wwPDB X-ray Structure Validation Summary Report (i)

Oct 17, 2021 – 06:25 AM EDT

PDB ID	:	1GHQ
Title	:	CR2-C3D COMPLEX STRUCTURE
Authors	:	Szakonyi, G.; Guthridge, J.M.; Li, D.; Holers, V.M.; Chen, X.S.
Deposited on	:	2001-01-11
Resolution	:	2.04 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*
A user guide is available at
https://www.wwpdb.org/validation/2017/XrayValidationReportHelp
with specific help available everywhere you see the (i) symbol.

The following versions of software and data (see references (i)) were used in the production of this report:

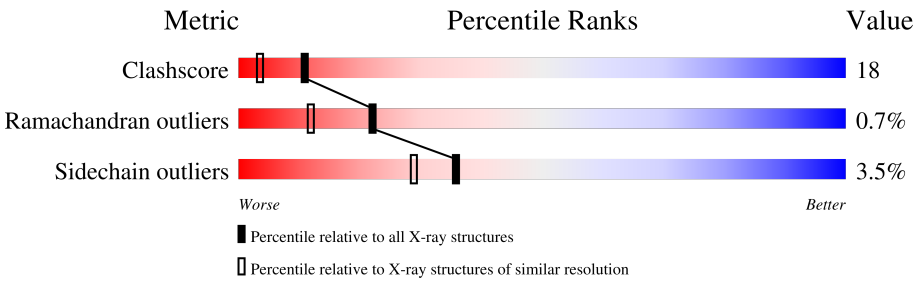
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	NOT EXECUTED
EDS	:	NOT EXECUTED
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.23.2

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

The reported resolution of this entry is 2.04 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))
Clashscore	141614	1773 (2.04-2.04)
Ramachandran outliers	138981	1752 (2.04-2.04)
Sidechain outliers	138945	1752 (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 of the lower bar indicate the fraction of residues that contain outliers for >=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 <=5%

Note EDS was not executed.

Mol	Chain	Length	Quality of chain
1	A	308	77% 20% •
2	B	134	66% 28% • •
2	C	134	66% 30% • •

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
4	NDG	C	701	-	-	X	-

2 Entry composition (i)

There are 5 unique types of molecules in this entry. The entry contains 5132 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 COMPLEMENT C3.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	307	2417	1550	406	452	9	0	0	0

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	MET	-	cloning artifact	UNP P01024
A	2	LEU	-	cloning artifact	UNP P01024
A	17	ALA	CYS	engineered mutation	UNP P01024
A	295	SER	-	insertion	UNP P01024

- Molecule 2 is a protein called CR2/CD121/C3D/EPSTEIN-BARR VIRUS RECEPTOR.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	B	129	987	626	169	182	10	0	0	0
2	C	134	1030	656	174	190	10	1	0	0

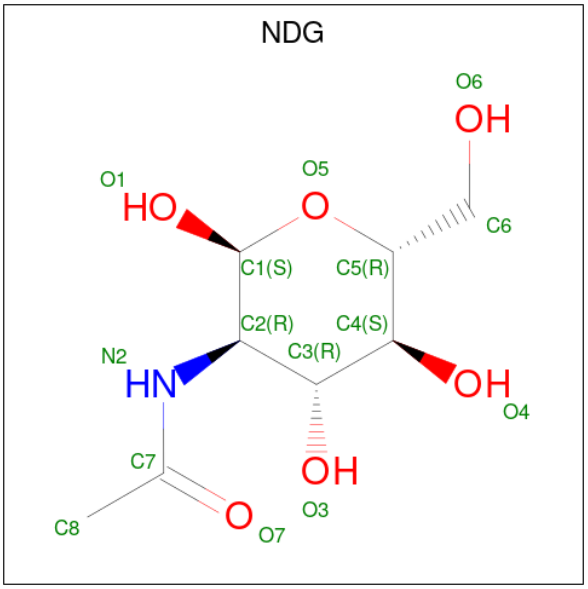
There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	1	ALA	-	cloning artifact	UNP P20023
C	1	ALA	-	cloning artifact	UNP P20023

- Molecule 3 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	Zn		
3	A	1	1	1	0	0
3	B	1	1	1	0	0

- Molecule 4 is 2-acetamido-2-deoxy-alpha-D-glucopyranose (three-letter code: NDG) (formula: $C_8H_{15}NO_6$).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
			Total	C	N	O		
4	B	1	15	8	1	6	0	0
			Total	C	N	O		
4	C	1	15	8	1	6	0	0

- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
5	A	372	372	372	0	0
			Total	O		
5	B	170	170	170	0	0
			Total	O		
5	C	124	124	124	0	0

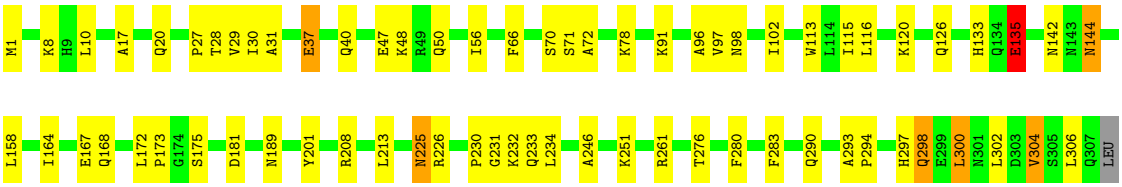
3 Residue-property plots (i)

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. 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. 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.

Note EDS was not executed.

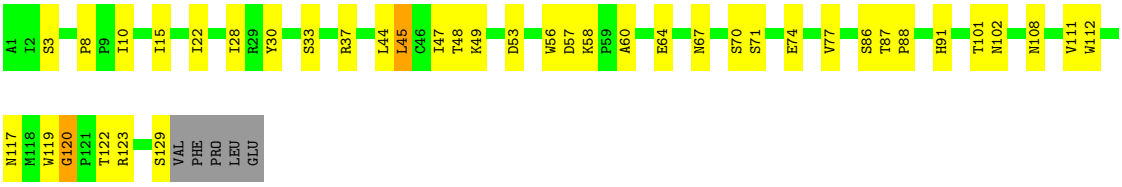
• Molecule 1: COMPLEMENT C3

Chain A:



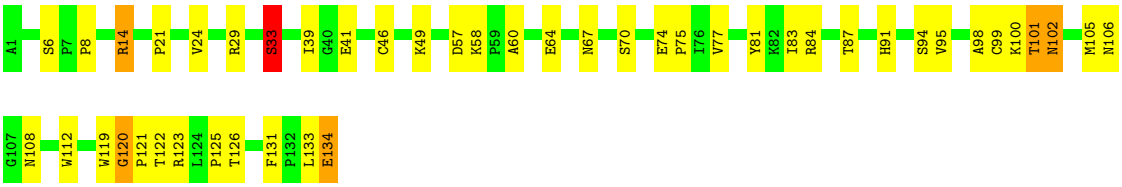
• Molecule 2: CR2/CD121/C3D/EPSTEIN-BARR VIRUS RECEPTOR

Chain B:



• Molecule 2: CR2/CD121/C3D/EPSTEIN-BARR VIRUS RECEPTOR

Chain C:



4 Data and refinement statistics (i)

Xtriage (Phenix) and EDS were not executed - this section is therefore incomplete.

Property	Value	Source
Space group	H 3 2	Depositor
Cell constants a, b, c, α, β, γ	170.00Å 170.00Å 174.00Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	25.00 – 2.04	Depositor
% Data completeness (in resolution range)	94.1 (25.00-2.04)	Depositor
R_{merge}	0.07	Depositor
R_{sym}	0.07	Depositor
Refinement program	CNS	Depositor
R, R_{free}	0.195 , 0.239	Depositor
Estimated twinning fraction	No twinning to report.	Xtriage
Total number of atoms	5132	wwPDB-VP
Average B, all atoms (Å^2)	33.0	wwPDB-VP

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: NDG, ZN

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	0.82	3/2467 (0.1%)	0.80	2/3344 (0.1%)
2	B	0.56	0/1016	0.74	0/1382
2	C	1.19	10/1061 (0.9%)	1.30	13/1443 (0.9%)
All	All	0.87	13/4544 (0.3%)	0.93	15/6169 (0.2%)

The worst 5 of 13 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	C	21	PRO	N-CD	-18.03	1.22	1.47
2	C	133	LEU	C-N	-14.21	1.01	1.34
1	A	135	GLU	CB-CG	14.05	1.78	1.52
2	C	134	GLU	C-O	12.87	1.47	1.23
2	C	74	GLU	CB-CG	-11.27	1.30	1.52

The worst 5 of 15 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	C	133	LEU	C-N-CA	20.43	172.77	121.70
2	C	133	LEU	O-C-N	-19.76	91.08	122.70
2	C	133	LEU	CA-C-N	14.02	148.05	117.20
2	C	33	SER	CB-CA-C	10.97	130.94	110.10
1	A	135	GLU	CG-CD-OE2	8.31	134.92	118.30

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	2417	0	2417	77	0
2	B	987	0	969	44	0
2	C	1030	0	1009	43	0
3	A	1	0	0	0	0
3	B	1	0	0	0	0
4	B	15	0	12	2	0
4	C	15	0	12	7	0
5	A	372	0	0	17	0
5	B	170	0	0	5	0
5	C	124	0	0	7	1
All	All	5132	0	4419	162	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 18.

The worst 5 of 162 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:135:GLU:CB	1:A:135:GLU:CG	1.78	1.60
2:C:102:ASN:HD21	4:C:701:NDG:C1	1.49	1.23
1:A:302:LEU:HD11	1:A:304:VAL:HG12	1.13	1.12
2:C:134:GLU:HG2	5:C:797:HOH:O	1.52	1.07
1:A:40:GLN:OE1	1:A:304:VAL:HG22	1.57	1.05

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)
5:C:757:HOH:O	5:C:757:HOH:O[4_555]	1.41	0.79

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	
1	A	305/308 (99%)	298 (98%)	7 (2%)	0	100	100
2	B	127/134 (95%)	123 (97%)	2 (2%)	2 (2%)	9	2
2	C	132/134 (98%)	128 (97%)	2 (2%)	2 (2%)	10	3
All	All	564/576 (98%)	549 (97%)	11 (2%)	4 (1%)	22	12

All (4) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	B	120	GLY
2	C	120	GLY
2	B	33	SER
2	C	33	SER

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	
1	A	255/257 (99%)	243 (95%)	12 (5%)	26	18
2	B	112/117 (96%)	110 (98%)	2 (2%)	59	55
2	C	117/117 (100%)	114 (97%)	3 (3%)	46	39
All	All	484/491 (99%)	467 (96%)	17 (4%)	36	29

5 of 17 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	C	14	ARG
2	C	87	THR
1	A	233	GLN
1	A	297	HIS
1	A	298	GLN

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 13 such sidechains are listed below:

Mol	Chain	Res	Type
2	B	102	ASN
2	B	108	ASN
2	C	114	GLN
2	C	91	HIS
2	C	102	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 monosaccharides in this entry.

5.6 Ligand geometry (i)

Of 4 ligands modelled in this entry, 2 are monoatomic - leaving 2 for Mogul analysis.

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	$\#\|Z\| > 2$
4	NDG	C	701	-	15,15,15	0.66	0	21,21,21	0.80	1 (4%)
4	NDG	B	702	-	15,15,15	0.59	0	21,21,21	0.89	1 (4%)

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
4	NDG	C	701	-	-	0/6/26/26	0/1/1/1
4	NDG	B	702	-	-	2/6/26/26	0/1/1/1

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
4	C	701	NDG	C2-N2-C7	-2.11	118.05	123.18
4	B	702	NDG	C4-C3-C2	2.01	113.29	110.34

There are no chirality outliers.

All (2) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
4	B	702	NDG	C8-C7-N2-C2
4	B	702	NDG	O7-C7-N2-C2

There are no ring outliers.

2 monomers are involved in 9 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	C	701	NDG	7	0
4	B	702	NDG	2	0

5.7 Other polymers (i)

There are no such residues in this entry.

5.8 Polymer linkage issues (i)

The following chains have linkage breaks:

Mol	Chain	Number of breaks
2	C	1

All chain breaks are listed below:

Model	Chain	Residue-1	Atom-1	Residue-2	Atom-2	Distance (Å)
1	C	133:LEU	C	134:GLU	N	1.01

6 Fit of model and data (i)

6.1 Protein, DNA and RNA chains (i)

EDS was not executed - this section is therefore empty.

6.2 Non-standard residues in protein, DNA, RNA chains (i)

EDS was not executed - this section is therefore empty.

6.3 Carbohydrates (i)

EDS was not executed - this section is therefore empty.

6.4 Ligands (i)

EDS was not executed - this section is therefore empty.

6.5 Other polymers (i)

EDS was not executed - this section is therefore empty.