wwPDB X-ray Structure Validation Summary Report ⓘ

May 23, 2020 – 05:26 pm BST

PDB ID	:	2Z9I
Title	:	Crystal structure of RV0983 from Mycobacterium tuberculosis- Proteolytically active form
Authors	:	Palaninathan, S.K.; Mohamedmohaideen, N.N.; Sacchettini, J.C.
Deposited on	:	2007-09-20
Resolution	:	2.00 Å(reported)

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

|		
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	NOT EXECUTED
EDS	:	NOT EXECUTED
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
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 ⓘ

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

The reported resolution of this entry is 2.00 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.

■ Percentile relative to all X-ray structures
▯ Percentile relative to X-ray structures of similar resolution
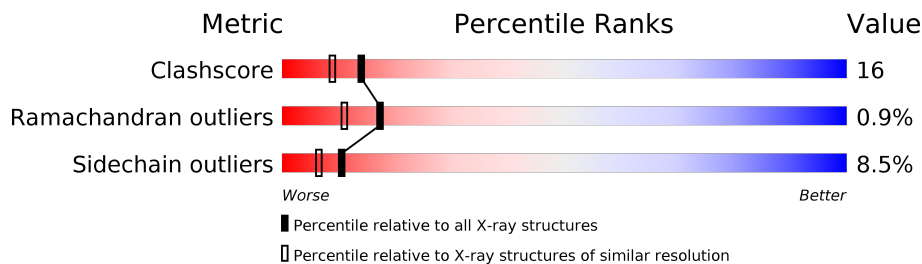
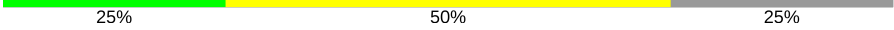| Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
Clashscore	141614	9178 (2.00-2.00)
Ramachandran outliers	138981	9054 (2.00-2.00)
Sidechain outliers	138945	9053 (2.00-2.00)

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%

Note EDS was not executed.

Mol	Chain	Length	Quality of chain
1	A	324	63% 19% 5% 12%
1	B	324	60% 20% • 17%
1	C	324	62% 15% • 20%
2	D	5	60% 20% 20%
2	E	5	20% 60% 20%
2	F	5	40% 40% 20%
3	G	4	25% 50% 25%
3	H	4	25% 25% 50%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
3	I	4	

2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 5920 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 PROBABLE SERINE PROTEASE PEPD.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	284	Total 1929	C 1197	N 338	O 392	Se 2	0	0	0
1	B	269	Total 1788	C 1107	N 314	O 365	Se 2	0	0	0
1	C	260	Total 1726	C 1070	N 301	O 353	Se 2	0	0	0

There are 24 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	317	LEU	-	EXPRESSION TAG	UNP O53896
A	318	GLU	-	EXPRESSION TAG	UNP O53896
A	319	HIS	-	EXPRESSION TAG	UNP O53896
A	320	HIS	-	EXPRESSION TAG	UNP O53896
A	321	HIS	-	EXPRESSION TAG	UNP O53896
A	322	HIS	-	EXPRESSION TAG	UNP O53896
A	323	HIS	-	EXPRESSION TAG	UNP O53896
A	324	HIS	-	EXPRESSION TAG	UNP O53896
B	317	LEU	-	EXPRESSION TAG	UNP O53896
B	318	GLU	-	EXPRESSION TAG	UNP O53896
B	319	HIS	-	EXPRESSION TAG	UNP O53896
B	320	HIS	-	EXPRESSION TAG	UNP O53896
B	321	HIS	-	EXPRESSION TAG	UNP O53896
B	322	HIS	-	EXPRESSION TAG	UNP O53896
B	323	HIS	-	EXPRESSION TAG	UNP O53896
B	324	HIS	-	EXPRESSION TAG	UNP O53896
C	317	LEU	-	EXPRESSION TAG	UNP O53896
C	318	GLU	-	EXPRESSION TAG	UNP O53896
C	319	HIS	-	EXPRESSION TAG	UNP O53896
C	320	HIS	-	EXPRESSION TAG	UNP O53896
C	321	HIS	-	EXPRESSION TAG	UNP O53896
C	322	HIS	-	EXPRESSION TAG	UNP O53896
C	323	HIS	-	EXPRESSION TAG	UNP O53896

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
C	324	HIS	-	EXPRESSION TAG	UNP O53896

- Molecule 2 is a protein called SVEQV.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
2	D	4	Total 32	C 20	N 5	O 7	0	0	0
2	E	4	Total 32	C 20	N 5	O 7	0	0	0
2	F	5	Total 38	C 23	N 6	O 9	0	0	0

- Molecule 3 is a protein called GATV.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
3	G	4	Total 24	C 14	N 4	O 6	0	0	0
3	H	2	Total 8	C 4	N 2	O 2	0	0	0
3	I	3	Total 16	C 10	N 3	O 3	0	0	0

- Molecule 4 is water.

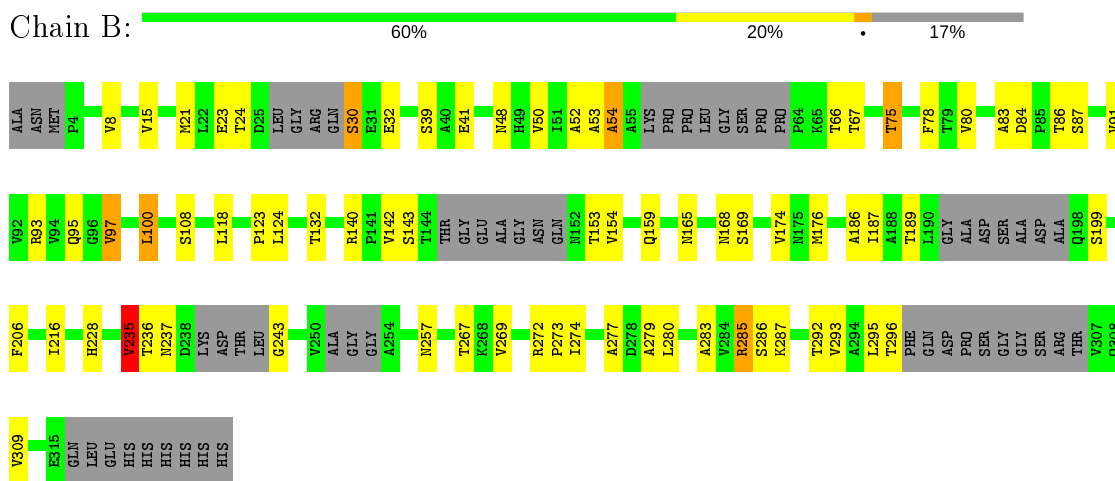| Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	111	Total 111	O 111	0	0
4	B	127	Total 127	O 127	0	0
4	C	84	Total 84	O 84	0	0
4	E	2	Total 2	O 2	0	0
4	F	1	Total 1	O 1	0	0
4	G	2	Total 2	O 2	0	0

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.
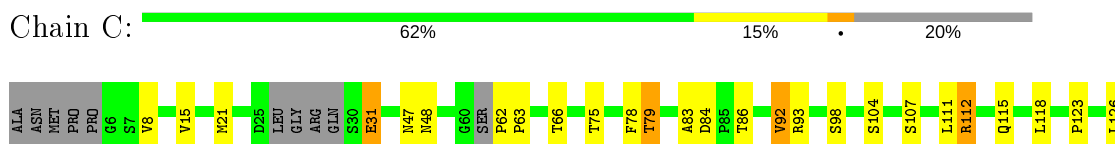
Note EDS was not executed.

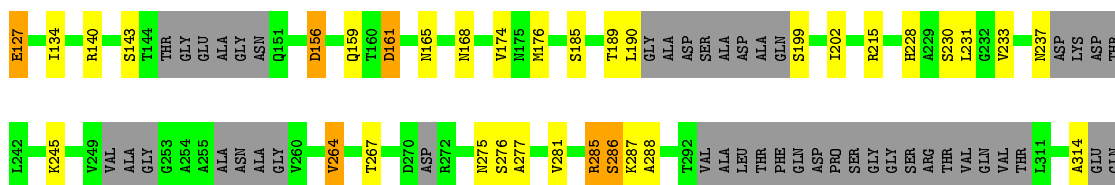- Molecule 1: PROBABLE SERINE PROTEASE PEPD
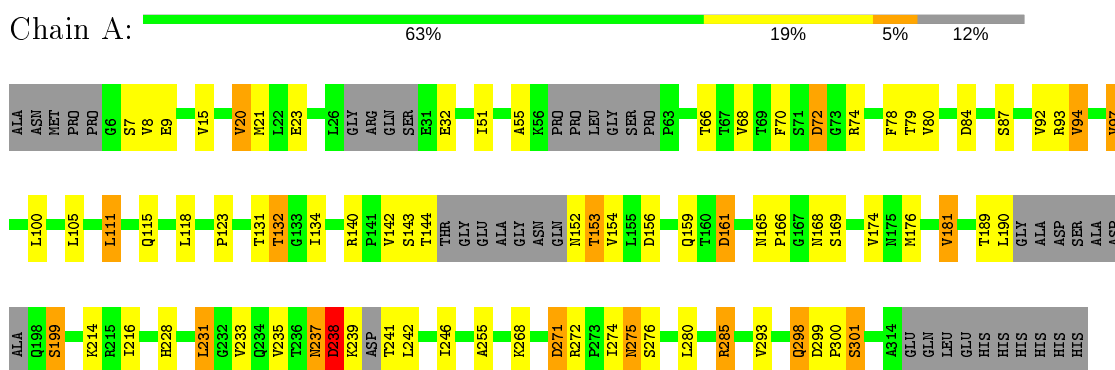
Chain A:

- Molecule 1: PROBABLE SERINE PROTEASE PEPD

Chain B:

- Molecule 1: PROBABLE SERINE PROTEASE PEPD

Chain C:

E127 I134 R140 S143 T144 THR GLY GLU ALA GLY ASN Q151 D156 Q159 T160 D161 N165 N168 V174 N175 M176 S185 T189 L190 GLY ALA ASP SER ALA ASP ALA GLN S199 I202 R215 H228 A229 S230 L231 G232 V233 N237 ASP LYS ASP THR

L242 K245 V249 VAL ALA GLY G253 A254 A255 ALA ASN ALA GLY V260 V264 T267 D270 ASP R272 N275 S276 A277 V281 R285 S286 K287 A288 T292 VAL ALA LEU THR PHE GLN ASP PRO SER GLY GLY SER ARG THR VAL GLN VAL THR L311 A314 GLU GLN

LEU GLU HIS HIS HIS HIS HIS HIS

• Molecule 2: SVEQV

Chain D:

| 60% | 20% | 20% |

SER V408 V411

• Molecule 2: SVEQV

Chain E:

| 20% | 60% | 20% |

SER V408 E409 Q410 V411

• Molecule 2: SVEQV

Chain F:

| 40% | 40% | 20% |

S407 V408 E409 Q410 V411

• Molecule 3: GATV

Chain G:

| 25% | 50% | 25% |

G1 A2 T3 V4

• Molecule 3: GATV

Chain H:

| 25% | 25% | 50% |

GLY ALA T3 V4

• Molecule 3: GATV

Chain I:

| 25% | 50% | 25% |

GLY A2 T3 V4

4 Data and refinement statistics (i)

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

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α, β, γ	149.58Å 89.07Å 69.41Å 90.00° 97.55° 90.00°	Depositor
Resolution (Å)	50.00 – 2.00	Depositor
% Data completeness (in resolution range)	98.8 (50.00-2.00)	Depositor
R_{merge}	0.05	Depositor
R_{sym}	(Not available)	Depositor
Refinement program	REFMAC 5.2.0019	Depositor
R, R_{free}	0.225 , 0.273	Depositor
Estimated twinning fraction	No twinning to report.	Xtriage
Total number of atoms	5920	wwPDB-VP
Average B, all atoms (Å^2)	47.0	wwPDB-VP

5 Model quality (i)

5.1 Standard geometry (i)

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	0.71	0/1942	0.92	3/2647 (0.1%)
1	B	0.68	0/1796	0.95	7/2445 (0.3%)
1	C	0.65	0/1734	0.83	2/2361 (0.1%)
2	D	0.76	0/31	1.32	0/41
2	E	1.09	0/31	1.89	0/41
2	F	0.90	0/37	1.85	1/49 (2.0%)
3	G	2.12	2/23 (8.7%)	2.37	1/29 (3.4%)
3	H	1.75	0/7	2.77	0/7
3	I	1.28	0/15	1.90	0/19
All	All	0.70	2/5616 (0.0%)	0.94	14/7639 (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	A	1	2
1	B	0	1
1	C	0	2
2	F	0	1
3	G	0	1
3	H	0	1
3	I	0	1
All	All	1	9

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	G	3	THR	CB-OG1	5.50	1.54	1.43
3	G	1	GLY	N-CA	5.23	1.53	1.46

The worst 5 of 14 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	53	ALA	C-N-CA	6.09	136.94	121.70
3	G	1	GLY	CA-C-N	-5.95	104.11	117.20
1	B	235	VAL	N-CA-C	5.61	126.15	111.00
1	B	54	ALA	CB-CA-C	5.47	118.31	110.10
1	B	93	ARG	O-C-N	5.37	131.29	122.70

All (1) chirality outliers are listed below:

Mol	Chain	Res	Type	Atom
1	A	238	ASP	CA

5 of 9 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	237	ASN	Mainchain
1	A	238	ASP	Mainchain
1	B	235	VAL	Mainchain
1	C	156	ASP	Mainchain
1	C	267	THR	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	A	1929	0	1950	73	0
1	B	1788	0	1763	63	0
1	C	1726	0	1696	42	0
2	D	32	0	31	2	0
2	E	32	0	31	13	0
2	F	38	0	36	1	0
3	G	24	0	26	1	0
3	H	8	0	1	0	0
3	I	16	0	11	5	0
4	A	111	0	0	3	0
4	B	127	0	0	4	0
4	C	84	0	0	3	0
4	E	2	0	0	0	0
4	F	1	0	0	0	0

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
4	G	2	0	0	0	0
All	All	5920	0	5545	175	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 16.

The worst 5 of 175 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:299:ASP:OD2	1:A:301:SER:HB2	1.46	1.15
1:A:239:LYS:C	1:A:241:THR:HG22	1.65	1.15
1:B:285:ARG:HG2	1:B:285:ARG:HH11	1.12	1.14
1:A:132:THR:HG22	4:C:455:HOH:O	1.52	1.09
1:A:285:ARG:HH11	1:A:285:ARG:HG2	0.95	1.09

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	A	272/324 (84%)	266 (98%)	4 (2%)	2 (1%)	22	16
1	B	253/324 (78%)	245 (97%)	6 (2%)	2 (1%)	19	13
1	C	240/324 (74%)	226 (94%)	12 (5%)	2 (1%)	19	13
2	D	2/5 (40%)	2 (100%)	0	0	100	100
2	E	2/5 (40%)	2 (100%)	0	0	100	100
2	F	3/5 (60%)	2 (67%)	0	1 (33%)	0	0
3	G	2/4 (50%)	2 (100%)	0	0	100	100
3	I	1/4 (25%)	1 (100%)	0	0	100	100
All	All	775/995 (78%)	746 (96%)	22 (3%)	7 (1%)	17	11

5 of 7 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	271	ASP
1	B	235	VAL
1	C	264	VAL
1	A	300	PRO
1	B	52	ALA

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	203/242 (84%)	181 (89%)	22 (11%)	6	3
1	B	180/242 (74%)	167 (93%)	13 (7%)	14	9
1	C	176/242 (73%)	163 (93%)	13 (7%)	13	9
2	D	4/5 (80%)	4 (100%)	0	100	100
2	E	4/5 (80%)	4 (100%)	0	100	100
2	F	5/5 (100%)	4 (80%)	1 (20%)	1	0
3	G	2/2 (100%)	2 (100%)	0	100	100
3	I	1/2 (50%)	1 (100%)	0	100	100
All	All	575/745 (77%)	526 (92%)	49 (8%)	10	6

5 of 49 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	30	SER
1	B	199	SER
1	C	231	LEU
1	B	95	GLN
1	B	237	ASN

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 20 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	95	GLN
1	B	159	GLN
1	C	168	ASN
1	A	275	ASN
1	A	298	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)

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)

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 is therefore empty.

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

EDS was not executed - this section is therefore empty.

6.3 Carbohydrates (i)

EDS was not executed - this section is therefore empty.

6.4 Ligands (i)

EDS was not executed - this section is therefore empty.

6.5 Other polymers (i)

EDS was not executed - this section is therefore empty.