



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 19, 2016 – 10:59 PM GMT

PDB ID : 5FGO  
Title : Crystal structure of D. melanogaster Pur-alpha repeat III.  
Authors : Windhager, A.; Janowski, R.; Niessing, D.  
Deposited on : 2015-12-21  
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](mailto:validation@mail.wwpdb.org)  
A user guide is available at  
<http://wwpdb.org/validation/2016/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 : unknown  
Xtriage (Phenix) : 1.9-1692  
EDS : rb-20026982  
Percentile statistics : 20151230.v01 (using entries in the PDB archive December 30th 2015)  
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-20026982

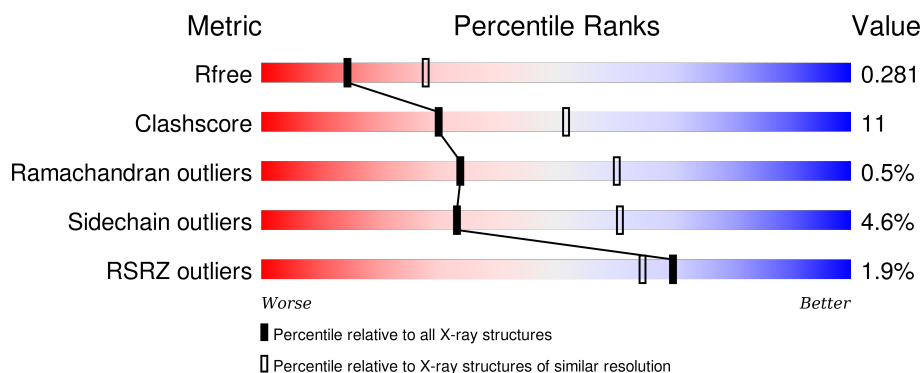
# 1 Overall quality at a glance

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	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
$R_{free}$	91344	2328 (2.60-2.60)
Clashscore	102246	2679 (2.60-2.60)
Ramachandran outliers	100387	2635 (2.60-2.60)
Sidechain outliers	100360	2635 (2.60-2.60)
RSRZ outliers	91569	2334 (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 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	81	<div> <div>2%</div> <div>53% 21% • 23%</div> </div>
1	B	81	<div> <div>60% 20% 20%</div> </div>
1	C	81	<div> <div>62% 10% • 25%</div> </div>
1	D	81	<div> <div>5% 54% 23% • 19%</div> </div>
1	E	81	<div> <div>% 51% 27% • 21%</div> </div>

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
1	F	81	 A horizontal bar chart showing the quality of chain F. The bar is divided into three segments: green (59%), yellow (17%), and grey (22%). A small black dot is located on the yellow segment.

## 2 Entry composition [i](#)

There are 3 unique types of molecules in this entry. The entry contains 3515 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 CG1507-PB, isoform B.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	62	Total	C	N	O	S	Se	0	0	0
			539	341	98	95	2	3			
1	B	65	Total	C	N	O	S	Se	3	0	0
			559	353	101	100	2	3			
1	C	61	Total	C	N	O	S	Se	1	1	0
			535	339	96	94	2	4			
1	D	66	Total	C	N	O	S	Se	0	5	0
			602	381	107	107	3	4			
1	E	64	Total	C	N	O	S	Se	0	1	0
			561	356	101	99	2	3			
1	F	63	Total	C	N	O	S	Se	1	0	0
			546	345	98	98	2	3			

There are 30 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	180	GLY	-	expression tag	UNP Q9V4D9
A	181	PRO	-	expression tag	UNP Q9V4D9
A	182	LEU	-	expression tag	UNP Q9V4D9
A	183	GLY	-	expression tag	UNP Q9V4D9
A	184	SER	-	expression tag	UNP Q9V4D9
B	180	GLY	-	expression tag	UNP Q9V4D9
B	181	PRO	-	expression tag	UNP Q9V4D9
B	182	LEU	-	expression tag	UNP Q9V4D9
B	183	GLY	-	expression tag	UNP Q9V4D9
B	184	SER	-	expression tag	UNP Q9V4D9
C	180	GLY	-	expression tag	UNP Q9V4D9
C	181	PRO	-	expression tag	UNP Q9V4D9
C	182	LEU	-	expression tag	UNP Q9V4D9
C	183	GLY	-	expression tag	UNP Q9V4D9
C	184	SER	-	expression tag	UNP Q9V4D9
D	180	GLY	-	expression tag	UNP Q9V4D9
D	181	PRO	-	expression tag	UNP Q9V4D9

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
D	182	LEU	-	expression tag	UNP Q9V4D9
D	183	GLY	-	expression tag	UNP Q9V4D9
D	184	SER	-	expression tag	UNP Q9V4D9
E	180	GLY	-	expression tag	UNP Q9V4D9
E	181	PRO	-	expression tag	UNP Q9V4D9
E	182	LEU	-	expression tag	UNP Q9V4D9
E	183	GLY	-	expression tag	UNP Q9V4D9
E	184	SER	-	expression tag	UNP Q9V4D9
F	180	GLY	-	expression tag	UNP Q9V4D9
F	181	PRO	-	expression tag	UNP Q9V4D9
F	182	LEU	-	expression tag	UNP Q9V4D9
F	183	GLY	-	expression tag	UNP Q9V4D9
F	184	SER	-	expression tag	UNP Q9V4D9

- Molecule 2 is CHLORIDE ION (three-letter code: CL) (formula: Cl).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	B	1	Total Cl 1 1	0	0
2	D	1	Total Cl 1 1	0	0

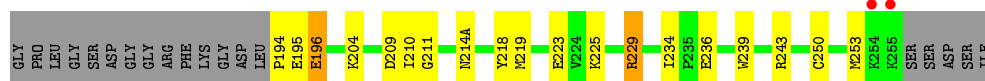
- Molecule 3 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	27	Total O 28 28	0	1
3	B	28	Total O 28 28	0	0
3	C	21	Total O 21 21	0	0
3	D	44	Total O 44 44	0	0
3	E	32	Total O 32 32	0	0
3	F	18	Total O 18 18	0	0

### 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 errors 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: CG1507-PB, isoform B



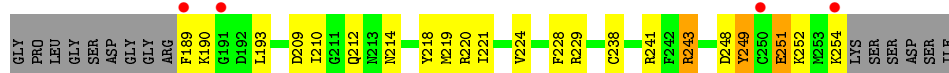
- Molecule 1: CG1507-PB, isoform B



- Molecule 1: CG1507-PB, isoform B



- Molecule 1: CG1507-PB, isoform B



- Molecule 1: CG1507-PB, isoform B



- Molecule 1: CG1507-PB, isoform B



GLY	PRO	LEU	GLY	SER	ASP	GLY	GLY	ARG	PHE	LYS	GLY	D192	E196	K200	K205	N213	N214	N215	K225	T230	F242	R243	D244	N247	C250	E251	K252	N253	K254	LYS	SER	SER	ASP	SER	ILE
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	-----	-----	-----	-----	-----	-----

## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	61.46Å 55.53Å 67.84Å 90.00° 95.65° 90.00°	Depositor
Resolution (Å)	47.73 – 2.60 47.72 – 2.60	Depositor EDS
% Data completeness (in resolution range)	96.4 (47.73-2.60) 96.6 (47.72-2.60)	Depositor EDS
$R_{merge}$	0.11	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.35 (at 2.61Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.8.2_1309)	Depositor
R, $R_{free}$	0.202 , 0.288 0.207 , 0.281	Depositor DCC
$R_{free}$ test set	687 reflections (5.27%)	DCC
Wilson B-factor (Å <sup>2</sup> )	30.3	Xtriage
Anisotropy	0.651	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.35 , 52.5	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.32$	Xtriage
Outliers	0 of 13726 reflections	Xtriage
$F_o, F_c$ correlation	0.92	EDS
Total number of atoms	3515	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	12.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.88% 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, 0.375 respectively for untwinned datasets, and 0.333, 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:  
CL

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	0.54	0/548	0.75	0/725
1	B	0.53	0/568	0.69	0/753
1	C	0.57	0/547	0.73	0/724
1	D	0.52	0/609	0.69	0/801
1	E	0.48	0/572	0.67	0/756
1	F	0.52	0/555	0.63	0/737
All	All	0.53	0/3399	0.70	0/4496

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	539	0	528	19	0
1	B	559	0	545	11	0
1	C	535	0	524	13	0
1	D	602	0	587	25	0
1	E	561	0	555	16	0
1	F	546	0	529	7	0
2	B	1	0	0	0	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	D	1	0	0	0	0
3	A	28	0	0	0	0
3	B	28	0	0	1	0
3	C	21	0	0	1	0
3	D	44	0	0	4	0
3	E	32	0	0	1	0
3	F	18	0	0	0	0
All	All	3515	0	3268	73	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 73 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:204:LYS:HG2	1:A:225:LYS:HB2	1.60	0.83
1:B:243:ARG:NH1	1:B:244:ASP:OD1	2.11	0.82
1:F:243:ARG:NH1	1:F:244:ASP:OD1	2.18	0.76
3:D:401:HOH:O	1:F:230:THR:HG21	1.93	0.68
1:C:210:ILE:HG12	1:C:219[A]:MSE:HG3	1.75	0.68

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	60/81 (74%)	56 (93%)	3 (5%)	1 (2%)	11	22
1	B	63/81 (78%)	56 (89%)	7 (11%)	0	100	100
1	C	60/81 (74%)	58 (97%)	2 (3%)	0	100	100
1	D	68/81 (84%)	62 (91%)	6 (9%)	0	100	100

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	E	63/81 (78%)	56 (89%)	7 (11%)	0	100	100
1	F	61/81 (75%)	55 (90%)	5 (8%)	1 (2%)	12	24
All	All	375/486 (77%)	343 (92%)	30 (8%)	2 (0%)	34	60

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	196	GLU
1	F	252	LYS

### 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	60/71 (84%)	59 (98%)	1 (2%)	68	88
1	B	62/71 (87%)	58 (94%)	4 (6%)	21	42
1	C	60/71 (84%)	57 (95%)	3 (5%)	30	56
1	D	67/71 (94%)	63 (94%)	4 (6%)	24	47
1	E	63/71 (89%)	61 (97%)	2 (3%)	46	74
1	F	61/71 (86%)	57 (93%)	4 (7%)	21	40
All	All	373/426 (88%)	355 (95%)	18 (5%)	33	58

5 of 18 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	D	243	ARG
1	D	249	TYR
1	F	196	GLU
1	C	200	LYS
1	C	203	ASN

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

### 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 carbohydrates in this entry.

### 5.6 Ligand geometry ⓘ

Of 2 ligands modelled in this entry, 2 are monoatomic - leaving 0 for Mogul analysis.

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 ⓘ

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 [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	59/81 (72%)	-0.11	2 (3%) 49 41	1, 6, 30, 39	3 (5%)
1	B	62/81 (76%)	-0.28	0 100 100	1, 5, 27, 37	3 (4%)
1	C	58/81 (71%)	-0.38	0 100 100	1, 3, 24, 35	2 (3%)
1	D	63/81 (77%)	0.06	4 (6%) 23 17	1, 6, 28, 57	1 (1%)
1	E	61/81 (75%)	0.19	1 (1%) 74 69	3, 14, 32, 52	2 (3%)
1	F	60/81 (74%)	0.27	0 100 100	8, 19, 33, 39	5 (8%)
All	All	363/486 (74%)	-0.04	7 (1%) 70 64	1, 10, 32, 57	16 (4%)

The worst 5 of 7 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	D	191	GLY	6.9
1	D	189	PHE	5.4
1	D	254[B]	LYS	4.5
1	A	255	LYS	3.5
1	E	198	HIS	2.8

### 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. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. The B-factors column lists the minimum, median, 95<sup>th</sup> 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	LLDF	B-factors( $\text{\AA}^2$ )	Q<0.9
2	CL	B	301	1/1	0.98	0.08	-1.11	20,20,20,20	0
2	CL	D	301	1/1	0.98	0.06	-	10,10,10,10	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.