wwPDB X-ray Structure Validation Summary Report (i)

Jan 31, 2016 – 10:15 PM GMT

PDB ID	:	1STF
Title	:	THE REFINED 2.4 ANGSTROMS X-RAY CRYSTAL STRUCTURE OF RECOMBINANT HUMAN STEFIN B IN COMPLEX WITH THE CYSTEINE PROTEINASE PAPAIN: A NOVEL TYPE OF PROTEINASE INHIBITOR INTERACTION
Authors	:	Stubbs, M.T.; Laber, B.; Bode, W.
Deposited on	:	1993-04-21
Resolution	:	2.37 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 (i)) were used in the production of this report:

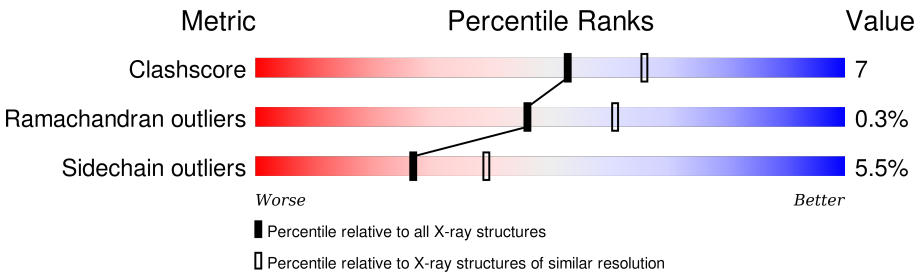
MolProbity	:	4.02b-467
Mogul	:	1.7 (RC4), CSD as536be (2015)
Xtriage (Phenix)	:	NOT EXECUTED
EDS	:	NOT EXECUTED
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
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 2.37 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))
Clashscore	102246	4595 (2.40-2.36)
Ramachandran outliers	100387	4520 (2.40-2.36)
Sidechain outliers	100360	4522 (2.40-2.36)

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.

Note EDS was not executed.

Mol	Chain	Length	Quality of chain
1	E	212	81% 17% •
2	I	98	70% 22% 5% •

2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 2590 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 PAPAIN.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	E	212	1659	1052	293	307	7	18	0	0

There are 3 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
E	47	GLN	GLU	CONFLICT	UNP P00784
E	118	GLN	GLU	CONFLICT	UNP P00784
E	135	GLN	GLU	CONFLICT	UNP P00784

- Molecule 2 is a protein called STEFIN B (CYSTATIN B).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	I	98	789	503	135	149	2	28	0	0

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
I	8	SER	CYS	CONFLICT	UNP P04080
I	36	TYR	GLU	CONFLICT	UNP P04080

- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
3	E	100	100	100	0	0
3	I	42	42	42	0	0

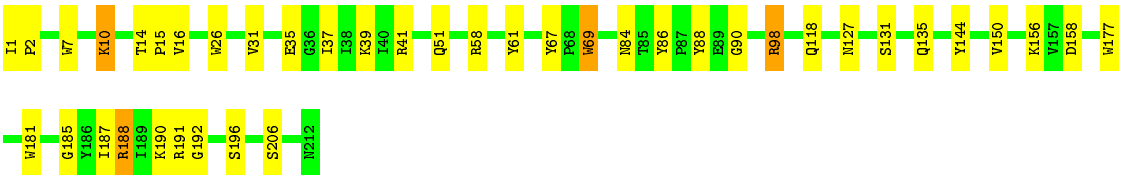
3 Residue-property plots (i)

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.

Note EDS was not executed.

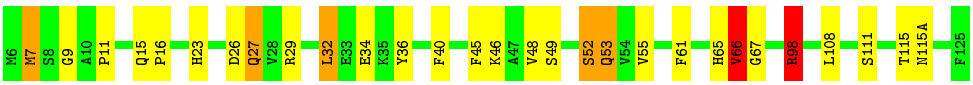
- Molecule 1: PAPAIN

Chain E:



- Molecule 2: STEFIN B (CYSTATIN B)

Chain I:



4 Data and refinement statistics (i)

Xtriage (Phenix) and EDS were not executed - this section will therefore be incomplete.

Property	Value	Source
Space group	P 31 2 1	Depositor
Cell constants a, b, c, α, β, γ	67.00Å 67.00Å 169.30Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	8.00 – 2.37	Depositor
% Data completeness (in resolution range)	(Not available) (8.00-2.37)	Depositor
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
Refinement program	EREF	Depositor
R, R_{free}	0.190 , (Not available)	Depositor
Estimated twinning fraction	No twinning to report.	Xtriage
Total number of atoms	2590	wwPDB-VP
Average B, all atoms ($Å^2$)	14.0	wwPDB-VP

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: CCS

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	E	0.92	5/1692 (0.3%)	1.29	6/2295 (0.3%)
2	I	0.93	0/809	1.40	4/1093 (0.4%)
All	All	0.92	5/2501 (0.2%)	1.33	10/3388 (0.3%)

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	E	0	2
2	I	0	5
All	All	0	7

All (5) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	E	26	TRP	NE1-CE2	-8.45	1.26	1.37
1	E	177	TRP	NE1-CE2	-7.89	1.27	1.37
1	E	181	TRP	NE1-CE2	-7.85	1.27	1.37
1	E	7	TRP	NE1-CE2	-7.04	1.28	1.37
1	E	69	TRP	NE1-CE2	-6.99	1.28	1.37

The worst 5 of 10 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	I	98	ARG	NE-CZ-NH1	6.44	123.52	120.30
1	E	191	ARG	NE-CZ-NH2	-6.32	117.14	120.30
1	E	158	ASP	CB-CG-OD1	6.07	123.76	118.30
2	I	26	ASP	CB-CG-OD1	5.80	123.52	118.30

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	E	41	ARG	NE-CZ-NH1	5.75	123.18	120.30

There are no chirality outliers.

5 of 7 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	E	188	ARG	Mainchain
1	E	61	TYR	Sidechain
2	I	27	GLN	Mainchain
2	I	65	HIS	Mainchain
2	I	66	VAL	Mainchain

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	E	1659	0	1597	18	1
2	I	789	0	769	14	1
3	E	100	0	0	1	0
3	I	42	0	0	1	0
All	All	2590	0	2366	31	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 7.

The worst 5 of 31 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:E:118:GLN:HE22	1:E:192:GLY:H	1.37	0.72
1:E:150:VAL:HG12	1:E:190:LYS:HD2	1.82	0.60
2:I:23:HIS:O	2:I:27:GLN:HG2	2.06	0.56
1:E:135:GLN:HG3	1:E:156:LYS:O	2.06	0.54
2:I:98:ARG:HD2	2:I:111:SER:HB2	1.89	0.53

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)
1:E:98:ARG:NH2	2:I:34:GLU:O[6_665]	1.91	0.29

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	E	209/212 (99%)	202 (97%)	7 (3%)	0	100	100
2	I	96/98 (98%)	89 (93%)	6 (6%)	1 (1%)	19	26
All	All	305/310 (98%)	291 (95%)	13 (4%)	1 (0%)	46	61

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	I	55	VAL

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	E	169/169 (100%)	163 (96%)	6 (4%)	42	61
2	I	87/87 (100%)	79 (91%)	8 (9%)	11	15
All	All	256/256 (100%)	242 (94%)	14 (6%)	27	39

5 of 14 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	I	7	MET
2	I	32	LEU
2	I	66	VAL
1	E	196	SER
2	I	53	GLN

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 8 such sidechains are listed below:

Mol	Chain	Res	Type
1	E	118	GLN
2	I	105(A)	ASN
2	I	27	GLN
1	E	92	GLN
1	E	127	ASN

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	$\#\|Z\| > 2$
1	CCS	E	25	1	5,9,10	1.31	1 (20%)	4,10,12	1.53	1 (25%)

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
1	CCS	E	25	1	-	0/4/8/10	0/0/0/0

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	E	25	CCS	CD-SG	-2.76	1.77	1.82

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	E	25	CCS	O-C-CA	-2.51	118.95	125.49

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)

There are no ligands in this entry.

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)

EDS was not executed - this section will therefore be empty.

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

EDS was not executed - this section will therefore be empty.

6.3 Carbohydrates (i)

EDS was not executed - this section will therefore be empty.

6.4 Ligands (i)

EDS was not executed - this section will therefore be empty.

6.5 Other polymers (i)

EDS was not executed - this section will therefore be empty.