



# Full wwPDB X-ray Structure Validation Report ⓘ

Jan 31, 2016 – 10:39 PM GMT

PDB ID : 1UO1  
Title : STRUCTURE BASED ENGINEERING OF INTERNAL MOLECULAR SURFACES OF FOUR HELIX BUNDLES  
Authors : Yadav, M.K.; Redman, J.E.; Alvarez-Gutierrez, J.M.; Zhang, Y.; Stout, C.D.; Ghadiri, M.R.  
Deposited on : 2003-09-15  
Resolution : 2.50 Å(reported)

This is a Full wwPDB X-ray Structure Validation 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 : 1.7 (RC4), CSD as536be (2015)  
Xtriage (Phenix) : 1.9-1692  
EDS : rb-20026688  
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) : trunk26865

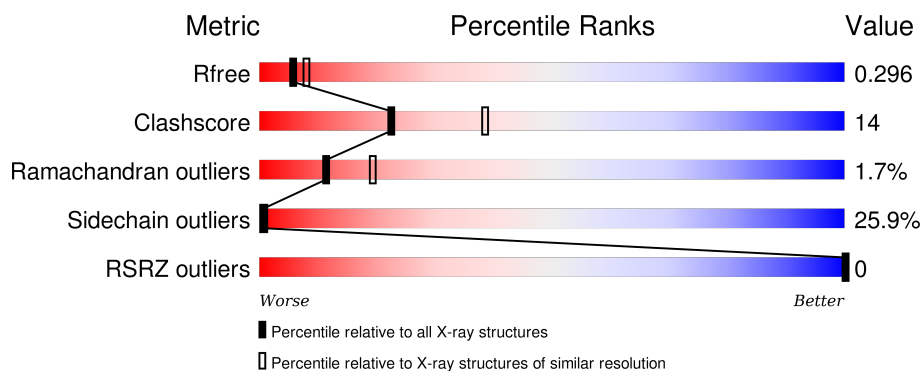
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

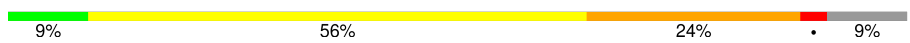
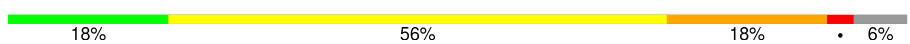
The reported resolution of this entry is 2.50 Å.

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	3553 (2.50-2.50)
Clashscore	102246	4242 (2.50-2.50)
Ramachandran outliers	100387	4156 (2.50-2.50)
Sidechain outliers	100360	4158 (2.50-2.50)
RSRZ outliers	91569	3562 (2.50-2.50)

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. 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	34	
1	B	34	

## 2 Entry composition

There is only 1 type of molecule in this entry. The entry contains 523 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 GENERAL CONTROL PROTEIN GCN4.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	31	Total	C	N	O	S	0	0	0
			261	166	46	48	1			
1	B	32	Total	C	N	O	S	0	0	1
			262	166	47	48	1			

There are 16 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	5	ILE	LEU	CONFLICT	UNP P03069
A	9	LEU	VAL	CONFLICT	UNP P03069
A	12	ILE	LEU	CONFLICT	UNP P03069
A	16	LEU	ASN	CONFLICT	UNP P03069
A	19	ILE	LEU	CONFLICT	UNP P03069
A	23	LEU	VAL	CONFLICT	UNP P03069
A	26	THR	LEU	ENGINEERED MUTATION	UNP P03069
A	30	LEU	VAL	CONFLICT	UNP P03069
B	5	ILE	LEU	CONFLICT	UNP P03069
B	9	LEU	VAL	CONFLICT	UNP P03069
B	12	ILE	LEU	CONFLICT	UNP P03069
B	16	LEU	ASN	CONFLICT	UNP P03069
B	19	ILE	LEU	CONFLICT	UNP P03069
B	23	LEU	VAL	CONFLICT	UNP P03069
B	26	THR	LEU	ENGINEERED MUTATION	UNP P03069
B	30	LEU	VAL	CONFLICT	UNP P03069

### 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: GENERAL CONTROL PROTEIN GCN4

Chain A: 



#### • Molecule 1: GENERAL CONTROL PROTEIN GCN4

Chain B: 



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 41 3 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	79.39Å 79.39Å 79.39Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	55.90 – 2.50 28.07 – 2.50	Depositor EDS
% Data completeness (in resolution range)	99.8 (55.90-2.50) 100.0 (28.07-2.50)	Depositor EDS
$R_{merge}$	0.09	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	5.36 (at 2.51Å)	Xtriage
Refinement program	REFMAC	Depositor
R, $R_{free}$	0.231 , 0.302 0.231 , 0.296	Depositor DCC
$R_{free}$ test set	141 reflections (4.53%)	DCC
Wilson B-factor (Å <sup>2</sup> )	35.1	Xtriage
Anisotropy	0.000	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.34 , 61.4	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.45$ , $\langle L^2 \rangle = 0.29$	Xtriage
Outliers	0 of 5594 reflections	Xtriage
$F_o, F_c$ correlation	0.92	EDS
Total number of atoms	523	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	35.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.77% 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

### 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	3.39	26/262 (9.9%)	2.30	12/346 (3.5%)
1	B	3.52	31/263 (11.8%)	2.70	24/348 (6.9%)
All	All	3.46	57/525 (10.9%)	2.51	36/694 (5.2%)

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	2

All (57) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	20	GLU	CD-OE2	19.35	1.47	1.25
1	B	11	GLU	CD-OE1	19.14	1.46	1.25
1	A	6	GLU	CD-OE2	14.91	1.42	1.25
1	B	22	GLU	CD-OE1	14.35	1.41	1.25
1	A	10	GLU	CD-OE2	11.96	1.38	1.25
1	B	6	GLU	CD-OE2	11.49	1.38	1.25
1	A	20	GLU	CD-OE1	11.45	1.38	1.25
1	B	17	TYR	CE1-CZ	-10.78	1.24	1.38
1	A	28	LYS	CD-CE	10.53	1.77	1.51
1	B	10	GLU	CD-OE1	9.52	1.36	1.25
1	A	8	LYS	CD-CE	9.37	1.74	1.51
1	B	25	ARG	NE-CZ	9.12	1.45	1.33
1	B	17	TYR	CG-CD2	-8.66	1.27	1.39
1	A	14	SER	CB-OG	-8.64	1.31	1.42
1	B	16	LEU	C-O	-8.60	1.07	1.23
1	B	20	GLU	CD-OE1	7.87	1.34	1.25
1	B	17	TYR	CB-CG	-7.84	1.39	1.51
1	B	22	GLU	CD-OE2	-7.81	1.17	1.25

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	8	LYS	CD-CE	7.62	1.70	1.51
1	B	26	THR	CA-CB	-7.36	1.34	1.53
1	B	25	ARG	CG-CD	7.23	1.70	1.51
1	A	22	GLU	CD-OE2	-7.22	1.17	1.25
1	A	17	TYR	CD2-CE2	7.18	1.50	1.39
1	B	20	GLU	CD-OE2	-7.16	1.17	1.25
1	B	26	THR	CA-C	-7.05	1.34	1.52
1	B	6	GLU	CG-CD	6.92	1.62	1.51
1	A	4	GLN	CD-OE1	6.87	1.39	1.24
1	B	23	LEU	CG-CD2	-6.84	1.26	1.51
1	A	28	LYS	CG-CD	6.80	1.75	1.52
1	B	28	LYS	CE-NZ	6.69	1.65	1.49
1	A	25	ARG	CB-CG	-6.66	1.34	1.52
1	B	17	TYR	CZ-OH	6.54	1.49	1.37
1	B	25	ARG	CD-NE	6.48	1.57	1.46
1	B	23	LEU	C-O	6.40	1.35	1.23
1	B	8	LYS	CG-CD	6.35	1.74	1.52
1	B	7	ASP	C-O	6.29	1.35	1.23
1	A	17	TYR	CG-CD2	-6.26	1.31	1.39
1	A	7	ASP	CB-CG	6.24	1.64	1.51
1	A	28	LYS	CE-NZ	6.19	1.64	1.49
1	B	28	LYS	CD-CE	6.08	1.66	1.51
1	A	24	ALA	CA-CB	6.07	1.65	1.52
1	A	8	LYS	C-O	-5.97	1.11	1.23
1	B	11	GLU	CG-CD	5.96	1.60	1.51
1	A	6	GLU	CD-OE1	5.93	1.32	1.25
1	B	8	LYS	CE-NZ	5.86	1.63	1.49
1	B	8	LYS	CB-CG	5.76	1.68	1.52
1	A	3	LYS	CA-C	-5.69	1.38	1.52
1	B	30	LEU	C-O	-5.65	1.12	1.23
1	A	11	GLU	C-O	-5.57	1.12	1.23
1	B	15	LYS	CE-NZ	5.44	1.62	1.49
1	A	21	ASN	C-O	-5.43	1.13	1.23
1	B	15	LYS	C-O	-5.38	1.13	1.23
1	A	18	HIS	CA-CB	-5.35	1.42	1.53
1	A	17	TYR	CZ-OH	5.10	1.46	1.37
1	A	9	LEU	CA-CB	5.07	1.65	1.53
1	A	6	GLU	CG-CD	5.06	1.59	1.51
1	A	26	THR	CB-CG2	5.02	1.69	1.52

All (36) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	31	GLY	O-C-N	-13.22	101.55	122.70
1	B	7	ASP	CB-CG-OD1	11.29	128.46	118.30
1	B	22	GLU	CG-CD-OE1	10.61	139.52	118.30
1	B	25	ARG	NE-CZ-NH2	9.70	125.15	120.30
1	B	13	LEU	CB-CG-CD2	-9.68	94.54	111.00
1	B	31	GLY	CA-C-N	8.99	136.98	117.20
1	A	16	LEU	CB-CG-CD2	-8.80	96.04	111.00
1	A	25	ARG	NE-CZ-NH1	-8.69	115.95	120.30
1	B	11	GLU	CG-CD-OE2	-8.55	101.20	118.30
1	B	16	LEU	CB-CG-CD2	-8.45	96.64	111.00
1	B	22	GLU	CG-CD-OE2	-8.30	101.71	118.30
1	A	22	GLU	OE1-CD-OE2	-8.17	113.50	123.30
1	B	10	GLU	OE1-CD-OE2	6.90	131.58	123.30
1	A	27	LYS	CD-CE-NZ	6.87	127.49	111.70
1	B	25	ARG	NH1-CZ-NH2	-6.44	112.31	119.40
1	A	17	TYR	CG-CD1-CE1	6.34	126.37	121.30
1	B	9	LEU	CB-CG-CD1	-6.33	100.23	111.00
1	A	10	GLU	CG-CD-OE1	-6.28	105.74	118.30
1	B	30	LEU	CB-CG-CD2	-6.18	100.50	111.00
1	B	28	LYS	CD-CE-NZ	5.93	125.33	111.70
1	A	14	SER	CB-CA-C	-5.91	98.87	110.10
1	B	11	GLU	CG-CD-OE1	5.82	129.95	118.30
1	B	2	MET	CG-SD-CE	5.74	109.39	100.20
1	B	22	GLU	OE1-CD-OE2	-5.65	116.52	123.30
1	A	22	GLU	CB-CA-C	-5.53	99.33	110.40
1	B	6	GLU	CG-CD-OE2	5.52	129.34	118.30
1	B	23	LEU	CA-CB-CG	5.46	127.86	115.30
1	B	10	GLU	CA-CB-CG	-5.43	101.46	113.40
1	A	21	ASN	CB-CA-C	-5.34	99.72	110.40
1	A	19	ILE	CA-CB-CG2	-5.34	100.22	110.90
1	A	4	GLN	CB-CG-CD	5.13	124.94	111.60
1	B	15	LYS	CB-CA-C	-5.11	100.19	110.40
1	B	17	TYR	CB-CG-CD1	-5.07	117.96	121.00
1	A	6	GLU	CB-CA-C	-5.05	100.30	110.40
1	B	17	TYR	CB-CG-CD2	5.02	124.01	121.00
1	B	20	GLU	N-CA-CB	-5.01	101.59	110.60

There are no chirality outliers.

All (2) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	B	26	THR	Mainchain
1	B	4	GLN	Mainchain



## 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	261	0	285	13	0
1	B	262	0	285	4	0
All	All	523	0	570	15	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 14.

All (15) 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:8:LYS:CD	1:A:8:LYS:CE	1.74	1.66
1:A:5:ILE:CG1	1:A:5:ILE:CD1	1.74	1.57
1:A:28:LYS:CE	1:A:28:LYS:CD	1.77	1.56
1:A:28:LYS:CG	1:A:28:LYS:CD	1.75	1.53
1:A:8:LYS:CD	1:A:8:LYS:NZ	2.55	0.70
1:B:1:ARG:HG2	1:B:1:ARG:HH11	1.55	0.70
1:A:3:LYS:HE2	1:A:7:ASP:OD1	1.92	0.69
1:A:8:LYS:CG	1:A:8:LYS:CE	2.67	0.69
1:A:5:ILE:CB	1:A:5:ILE:CD1	2.73	0.60
1:B:8:LYS:O	1:B:12:ILE:HG13	2.01	0.60
1:A:28:LYS:CD	1:A:28:LYS:CB	2.77	0.59
1:A:1:ARG:HG3	1:A:3:LYS:H	1.69	0.57
1:A:30:LEU:O	1:A:31:GLY:C	2.51	0.47
1:A:6:GLU:HG3	1:B:5:ILE:HG13	1.97	0.46
1:A:2:MET:HG3	1:B:5:ILE:CD1	2.47	0.44

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	29/34 (85%)	25 (86%)	4 (14%)	0	100	100
1	B	30/34 (88%)	24 (80%)	5 (17%)	1 (3%)	5	6
All	All	59/68 (87%)	49 (83%)	9 (15%)	1 (2%)	11	19

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	6	GLU

### 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	29/31 (94%)	22 (76%)	7 (24%)	1	1
1	B	29/31 (94%)	21 (72%)	8 (28%)	0	0
All	All	58/62 (94%)	43 (74%)	15 (26%)	0	1

All (15) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	2	MET
1	A	4	GLN
1	A	13	LEU
1	A	14	SER
1	A	23	LEU
1	A	28	LYS
1	A	30	LEU
1	B	2	MET
1	B	3	LYS
1	B	8	LYS
1	B	15	LYS
1	B	23	LEU

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	B	27	LYS
1	B	29	LEU
1	B	30	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
1	B	21	ASN

### 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	31/34 (91%)	-0.36	0 <a href="#">100</a> <a href="#">100</a>	17, 30, 57, 62	0
1	B	32/34 (94%)	-0.33	0 <a href="#">100</a> <a href="#">100</a>	15, 30, 64, 66	0
All	All	63/68 (92%)	-0.34	0 <a href="#">100</a> <a href="#">100</a>	15, 30, 62, 66	0

There are no RSRZ outliers to report.

### 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.