



# Full wwPDB X-ray Structure Validation Report ⓘ

Feb 13, 2017 – 02:03 pm GMT

PDB ID : 5DY8  
Title : Crystal Structure of the ER-alpha Ligand-binding Domain in Complex with the Cyclofenil Derivative 4,4'-{[(3S)-3-ethylcyclohexylidene]methanediyl}diphenol  
Authors : Nwachukwu, J.C.; Srinivasan, S.; Zheng, Y.; Wang, S.; Min, J.; Dong, C.; Liao, Z.; Cavett, V.; Nowak, J.; Houtman, R.; Carlson, K.E.; Josan, J.S.; Elemento, O.; Katzenellenbogen, J.A.; Zhou, H.B.; Nettles, K.W.  
Deposited on : 2015-09-24  
Resolution : 2.03 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	:	1.7.2 (RC1), CSD as538be (2017)
Xtriage (Phenix)	:	1.9-1692
EDS	:	trunk28620
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)	:	recalc28949

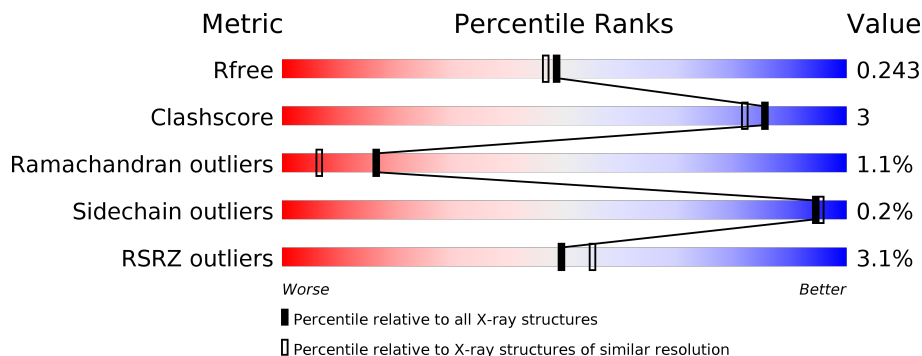
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 2.03 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))
$R_{free}$	100719	1316 (2.04-2.04)
Clashscore	112137	1394 (2.04-2.04)
Ramachandran outliers	110173	1383 (2.04-2.04)
Sidechain outliers	110143	1383 (2.04-2.04)
RSRZ outliers	101464	1319 (2.04-2.04)

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	257	<div> <div>3%</div> <div>83%</div> <div>7%</div> <div>9%</div> </div>
1	B	257	<div> <div>3%</div> <div>87%</div> <div>9%</div> </div>
2	C	14	<div> <div>7%</div> <div>50%</div> <div>21%</div> <div>29%</div> </div>
2	D	14	<div> <div>64%</div> <div>36%</div> </div>

## 2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 4038 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 Estrogen receptor.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	233	Total	C	N	O	S	0	0	0
			1818	1163	314	322	19			
1	B	235	Total	C	N	O	S	0	0	0
			1809	1157	313	321	18			

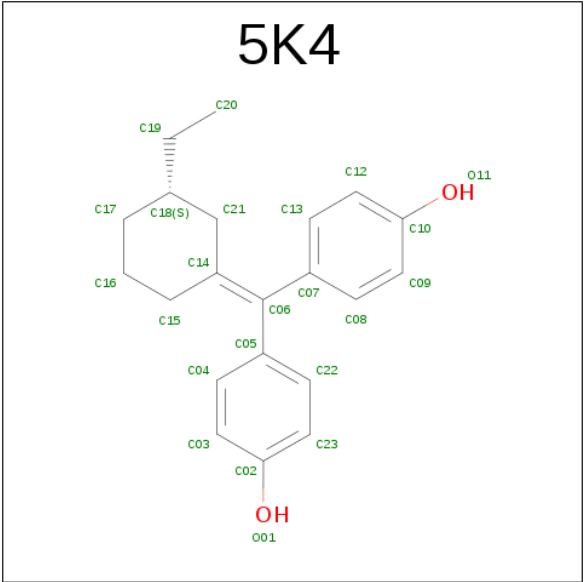
There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	537	SER	TYR	engineered mutation	UNP P03372
B	537	SER	TYR	engineered mutation	UNP P03372

- Molecule 2 is a protein called Nuclear receptor coactivator 2.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
2	C	10	Total	C	N	O	0	0	0
			89	57	19	13			
2	D	9	Total	C	N	O	0	0	0
			79	51	16	12			

- Molecule 3 is 4,4'-{[(3S)-3-ethylcyclohexylidene]methanediyl}diphenol (three-letter code: 5K4) (formula: C<sub>21</sub>H<sub>24</sub>O<sub>2</sub>).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
3	A	1	Total	C	O	0	0
			23	21	2		
3	B	1	Total	C	O	0	0
			23	21	2		

- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	106	Total	O	0	0
			106	106		
4	B	90	Total	O	0	0
			90	90		
4	C	1	Total	O	0	0
			1	1		



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	54.38Å 81.28Å 58.23Å 90.00° 110.28° 90.00°	Depositor
Resolution (Å)	46.09 – 2.03 46.09 – 2.03	Depositor EDS
% Data completeness (in resolution range)	93.2 (46.09-2.03) 93.0 (46.09-2.03)	Depositor EDS
$R_{merge}$	0.07	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.82 (at 2.03Å)	Xtriage
Refinement program	PHENIX	Depositor
R, $R_{free}$	0.189 , 0.243 0.190 , 0.243	Depositor DCC
$R_{free}$ test set	1915 reflections (6.69%)	DCC
Wilson B-factor (Å <sup>2</sup> )	32.1	Xtriage
Anisotropy	0.242	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.34 , 50.5	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.50$ , $\langle L^2 \rangle = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.96	EDS
Total number of atoms	4038	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	43.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.86% 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](#)

Bond lengths and bond angles in the following residue types are not validated in this section: 5K4

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	0.21	0/1852	0.37	0/2510
1	B	0.21	0/1841	0.38	0/2495
2	C	0.19	0/90	0.42	0/119
2	D	0.18	0/79	0.32	0/104
All	All	0.21	0/3862	0.38	0/5228

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	1818	0	1835	11	0
1	B	1809	0	1804	9	0
2	C	89	0	95	2	0
2	D	79	0	88	0	0
3	A	23	0	0	0	0
3	B	23	0	0	0	0
4	A	106	0	0	2	0
4	B	90	0	0	2	0
4	C	1	0	0	0	0
All	All	4038	0	3822	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 3.

All (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)
1:A:417:CYS:SG	4:A:1082:HOH:O	2.31	0.89
1:B:533:VAL:HG22	1:B:534:VAL:HG23	1.60	0.83
1:B:534:VAL:O	4:B:1001:HOH:O	2.04	0.76
1:B:374:ASP:OD2	4:B:1002:HOH:O	2.09	0.71
1:A:473:ASP:OD2	1:A:477:ARG:NH1	2.34	0.61
1:B:456:SER:HA	1:B:515:ARG:NH2	2.23	0.53
1:A:339:GLU:HA	1:A:418:VAL:HG22	1.91	0.52
1:A:400:GLY:O	4:A:1001:HOH:O	2.20	0.49
2:C:688:LYS:HB2	2:C:691:HIS:ND1	2.28	0.48
1:B:539:LEU:O	1:B:543:MET:HG2	2.16	0.46
1:B:473:ASP:O	1:B:477:ARG:HG3	2.17	0.44
1:A:498:GLN:HA	1:A:501:HIS:NE2	2.33	0.43
1:A:456:SER:HA	1:A:515:ARG:NH2	2.33	0.43
1:B:532:ASN:O	1:B:534:VAL:N	2.52	0.43
1:A:376:VAL:HG13	2:C:690:LEU:HD23	2.00	0.42
1:A:498:GLN:HA	1:A:501:HIS:CD2	2.55	0.42
1:A:459:TYR:HE2	1:B:513:HIS:ND1	2.17	0.42
1:B:531:LYS:O	1:B:533:VAL:N	2.42	0.42
1:A:539:LEU:O	1:A:543:MET:HG2	2.20	0.42
1:A:474:HIS:O	1:A:478:VAL:HG23	2.21	0.41

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	229/257 (89%)	221 (96%)	4 (2%)	4 (2%)	<b>11</b> <b>2</b>

*Continued on next page...*

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	B	229/257 (89%)	223 (97%)	5 (2%)	1 (0%)	38	27
2	C	8/14 (57%)	8 (100%)	0	0	100	100
2	D	7/14 (50%)	7 (100%)	0	0	100	100
All	All	473/542 (87%)	459 (97%)	9 (2%)	5 (1%)	17	6

All (5) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	418	VAL
1	B	533	VAL
1	A	331	TYR
1	A	417	CYS
1	A	332	ASP

### 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	200/232 (86%)	199 (100%)	1 (0%)	91	91
1	B	194/232 (84%)	194 (100%)	0	100	100
2	C	10/14 (71%)	10 (100%)	0	100	100
2	D	9/14 (64%)	9 (100%)	0	100	100
All	All	413/492 (84%)	412 (100%)	1 (0%)	94	95

All (1) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	533	VAL

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

Mol	Chain	Res	Type
1	A	498	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](#)

2 ligands are modelled in this entry.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the chemical component dictionary. The Link column lists molecule types, if any, to which the group is linked. 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  $|Z| > 2$  is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\# Z  > 2$	Counts	RMSZ	$\# Z  > 2$
3	5K4	A	900	-	25,25,25	0.73	0	31,34,34	0.79	1 (3%)
3	5K4	B	901	-	25,25,25	0.72	0	31,34,34	0.90	1 (3%)

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the chemical component dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
3	5K4	A	900	-	-	0/14/24/24	0/3/3/3
3	5K4	B	901	-	-	0/14/24/24	0/3/3/3

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	B	901	5K4	C07-C06-C14	-2.83	119.32	122.71
3	A	900	5K4	C21-C14-C06	-2.03	121.54	124.32

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 [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	233/257 (90%)	-0.17	7 (3%) 51 56	22, 38, 71, 93	0
1	B	235/257 (91%)	-0.17	7 (2%) 51 56	25, 42, 85, 110	0
2	C	10/14 (71%)	0.05	1 (10%) 8 8	37, 41, 75, 81	0
2	D	9/14 (64%)	-0.01	0 100 100	41, 52, 64, 72	0
All	All	487/542 (89%)	-0.16	15 (3%) 49 55	22, 40, 76, 110	0

All (15) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	459	TYR	9.5
1	A	459	TYR	5.8
1	B	533	VAL	5.4
1	A	420	GLY	3.4
1	B	549	LEU	3.2
1	B	420	GLY	3.0
1	A	533	VAL	2.7
1	B	458	VAL	2.6
1	B	460	THR	2.5
1	A	472	LYS	2.3
1	A	532	ASN	2.3
2	C	687	HIS	2.2
1	A	526	TYR	2.1
1	B	419	GLU	2.1
1	A	306	LEU	2.0

### 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
3	5K4	B	901	23/23	0.96	0.14	0.59	28,33,45,49	0
3	5K4	A	900	23/23	0.94	0.12	-0.11	28,34,44,52	0

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

There are no such residues in this entry.