W O R L D W I D E
wwPDB
PROTEIN DATA BANK

wwPDB X-ray Structure Validation Summary Report (i)

Jan 31, 2016 – 10:11 PM GMT

PDB ID	:	1SHY
Title	:	The Crystal Structure of HGF beta-chain in Complex with the Sema Domain of the Met Receptor.
Authors	:	Stamos, J.; Wiesmann, C.
Deposited on	:	2004-02-26
Resolution	:	3.22 Å(reported)

The following versions of software and data (see references (1)) 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 3.22 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	Percentile Ranks	Value
Rfree		0.246
Clashscore		16
Ramachandran outliers		3.6%
Sidechain outliers		25.2%
RSRZ outliers		2.6%

Worse .... *Better*

■ Percentile relative to all X-ray structures
▯ Percentile relative to X-ray structures of similar resolution
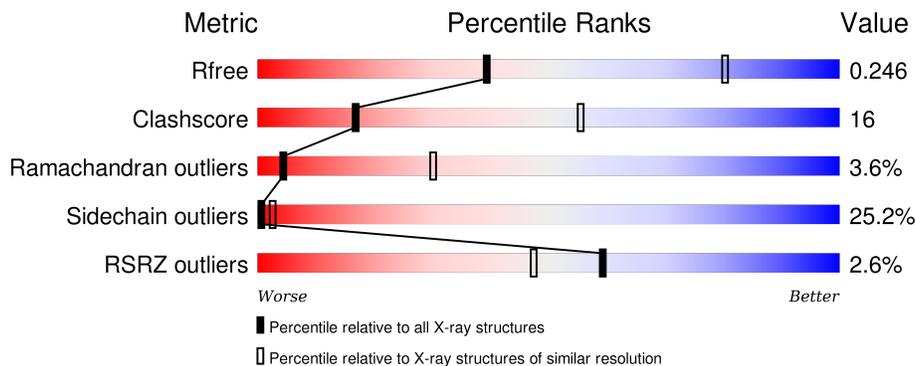
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R$_{free}$	91344	1095 (3.26-3.18)
Clashscore	102246	1046 (3.24-3.20)
Ramachandran outliers	100387	1026 (3.24-3.20)
Sidechain outliers	100360	1025 (3.24-3.20)
RSRZ outliers	91569	1100 (3.26-3.18)

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 <=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	234	2% 56% 31% 9% • •
2	B	551	3% 45% 35% 10% • 9%

2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 5735 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 Hepatocyte growth factor.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	228	Total 1773	C 1127	N 315	O 317	S 14	0	0	0

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	604	SER	CYS	ENGINEERED	UNP P14210

- Molecule 2 is a protein called Hepatocyte growth factor receptor.

| Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	499	Total 3962	C 2517	N 674	O 742	S 29	0	0	0

There are 13 discrepancies between the modelled and reference sequences:

| Chain	Residue	Modelled	Actual	Comment	Reference
B	303	LEU	LYS	SEE REMARK 999	UNP P08581
B	304	VAL	ARG	SEE REMARK 999	UNP P08581
B	305	PRO	LYS	SEE REMARK 999	UNP P08581
B	306	ARG	LYS	SEE REMARK 999	UNP P08581
B	307	GLY	ARG	SEE REMARK 999	UNP P08581
B	568	HIS	-	EXPRESSION TAG	UNP P08581
B	569	HIS	-	EXPRESSION TAG	UNP P08581
B	570	HIS	-	EXPRESSION TAG	UNP P08581
B	571	HIS	-	EXPRESSION TAG	UNP P08581
B	572	HIS	-	EXPRESSION TAG	UNP P08581
B	573	HIS	-	EXPRESSION TAG	UNP P08581
B	574	HIS	-	EXPRESSION TAG	UNP P08581
B	575	HIS	-	EXPRESSION TAG	UNP P08581

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: Hepatocyte growth factor
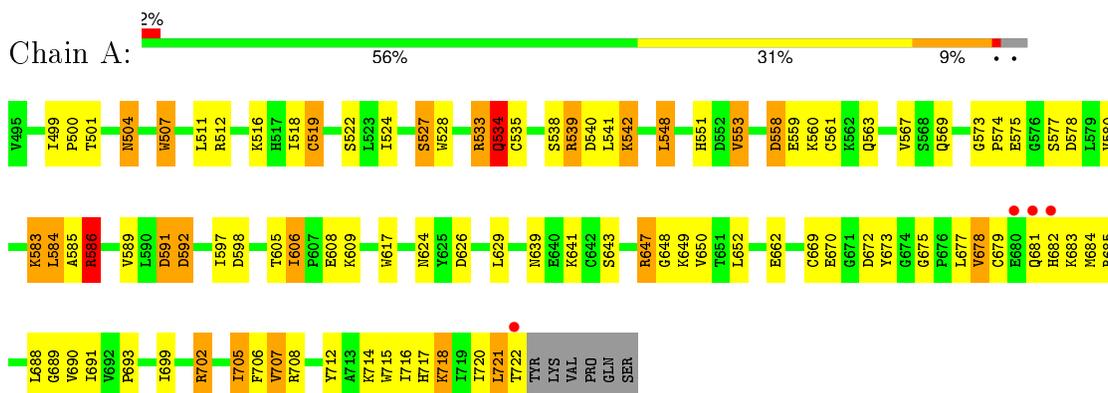
- Molecule 2: Hepatocyte growth factor receptor

L496 N497 T502 L503 V504 I505 T506 G507 K508 K509 I510 T511 K512 I513 P514 L518 G519 C520 R521 Q524 S525 C526 S527 Q528 C529 L530 S531 F535 V536 Q537 W540 D543 V546 R547 E550 C551 L552 T555 W556 T557 Q558 Q559 I560 C561 L562 P563 A564 ILE TYR LYS HIS

HIS HIS HIS HIS HIS HIS
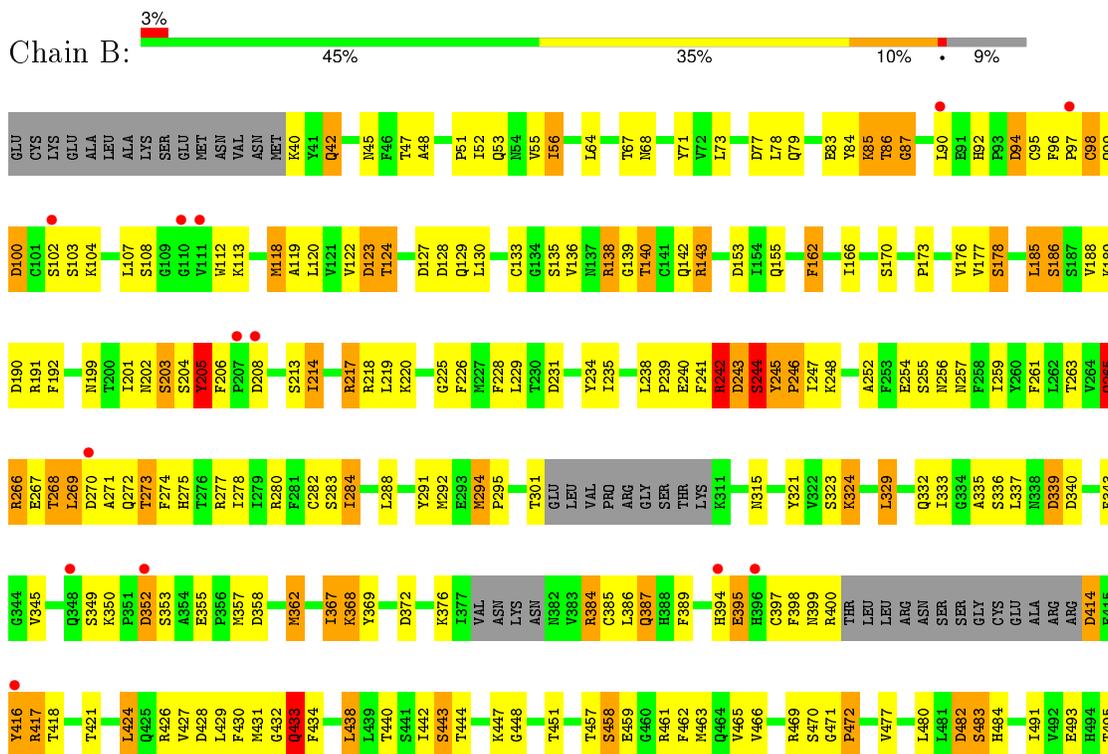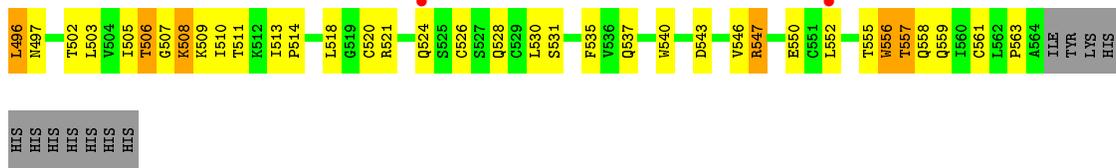
4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 2	Depositor
Cell constants a, b, c, α, β, γ	137.05Å 186.35Å 66.66Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	30.00 – 3.22 47.78 – 3.22	Depositor EDS
% Data completeness (in resolution range)	99.1 (30.00-3.22) 99.0 (47.78-3.22)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.06	Depositor
$<I/\sigma(I)>$ [1]	2.01 (at 3.19Å)	Xtriage
Refinement program	REFMAC 5.1.24	Depositor
R, R_{free}	0.209 , 0.270 0.194 , 0.246	Depositor DCC
R_{free} test set	1414 reflections (5.28%)	DCC
Wilson B-factor (Å^2)	103.0	Xtriage
Anisotropy	0.167	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.31 , 86.7	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning[2]	$<|L|> = 0.50$, $<L^2> = 0.34$	Xtriage
Outliers	0 of 28267 reflections	Xtriage
F_o,F_c correlation	0.93	EDS
Total number of atoms	5735	wwPDB-VP
Average B, all atoms (Å^2)	55.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.44% of the height of the origin peak. No significant pseudotranslation is detected.*

---

[1]Intensities estimated from amplitudes.
[2]Theoretical values of $<|L|>$, $<L^2>$ for acentric reflections are 0.5, 0.375 respectively for untwinned datasets, and 0.333, 0.2 for perfectly twinned datasets.

5 Model quality (i)

5.1 Standard geometry (i)

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	0.62	0/1815	0.86	3/2459 (0.1%)
2	B	0.56	1/4062 (0.0%)	0.86	18/5515 (0.3%)
All	All	0.58	1/5877 (0.0%)	0.86	21/7974 (0.3%)

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
2	B	0	4

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	B	561	CYS	CB-SG	5.54	1.91	1.82

The worst 5 of 21 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	B	543	ASP	CB-CG-OD2	6.80	124.42	118.30
2	B	128	ASP	CB-CG-OD2	6.58	124.22	118.30
1	A	558	ASP	CB-CG-OD2	6.34	124.01	118.30
2	B	100	ASP	CB-CG-OD2	6.28	123.95	118.30
2	B	358	ASP	CB-CG-OD2	6.06	123.75	118.30

There are no chirality outliers.

All (4) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
2	B	135	SER	Peptide
2	B	244	SER	Peptide

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Group
2	B	265	GLN	Peptide
2	B	274	PHE	Peptide

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	1773	0	1771	54	0
2	B	3962	0	3806	132	0
All	All	5735	0	5577	186	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 16.

The worst 5 of 186 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)
2:B:245:TYR:HB2	2:B:246:PRO:HD3	1.36	1.06
2:B:85:LYS:H	2:B:85:LYS:HD2	1.35	0.92
2:B:256:ASN:HD22	2:B:367:ILE:HD11	1.38	0.86
2:B:332:GLN:NE2	2:B:469:ARG:H	1.77	0.83
2:B:269:LEU:H	2:B:269:LEU:CD2	1.92	0.83

There are no symmetry-related clashes.

5.3 Torsion angles ⓘ

5.3.1 Protein backbone ⓘ

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	226/234 (97%)	203 (90%)	16 (7%)	7 (3%)	5	34
2	B	491/551 (89%)	415 (84%)	57 (12%)	19 (4%)	4	28
All	All	717/785 (91%)	618 (86%)	73 (10%)	26 (4%)	4	30

5 of 26 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	586	ARG
2	B	87	GLY
2	B	270	ASP
2	B	272	GLN
2	B	483	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	192/198 (97%)	143 (74%)	49 (26%)	1	2
2	B	450/498 (90%)	337 (75%)	113 (25%)	1	2
All	All	642/696 (92%)	480 (75%)	162 (25%)	1	2

5 of 162 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	B	138	ARG
2	B	205	TYR
2	B	480	LEU
2	B	142	GLN
2	B	178	SER

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 12 such sidechains are listed below:

Mol	Chain	Res	Type
2	B	129	GLN

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
2	B	256	ASN
2	B	332	GLN
2	B	61	HIS
2	B	318	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, 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	228/234 (97%)	-0.18	4 (1%)	71	60	23, 44, 92, 126	0
2	B	499/551 (90%)	0.06	15 (3%)	54	41	6, 52, 111, 135	0
All	All	727/785 (92%)	-0.01	19 (2%)	59	47	6, 49, 105, 135	0

The worst 5 of 19 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	680	GLU	3.7
2	B	207	PRO	3.4
2	B	416	TYR	2.9
1	A	722	THR	2.8
2	B	111	VAL	2.7

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.