WORLDWIDE wwPDB PROTEIN DATA BANK

wwPDB X-ray Structure Validation Summary Report ⓘ

Dec 18, 2023 – 09:14 AM EST

PDB ID	:	1VHK
Title	:	Crystal structure of an hypothetical protein
Authors	:	Structural GenomiX
Deposited on	:	2003-12-01
Resolution	:	2.60 Å(reported)

This is a wwPDB X-ray Structure Validation Summary 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 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:

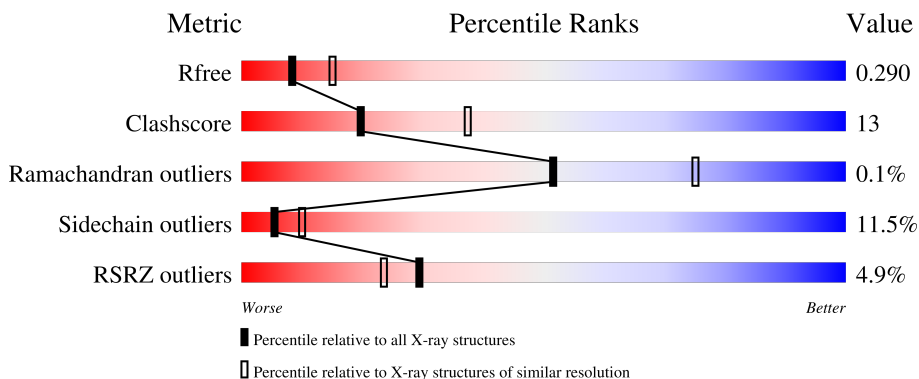|		
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (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 2.60 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		0.290
Clashscore		13
Ramachandran outliers		0.1%
Sidechain outliers		11.5%
RSRZ outliers		4.9%

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	3163 (2.60-2.60)
Clashscore	141614	3518 (2.60-2.60)
Ramachandran outliers	138981	3455 (2.60-2.60)
Sidechain outliers	138945	3455 (2.60-2.60)
RSRZ outliers	127900	3104 (2.60-2.60)

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	268	4% / 57% 25% 5% 12%
1	B	268	6% / 50% 19% • 27%
1	C	268	% / 53% 32% 6% 9%
1	D	268	4% / 49% 16% • 34%

2 Entry composition ⓘ

There are 2 unique types of molecules in this entry. The entry contains 6788 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 Hypothetical protein yqeU.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	235	Total	C	N	O	S	Se	0	0	0
			1847	1168	314	355	6	4			
1	B	196	Total	C	N	O	S	Se	0	0	0
			1534	975	262	289	6	2			
1	C	244	Total	C	N	O	S	Se	0	1	0
			1923	1216	331	366	6	4			
1	D	178	Total	C	N	O	S	Se	0	0	0
			1397	886	241	264	4	2			

There are 68 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-1	MSE	-	cloning artifact	UNP P54461
A	0	SER	-	cloning artifact	UNP P54461
A	1	LEU	-	cloning artifact	UNP P54461
A	33	MSE	MET	modified residue	UNP P54461
A	35	MSE	MET	modified residue	UNP P54461
A	152	MSE	MET	modified residue	UNP P54461
A	164	MSE	MET	modified residue	UNP P54461
A	257	GLU	-	cloning artifact	UNP P54461
A	258	GLY	-	cloning artifact	UNP P54461
A	259	GLY	-	cloning artifact	UNP P54461
A	260	SER	-	cloning artifact	UNP P54461
A	261	HIS	-	cloning artifact	UNP P54461
A	262	HIS	-	cloning artifact	UNP P54461
A	263	HIS	-	cloning artifact	UNP P54461
A	264	HIS	-	cloning artifact	UNP P54461
A	265	HIS	-	cloning artifact	UNP P54461
A	266	HIS	-	cloning artifact	UNP P54461
B	-1	MSE	-	cloning artifact	UNP P54461
B	0	SER	-	cloning artifact	UNP P54461
B	1	LEU	-	cloning artifact	UNP P54461
B	33	MSE	MET	modified residue	UNP P54461

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	35	MSE	MET	modified residue	UNP P54461
B	152	MSE	MET	modified residue	UNP P54461
B	164	MSE	MET	modified residue	UNP P54461
B	257	GLU	-	cloning artifact	UNP P54461
B	258	GLY	-	cloning artifact	UNP P54461
B	259	GLY	-	cloning artifact	UNP P54461
B	260	SER	-	cloning artifact	UNP P54461
B	261	HIS	-	cloning artifact	UNP P54461
B	262	HIS	-	cloning artifact	UNP P54461
B	263	HIS	-	cloning artifact	UNP P54461
B	264	HIS	-	cloning artifact	UNP P54461
B	265	HIS	-	cloning artifact	UNP P54461
B	266	HIS	-	cloning artifact	UNP P54461
C	-1	MSE	-	cloning artifact	UNP P54461
C	0	SER	-	cloning artifact	UNP P54461
C	1	LEU	-	cloning artifact	UNP P54461
C	33	MSE	MET	modified residue	UNP P54461
C	35	MSE	MET	modified residue	UNP P54461
C	152	MSE	MET	modified residue	UNP P54461
C	164	MSE	MET	modified residue	UNP P54461
C	257	GLU	-	cloning artifact	UNP P54461
C	258	GLY	-	cloning artifact	UNP P54461
C	259	GLY	-	cloning artifact	UNP P54461
C	260	SER	-	cloning artifact	UNP P54461
C	261	HIS	-	cloning artifact	UNP P54461
C	262	HIS	-	cloning artifact	UNP P54461
C	263	HIS	-	cloning artifact	UNP P54461
C	264	HIS	-	cloning artifact	UNP P54461
C	265	HIS	-	cloning artifact	UNP P54461
C	266	HIS	-	cloning artifact	UNP P54461
D	-1	MSE	-	cloning artifact	UNP P54461
D	0	SER	-	cloning artifact	UNP P54461
D	1	LEU	-	cloning artifact	UNP P54461
D	33	MSE	MET	modified residue	UNP P54461
D	35	MSE	MET	modified residue	UNP P54461
D	152	MSE	MET	modified residue	UNP P54461
D	164	MSE	MET	modified residue	UNP P54461
D	257	GLU	-	cloning artifact	UNP P54461
D	258	GLY	-	cloning artifact	UNP P54461
D	259	GLY	-	cloning artifact	UNP P54461
D	260	SER	-	cloning artifact	UNP P54461
D	261	HIS	-	cloning artifact	UNP P54461

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
D	262	HIS	-	cloning artifact	UNP P54461
D	263	HIS	-	cloning artifact	UNP P54461
D	264	HIS	-	cloning artifact	UNP P54461
D	265	HIS	-	cloning artifact	UNP P54461
D	266	HIS	-	cloning artifact	UNP P54461

- Molecule 2 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	21	Total 21	O 21	0	0
2	B	24	Total 24	O 24	0	0
2	C	32	Total 32	O 32	0	0
2	D	10	Total 10	O 10	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: Hypothetical protein yqeU

Chain A:

- Molecule 1: Hypothetical protein yqeU

Chain B:

- Molecule 1: Hypothetical protein yqeU

Chain C:

S187 A188 I189 V190 S191 S192 L193 P194 S197 S198 L199 L200 I201 G204 G207 G208 L209 T210 E211 A212 E213 V214 E215 R216 L217 T218 E219 Q220 D221 G222 R230 I231 L232 R233 T234 E235 T236 A237 P238 L239 L242 Y247 Q248 T249 L252 R253 GLY ASP GLN GLU GLY SER

HIS HIS HIS HIS HIS

● Molecule 1: Hypothetical protein yqeU

Chain D:

4%

49%   16%   ●   34%

MSE SER LEU Q2 R3 I6 GLU LEU THR LYS GLN GLN ILE GLU GLU ALA PRO THR PHE SER ILE THR GLY GLU GLU VAL HIS HIS ILE VAL ASN VAL MSE ARG MSE ASN GLU GLY ASP GLN ILE ILE C43 C44 S45 Q46 D47 G48 F49 E50 A51 LYS CYS GLU LEU GLN SER VAL SER

LYS ASP LYS VAL SER CYS LEU VAL ILE GLU TRP THR ASN E73 N74 R75 E76 L77 P78 I79 L87 K92 L93 E94 W95 K99 G104 A105 I109 V117 VAL LYS LEU ASP ASP LYS LYS ALA LYS LYS K128 R129 E130 R131 W132 E141 E147 V148 P149 R150 V151 M152

D153 V154 L161 Q162 Q165 E176 SER SER LYS GLN GLY GLU ILE S184 S187 A188 I189 V190 S191 S192 L193 P194 K195 E206 R216 Q220 L227 G228 P229 R233 T234 E235 T236 A237 P238 S246 Y247 Q248 T249 R253 GLY ASP GLN GLU GLY GLY SER 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	62.71Å 145.05Å 66.01Å 90.00° 109.40° 90.00°	Depositor
Resolution (Å)	38.19 – 2.60 38.19 – 2.60	Depositor EDS
% Data completeness (in resolution range)	(Not available) (38.19-2.60) 99.6 (38.19-2.60)	Depositor EDS
R_{merge}	0.11	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	3.62 (at 2.61Å)	Xtriage
Refinement program	REFMAC 4.0	Depositor
R, R_{free}	0.270 , 0.327 0.238 , 0.290	Depositor DCC
R_{free} test set	1684 reflections (4.95%)	wwPDB-VP
Wilson B-factor (Å^2)	43.6	Xtriage
Anisotropy	0.133	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.34 , 63.0	EDS
L-test for twinning[2]	$< |L| > = 0.49, < L^2 > = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.90	EDS
Total number of atoms	6788	wwPDB-VP
Average B, all atoms (Å^2)	43.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.28% 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 ⓘ

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	0.59	0/1874	1.03	3/2524 (0.1%)
1	B	0.67	0/1557	1.08	5/2099 (0.2%)
1	C	0.63	0/1956	1.07	6/2631 (0.2%)
1	D	0.59	0/1418	0.98	1/1907 (0.1%)
All	All	0.62	0/6805	1.04	15/9161 (0.2%)

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
1	C	0	1
All	All	0	2

There are no bond length outliers.

The worst 5 of 15 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	C	75	ARG	NE-CZ-NH1	-7.38	116.61	120.30
1	B	75	ARG	NE-CZ-NH1	6.96	123.78	120.30
1	C	230	ARG	NE-CZ-NH2	6.92	123.76	120.30
1	A	242	LEU	CA-CB-CG	6.27	129.71	115.30
1	C	47	ASP	CB-CG-OD2	6.09	123.78	118.30

There are no chirality outliers.

All (2) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	81	VAL	Mainchain
1	C	201	ILE	Mainchain

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	1847	0	1837	55	0
1	B	1534	0	1521	41	0
1	C	1923	0	1932	53	0
1	D	1397	0	1383	28	0
2	A	21	0	0	2	0
2	B	24	0	0	1	0
2	C	32	0	0	1	0
2	D	10	0	0	1	0
All	All	6788	0	6673	170	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 13.

The worst 5 of 170 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:C:75:ARG:HG2	1:D:229:PRO:HB2	1.39	1.01
1:A:57:SER:HB3	1:A:64:SER:HB2	1.46	0.94
1:C:9:THR:HG22	1:C:12:GLN:H	1.40	0.84
1:B:210:THR:HG22	1:B:213:GLU:H	1.42	0.83
1:B:6:ILE:HD11	1:B:43:CYS:HB3	1.61	0.83

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	229/268 (85%)	217 (95%)	12 (5%)	0	100	100
1	B	186/268 (69%)	178 (96%)	8 (4%)	0	100	100
1	C	241/268 (90%)	228 (95%)	12 (5%)	1 (0%)	34	57
1	D	168/268 (63%)	160 (95%)	8 (5%)	0	100	100
All	All	824/1072 (77%)	783 (95%)	40 (5%)	1 (0%)	51	75

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	C	61	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	202/227 (89%)	177 (88%)	25 (12%)	4	8
1	B	164/227 (72%)	149 (91%)	15 (9%)	9	18
1	C	210/227 (92%)	177 (84%)	33 (16%)	2	4
1	D	148/227 (65%)	138 (93%)	10 (7%)	16	32
All	All	724/908 (80%)	641 (88%)	83 (12%)	5	10

5 of 83 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	123	LYS
1	C	242	LEU
1	C	129	ARG
1	C	189	ILE
1	D	150	ARG

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 11 such sidechains are listed below:

Mol	Chain	Res	Type
1	C	56	GLN
1	C	106	HIS
1	D	165	GLN
1	C	220	GLN
1	B	112	GLN

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 ⓘ

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(Å²)	Q<0.9
1	A	231/268 (86%)	0.17	12 (5%)	27	21	17, 40, 76, 92	0
1	B	194/268 (72%)	0.26	17 (8%)	10	7	17, 34, 71, 92	0
1	C	240/268 (89%)	0.10	2 (0%)	86	84	21, 41, 61, 70	0
1	D	176/268 (65%)	0.28	10 (5%)	23	18	22, 45, 80, 91	0
All	All	841/1072 (78%)	0.19	41 (4%)	29	23	17, 40, 73, 92	0

The worst 5 of 41 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	41	ILE	5.2
1	B	66	LEU	5.1
1	B	65	CYS	4.6
1	D	47	ASP	4.2
1	B	43	CYS	3.8

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.