers	138981	1029 (1.28-1.24)
Sidechain outliers	138945	1028 (1.28-1.24)

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%

Mol	Chain	Length	Quality of chain
1	A	36	50% 17% 8% • 22%
2	B	259	80% 15% • •
3	H	10	100%

2 Entry composition (i)

There are 6 unique types of molecules in this entry. The entry contains 2738 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 HUMAN ALPHA THROMBIN.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	28	236	149	38	48	1	0	1	0

- Molecule 2 is a protein called THROMBIN.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	B	252	2060	1315	362	368	15	0	7	0

- Molecule 3 is a protein called HIRUDIN I.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
3	H	10	94	59	10	24	1	0	0	0

- Molecule 4 is SODIUM ION (three-letter code: NA) (formula: Na).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	Na		
4	B	1	1	1	0	0

- Molecule 5 is CALCIUM ION (three-letter code: CA) (formula: Ca).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	Ca		
5	B	1	1	1	0	0

- Molecule 6 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
6	A	42	42	42	0	0

Continued on next page...

Continued from previous page...

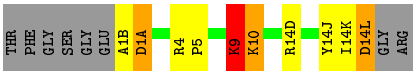
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
6	B	297	Total O 297 297	0	0
6	H	7	Total O 7 7	0	0

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. 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. 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.

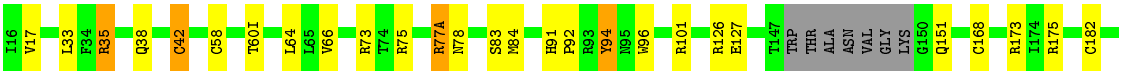
- Molecule 1: HUMAN ALPHA THROMBIN

Chain A:



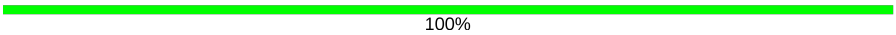
- Molecule 2: THROMBIN

Chain B:



- Molecule 3: HIRUDIN I

Chain H:



There are no outlier residues recorded for this chain.

4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α, β, γ	70.28Å 71.29Å 72.04Å 90.00° 99.90° 90.00°	Depositor
Resolution (Å)	49.39 – 1.26 49.66 – 1.26	Depositor EDS
% Data completeness (in resolution range)	93.0 (49.39-1.26) 94.7 (49.66-1.26)	Depositor EDS
R_{merge}	0.06	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.37 (at 1.26Å)	Xtriage
Refinement program	SHELXL-97	Depositor
R, R_{free}	0.146 , 0.194 0.240 , 0.260	Depositor DCC
R_{free} test set	4490 reflections (5.02%)	wwPDB-VP
Wilson B-factor (Å^2)	13.0	Xtriage
Anisotropy	0.174	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.36 , 45.1	EDS
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	2738	wwPDB-VP
Average B, all atoms (Å^2)	23.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.82% 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: NA, CA, TYS

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	0.86	0/241	1.85	4/320 (1.2%)
2	B	0.96	6/2139 (0.3%)	1.47	26/2890 (0.9%)
3	H	0.93	0/78	1.24	0/103
All	All	0.95	6/2458 (0.2%)	1.50	30/3313 (0.9%)

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	1	2

The worst 5 of 6 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	B	247	GLU	C-O	23.41	1.67	1.23
2	B	186	PRO	C-N	-11.86	1.06	1.34
2	B	186(A)	ASP	N-CA	7.89	1.62	1.46
2	B	186(A)	ASP	CB-CG	5.71	1.63	1.51
2	B	168	CYS	CB-SG	-5.41	1.73	1.81

The worst 5 of 30 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	9[A]	LYS	CB-CA-C	-14.25	81.90	110.40
1	A	9[B]	LYS	CB-CA-C	-14.25	81.90	110.40
2	B	75	ARG	NE-CZ-NH1	-14.23	113.18	120.30
2	B	173	ARG	NE-CZ-NH1	-13.15	113.72	120.30
2	B	126	ARG	NE-CZ-NH1	12.03	126.31	120.30

All (1) chirality outliers are listed below:

Mol	Chain	Res	Type	Atom
1	A	14(L)	ASP	CA

All (2) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	14(J)	TYR	Mainchain
1	A	9[A]	LYS	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	236	0	242	21	0
2	B	2060	0	2038	25	0
3	H	94	0	73	0	0
4	B	1	0	0	0	0
5	B	1	0	0	0	0
6	A	42	0	0	6	0
6	B	297	0	0	8	1
6	H	7	0	0	0	0
All	All	2738	0	2353	46	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 10.

The worst 5 of 46 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:B:247:GLU:O	2:B:247:GLU:C	1.67	1.30
2:B:33[B]:LEU:HD23	2:B:42[B]:CYS:SG	2.01	1.00
1:A:5:PRO:HA	1:A:9[B]:LYS:HD2	1.49	0.95
1:A:9[B]:LYS:HE3	6:A:1012:HOH:O	1.67	0.92
1:A:14(K):ILE:HG13	1:A:14(L):ASP:N	1.90	0.87

All (1) symmetry-related close contacts are listed below. The label for Atom-2 includes the sym-

metry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
6:B:1098:HOH:O	6:B:1098:HOH:O[2_657]	0.99	1.21

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	27/36 (75%)	25 (93%)	1 (4%)	1 (4%)	3	0
2	B	255/259 (98%)	247 (97%)	8 (3%)	0	100	100
3	H	7/10 (70%)	7 (100%)	0	0	100	100
All	All	289/305 (95%)	279 (96%)	9 (3%)	1 (0%)	34	16

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	10	LYS

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	27/31 (87%)	26 (96%)	1 (4%)	34	4
2	B	226/225 (100%)	223 (99%)	3 (1%)	69	32
3	H	9/9 (100%)	9 (100%)	0	100	100
All	All	262/265 (99%)	258 (98%)	4 (2%)	62	28

All (4) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	14(L)	ASP
2	B	64	LEU
2	B	182	CYS
2	B	240	LYS

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

Mol	Chain	Res	Type
2	B	244	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)

1 non-standard protein/DNA/RNA residue is 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	$\#\|Z\| > 2$
3	TYS	H	63	3	15,16,17	1.86	5 (33%)	18,22,24	1.51	4 (22%)

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
3	TYS	H	63	3	-	0/10/11/13	0/1/1/1

All (5) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	H	63	TYS	OH-CZ	-4.75	1.34	1.42
3	H	63	TYS	CE1-CD1	-2.68	1.33	1.38
3	H	63	TYS	CE2-CD2	-2.30	1.34	1.38
3	H	63	TYS	CE2-CZ	-2.24	1.34	1.38
3	H	63	TYS	OH-S	2.05	1.61	1.58

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
3	H	63	TYS	O3-S-O2	4.29	123.39	108.49
3	H	63	TYS	OH-S-O1	-2.55	100.28	107.71
3	H	63	TYS	O3-S-OH	2.38	111.57	105.83
3	H	63	TYS	OH-S-O2	-2.36	100.83	107.71

There are no chirality outliers.

There are no torsion outliers.

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)

Of 2 ligands modelled in this entry, 2 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 (i)

There are no such residues in this entry.

5.8 Polymer linkage issues (i)

The following chains have linkage breaks:

Mol	Chain	Number of breaks
2	B	1

All chain breaks are listed below:

Model	Chain	Residue-1	Atom-1	Residue-2	Atom-2	Distance (Å)
1	B	186:PRO	C	186(A):ASP	N	1.06

6 Fit of model and data (i)

6.1 Protein, DNA and RNA chains (i)

Unable to reproduce the depositors R factor - this section is therefore empty.

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

Unable to reproduce the depositors R factor - this section is therefore empty.

6.3 Carbohydrates (i)

Unable to reproduce the depositors R factor - this section is therefore empty.

6.4 Ligands (i)

Unable to reproduce the depositors R factor - this section is therefore empty.

6.5 Other polymers (i)

Unable to reproduce the depositors R factor - this section is therefore empty.