Full wwPDB NMR Structure Validation Report (i)

Apr 27, 2016 – 02:15 AM BST

PDB ID	:	2M0B
Title	:	Homodimeric transmembrane domain of the human receptor tyrosine kinase ErbB1 (EGFR, HER1) in micelles
Authors	:	Lesovoy, D.M.; Bocharov, E.V.; Pustovalova, Y.E.; Bocharova, O.V.; Arseniev, A.S.
Deposited on	:	2012-10-24

This is a Full wwPDB NMR Structure Validation 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/NMRValidationReportHelp
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:

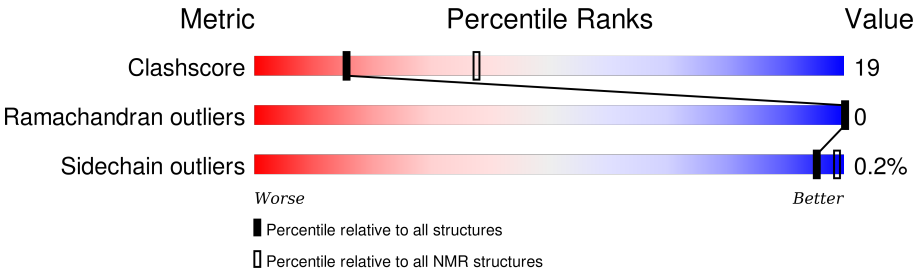
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Mogul	:	unknown
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	rb-20027457
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	rb-20027457

1 Overall quality at a glance

The following experimental techniques were used to determine the structure:
SOLUTION NMR

The overall completeness of chemical shifts assignment is 58%.

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)	NMR archive (#Entries)
Clashscore	114402	11133
Ramachandran outliers	111179	9975
Sidechain outliers	111093	9958

The table below summarises the geometric issues observed across the polymeric chains and their fit to the experimental data. The red, orange, yellow and green segments indicate the fraction of residues that contain outliers for >=3, 2, 1 and 0 types of geometric quality criteria. A cyan segment indicates the fraction of residues that are not part of the well-defined cores, and 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%

Mol	Chain	Length	Quality of chain
1	A	44	18% 20% 61%
1	B	44	9% 18% 73%

2 Ensemble composition and analysis

This entry contains 20 models. Model 4 is the overall representative, medoid model (most similar to other models). The authors have identified model 1 as representative, based on the following criterion: *fewest violations*.

The following residues are included in the computation of the global validation metrics.

Well-defined (core) protein residues			
Well-defined core	**Residue range (total)**	**Backbone RMSD (Å)**	**Medoid model**
1	A:650-A:666, B:651-B:662 (29)	0.12	4

Ill-defined regions of proteins are excluded from the global statistics.

Ligands and non-protein polymers are included in the analysis.

The models can be grouped into 4 clusters. No single-model clusters were found.

Cluster number	Models
1	4, 5, 8, 11, 12, 14, 16, 17
2	2, 3, 9, 10, 13, 15, 19
3	6, 18, 20
4	1, 7

3 Entry composition

There is only 1 type of molecule in this entry. The entry contains 1402 atoms, of which 744 are hydrogens and 0 are deuteriums.

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

Mol	Chain	Residues	Atoms						Trace
1	A	44	Total	C	H	N	O	S	0
			701	212	372	64	50	3	
1	B	44	Total	C	H	N	O	S	0
			701	212	372	64	50	3	

4 Residue-property plots (i)

4.1 Average score per residue in the NMR ensemble

These plots are provided for all protein, RNA and DNA chains in the entry. The first graphic is the same as shown in the summary in section 1 of this report. The second graphic shows the sequence where residues are colour-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 outliers are shown as green connectors. Residues which are classified as ill-defined in the NMR ensemble, are shown in cyan with an underline colour-coded according to the previous scheme. Residues which were present in the experimental sample, but not modelled in the final structure are shown in grey.

• Molecule 1: Epidermal growth factor receptor

Chain A:

• Molecule 1: Epidermal growth factor receptor

Chain B:

4.2 Scores per residue for each member of the ensemble

Colouring as in section 4.1 above.

4.2.1 Score per residue for model 1

• Molecule 1: Epidermal growth factor receptor

Chain A:

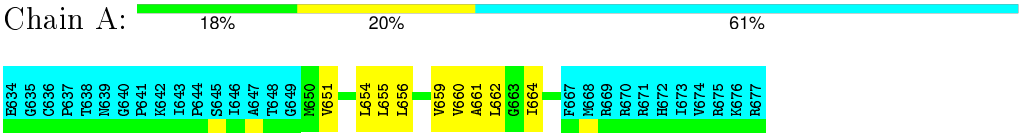
• Molecule 1: Epidermal growth factor receptor

Chain B:

4.2.2 Score per residue for model 2

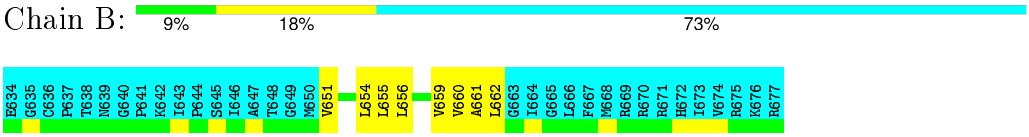
• Molecule 1: Epidermal growth factor receptor

Chain A:



• Molecule 1: Epidermal growth factor receptor

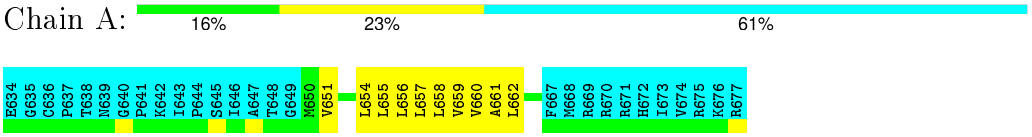
Chain B:



4.2.3 Score per residue for model 3

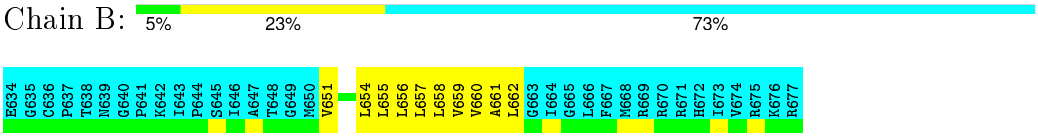
• Molecule 1: Epidermal growth factor receptor

Chain A:



• Molecule 1: Epidermal growth factor receptor

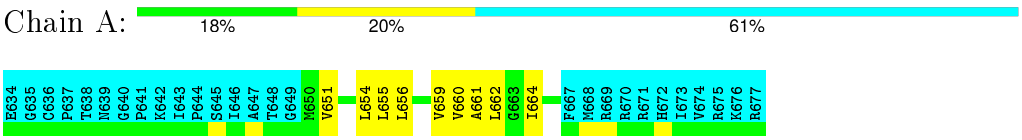
Chain B:



4.2.4 Score per residue for model 4 (medoid)

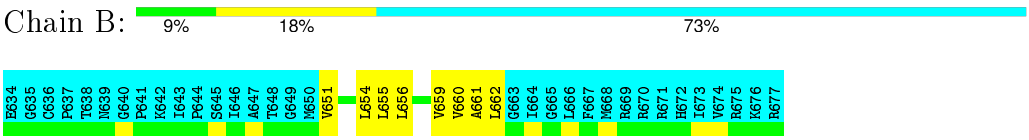
• Molecule 1: Epidermal growth factor receptor

Chain A:



• Molecule 1: Epidermal growth factor receptor

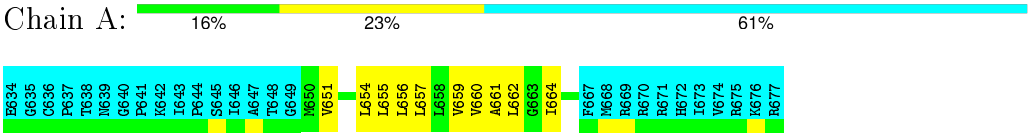
Chain B:



4.2.5 Score per residue for model 5

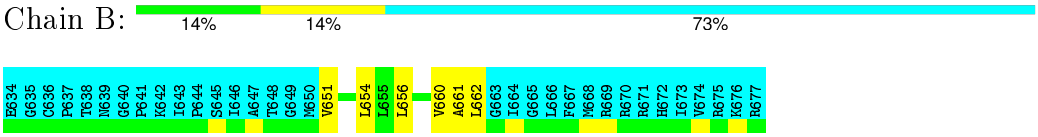
• Molecule 1: Epidermal growth factor receptor

Chain A:



• Molecule 1: Epidermal growth factor receptor

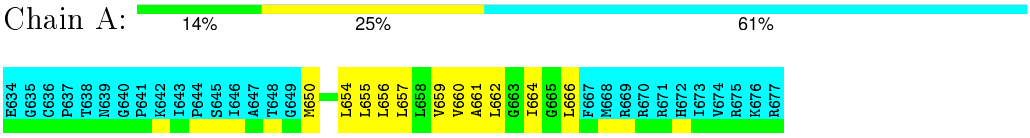
Chain B:



4.2.6 Score per residue for model 6

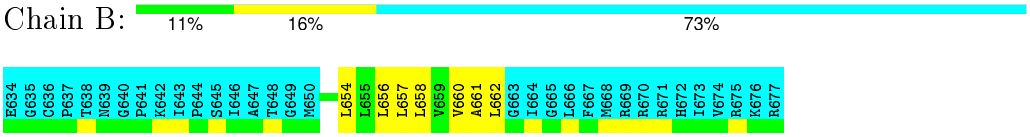
• Molecule 1: Epidermal growth factor receptor

Chain A:



• Molecule 1: Epidermal growth factor receptor

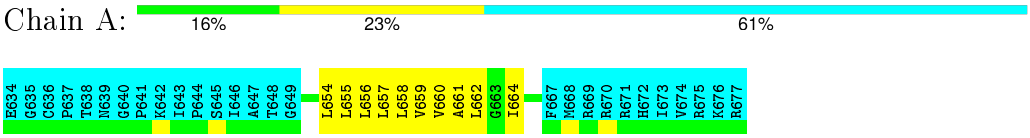
Chain B:



4.2.7 Score per residue for model 7

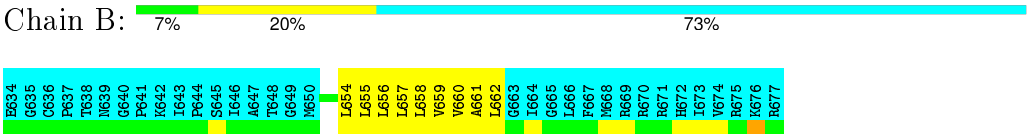
• Molecule 1: Epidermal growth factor receptor

Chain A:



• Molecule 1: Epidermal growth factor receptor

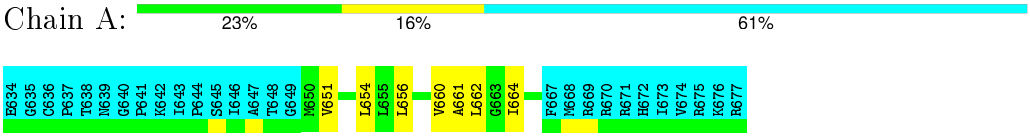
Chain B:



4.2.8 Score per residue for model 8

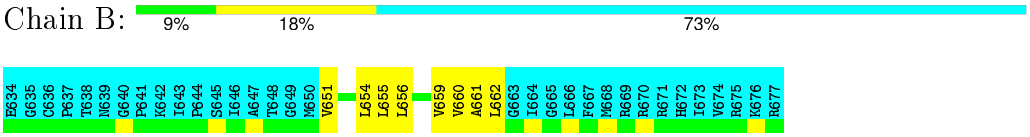
• Molecule 1: Epidermal growth factor receptor

Chain A:



• Molecule 1: Epidermal growth factor receptor

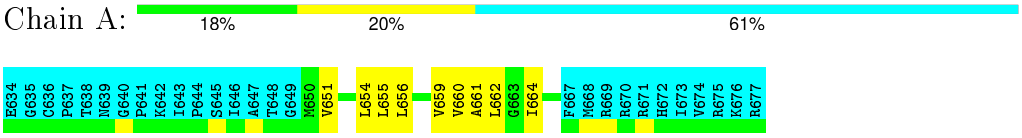
Chain B:



4.2.9 Score per residue for model 9

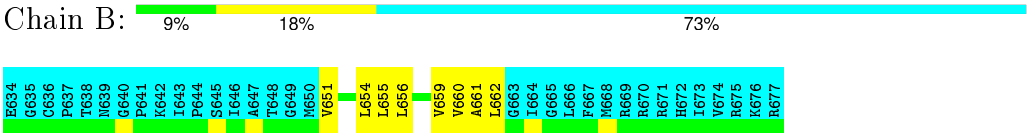
• Molecule 1: Epidermal growth factor receptor

Chain A:



• Molecule 1: Epidermal growth factor receptor

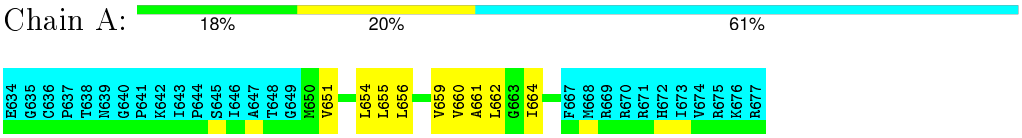
Chain B:



4.2.10 Score per residue for model 10

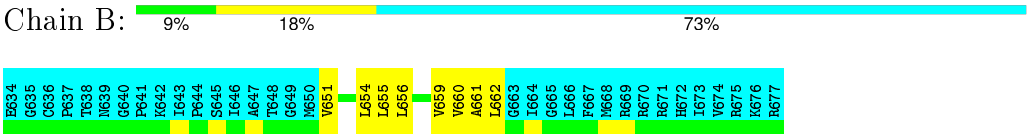
• Molecule 1: Epidermal growth factor receptor

Chain A:



• Molecule 1: Epidermal growth factor receptor

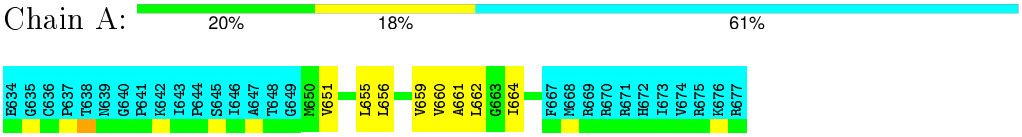
Chain B:



4.2.11 Score per residue for model 11

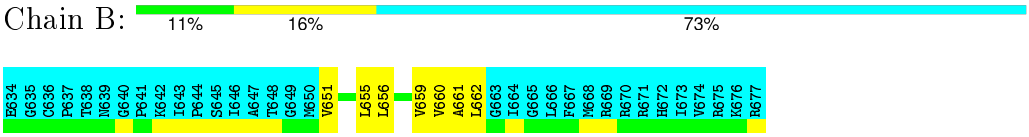
• Molecule 1: Epidermal growth factor receptor

Chain A:



• Molecule 1: Epidermal growth factor receptor

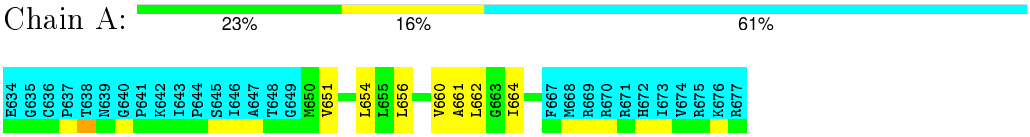
Chain B:



4.2.12 Score per residue for model 12

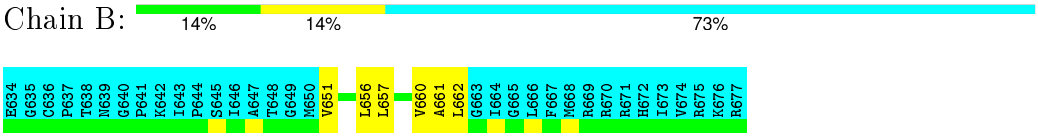
• Molecule 1: Epidermal growth factor receptor

Chain A:



• Molecule 1: Epidermal growth factor receptor

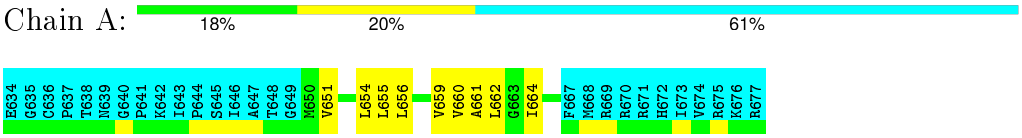
Chain B:



4.2.13 Score per residue for model 13

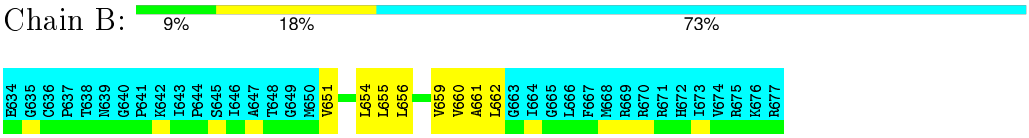
• Molecule 1: Epidermal growth factor receptor

Chain A:



• Molecule 1: Epidermal growth factor receptor

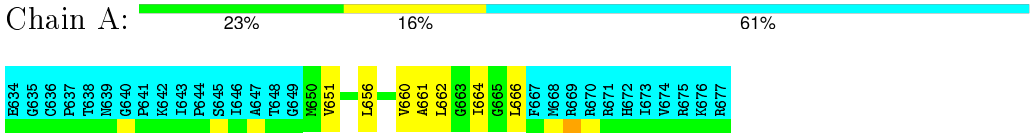
Chain B:



4.2.14 Score per residue for model 14

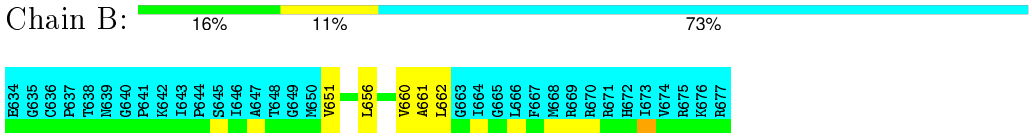
• Molecule 1: Epidermal growth factor receptor

Chain A:



• Molecule 1: Epidermal growth factor receptor

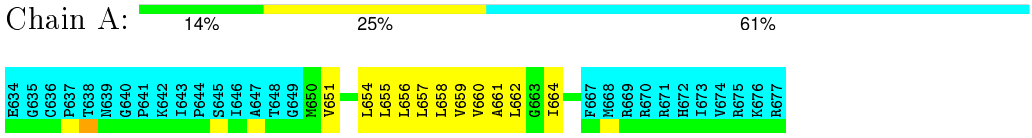
Chain B:



4.2.15 Score per residue for model 15

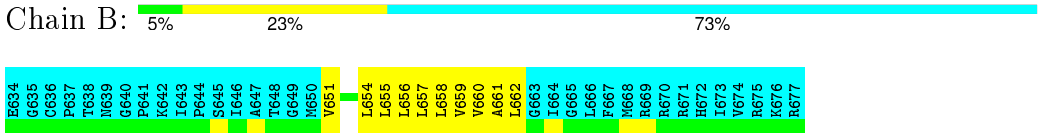
• Molecule 1: Epidermal growth factor receptor

Chain A:



• Molecule 1: Epidermal growth factor receptor

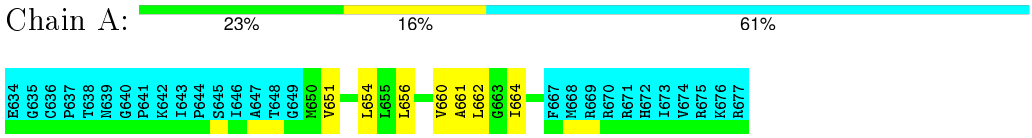
Chain B:



4.2.16 Score per residue for model 16

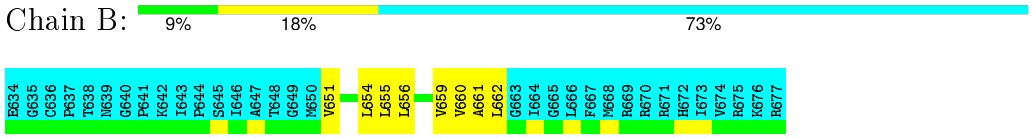
• Molecule 1: Epidermal growth factor receptor

Chain A:



• Molecule 1: Epidermal growth factor receptor

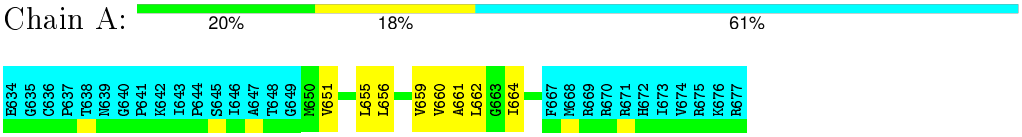
Chain B:



4.2.17 Score per residue for model 17

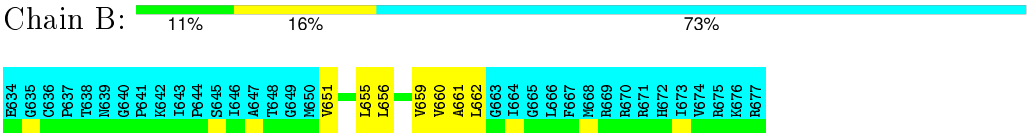
- Molecule 1: Epidermal growth factor receptor

Chain A:



- Molecule 1: Epidermal growth factor receptor

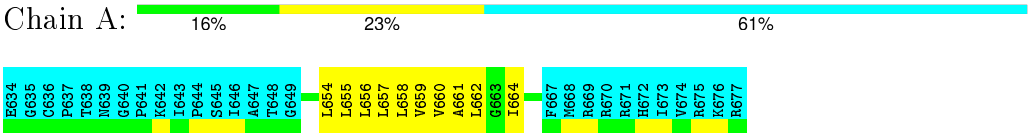
Chain B:



4.2.18 Score per residue for model 18

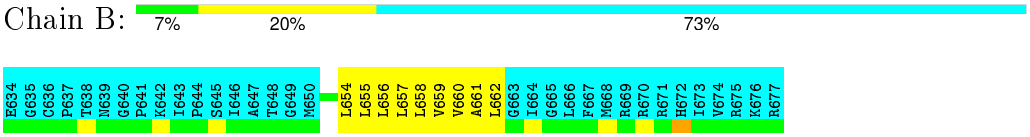
- Molecule 1: Epidermal growth factor receptor

Chain A:



- Molecule 1: Epidermal growth factor receptor

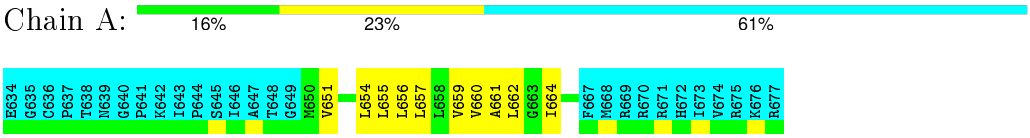
Chain B:



4.2.19 Score per residue for model 19

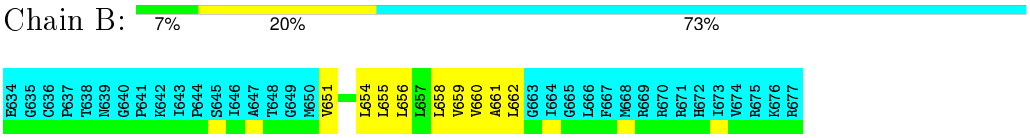
- Molecule 1: Epidermal growth factor receptor

Chain A:



- Molecule 1: Epidermal growth factor receptor

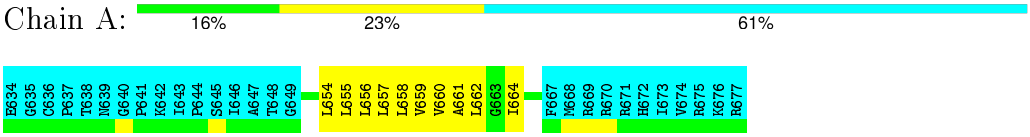
Chain B:



4.2.20 Score per residue for model 20

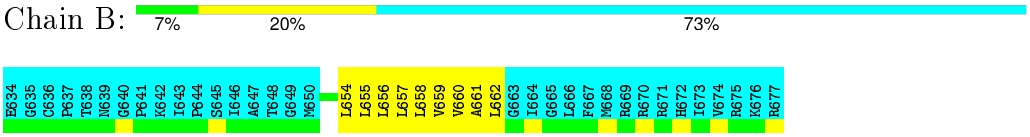
- Molecule 1: Epidermal growth factor receptor

Chain A:



- Molecule 1: Epidermal growth factor receptor

Chain B:



5 Refinement protocol and experimental data overview

The models were refined using the following method: *simulated annealing, TORSION ANGLE DYNAMICS.*

Of the 200 calculated structures, 20 were deposited, based on the following criterion: *target function.*

The following table shows the software used for structure solution, optimisation and refinement.

Software name	Classification	Version
CYANA	structure solution	3.0
CYANA	refinement	3.0

The following table shows chemical shift validation statistics as aggregates over all chemical shift files. Detailed validation can be found in section 7 of this report.

Chemical shift file(s)	2m0b_cs.str
Number of chemical shift lists	1
Total number of shifts	594
Number of shifts mapped to atoms	594
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	58%

No validations of the models with respect to experimental NMR restraints is performed at this time.

6 Model quality

6.1 Standard geometry

There are no covalent bond-length or bond-angle outliers.

There are no bond-length outliers.

There are no bond-angle outliers.

There are no chirality outliers.

There are no planarity outliers.

6.2 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 each 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 averaged over the ensemble.

Mol	Chain	Non-H	H(model)	H(added)	Clashes
1	A	115	143	143	7±2
1	B	83	106	106	7±2
All	All	3960	4980	4980	171

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 19.

All unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:658:LEU:HD11	1:B:657:LEU:HD22	0.74	1.57	18	6
1:A:661:ALA:HB3	1:B:661:ALA:HB3	0.74	1.59	3	10
1:A:657:LEU:HD22	1:B:658:LEU:HD11	0.66	1.65	20	8
1:B:656:LEU:O	1:B:660:VAL:HG23	0.62	1.95	12	20
1:A:654:LEU:HD23	1:B:654:LEU:HD23	0.61	1.72	18	16
1:A:656:LEU:O	1:A:660:VAL:HG23	0.61	1.95	6	20
1:A:662:LEU:HG	1:B:661:ALA:HB1	0.60	1.73	18	20
1:A:661:ALA:HB1	1:B:662:LEU:HG	0.59	1.73	5	20
1:A:655:LEU:O	1:A:659:VAL:HG23	0.58	1.99	3	16
1:B:655:LEU:O	1:B:659:VAL:HG23	0.56	2.00	15	16
1:A:661:ALA:HB3	1:B:661:ALA:CB	0.51	2.35	3	5
1:A:661:ALA:CB	1:B:661:ALA:HB3	0.50	2.34	3	5
1:A:662:LEU:HD23	1:B:661:ALA:O	0.42	2.14	3	1

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:654:LEU:HD22	1:B:657:LEU:CD1	0.42	2.45	12	4
1:A:657:LEU:CD1	1:B:654:LEU:HD22	0.41	2.45	20	4

6.3 Torsion angles (i)

6.3.1 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 PDB entries followed by that with respect to all NMR entries. 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	17/44 (39%)	17±0 (100±0%)	0±0 (0±0%)	0±0 (0±0%)	100	100
1	B	12/44 (27%)	12±0 (100±0%)	0±0 (0±0%)	0±0 (0±0%)	100	100
All	All	580/1760 (33%)	580 (100%)	0 (0%)	0 (0%)	100	100

There are no Ramachandran outliers.

6.3.2 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 PDB entries followed by that with respect to all NMR entries. 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	12/35 (34%)	12±0 (100±2%)	0±0 (0±2%)	94	98
1	B	9/35 (26%)	9±0 (100±0%)	0±0 (0±0%)	100	100
All	All	420/1400 (30%)	419 (100%)	1 (0%)	95	99

All 1 unique residues with a non-rotameric sidechain are listed below.

Mol	Chain	Res	Type	Models (Total)
1	A	650	MET	1

6.3.3 RNA

There are no RNA molecules in this entry.

6.4 Non-standard residues in protein, DNA, RNA chains

There are no non-standard protein/DNA/RNA residues in this entry.

6.5 Carbohydrates

There are no carbohydrates in this entry.

6.6 Ligand geometry

There are no ligands in this entry.

6.7 Other polymers

There are no such molecules in this entry.

6.8 Polymer linkage issues

There are no chain breaks in this entry.

7 Chemical shift validation

The completeness of assignment taking into account all chemical shift lists is 58% for the well-defined parts and 46% for the entire structure.

7.1 Chemical shift list 1

File name: 2m0b_cs.str

Chemical shift list name: *rcsb103047-deposited-cs.cif-20131011*

7.1.1 Bookkeeping

The following table shows the results of parsing the chemical shift list and reports the number of nuclei with statistically unusual chemical shifts.

Total number of shifts	594
Number of shifts mapped to atoms	594
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	0

7.1.2 Chemical shift referencing

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_{\alpha}$	44	0.42 ± 0.21	None needed (< 0.5 ppm)
$^{13}C_{\beta}$	38	0.93 ± 0.18	Should be applied
$^{13}C'$	44	1.07 ± 0.17	Should be applied
^{15}N	40	0.45 ± 0.38	None needed (< 0.5 ppm)

7.1.3 Completeness of resonance assignments

The following table shows the completeness of the chemical shift assignments for the well-defined regions of the structure. The overall completeness is 58%, i.e. 187 atoms were assigned a chemical shift out of a possible 323. 10 out of 19 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	85/145 (59%)	34/58 (59%)	34/58 (59%)	17/29 (59%)
Sidechain	102/178 (57%)	56/97 (58%)	46/81 (57%)	0/0 (—%)

Continued on next page...

Continued from previous page...

	Total	^{1}H	^{13}C	^{15}N
Aromatic	0/0 (—%)	0/0 (—%)	0/0 (—%)	0/0 (—%)
Overall	187/323 (58%)	90/155 (58%)	80/139 (58%)	17/29 (59%)

The following table shows the completeness of the chemical shift assignments for the full structure. The overall completeness is 46%, i.e. 509 atoms were assigned a chemical shift out of a possible 1104. 11 out of 22 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	212/428 (50%)	84/170 (49%)	88/176 (50%)	40/82 (49%)
Sidechain	285/644 (44%)	173/378 (46%)	106/230 (46%)	6/36 (17%)
Aromatic	12/32 (38%)	7/18 (39%)	5/12 (42%)	0/2 (0%)
Overall	509/1104 (46%)	264/566 (47%)	199/418 (48%)	46/120 (38%)

7.1.4 Statistically unusual chemical shifts

There are no statistically unusual chemical shifts.

7.1.5 Random Coil Index (RCI) plots

The image below reports *random coil index* values for the protein chains in the structure. The height of each bar gives a probability of a given residue to be disordered, as predicted from the available chemical shifts and the amino acid sequence. A value above 0.2 is an indication of significant predicted disorder. The colour of the bar shows whether the residue is in the well-defined core (black) or in the ill-defined residue ranges (cyan), as described in section 2 on ensemble composition.

Random coil index (RCI) for chain A: