



# wwPDB X-ray Structure Validation Summary Report

Feb 27, 2014 – 08:56 AM GMT

PDB ID : 3S5K  
Title : Crystal structures of falcilysin, a M16 metalloprotease from the malaria parasite *Plasmodium falciparum*  
Authors : Morgunova, E.; Ponpuak, M.; Istvan, E.; Popov, A.; Goldberg, D.; Eneqvist, T.  
Deposited on : 2011-05-23  
Resolution : 3.20 Å(reported)

This is a wwPDB validation summary report for a publicly released PDB entry.  
We welcome your comments at [validation@mail.wwpdb.org](mailto:validation@mail.wwpdb.org)  
A user guide is available at <http://wwpdb.org/ValidationPDFNotes.html>

---

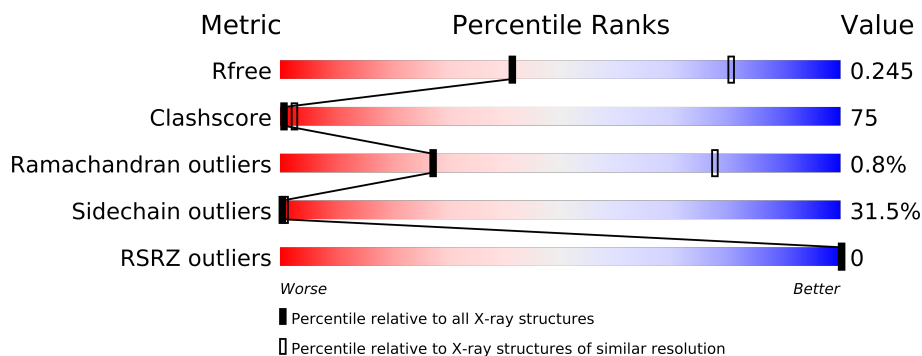
The following versions of software and data (see [references](#)) were used in the production of this report:

MolProbity : 4.02b-467  
Mogul : 1.15 2013  
Xtriage (Phenix) : dev-1323  
EDS : stable22639  
Percentile statistics : 21963  
Refmac : 5.8.0049  
CCP4 : 6.3.0 (Settle)  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et. al. (1996)  
Validation Pipeline (wwPDB-VP) : stable22683

# 1 Overall quality at a glance

The reported resolution of this entry is 3.20 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))
$R_{free}$	66092	1824 (3.30-3.10)
Clashscore	79885	1078 (3.26-3.14)
Ramachandran outliers	78287	1059 (3.26-3.14)
Sidechain outliers	78261	1058 (3.26-3.14)
RSRZ outliers	66119	1825 (3.30-3.10)

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  $\geq 3$ , 2, 1 and 0 types of geometric quality criteria. The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density.

Mol	Chain	Length	Quality of chain
1	A	1193	

## 2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 8741 atoms, of which 0 are hydrogen and 0 are deuterium.

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 Falcilysin.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	1053	Total	C	N	O	S	0	0	0
			8674	5577	1418	1653	26			

- Molecule 2 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	1	Total	Zn	0	0
			1	1		

- Molecule 3 is water.

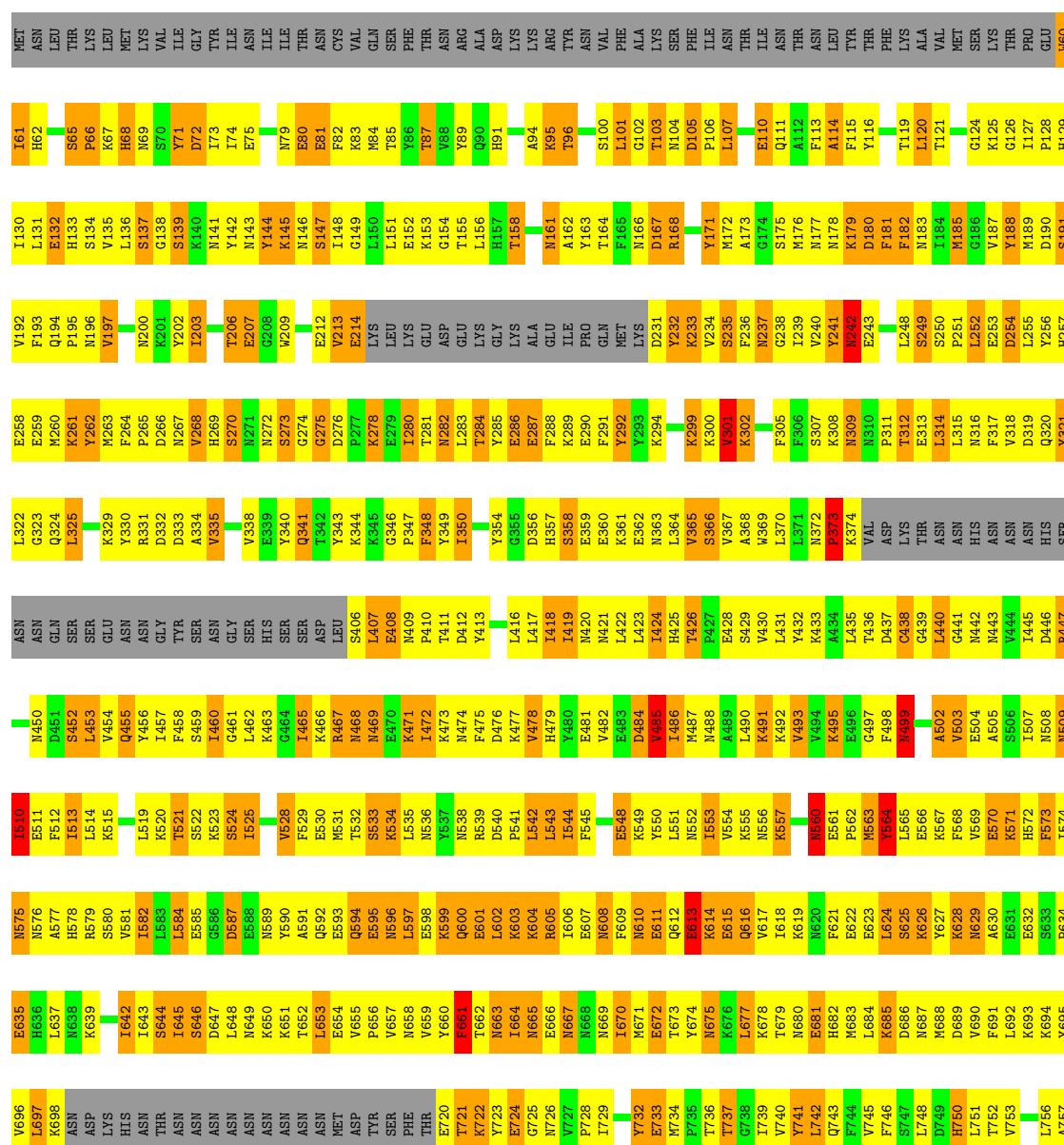
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	66	Total	O	0	0
			66	66		

### 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: Falcilysin

Chain A:



K1151	F1152	A1153	D1154	L1155	L1156	E1157	S1158	K1159	V1160	N1161	E1162	F1163	E1164	K1165	V1168	I1169	I1170	T1171	T1172	K1173	E1174	K1175	A1176	N1177	E1178	Y1179	I1180	A1181	N1182	V1183	D1184	G1185	E1186	F1187	K1188	K1189	V1190	L1191	I1192	E1193																			
L1085	R1086	K1087	M1088	T1091	M1092	T1093	E1094	N1095	D1096	L1097	L1098	R1099	Y1100	I1101	I1102	N1103	T1104	T1105	G1106	T1107	T1108	D1109	K1110	P1111	R1112	R1113	G1114	I1115	E1116	L1117	S1118	K1119	L1120	S1121	F1122	I1126	S1127	N1128	Q1132	D1133	R1134	F1137	R1138	K1139	R1140	I1141	M1142	N1143	T1144	K1145	E1146	E1147	D1148	F1149	Y1150				
G1019	E1020	Y1021	L1022	P1023	P1024	S1025	F1026	T1027	V1028	I1029	V1030	A1031	A1032	L1033	K1034	N1035	Y1037	L1038	W1039	D1040	T1041	V1042	R1043	G1047	A1048	Y1049	G1050	V1051	F1052	A1053	D1054	E1055	Y1056	D1057	D1058	G1059	S1060	V1061	V1062	F1063	L1064	S1065	A1066	R1067	D1068	P1069	N1070	L1071	E1072	K1073	T1077	F1078	R1079	K1083	G1084				
S956	Y957	F958	E959	E960	N961	D962	K963	Y964	Y965	ASN	ASP	MET	GLN	ASN	D908	F909	K910	Y911	L912	R915	L916	Y917	R918	T919	R920	N921	K922	Y923	F924	N925	K926	K927	N928	L929	N930	Y933	T934	S935	D936	Y937	G938	L940	K941	H942	T1004	F1005	F944	V1006	N1007	Y946	S947	S1010	N948	M1011	I1014	L1015	F1016	K1017	P1018
C826	N827	D828	I832	A833	L834	E835	A836	L837	K838	E839	S840	D841	S843	K846	K847	V848	I849	D850	I851	L852	K853	R854	K855	L856	N857	G858	K859	K860	T861	T862	F863	S864	E865	Y868	A869	I870	L871	M872	K873	Y874	V875	K876	K882	A885	H886	N887	I888	E889	Y890	G891	L892	S893							
Y758	L759	N760	L761	F762	K763	T764	L765	L766	L767	E768	N769	K770	T771	R774	F779	W780	L781	L782	R783	N786	L787	G788	S789	W790	S791	A792	Y797	S798	K799	D800	D801	L802	N803	N804	V805	T806	D807	K808	Y809	N810	A811	Q812	A813	L814	F815	N816	L817	E818	X819	H820	Y821	L822	S823						

## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	93.78Å 105.29Å 114.76Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	42.83 – 3.20 42.83 – 2.84	Depositor EDS
% Data completeness (in resolution range)	99.9 (42.83-3.20) 95.4 (42.83-2.84)	Depositor EDS
$R_{merge}$	0.26	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.24 (at 2.86Å)	Xtriage
Refinement program	PHENIX (PHENIX.REFINE: 1.6_289)	Depositor
R, $R_{free}$	0.210 , 0.316 0.220 , 0.245	Depositor DCC
$R_{free}$ test set	800 reflections (4.32%)	DCC
Wilson B-factor (Å <sup>2</sup> )	50.1	Xtriage
Anisotropy	0.258	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.36 , 50.7	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning	$\langle  L  \rangle = 0.42$ , $\langle L^2 \rangle = 0.24$	Xtriage
Outliers	0 of 26160 reflections	Xtriage
$F_o, F_c$ correlation	0.89	EDS
Total number of atoms	8741	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	36.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.34% of the height of the origin peak. No significant pseudotranslation is detected.*

<sup>1</sup>Intensities estimated from amplitudes.

## 5 Model quality ⓘ

### 5.1 Standard geometry ⓘ

Bond lengths and bond angles in the following residue types are not validated in this section: 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	1.79	58/8849 (0.7%)	1.11	36/11918 (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
1	A	0	1

The worst 5 of 58 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	438	CYS	CB-SG	-10.15	1.65	1.82
1	A	826	CYS	CB-SG	-7.60	1.69	1.82
1	A	818	GLU	CG-CD	-7.31	1.41	1.51
1	A	188	TYR	CE1-CZ	-6.19	1.30	1.38
1	A	821	VAL	CA-CB	-6.18	1.41	1.54

The worst 5 of 36 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	915	ILE	CB-CA-C	-7.40	96.80	111.60
1	A	365	VAL	CB-CA-C	-6.51	99.02	111.40
1	A	945	VAL	CB-CA-C	-6.49	99.07	111.40
1	A	837	VAL	CB-CA-C	-6.44	99.16	111.40
1	A	856	ILE	CB-CA-C	-6.38	98.83	111.60

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	560	ASN	Peptide

## 5.2 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 hydrogens added by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, and the number in parentheses is this value normalized per 1000 atoms of the molecule in the chain. The Symm-Clashes column gives symmetry related clashes, in the same way as for the Clashes column.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	8674	0	8613	1304	2
2	A	1	0	0	0	0
3	A	66	0	0	6	0
All	All	8741	0	8613	1304	2

Clashscore is defined as the number of clashes calculated for the entry per 1000 atoms (including hydrogens) of the entry. The overall clashscore for this entry is 75.

The worst 5 of 1304 close contacts within the same asymmetric unit are listed below.

Atom-1	Atom-2	Distance(Å)	Clash(Å)
1:A:263:MET:CE	1:A:457:ILE:HD12	1.14	1.54
1:A:610:ASN:C	1:A:613:GLU:CG	1.75	1.50
1:A:232:TYR:HB2	1:A:614:LYS:NZ	1.30	1.45
1:A:151:LEU:O	1:A:155:THR:CG2	1.67	1.42
1:A:612:GLN:CB	1:A:616:GLN:HB3	1.49	1.42

All (2) 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	Distance(Å)	Clash(Å)
1:A:608:ASN:ND2	1:A:988:LYS:CG[2_554]	1.98	0.22
1:A:538:ASN:O	1:A:679:THR:OG1[2_654]	2.10	0.10

## 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	1043/1193 (87%)	998 (96%)	37 (4%)	8 (1%)	27 77

5 of 8 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	213	VAL
1	A	610	ASN
1	A	613	GLU
1	A	1128	ASN
1	A	373	PRO

### 5.3.2 Protein sidechains ⓘ

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	972/1104 (88%)	666 (68%)	306 (32%)	0 1

5 of 306 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	564	TYR
1	A	645	ILE
1	A	1127	SER
1	A	571	LYS
1	A	604	LYS

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 33 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	538	ASN
1	A	600	GLN
1	A	953	ASN
1	A	560	ASN
1	A	572	HIS

### 5.3.3 RNA ⓘ

There are no RNA chains in this entry.

### 5.4 Non-standard residues in protein, DNA, RNA chains ⓘ

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

### 5.5 Carbohydrates ⓘ

There are no carbohydrates in this entry.

### 5.6 Ligand geometry ⓘ

Of 1 ligands modelled in this entry, 1 is monoatomic - leaving 0 for Mogul analysis.

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

### 5.7 Other polymers ⓘ

There are no such residues in this entry.

### 5.8 Polymer linkage issues

There are no chain breaks in this entry.

## 6 Fit of model and data ⓘ

### 6.1 Protein, DNA and RNA chains ⓘ

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<sup>th</sup> 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 <sup>2</sup> )	Q<0.9
1	A	1053/1193 (88%)	-0.37	0 100 100	9, 32, 70, 128	0

There are no RSRZ outliers to report.

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

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

### 6.3 Carbohydrates ⓘ

There are no carbohydrates in this entry.

### 6.4 Ligands ⓘ

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, 95<sup>th</sup> 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	RSR	LLDF	B-factors(Å <sup>2</sup> )	Q<0.9
2	ZN	A	1194	1/1	0.21	-0.41	38,38,38,38	0

### 6.5 Other polymers ⓘ

There are no such residues in this entry.