



# wwPDB X-ray Structure Validation Summary Report ⓘ

Mar 9, 2018 – 06:19 pm GMT

PDB ID : 1G2Y  
Title : HNF-1ALPHA DIMERIZATION DOMAIN, WITH SELENOMETHIONINE  
SUBSTITUED AT LEU 12  
Authors : Rose, R.B.; Endrizzi, J.A.; Cronk, J.D.; Holton, J.; Alber, T.  
Deposited on : 2000-10-23  
Resolution : 1.00 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](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 following versions of software and data (see [references ⓘ](#)) were used in the production of this report:

MolProbity : 4.02b-467  
Xtriage (Phenix) : 1.13  
EDS : trunk30967  
Percentile statistics : 20171227.v01 (using entries in the PDB archive December 27th 2017)  
Refmac : 5.8.0158  
CCP4 : 7.0 (Gargrove)  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : trunk30967

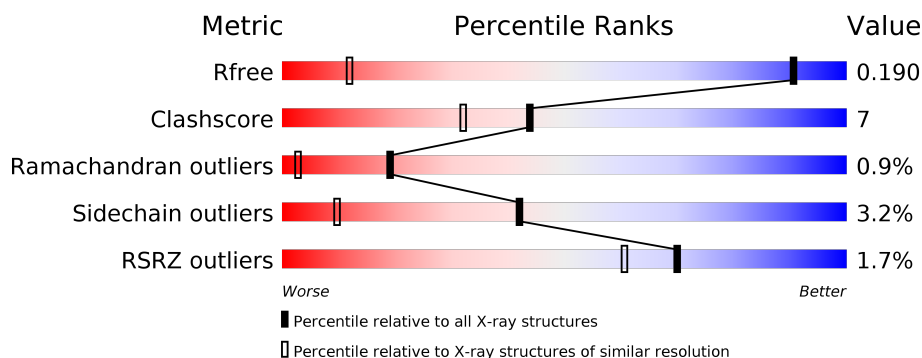
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 1.00 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))
$R_{free}$	111664	1020 (1.08-0.92)
Clashscore	122126	1084 (1.08-0.92)
Ramachandran outliers	120053	1010 (1.08-0.92)
Sidechain outliers	120020	1012 (1.08-0.92)
RSRZ outliers	108989	1485 (1.10-0.90)

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  $\geq 3$ , 2, 1 and 0 types of geometric quality criteria. 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	32	<div> <div>3%</div> <div> <div></div> <div>75%</div> <div>19%</div> <div>• •</div> </div> </div>
1	B	32	<div> <div>88%</div> <div>6%</div> <div>6%</div> </div>
1	C	32	<div> <div>72%</div> <div>16%</div> <div>•</div> <div>9%</div> </div>
1	D	32	<div> <div>3%</div> <div>84%</div> <div>6%</div> <div>•</div> <div>6%</div> </div>

## 2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 1039 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 HEPATOCYTE NUCLEAR FACTOR 1-ALPHA.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	31	Total	C	N	O	Se	0	0	0
			223	141	36	45	1			
1	B	30	Total	C	N	O	S	Se	0	0
			214	137	34	41	1	1		
1	C	29	Total	C	N	O	Se	0	0	0
			211	133	34	43	1			
1	D	30	Total	C	N	O	Se	0	0	0
			215	137	35	42	1			

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	12	MSE	LEU	Engineered	UNP P22361
B	12	MSE	LEU	Engineered	UNP P22361
C	12	MSE	LEU	Engineered	UNP P22361
D	12	MSE	LEU	Engineered	UNP P22361

- Molecule 2 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	50	Total	O	0	0
			50	50		
2	B	43	Total	O	0	0
			43	43		
2	C	41	Total	O	0	0
			41	41		
2	D	42	Total	O	0	0
			42	42		

### 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: HEPATOCYTE NUCLEAR FACTOR 1-ALPHA

Chain A: 



- Molecule 1: HEPATOCYTE NUCLEAR FACTOR 1-ALPHA

Chain B: 




- Molecule 1: HEPATOCYTE NUCLEAR FACTOR 1-ALPHA

Chain C: 



- Molecule 1: HEPATOCYTE NUCLEAR FACTOR 1-ALPHA

Chain D: 



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	31.27Å 47.88Å 40.47Å 90.00° 93.60° 90.00°	Depositor
Resolution (Å)	30.88 – 1.00 30.87 – 1.00	Depositor EDS
% Data completeness (in resolution range)	85.5 (30.88-1.00) 85.2 (30.87-1.00)	Depositor EDS
$R_{merge}$	0.05	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	12.50 (at 1.00Å)	Xtriage
Refinement program	CNS	Depositor
R, $R_{free}$	0.195 , 0.198 0.198 , 0.190	Depositor DCC
$R_{free}$ test set	1657 reflections (3.00%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	8.5	Xtriage
Anisotropy	0.242	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.35 , 110.0	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.48$ , $\langle L^2 \rangle = 0.31$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.96	EDS
Total number of atoms	1039	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	16.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 15.23% 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](#)

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	1.14	3/221 (1.4%)	0.98	0/294
1	B	1.02	2/212 (0.9%)	0.81	0/282
1	C	1.15	2/209 (1.0%)	0.83	0/278
1	D	1.10	2/213 (0.9%)	0.93	0/284
All	All	1.11	9/855 (1.1%)	0.89	0/1138

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	1	0

The worst 5 of 9 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	C	18	GLU	CD-OE2	7.52	1.33	1.25
1	A	18	GLU	CD-OE2	6.94	1.33	1.25
1	B	18	GLU	CD-OE2	6.56	1.32	1.25
1	A	24	GLU	CD-OE2	6.35	1.32	1.25
1	C	24	GLU	CD-OE2	5.53	1.31	1.25

There are no bond angle outliers.

All (1) chirality outliers are listed below:

Mol	Chain	Res	Type	Atom
1	A	2	VAL	CA

There are no planarity outliers.

## 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	223	0	242	4	0
1	B	214	0	236	0	0
1	C	211	0	225	8	1
1	D	215	0	235	1	0
2	A	50	0	0	0	1
2	B	43	0	0	0	1
2	C	41	0	0	0	0
2	D	42	0	0	0	1
All	All	1039	0	938	13	2

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 7.

The worst 5 of 13 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:23:LYS:HD2	1:C:23:LYS:H	1.14	1.09
1:C:23:LYS:HD2	1:C:23:LYS:N	1.95	0.81
1:C:8:LEU:CG	1:C:12:MSE:HE3	2.12	0.79
1:A:21:LEU:HD23	1:A:26:LEU:HD13	1.68	0.76
1:C:8:LEU:HD11	1:C:12:MSE:HE3	1.68	0.75

All (2) 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)
2:A:36:HOH:O	2:B:54:HOH:O[2_645]	2.02	0.18
1:C:2:VAL:N	2:D:36:HOH:O[2_555]	2.18	0.02

## 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	29/32 (91%)	29 (100%)	0	0	100	100
1	B	28/32 (88%)	28 (100%)	0	0	100	100
1	C	27/32 (84%)	26 (96%)	1 (4%)	0	100	100
1	D	28/32 (88%)	27 (96%)	0	1 (4%)	4	0
All	All	112/128 (88%)	110 (98%)	1 (1%)	1 (1%)	19	3

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	D	2	VAL

### 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	24/25 (96%)	22 (92%)	2 (8%)	12	1
1	B	23/25 (92%)	23 (100%)	0	100	100
1	C	23/25 (92%)	22 (96%)	1 (4%)	32	5
1	D	23/25 (92%)	23 (100%)	0	100	100
All	All	93/100 (93%)	90 (97%)	3 (3%)	42	10

All (3) residues with a non-rotameric sidechain are listed below:



Mol	Chain	Res	Type
1	A	2	VAL
1	A	28	GLN
1	C	23	LYS

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (2) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	28	GLN
1	D	7	GLN

### 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 carbohydrates in this entry.

## 5.6 Ligand geometry [i](#)

There are no ligands in this entry.

## 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	30/32 (93%)	0.14	1 (3%) 46 38	7, 12, 28, 37	0
1	B	29/32 (90%)	0.01	0 100 100	7, 11, 19, 23	0
1	C	28/32 (87%)	0.11	0 100 100	7, 10, 38, 46	0
1	D	29/32 (90%)	0.30	1 (3%) 45 37	7, 11, 34, 66	0
All	All	116/128 (90%)	0.14	2 (1%) 70 60	7, 11, 35, 66	0

All (2) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	D	1	MET	6.5
1	A	1	MET	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 carbohydrates in this entry.

### 6.4 Ligands [i](#)

There are no ligands in this entry.

### 6.5 Other polymers [i](#)

There are no such residues in this entry.