Full wwPDB X-ray Structure Validation Report (i)

May 29, 2024 – 11:57 AM EDT

PDB ID	:	4SGB
Title	:	STRUCTURE OF THE COMPLEX OF STREPTOMYCES GRISEUS PROTEINASE B AND POLYPEPTIDE CHYMOTRYPSIN INHIBITOR-1 FROM RUSSET BURBANK POTATO TUBERS AT 2.1 ANGSTROMS RESOLUTION
Authors	:	James, M.; Greenblatt, H.
Deposited on	:	1989-09-21
Resolution	:	2.10 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
https://www.wwpdb.org/validation/2017/XrayValidationReportHelp
with specific help available everywhere you see the (i) symbol.

The types of validation reports are described at
http://www.wwpdb.org/validation/2017/FAQs#types.

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

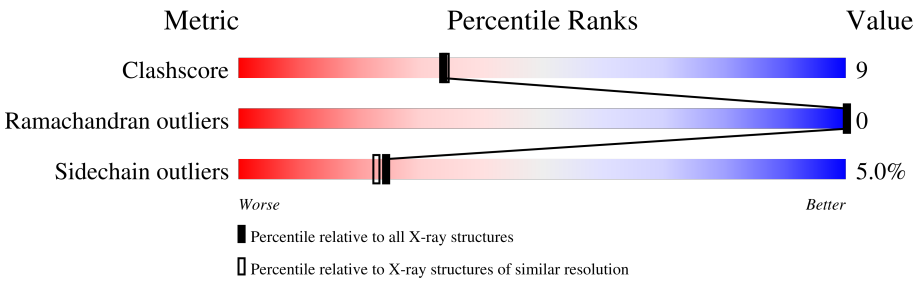
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	NOT EXECUTED
EDS	:	NOT EXECUTED
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36.2

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 2.10 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	141614	5710 (2.10-2.10)
Ramachandran outliers	138981	5647 (2.10-2.10)
Sidechain outliers	138945	5648 (2.10-2.10)

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%

Note EDS was not executed.

Mol	Chain	Length	Quality of chain
1	E	185	85% 12% •
2	I	51	63% 25% 10% •

2 Entry composition (i)

There are 5 unique types of molecules in this entry. The entry contains 1880 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 SERINE PROTEINASE B.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	E	185	1310	801	228	275	6	0	0	0

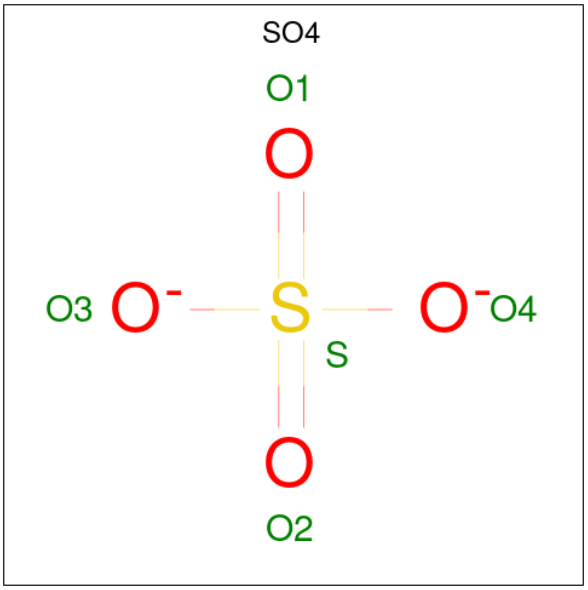
There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
E	236	VAL	SER	conflict	UNP P00777

- Molecule 2 is a protein called POTATO INHIBITOR, PCI-1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	I	51	380	235	66	71	8	0	0	0

- Molecule 3 is SULFATE ION (three-letter code: SO4) (formula: O_4S).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	E	1	Total O S 5 4 1	0	0
3	E	1	Total O S 5 4 1	0	0

- Molecule 4 is CALCIUM ION (three-letter code: CA) (formula: Ca).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	E	1	Total Ca 1 1	0	0

- Molecule 5 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	E	141	Total O 141 141	0	0
5	I	38	Total O 38 38	0	0

3 Residue-property plots (i)

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. 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.

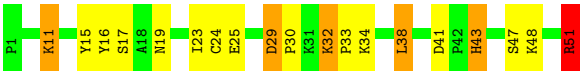
Note EDS was not executed.

• Molecule 1: SERINE PROTEINASE B

Chain E:

• Molecule 2: POTATO INHIBITOR, PCI-1

Chain I:



4 Data and refinement statistics (i)

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

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	50.92Å 46.20Å 52.53Å 90.00° 117.08° 90.00°	Depositor
Resolution (Å)	8.00 – 2.10	Depositor
% Data completeness (in resolution range)	(Not available) (8.00-2.10)	Depositor
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
Refinement program	PROLSQ	Depositor
R, R_{free}	0.142 , (Not available)	Depositor
Estimated twinning fraction	No twinning to report.	Xtriage
Total number of atoms	1880	wwPDB-VP
Average B, all atoms ($Å^2$)	15.0	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: CA, SO4

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	1.04	0/1335	1.73	18/1820 (1.0%)
2	I	0.97	0/391	1.85	7/527 (1.3%)
All	All	1.02	0/1726	1.76	25/2347 (1.1%)

There are no bond length outliers.

All (25) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	E	107	ARG	NE-CZ-NH1	18.54	129.57	120.30
1	E	107	ARG	CD-NE-CZ	12.33	140.86	123.60
1	E	48	ARG	NE-CZ-NH1	12.32	126.46	120.30
1	E	182	ARG	NE-CZ-NH2	-10.23	115.18	120.30
1	E	107	ARG	NE-CZ-NH2	-10.00	115.30	120.30
2	I	38	LEU	CA-CB-CG	9.01	136.03	115.30
1	E	139	ARG	NE-CZ-NH2	8.76	124.68	120.30
1	E	139	ARG	NE-CZ-NH1	-8.21	116.20	120.30
2	I	43	HIS	CA-CB-CG	-7.96	100.07	113.60
2	I	29	ASP	CB-CG-OD2	-7.11	111.90	118.30
2	I	41	ASP	CB-CG-OD1	7.06	124.66	118.30
1	E	34	TYR	CB-CG-CD1	-6.90	116.86	121.00
1	E	41	ARG	NE-CZ-NH1	-6.55	117.03	120.30
1	E	52	PHE	N-CA-CB	6.12	121.61	110.60
1	E	41	ARG	NE-CZ-NH2	6.01	123.30	120.30
2	I	29	ASP	CB-CG-OD1	6.01	123.71	118.30
1	E	158	HIS	CA-CB-CG	-5.97	103.45	113.60
1	E	164	ALA	N-CA-CB	5.85	118.29	110.10
2	I	51	ARG	NE-CZ-NH2	-5.55	117.52	120.30
1	E	34	TYR	CB-CG-CD2	5.33	124.19	121.00
2	I	24	CYS	O-C-N	5.25	131.10	122.70
1	E	161	SER	O-C-N	5.23	131.07	122.70

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	E	31	ASP	CB-CG-OD2	5.18	122.96	118.30
1	E	107	ARG	CG-CD-NE	5.06	122.43	111.80
1	E	138	ARG	NE-CZ-NH2	-5.02	117.79	120.30

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	E	1310	0	1231	13	0
2	I	380	0	355	15	0
3	E	10	0	0	0	0
4	E	1	0	0	0	0
5	E	141	0	0	2	0
5	I	38	0	0	4	0
All	All	1880	0	1586	28	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 9.

All (28) 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:I:19:ASN:HB3	5:I:87:HOH:O	1.70	0.90
2:I:48:LYS:HE3	5:I:82:HOH:O	1.86	0.76
1:E:166:ASN:HD22	1:E:179:GLY:HA2	1.53	0.73
2:I:32:LYS:HG2	2:I:32:LYS:O	1.90	0.72
2:I:25:GLU:OE1	5:I:84:HOH:O	2.07	0.71
2:I:11:LYS:NZ	2:I:29:ASP:HB2	2.11	0.65
1:E:222:SER:HB2	5:E:360:HOH:O	1.97	0.64
1:E:232:THR:O	1:E:236:VAL:HG23	1.97	0.63
1:E:113:ILE:HG13	1:E:113(A):PRO:HD2	1.85	0.58
1:E:120(J):GLN:HG2	1:E:200:TYR:CE1	2.38	0.58
1:E:110:ASN:ND2	1:E:113:ILE:H	2.03	0.57

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
2:I:11:LYS:HZ1	2:I:29:ASP:HB2	1.70	0.55
2:I:51:ARG:NH2	5:I:78:HOH:O	2.41	0.53
2:I:15:TYR:CD1	2:I:47:SER:HB3	2.45	0.51
1:E:166:ASN:ND2	1:E:179:GLY:HA2	2.24	0.50
2:I:32:LYS:O	2:I:32:LYS:CG	2.60	0.49
2:I:17:SER:HA	2:I:43:HIS:O	2.12	0.49
2:I:32:LYS:N	2:I:33:PRO:HD3	2.28	0.49
1:E:110:ASN:C	1:E:110:ASN:HD22	2.21	0.45
1:E:34:TYR:CE1	1:E:41:ARG:HG3	2.52	0.44
2:I:16:TYR:CE2	2:I:48:LYS:HD2	2.53	0.44
1:E:180:MET:HE1	1:E:214:SER:HB2	2.00	0.44
2:I:29:ASP:HA	2:I:30:PRO:HD2	1.76	0.42
2:I:17:SER:HB3	2:I:23:ILE:HD11	2.02	0.41
1:E:241:SER:HB3	5:E:263:HOH:O	2.19	0.41
2:I:11:LYS:HZ2	2:I:29:ASP:HB2	1.83	0.41
1:E:160:GLY:HA3	1:E:185:VAL:HG22	2.01	0.41
1:E:180:MET:CE	1:E:214:SER:CB	3.00	0.40

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	E	183/185 (99%)	176 (96%)	7 (4%)	0	100	100
2	I	49/51 (96%)	47 (96%)	2 (4%)	0	100	100
All	All	232/236 (98%)	223 (96%)	9 (4%)	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	E	138/138 (100%)	134 (97%)	4 (3%)	42 46
2	I	42/42 (100%)	37 (88%)	5 (12%)	5 2
All	All	180/180 (100%)	171 (95%)	9 (5%)	24 23

All (9) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	E	48(C)	SER
1	E	107	ARG
1	E	110	ASN
1	E	241	SER
2	I	11	LYS
2	I	32	LYS
2	I	34	LYS
2	I	38	LEU
2	I	51	ARG

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (3) such sidechains are listed below:

Mol	Chain	Res	Type
1	E	110	ASN
1	E	166	ASN
2	I	19	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)

There are no non-standard protein/DNA/RNA residues in this entry.

5.5 Carbohydrates (i)

There are no monosaccharides in this entry.

5.6 Ligand geometry (i)

Of 3 ligands modelled in this entry, 1 is monoatomic - leaving 2 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	$\#\|Z\| > 2$
3	SO4	E	245	-	4,4,4	0.24	0	6,6,6	0.43	0
3	SO4	E	244	-	4,4,4	0.71	0	6,6,6	1.06	0

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)

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.