



# wwPDB X-ray Structure Validation Summary Report ⓘ

Oct 14, 2021 – 02:09 PM EDT

PDB ID : 2R5Y  
Title : Structure of Scr/Exd complex bound to a consensus Hox-Exd site  
Authors : Aggarwal, A.K.; Passner, J.M.; Jain, R.  
Deposited on : 2007-09-04  
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

<https://www.wwpdb.org/validation/2017/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
Xtriage (Phenix)	:	1.13
EDS	:	2.23.2
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.23.2

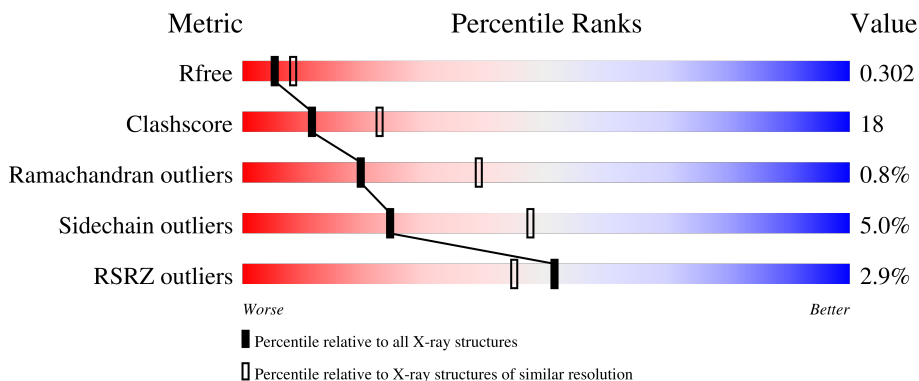
# 1 Overall quality at a glance

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))
$R_{free}$	130704	3163 (2.60-2.60)
Clashscore	141614	3518 (2.60-2.60)
Ramachandran outliers	138981	3455 (2.60-2.60)
Sidechain outliers	138945	3455 (2.60-2.60)
RSRZ outliers	127900	3104 (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 of the lower bar indicate the fraction of residues that contain outliers for  $\geq 3$ , 2, 1 and 0 types of geometric quality criteria respectively. 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	C	20	<div> <div>5%</div> <div>45%</div> <div>55%</div> </div>
2	D	20	<div> <div>40%</div> <div>60%</div> </div>
3	A	88	<div> <div>2%</div> <div>55%</div> <div>25%</div> <div>17%</div> </div>
4	B	63	<div> <div>3%</div> <div>70%</div> <div>22%</div> <div>6%</div> </div>

## 2 Entry composition

There are 5 unique types of molecules in this entry. The entry contains 2066 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 DNA chain called DNA (5'-D(\*DAP\*DCP\*DTP\*DCP\*DTP\*DAP\*DTP\*DG\*GP\*DAP\*DTP\*DTP\*DTP\*DAP\*DTP\*DGP\*DGP\*DGP\*DCP\*DTP\*DG)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	C	20	Total	C	N	O	P	0	0	0
			408	197	70	122	19			

- Molecule 2 is a DNA chain called DNA (5'-D(\*DTP\*DCP\*DAP\*DGP\*DCP\*DCP\*DCP\*DAP\*DTP\*DAP\*DAP\*DAP\*DTP\*DCP\*DAP\*DTP\*DAP\*DGP\*DAP\*DG)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	D	20	Total	C	N	O	P	0	0	0
			406	195	78	114	19			

- Molecule 3 is a protein called Homeotic protein Sex combs reduced.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	A	73	Total	C	N	O	S	0	0	0
			652	417	130	103	2			

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	74	GLY	-	expression tag	UNP P09077
A	139	SER	CYS	engineered mutation	UNP P09077

- Molecule 4 is a protein called Homeobox protein extradenticle.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	B	59	Total	C	N	O	S	0	0	0
			497	313	91	92	1			

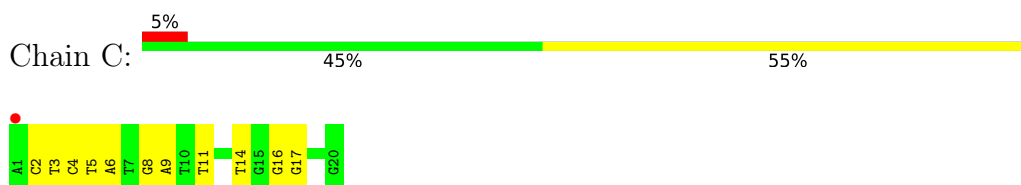
- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	C	15	Total 15	O 15	0	0
5	D	17	Total 17	O 17	0	0
5	A	33	Total 33	O 33	0	0
5	B	38	Total 38	O 38	0	0

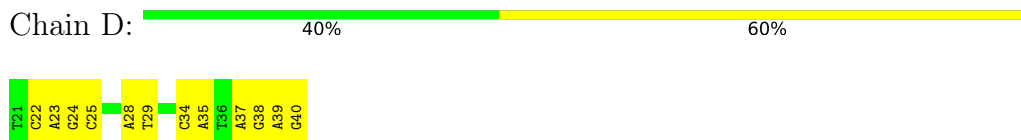
### 3 Residue-property plots

These plots are drawn for all protein, RNA, DNA and oligosaccharide 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.

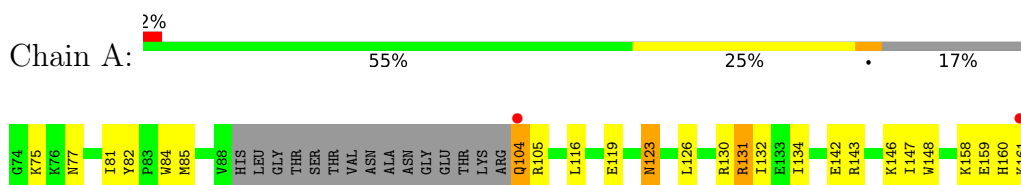
- Molecule 1: DNA (5'-D(\*DAP\*DCP\*DTP\*DCP\*DTP\*DAP\*DTP\*DGP\*DAP\*DTP\*DTP\*DT\*  
TP\*DAP\*DTP\*DGP\*DGP\*DGP\*DCP\*DTP\*DG)-3')



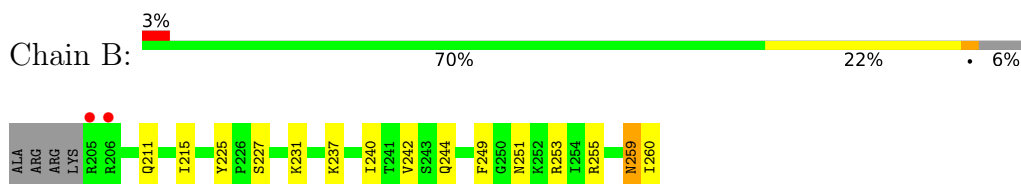
- Molecule 2: DNA (5'-D(\*DTP\*DCP\*DAP\*DGP\*DCP\*DCP\*DCP\*DAP\*DTP\*DAP\*DAP\*D  
AP\*DTP\*DCP\*DAP\*DTP\*DAP\*DGP\*DAP\*DG)-3')



- Molecule 3: Homeotic protein Sex combs reduced



- Molecule 4: Homeobox protein extradenticle



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 43 21 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	65.31Å 65.31Å 200.30Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	11.96 – 2.60 19.96 – 2.60	Depositor EDS
% Data completeness (in resolution range)	96.3 (11.96-2.60) 96.3 (19.96-2.60)	Depositor EDS
$R_{merge}$	0.10	Depositor
$R_{sym}$	0.10	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	7.89 (at 2.59Å)	Xtriage
Refinement program	CNS 1.1	Depositor
R, $R_{free}$	0.255 , 0.299 0.254 , 0.302	Depositor DCC
$R_{free}$ test set	1076 reflections (7.67%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	58.7	Xtriage
Anisotropy	0.015	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.29 , 41.1	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.91	EDS
Total number of atoms	2066	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	57.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.23% 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](#)

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	C	0.40	0/456	0.71	0/703
2	D	0.47	0/456	0.77	0/701
3	A	0.39	0/669	0.56	0/894
4	B	0.42	0/508	0.57	0/679
All	All	0.42	0/2089	0.65	0/2977

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
2	D	0	1

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
2	D	37	DA	Sidechain

### 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	C	408	0	230	11	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	D	406	0	226	9	0
3	A	652	0	656	39	0
4	B	497	0	487	12	0
5	A	33	0	0	4	0
5	B	38	0	0	2	0
5	C	15	0	0	3	0
5	D	17	0	0	1	0
All	All	2066	0	1599	63	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 18.

The worst 5 of 63 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)
3:A:131:ARG:HE	3:A:142:GLU:HG2	1.39	0.86
3:A:159:GLU:O	3:A:161:LYS:N	2.13	0.81
5:C:808:HOH:O	3:A:147:ILE:HD12	1.81	0.79
2:D:38:DG:H2''	2:D:39:DA:N7	2.03	0.73
3:A:131:ARG:HG3	3:A:131:ARG:HH11	1.55	0.72

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	
3	A	69/88 (78%)	64 (93%)	4 (6%)	1 (1%)	11	22
4	B	57/63 (90%)	54 (95%)	3 (5%)	0	100	100
All	All	126/151 (83%)	118 (94%)	7 (6%)	1 (1%)	19	39

All (1) Ramachandran outliers are listed below:



Mol	Chain	Res	Type
3	A	160	HIS

### 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	
3	A	67/82 (82%)	63 (94%)	4 (6%)	19	39
4	B	54/57 (95%)	52 (96%)	2 (4%)	34	60
All	All	121/139 (87%)	115 (95%)	6 (5%)	24	47

5 of 6 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	A	131	ARG
4	B	242	VAL
4	B	259	ASN
3	A	123	ASN
3	A	104	GLN

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 6 such sidechains are listed below:

Mol	Chain	Res	Type
4	B	217	ASN
4	B	251	ASN
4	B	259	ASN
3	A	123	ASN
3	A	104	GLN

### 5.3.3 RNA ⓘ

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 monosaccharides 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	C	20/20 (100%)	-0.15	1 (5%) 28 23	35, 58, 102, 103	0
2	D	20/20 (100%)	-0.25	0 100 100	36, 58, 95, 95	0
3	A	73/88 (82%)	0.10	2 (2%) 54 48	37, 58, 86, 103	0
4	B	59/63 (93%)	-0.32	2 (3%) 45 38	30, 42, 59, 91	0
All	All	172/191 (90%)	-0.12	5 (2%) 51 45	30, 52, 92, 103	0

All (5) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
4	B	206	ARG	3.7
3	A	104	GLN	2.6
1	C	1	DA	2.5
3	A	161	LYS	2.5
4	B	205	ARG	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 monosaccharides 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.