



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 13, 2017 – 02:31 am GMT

PDB ID : 2VBJ  
Title : Molecular basis of human XPC gene recognition and cleavage by engineered homing endonuclease heterodimers  
Authors : Redondo, P.; Prieto, J.; Munoz, I.G.; Alibes, A.; Stricher, F.; Serrano, L.; Arnould, S.; Perez, C.; Cabaniols, J.P.; Duchateau, P.; Paques, F.; Blanco, F.J.; Montoya, G.  
Deposited on : 2007-09-14  
Resolution : 1.95 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
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

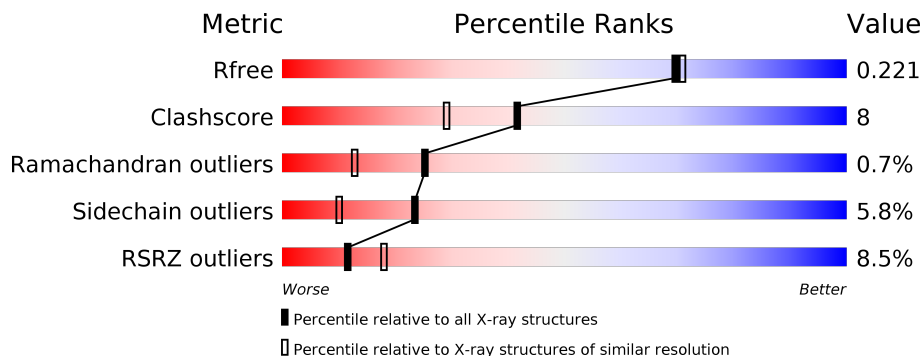
# 1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 1.95 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	2004 (1.96-1.96)
Clashscore	112137	2136 (1.96-1.96)
Ramachandran outliers	110173	2117 (1.96-1.96)
Sidechain outliers	110143	2117 (1.96-1.96)
RSRZ outliers	101464	2018 (1.96-1.96)

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	152	<div> <div>11%</div> <div> <div></div> <div>80%</div> <div>16%</div> <div>.</div> </div> </div>
2	B	152	<div> <div>9%</div> <div> <div></div> <div>80%</div> <div>19%</div> <div>.</div> </div> </div>
3	C	24	<div> <div>8%</div> <div> <div></div> <div>71%</div> <div>21%</div> </div> </div>
4	E	24	<div> <div>13%</div> <div> <div></div> <div>67%</div> <div>21%</div> </div> </div>

## 2 Entry composition

There are 6 unique types of molecules in this entry. The entry contains 3766 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 DNA ENDONUCLEASE I-CREI.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	152	Total	C	N	O	S	0	0	0
			1232	792	210	229	1			

There are 11 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	28	GLU	LYS	CONFLICT	UNP P05725
A	38	ARG	GLN	CONFLICT	UNP P05725
A	40	LYS	SER	CONFLICT	UNP P05725
A	42	THR	ALA	CONFLICT	UNP P05725
A	44	LYS	GLN	CONFLICT	UNP P05725
A	70	GLU	ARG	CONFLICT	UNP P05725
A	75	ASN	ASP	CONFLICT	UNP P05725
A	85	ARG	HIS	CONFLICT	UNP P05725
A	109	THR	ILE	CONFLICT	UNP P05725
A	110	GLU	TRP	CONFLICT	UNP P05725
A	111	GLN	ARG	CONFLICT	UNP P05725

- Molecule 2 is a protein called DNA ENDONUCLEASE I-CREI.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	152	Total	C	N	O	S	0	0	0
			1224	787	210	226	1			

There are 10 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	19	ALA	GLY	CONFLICT	UNP P05725
B	28	GLU	LYS	CONFLICT	UNP P05725
B	33	SER	TYR	CONFLICT	UNP P05725
B	38	ARG	GLN	CONFLICT	UNP P05725
B	40	LYS	SER	CONFLICT	UNP P05725

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
B	42	THR	ALA	CONFLICT	UNP P05725
B	70	SER	ARG	CONFLICT	UNP P05725
B	75	ASN	ASP	CONFLICT	UNP P05725
B	110	GLU	TRP	CONFLICT	UNP P05725
B	111	GLN	ARG	CONFLICT	UNP P05725

- Molecule 3 is a DNA chain called 5'-D(\*TP\*CP\*TP\*GP\*CP\*CP\*TP\*TP\*TP\*TP \*TP\*T  
P\*GP\*AP\*AP\*GP\*GP\*AP\*TP\*CP\*CP\*TP\*AP\*A)-3'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	C	24	Total	C	N	O	P	0	0	0
			485	235	80	147	23			

- Molecule 4 is a DNA chain called 5'-D(\*TP\*TP\*AP\*GP\*GP\*AP\*TP\*CP\*CP\*TP \*TP\*C  
P\*AP\*AP\*AP\*AP\*AP\*AP\*GP\*GP\*CP\*AP\*GP\*A)-3'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	E	24	Total	C	N	O	P	0	0	0
			493	236	97	137	23			

- Molecule 5 is CALCIUM ION (three-letter code: CA) (formula: Ca).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	2	Total	Ca	0	0
			2	2		

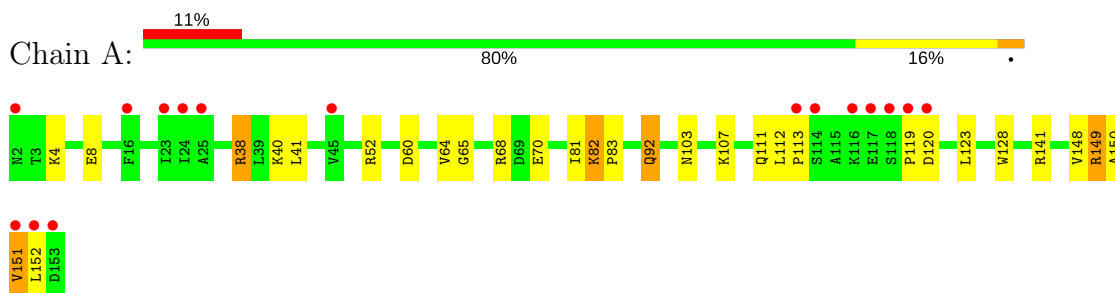
- Molecule 6 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	A	89	Total	O	0	0
			89	89		
6	B	110	Total	O	0	0
			110	110		
6	C	74	Total	O	0	0
			74	74		
6	E	57	Total	O	0	0
			57	57		

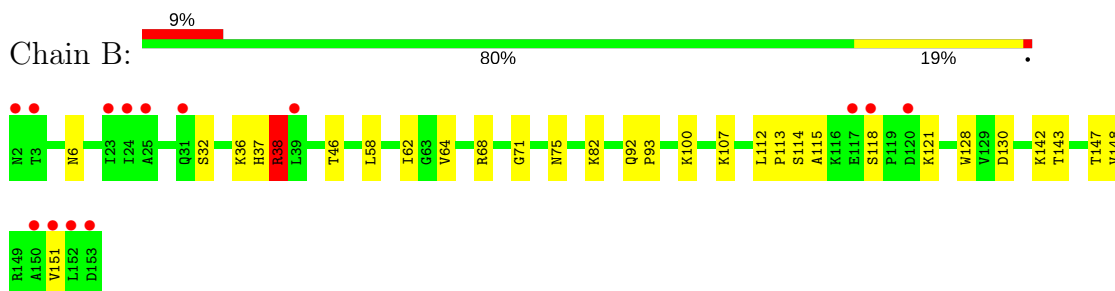
### 3 Residue-property plots [i](#)

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of 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.

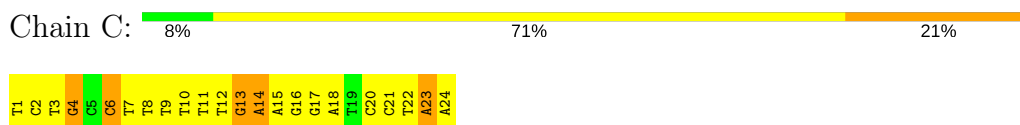
- Molecule 1: DNA ENDONUCLEASE I-CREI



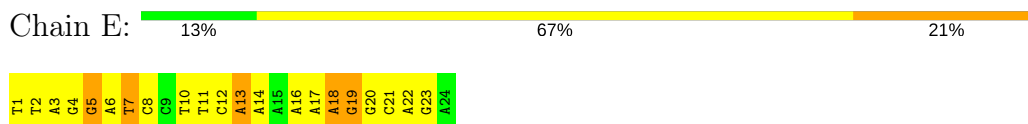
- Molecule 2: DNA ENDONUCLEASE I-CREI



- Molecule 3: 5'-D(\*TP\*CP\*TP\*GP\*CP\*CP\*TP\*TP\*TP\*TP \*TP\*TP\*GP\*AP\*AP\*GP\*GP\*AP\*TP\*CP\*CP\*TP\*AP\*A)-3'



- Molecule 4: 5'-D(\*TP\*TP\*AP\*GP\*GP\*AP\*TP\*CP\*CP\*TP \*TP\*CP\*AP\*AP\*AP\*AP\*AP\*AP\*GP\*GP\*CP\*AP\*GP\*A)-3'



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	44.55Å 68.70Å 88.73Å 90.00° 95.36° 90.00°	Depositor
Resolution (Å)	32.88 – 1.95 32.87 – 1.95	Depositor EDS
% Data completeness (in resolution range)	96.7 (32.88-1.95) 96.7 (32.87-1.95)	Depositor EDS
$R_{merge}$	0.05	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	7.95 (at 1.95Å)	Xtriage
Refinement program	REFMAC 5.2.0019	Depositor
R, $R_{free}$	0.147 , 0.216 0.158 , 0.221	Depositor DCC
$R_{free}$ test set	1950 reflections (5.46%)	DCC
Wilson B-factor (Å <sup>2</sup> )	27.3	Xtriage
Anisotropy	0.194	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.39 , 62.3	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.48$ , $\langle L^2 \rangle = 0.31$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.96	EDS
Total number of atoms	3766	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	42.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.37% 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: CA

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	0.63	0/1254	0.85	2/1692 (0.1%)
2	B	0.70	0/1246	0.89	4/1681 (0.2%)
3	C	1.13	0/541	2.49	48/833 (5.8%)
4	E	1.16	0/555	2.42	46/855 (5.4%)
All	All	0.84	0/3596	1.59	100/5061 (2.0%)

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

There are no bond length outliers.

The worst 5 of 100 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
4	E	13	DA	O4'-C1'-N9	-12.51	99.24	108.00
3	C	10	DT	O4'-C1'-N1	-12.01	99.59	108.00
3	C	13	DG	O4'-C1'-N9	-10.04	100.97	108.00
4	E	13	DA	O4'-C4'-C3'	-9.81	100.11	106.00
3	C	13	DG	N1-C6-O6	-9.46	114.22	119.90

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	149	ARG	Peptide

## 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	1232	0	1263	25	0
2	B	1224	0	1261	20	1
3	C	485	0	276	8	1
4	E	493	0	271	5	0
5	A	2	0	0	0	0
6	A	89	0	0	4	0
6	B	110	0	0	3	0
6	C	74	0	0	3	0
6	E	57	0	0	1	0
All	All	3766	0	3071	52	1

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 8.

The worst 5 of 52 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)
2:B:37:HIS:HE1	2:B:151:VAL:HG21	1.15	1.11
1:A:150:ALA:HB3	6:A:1088:HOH:O	1.58	1.02
3:C:23:DA:H2'	3:C:24:DA:C8	1.95	1.00
2:B:37:HIS:CE1	2:B:151:VAL:HG21	2.01	0.95
1:A:68:ARG:NH1	4:E:7:DT:OP2	2.05	0.89

All (1) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
2:B:100:LYS:NZ	3:C:24:DA:OP2[2_655]	1.92	0.28



## 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	150/152 (99%)	140 (93%)	9 (6%)	1 (1%)	25	13
2	B	150/152 (99%)	143 (95%)	6 (4%)	1 (1%)	25	13
All	All	300/304 (99%)	283 (94%)	15 (5%)	2 (1%)	25	13

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	149	ARG
2	B	71	GLY

### 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	137/139 (99%)	127 (93%)	10 (7%)	16	5
2	B	137/139 (99%)	131 (96%)	6 (4%)	33	18
All	All	274/278 (99%)	258 (94%)	16 (6%)	23	10

5 of 16 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	92	GLN
1	A	123	LEU
2	B	38	ARG
1	A	82	LYS

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
2	B	58	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (4) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	75	ASN
2	B	37	HIS
2	B	75	ASN
2	B	111	GLN

### 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	152/152 (100%)	0.40	16 (10%) 7 11	24, 40, 77, 88	0
2	B	152/152 (100%)	0.26	14 (9%) 10 16	23, 34, 70, 81	0
3	C	24/24 (100%)	-0.28	0 100 100	27, 41, 58, 60	0
4	E	24/24 (100%)	-0.41	0 100 100	27, 43, 49, 50	0
All	All	352/352 (100%)	0.24	30 (8%) 11 18	23, 38, 70, 88	0

The worst 5 of 30 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	151	VAL	5.0
1	A	117	GLU	3.7
2	B	153	ASP	3.7
2	B	2	ASN	3.6
1	A	24	ILE	3.5

### 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
5	CA	A	1155	1/1	1.00	0.12	0.15	25,25,25,25	0
5	CA	A	1154	1/1	1.00	0.10	-1.01	25,25,25,25	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.