WORLDWIDE PDB

PROTEIN DATA BANK

wwPDB X-ray Structure Validation Summary Report (i)

Jan 16, 2018 – 06:33 AM EST

PDB ID	:	5NVK
Title	:	Crystal structure of the human 4EHP-GIGYF1 complex
Authors	:	Peter, D.; Valkov, E.
Deposited on	:	2017-05-04
Resolution	:	2.90 Å(reported)

The following versions of software and data (see references (1)) were used in the production of this report:

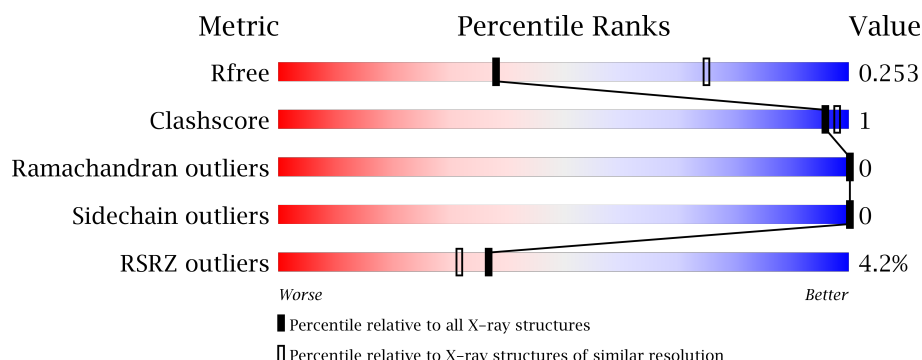|		
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20030736
Percentile statistics	:	20161228.v01 (using entries in the PDB archive December 28th 2016)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	rb-20030736

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 2.90 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		0.253
Clashscore		1
Ramachandran outliers		0
Sidechain outliers		0
RSRZ outliers		4.2%

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}	100719	1586 (2.90-2.90)
Clashscore	112137	1807 (2.90-2.90)
Ramachandran outliers	110173	1768 (2.90-2.90)
Sidechain outliers	110143	1770 (2.90-2.90)
RSRZ outliers	101464	1596 (2.90-2.90)

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 >=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 <=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	189	2% 88% • 9%
1	C	189	4% 85% • 11%
1	E	189	3% 86% • 11%
1	G	189	.% 89% • 9%
2	B	75	84% 7% 9%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
2	D	75	16% ▬▬▬ 92% ▬▬▬▬ 7% ·
2	F	75	11% ▬▬ 85% ▬▬▬ · 11%
2	H	75	3% ▬ 84% ▬▬▬ · 12%

2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 15534 atoms, of which 7701 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 Eukaryotic translation initiation factor 4E type 2.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	172	Total	C	H	N	O	S	0	0	0
			2805	899	1392	255	251	8			
1	C	168	Total	C	H	N	O	S	0	0	0
			2722	876	1341	247	251	7			
1	E	168	Total	C	H	N	O	S	0	0	0
			2721	876	1340	247	251	7			
1	G	172	Total	C	H	N	O	S	0	0	0
			2807	900	1395	254	250	8			

There are 24 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	46	GLY	-	expression tag	UNP O60573
A	47	PRO	-	expression tag	UNP O60573
A	48	HIS	-	expression tag	UNP O60573
A	49	MET	-	expression tag	UNP O60573
A	50	LEU	-	expression tag	UNP O60573
A	51	GLU	-	expression tag	UNP O60573
C	46	GLY	-	expression tag	UNP O60573
C	47	PRO	-	expression tag	UNP O60573
C	48	HIS	-	expression tag	UNP O60573
C	49	MET	-	expression tag	UNP O60573
C	50	LEU	-	expression tag	UNP O60573
C	51	GLU	-	expression tag	UNP O60573
E	46	GLY	-	expression tag	UNP O60573
E	47	PRO	-	expression tag	UNP O60573
E	48	HIS	-	expression tag	UNP O60573
E	49	MET	-	expression tag	UNP O60573
E	50	LEU	-	expression tag	UNP O60573
E	51	GLU	-	expression tag	UNP O60573
G	46	GLY	-	expression tag	UNP O60573
G	47	PRO	-	expression tag	UNP O60573
G	48	HIS	-	expression tag	UNP O60573

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
G	49	MET	-	expression tag	UNP O60573
G	50	LEU	-	expression tag	UNP O60573
G	51	GLU	-	expression tag	UNP O60573

- Molecule 2 is a protein called GRB10-interacting GYF protein 1.

| Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
2	B	68	Total 1103	C 351	H 550	N 90	O 110	S 2	0	0	0
2	D	74	Total 1217	C 388	H 609	N 100	O 117	S 3	0	0	0
2	F	67	Total 1085	C 345	H 540	N 89	O 109	S 2	0	0	0
2	H	66	Total 1074	C 342	H 534	N 88	O 108	S 2	0	0	0

There are 16 discrepancies between the modelled and reference sequences:

| Chain	Residue	Modelled	Actual	Comment	Reference
B	29	GLY	-	expression tag	UNP O75420
B	30	PRO	-	expression tag	UNP O75420
B	31	HIS	-	expression tag	UNP O75420
B	32	MET	-	expression tag	UNP O75420
D	29	GLY	-	expression tag	UNP O75420
D	30	PRO	-	expression tag	UNP O75420
D	31	HIS	-	expression tag	UNP O75420
D	32	MET	-	expression tag	UNP O75420
F	29	GLY	-	expression tag	UNP O75420
F	30	PRO	-	expression tag	UNP O75420
F	31	HIS	-	expression tag	UNP O75420
F	32	MET	-	expression tag	UNP O75420
H	29	GLY	-	expression tag	UNP O75420
H	30	PRO	-	expression tag	UNP O75420
H	31	HIS	-	expression tag	UNP O75420
H	32	MET	-	expression tag	UNP O75420

3 Residue-property plots ⓘ

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.
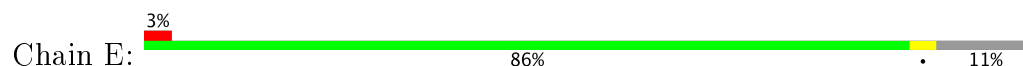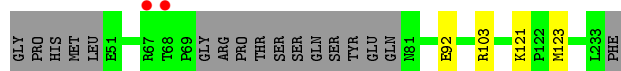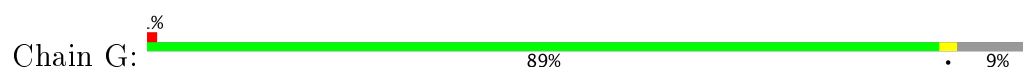
- Molecule 1: Eukaryotic translation initiation factor 4E type 2
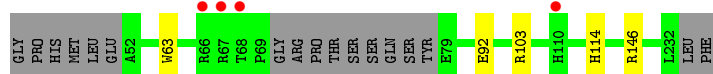
- Molecule 1: Eukaryotic translation initiation factor 4E type 2
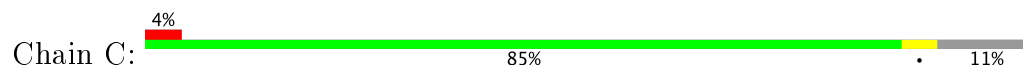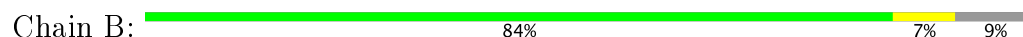

- Molecule 1: Eukaryotic translation initiation factor 4E type 2

- Molecule 1: Eukaryotic translation initiation factor 4E type 2
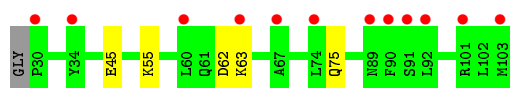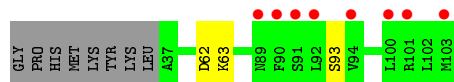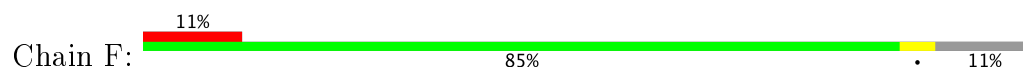

- Molecule 2: GRB10-interacting GYF protein 1

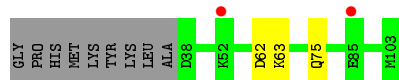- Molecule 2: GRB10-interacting GYF protein 1
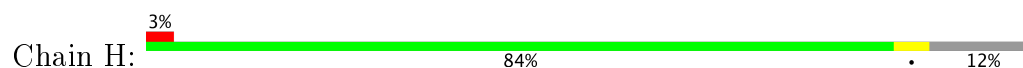
• Molecule 2: GRB10-interacting GYF protein 1

Chain F:

• Molecule 2: GRB10-interacting GYF protein 1

Chain H:

4   Data and refinement statistics   (i)

Property	Value	Source
Space group	P 42	Depositor
Cell constants a, b, c, α, β, γ	135.32Å 135.32Å 60.91Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	47.84 – 2.90 47.84 – 2.90	Depositor EDS
% Data completeness (in resolution range)	99.7 (47.84-2.90) 99.7 (47.84-2.90)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.11	Depositor
$< I/\sigma(I) >$ [1]	2.02 (at 2.91Å)	Xtriage
Refinement program	PHENIX (1.10.1_2155: ???)	Depositor
R, R_{free}	0.204 , 0.254 0.204 , 0.253	Depositor DCC
R_{free} test set	1217 reflections (4.93%)	DCC
Wilson B-factor (Å^2)	51.9	Xtriage
Anisotropy	0.165	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.38 , 29.1	EDS
L-test for twinning[2]	$< |L| > = 0.49, < L^2 > = 0.32$	Xtriage
Estimated twinning fraction	0.031 for h,-k,-l	Xtriage
F_o,F_c correlation	0.92	EDS
Total number of atoms	15534	wwPDB-VP
Average B, all atoms (Å^2)	54.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 10.73% 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 (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	0.26	0/1449	0.43	0/1957
1	C	0.25	0/1418	0.43	0/1919
1	E	0.24	0/1418	0.43	0/1919
1	G	0.26	0/1448	0.43	0/1956
2	B	0.26	0/561	0.40	0/759
2	D	0.24	0/619	0.39	0/835
2	F	0.26	0/553	0.41	0/748
2	H	0.25	0/548	0.40	0/741
All	All	0.25	0/8014	0.42	0/10834

There are no bond length outliers.

There are no bond angle outliers.

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	1413	1392	1391	4	0
1	C	1381	1341	1340	5	0
1	E	1381	1340	1340	3	0
1	G	1412	1395	1394	3	0
2	B	553	550	550	4	0
2	D	608	609	609	4	0
2	F	545	540	539	2	0
2	H	540	534	534	2	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
All	All	7833	7701	7697	20	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 1.

The worst 5 of 20 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)
2:B:40:ARG:NE	2:D:45:GLU:OE2	2.29	0.62
2:F:62:ASP:OD1	2:F:63:LYS:N	2.35	0.60
1:C:103:ARG:NH1	2:D:75:GLN:O	2.40	0.54
1:A:103:ARG:NH2	2:B:75:GLN:O	2.41	0.54
1:C:100:HIS:NE2	2:D:55:LYS:O	2.42	0.53

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	168/189 (89%)	160 (95%)	8 (5%)	0	100	100
1	C	166/189 (88%)	159 (96%)	7 (4%)	0	100	100
1	E	166/189 (88%)	161 (97%)	5 (3%)	0	100	100
1	G	168/189 (89%)	159 (95%)	9 (5%)	0	100	100
2	B	66/75 (88%)	64 (97%)	2 (3%)	0	100	100
2	D	72/75 (96%)	68 (94%)	4 (6%)	0	100	100
2	F	65/75 (87%)	64 (98%)	1 (2%)	0	100	100
2	H	64/75 (85%)	62 (97%)	2 (3%)	0	100	100
All	All	935/1056 (88%)	897 (96%)	38 (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	151/166 (91%)	151 (100%)	0	100	100
1	C	148/166 (89%)	148 (100%)	0	100	100
1	E	148/166 (89%)	148 (100%)	0	100	100
1	G	151/166 (91%)	151 (100%)	0	100	100
2	B	61/67 (91%)	61 (100%)	0	100	100
2	D	67/67 (100%)	67 (100%)	0	100	100
2	F	60/67 (90%)	60 (100%)	0	100	100
2	H	60/67 (90%)	60 (100%)	0	100	100
All	All	846/932 (91%)	846 (100%)	0	100	100

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

Some 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 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, 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	172/189 (91%)	0.21	4 (2%)	61	57	25, 34, 70, 139	0
1	C	168/189 (88%)	0.43	7 (4%)	37	32	26, 46, 101, 153	0
1	E	168/189 (88%)	0.32	5 (2%)	51	44	29, 47, 81, 108	0
1	G	172/189 (91%)	0.20	2 (1%)	79	77	25, 36, 71, 139	0
2	B	68/75 (90%)	0.29	0	100	100	28, 46, 77, 84	0
2	D	74/75 (98%)	1.01	12 (16%)	2	1	32, 63, 106, 193	0
2	F	67/75 (89%)	0.70	8 (11%)	5	3	29, 51, 105, 150	0
2	H	66/75 (88%)	0.37	2 (3%)	51	44	30, 50, 81, 106	0
All	All	955/1056 (90%)	0.38	40 (4%)	37	32	25, 43, 90, 193	0

The worst 5 of 40 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	D	90	PHE	10.1
1	G	68	THR	5.2
2	D	30	PRO	4.9
1	C	69	PRO	4.7
2	F	90	PHE	4.2

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.