Full wwPDB X-ray Structure Validation Report (i)

Feb 1, 2016 – 01:58 AM GMT

PDB ID	:	2F1G
Title	:	Cathepsin S in complex with non-covalent 2-(Benzoxazol-2-ylamino)-acetamide
Authors	:	Spraggon, G.; Hornsby, M.; Lesley, S.A.; Tully, D.C.; Harris, J.L.; Karenewsky, D.S.; Kulathila, R.; Clark, K.
Deposited on	:	2005-11-14
Resolution	:	1.90 Å(reported)

This is a Full wwPDB X-ray Structure Validation 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:

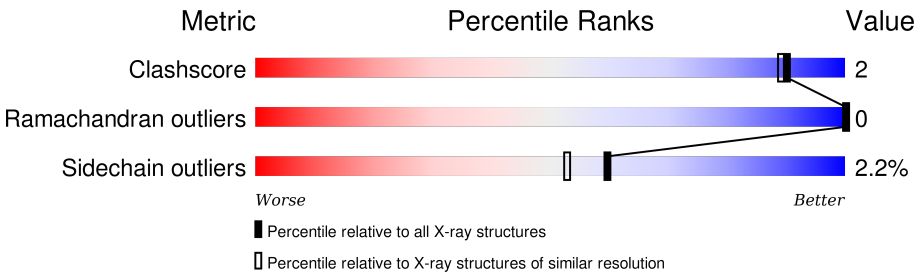
MolProbity	:	4.02b-467
Mogul	:	1.7 (RC4), CSD as536be (2015)
Xtriage (Phenix)	:	NOT EXECUTED
EDS	:	NOT EXECUTED
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
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 1.90 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	5398 (1.90-1.90)
Ramachandran outliers	100387	5338 (1.90-1.90)
Sidechain outliers	100360	5339 (1.90-1.90)

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	A	220	96% •
1	B	220	91% 8% •

2 Entry composition (i)

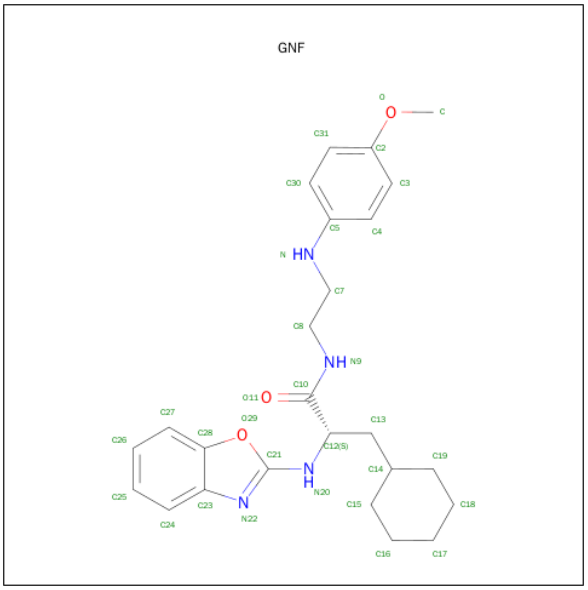
There are 4 unique types of molecules in this entry. The entry contains 3820 atoms, of which 0 are hydrogens and 0 are deuteriums.

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 Cathepsin S.

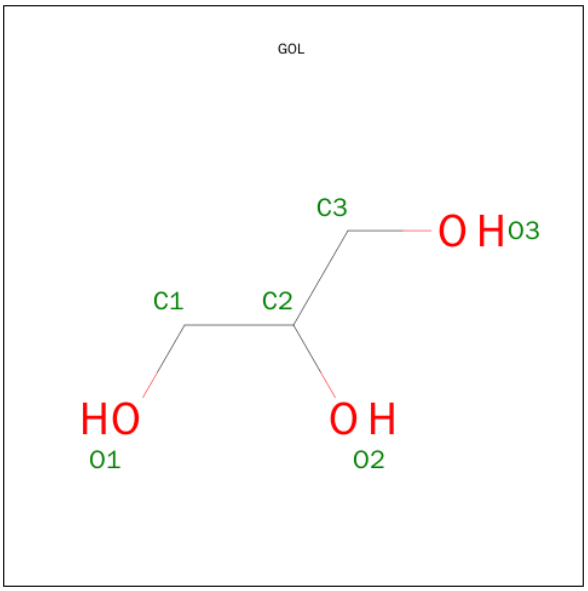
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	220	1731	1093	297	329	12	4	3	0
1	B	217	1693	1069	288	324	12	4	1	0

- Molecule 2 is N 2 -1,3-BENZOXAZOL-2-YL-3-CYCLOHEXYL-N-{2-[(4-METHOXYPHENYL)AMINO]ETHYL}-L-ALANINAMIDE (three-letter code: GNF) (formula: $C_{25}H_{32}N_4O_3$).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
			Total	C	N	O		
2	B	1	32	25	4	3	0	0
2	A	1	32	25	4	3	0	0

- Molecule 3 is GLYCEROL (three-letter code: GOL) (formula: $C_3H_8O_3$).



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

- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
4	A	165	165	165	0	0
			Total	O		
4	B	161	161	161	0	0

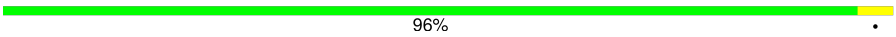
3 Residue-property plots (i)

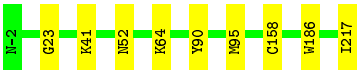
These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are color-coded according 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: Cathepsin S

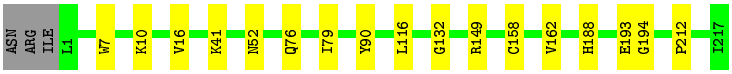
Chain A:





- Molecule 1: Cathepsin S

Chain B:



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 41 2 2	Depositor
Cell constants a, b, c, α, β, γ	85.25Å 85.25Å 151.61Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	84.51 – 1.90	Depositor
% Data completeness (in resolution range)	92.4 (84.51-1.90)	Depositor
R_{merge}	0.11	Depositor
R_{sym}	0.11	Depositor
Refinement program	REFMAC 5.2.0005	Depositor
R, R_{free}	0.189 , 0.230	Depositor
Estimated twinning fraction	No twinning to report.	Xtriage
Total number of atoms	3820	wwPDB-VP
Average B, all atoms ($Å^2$)	21.0	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: GOL, GNF

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	0.67	0/1783	0.67	0/2408
1	B	0.69	0/1739	0.69	0/2350
All	All	0.68	0/3522	0.68	0/4758

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no planarity outliers.

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	1731	0	1659	6	0
1	B	1693	0	1614	8	0
2	A	32	0	32	6	0
2	B	32	0	32	2	0
3	B	6	0	8	0	0
4	A	165	0	0	0	0
4	B	161	0	0	2	0
All	All	3820	0	3345	16	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 2.

All (16) 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:A:186:TRP:HE1	2:A:1002:GNF:H1	1.35	0.92
1:A:186:TRP:NE1	2:A:1002:GNF:H1	1.98	0.77
1:B:10:LYS:NZ	4:B:1041:HOH:O	2.29	0.66
1:A:64:LYS:CE	2:A:1002:GNF:C25	2.84	0.56
1:B:149:ARG:NE	4:B:1138:HOH:O	2.33	0.56
1:A:64:LYS:HE2	2:A:1002:GNF:C25	2.37	0.55
2:A:1002:GNF:H82	2:A:1002:GNF:C30	2.37	0.55
1:B:162:VAL:HG23	2:B:1001:GNF:H182	1.88	0.54
1:B:188:HIS:O	1:B:193:GLU:HA	2.08	0.54
1:B:116:LEU:HD22	1:B:212:PRO:HB2	2.00	0.44
1:A:41:LYS:HD2	1:A:217:ILE:HG21	1.99	0.43
1:A:23:GLY:HA2	2:A:1002:GNF:H31	2.00	0.43
2:B:1001:GNF:H4	2:B:1001:GNF:H82	2.01	0.43
1:B:76:GLN:HA	1:B:79:ILE:HD12	2.01	0.42
1:B:7:TRP:CZ2	1:B:132:GLY:HA2	2.55	0.41
1:B:16:VAL:HG21	1:B:194:GLY:HA3	2.03	0.41

There are no symmetry-related clashes.

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	221/220 (100%)	217 (98%)	4 (2%)	0	100	100
1	B	216/220 (98%)	211 (98%)	5 (2%)	0	100	100
All	All	437/440 (99%)	428 (98%)	9 (2%)	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	183/181 (101%)	179 (98%)	4 (2%)	60	53
1	B	179/181 (99%)	175 (98%)	4 (2%)	60	53
All	All	362/362 (100%)	354 (98%)	8 (2%)	60	53

All (8) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	52	ASN
1	A	90	TYR
1	A	95	MET
1	A	158	CYS
1	B	41	LYS
1	B	52	ASN
1	B	90	TYR
1	B	158	CYS

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	GNF	A	1002	-	32,35,35	0.75	0	35,46,46	0.97	2 (5%)
2	GNF	B	1001	-	32,35,35	0.83	1 (3%)	35,46,46	1.25	6 (17%)
3	GOL	B	1002	-	5,5,5	0.39	0	5,5,5	0.65	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	GNF	A	1002	-	-	0/19/29/29	0/3/4/4
2	GNF	B	1001	-	-	0/19/29/29	0/3/4/4
3	GOL	B	1002	-	-	0/4/4/4	0/0/0/0

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	B	1001	GNF	C4-C5	2.69	1.43	1.39

All (8) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	B	1001	GNF	C30-C5-N	-3.42	114.50	121.06
2	A	1002	GNF	C10-C12-N20	-2.52	104.17	111.26
2	A	1002	GNF	C-O-C2	-2.43	111.83	117.51
2	B	1001	GNF	C13-C12-C10	-2.22	104.82	110.60
2	B	1001	GNF	C16-C15-C14	-2.15	108.76	112.22
2	B	1001	GNF	C8-N9-C10	-2.09	118.40	122.53
2	B	1001	GNF	C12-C10-N9	2.22	121.08	116.72
2	B	1001	GNF	C4-C5-N	2.63	126.09	121.06

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

2 monomers are involved in 8 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	A	1002	GNF	6	0
2	B	1001	GNF	2	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)

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.