wwPDB X-ray Structure Validation Summary Report (i)

Sep 27, 2023 – 04:35 PM EDT

PDB ID	:	1K32
Title	:	Crystal structure of the tricorn protease
Authors	:	Brandstetter, H.; Kim, J.-S.; Groll, M.; Huber, R.
Deposited on	:	2001-10-01
Resolution	:	2.00 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
https://www.wwpdb.org/validation/2017/XrayValidationReportHelp
with specific help available everywhere you see the (i) symbol.

The types of validation reports are described at
http://www.wwpdb.org/validation/2017/FAQs#types.

The following versions of software and data (see references (1)) were used in the production of this report:

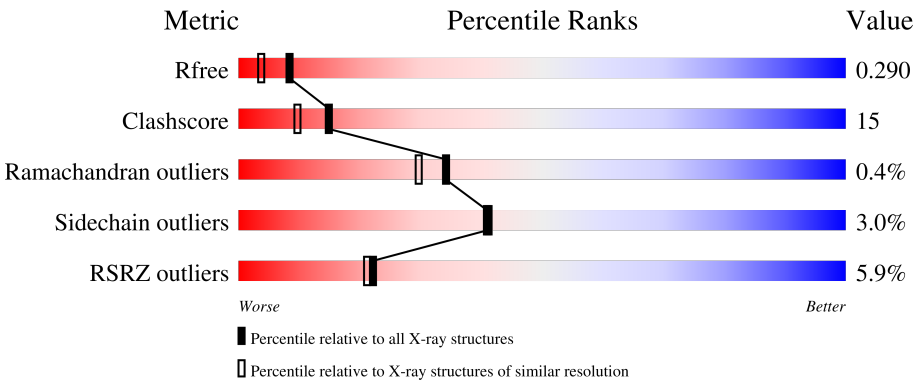
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.35.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.35.1

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

The reported resolution of this entry is 2.00 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.



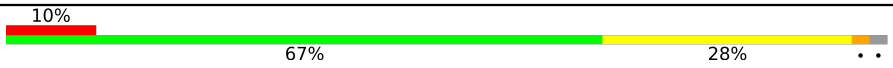
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	8085 (2.00-2.00)
Clashscore	141614	9178 (2.00-2.00)
Ramachandran outliers	138981	9054 (2.00-2.00)
Sidechain outliers	138945	9053 (2.00-2.00)
RSRZ outliers	127900	7900 (2.00-2.00)

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 of the lower bar indicate the fraction of residues that contain outliers for >=3, 2, 1 and 0 types of geometric quality criteria respectively. 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	1045	3% 72% 23% • •
1	B	1045	3% 71% 24% • •
1	C	1045	11% 70% 26% • •
1	D	1045	4% 73% 23% • •
1	E	1045	3% 73% 23% • •

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	F	1045	

2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 51456 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 tricorn protease.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	1023	8177	5196	1402	1551	28	95	0	0
1	B	1023	8177	5196	1402	1551	28	95	0	0
1	C	1023	8177	5196	1402	1551	28	95	0	0
1	D	1023	8177	5196	1402	1551	28	95	0	0
1	E	1023	8177	5196	1402	1551	28	95	0	0
1	F	1023	8177	5196	1402	1551	28	95	0	0

- Molecule 2 is water.

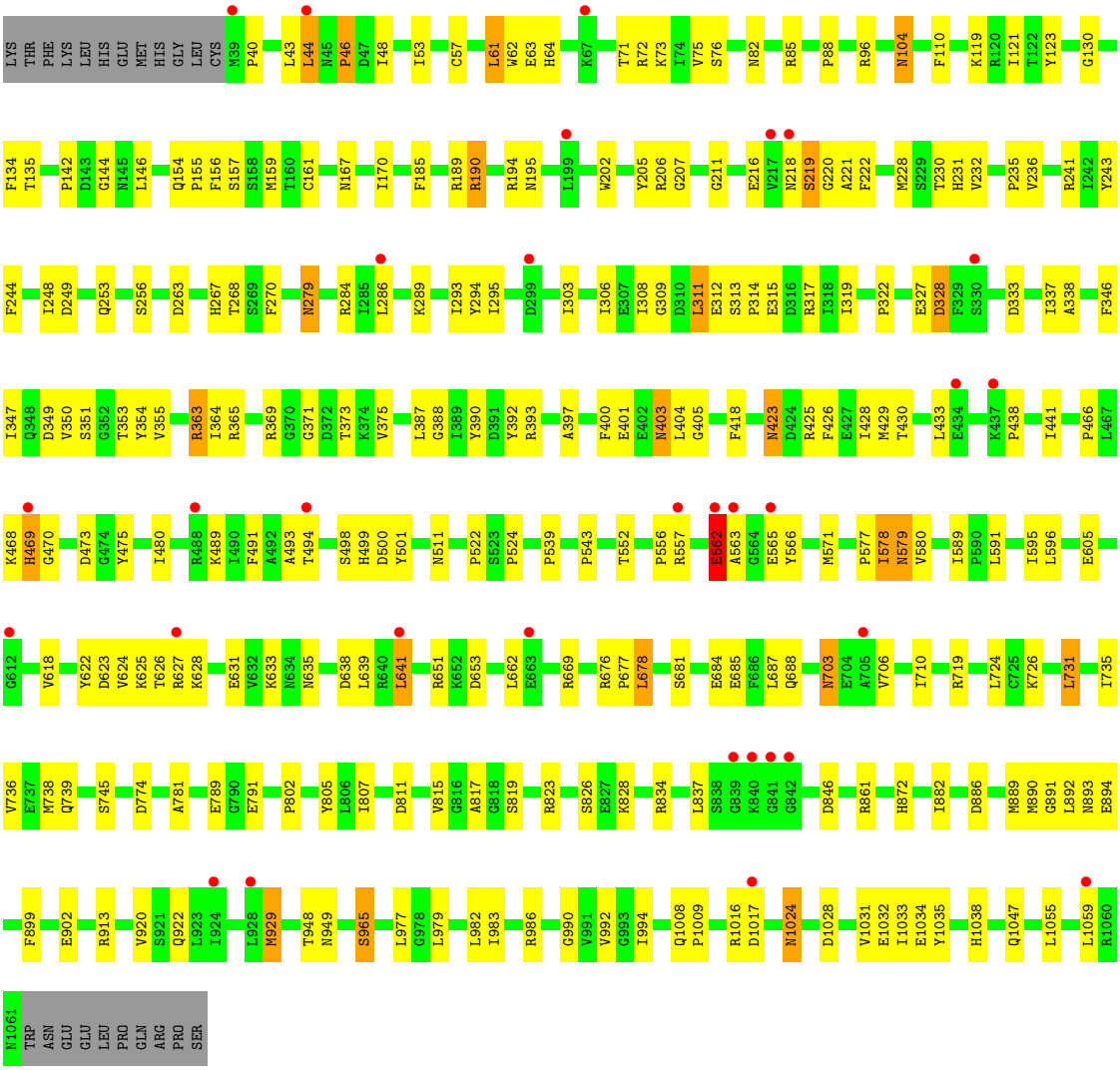
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
2	A	401	401	401	0	0
2	B	395	395	395	0	0
2	C	398	398	398	0	0
2	D	401	401	401	0	0
2	E	405	405	405	0	0
2	F	394	394	394	0	0

3 Residue-property plots (i)

These plots are drawn for all protein, RNA, DNA and oligosaccharide 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 (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.

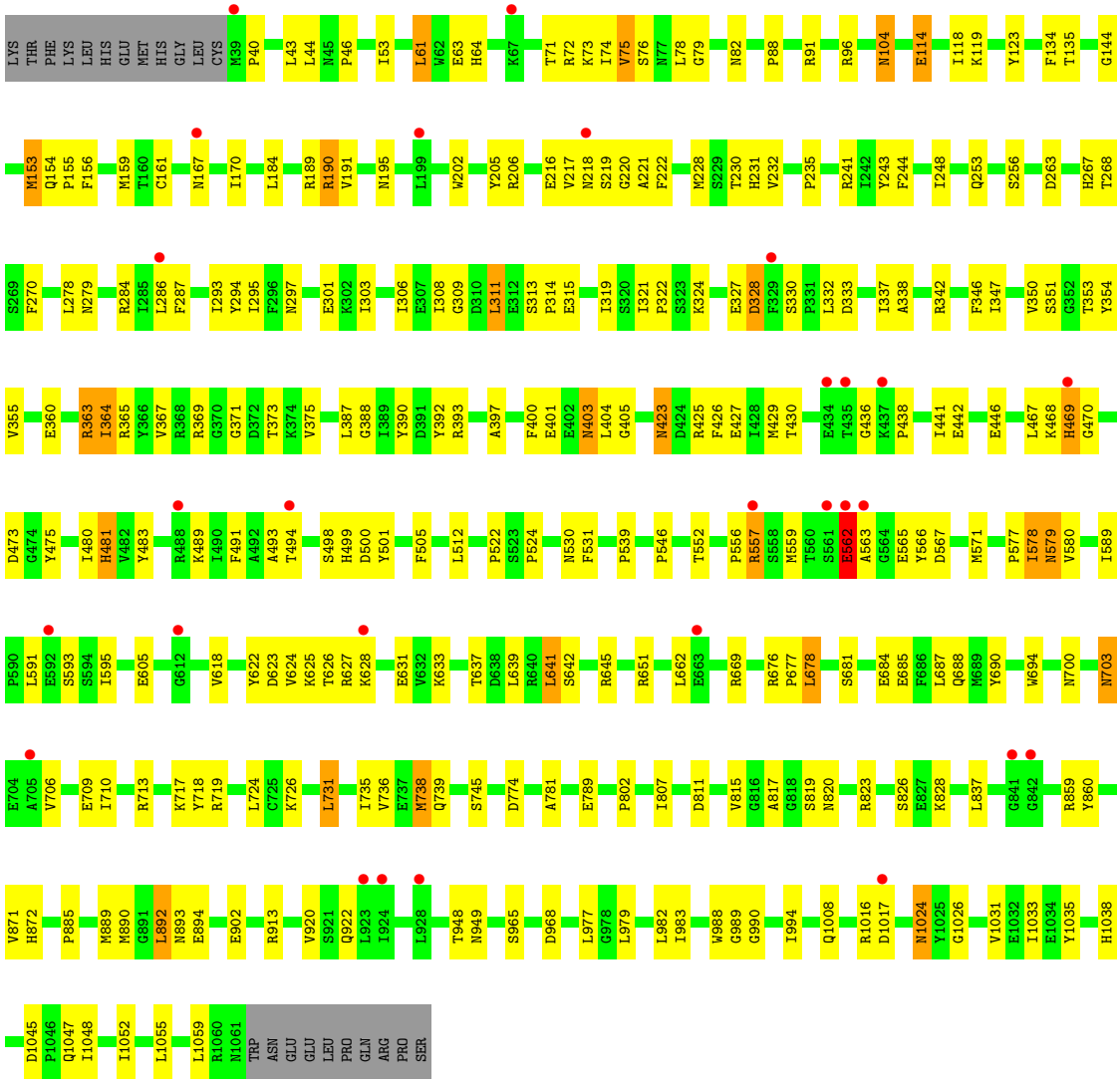
• Molecule 1: tricorn protease

Chain A: 3% 72% 23%



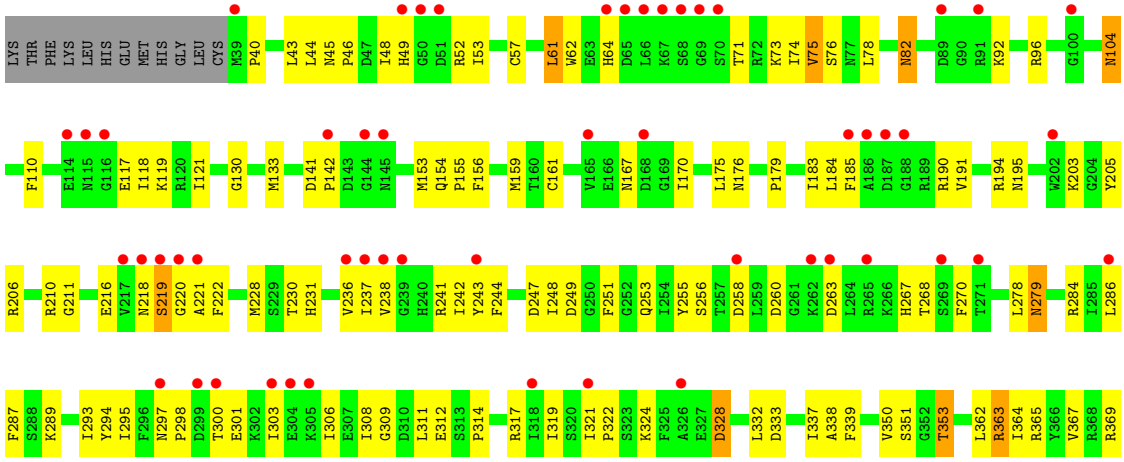
• Molecule 1: tricorn protease

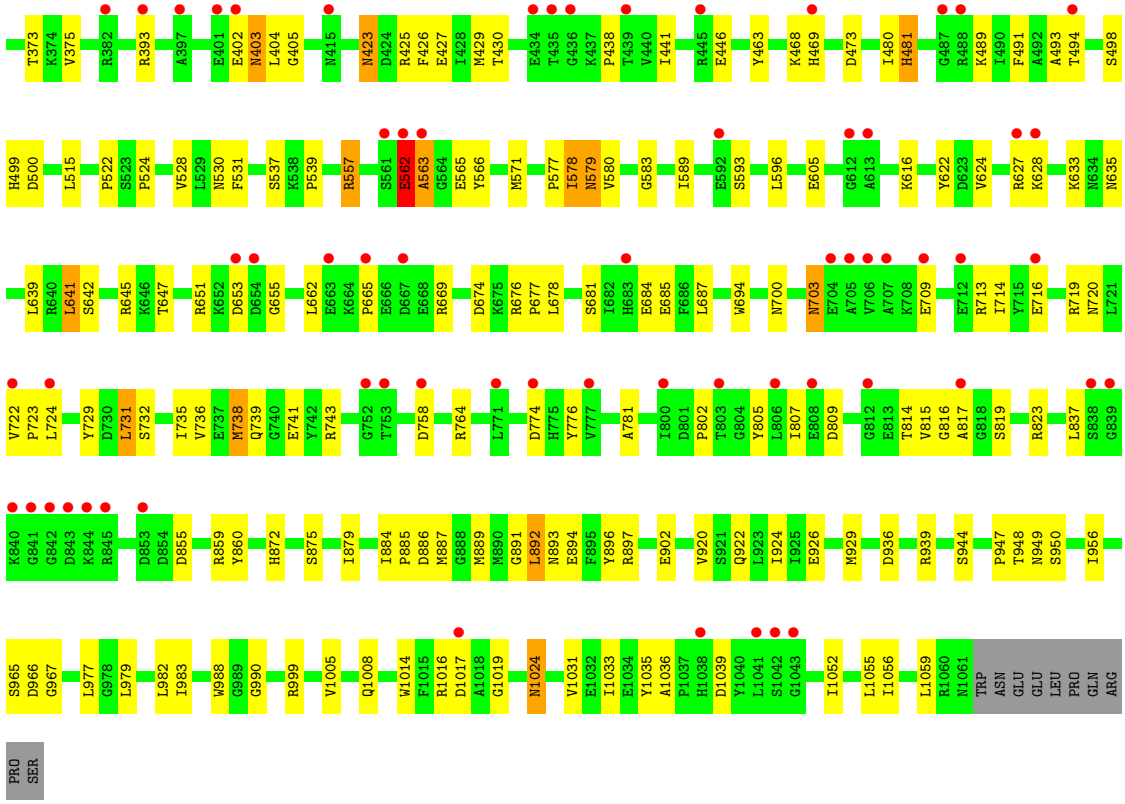
Chain B: 3% 71% 24%



- Molecule 1: tricorn protease

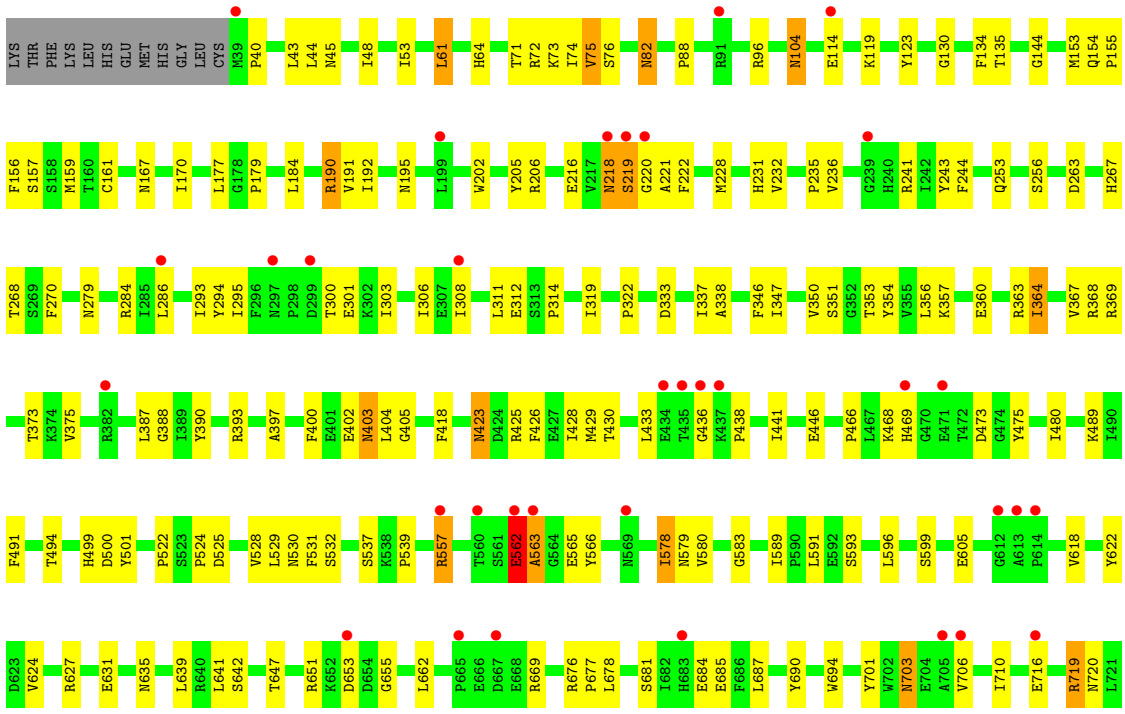
Chain C:

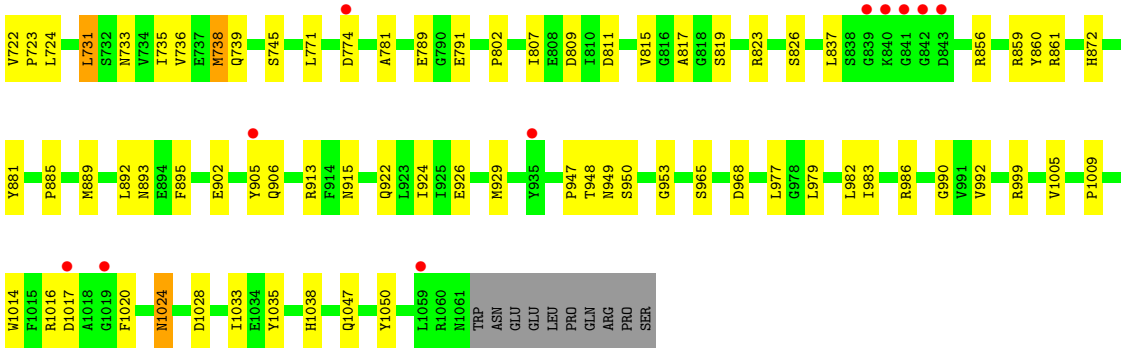




- Molecule 1: tricorn protease

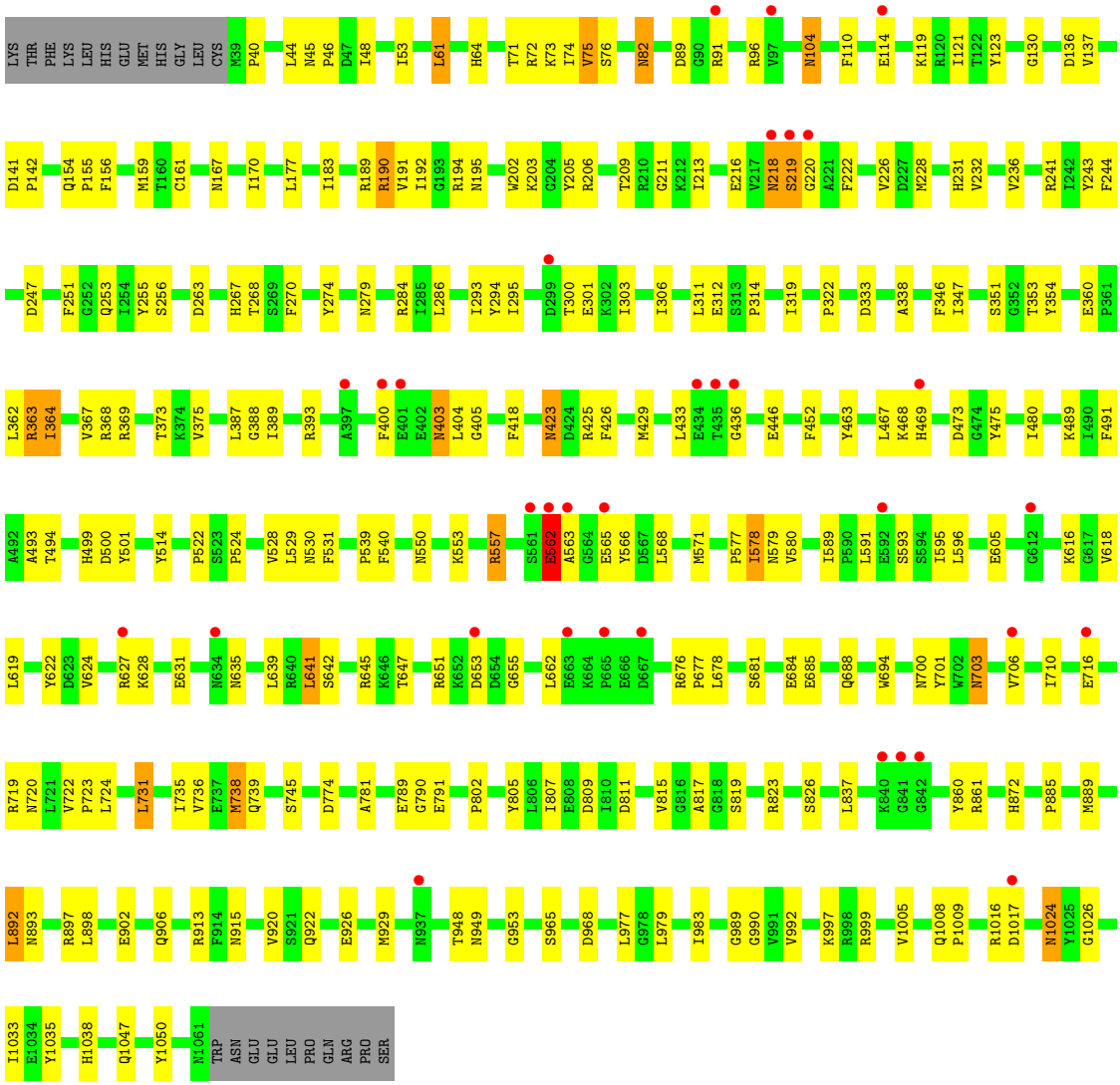
Chain D:





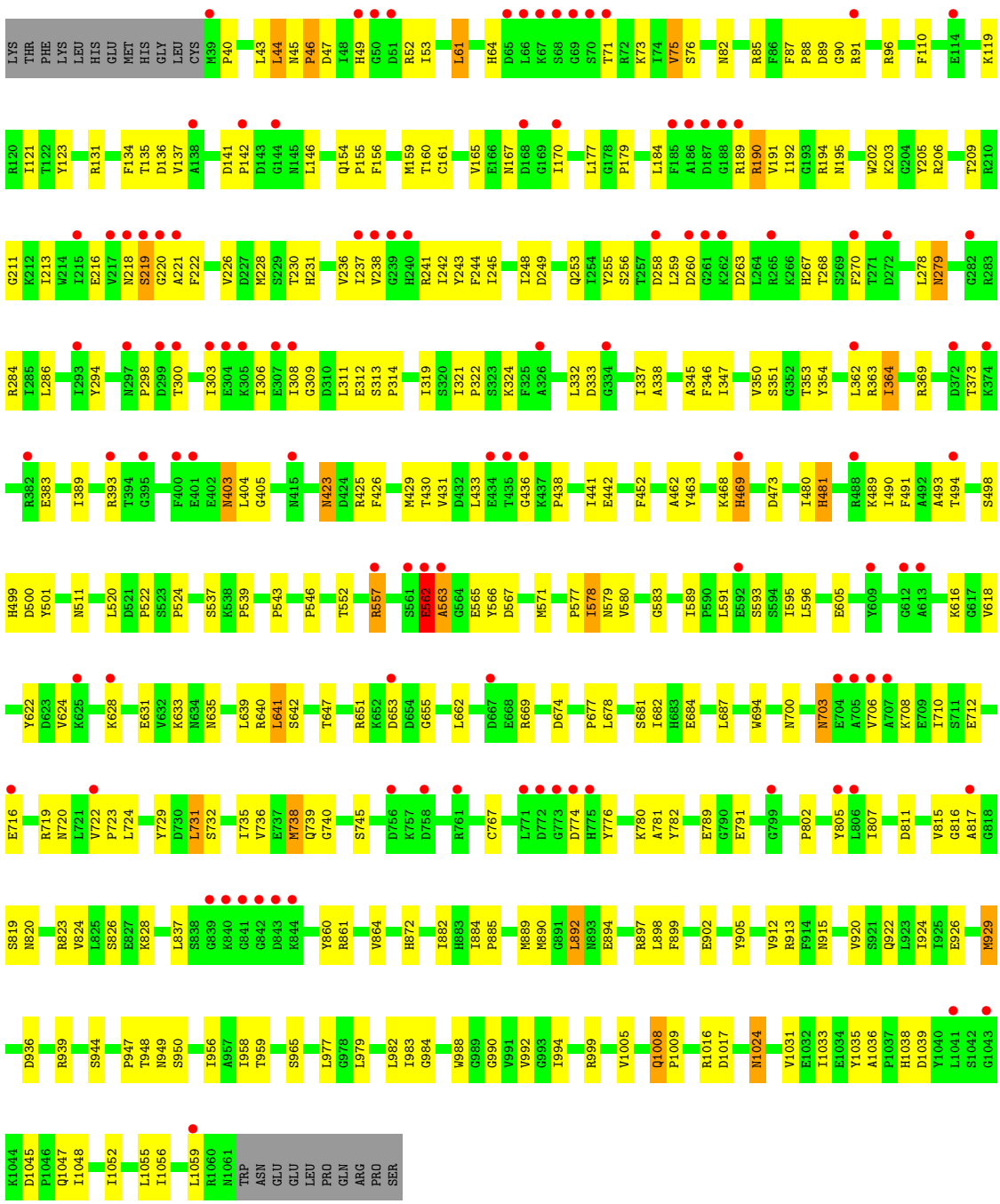
- Molecule 1: tricorn protease

Chain E:



- Molecule 1: tricorn protease

Chain F:


LYS THR PHE LYS LEU HIS GLU MET HIS GLY LEU CYS M39 P40 L43 L44 N45 P46 D47 I48 H49 G50 D51 R52 I53 L61 H64 D65 L66 K67 S68 G69 S70 T71 R72 K73 I74 V75 S76 N82 R85 F86 F87 P88 D89 G90 R91 R96 F110 E114 K119
R120 I121 T122 Y123 R131 F134 T135 D136 V137 A138 D141 P142 D143 G144 N145 L146 Q154 P155 F156 M159 T160 C161 V165 E166 N167 D168 G169 I170 L177 G178 P179 L184 F185 A186 D187 G188 R189 R190 V191 I192 G193 R194 N195 W202 K203 G204 Y205 R206 T209 R210
G211 K212 I213 W214 I215 E216 V217 N218 S219 G220 A221 F222 V226 D227 M228 S229 T230 H231 V236 I237 V238 G239 H240 R241 I242 Y243 F244 I245 I248 D249 Q253 I254 Y255 S256 T257 D258 L259 D260 G261 K262 D263 L264 R265 K266 H267 T268 S269 F270 T271 D272 L278 N279 G282 R283
R284 I285 L286 I293 Y294 N297 P298 D299 T300 I303 E304 K305 I306 E307 I308 G309 D310 L311 E312 S313 P314 I319 S320 I321 P322 S323 K324 F325 A326 L332 D333 G334 I337 A338 A345 F346 I347 V350 S351 G352 T353 Y354 L362 R363 I364 R369 D372 T373 K374
R382 E383 I389 R393 T394 G395 F400 E401 E402 N403 L404 G405 N415 N423 D424 R425 F426 M429 T430 V431 D432 L433 E434 T435 G436 K437 P438 I441 E442 F452 A462 Y463 K468 H469 D473 I480 H481 R488 K489 I490 F491 A492 A493 T494 S498
H499 D500 Y501 N511 L520 D521 P522 S523 P524 S537 K538 P539 P543 P546 T552 R557 S561 E562 A563 G564 E565 Y566 D567 M571 P577 I578 N579 V580 G583 I589 P590 L591 E592 S593 S594 I595 L596 E605 Y609 G612 A613 K616 G617 V618
Y622 D623 V624 K625 K628 E631 V632 K633 N634 N635 L639 R640 L641 S642 T647 R651 K652 D653 D654 G655 L662 D667 E668 R669 D674 P677 L678 S681 I682 H683 E684 L687 W694 N700 N703 E704 A705 V706 A707 K708 E709 I710 S711 E712
E716 R719 N720 L721 V722 P723 L724 Y729 D730 L731 S732 I735 V736 E737 M738 Q739 G740 S745 D756 K757 D758 R761 C767 L771 D772 G773 D774 H775 Y776 K780 A781 Y782 E789 G790 E791 G799 P802 Y805 L806 I807 D811 V815 G816 A817 G818
S819 N820 R823 V824 L825 S826 E827 K828 L837 S838 G839 K840 G841 G842 D843 K844 Y860 R861 V864 H872 I882 H883 I884 P885 M889 M890 G891 L892 N893 E894 R897 L898 F899 E902 Y905 V912 R913 F914 N915 V920 S921 Q922 L923 I924 I925 E926 M929
D936 R939 S944 P947 T948 N949 S950 I956 A957 I958 T959 S965 L977 G978 L979 L982 I983 G984 W988 G989 G990 V991 V992 G993 I994 R999 V1005 Q1008 P1009 R1016 D1017 N1024 V1031 E1032 I1033 E1034 Y1035 A1036 P1037 H1038 D1039 Y1040 L1041 S1042 G1043
K1044 D1045 P1046 Q1047 I1048 I1052 L1055 I1056 L1059 R1060 N1061 TRP ASN GLU GLU LEU PRO GLN ARG PRO SER

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	95.86Å 246.00Å 159.04Å 90.00° 105.30° 90.00°	Depositor
Resolution (Å)	20.00 – 2.00 37.48 – 1.98	Depositor EDS
% Data completeness (in resolution range)	78.6 (20.00-2.00) 81.6 (37.48-1.98)	Depositor EDS
R_{merge}	0.10	Depositor
R_{sym}	0.10	Depositor
$< I/\sigma(I) >$ [1]	1.33 (at 1.98Å)	Xtriage
Refinement program	CNS 1.0	Depositor
R, R_{free}	0.245 , 0.264 0.253 , 0.290	Depositor DCC
R_{free} test set	20086 reflections (5.02%)	wwPDB-VP
Wilson B-factor (Å^2)	25.8	Xtriage
Anisotropy	0.345	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.39 , 54.3	EDS
L-test for twinning[2]	$< \|L\| > = 0.48$, $< L^2 > = 0.31$	Xtriage
Estimated twinning fraction	0.054 for h,-k,-h-l	Xtriage
F_o,F_c correlation	0.91	EDS
Total number of atoms	51456	wwPDB-VP
Average B, all atoms (Å^2)	30.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.98% 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, 0.333 respectively for untwinned datasets, and 0.375, 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	0.40	0/8367	0.66	1/11311 (0.0%)
1	B	0.40	0/8367	0.66	2/11311 (0.0%)
1	C	0.36	0/8367	0.62	1/11311 (0.0%)
1	D	0.39	0/8367	0.65	1/11311 (0.0%)
1	E	0.38	0/8367	0.65	1/11311 (0.0%)
1	F	0.36	0/8367	0.62	1/11311 (0.0%)
All	All	0.38	0/50202	0.64	7/67866 (0.0%)

There are no bond length outliers.

The worst 5 of 7 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	C	364	ILE	N-CA-C	-6.20	94.25	111.00
1	A	364	ILE	N-CA-C	-5.63	95.81	111.00
1	E	364	ILE	N-CA-C	-5.41	96.39	111.00
1	B	79	GLY	N-CA-C	-5.28	99.89	113.10
1	D	364	ILE	N-CA-C	-5.23	96.87	111.00

There are no chirality outliers.

There are no planarity outliers.

5.2 Too-close contacts (i)

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	8177	0	8003	256	0
1	B	8177	0	8003	265	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	C	8177	0	8003	272	0
1	D	8177	0	8003	226	0
1	E	8177	0	8003	249	0
1	F	8177	0	8003	275	0
2	A	401	0	0	12	0
2	B	395	0	0	8	0
2	C	398	0	0	18	0
2	D	401	0	0	9	0
2	E	405	0	0	10	0
2	F	394	0	0	10	0
All	All	51456	0	48018	1453	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 15.

The worst 5 of 1453 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:F:155:PRO:HG2	1:F:159:MET:HE1	1.33	1.09
1:B:155:PRO:HG2	1:B:159:MET:HE1	1.36	1.07
1:C:155:PRO:HG2	1:C:159:MET:HE1	1.37	1.06
1:D:155:PRO:HG2	1:D:159:MET:HE1	1.35	1.03
1:B:983:ILE:HG23	1:B:1033:ILE:HD13	1.41	1.02

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	1021/1045 (98%)	979 (96%)	38 (4%)	4 (0%)	34 30
1	B	1021/1045 (98%)	977 (96%)	41 (4%)	3 (0%)	41 37

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	C	1021/1045 (98%)	961 (94%)	53 (5%)	7 (1%)	22	16
1	D	1021/1045 (98%)	979 (96%)	38 (4%)	4 (0%)	34	30
1	E	1021/1045 (98%)	975 (96%)	43 (4%)	3 (0%)	41	37
1	F	1021/1045 (98%)	971 (95%)	44 (4%)	6 (1%)	25	19
All	All	6126/6270 (98%)	5842 (95%)	257 (4%)	27 (0%)	34	30

5 of 27 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	328	ASP
1	B	328	ASP
1	C	219	SER
1	C	579	ASN
1	D	219	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	883/904 (98%)	857 (97%)	26 (3%)	42	43
1	B	883/904 (98%)	853 (97%)	30 (3%)	37	36
1	C	883/904 (98%)	860 (97%)	23 (3%)	46	48
1	D	883/904 (98%)	856 (97%)	27 (3%)	40	40
1	E	883/904 (98%)	857 (97%)	26 (3%)	42	43
1	F	883/904 (98%)	857 (97%)	26 (3%)	42	43
All	All	5298/5424 (98%)	5140 (97%)	158 (3%)	41	41

5 of 158 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	E	311	LEU
1	F	481	HIS

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	E	469	HIS
1	E	1024	ASN
1	F	719	ARG

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 193 such sidechains are listed below:

Mol	Chain	Res	Type
1	D	867	ASN
1	E	635	ASN
1	D	937	ASN
1	E	253	GLN
1	E	901	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 monosaccharides 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	1022/1045 (97%)	0.23	31 (3%)	50	49	14, 26, 43, 56	19 (1%)
1	B	1022/1045 (97%)	0.22	28 (2%)	54	53	16, 26, 43, 55	19 (1%)
1	C	1022/1045 (97%)	0.68	117 (11%)	5	4	20, 33, 46, 56	19 (1%)
1	D	1022/1045 (97%)	0.27	45 (4%)	34	33	18, 28, 45, 54	19 (1%)
1	E	1022/1045 (97%)	0.24	33 (3%)	47	46	17, 28, 45, 55	19 (1%)
1	F	1022/1045 (97%)	0.63	106 (10%)	6	5	21, 32, 46, 55	19 (1%)
All	All	6132/6270 (97%)	0.38	360 (5%)	22	21	14, 30, 45, 56	114 (1%)

The worst 5 of 360 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	219	SER	9.0
1	F	219	SER	8.9
1	A	841	GLY	8.7
1	F	220	GLY	8.1
1	D	219	SER	7.9

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 monosaccharides 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.