



Full wwPDB X-ray Structure Validation Report ⓘ

May 21, 2020 – 11:09 pm BST

PDB ID : 2I4E
Title : Structural studies of protein tyrosine phosphatase beta catalytic domain in complex with inhibitors
Authors : Evdokimov, A.G.; Pokross, M.E.; Walter, R.L.; Mekel, M.
Deposited on : 2006-08-21
Resolution : 1.75 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 ⓘ symbol.

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

MolProbity : 4.02b-467
Mogul : 1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix) : 1.13
EDS : 2.11
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) : 2.11

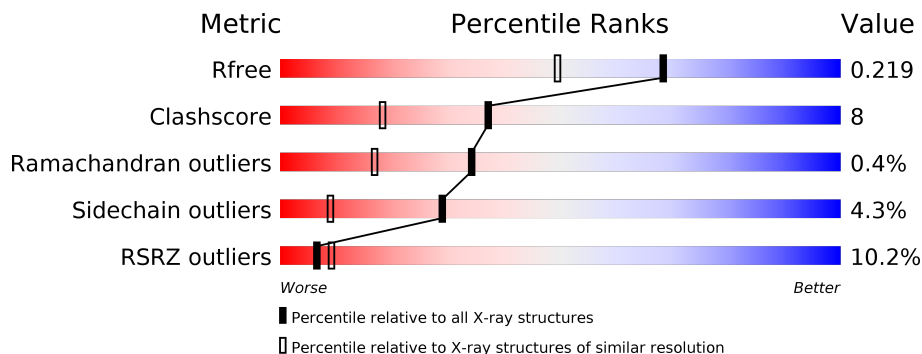
1 Overall quality at a glance i

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 1.75 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}	130704	2340 (1.76-1.76)
Clashscore	141614	2466 (1.76-1.76)
Ramachandran outliers	138981	2437 (1.76-1.76)
Sidechain outliers	138945	2437 (1.76-1.76)
RSRZ outliers	127900	2298 (1.76-1.76)

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 on 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 $\leq 5\%$. 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	313	 9% 76% 12% • 10%
1	B	313	 10% 73% 14% • 11%

2 Entry composition [i](#)

There are 3 unique types of molecules in this entry. The entry contains 4919 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.

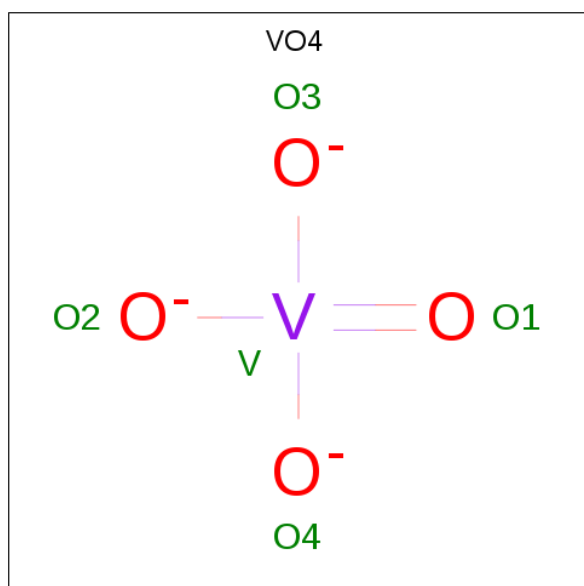
- Molecule 1 is a protein called Receptor-type tyrosine-protein phosphatase beta.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	281	2323	1469	413	425	16	0	6	0
1	B	280	2310	1462	415	418	15	0	4	0

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1661	SER	-	CLONING ARTIFACT	UNP P23467
B	1661	SER	-	CLONING ARTIFACT	UNP P23467

- Molecule 2 is VANADATE ION (three-letter code: VO4) (formula: O₄V).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
			Total	O	V		
2	A	1	5	4	1	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
2	B	1	Total	O	V	0	0
			5	4	1		

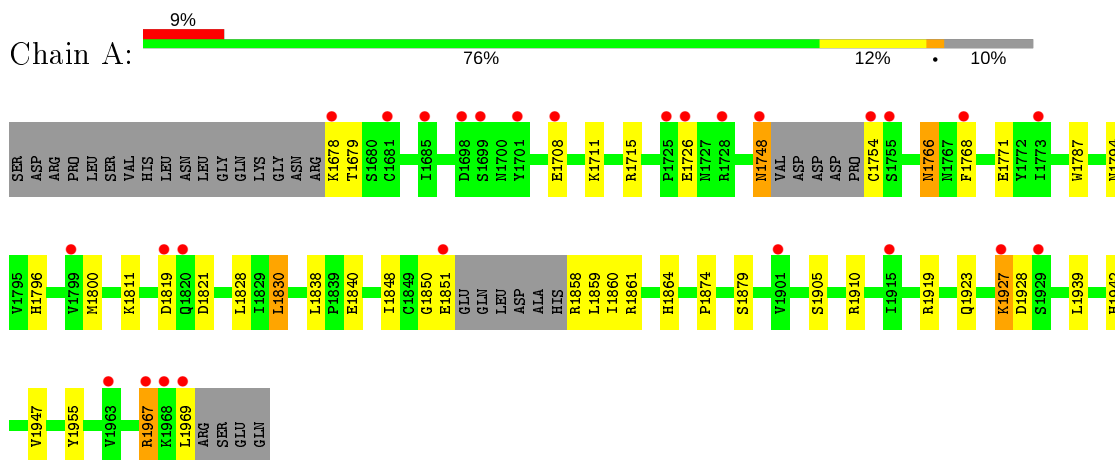
- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	165	Total	O	0	0
			165	165		
3	B	111	Total	O	0	0
			111	111		

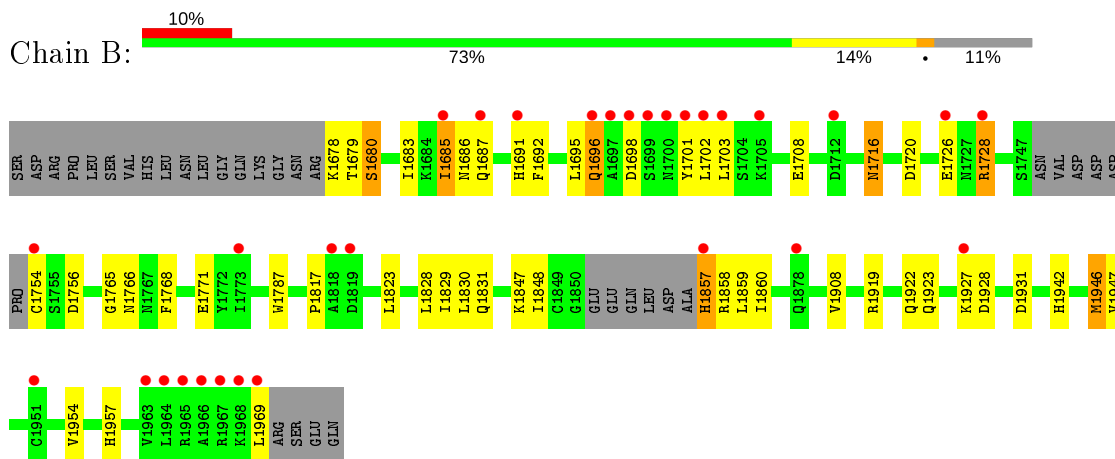
3 Residue-property plots [i](#)

These plots are drawn for all protein, RNA and DNA 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: Receptor-type tyrosine-protein phosphatase beta



- Molecule 1: Receptor-type tyrosine-protein phosphatase beta



4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α , β , γ	62.06Å 71.72Å 69.86Å 90.00° 93.31° 90.00°	Depositor
Resolution (Å)	38.10 – 1.75 38.15 – 1.75	Depositor EDS
% Data completeness (in resolution range)	97.1 (38.10-1.75) 97.1 (38.15-1.75)	Depositor EDS
R_{merge}	0.03	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	5.29 (at 1.75Å)	Xtrriage
Refinement program	REFMAC	Depositor
R, R_{free}	0.183 , 0.221 0.181 , 0.219	Depositor DCC
R_{free} test set	3035 reflections (5.08%)	wwPDB-VP
Wilson B-factor (Å ²)	24.9	Xtrriage
Anisotropy	0.374	Xtrriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.43 , 62.6	EDS
L-test for twinning ²	$\langle L \rangle = 0.49$, $\langle L^2 \rangle = 0.33$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
F_o, F_c correlation	0.96	EDS
Total number of atoms	4919	wwPDB-VP
Average B, all atoms (Å ²)	29.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.41% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

²Theoretical 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: VO4

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	0/2391	0.89	3/3240 (0.1%)
1	B	0.52	0/2373	0.83	2/3214 (0.1%)
All	All	0.58	0/4764	0.86	5/6454 (0.1%)

There are no bond length outliers.

All (5) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	1946	MET	CG-SD-CE	-14.01	77.78	100.20
1	A	1919	ARG	NE-CZ-NH1	8.01	124.30	120.30
1	A	1830	LEU	CA-CB-CG	5.83	128.70	115.30
1	B	1756	ASP	CB-CG-OD1	5.37	123.13	118.30
1	A	1910	ARG	NE-CZ-NH2	-5.25	117.68	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	A	2323	0	2285	35	0
1	B	2310	0	2276	37	0
2	A	5	0	0	0	0
2	B	5	0	0	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
3	A	165	0	0	10	0
3	B	111	0	0	6	0
All	All	4919	0	4561	72	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 8.

All (72) 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:1726[A]:GLU:CD	1:A:1754:CYS:N	2.15	1.00
1:B:1828:LEU:HD13	1:B:1848:ILE:HD11	1.48	0.95
1:A:1726[A]:GLU:OE2	1:A:1754:CYS:N	2.13	0.81
1:A:1754:CYS:HB2	3:A:56:HOH:O	1.82	0.78
1:B:1720:ASP:OD2	3:B:95:HOH:O	2.02	0.77
1:A:1850:GLY:O	1:A:1851:GLU:HB2	1.86	0.76
1:A:1711:LYS:CB	3:A:157:HOH:O	2.33	0.75
1:B:1728:ARG:HD2	3:B:283:HOH:O	1.89	0.72
1:B:1716:ASN:H	1:B:1716:ASN:HD22	1.39	0.70
1:A:1726[A]:GLU:OE1	1:A:1754:CYS:N	2.25	0.70
1:B:1768:PHE:CE1	1:B:1771:GLU:HA	2.28	0.69
1:B:1923:GLN:HE22	1:B:1931:ASP:H	1.40	0.68
1:B:1766:ASN:OD1	1:B:1919[A]:ARG:HD3	1.95	0.67
1:A:1708:GLU:O	1:A:1711:LYS:HG2	1.95	0.66
1:A:1711:LYS:HB3	3:A:157:HOH:O	1.96	0.64
1:B:1701:TYR:CE2	1:B:1702:LEU:HD23	2.34	0.63
1:B:1716:ASN:ND2	1:B:1716:ASN:H	1.96	0.63
1:B:1754:CYS:HA	3:B:48:HOH:O	1.99	0.62
1:A:1768:PHE:CE1	1:A:1771:GLU:HA	2.35	0.61
1:A:1828:LEU:HD13	1:A:1848:ILE:HD11	1.84	0.60
1:B:1687:GLN:HG3	1:B:1687:GLN:O	2.00	0.60
1:B:1695:LEU:HD23	1:B:1703:LEU:HD13	1.85	0.58
1:B:1923:GLN:NE2	1:B:1931:ASP:H	2.02	0.57
1:B:1680:SER:HB2	1:B:1931:ASP:OD2	2.06	0.56
1:B:1692:PHE:O	1:B:1696:GLN:HG3	2.05	0.56
1:A:1796:HIS:CD2	1:A:1858:ARG:HD3	2.41	0.55
1:A:1787:TRP:CZ2	1:A:1830:LEU:HD21	2.42	0.55
1:B:1787:TRP:CZ2	1:B:1830:LEU:HD21	2.42	0.55
1:A:1923:GLN:O	1:A:1927:LYS:HG3	2.08	0.54
1:B:1848:ILE:O	1:B:1857:HIS:HA	2.09	0.53
1:B:1766:ASN:ND2	1:B:1922:GLN:HB2	2.23	0.53

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:B:1823:LEU:C	1:B:1829[A]:ILE:HD12	2.29	0.53
1:A:1923:GLN:HA	1:A:1927:LYS:HG3	1.92	0.52
1:B:1942:HIS:HD2	3:B:221:HOH:O	1.92	0.51
1:B:1857:HIS:N	3:B:174:HOH:O	2.43	0.51
1:A:1942:HIS:HD2	3:A:220:HOH:O	1.94	0.50
1:B:1703:LEU:CD2	1:B:1954:VAL:HG22	2.43	0.49
1:A:1708:GLU:HG2	3:A:219:HOH:O	2.15	0.46
1:A:1715:ARG:HA	3:A:182:HOH:O	2.15	0.46
1:B:1685:ILE:HG22	1:B:1686:ASN:N	2.31	0.46
1:A:1711:LYS:HB2	3:A:157:HOH:O	2.07	0.46
1:B:1683:ILE:HD13	1:B:1691:HIS:CD2	2.50	0.46
1:A:1748:ASN:ND2	1:A:1748:ASN:H	2.14	0.46
1:B:1847:LYS:HE3	1:B:1857:HIS:HB3	1.97	0.46
1:A:1811:LYS:HB3	1:A:1905:SER:HB2	1.98	0.45
1:B:1703:LEU:HD23	1:B:1954:VAL:HG22	1.98	0.45
1:A:1840:GLU:OE2	1:A:1879:SER:OG	2.22	0.45
1:B:1858:ARG:HG2	1:B:1859:LEU:N	2.31	0.45
1:A:1821:ASP:O	1:A:1821:ASP:OD1	2.34	0.45
1:B:1831:GLN:HG3	1:B:1847:LYS:HB3	1.98	0.45
1:A:1858:ARG:HD2	1:A:1860:ILE:HD11	1.99	0.44
1:B:1692:PHE:O	1:B:1696:GLN:CG	2.66	0.44
1:B:1698:ASP:HB3	1:B:1701:TYR:HB3	2.00	0.44
1:B:1728:ARG:CD	3:B:283:HOH:O	2.59	0.43
1:B:1766:ASN:HD22	1:B:1922:GLN:HB2	1.83	0.43
1:A:1874:PRO:HD2	1:A:1955:TYR:CZ	2.52	0.43
1:A:1708:GLU:O	1:A:1711:LYS:CG	2.64	0.43
1:B:1692:PHE:CE1	1:B:1957:HIS:HB3	2.52	0.43
1:B:1817:PRO:HD3	1:B:1830:LEU:HD23	2.00	0.43
1:A:1858:ARG:HD2	1:A:1860:ILE:CD1	2.49	0.43
1:A:1858:ARG:HG2	1:A:1859:LEU:N	2.35	0.42
1:B:1765:GLY:HA2	1:B:1919[A]:ARG:HE	1.84	0.42
1:A:1967:ARG:O	1:A:1969:LEU:N	2.52	0.42
1:A:1766:ASN:ND2	3:A:123:HOH:O	2.53	0.42
1:A:1800:MET:HG2	1:A:1864:HIS:CD2	2.54	0.42
1:A:1861:ARG:HG3	1:A:1861:ARG:HH11	1.84	0.41
1:A:1748:ASN:ND2	1:A:1748:ASN:N	2.69	0.41
1:A:1794:ASN:HB3	3:A:289:HOH:O	2.21	0.41
1:A:1923:GLN:O	1:A:1927:LYS:CG	2.69	0.40
1:B:1726:GLU:HB2	1:B:1754:CYS:SG	2.62	0.40
1:A:1928:ASP:HA	3:A:127:HOH:O	2.21	0.40
1:B:1908:VAL:HG12	1:B:1946:MET:HE1	2.03	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	A	281/313 (90%)	270 (96%)	10 (4%)	1 (0%)	34	17
1	B	278/313 (89%)	270 (97%)	7 (2%)	1 (0%)	34	17
All	All	559/626 (89%)	540 (97%)	17 (3%)	2 (0%)	34	17

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	1947	VAL
1	B	1947	VAL

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	261/284 (92%)	252 (97%)	9 (3%)	37	14
1	B	258/284 (91%)	245 (95%)	13 (5%)	24	6
All	All	519/568 (91%)	497 (96%)	22 (4%)	29	9

All (22) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	1678	LYS

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	A	1679	THR
1	A	1748	ASN
1	A	1766	ASN
1	A	1819	ASP
1	A	1838	LEU
1	A	1927	LYS
1	A	1939	LEU
1	A	1967	ARG
1	B	1678	LYS
1	B	1679	THR
1	B	1680	SER
1	B	1685	ILE
1	B	1696	GLN
1	B	1708	GLU
1	B	1716	ASN
1	B	1728	ARG
1	B	1857	HIS
1	B	1860	ILE
1	B	1927	LYS
1	B	1928	ASP
1	B	1969	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (12) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	1734	ASN
1	A	1748	ASN
1	A	1766	ASN
1	A	1796	HIS
1	A	1882	GLN
1	A	1948	GLN
1	A	1958	GLN
1	B	1716	ASN
1	B	1831	GLN
1	B	1882	GLN
1	B	1923	GLN
1	B	1948	GLN

5.3.3 RNA

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 carbohydrates in this entry.

5.6 Ligand geometry [i](#)

2 ligands are 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	$\# Z > 2$
2	VO4	A	1	1	1,4,4	1.98	0	-		
2	VO4	B	2	1	1,4,4	1.98	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](#)

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	281/313 (89%)	0.61	27 (9%) 8 10	15, 25, 43, 60	0
1	B	280/313 (89%)	0.65	30 (10%) 6 8	16, 30, 51, 63	0
All	All	561/626 (89%)	0.63	57 (10%) 6 9	15, 27, 48, 63	0

All (57) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	1754	CYS	6.1
1	B	1754	CYS	5.2
1	A	1701	TYR	5.2
1	B	1701	TYR	5.2
1	B	1698	ASP	4.8
1	A	1748	ASN	4.4
1	B	1857	HIS	4.4
1	A	1678	LYS	4.2
1	B	1702	LEU	4.2
1	A	1968	LYS	4.1
1	B	1699	SER	4.1
1	B	1969	LEU	3.8
1	A	1726[A]	GLU	3.6
1	B	1703	LEU	3.5
1	B	1705	LYS	3.4
1	B	1967	ARG	3.3
1	B	1697	ALA	3.3
1	B	1968	LYS	3.2
1	A	1901	VAL	3.1
1	A	1685	ILE	3.0
1	A	1773	ILE	3.0
1	A	1799	VAL	2.9
1	B	1728	ARG	2.9
1	A	1969	LEU	2.9

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	B	1927	LYS	2.7
1	B	1951	CYS	2.7
1	B	1963	VAL	2.6
1	B	1878	GLN	2.5
1	B	1964	LEU	2.5
1	B	1965	ARG	2.5
1	B	1691	HIS	2.5
1	A	1967	ARG	2.4
1	B	1700	ASN	2.4
1	A	1755	SER	2.4
1	B	1712	ASP	2.4
1	A	1929	SER	2.3
1	A	1768	PHE	2.3
1	B	1818	ALA	2.2
1	A	1819	ASP	2.2
1	B	1966	ALA	2.2
1	B	1687	GLN	2.2
1	A	1728	ARG	2.2
1	B	1696	GLN	2.2
1	B	1726	GLU	2.2
1	A	1699	SER	2.2
1	A	1820	GLN	2.2
1	A	1851	GLU	2.2
1	A	1725	PRO	2.2
1	A	1698	ASP	2.2
1	A	1915	ILE	2.1
1	A	1708	GLU	2.1
1	A	1681[A]	CYS	2.1
1	B	1685	ILE	2.1
1	B	1773	ILE	2.1
1	B	1819	ASP	2.1
1	A	1927	LYS	2.0
1	A	1963	VAL	2.0

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 carbohydrates 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, 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	VO4	A	1	5/5	0.99	0.06	19,20,20,20	0
2	VO4	B	2	5/5	0.99	0.05	22,23,25,25	0

6.5 Other polymers [i](#)

There are no such residues in this entry.