wwPDB X-ray Structure Validation Summary Report ⓘ

Oct 21, 2024 – 06:31 PM EDT

PDB ID	:	9D2X
Title	:	SpiD ETS-domain (168-273) in complex with the DNA sequence d(AATAAAAGGAAGTGGG)
Authors	:	Terrell, J.R.; Vernon, T.N.; Poon, G.M.K.
Deposited on	:	2024-08-09
Resolution	:	2.29 Å(reported)

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

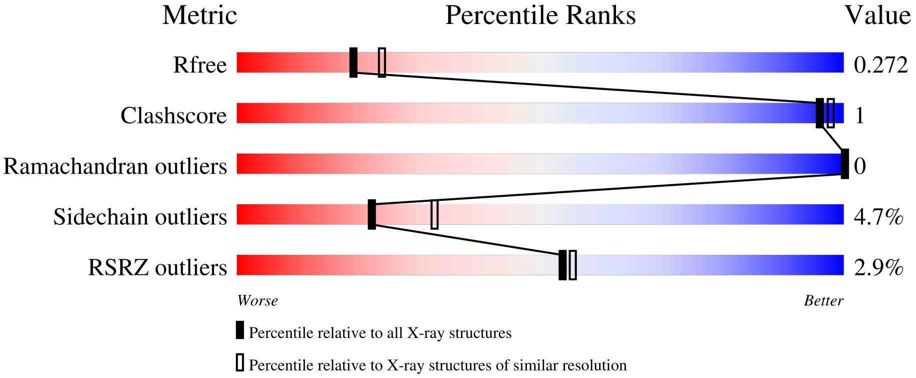
MolProbity	:	4.02b-467
Mogul	:	2022.3.0, CSD as543be (2022)
Xtriage (Phenix)	:	1.20.1
EDS	:	3.0
Percentile statistics	:	20231227.v01 (using entries in the PDB archive December 27th 2023)
CCP4	:	9.0.003 (Gargrove)
Density-Fitness	:	1.0.11
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.39

1 Overall quality at a glance ⓘ

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

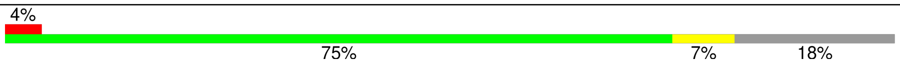
The reported resolution of this entry is 2.29 Å.

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))**
R_{free}	164625	5963 (2.30-2.30)
Clashscore	180529	6698 (2.30-2.30)
Ramachandran outliers	177936	6640 (2.30-2.30)
Sidechain outliers	177891	6640 (2.30-2.30)
RSRZ outliers	164620	5963 (2.30-2.30)

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	16	
1	C	16	
2	B	16	
2	D	16	
3	E	107	

Continued on next page...

Continued from previous page...

Mol	**Chain**	**Length**	**Quality of chain**
3	F	107	

2 Entry composition ⓘ

There are 5 unique types of molecules in this entry. The entry contains 5064 atoms, of which 2233 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 DNA chain called DNA (5'-D(*AP*AP*TP*AP*AP*AP*AP*GP*GP*AP*AP*GP*TP*GP*GP*G)-3').

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	H	N	O	P			
1	C	16	517	160	180	74	88	15	0	0	0
1	A	16	517	160	180	74	88	15	0	0	0

- Molecule 2 is a DNA chain called DNA (5'-D(*TP*CP*CP*CP*AP*CP*TP*TP*CP*CP*TP*TP*TP*TP*AP*T)-3').

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	H	N	O	P			
2	D	16	499	154	186	44	100	15	0	0	0
2	B	16	499	154	186	44	100	15	0	0	0

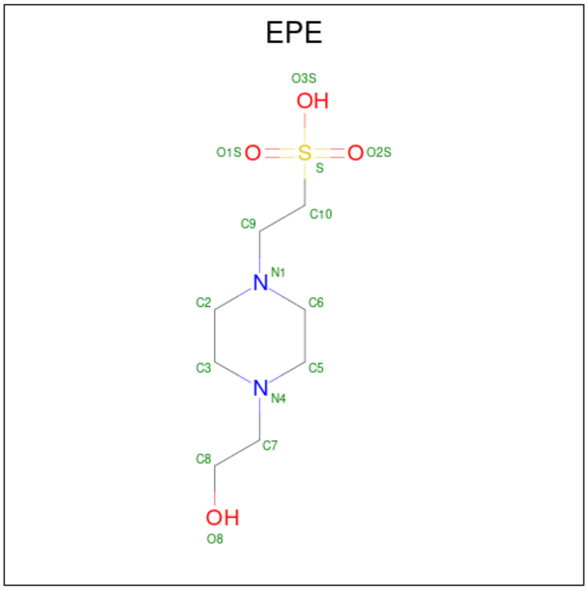
- Molecule 3 is a protein called SpiD.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	H	N	O	S			
3	F	88	1486	472	747	138	122	7	0	0	0
3	E	88	1470	469	737	135	122	7	0	0	0

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
F	164	MET	-	initiating methionine	UNP Q9DEW5
E	164	MET	-	initiating methionine	UNP Q9DEW5

- Molecule 4 is 4-(2-HYDROXYETHYL)-1-PIPERAZINE ETHANESULFONIC ACID (three-letter code: EPE) (formula: $C_8H_{18}N_2O_4S$).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	B	1	Total C H N O S 32 8 17 2 4 1	0	0

- Molecule 5 is water.

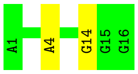
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	C	7	Total O 7 7	0	0
5	D	9	Total O 9 9	0	0
5	F	8	Total O 8 8	0	0
5	A	3	Total O 3 3	0	0
5	B	11	Total O 11 11	0	0
5	E	6	Total O 6 6	0	0

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.

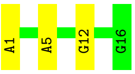
- Molecule 1: DNA (5'-D(*AP*AP*TP*AP*AP*AP*AP*GP*GP*AP*AP*GP*TP*GP*GP*G)-3')

Chain C:

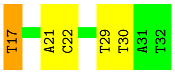
- Molecule 1: DNA (5'-D(*AP*AP*TP*AP*AP*AP*AP*GP*GP*AP*AP*GP*TP*GP*GP*G)-3')

Chain A:

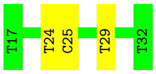
- Molecule 2: DNA (5'-D(*TP*CP*CP*CP*AP*CP*TP*TP*CP*CP*TP*TP*TP*TP*AP*T)-3 ')

Chain D:

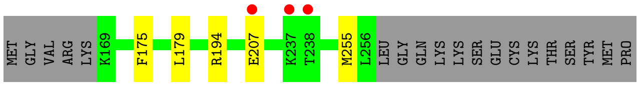
- Molecule 2: DNA (5'-D(*TP*CP*CP*CP*AP*CP*TP*TP*CP*CP*TP*TP*TP*TP*AP*T)-3 ')

Chain B:

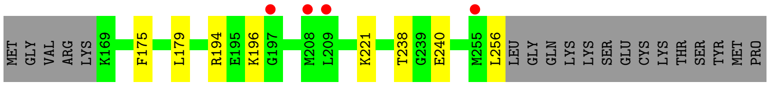
- Molecule 3: SpiD

Chain F:

- Molecule 3: SpiD

Chain E:



4 Data and refinement statistics ⓘ

Property	**Value**	**Source**
Space group	P 1	Depositor
Cell constants a, b, c, α, β, γ	36.93Å 42.94Å 62.53Å 79.21° 81.53° 76.19°	Depositor
Resolution (Å)	32.35 – 2.29 32.35 – 2.29	Depositor EDS
% Data completeness (in resolution range)	91.9 (32.35-2.29) 91.9 (32.35-2.29)	Depositor EDS
R_{merge}	0.04	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.22 (at 2.29Å)	Xtriage
Refinement program	PHENIX 1.19.2_4158	Depositor
R, R_{free}	0.221 , 0.270 0.222 , 0.272	Depositor DCC
R_{free} test set	14885 reflections (10.01%)	wwPDB-VP
Wilson B-factor (Å^2)	45.8	Xtriage
Anisotropy	0.354	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.37 , 37.8	EDS
L-test for twinning[2]	$< \vert L \vert > = 0.49$, $< L^2 > = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	5064	wwPDB-VP
Average B, all atoms (Å^2)	59.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 9.16% 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)

Bond lengths and bond angles in the following residue types are not validated in this section: EPE

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	1.35	2/382 (0.5%)	1.00	0/590
1	C	1.62	2/382 (0.5%)	1.09	1/590 (0.2%)
2	B	1.40	2/346 (0.6%)	1.21	0/530
2	D	1.51	6/346 (1.7%)	1.23	0/530
3	E	0.47	0/748	0.55	0/995
3	F	0.56	0/754	0.58	0/1002
All	All	1.10	12/2958 (0.4%)	0.91	1/4237 (0.0%)

The worst 5 of 12 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	B	24	DT	C5-C7	-8.28	1.45	1.50
1	C	14	DG	C2'-C1'	5.83	1.58	1.52
1	C	4	DA	N7-C5	-5.66	1.35	1.39
2	D	29	DT	C5-C7	-5.61	1.46	1.50
2	D	22	DC	C2-O2	-5.55	1.19	1.24

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	C	14	DG	O4'-C1'-N9	5.30	111.71	108.00

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	337	180	180	0	1
1	C	337	180	180	0	0
2	B	313	186	186	1	0
2	D	313	186	186	0	1
3	E	733	737	736	2	0
3	F	739	747	747	1	0
4	B	15	17	17	0	0
5	A	3	0	0	0	0
5	B	11	0	0	1	0
5	C	7	0	0	0	0
5	D	9	0	0	0	0
5	E	6	0	0	0	0
5	F	8	0	0	0	0
All	All	2831	2233	2232	4	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 1.

All (4) 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:B:25:DC:OP2	5:B:201:HOH:O	2.14	0.65
3:E:175:PHE:CZ	3:E:179:LEU:HD11	2.49	0.47
3:F:175:PHE:CZ	3:F:179:LEU:HD11	2.50	0.47
3:E:238:THR:OG1	3:E:240:GLU:OE1	2.22	0.46

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)
2:D:17:DT:O4	1:A:1:DA:H61[1_456]	1.57	0.03

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	
3	E	86/107 (80%)	83 (96%)	3 (4%)	0	100	100
3	F	86/107 (80%)	84 (98%)	2 (2%)	0	100	100
All	All	172/214 (80%)	167 (97%)	5 (3%)	0	100	100

There are no Ramachandran outliers to report.

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	
3	E	74/95 (78%)	70 (95%)	4 (5%)	18	27
3	F	75/95 (79%)	72 (96%)	3 (4%)	27	40
All	All	149/190 (78%)	142 (95%)	7 (5%)	22	32

5 of 7 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	E	194	ARG
3	E	196	LYS
3	E	256	LEU
3	E	221	LYS
3	F	255	MET

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

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 ⓘ

1 ligand 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	$\#\|Z\| > 2$
4	EPE	B	101	-	15,15,15	0.36	0	19,20,20	0.30	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
4	EPE	B	101	-	-	2/9/19/19	0/1/1/1

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (2) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
4	B	101	EPE	N4-C7-C8-O8
4	B	101	EPE	C10-C9-N1-C6

There are no ring outliers.

No monomer is involved in short contacts.

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	16/16 (100%)	0.05	0	100	100	50, 59, 76, 85	0
1	C	16/16 (100%)	-0.24	0	100	100	43, 53, 68, 74	0
2	B	16/16 (100%)	-0.09	0	100	100	40, 64, 71, 77	0
2	D	16/16 (100%)	-0.39	0	100	100	37, 54, 65, 70	0
3	E	88/107 (82%)	0.71	4 (4%)	39	40	37, 66, 94, 99	0
3	F	88/107 (82%)	0.41	3 (3%)	48	50	36, 53, 80, 109	0
All	All	240/278 (86%)	0.36	7 (2%)	54	55	36, 59, 89, 109	0

The worst 5 of 7 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
3	F	207	GLU	2.7
3	F	238	THR	2.3
3	E	209	LEU	2.2
3	E	208	MET	2.2
3	E	255	MET	2.2

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 ⓘ

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95th percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å^2)	Q<0.9
4	EPE	B	101	15/15	0.88	0.10	53,66,71,79	0

6.5 Other polymers ⓘ

There are no such residues in this entry.