



wwPDB X-ray Structure Validation Summary Report ⓘ

Oct 5, 2017 – 08:33 AM EDT

PDB ID : 5XGN
Title : Crystal structure of EGFR 696-1022 T790M/C797S in complex with Go6976
Authors : Kong, L.L.; Yun, C.H.
Deposited on : unknown
Resolution : 3.00 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

<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 : 1.7.2 (RC1), CSD as538be (2017)
Xtriage (Phenix) : 1.9-1692
EDS : rb-20030345
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) : rb-20030345

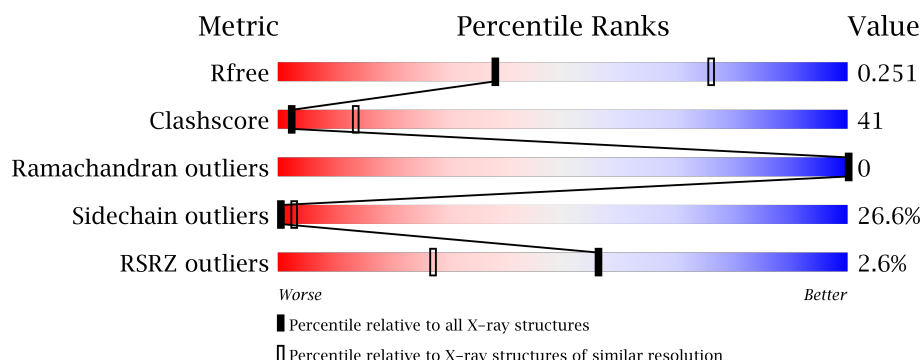
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 3.00 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	1692 (3.00-3.00)
Clashscore	112137	2037 (3.00-3.00)
Ramachandran outliers	110173	1973 (3.00-3.00)
Sidechain outliers	110143	1976 (3.00-3.00)
RSRZ outliers	101464	1716 (3.00-3.00)

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 ≥ 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	331	<div> <div>2%</div> <div> <div></div> <div>36%</div> <div>43%</div> <div>13%</div> <div>9%</div> </div> </div>
1	B	331	<div> <div>3%</div> <div> <div></div> <div>42%</div> <div>39%</div> <div>11%</div> <div>7%</div> </div> </div>

2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 4819 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.

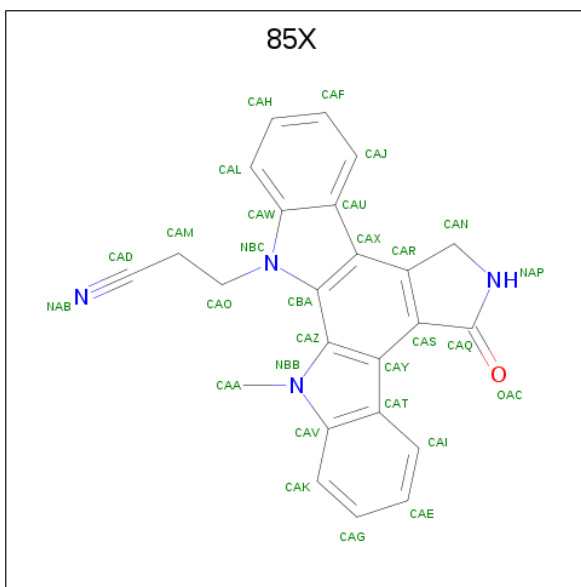
- Molecule 1 is a protein called Epidermal growth factor receptor.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	302	Total	C	N	O	S	0	0	0
			2344	1509	394	425	16			
1	B	308	Total	C	N	O	S	0	0	0
			2389	1532	400	440	17			

There are 12 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	692	GLY	-	expression tag	UNP P00533
A	693	ALA	-	expression tag	UNP P00533
A	694	MET	-	expression tag	UNP P00533
A	695	GLY	-	expression tag	UNP P00533
A	790	MET	THR	engineered mutation	UNP P00533
A	797	SER	CYS	engineered mutation	UNP P00533
B	692	GLY	-	expression tag	UNP P00533
B	693	ALA	-	expression tag	UNP P00533
B	694	MET	-	expression tag	UNP P00533
B	695	GLY	-	expression tag	UNP P00533
B	790	MET	THR	engineered mutation	UNP P00533
B	797	SER	CYS	engineered mutation	UNP P00533

- Molecule 2 is 12-(2-Cyanoethyl)-6,7,12,13-tetrahydro-13-methyl-5-oxo-5H-indolo[2,3-a]pyrrolo[3,4-c]carbazole (three-letter code: 85X) (formula: C₂₄H₁₈N₄O).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
2	A	1	Total	C	N	O	0	0
			29	24	4	1		

- Molecule 3 is CHLORIDE ION (three-letter code: CL) (formula: Cl).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	B	2	Total	Cl	0	0
			2	2		

- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	24	Total	O	0	0
			24	24		
4	B	31	Total	O	0	0
			31	31		

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 ($\text{RSRZ} > 2$). 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.

Chain A:

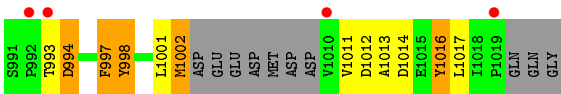
2% 36% 43% 13% 9%

Amino Acid	Percentage
GLY	2%
ALA	36%
MET	43%
GLY	13%
E697	9%
A698	0%
A702	0%
L703	0%
L704	0%
L705	0%
L706	0%
L707	0%
K708	0%
E709	0%
T710	0%
E711	0%
F712	0%
K713	0%
K714	0%
L715	0%
K716	0%
V717	0%
L718	0%
F723	0%
G724	0%
T725	0%
V726	0%
T727	0%
K728	0%
G729	0%
L730	0%
F733	0%
E734	0%
F735	0%
E736	0%
K739	0%
L740	0%
F741	0%
V742	0%
L744	0%
K745	0%
E746	0%
R748	0%
E749	0%
E750	0%
T751	0%
S752	0%
F753	0%
K754	0%
A755	0%
N756	0%
K756	0%
F756	0%
T759	0%
L760	0%
D761	0%
E762	0%
A763	0%
V765	0%
K766	0%
D767	0%
S768	0%
V769	0%
D770	0%
N771	0%
F772	0%
K773	0%
V774	0%
C775	0%
R776	0%
L777	0%
L778	0%
G781	0%
L782	0%
T783	0%
S784	0%
V785	0%
V786	0%
K787	0%
L788	0%
L789	0%
M790	0%
K791	0%
L792	0%
K793	0%
M794	0%
F795	0%
G796	0%
S797	0%
L798	0%
E804	0%
H805	0%
K806	0%
D807	0%
K808	0%
K809	0%
G810	0%
S811	0%
Q812	0%
L813	0%
L814	0%
L815	0%
K823	0%
N826	0%
Y827	0%
L828	0%
E829	0%
E830	0%
R831	0%
R832	0%
H833	0%
H834	0%
S835	0%
H836	0%
D837	0%
L838	0%
A839	0%
D840	0%
R841	0%
L842	0%
V843	0%
L844	0%
E845	0%
S846	0%
T847	0%
R848	0%
Q849	0%
H850	0%
V851	0%
R852	0%
Q853	0%
R854	0%
G857	0%
L858	0%
A859	0%
K860	0%
L861	0%
L862	0%
G863	0%
A864	0%
Y869	0%
H870	0%
A871	0%
E872	0%
G873	0%
G874	0%
R875	0%
V876	0%
P877	0%
L878	0%
E881	0%
E882	0%
E883	0%
R884	0%
S885	0%
H886	0%
R889	0%
L890	0%
Y891	0%
R892	0%
H893	0%
Q894	0%
V897	0%
A898	0%
R899	0%
K900	0%
V904	0%
M905	0%
K913	0%
P914	0%
Y915	0%
D916	0%
G917	0%
I918	0%
P919	0%
E922	0%
R923	0%
S924	0%
S925	0%
I926	0%
E928	0%
K929	0%
ASN	0%
E931	0%
R932	0%
I933	0%
P934	0%
K935	0%
P936	0%
P937	0%
I938	0%
C939	0%
T940	0%
I941	0%
D942	0%
V943	0%
Y944	0%
M947	0%
V948	0%
E949	0%
M952	0%
I953	0%
D954	0%
A955	0%
R958	0%
K959	0%
K960	0%
P961	0%
E963	0%
L964	0%
I965	0%
I966	0%
K970	0%
M971	0%
A972	0%
R973	0%
R974	0%
P975	0%
Y978	0%
V979	0%
V980	0%

Chain B:

42% 39% 11% 7%

GLY
ALA
MET
GLY
E697
A698
P699
N700
L704
R705
I706
L707
K708
E711
F712
K713
K714
I715
K716
V717
L718
G724
T725
K728
G729
L730
W731
I732
F733
GLU
G735
E736
K737
I740
P741
I744
K745
E746
L747
R748
S752
N756
I759
L760
D761
E762
A763
V764
W765
F766
A767
S768
V769
G775
R776
L777
L778
G779
I780
C781
G782
L783
S784
T785
V786
Q787
L788
I789
M790
Q791
L792
V793
W794
F795
L796
L799
E804
K805
G806
D807
N808
I809
G810
S811
L815
N816
V819
Q820
I821
A822
N826
Y827
L828
E829
D830
R831
R832
L833
V834
E835
R836
D837
L838
A839
I840
V843
E844
S845
H846
T847
Q848
H849
H850
K851
K852
K853
T854
D855
F856
E857
L858
L861
LEU
GLY
ALA
GLU
E866
H870
A871
E872
A873
V876
P877
D878
H879
H881
A882
L883
E884
S885
H888
R889
T890
Y891
T892
H893
Q894
V897
T898
L899
I900
Y900
H905
E906
L907
N908
L909
H910
L911
L912
L913
L914
L915
L916
L917
L918
L919
L920
L921
L922
L923
L924
L925
L926
L927
L928
L929
L930
L931
L932
L933
L934
L935
L936
L937
L938
L939
L940
L941
L942
L943
L944
L945
L946
L947
L948
L949
L950
L951
L952
L953
L954
L955
L956
L957
L958
L959
L960
L961
L962
L963
L964
L965
L966
L967
L968
L969
L970
L971
L972
L973
L974
L975
L976
L977
L978
L979
L980
L981
L982
L983
L984
L985
L986
L987
L988
L989
L990
L991
L992
L993
L994
L995
L996
L997
L998
L999
L1000



4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α , β , γ	49.43 Å 90.51 Å 164.22 Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	36.69 – 3.00 36.69 – 3.00	Depositor EDS
% Data completeness (in resolution range)	96.7 (36.69-3.00) 97.2 (36.69-3.00)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	3.20 (at 3.00 Å)	Xtriage
Refinement program	PHENIX (1.10.1_2155)	Depositor
R, R_{free}	0.221 , 0.251 0.224 , 0.251	Depositor DCC
R_{free} test set	772 reflections (5.43%)	DCC
Wilson B-factor (Å ²)	79.2	Xtriage
Anisotropy	0.959	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.33 , 66.9	EDS
L-test for twinning ²	$\langle L \rangle = 0.49$, $\langle L^2 \rangle = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.94	EDS
Total number of atoms	4819	wwPDB-VP
Average B, all atoms (Å ²)	81.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.43% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

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.

5 Model quality

5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: 85X, CL

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	0.48	0/2396	0.67	0/3250
1	B	0.48	0/2437	0.63	0/3303
All	All	0.48	0/4833	0.65	0/6553

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

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	2344	0	2316	229	0
1	B	2389	0	2340	166	0
2	A	29	0	0	6	0
3	B	2	0	0	0	0
4	A	24	0	0	6	0
4	B	31	0	0	5	0
All	All	4819	0	4656	388	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 41.

The worst 5 of 388 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:B:935:GLN:HA	1:B:944:TYR:CE2	1.54	1.42
1:B:935:GLN:CA	1:B:944:TYR:HE2	1.48	1.24
1:A:1017:LEU:HD12	1:A:1018:ILE:N	1.55	1.20
1:A:1017:LEU:C	1:A:1017:LEU:HD12	1.63	1.18
1:A:846:LYS:HG3	1:A:847:THR:HG22	1.25	1.16

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	
1	A	298/331 (90%)	292 (98%)	6 (2%)	0	100	100
1	B	298/331 (90%)	289 (97%)	9 (3%)	0	100	100
All	All	596/662 (90%)	581 (98%)	15 (2%)	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	A	245/288 (85%)	176 (72%)	69 (28%)	0	2
1	B	251/288 (87%)	188 (75%)	63 (25%)	0	3

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
All	All	496/576 (86%)	364 (73%)	132 (27%)	0 3

5 of 132 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	964	LEU
1	B	718	LEU
1	B	977	ARG
1	A	980	VAL
1	A	1017	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (2) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	787	GLN
1	B	808	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 carbohydrates in this entry.

5.6 Ligand geometry [i](#)

Of 3 ligands modelled in this entry, 2 are monoatomic - leaving 1 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
2	85X	A	1101	-	25,34,34	4.52	14 (56%)	24,52,52	1.86	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
2	85X	A	1101	-	-	2/4/13/13	0/6/6/6

The worst 5 of 14 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	A	1101	85X	CAN-CAR	-12.89	1.39	1.50
2	A	1101	85X	CAS-CAY	-7.64	1.32	1.43
2	A	1101	85X	CAY-CAZ	-7.55	1.32	1.42
2	A	1101	85X	CAX-CBA	-7.41	1.32	1.42
2	A	1101	85X	CAR-CAX	-5.28	1.33	1.42

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	A	1101	85X	CAN-NAP-CAQ	-7.51	105.68	113.78

There are no chirality outliers.

All (2) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	A	1101	85X	CAM-CAO-NBC-CAW
2	A	1101	85X	CAM-CAO-NBC-CBA

There are no ring outliers.

1 monomer is involved in 6 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	A	1101	85X	6	0

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, 95th 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	A	302/331 (91%)	-0.04	5 (1%) 70 42	62, 79, 115, 145	0
1	B	308/331 (93%)	-0.00	11 (3%) 43 18	58, 77, 115, 134	1 (0%)
All	All	610/662 (92%)	-0.02	16 (2%) 56 27	58, 78, 116, 145	1 (0%)

The worst 5 of 16 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	862	LEU	3.6
1	B	730	LEU	3.3
1	B	873	GLY	3.3
1	B	1019	PRO	3.0
1	B	917	GLY	3.0

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 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, 95th 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(\AA^2)	Q<0.9
2	85X	A	1101	29/29	0.89	0.21	-0.21	64,75,84,90	0
3	CL	B	1101	1/1	0.86	0.86	-	89,89,89,89	0
3	CL	B	1102	1/1	0.64	0.30	-	80,80,80,80	0

6.5 Other polymers [i](#)

There are no such residues in this entry.