



# wwPDB X-ray Structure Validation Summary Report i

Feb 28, 2014 – 01:52 PM GMT

PDB ID : 2GCJ  
Title : Crystal Structure of the Pob3 middle domain  
Authors : VanDemark, A.P.  
Deposited on : 2006-03-14  
Resolution : 2.55 Å(reported)

This is a wwPDB 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/ValidationPDFNotes.html>

---

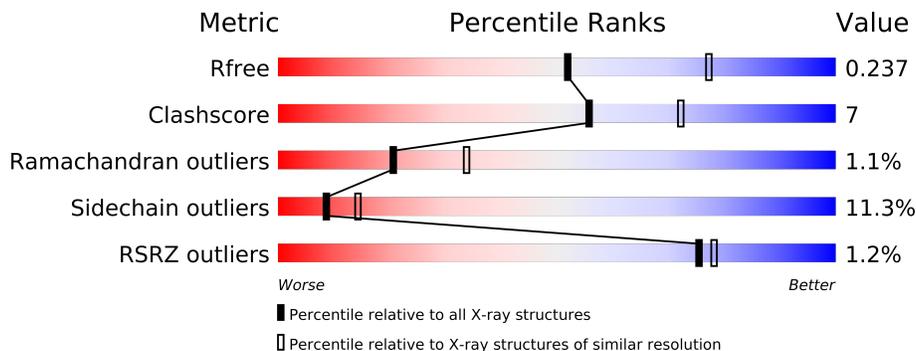
The following versions of software and data (see [references](#)) were used in the production of this report:

MolProbity : 4.02b-467  
Mogul : 1.15 2013  
Xtriage (Phenix) : dev-1323  
EDS : stable22639  
Percentile statistics : 21963  
Refmac : 5.8.0049  
CCP4 : 6.3.0 (Settle)  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et. al. (1996)  
Validation Pipeline (wwPDB-VP) : stable22683

# 1 Overall quality at a glance

The reported resolution of this entry is 2.55 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))
$R_{free}$	66092	3413 (2.58-2.50)
Clashscore	79885	4284 (2.58-2.50)
Ramachandran outliers	78287	4193 (2.58-2.50)
Sidechain outliers	78261	4195 (2.58-2.50)
RSRZ outliers	66119	3414 (2.58-2.50)

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. The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density.

Mol	Chain	Length	Quality of chain
1	A	261	
1	B	261	
1	C	261	
1	D	261	

## 2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 7649 atoms, of which 0 are hydrogen and 0 are deuterium.

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 63.0 kDa protein in DAK1-ORC1 intergenic region.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	229	Total 1888	C 1210	N 321	O 354	S 3	0	0	0
1	B	230	Total 1895	C 1215	N 322	O 355	S 3	3	0	0
1	C	231	Total 1904	C 1220	N 323	O 358	S 3	5	0	0
1	D	229	Total 1882	C 1207	N 318	O 354	S 3	0	0	0

There are 12 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	218	GLY	-	CLONING ARTIFACT	UNP Q04636
A	219	HIS	-	CLONING ARTIFACT	UNP Q04636
A	308	LYS	GLN	ENGINEERED	UNP Q04636
B	218	GLY	-	CLONING ARTIFACT	UNP Q04636
B	219	HIS	-	CLONING ARTIFACT	UNP Q04636
B	308	LYS	GLN	ENGINEERED	UNP Q04636
C	218	GLY	-	CLONING ARTIFACT	UNP Q04636
C	219	HIS	-	CLONING ARTIFACT	UNP Q04636
C	308	LYS	GLN	ENGINEERED	UNP Q04636
D	218	GLY	-	CLONING ARTIFACT	UNP Q04636
D	219	HIS	-	CLONING ARTIFACT	UNP Q04636
D	308	LYS	GLN	ENGINEERED	UNP Q04636

- Molecule 2 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	21	Total 21	O 21	0	0
2	B	25	Total 25	O 25	0	0

*Continued on next page...*

*Continued from previous page...*

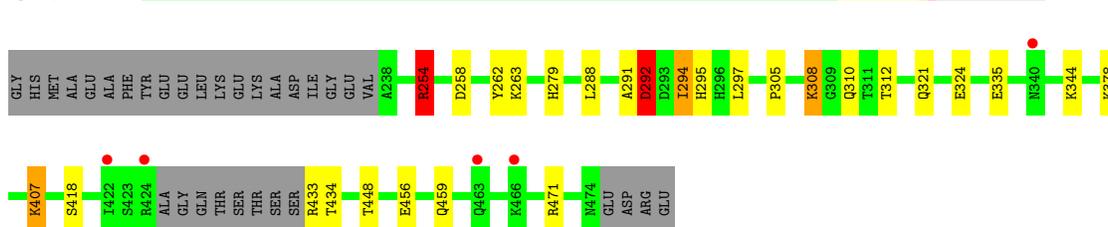
<b>Mol</b>	<b>Chain</b>	<b>Residues</b>	<b>Atoms</b>		<b>ZeroOcc</b>	<b>AltConf</b>
2	C	18	Total	O	0	0
			18	18		
2	D	16	Total	O	0	0
			16	16		

### 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 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.

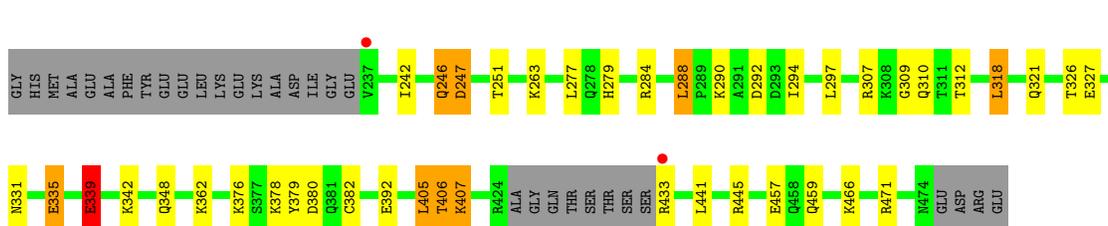
- Molecule 1: Hypothetical 63.0 kDa protein in DAK1-ORC1 intergenic region

Chain A:



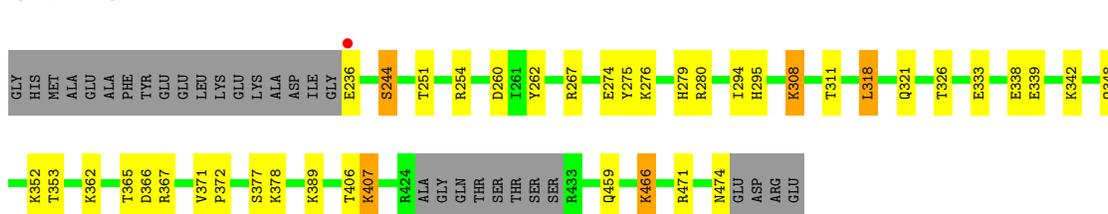
- Molecule 1: Hypothetical 63.0 kDa protein in DAK1-ORC1 intergenic region

Chain B:



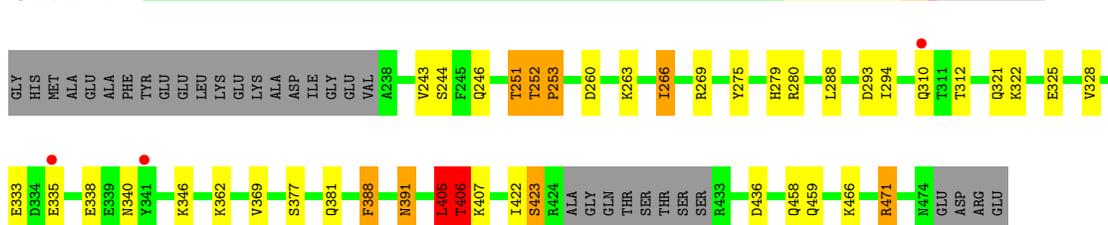
- Molecule 1: Hypothetical 63.0 kDa protein in DAK1-ORC1 intergenic region

Chain C:



- Molecule 1: Hypothetical 63.0 kDa protein in DAK1-ORC1 intergenic region

Chain D:



## 4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	57.87Å 157.53Å 57.79Å 90.00° 89.77° 90.00°	Depositor
Resolution (Å)	78.81 – 2.55 32.30 – 2.49	Depositor EDS
% Data completeness (in resolution range)	91.7 (78.81-2.55) 91.1 (32.30-2.49)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.08 (at 2.51Å)	Xtrriage
Refinement program	REFMAC 5.2.0005	Depositor
R, $R_{free}$	0.215 , 0.303 0.206 , 0.237	Depositor DCC
$R_{free}$ test set	1638 reflections (5.25%)	DCC
Wilson B-factor (Å <sup>2</sup> )	45.9	Xtrriage
Anisotropy	0.360	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.33 , 7.9	EDS
Estimated twinning fraction	0.000 for l,k,-h 0.460 for -h,-k,l 0.000 for -l,-k,-h	Xtrriage
L-test for twinning	$\langle  L  \rangle = 0.52$ , $\langle L^2 \rangle = 0.35$	Xtrriage
Outliers	2 of 32801 reflections (0.006%)	Xtrriage
$F_o, F_c$ correlation	0.95	EDS
Total number of atoms	7649	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	38.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 38.28 % of the origin peak, indicating pseudo translational symmetry. The chance of finding a peak of this or larger height randomly in a structure without pseudo translational symmetry is equal to 3.7782e-04. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

<sup>1</sup>Intensities estimated from amplitudes.

## 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	0.57	0/1926	0.70	1/2599 (0.0%)
1	B	0.63	0/1933	0.74	0/2609
1	C	0.62	0/1942	0.75	0/2621
1	D	0.57	0/1920	0.70	0/2592
All	All	0.60	0/7721	0.72	1/10421 (0.0%)

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	B	0	3
1	C	0	1
1	D	0	4
All	All	0	9

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	254	ARG	NE-CZ-NH1	5.29	122.94	120.30

There are no chirality outliers.

5 of 9 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	407	LYS	Peptide
1	B	339	GLU	Peptide
1	B	382	CYS	Peptide
1	B	405	LEU	Peptide

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Group
1	C	407	LYS	Peptide

## 5.2 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 hydrogens added by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, and the number in parentheses is this value normalized per 1000 atoms of the molecule in the chain. The Symm-Clashes column gives symmetry related clashes, in the same way as for the Clashes column.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	1888	0	0	8	0
1	B	1895	0	0	12	0
1	C	1904	0	0	17	0
1	D	1882	0	0	17	0
2	A	21	0	0	0	0
2	B	25	0	0	1	0
2	C	18	0	0	0	0
2	D	16	0	0	0	0
All	All	7649	0	0	50	0

Clashscore is defined as the number of clashes calculated for the entry per 1000 atoms (including hydrogens) of the entry. The overall clashscore for this entry is 7.

The worst 5 of 50 close contacts within the same asymmetric unit are listed below.

Atom-1	Atom-2	Distance(Å)	Clash(Å)
1:D:406:THR:CG2	1:D:407:LYS:N	2.30	0.94
1:C:294:ILE:CD1	1:C:295:HIS:CE1	2.51	0.93
1:B:339:GLU:OE2	1:B:339:GLU:CA	2.30	0.78
1:D:406:THR:O	1:D:407:LYS:CG	2.39	0.70
1:A:292:ASP:OD1	1:A:294:ILE:CG2	2.46	0.64

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	225/261 (86%)	197 (88%)	26 (12%)	2 (1%)	25	41
1	B	226/261 (87%)	203 (90%)	21 (9%)	2 (1%)	25	41
1	C	227/261 (87%)	212 (93%)	13 (6%)	2 (1%)	25	41
1	D	225/261 (86%)	210 (93%)	11 (5%)	4 (2%)	13	20
All	All	903/1044 (86%)	822 (91%)	71 (8%)	10 (1%)	21	34

5 of 10 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	406	THR
1	D	322	LYS
1	D	406	THR
1	A	292	ASP
1	C	366	ASP

### 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	211/237 (89%)	192 (91%)	19 (9%)	14	24
1	B	212/237 (90%)	182 (86%)	30 (14%)	5	8
1	C	213/237 (90%)	190 (89%)	23 (11%)	9	16
1	D	210/237 (89%)	186 (89%)	24 (11%)	8	14
All	All	846/948 (89%)	750 (89%)	96 (11%)	9	14

5 of 96 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	433	ARG
1	C	308	LYS
1	D	340	ASN
1	B	441	LEU
1	B	471	ARG

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 chains 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 ⓘ

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 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	229/261 (87%)	0.22	5 (2%) 59 62	15, 41, 75, 105	2 (0%)
1	B	230/261 (88%)	0.03	2 (0%) 81 83	9, 31, 65, 94	1 (0%)
1	C	231/261 (88%)	-0.02	1 (0%) 90 92	7, 30, 65, 100	2 (0%)
1	D	229/261 (87%)	0.13	3 (1%) 74 77	15, 41, 77, 113	0
All	All	919/1044 (88%)	0.09	11 (1%) 75 78	7, 36, 73, 113	5 (0%)

The worst 5 of 11 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	237	VAL	3.7
1	A	422	ILE	3.0
1	A	466	LYS	2.9
1	A	340	ASN	2.8
1	D	310	GLN	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

There are no such residues in this entry.