



# wwPDB X-ray Structure Validation Summary Report

Feb 27, 2014 – 11:54 PM GMT

PDB ID : 3ISR  
Title : The Crystal Structure of a Putative Cysteine Protease from *Cytophaga hutchinsonii* to 1.9Å  
Authors : Stein, A.J.; Bigelow, L.; Trevino, D.; Buck, K.; Joachimiak, A.; Midwest Center for Structural Genomics (MCSG)  
Deposited on : 2009-08-27  
Resolution : 1.90 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>

---

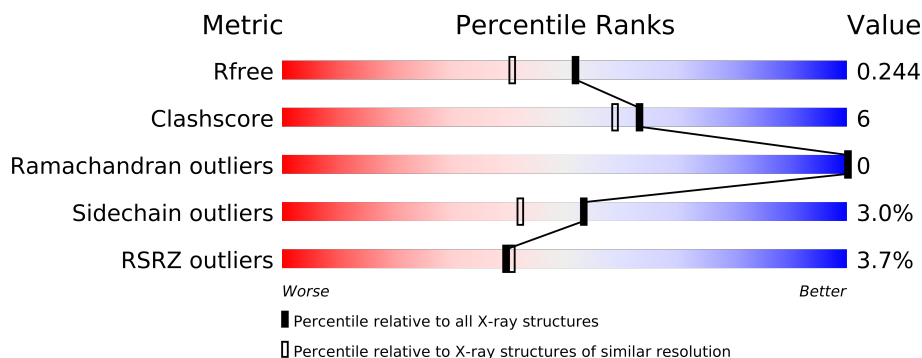
The following versions of software and data (see [references](#)) were used in the production of this report:

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 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 (#Entries)	Similar resolution (#Entries, resolution range(Å))
$R_{free}$	66092	3684 (1.90-1.90)
Clashscore	79885	4465 (1.90-1.90)
Ramachandran outliers	78287	4413 (1.90-1.90)
Sidechain outliers	78261	4414 (1.90-1.90)
RSRZ outliers	66119	3686 (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 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	293	
1	B	293	
1	C	293	
1	D	293	

## 2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 9509 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 Transglutaminase-likeenzymes, putative cysteine protease.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	286	Total	C	N	O	S	Se	0	4	0
			2228	1436	362	421	5	4			
1	B	289	Total	C	N	O	S	Se	0	0	0
			2234	1437	364	424	5	4			
1	C	285	Total	C	N	O	S	Se	0	1	0
			2228	1430	364	425	5	4			
1	D	289	Total	C	N	O	S	Se	0	0	0
			2254	1451	368	427	5	3			

There are 44 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-10	GLY	-	expression tag	UNP Q11RW0
A	-9	THR	-	expression tag	UNP Q11RW0
A	-8	GLU	-	expression tag	UNP Q11RW0
A	-7	ASN	-	expression tag	UNP Q11RW0
A	-6	LEU	-	expression tag	UNP Q11RW0
A	-5	TYR	-	expression tag	UNP Q11RW0
A	-4	PHE	-	expression tag	UNP Q11RW0
A	-3	GLN	-	expression tag	UNP Q11RW0
A	-2	SER	-	expression tag	UNP Q11RW0
A	-1	ASN	-	expression tag	UNP Q11RW0
A	0	ALA	-	expression tag	UNP Q11RW0
B	-10	GLY	-	expression tag	UNP Q11RW0
B	-9	THR	-	expression tag	UNP Q11RW0
B	-8	GLU	-	expression tag	UNP Q11RW0
B	-7	ASN	-	expression tag	UNP Q11RW0
B	-6	LEU	-	expression tag	UNP Q11RW0
B	-5	TYR	-	expression tag	UNP Q11RW0
B	-4	PHE	-	expression tag	UNP Q11RW0
B	-3	GLN	-	expression tag	UNP Q11RW0
B	-2	SER	-	expression tag	UNP Q11RW0
B	-1	ASN	-	expression tag	UNP Q11RW0

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
B	0	ALA	-	expression tag	UNP Q11RW0
C	-10	GLY	-	expression tag	UNP Q11RW0
C	-9	THR	-	expression tag	UNP Q11RW0
C	-8	GLU	-	expression tag	UNP Q11RW0
C	-7	ASN	-	expression tag	UNP Q11RW0
C	-6	LEU	-	expression tag	UNP Q11RW0
C	-5	TYR	-	expression tag	UNP Q11RW0
C	-4	PHE	-	expression tag	UNP Q11RW0
C	-3	GLN	-	expression tag	UNP Q11RW0
C	-2	SER	-	expression tag	UNP Q11RW0
C	-1	ASN	-	expression tag	UNP Q11RW0
C	0	ALA	-	expression tag	UNP Q11RW0
D	-10	GLY	-	expression tag	UNP Q11RW0
D	-9	THR	-	expression tag	UNP Q11RW0
D	-8	GLU	-	expression tag	UNP Q11RW0
D	-7	ASN	-	expression tag	UNP Q11RW0
D	-6	LEU	-	expression tag	UNP Q11RW0
D	-5	TYR	-	expression tag	UNP Q11RW0
D	-4	PHE	-	expression tag	UNP Q11RW0
D	-3	GLN	-	expression tag	UNP Q11RW0
D	-2	SER	-	expression tag	UNP Q11RW0
D	-1	ASN	-	expression tag	UNP Q11RW0
D	0	ALA	-	expression tag	UNP Q11RW0

- Molecule 2 is water.

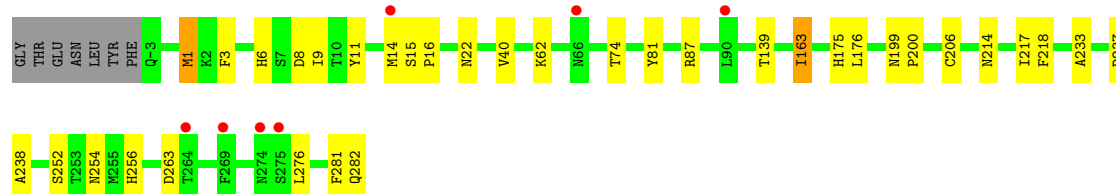
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	160	Total O 160 160	0	0
2	B	152	Total O 152 152	0	0
2	C	138	Total O 138 138	0	0
2	D	115	Total O 115 115	0	0

### 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.

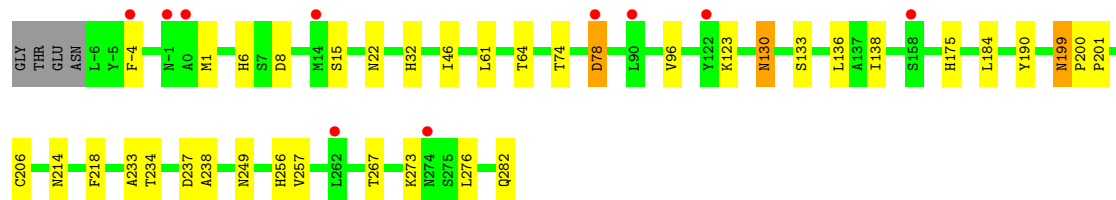
- Molecule 1: Transglutaminase-likeenzymes, putative cysteine protease

Chain A: 



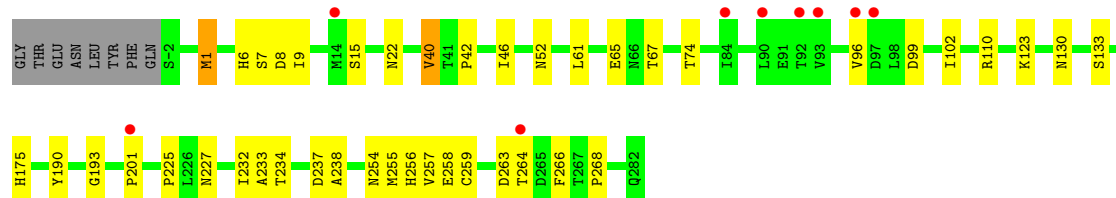
- Molecule 1: Transglutaminase-likeenzymes, putative cysteine protease

Chain B: 



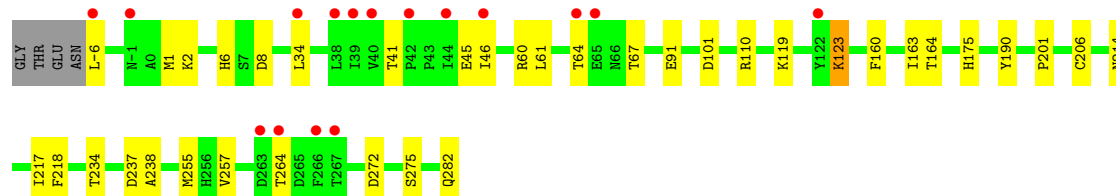
- Molecule 1: Transglutaminase-likeenzymes, putative cysteine protease

Chain C: 



- Molecule 1: Transglutaminase-likeenzymes, putative cysteine protease

Chain D: 



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	90.72Å 115.24Å 131.26Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	48.28 – 1.90 48.28 – 1.90	Depositor EDS
% Data completeness (in resolution range)	99.3 (48.28-1.90) 99.3 (48.28-1.90)	Depositor EDS
$R_{merge}$	0.07	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.97 (at 1.90Å)	Xtriage
Refinement program	REFMAC 5.5.0102	Depositor
R, $R_{free}$	0.187 , 0.220 0.214 , 0.244	Depositor DCC
$R_{free}$ test set	5392 reflections (4.98%)	DCC
Wilson B-factor (Å <sup>2</sup> )	31.9	Xtriage
Anisotropy	0.068	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.34 , 24.7	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning	$\langle  L  \rangle = 0.50$ , $\langle L^2 \rangle = 0.33$	Xtriage
Outliers	2 of 108299 reflections (0.002%)	Xtriage
$F_o, F_c$ correlation	0.95	EDS
Total number of atoms	9509	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	17.0	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 27.39 % 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 2.2201e-03. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

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

## 5 Model quality

### 5.1 Standard geometry

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	0.57	0/2291	0.63	0/3118
1	B	0.55	0/2287	0.60	0/3116
1	C	0.52	0/2282	0.59	0/3103
1	D	0.52	0/2307	0.60	0/3139
All	All	0.54	0/9167	0.61	0/12476

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no planarity outliers.

### 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	2228	0	2147	29	0
1	B	2234	0	2116	22	0
1	C	2228	0	2142	29	0
1	D	2254	0	2157	22	0
2	A	160	0	0	1	0
2	B	152	0	0	1	0
2	C	138	0	0	1	0
2	D	115	0	0	1	0
All	All	9509	0	8562	97	0

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 6.

The worst 5 of 97 close contacts within the same asymmetric unit are listed below.

Atom-1	Atom-2	Distance(Å)	Clash(Å)
1:A:1:MSE:HE1	1:A:81:TYR:CE2	2.11	0.85
1:C:175:HIS:HE1	1:C:238:ALA:H	1.32	0.75
1:A:1:MSE:HE1	1:A:81:TYR:HE2	1.52	0.75
1:A:9:ILE:HD11	1:A:252:SER:HB2	1.69	0.74
1:D:175:HIS:HE1	1:D:238:ALA:H	1.35	0.74

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	288/293 (98%)	282 (98%)	6 (2%)	0	100	100
1	B	287/293 (98%)	279 (97%)	8 (3%)	0	100	100
1	C	284/293 (97%)	277 (98%)	7 (2%)	0	100	100
1	D	287/293 (98%)	282 (98%)	5 (2%)	0	100	100
All	All	1146/1172 (98%)	1120 (98%)	26 (2%)	0	100	100

There are no Ramachandran outliers to report.

### 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	238/251 (95%)	234 (98%)	4 (2%)	73	68

*Continued on next page...*



*Continued from previous page...*

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	B	236/251 (94%)	228 (97%)	8 (3%)	49	36
1	C	240/251 (96%)	234 (98%)	6 (2%)	60	50
1	D	240/251 (96%)	230 (96%)	10 (4%)	40	27
All	All	954/1004 (95%)	926 (97%)	28 (3%)	53	44

5 of 28 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	1	MSE
1	C	123	LYS
1	D	264	THR
1	C	40	VAL
1	C	67	THR

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 27 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	199	ASN
1	C	6	HIS
1	D	175	HIS
1	B	214	ASN
1	A	214	ASN

### 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 ⓘ

There are no ligands in this entry.

## 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	286/293 (97%)	0.44	7 (2%) 56 57	3, 16, 22, 30	0
1	B	289/293 (98%)	0.37	10 (3%) 42 43	5, 16, 21, 29	0
1	C	285/293 (97%)	0.25	9 (3%) 45 46	4, 17, 23, 28	0
1	D	289/293 (98%)	0.50	16 (5%) 24 24	8, 18, 22, 25	0
All	All	1149/1172 (98%)	0.39	42 (3%) 39 41	3, 17, 22, 30	0

The worst 5 of 42 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	D	266	PHE	4.6
1	C	96	VAL	4.5
1	D	44	ILE	4.3
1	D	46	ILE	3.7
1	B	0	ALA	3.3

### 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

There are no ligands in this entry.

## 6.5 Other polymers ⓘ

There are no such residues in this entry.