W O R L D W I D E

wwPDB

P R O T E I N   D A T A   B A N K

wwPDB X-ray Structure Validation Summary Report (i)

Nov 2, 2024 – 11:34 AM EDT

|		
PDB ID	:	5ELP
Title	:	Ketosynthase from module 1 of the bacillaene synthase from Bacillus amyloliquefaciens FZB42
Authors	:	Wagner, D.T.; Gay, D.C.; Keatinge-Clay, A.T.; Zogzas, C.E.
Deposited on	:	2015-11-04
Resolution	:	2.93 Å(reported)

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 2.93 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	Percentile Ranks	Value
Rfree		0.298
Clashscore		39
Ramachandran outliers		1.6%
Sidechain outliers		5.1%
RSRZ outliers		6.8%

Worse   *Better*

▮ Percentile relative to all X-ray structures

▯ Percentile relative to X-ray structures of similar resolution
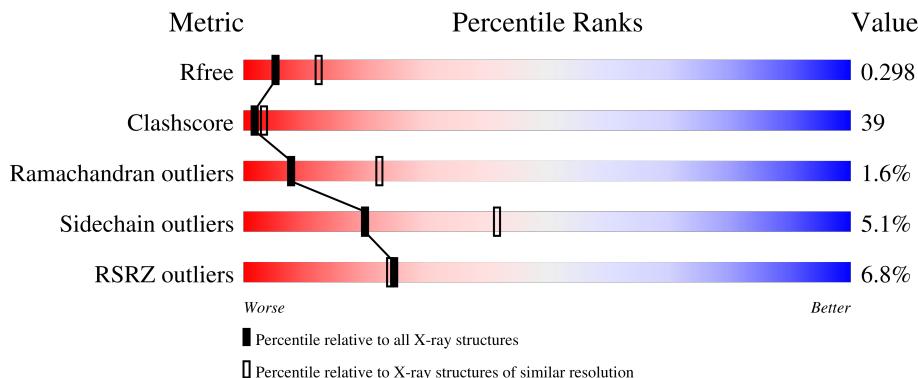
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	164625	1067 (2.96-2.92)
Clashscore	180529	1122 (2.96-2.92)
Ramachandran outliers	177936	1075 (2.96-2.92)
Sidechain outliers	177891	1075 (2.96-2.92)
RSRZ outliers	164620	1067 (2.96-2.92)

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% 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.

Mol	Chain	Length	Quality of chain
1	A	622	3%, 47%, 36%, ·· 13%
1	B	622	5%, 43%, 41%, · 12%
1	C	622	9%, 47%, 35%, 5% · 10%
1	D	622	8%, 40%, 38%, 6% · 16%

2 Entry composition ⓘ

There is only 1 type of molecule in this entry. The entry contains 16749 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 NRPS/PKS protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	C	557	Total	C	N	O	S	0	0	0
			4317	2757	722	821	17			
1	A	540	Total	C	N	O	S	0	0	0
			4169	2656	704	792	17			
1	B	547	Total	C	N	O	S	0	0	0
			4213	2691	699	806	17			
1	D	523	Total	C	N	O	S	0	0	0
			4050	2591	677	766	16			

There are 80 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	-16	MET	-	initiating methionine	UNP Q1RS73
C	-15	GLY	-	expression tag	UNP Q1RS73
C	-14	SER	-	expression tag	UNP Q1RS73
C	-13	SER	-	expression tag	UNP Q1RS73
C	-12	HIS	-	expression tag	UNP Q1RS73
C	-11	HIS	-	expression tag	UNP Q1RS73
C	-10	HIS	-	expression tag	UNP Q1RS73
C	-9	HIS	-	expression tag	UNP Q1RS73
C	-8	HIS	-	expression tag	UNP Q1RS73
C	-7	HIS	-	expression tag	UNP Q1RS73
C	-6	SER	-	expression tag	UNP Q1RS73
C	-5	SER	-	expression tag	UNP Q1RS73
C	-4	GLY	-	expression tag	UNP Q1RS73
C	-3	LEU	-	expression tag	UNP Q1RS73
C	-2	VAL	-	expression tag	UNP Q1RS73
C	-1	PRO	-	expression tag	UNP Q1RS73
C	0	ARG	-	expression tag	UNP Q1RS73
C	1	GLY	-	expression tag	UNP Q1RS73
C	2	SER	-	expression tag	UNP Q1RS73
C	3	SER	-	expression tag	UNP Q1RS73
A	-16	MET	-	initiating methionine	UNP Q1RS73

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
A	-15	GLY	-	expression tag	UNP Q1RS73
A	-14	SER	-	expression tag	UNP Q1RS73
A	-13	SER	-	expression tag	UNP Q1RS73
A	-12	HIS	-	expression tag	UNP Q1RS73
A	-11	HIS	-	expression tag	UNP Q1RS73
A	-10	HIS	-	expression tag	UNP Q1RS73
A	-9	HIS	-	expression tag	UNP Q1RS73
A	-8	HIS	-	expression tag	UNP Q1RS73
A	-7	HIS	-	expression tag	UNP Q1RS73
A	-6	SER	-	expression tag	UNP Q1RS73
A	-5	SER	-	expression tag	UNP Q1RS73
A	-4	GLY	-	expression tag	UNP Q1RS73
A	-3	LEU	-	expression tag	UNP Q1RS73
A	-2	VAL	-	expression tag	UNP Q1RS73
A	-1	PRO	-	expression tag	UNP Q1RS73
A	0	ARG	-	expression tag	UNP Q1RS73
A	1	GLY	-	expression tag	UNP Q1RS73
A	2	SER	-	expression tag	UNP Q1RS73
A	3	SER	-	expression tag	UNP Q1RS73
B	-16	MET	-	initiating methionine	UNP Q1RS73
B	-15	GLY	-	expression tag	UNP Q1RS73
B	-14	SER	-	expression tag	UNP Q1RS73
B	-13	SER	-	expression tag	UNP Q1RS73
B	-12	HIS	-	expression tag	UNP Q1RS73
B	-11	HIS	-	expression tag	UNP Q1RS73
B	-10	HIS	-	expression tag	UNP Q1RS73
B	-9	HIS	-	expression tag	UNP Q1RS73
B	-8	HIS	-	expression tag	UNP Q1RS73
B	-7	HIS	-	expression tag	UNP Q1RS73
B	-6	SER	-	expression tag	UNP Q1RS73
B	-5	SER	-	expression tag	UNP Q1RS73
B	-4	GLY	-	expression tag	UNP Q1RS73
B	-3	LEU	-	expression tag	UNP Q1RS73
B	-2	VAL	-	expression tag	UNP Q1RS73
B	-1	PRO	-	expression tag	UNP Q1RS73
B	0	ARG	-	expression tag	UNP Q1RS73
B	1	GLY	-	expression tag	UNP Q1RS73
B	2	SER	-	expression tag	UNP Q1RS73
B	3	SER	-	expression tag	UNP Q1RS73
D	-16	MET	-	initiating methionine	UNP Q1RS73
D	-15	GLY	-	expression tag	UNP Q1RS73
D	-14	SER	-	expression tag	UNP Q1RS73

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
D	-13	SER	-	expression tag	UNP Q1RS73
D	-12	HIS	-	expression tag	UNP Q1RS73
D	-11	HIS	-	expression tag	UNP Q1RS73
D	-10	HIS	-	expression tag	UNP Q1RS73
D	-9	HIS	-	expression tag	UNP Q1RS73
D	-8	HIS	-	expression tag	UNP Q1RS73
D	-7	HIS	-	expression tag	UNP Q1RS73
D	-6	SER	-	expression tag	UNP Q1RS73
D	-5	SER	-	expression tag	UNP Q1RS73
D	-4	GLY	-	expression tag	UNP Q1RS73
D	-3	LEU	-	expression tag	UNP Q1RS73
D	-2	VAL	-	expression tag	UNP Q1RS73
D	-1	PRO	-	expression tag	UNP Q1RS73
D	0	ARG	-	expression tag	UNP Q1RS73
D	1	GLY	-	expression tag	UNP Q1RS73
D	2	SER	-	expression tag	UNP Q1RS73
D	3	SER	-	expression tag	UNP Q1RS73

3   Residue-property plots ⓘ

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 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.

• Molecule 1: NRPS/PKS protein

Chain C:

• Molecule 1: NRPS/PKS protein

Chain A:

● Molecule 1: NRPS/PKS protein

• Molecule 1: NRPS/PKS protein

Chain D:
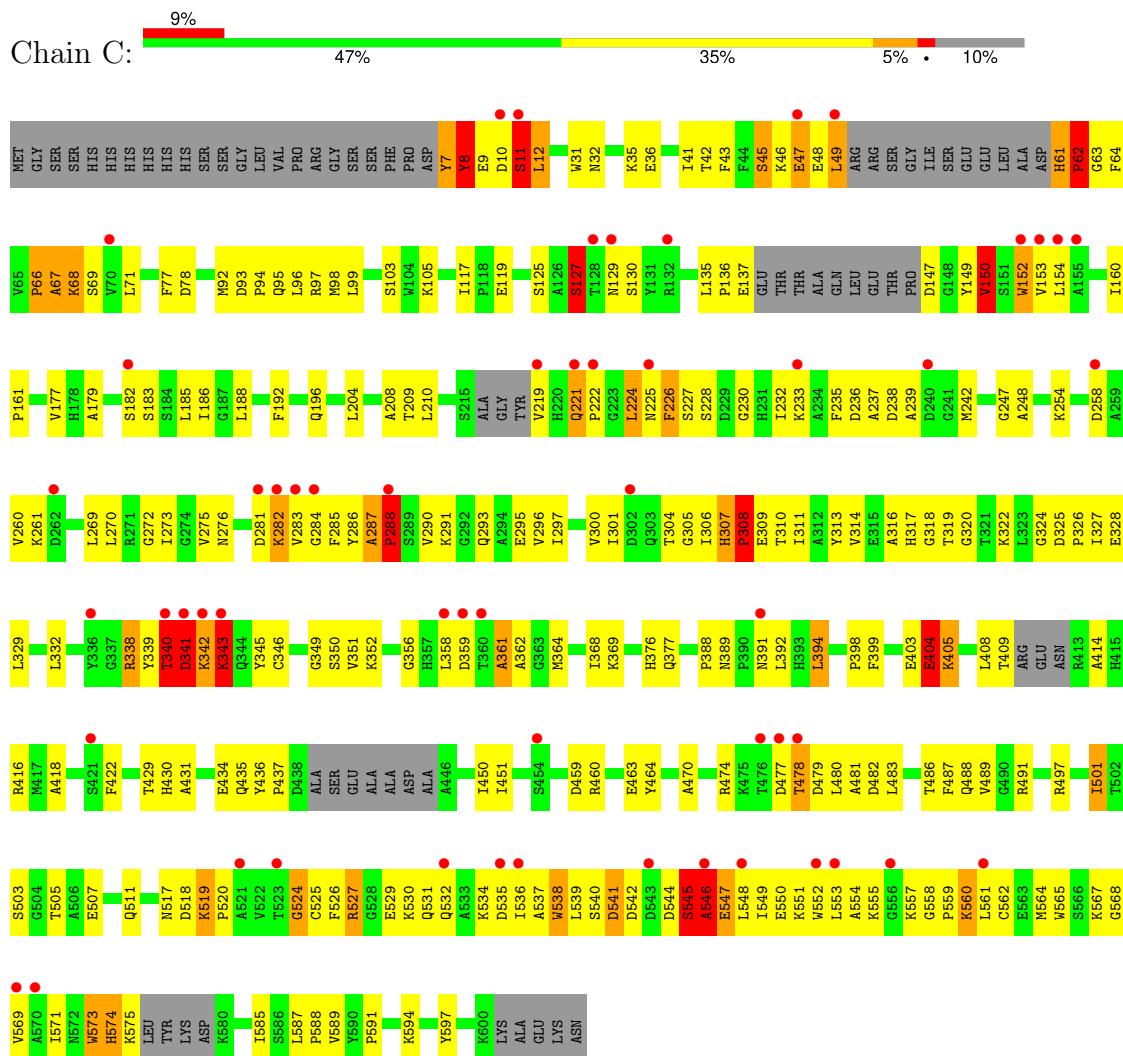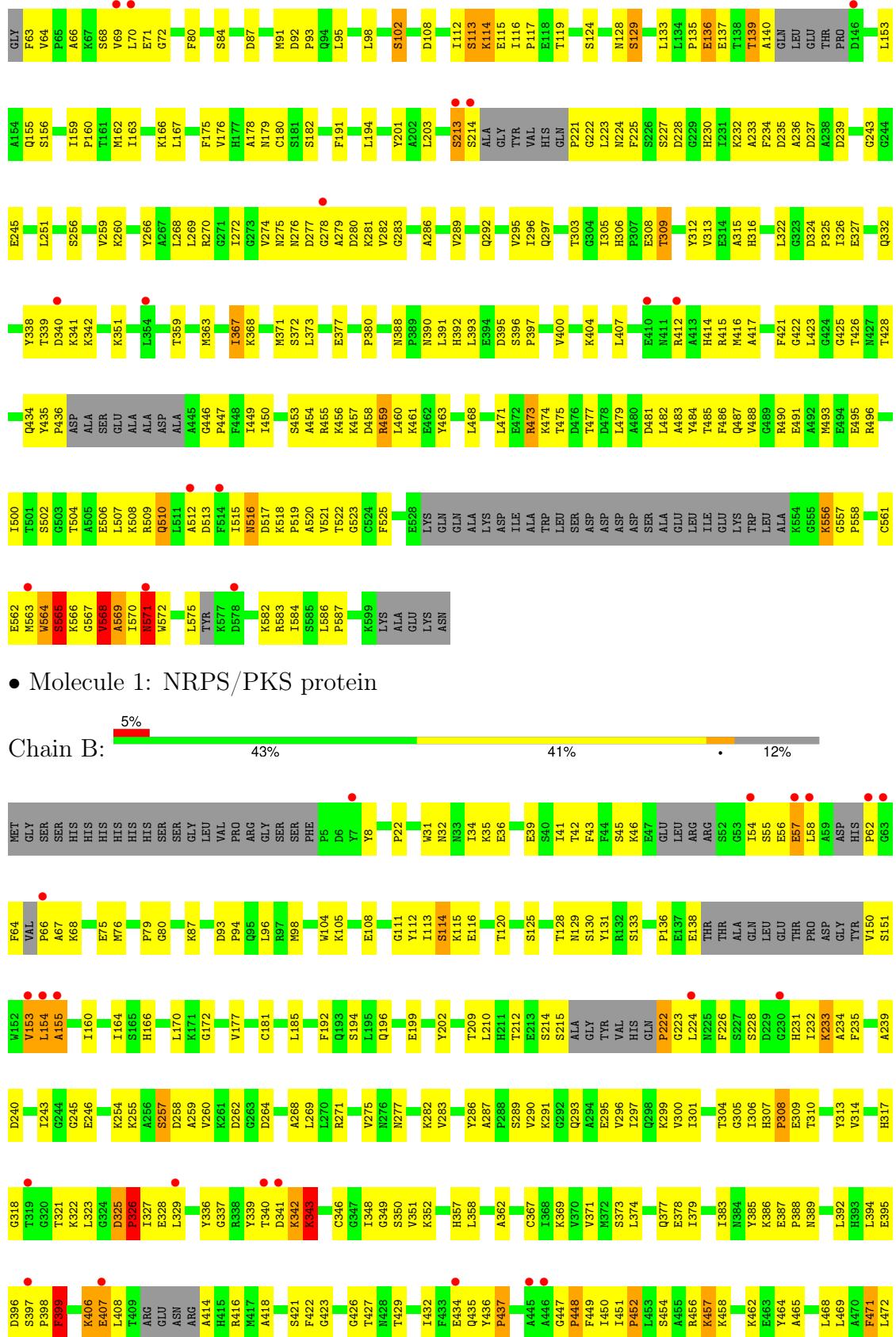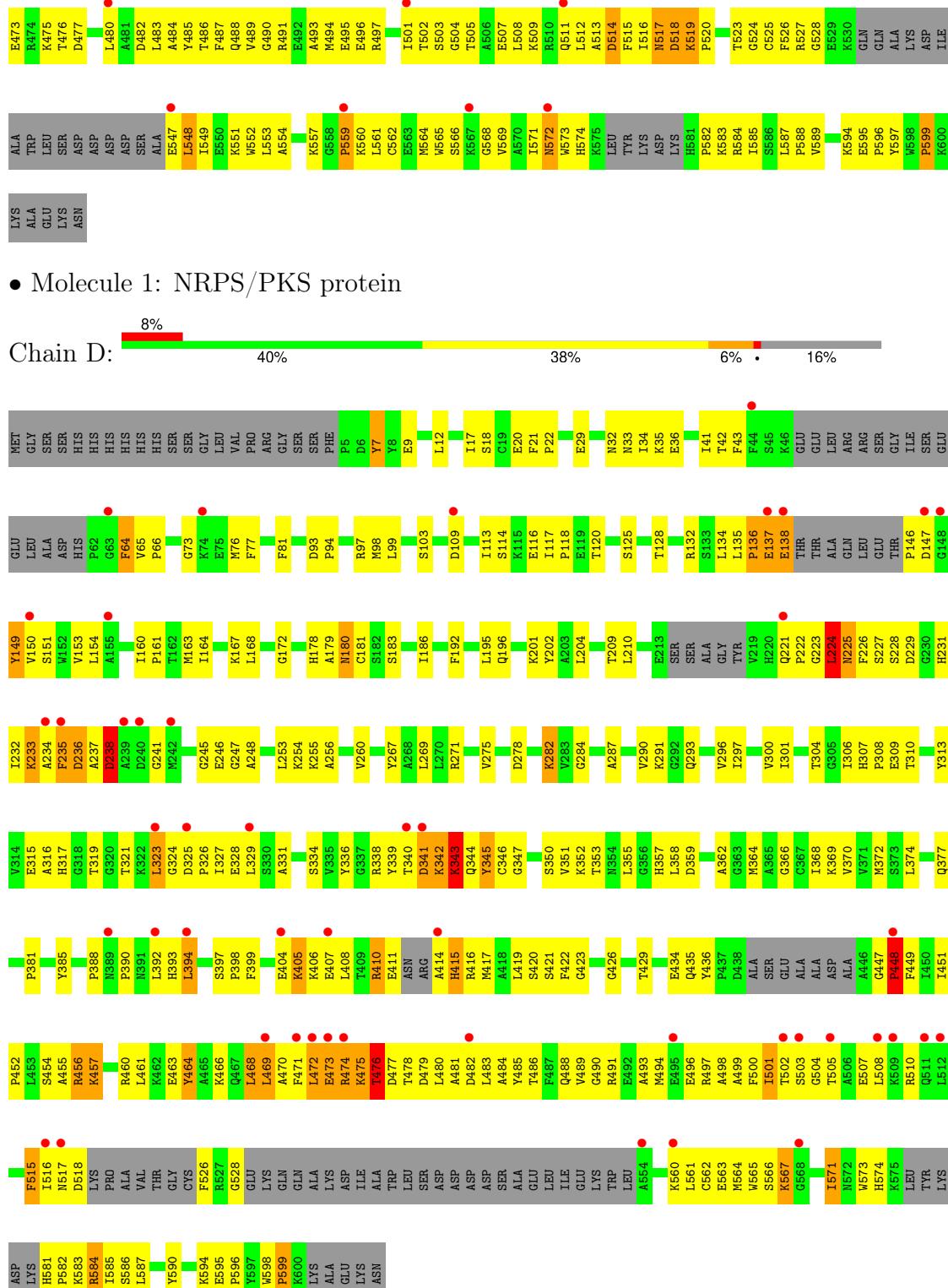
4   Data and refinement statistics   ⓘ

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, α, β, γ	61.62Å 99.95Å 100.84Å 91.93° 88.18° 96.04°	Depositor
Resolution (Å)	99.35 – 2.93 99.35 – 2.93	Depositor EDS
% Data completeness (in resolution range)	94.9 (99.35-2.93) 94.9 (99.35-2.93)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.80 (at 2.91Å)	Xtriage
Refinement program	REFMAC 5.8.0107	Depositor
R, R_{free}	0.270 , 0.303 0.269 , 0.298	Depositor DCC
R_{free} test set	2426 reflections (4.98%)	wwPDB-VP
Wilson B-factor (Å^2)	64.1	Xtriage
Anisotropy	0.098	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.32 , 58.3	EDS
L-test for twinning[2]	$< |L| > = 0.46, < L^2 > = 0.29$	Xtriage
Estimated twinning fraction	0.026 for -h,-k,l	Xtriage
F_o,F_c correlation	0.88	EDS
Total number of atoms	16749	wwPDB-VP
Average B, all atoms (Å^2)	61.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.16% 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)

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	0.63	1/4264 (0.0%)	0.97	17/5749 (0.3%)
1	B	0.56	1/4311 (0.0%)	0.86	15/5813 (0.3%)
1	C	0.68	3/4421 (0.1%)	1.09	58/5968 (1.0%)
1	D	0.62	6/4148 (0.1%)	1.01	35/5595 (0.6%)
All	All	0.62	11/17144 (0.1%)	0.98	125/23125 (0.5%)

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	0	3
1	B	0	3
1	C	0	4
1	D	0	1
All	All	0	11

The worst 5 of 11 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	C	308	PRO	N-CD	13.48	1.66	1.47
1	A	309	THR	CB-OG1	12.87	1.69	1.43
1	C	288	PRO	N-CD	10.09	1.61	1.47
1	D	414	ALA	CA-C	6.99	1.71	1.52
1	D	414	ALA	N-CA	-6.51	1.33	1.46

The worst 5 of 125 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	309	THR	CA-CB-OG1	24.69	160.85	109.00
1	A	565	SER	CB-CA-C	16.56	141.57	110.10
1	B	57	GLU	CB-CA-C	16.45	143.30	110.40

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	45	SER	C-N-CA	16.14	162.05	121.70
1	C	477	ASP	CB-CA-C	-15.74	78.93	110.40

There are no chirality outliers.

5 of 11 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	45	SER	Peptide
1	C	11	SER	Mainchain
1	C	546	ALA	Peptide
1	C	568	GLY	Peptide
1	C	8	TYR	Mainchain

5.2   Too-close contacts   ⓘ

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	4169	0	4098	248	1
1	B	4213	0	4124	331	6
1	C	4317	0	4213	330	4
1	D	4050	0	3958	391	3
All	All	16749	0	16393	1294	7

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 39.

The worst 5 of 1294 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:309:THR:OG1	1:A:309:THR:CB	1.69	1.37
1:D:308:PRO:O	1:D:345:TYR:OH	1.54	1.23
1:D:345:TYR:N	1:D:398:PRO:O	1.78	1.15
1:D:340:THR:HA	1:D:341:ASP:HB2	1.18	1.14
1:D:342:LYS:HD3	1:D:342:LYS:H	1.13	1.13

The worst 5 of 7 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:A:155:GLN:OE1	1:D:128:THR:OG1[1_556]	1.71	0.49
1:C:284:GLY:O	1:B:166:HIS:CD2[1_565]	1.78	0.42
1:B:477:ASP:OD2	1:D:223:GLY:O[1_655]	1.95	0.25
1:C:284:GLY:C	1:B:166:HIS:NE2[1_565]	2.03	0.17
1:C:284:GLY:O	1:B:166:HIS:CE1[1_565]	2.06	0.14

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	526/622 (85%)	483 (92%)	34 (6%)	9 (2%)	7	21
1	B	529/622 (85%)	473 (89%)	48 (9%)	8 (2%)	8	24
1	C	543/622 (87%)	485 (89%)	48 (9%)	10 (2%)	7	20
1	D	505/622 (81%)	461 (91%)	38 (8%)	6 (1%)	11	28
All	All	2103/2488 (84%)	1902 (90%)	168 (8%)	33 (2%)	8	22

5 of 33 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	C	226	PHE
1	C	288	PRO
1	C	554	ALA
1	A	447	PRO
1	A	571	ASN

5.3.2 Protein sidechains ⓘ

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	439/506 (87%)	417 (95%)	22 (5%)	20	42
1	B	442/506 (87%)	426 (96%)	16 (4%)	30	54
1	C	453/506 (90%)	427 (94%)	26 (6%)	17	37
1	D	425/506 (84%)	399 (94%)	26 (6%)	15	35
All	All	1759/2024 (87%)	1669 (95%)	90 (5%)	20	42

5 of 90 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	406	LYS
1	D	224	LEU
1	B	464	TYR
1	D	20	GLU
1	D	342	LYS

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 28 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	231	HIS
1	D	581	HIS
1	B	435	GLN
1	D	391	ASN
1	B	317	HIS

5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains ⓘ

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

5.5 Carbohydrates ⓘ

There are no oligosaccharides in this entry.

5.6 Ligand geometry ⓘ

There are no ligands in this entry.

5.7 Other polymers ⓘ

There are no such residues in this entry.

5.8 Polymer linkage issues ⓘ

There are no chain breaks in this entry.

6 Fit of model and data ⓘ

6.1 Protein, DNA and RNA chains ⓘ

In the following table, the column labelled '#RSRZ> 2' contains the number (and percentage) of RSRZ outliers, followed by percent RSRZ outliers for the chain as percentile scores relative to all X-ray entries and entries of similar resolution. The OWAB column contains the minimum, median, 95th percentile and maximum values of the occupancy-weighted average B-factor per residue. The column labelled 'Q< 0.9' lists the number of (and percentage) of residues with an average occupancy less than 0.9.

Mol	Chain	Analysed	<RSRZ>	#RSRZ>2			OWAB(Å^2)	Q<0.9
1	A	540/622 (86%)	0.20	17 (3%)	51	48	23, 48, 87, 109	0
1	B	547/622 (87%)	0.53	28 (5%)	34	31	24, 62, 100, 121	0
1	C	557/622 (89%)	0.57	55 (9%)	14	16	23, 55, 104, 162	0
1	D	523/622 (84%)	0.66	47 (8%)	17	18	24, 63, 108, 133	0
All	All	2167/2488 (87%)	0.49	147 (6%)	25	24	23, 57, 101, 162	0

The worst 5 of 147 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	341	ASP	7.2
1	D	469	LEU	5.2
1	D	341	ASP	4.9
1	C	548	LEU	4.6
1	C	523	THR	4.6

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

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

6.3 Carbohydrates ⓘ

There are no monosaccharides in this entry.

6.4 Ligands ⓘ

There are no ligands in this entry.

6.5 Other polymers ⓘ

There are no such residues in this entry.