



# Full wwPDB X-ray Structure Validation Report ⓘ

Sep 28, 2024 – 11:54 AM EDT

PDB ID : 4HQS  
Title : Crystal structure of the pneumococcal exposed lipoprotein thioredoxin sp\_0659 (Etrx1) from Streptococcus pneumoniae strain TIGR4  
Authors : Bartual, S.G.; Saleh, M.; Abdullah, M.R.; Jensch, I.; Asmat, T.M.; Petruschka, L.; Pribyl, T.; Hermoso, J.A.; Hammerschmidt, S.  
Deposited on : 2012-10-26  
Resolution : 1.48 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](mailto: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 ⓘ 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 ⓘ](#)) 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

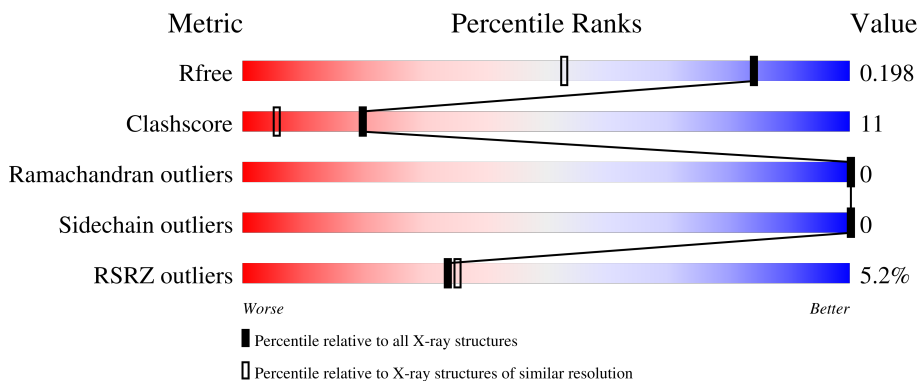
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 1.48 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	6131 (1.50-1.46)
Clashscore	180529	6623 (1.50-1.46)
Ramachandran outliers	177936	6521 (1.50-1.46)
Sidechain outliers	177891	6518 (1.50-1.46)
RSRZ outliers	164620	6132 (1.50-1.46)

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  $\geq 3$ , 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  $\leq 5\%$ . 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	190	

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
2	GOL	A	201	-	-	X	-

*Continued on next page...*

*Continued from previous page...*

<b>Mol</b>	<b>Type</b>	<b>Chain</b>	<b>Res</b>	<b>Chirality</b>	<b>Geometry</b>	<b>Clashes</b>	<b>Electron density</b>
2	GOL	A	214	-	-	X	-

## 2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 2437 atoms, of which 1103 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 Thioredoxin family protein.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	H	N	O	S			
1	A	135	2135	694	1063	169	205	4	0	0	0

There are 22 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-1	MET	-	expression tag	UNP Q97RX4
A	0	GLY	-	expression tag	UNP Q97RX4
A	1	SER	-	expression tag	UNP Q97RX4
A	2	SER	-	expression tag	UNP Q97RX4
A	3	HIS	-	expression tag	UNP Q97RX4
A	4	HIS	-	expression tag	UNP Q97RX4
A	5	HIS	-	expression tag	UNP Q97RX4
A	6	HIS	-	expression tag	UNP Q97RX4
A	7	HIS	-	expression tag	UNP Q97RX4
A	8	HIS	-	expression tag	UNP Q97RX4
A	9	MET	-	expression tag	UNP Q97RX4
A	10	SER	-	expression tag	UNP Q97RX4
A	11	GLY	-	expression tag	UNP Q97RX4
A	12	GLU	-	expression tag	UNP Q97RX4
A	13	ASN	-	expression tag	UNP Q97RX4
A	14	LEU	-	expression tag	UNP Q97RX4
A	15	TYR	-	expression tag	UNP Q97RX4
A	16	PHE	-	expression tag	UNP Q97RX4
A	17	GLN	-	expression tag	UNP Q97RX4
A	18	GLY	-	expression tag	UNP Q97RX4
A	19	ALA	-	expression tag	UNP Q97RX4
A	20	SER	-	expression tag	UNP Q97RX4

- Molecule 2 is GLYCEROL (three-letter code: GOL) (formula: C<sub>3</sub>H<sub>8</sub>O<sub>3</sub>).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
2	A	1	Total	C	H	O	0	0
			14	3	8	3		
2	A	1	Total	C	H	O	0	0
			14	3	8	3		
2	A	1	Total	C	H	O	0	0
			14	3	8	3		
2	A	1	Total	C	H	O	0	0
			14	3	8	3		
2	A	1	Total	C	H	O	0	0
			14	3	8	3		

- Molecule 3 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	9	Total	Mg	0	0
			9	9		

- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	223	Total	O	0	0
			223	223		



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 43 21 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	62.85Å 62.85Å 89.60Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	28.11 – 1.48 28.11 – 1.48	Depositor EDS
% Data completeness (in resolution range)	99.8 (28.11-1.48) 98.9 (28.11-1.48)	Depositor EDS
$R_{merge}$	0.09	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	5.97 (at 1.48Å)	Xtrriage
Refinement program	PHENIX (phenix.refine: 1.7.3_928)	Depositor
R, $R_{free}$	0.182 , 0.197 0.181 , 0.198	Depositor DCC
$R_{free}$ test set	1544 reflections (5.04%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	15.8	Xtrriage
Anisotropy	0.013	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.39 , 38.5	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.47$ , $\langle L^2 \rangle = 0.30$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
$F_o, F_c$ correlation	0.96	EDS
Total number of atoms	2437	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	21.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 8.32% of the height of the origin peak. No significant pseudotranslation is detected.*

<sup>1</sup>Intensities estimated from amplitudes.

<sup>2</sup>Theoretical values of  $\langle |L| \rangle$ ,  $\langle L^2 \rangle$  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: GOL, MG

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	0.55	0/1098	0.69	1/1484 (0.1%)

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	175	MET	CG-SD-CE	-7.42	88.32	100.20

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	1072	1063	1061	25	4
2	A	30	40	40	16	0
3	A	9	0	0	0	0
4	A	223	0	0	3	4
All	All	1334	1103	1101	25	4

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 11.

All (25) 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:145:LYS:O	1:A:148:GLU:HG2	1.78	0.83
1:A:186:GLU:O	2:A:214:GOL:O2	2.13	0.67
1:A:186:GLU:HB3	2:A:214:GOL:H12	1.79	0.65
1:A:77:LEU:HD11	1:A:110:VAL:HG23	1.82	0.61
1:A:77:LEU:HD11	1:A:110:VAL:CG2	2.32	0.59
1:A:134:LYS:HZ3	2:A:203:GOL:H11	1.69	0.56
1:A:187:LEU:C	2:A:201:GOL:O2	2.45	0.54
1:A:153:ARG:HG2	4:A:363:HOH:O	2.06	0.54
1:A:145:LYS:O	1:A:148:GLU:CG	2.54	0.53
1:A:66:ARG:NH1	4:A:494:HOH:O	2.44	0.51
1:A:134:LYS:HZ3	2:A:203:GOL:C1	2.24	0.50
1:A:99:LYS:CE	2:A:204:GOL:O2	2.61	0.49
1:A:134:LYS:NZ	2:A:203:GOL:H11	2.28	0.48
1:A:163:LYS:CE	2:A:201:GOL:H31	2.44	0.48
1:A:187:LEU:C	2:A:201:GOL:H2	2.37	0.45
1:A:163:LYS:HG3	2:A:201:GOL:H11	1.99	0.45
1:A:153:ARG:NH2	4:A:311:HOH:O	2.42	0.45
1:A:86:ILE:HG13	2:A:202:GOL:O3	2.18	0.44
1:A:186:GLU:CA	2:A:214:GOL:O2	2.66	0.43
1:A:99:LYS:NZ	2:A:204:GOL:O2	2.49	0.43
1:A:134:LYS:HB2	2:A:204:GOL:H32	2.01	0.43
1:A:75:VAL:HB	1:A:161:ILE:HB	2.01	0.43
1:A:186:GLU:HA	2:A:214:GOL:O2	2.20	0.41
1:A:73:LYS:HD3	1:A:106:VAL:HG23	2.04	0.40
1:A:163:LYS:NZ	2:A:201:GOL:H31	2.37	0.40

All (4) 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:66:ARG:NH1	4:A:340:HOH:O[3_554]	1.41	0.79
1:A:66:ARG:HH12	4:A:340:HOH:O[3_554]	1.14	0.46
1:A:66:ARG:CZ	4:A:340:HOH:O[3_554]	1.78	0.42
1:A:66:ARG:NH2	4:A:340:HOH:O[3_554]	1.96	0.24

## 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	A	133/190 (70%)	133 (100%)	0	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	A	116/161 (72%)	116 (100%)	0	100 100

There are no protein residues with a non-rotameric sidechain to report.

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

### 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 oligosaccharides in this entry.

## 5.6 Ligand geometry [i](#)

Of 14 ligands modelled in this entry, 9 are monoatomic - leaving 5 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
2	GOL	A	203	-	5,5,5	0.47	0	5,5,5	0.31	0
2	GOL	A	201	-	5,5,5	0.33	0	5,5,5	0.63	0
2	GOL	A	202	-	5,5,5	0.52	0	5,5,5	0.30	0
2	GOL	A	204	-	5,5,5	0.46	0	5,5,5	0.26	0
2	GOL	A	214	-	5,5,5	0.60	0	5,5,5	0.56	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
2	GOL	A	203	-	-	2/4/4/4	-
2	GOL	A	201	-	-	0/4/4/4	-
2	GOL	A	202	-	-	0/4/4/4	-
2	GOL	A	204	-	-	4/4/4/4	-
2	GOL	A	214	-	-	4/4/4/4	-

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (10) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	A	204	GOL	O1-C1-C2-O2
2	A	204	GOL	O1-C1-C2-C3
2	A	204	GOL	C1-C2-C3-O3
2	A	214	GOL	O1-C1-C2-C3
2	A	214	GOL	C1-C2-C3-O3
2	A	203	GOL	O1-C1-C2-O2
2	A	214	GOL	O1-C1-C2-O2
2	A	203	GOL	O1-C1-C2-C3
2	A	204	GOL	O2-C2-C3-O3
2	A	214	GOL	O2-C2-C3-O3

There are no ring outliers.

5 monomers are involved in 16 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	A	203	GOL	3	0
2	A	201	GOL	5	0
2	A	202	GOL	1	0
2	A	204	GOL	3	0
2	A	214	GOL	4	0

## 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](#)

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, 95<sup>th</sup> 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 <sup>2</sup> )	Q<0.9
1	A	135/190 (71%)	-0.02	7 (5%) 34 35	9, 17, 37, 52	0

All (7) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	53	ALA	5.6
1	A	148	GLU	3.7
1	A	66	ARG	3.2
1	A	153	ARG	3.0
1	A	104	ASP	2.9
1	A	165	GLY	2.5
1	A	166	LYS	2.4

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

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

### 6.3 Carbohydrates [i](#)

There are no monosaccharides in this entry.

### 6.4 Ligands [i](#)

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, 95<sup>th</sup> 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 <sup>2</sup> )	Q<0.9
2	GOL	A	201	6/6	0.56	0.26	39,49,58,59	0

*Continued on next page...*

*Continued from previous page...*

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors( $\text{\AA}^2$ )	Q<0.9
2	GOL	A	214	6/6	0.82	0.16	15,28,49,54	0
2	GOL	A	202	6/6	0.84	0.14	28,38,48,48	0
2	GOL	A	204	6/6	0.87	0.13	25,40,60,62	0
2	GOL	A	203	6/6	0.89	0.12	22,38,48,49	0
3	MG	A	212	1/1	0.89	0.14	33,33,33,33	0
3	MG	A	206	1/1	0.91	0.10	24,24,24,24	0
3	MG	A	211	1/1	0.94	0.10	29,29,29,29	0
3	MG	A	213	1/1	0.94	0.08	26,26,26,26	0
3	MG	A	209	1/1	0.97	0.09	21,21,21,21	0
3	MG	A	210	1/1	0.99	0.03	20,20,20,20	0
3	MG	A	207	1/1	0.99	0.05	23,23,23,23	0
3	MG	A	208	1/1	0.99	0.06	18,18,18,18	0
3	MG	A	205	1/1	0.99	0.06	17,17,17,17	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.