



# wwPDB X-ray Structure Validation Summary Report ⓘ

Jan 31, 2016 – 09:11 PM GMT

PDB ID : 1NWQ  
Title : CRYSTAL STRUCTURE OF C/EBPALPHA-DNA COMPLEX  
Authors : Miller, M.; Shuman, J.D.; Sebastian, T.; Dauter, Z.; Johnson, P.F.  
Deposited on : 2003-02-06  
Resolution : 2.80 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  
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

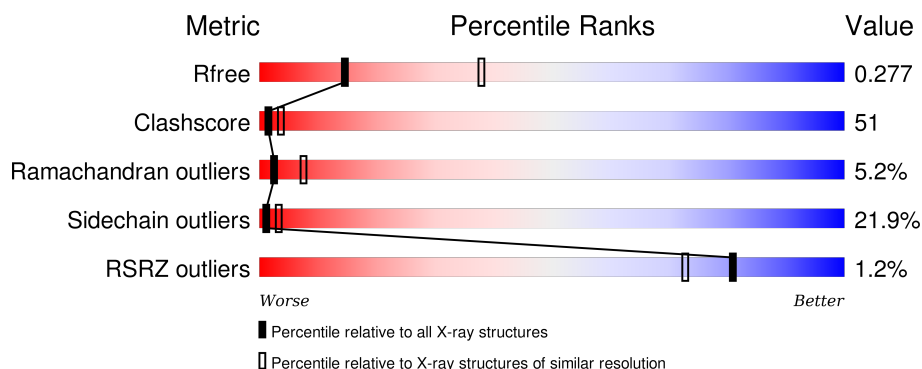
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 2.80 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	2393 (2.80-2.80)
Clashscore	102246	2827 (2.80-2.80)
Ramachandran outliers	100387	2782 (2.80-2.80)
Sidechain outliers	100360	2784 (2.80-2.80)
RSRZ outliers	91569	2404 (2.80-2.80)

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	B	21	
2	D	21	
3	A	62	
3	C	62	

## 2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 1909 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 5'-D(\*TP\*TP\*CP\*CP\*TP\*AP\*TP\*TP\*GP\*CP\*GP\*CP\*AP\*AP\*TP\*CP\*CP\*AP\*GP\*TP\*T)-3'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	B	21	Total	C	N	O	P	0	0	0
			421	204	69	128	20			

- Molecule 2 is a DNA chain called 5'-D(\*AP\*AP\*AP\*CP\*TP\*GP\*GP\*AP\*TP\*TP\*GP\*C P\*GP\*CP\*AP\*AP\*TP\*AP\*GP\*GP\*A)-3'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	D	21	Total	C	N	O	P	0	0	0
			434	207	87	120	20			

- Molecule 3 is a protein called CCAAT/enhancer binding protein alpha.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
3	A	60	Total	C	N	O	0	0	0
			510	300	111	99			
3	C	60	Total	C	N	O	0	0	0
			510	300	111	99			

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	279	GLY	-	CLONING ARTIFACT	UNP P05554
A	280	SER	-	CLONING ARTIFACT	UNP P05554
C	279	GLY	-	CLONING ARTIFACT	UNP P05554
C	280	SER	-	CLONING ARTIFACT	UNP P05554

- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	6	Total	O	0	0
			6	6		

*Continued on next page...*

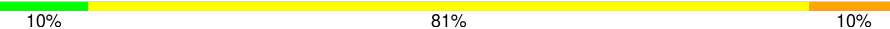
*Continued from previous page...*

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	B	6	Total 6	O 6	0	0
4	C	17	Total 17	O 17	0	0
4	D	5	Total 5	O 5	0	0

### 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: 5'-D(\*TP\*TP\*CP\*CP\*TP\*AP\*TP\*TP\*GP\*CP\*GP\*CP\*AP\*AP\*TP\*CP\*CP\*AP\*GP\*TP\*T)-3'

Chain B: 



- Molecule 2: 5'-D(\*AP\*AP\*AP\*CP\*TP\*GP\*GP\*AP\*TP\*TP\*GP\*CP\*GP\*CP\*AP\*AP\*TP\*AP\*GP\*GP\*A)-3'

Chain D: 



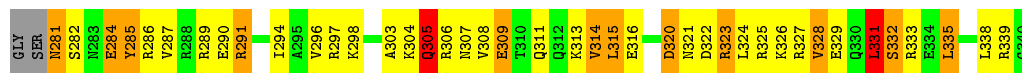
- Molecule 3: CCAAT/enhancer binding protein alpha

Chain A: 



- Molecule 3: CCAAT/enhancer binding protein alpha

Chain C: 



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 2 <sub>1</sub> 2 <sub>1</sub> 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	140.89 Å   53.09 Å   67.41 Å 90.00°   90.00°   90.00°	Depositor
Resolution (Å)	38.53 – 2.80 38.53 – 2.80	Depositor EDS
% Data completeness (in resolution range)	95.4 (38.53-2.80) 95.4 (38.53-2.80)	Depositor EDS
$R_{merge}$	0.08	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.26 (at 2.81 Å)	Xtriage
Refinement program	CNS 1.0	Depositor
R, $R_{free}$	0.221   ,   0.275 0.221   ,   0.277	Depositor DCC
$R_{free}$ test set	630 reflections (5.07%)	DCC
Wilson B-factor (Å <sup>2</sup> )	88.9	Xtriage
Anisotropy	0.582	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.26 , 58.5	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.50$ , $\langle L^2 \rangle = 0.33$	Xtriage
Outliers	0 of 12416 reflections	Xtriage
$F_o, F_c$ correlation	0.96	EDS
Total number of atoms	1909	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	92.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 14.21% 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	B	1.27	1/469 (0.2%)	1.12	0/721
2	D	1.20	0/489	1.07	2/754 (0.3%)
3	A	1.11	0/510	1.14	1/676 (0.1%)
3	C	0.97	0/510	1.23	4/676 (0.6%)
All	All	1.14	1/1978 (0.1%)	1.14	7/2827 (0.2%)

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	1
2	D	0	1
3	C	0	1
All	All	0	3

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	-1	DC	N1-C2	-5.25	1.34	1.40

The worst 5 of 7 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	C	315	LEU	CA-CB-CG	-10.54	91.07	115.30
2	D	5	DT	C3'-C2'-C1'	-6.81	94.33	102.50
3	A	325	ARG	NE-CZ-NH1	5.69	123.14	120.30
3	C	335	LEU	CB-CG-CD2	-5.61	101.46	111.00
3	C	331	LEU	CB-CG-CD2	-5.39	101.84	111.00

There are no chirality outliers.

All (3) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	B	-7	DC	Sidechain
3	C	285	TYR	Sidechain
2	D	-7	DG	Sidechain

## 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	B	421	0	241	48	0
2	D	434	0	237	18	0
3	A	510	0	525	69	0
3	C	510	0	525	56	0
4	A	6	0	0	1	0
4	B	6	0	0	0	0
4	C	17	0	0	0	0
4	D	5	0	0	0	0
All	All	1909	0	1528	173	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 51.

The worst 5 of 173 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:C:291:ARG:HG3	3:C:291:ARG:HH11	1.19	1.08
1:B:4:DA:H2''	1:B:5:DT:H5''	1.32	1.06
1:B:7:DC:H2''	1:B:8:DA:H5''	1.39	1.01
1:B:10:DT:H5'	1:B:10:DT:H6	1.18	1.01
1:B:4:DA:C2'	1:B:5:DT:H5''	1.92	0.99

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	58/62 (94%)	39 (67%)	15 (26%)	4 (7%)	1	3
3	C	58/62 (94%)	42 (72%)	14 (24%)	2 (3%)	5	16
All	All	116/124 (94%)	81 (70%)	29 (25%)	6 (5%)	2	7

5 of 6 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
3	A	338	LEU
3	A	285	TYR
3	C	305	GLN
3	A	339	ARG
3	C	309	GLU

### 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	
3	A	57/58 (98%)	47 (82%)	10 (18%)	2	7
3	C	57/58 (98%)	42 (74%)	15 (26%)	0	1
All	All	114/116 (98%)	89 (78%)	25 (22%)	1	3

5 of 25 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	C	284	GLU

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
3	C	287	VAL
3	C	332	SER
3	C	286	ARG
3	C	291	ARG

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (5) such sidechains are listed below:

Mol	Chain	Res	Type
3	A	312	GLN
3	C	281	ASN
3	C	283	ASN
3	C	307	ASN
3	C	312	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](#)

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	B	21/21 (100%)	-0.95	0	100 100	65, 86, 125, 142	0
2	D	21/21 (100%)	-0.98	0	100 100	67, 87, 129, 142	0
3	A	60/62 (96%)	-0.35	2 (3%)	50 38	62, 86, 123, 165	0
3	C	60/62 (96%)	-0.46	0	100 100	64, 85, 117, 128	0
All	All	162/166 (97%)	-0.55	2 (1%)	81 73	62, 86, 128, 165	0

All (2) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
3	A	282	SER	3.8
3	A	281	ASN	2.8

### 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.