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)

Dec 2, 2023 – 01:42 pm GMT

PDB ID	:	1OID
Title	:	5'-Nucleotidase (E. coli) with an Engineered Disulfide Bridge (S228C, P513C)
Authors	:	Schultz-Heienbrok, R.; Maier, T.; Straeter, N.
Deposited on	:	2003-06-13
Resolution	:	2.10 Å(reported)

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

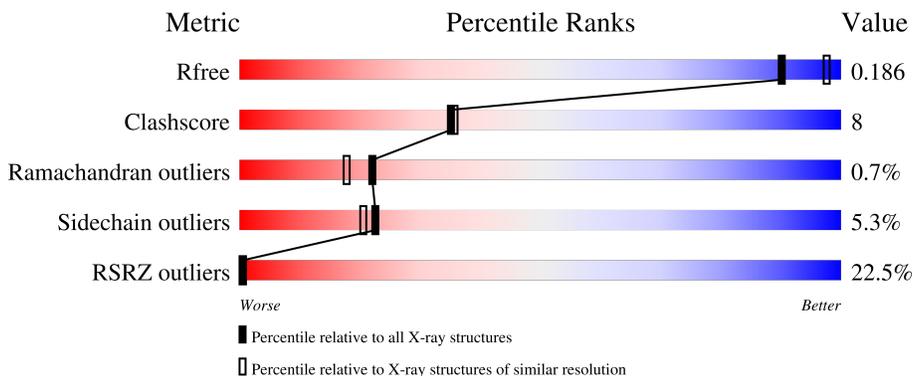|		
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36

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 2.10 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	Percentile Ranks	Value
Rfree		0.186
Clashscore		8
Ramachandran outliers		0.7%
Sidechain outliers		5.3%
RSRZ outliers		22.5%

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$_{free}$	130704	5197 (2.10-2.10)
Clashscore	141614	5710 (2.10-2.10)
Ramachandran outliers	138981	5647 (2.10-2.10)
Sidechain outliers	138945	5648 (2.10-2.10)
RSRZ outliers	127900	5083 (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% 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	532	14% / 77% / 19%
1	B	532	31% / 76% / 19%

2 Entry composition ⓘ

There are 3 unique types of molecules in this entry. The entry contains 8920 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 PROTEIN USHA.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	525	Total	C	N	O	S	0	0	0
			4098	2588	703	788	19			
1	B	523	Total	C	N	O	S	0	0	0
			4075	2572	699	785	19			

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	228	CYS	SER	engineered mutation	UNP P07024
A	513	CYS	PRO	engineered mutation	UNP P07024
B	228	CYS	SER	engineered mutation	UNP P07024
B	513	CYS	PRO	engineered mutation	UNP P07024

- Molecule 2 is NICKEL (II) ION (three-letter code: NI) (formula: Ni).
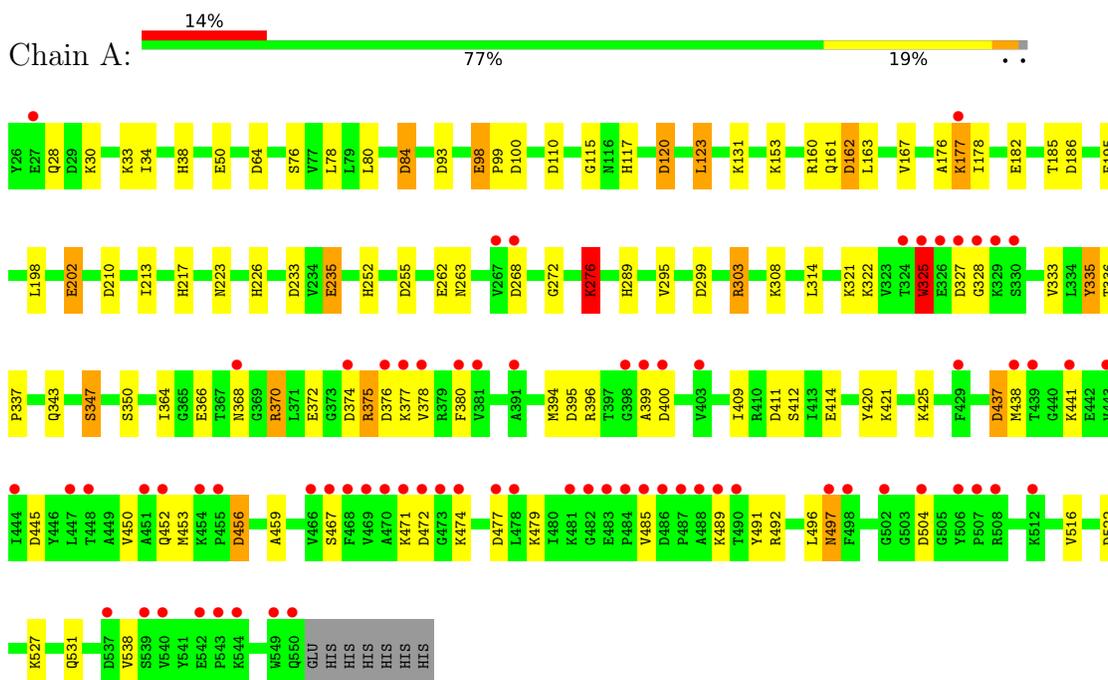
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	1	Total	Ni	0	0
			1	1		

- Molecule 3 is water.

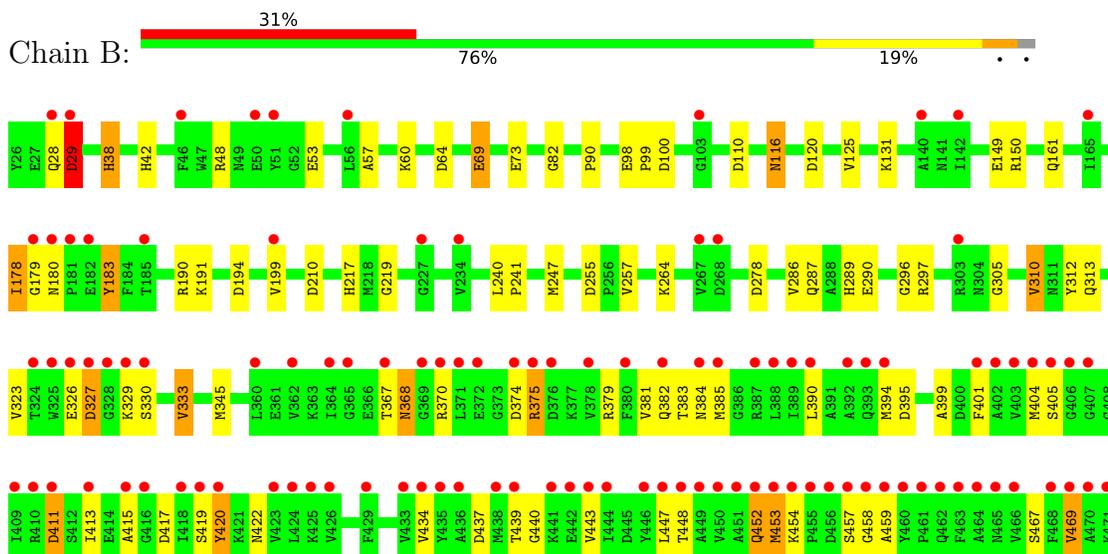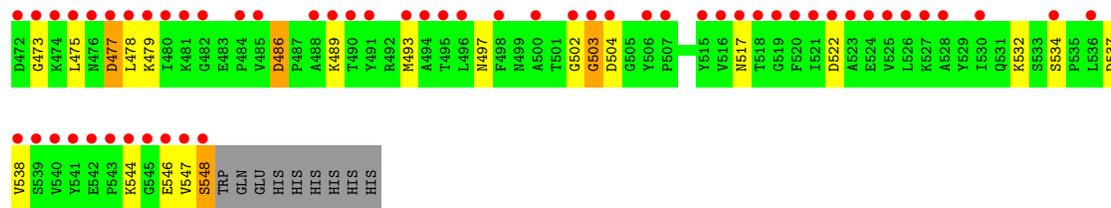| Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	408	Total	O	0	0
			408	408		
3	B	338	Total	O	0	0
			338	338		

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: PROTEIN USHA

- Molecule 1: PROTEIN USHA

D472 G473 K474 L475 N476 D477 L478 K479 I480 K481 G482 E483 P484 V485 D486 P487 A488 K489 T490 Y491 R492 M493 A494 T495 L496 N497 F498 N499 A500 T501 G502 G503 D504 G505 Y506 P507 Y515 V516 N517 T518 G519 F520 I521 D522 A523 E524 V525 L526 K527 A528 Y529 I530 Q531 K532 S533 S534 P535 L536 D537

V538 S539 V540 Y541 E542 P543 K544 G545 E546 V547 S548 TRP GLN GLU HIS HIS HIS HIS HIS HIS

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	80.50Å 93.71Å 82.90Å 90.00° 97.71° 90.00°	Depositor
Resolution (Å)	30.00 – 2.10 19.94 – 2.10	Depositor EDS
% Data completeness (in resolution range)	98.3 (30.00-2.10) 98.4 (19.94-2.10)	Depositor EDS
R_{merge}	0.06	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.49 (at 2.09Å)	Xtriage
Refinement program	REFMAC 5.1.19	Depositor
R, R_{free}	0.166 , 0.215 0.193 , 0.186	Depositor DCC
R_{free} test set	3530 reflections (5.04%)	wwPDB-VP
Wilson B-factor (Å^2)	32.3	Xtriage
Anisotropy	0.061	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.33 , 66.4	EDS
L-test for twinning[2]	$< |L| > = 0.49, < L^2 > = 0.32$	Xtriage
Estimated twinning fraction	0.019 for l,-k,h	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	8920	wwPDB-VP
Average B, all atoms (Å^2)	42.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.29% 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 ⓘ

5.1 Standard geometry ⓘ

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

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	1.44	15/4185 (0.4%)	1.24	39/5662 (0.7%)
1	B	1.28	17/4160 (0.4%)	1.16	27/5627 (0.5%)
All	All	1.36	32/8345 (0.4%)	1.20	66/11289 (0.6%)

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	1

The worst 5 of 32 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	333	VAL	CB-CG2	-10.24	1.31	1.52
1	B	333	VAL	CB-CG2	-9.18	1.33	1.52
1	A	347	SER	CB-OG	-8.95	1.30	1.42
1	A	235	GLU	CD-OE2	-8.45	1.16	1.25
1	B	69	GLU	CD-OE1	8.01	1.34	1.25

The worst 5 of 66 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	235	GLU	OE1-CD-OE2	-13.50	107.11	123.30
1	B	48	ARG	NE-CZ-NH1	13.28	126.94	120.30
1	B	64	ASP	CB-CG-OD1	11.16	128.34	118.30
1	B	210	ASP	CB-CG-OD2	10.76	127.99	118.30
1	A	162	ASP	CB-CG-OD2	10.71	127.94	118.30

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	437	ASP	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	4098	0	4015	58	1
1	B	4075	0	3997	74	0
2	A	1	0	0	0	0
3	A	408	0	0	13	1
3	B	338	0	0	9	0
All	All	8920	0	8012	132	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 8.

The worst 5 of 132 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:B:38:HIS:CE1	1:B:296:GLY:HA3	1.80	1.15
1:B:401:PHE:CE2	1:B:493:MET:HG3	1.83	1.14
1:B:73:GLU:HG2	3:B:2058:HOH:O	1.59	1.02
1:A:84:ASP:CG	1:A:217:HIS:HE1	1.67	0.98
1:B:217:HIS:HE1	3:B:2099:HOH:O	1.51	0.94

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)
1:A:414:GLU:OE2	3:A:2286:HOH:O[2_945]	1.99	0.21

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	523/532 (98%)	495 (95%)	25 (5%)	3 (1%)	25	21
1	B	521/532 (98%)	496 (95%)	21 (4%)	4 (1%)	19	15
All	All	1044/1064 (98%)	991 (95%)	46 (4%)	7 (1%)	22	18

5 of 7 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	328	GLY
1	B	183	TYR
1	B	503	GLY
1	A	289	HIS
1	B	29	ASP

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	433/440 (98%)	409 (94%)	24 (6%)	21	19
1	B	431/440 (98%)	409 (95%)	22 (5%)	24	22
All	All	864/880 (98%)	818 (95%)	46 (5%)	22	20

5 of 46 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	310	VAL

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	B	383	THR
1	B	329	LYS
1	B	368	ASN
1	B	420	TYR

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 17 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	422	ASN
1	B	452	GLN
1	A	531	GLN
1	B	38	HIS
1	B	42	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 monosaccharides in this entry.

5.6   Ligand geometry ⓘ

Of 1 ligands modelled in this entry, 1 is monoatomic - leaving 0 for Mogul analysis.

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 ⓘ

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	525/532 (98%)	0.53	72 (13%)	3	4	18, 32, 63, 82	0
1	B	523/532 (98%)	1.51	164 (31%)	0	0	22, 38, 80, 94	0
All	All	1048/1064 (98%)	1.02	236 (22%)	0	0	18, 35, 75, 94	0

The worst 5 of 236 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	401	PHE	9.7
1	B	463	PHE	9.3
1	B	454	LYS	9.0
1	B	491	TYR	8.9
1	B	389	ILE	8.7

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
2	NI	A	1551	1/1	0.97	0.11	66,66,66,66	0

6.5 Other polymers ⓘ

There are no such residues in this entry.

WORLDWIDE
PDB
PROTEIN DATA BANK