wwPDB X-ray Structure Validation Summary Report (i)

Jun 26, 2024 – 11:02 AM EDT

PDB ID	:	6TVH
Title	:	Selenomethionine-substituted HPF1 from Nematostella vectensis
Authors	:	Ariza, A.
Deposited on	:	2020-01-09
Resolution	:	2.65 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 (i) 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 (1)) were used in the production of this report:

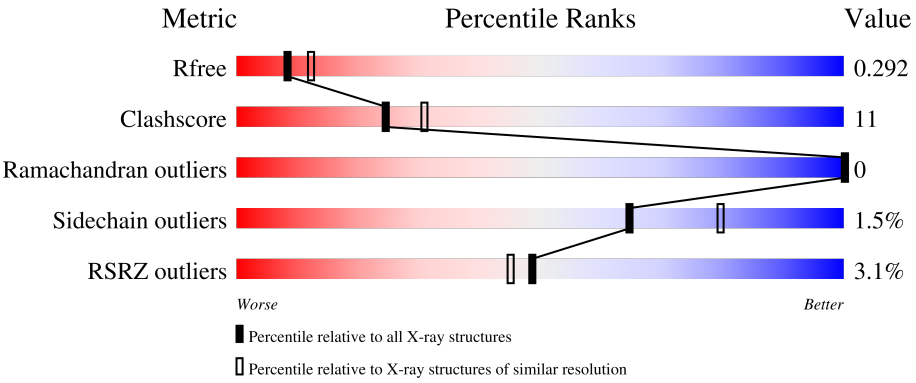
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.37.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.37.1

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 2.65 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	1332 (2.68-2.64)
Clashscore	141614	1374 (2.68-2.64)
Ramachandran outliers	138981	1349 (2.68-2.64)
Sidechain outliers	138945	1349 (2.68-2.64)
RSRZ outliers	127900	1318 (2.68-2.64)

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	336	4% 65% 24% · 10%
1	B	336	% 75% 14% 11%

2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 4835 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 Predicted protein.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	301	Total	C	N	O	S	Se	0	3	0
			2419	1546	399	460	4	10			
1	B	300	Total	C	N	O	S	Se	0	2	0
			2401	1535	393	459	4	10			

There are 18 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-1	MSE	-	expression tag	UNP A7RS11
A	0	GLY	-	expression tag	UNP A7RS11
A	328	LYS	-	expression tag	UNP A7RS11
A	329	HIS	-	expression tag	UNP A7RS11
A	330	HIS	-	expression tag	UNP A7RS11
A	331	HIS	-	expression tag	UNP A7RS11
A	332	HIS	-	expression tag	UNP A7RS11
A	333	HIS	-	expression tag	UNP A7RS11
A	334	HIS	-	expression tag	UNP A7RS11
B	-1	MSE	-	expression tag	UNP A7RS11
B	0	GLY	-	expression tag	UNP A7RS11
B	328	LYS	-	expression tag	UNP A7RS11
B	329	HIS	-	expression tag	UNP A7RS11
B	330	HIS	-	expression tag	UNP A7RS11
B	331	HIS	-	expression tag	UNP A7RS11
B	332	HIS	-	expression tag	UNP A7RS11
B	333	HIS	-	expression tag	UNP A7RS11
B	334	HIS	-	expression tag	UNP A7RS11

- Molecule 2 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	9	Total	O	0	0
			9	9		

Continued on next page...

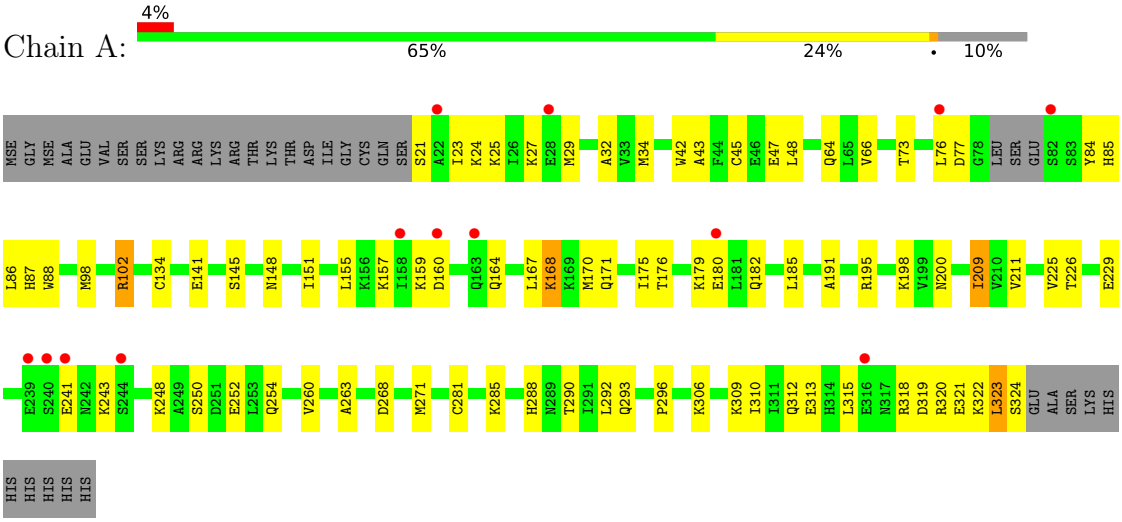
Continued from previous page...

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	B	6	Total	O	0	0
			6	6		

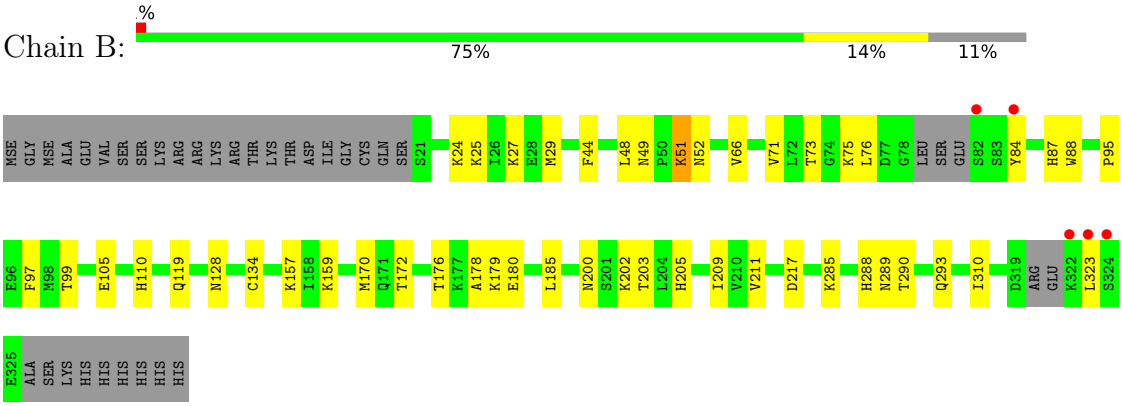
3 Residue-property plots (i)

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: Predicted protein



• Molecule 1: Predicted protein



4 Data and refinement statistics (i)

Property	Value	Source
Space group	H 3	Depositor
Cell constants a, b, c, α, β, γ	88.95Å 88.95Å 224.10Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	72.85 – 2.65 72.85 – 2.65	Depositor EDS
% Data completeness (in resolution range)	96.9 (72.85-2.65) 91.6 (72.85-2.65)	Depositor EDS
R_{merge}	0.28	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.22 (at 2.65Å)	Xtriage
Refinement program	REFMAC 5.8.0253, PHENIX dev_3126	Depositor
R, R_{free}	0.223 , 0.293 0.224 , 0.292	Depositor DCC
R_{free} test set	863 reflections (4.64%)	wwPDB-VP
Wilson B-factor (Å^2)	56.9	Xtriage
Anisotropy	0.181	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.30 , 38.0	EDS
L-test for twinning[2]	$< \|L\| > = 0.48$, $< L^2 > = 0.30$	Xtriage
Estimated twinning fraction	0.018 for -1/3*h+1/3*k+1/3*l,-k,8/3*h+4/3*k+1/3*l 0.025 for -2/3*h-1/3*k-1/3*l,-1/3*h-2/3*k+1/3*l,-4/3*h+4/3*k+1/3*l 0.022 for -h,1/3*h-1/3*k-1/3*l,-4/3*h-8/3*k+1/3*l 0.025 for -h,2/3*h+1/3*k+1/3*l,4/3*h+8/3*k-1/3*l 0.015 for 1/3*h+2/3*k-1/3*l,-k,-8/3*h-4/3*k-1/3*l 0.023 for -1/3*h-2/3*k+1/3*l,-2/3*h-1/3*k-1/3*l,4/3*h-4/3*k-1/3*l 0.054 for -h-k,k,-l	Xtriage
F_o,F_c correlation	0.93	EDS
Total number of atoms	4835	wwPDB-VP
Average B, all atoms (Å^2)	76.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 12.31% 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	0.30	0/2464	0.65	7/3301 (0.2%)
1	B	0.31	0/2441	0.48	1/3269 (0.0%)
All	All	0.31	0/4905	0.57	8/6570 (0.1%)

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	B	0	1

There are no bond length outliers.

The worst 5 of 8 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	24	LYS	CD-CE-NZ	-12.62	82.67	111.70
1	A	76	LEU	CA-CB-CG	10.01	138.31	115.30
1	A	24	LYS	CA-CB-CG	8.71	132.57	113.40
1	A	168	LYS	CA-CB-CG	8.11	131.24	113.40
1	B	76	LEU	CA-CB-CG	7.28	132.05	115.30

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	B	179	LYS	Peptide

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	2419	0	2435	70	0
1	B	2401	0	2412	36	0
2	A	9	0	0	1	0
2	B	6	0	0	0	0
All	All	4835	0	4847	105	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 11.

The worst 5 of 105 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:178:ALA:HB1	1:B:185:LEU:HD11	1.57	0.85
1:A:23:ILE:HG23	1:A:27:LYS:HE2	1.60	0.83
1:B:25:LYS:HE3	1:B:29:MSE:SE	2.30	0.80
1:B:44:PHE:O	1:B:48:LEU:HD12	1.80	0.80
1:B:178:ALA:CB	1:B:185:LEU:HD11	2.14	0.76

There are no symmetry-related clashes.

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	300/336 (89%)	289 (96%)	11 (4%)	0	100	100
1	B	296/336 (88%)	283 (96%)	13 (4%)	0	100	100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles
All	All	596/672 (89%)	572 (96%)	24 (4%)	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	268/284 (94%)	264 (98%)	4 (2%)	65 80
1	B	266/284 (94%)	261 (98%)	5 (2%)	57 74
All	All	534/568 (94%)	525 (98%)	9 (2%)	65 77

5 of 9 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	134	CYS
1	B	217	ASP
1	A	323	LEU
1	B	27	LYS
1	B	51	LYS

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	B	163	GLN
1	B	164	GLN
1	B	293	GLN
1	B	200	ASN
1	A	300	GLN

5.3.3 RNA (i)

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, 95^{th} 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	291/336 (86%)	0.20	13 (4%) 33 30	41, 73, 141, 203	0
1	B	290/336 (86%)	0.02	5 (1%) 70 67	37, 63, 116, 182	0
All	All	581/672 (86%)	0.11	18 (3%) 49 45	37, 68, 130, 203	0

The worst 5 of 18 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	82	SER	6.5
1	A	82	SER	6.3
1	A	160	ASP	4.3
1	B	323	LEU	3.6
1	A	22	ALA	3.2

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

There are no ligands in this entry.

6.5 Other polymers

There are no such residues in this entry.