

Full wwPDB X-ray Structure Validation Report (i)

Feb 3, 2025 - 06:05 pm GMT

PDB ID	:	9GC6
Title	:	Highly optimized CNS penetrant inhibitors of EGFR Exon20 Insertion Muta-
		tions
Authors	:	Hargreaves, D.
Deposited on	:	2024-08-01
Resolution	:	1.90 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* A user guide is available at https://www.wwpdb.org/validation/2017/XrayValidationReportHelp with specific help available everywhere you see the (i) symbol.

The types of validation reports are described at http://www.wwpdb.org/validation/2017/FAQs#types.

The following versions of software and data (see references (1)) were used in the production of this report:

MolProbity	:	4.02b-467
Mogul	:	1.8.4, CSD as 541 be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	3.0
buster-report	:	1.1.7(2018)
Percentile statistics	:	20231227.v01 (using entries in the PDB archive December 27th 2023)
CCP4	:	9.0.003 (Gargrove)
Density-Fitness	:	1.0.11
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.40

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure: $X\text{-}RAY\;DIFFRACTION$

The reported resolution of this entry is 1.90 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	Similar resolution		
	$(\# { m Entries})$	$(\# { m Entries}, { m resolution} { m range}({ m A}))$		
R_{free}	164625	7293 (1.90-1.90)		
Clashscore	180529	8090 (1.90-1.90)		
Ramachandran outliers	177936	8022 (1.90-1.90)		
Sidechain outliers	177891	8022 (1.90-1.90)		
RSRZ outliers	164620	7292 (1.90-1.90)		

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% 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		
			18%		
1	А	330	82%	9%	10%
			16%		
1	В	330	81%	10%	9%



9GC6

2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 5047 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.

Mol	Chain	Residues		Atoms				ZeroOcc	AltConf	Trace
1	Δ	208	Total	С	Ν	0	\mathbf{S}	0	Ο	0
I A	290	2400	1539	407	436	18	0	0		
1	1 D	D 200	Total	С	Ν	0	S	0	0	0
I D	500	2415	1549	410	438	18	0	0	0	

• Molecule 1 is a protein called Epidermal growth factor receptor.

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
А	693	GLY	-	expression tag	UNP P00533
А	694	SER	-	expression tag	UNP P00533
А	948	ARG	VAL	engineered mutation	UNP P00533
В	693	GLY	-	expression tag	UNP P00533
В	694	SER	-	expression tag	UNP P00533
В	948	ARG	VAL	engineered mutation	UNP P00533

• Molecule 2 is 1-[2-[5-fluoranyl-4-(2-fluorophenyl)pyridin-2-yl]-3-pyrimidin-4-yl-4,6-dihydrop yrrolo[3,4-d]imidazol-5-yl]propan-1-one (three-letter code: A1IZ9) (formula: C₂₃H₁₈F₂N₆O) (labeled as "Ligand of Interest" by depositor).





Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	
9	Λ	1	Total	С	F	Ν	Ο	0	0
	1	32	23	2	6	1	0	0	
0	Р	1	Total	С	F	Ν	0	0	0
	L	32	23	2	6	1	0	0	

• Molecule 3 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	А	82	Total O 82 82	0	0
3	В	86	Total O 86 86	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 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: Epidermal growth factor receptor



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants	76.44Å 70.93Å 78.23Å	Deperitor
a, b, c, α , β , γ	90.00° 112.88° 90.00°	Depositor
$\mathbf{P}_{\text{assolution}}(\hat{\mathbf{A}})$	64.42 - 1.90	Depositor
Resolution (A)	$64.42 \ - \ 1.90$	EDS
% Data completeness	55.7 (64.42-1.90)	Depositor
(in resolution range)	55.7(64.42 - 1.90)	EDS
R _{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) > 1$	$1.29 (at 1.90 \text{\AA})$	Xtriage
Refinement program	BUSTER 2.11.8 (26-JUL-2023)	Depositor
D D.	0.266 , 0.323	Depositor
Π, Π_{free}	0.259 , 0.310	DCC
R_{free} test set	28622 reflections $(4.96%)$	wwPDB-VP
Wilson B-factor $(Å^2)$	25.1	Xtriage
Anisotropy	0.078	Xtriage
Bulk solvent $k_{sol}(e/Å^3), B_{sol}(Å^2)$	0.32 , 36.4	EDS
L-test for twinning ²	$< L >=0.53, < L^2>=0.37$	Xtriage
Estimated twinning fraction	0.000 for l,-k,h	Xtriage
F_o, F_c correlation	0.92	EDS
Total number of atoms	5047	wwPDB-VP
Average B, all atoms $(Å^2)$	29.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: The analyses of the Patterson function reveals a significant off-origin peak that is 56.66 % of the origin peak, indicating pseudo-translational symmetry. The chance of finding a peak of this or larger height randomly in a structure without pseudo-translational symmetry is equal to 2.6507e-05. The detected translational NCS is most likely also responsible for the elevated intensity ratio.

²Theoretical 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.



¹Intensities estimated from amplitudes.

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: A1IZ9 $\,$

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).

Mal	Chain	Bond	lengths	Bond angles		
		RMSZ	# Z > 5	RMSZ	# Z > 5	
1	А	0.43	0/2451	0.59	0/3316	
1	В	0.42	0/2466	0.58	0/3334	
All	All	0.43	0/4917	0.58	0/6650	

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	А	2400	0	2431	18	0
1	В	2415	0	2452	16	0
2	А	32	0	0	0	0
2	В	32	0	0	0	0
3	А	82	0	0	0	0
3	В	86	0	0	0	0
All	All	5047	0	4883	34	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 (34) close contacts within the same asymmetric unit are listed below, sorted by their clash



magnitude.

Atom 1	Atom 2	Interatomic	Clash
Atom-1	Atom-2	$distance ({ m \AA})$	overlap (Å)
1:A:960:LYS:NZ	1:A:962:ARG:HH21	1.73	0.86
1:A:960:LYS:HZ2	1:A:962:ARG:HH21	1.30	0.80
1:B:747:LEU:HD13	1:B:862:LEU:HD11	1.70	0.72
1:A:960:LYS:NZ	1:A:962:ARG:NH2	2.41	0.69
1:A:960:LYS:HZ2	1:A:962:ARG:NH2	1.96	0.62
1:A:813:TYR:OH	1:A:990:PRO:HD3	2.08	0.54
1:B:949:LYS:HB3	1:B:959:PRO:HD3	1.89	0.53
1:B:905:TRP:CD1	1:B:947:MET:HE1	2.43	0.53
1:B:905:TRP:HD1	1:B:947:MET:HE1	1.74	0.51
1:A:960:LYS:HZ1	1:A:962:ARG:HH21	1.54	0.51
1:A:731:TRP:CZ2	1:A:733:PRO:HB3	2.48	0.48
1:A:905:TRP:HD1	1:A:947:MET:CE	2.28	0.47
1:A:744:ILE:HG12	1:A:789:ILE:HG13	1.96	0.46
1:A:905:TRP:HD1	1:A:947:MET:HE1	1.81	0.46
1:B:882:ALA:HA	1:B:898:TRP:CD2	2.51	0.46
1:B:918:ILE:HD13	1:B:926:ILE:HD13	1.99	0.45
1:B:731:TRP:CZ2	1:B:733:PRO:HB3	2.52	0.45
1:B:769:VAL:HG23	1:B:769:VAL:O	2.16	0.45
1:B:812:GLN:NE2	1:B:1013:ALA:H	2.15	0.44
1:A:726:VAL:HG22	1:A:745:LYS:HE3	2.00	0.44
1:A:904:VAL:O	1:A:908:MET:HG2	2.18	0.44
1:B:708:LYS:HB2	1:B:711:GLU:HG3	1.98	0.43
1:B:707:LEU:HD12	1:B:789:ILE:HD13	1.99	0.43
1:B:806:LYS:HA	1:B:807:ASP:HA	1.80	0.43
1:A:703:LEU:HD13	1:A:768:SER:HA	2.01	0.43
1:B:813:TYR:OH	1:B:990:PRO:HD3	2.18	0.43
1:A:841:ARG:HH12	1:A:877:PRO:HB3	1.84	0.42
1:A:882:ALA:HA	1:A:898:TRP:CD2	2.54	0.42
1:A:829:GLU:HA	1:A:893:HIS:CE1	2.55	0.42
1:A:919:PRO:HG2	1:A:922:GLU:HG2	2.01	0.42
1:B:961:PHE:O	1:B:965:ILE:HG13	2.20	0.42
1:B:802:VAL:HA	1:B:809:ILE:HD11	2.02	0.41
1:B:905:TRP:HD1	1:B:947:MET:CE	2.33	0.41
1:A:811:SER:OG	1:A:975:PRO:HB2	2.20	0.40

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	Perce	ntiles
1	А	292/330~(88%)	286~(98%)	6(2%)	0	100	100
1	В	294/330~(89%)	285~(97%)	9~(3%)	0	100	100
All	All	586/660~(89%)	571 (97%)	15 (3%)	0	100	100

There are no Ramachandran outliers to report.

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	А	266/289~(92%)	263~(99%)	3 (1%)	70 71	
1	В	267/289~(92%)	262~(98%)	5(2%)	52 49	
All	All	533/578~(92%)	525 (98%)	8 (2%)	60 59	

All (8) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	А	748	ARG
1	А	858	LEU
1	А	956	ASP
1	В	705	ARG
1	В	778	LEU
1	В	804	GLU
1	В	956	ASP
1	В	1008	ASP



Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (3) such sidechains are listed below:

Mol	Chain	Res	Type
1	А	773	HIS
1	В	773	HIS
1	В	812	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 oligosaccharides 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	Turne	Chain	hain Dec	T:nl.	Bond lengths			Bond angles		
INIOI	I Type Chain Res	nes	LIIIK	Counts	RMSZ	# Z > 2	Counts	RMSZ	# Z > 2	
2	A1IZ9	В	1101	1	32,36,36	0.41	0	36,52,52	1.21	2 (5%)
2	A1IZ9	А	1101	1	32,36,36	0.47	0	36,52,52	1.02	2 (5%)

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
2	A1IZ9	В	1101	1	-	5/15/26/26	0/5/5/5
2	A1IZ9	А	1101	1	-	5/15/26/26	0/5/5/5

There are no bond length outliers.

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	$Observed(^{o})$	$Ideal(^{o})$
2	В	1101	A1IZ9	C16-N3-C11	4.51	130.63	124.57
2	А	1101	A1IZ9	C16-N3-C11	4.15	130.15	124.57
2	В	1101	A1IZ9	C5-C6-C10	4.00	126.45	122.64
2	А	1101	A1IZ9	C12-C13-N2	2.03	104.22	102.35

There are no chirality outliers.

All (10) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	А	1101	A1IZ9	N1-C11-C8-C7
2	В	1101	A1IZ9	N1-C11-C8-C7
2	В	1101	A1IZ9	N3-C11-C8-N
2	В	1101	A1IZ9	N2-C20-C21-C22
2	А	1101	A1IZ9	N1-C11-C8-N
2	В	1101	A1IZ9	N1-C11-C8-N
2	А	1101	A1IZ9	C4-C5-C6-C10
2	В	1101	A1IZ9	N3-C11-C8-C7
2	А	1101	A1IZ9	N3-C11-C8-N
2	А	1101	A1IZ9	N3-C11-C8-C7

There are no ring outliers.

No monomer is involved in short contacts.

The following is a two-dimensional graphical depiction of Mogul quality analysis of bond lengths, bond angles, torsion angles, and ring geometry for all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the validation Tables will also be included. For torsion angles, if less then 5% of the Mogul distribution of torsion angles is within 10 degrees of the torsion angle in question, then that torsion angle is considered an outlier. Any bond that is central to one or more torsion angles identified as an outlier by Mogul will be highlighted in the graph. For rings, the root-mean-square deviation (RMSD) between the ring in question and similar rings identified by Mogul is calculated over all ring torsion angles. If the average RMSD is greater than 60 degrees and the minimal RMSD between the ring in question and any Mogul-identified rings is also greater than 60 degrees, then that ring is considered an outlier. The outliers are highlighted in purple. The color gray indicates Mogul did not find sufficient equivalents in the CSD to analyse the geometry.







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^{th} 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	$OWAB(Å^2)$	Q<0.9
1	А	298/330~(90%)	1.23	59 (19%) 3 3	10, 28, 50, 66	0
1	В	300/330~(90%)	1.12	53 (17%) 4 4	12, 28, 49, 63	0
All	All	598/660~(90%)	1.17	112 (18%) 4 4	10, 28, 49, 66	0

All (112) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	В	861	LEU	6.5
1	В	699	PRO	6.4
1	А	755	ALA	5.9
1	В	862	LEU	5.7
1	А	876	VAL	5.7
1	А	861	LEU	5.6
1	А	749	GLU	5.6
1	А	889	ARG	5.6
1	В	750	ALA	5.1
1	А	1011	VAL	5.0
1	А	737	LYS	4.5
1	А	1010	VAL	4.4
1	В	698	ALA	4.3
1	А	806	LYS	4.3
1	А	977	ARG	4.3
1	А	986	ARG	4.2
1	А	859	ALA	4.2
1	А	698	ALA	4.0
1	В	986	ARG	3.6
1	A	725	THR	3.5
1	A	807	ASP	3.4
1	В	860	LYS	3.3
1	В	757	LYS	3.3
1	А	1008	ASP	3.2

Continued on next page...



Mol	Chain	Res	Type	RSRZ
1	А	918	ILE	3.2
1	В	977	ARG	3.2
1	В	1008	ASP	3.2
1	В	754	LYS	3.2
1	А	699	PRO	3.2
1	А	988	HIS	3.2
1	В	1011	VAL	3.1
1	В	945	MET	3.1
1	В	806	LYS	3.1
1	А	715	ILE	3.0
1	В	889	ARG	3.0
1	В	849	GLN	3.0
1	А	759	ILE	3.0
1	В	714	LYS	2.9
1	В	809	ILE	2.9
1	А	989	LEU	2.9
1	А	717	VAL	2.9
1	В	747	LEU	2.8
1	В	962	ARG	2.8
1	В	737	LYS	2.8
1	А	987	MET	2.8
1	В	1006	ASP	2.8
1	В	755	ALA	2.8
1	А	982	GLN	2.7
1	А	747	LEU	2.7
1	А	760	LEU	2.7
1	А	910	PHE	2.7
1	В	723	PHE	2.7
1	В	988	HIS	2.7
1	А	922	GLU	2.7
1	A	738	VAL	2.7
1	А	1006	ASP	2.7
1	В	807	ASP	2.7
1	В	985	GLU	2.7
1	A	857	GLY	2.6
1	В	890	ILE	2.6
1	А	808	ASN	2.6
1	A	993	THR	2.6
1	A	1007	MET	2.6
1	А	718	LEU	2.6
1	В	707	LEU	2.6
1	В	858	LEU	2.6

Continued from previous page...

Continued on next page...



9	G	С	6

Mol	Chain	Res	Type	RSRZ	
1	В	701	GLN	2.6	
1	В	1014 ASP		2.6	
1	В	705 ARG		2.5	
1	А	727	TYR	2.5	
1	А	990	PRO	2.5	
1	В	765	VAL	2.5	
1	А	925	SER	2.4	
1	А	917	GLY	2.4	
1	А	809	ILE	2.4	
1	А	976	GLN	2.4	
1	В	882	ALA	2.4	
1	A	741	PRO	2.4	
1	В	810	GLY	2.4	
1	А	757	LYS	2.3	
1	В	743	ALA	2.3	
1	В	1013	ALA	2.3	
1	А	1014	ASP	2.3	
1	В	741	PRO	2.3	
1	А	748	ARG	2.3	
1	В	918	ILE	2.3	
1	В	711	GLU	2.3	
1	В	832	ARG	2.3	
1	В	836	ARG	2.3	
1	А	723	PHE	2.3	
1	А	858	LEU	2.3	
1	В	749	GLU	2.3	
1	А	836	ARG	2.2	
1	А	734	GLU	2.2	
1	А	790	THR	2.2	
1	В	738	VAL	2.2	
1	А	713	LYS	2.2	
1	А	786	VAL	2.2	
1	А	1013	ALA	2.2	
1	В	782	LEU	2.2	
1	В	877	PRO	2.2	
1	В	859	ALA	2.1	
1	В	748	ARG	2.1	
1	А	890	ILE	2.1	
1	А	945	MET	2.1	
1	А	742	VAL	2.0	
1	В	828	LEU	2.0	
1	В	808	ASN	2.0	

Continued from previous page...

Continued on next page...



Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	А	962	ARG	2.0
1	В	917	GLY	2.0
1	В	987	MET	2.0
1	А	996	ASN	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 monosaccharides 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. The B-factors column lists the minimum, median, 95^{th} 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	$B-factors(Å^2)$	Q<0.9
2	A1IZ9	В	1101	32/32	0.85	0.12	$23,\!26,\!28,\!28$	0
2	A1IZ9	А	1101	32/32	0.86	0.10	29,31,31,32	0

The following is a graphical depiction of the model fit to experimental electron density of all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the geometry validation Tables will also be included. Each fit is shown from different orientation to approximate a three-dimensional view.







6.5 Other polymers (i)

There are no such residues in this entry.

