Ricin RTA-chain mutagenesis to prevent the onset of the Vascular Leak Syndrome

Riccardo Carangelo

Theoretical Introduction to the Protein

Ricin is an rRNA N-glycosidase (EC number 3.2.2.22, UniProt ID P02879). In particular, ricin is a protein with a very high cytotoxicity produced in the seeds of Ricinus communis. It has a molecular weight of about 62 kDa and consists of two chains connected by a peptide linker (Endo & Tsurugi, 1988).

The B chain (also known as RTB chain, Ricin Toxin B chain) is able of allowing the entry into the cell of the A chain (also known as RTA chain, Ricin Toxin A chain), by binding to the complex carbohydrates present on the cell surface that contain terminal N-acetylglucosamine or galactose residues (Magnusson et al., 1993).

The RTA chain possesses cytotoxic activity, as it is able of inactivating type II ribosome, acting as an N-glycosidase that removes a specific adenine residue contained within the 28S rRNA. This adenine in mammals is positioned at 4324 and is contained within a highly conserved sequence of 12 nucleotides (5'-AGUACGAGAGGA-3') that are universally present in eukaryotic ribosomes and that characterize the Sarcin – Ricin Loop (SRL). This sequence constitutes an important binding site for elongation factors involved in protein synthesis (Endo & Tsurugi, 1988, Sperti *et al.*, 1973). The RTA chain is capable of completely inactivating the ribosome, causing toxicity due to the complete inhibition of protein synthesis. The very high toxicity of ricin is due to the fact that the RTA chain can inactivate about 1500 ribosomes per minute within the cell, which is faster than the cell can regenerate them. In doing so, a single RTA chain is capable of rapidly killing an entire animal cell (Olsnes *et al.*, 1975).

Despite the dangerousness of this molecule, studies are underway for its potential use in anticancer treatments (Lord *et al.*, 2003, Akbari *et al.*, 2017). A deglycosylated version of the RTA chain could be linked to monoclonal antibodies to produce an immunotoxin that can intercept cancer cells with great precision (Akbari *et al.*, 2017). Although there are still no advanced studies on its effectiveness as a treatment method, from such observations, it is deduced that ricin is a protein of great medical and pharmaceutical interest.

Study of the Intrinsic Characteristics of Ricin

Structural and Sequence Observations

Before investigating the biotechnological issue of engineered ricin as an immunotoxin, it is necessary to focus on the basic characteristics of this protein.
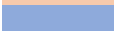
Below it is possible to examine the amino acid sequence in FASTA format of the protein obtained from the UniProt (Universal Protein) database (https://www.uniprot.org/) (576 amino acids):

MKPGGNTIVIWMYAVATWLCFGSTSGWSFTLEDNNIFPKQYPIINFTTAGATVQSYTNFIRA
VRGRLTTGADVRHEIPVLPNRVGLPINQRFILVELSNHAELSVTLALDVTNAYVVGYRAGNS
AYFFHPDNQEDAEAITHLFTDVQNRYTFAFGGNYDRLEQLAGNLRENIELGNGPLEEAISA
LYYYSTGGTQLPTLARSFIICIQMISEAARFQYIEGEMRTRIRYNRRSAPDPSVITLENSWGRLS
TAIQESNQGAFASPIQLQRRNGSKFSVYDVSILIPIIALMVYRCAPPPSSQFSLLIRPVVPNFNAD
VCMDPEPIVRIVGRNGLCVDVRDGRFHNGNAIQLWPCKSNTDANQLWTLKRDNTIRSNGK
CLTTYGYSPGVYVMIYDCNTAATDATRWQIWDNGTIINPRSSLVLAATSGNSGTTLTVQTNI
YAVSQGWLPTNNTQPFVTTIVGLYGLCLQANSGQVWIEDCSSEKAEQQWALYADGSIRPQ
QNRDNCLTSDSNIRETVVKILSCGPASSGQRWMFKNDGTILNLYSGLVLDVRASDPSLKQIIL
YPLHGDPNQIWLPLF-

Below it is possible to see the encoding sequence of the protein obtained from the ENA (European Nucleotide Archive) database (https://www.ebi.ac.uk/ena) (1731 nucleotides):

ATGAAACCGGGGAGGAAATACTATTGTAATATGGATGTATGCAGTGGCAACATGGCTTT
GTTTTGGATCCACCTCAGGGTGGTCTTTCACATTAGAGGATAACAACATATTCCCCAAAC
AATACCCAATTATAAACTTTACCACAGCGGGTGCCACTGTGCAAAGCTACACAAACTTTA
TCAGAGCTGTTCGCGGTCGTTTAACAACTGGAGCTGATGTGAGACATGAAATACCAGTG
TTGCCAAACAGAGTTGGTTTGCCTATAAACCAACGGTTTATTTTAGTTGAACTCTCAAAT
CATGCAGAGCTTTCTGTTACATTAGCGCTGGATGTCACCAATGCATATGTGGTCGGCTA
CCGTGCTGGAAATAGCGCATATTTCTTTCATCCTGACAATCAGGAAGATGCAGAAGCAA
TCACTCATCTTTTCACTGATGTTCAAAATCGATATACATTCGCCTTTGGAGGTAATTATG
ATAGACTTGAACAACTTGCTGGTAATCTGAGAGAAAATATCGAGTTGGGAAATGGTCCA
CTAGAGGAGGCTATCTCAGCGCTTTATTATTACAGTACTGGTGGCACTCAGCTTCCAACT
CTGGCTCGTTCCTTTATAATTTGCATCCAAATGATTTCAGAAGCAGCAAGATTCCAATAT
ATTGAGGGAGAAATGCGCACGAGAATTAGGTACAACCGGAGATCTGCACCAGATCCTA
GCGTAATTACACTTGAGAATAGTTGGGGGGAGACTTTCCACTGCAATTCAAGAGTCTAAC
CAAGGAGCCTTTGCTAGTCCAATTCAACTGCAAAGACGTAATGGTTCCAAATTCAGTGT
GTACGATGTGAGTATATTAATCCCTATCATAGCTCTCATGGTGTATAGATGCGCACCTCC
ACCATCGTCACAGTTTTCTTTGCTTATAAGGCCAGTGGTACCAAATTTTAATGCTGATGT
TTGTATGGATCCTGAGCCCATAGTGCGTATCGTAGGTCGAAATGGTCTATGTGTTGATG
TTAGGGATGGAAGATTCCACAACGGAAACGCAATACAGTTGTGGCCATGCAAGTCTAAT
ACAGATGCAAATCAGCTCTGGACTTTGAAAAGAGACAATACTATTCGATCTAATGGAAA
GTGTTTAACTACTTACGGGTACAGTCCGGGAGTCTATGTGATGATCTATGATTGCAATA
CTGCTGCAACTGATGCCACCCGCTGGCAAATATGGGATAATGGAACCATCATAAATCCC
AGATCTAGTCTAGTTTTAGCAGCGACATCAGGGAACAGTGGTACCACACTTACAGTGCA
AACCAACATTTATGCCGTTAGTCAAGGTTGGCTTCCTACTAATAATACACAACCTTTTGT
GACAACCATTGTTGGGCTATATGGTCTGTGCTTGCAAGCAAATAGTGGACAAGTATGG
ATAGAGGACTGTAGCAGTGAAAAGGCTGAACAACAGTGGGCTCTTTATGCAGATGGTT
CAATACGTCCTCAGCAAAACCGAGATAATTGCCTTACAAGTGATTCTAATATACGGGAAA
CAGTTGTCAAGATCCTCTCTTGTGGCCCTGCATCCTCTGGCCAACGATGGATGTTCAAG
AATGATGGAACCATTTTAAATTTGTATAGTGGGTTGGTGTTAGATGTGAGGGCATCGG
ATCCGAGCCTTAAACAAATCATTCTTTACCCTCTCCATGGTGACCCAAACCAAATATGGTT
ACCATTATTTTGA

Legend of the portions highlighted in the two sequences:

- signal peptide(1 – 35 positions)
- RTA chain (36 – 302 positions)
- linker peptide (303 – 314 positions)
- RTB chain (315 – 576 positions)
- stop codon (opal)

Using the translation tool of ExPASy (https://web.expasy.org/translate/), it is possible to translate the sequence found on the ENA database. Using the online tool BLAST (Basic Local Alignment Search Tool) (https://blast.ncbi.nlm.nih.gov/Blast.cgi), the sequence obtained from the translation tool was subsequently aligned with the sequence found on UniProt, verifying a 100% identity and confirming that the nucleotide sequence accurately encodes the sequence provided by UniProt.

Medical research on ricin mostly focuses on the RTA chain, as it is this part that exhibits cytotoxic activity. Therefore, from now on, the focus will be on it, explaining its operating characteristics in the next chapter. All amino acid numbering refers only to the indexing of the RTA chain.

Mechanism of Action of the RTA Chain's Active Site

As proposed by Weston *et al.* in 1994, the mechanism involves the following steps (the active site is highlighted in *Figure 1*):

(1) The Sarcin – Ricin Loop binds to the active site through the interaction of adenine 4324 with residues Tyr80 and Tyr123;

(2) The residue Arg180 is positioned to protonate N-3 of the adenine and break the N-9 bond of the adenine ring and the C-1' of the ribose;

(3) The residue Glu177 meanwhile stabilizes the ionic form of the ribose generated by the cleavage of the bond;

(4) At this point, a water molecule is included and deprotonated by N-3, previously protonated by Arg180;

(5) A depurination of nucleotide 4324 occurs, resulting in a ribose on an intact nucleotide sequence.
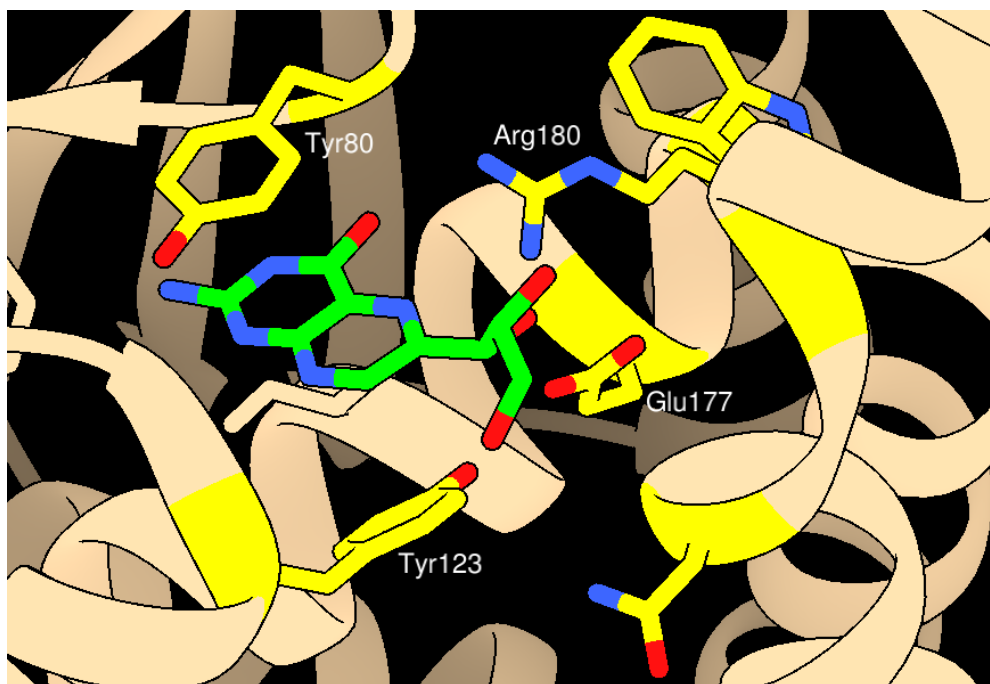
Figure 1 – Detail of the three-dimensional model of the RTA chain of ricin viewed through UCSF Chimera (reference file obtained from the PDB database, Protein Data Bank, https://www.rcsb.org/, protein structure ID 1BR5), highlighting the active site; in yellow, the amino acids directly involved in the catalytic activity are observable (specifically, the amino acids described in the catalytic process proposed by Weston *et al.* in 1994 have been labeled; the neopterin substrate present in the protein crystal has been colored in green

Engineering of the Ricin RTA Chain

Despite the research interest in the medical field regarding the RTA chain, various side effects have been experimented in animal model studies, some of which are in the process of being resolved. A significant side effect under study in recent years is the onset of Vascular Leak Syndrome (VLS) in studies conducted on mice (Smallshaw *et al.*, 2003). This document presents and discusses a proposal for engineering through rational design performed on the RTA chain of ricin to overcome the issue induced by such a side effect.

A study in 2000 by Baluna *et al.* highlights how a triad of consecutive amino acids not part of the active site (L74, D75, and V76), called the LDV region, is responsible for the onset of VLS. A subsequent mutagenesis study conducted by Smallshaw *et al.* in 2003 experimented with mutations on the individual amino acids of this triad and on neighboring amino acids to generate a protein incapable of causing VLS phenomena. Specifically, the mutations of the 2003 study by Smallshaw *et al.* concerning the LDV triad are as follows (their effects on protein activity are also briefly reported):

Amino acid	Substitution	Effect on the enzyme activity	Effect on the VLS
L74	A	No effect	No effect
L74	M	No effect	No effect
D75	A	Suppression	NA
D75	E	Suppression	NA
D75	N	Suppression	NA
V76	A	No effect	No effect
V76	M	No effect	No effect

In the same paper, it is also possible to read about the negative results regarding the mutation of two amino acids both close too the LDV sequence, with the following effects:

| Amino acid	Substitution	Effect on the enzyme activity	Effect on the VLS
R48	A	No effect	Suppression
N76	A	No effect	Suppression

In the paper, any mutation made on D75 is immediately excluded, as it is capable of completely inhibiting the catalytic activity of the RTA chain, making it unusable. The amino acids L74 and V76 of the triad, on the other hand, did not have any effects on the catalytic activity, nor did they show inhibitory outcomes in the onset of VLS. The mutations of greater interest turned out to be those on R48 and N76, which, in addition to not having effects on the catalytic activity of the RTA chain, did not cause the appearance of VLS. The paper selects the two mutants R48A and N76A, conducting a study on therapeutic efficacy, but without any mention of the intrinsic reasons that led to the protein's improvement.

To go deeper into the matter, the two mutants R48A and N76A were modeled using the online modeling tool Phyre² (Protein Homology/AnalogY Recognition Engine V2.0, http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index), which allows for the modeling of a given amino acid sequence through a fold recognition algorithm. The sequence submitted to the tool corresponds to the FASTA sequence obtained from the UniProt database, to which the mutations R48A and N76A were manually added, one for each model. Both mutant sequences were modeled with Confidence values of 100% and Coverage of 99%.

The two derived models were compared with the wildtype model of the ricin A chain, using the MatchMaker function of UCSF Chimera. Both models provide a Root Mean Square Deviation (RMSD) value of 1.566, indicating generally a good level of structural overlap (*Figures 2* and *4*).

In the N97A mutant, the substitution of an asparagine (a polar amino acid with a relatively bulky side chain) with an alanine, an amino acid lacking charge and polarity and with a less bulky side chain, is

observed (*Figure 3*). A similar substitution is observed in the R48A mutant, where an arginine (a positively charged amino acid with a relatively bulky side chain) was replaced with an alanine (*Figure 5*).
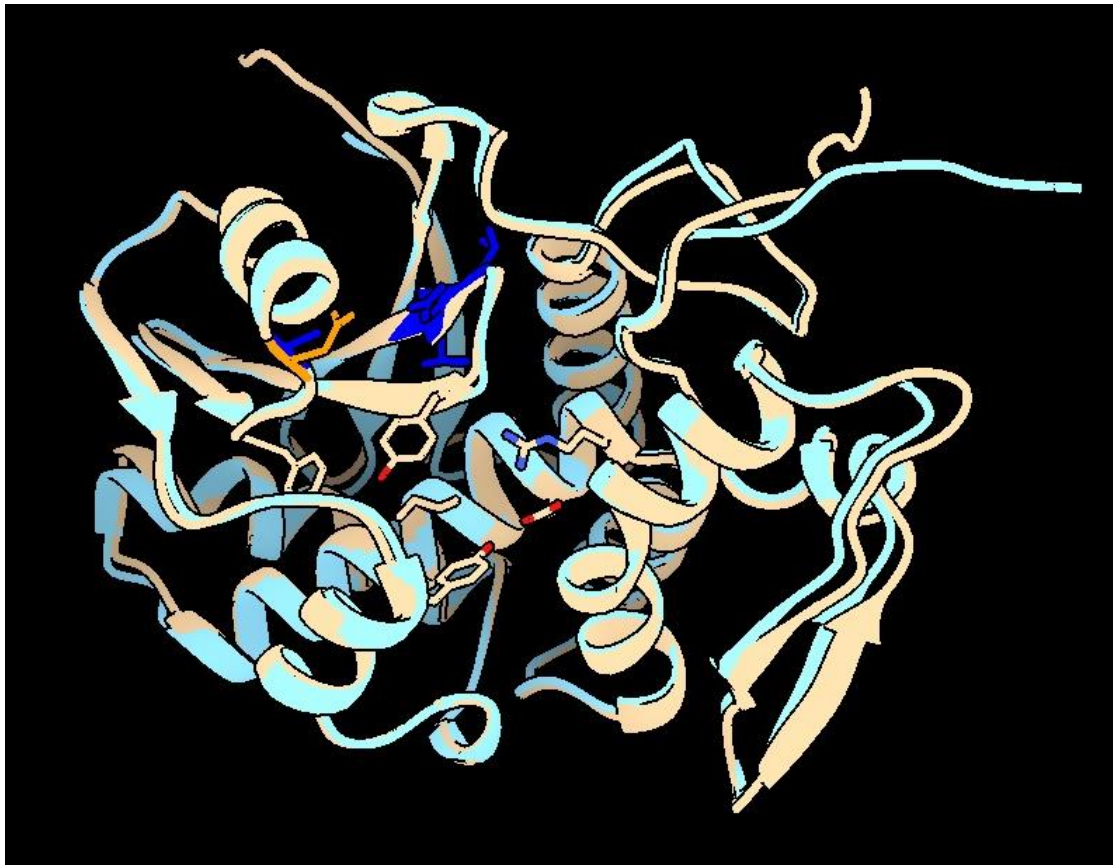
Figure 2 – Structural overlap obtained with UCSF Chimera performed between the wildtype RTA chain (in light orange) and the N97A mutant modeled with Phyre² (in light blue); high structural similarity is observed
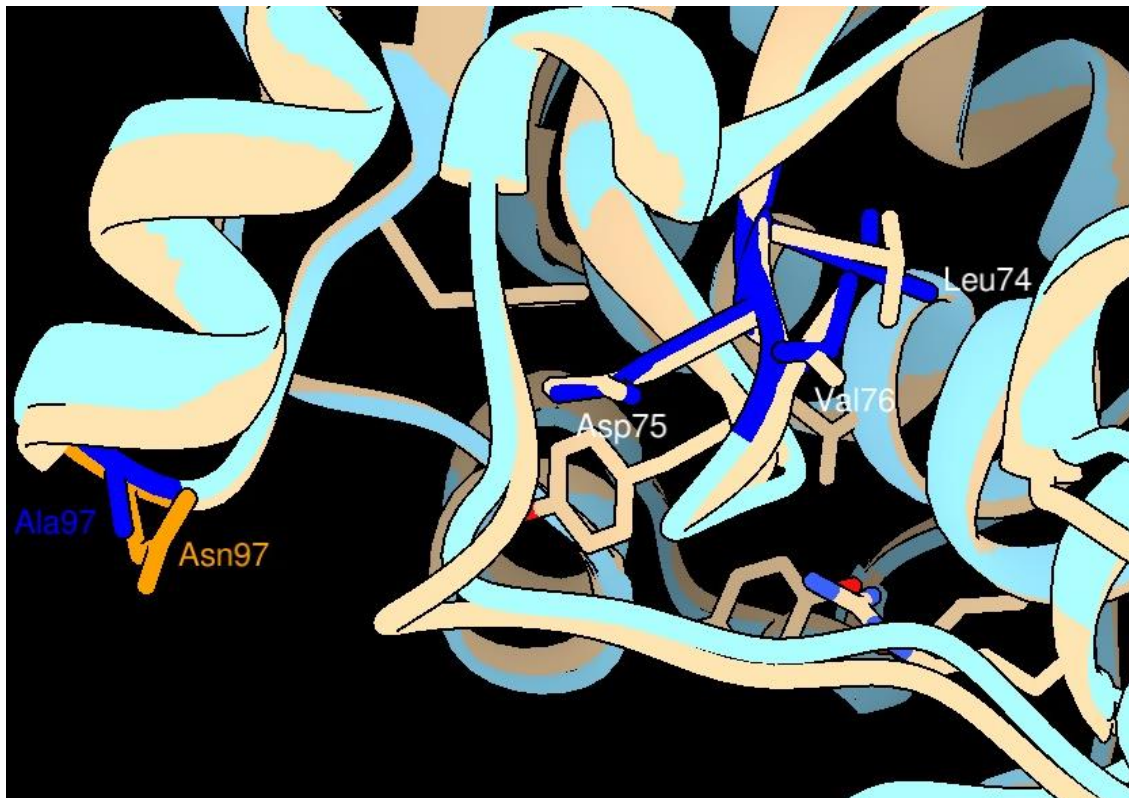
Figure 3 – Detail of the structural overlap shown in Figure 2; on the right (in blue), the three amino acids of the LDV triad are observed, while on the left, the single mutation that replaced Asn97 (in orange) in the wild type RTA with an Ala (in blue) to generate the N97A mutant can be seen

Figure 4 – Structural overlap obtained with UCSF Chimera performed between the wildtype RTA chain (in light orange) and the R48A mutant modeled with Phyre² (in light blue); in this case, too, high structural similarity is observed
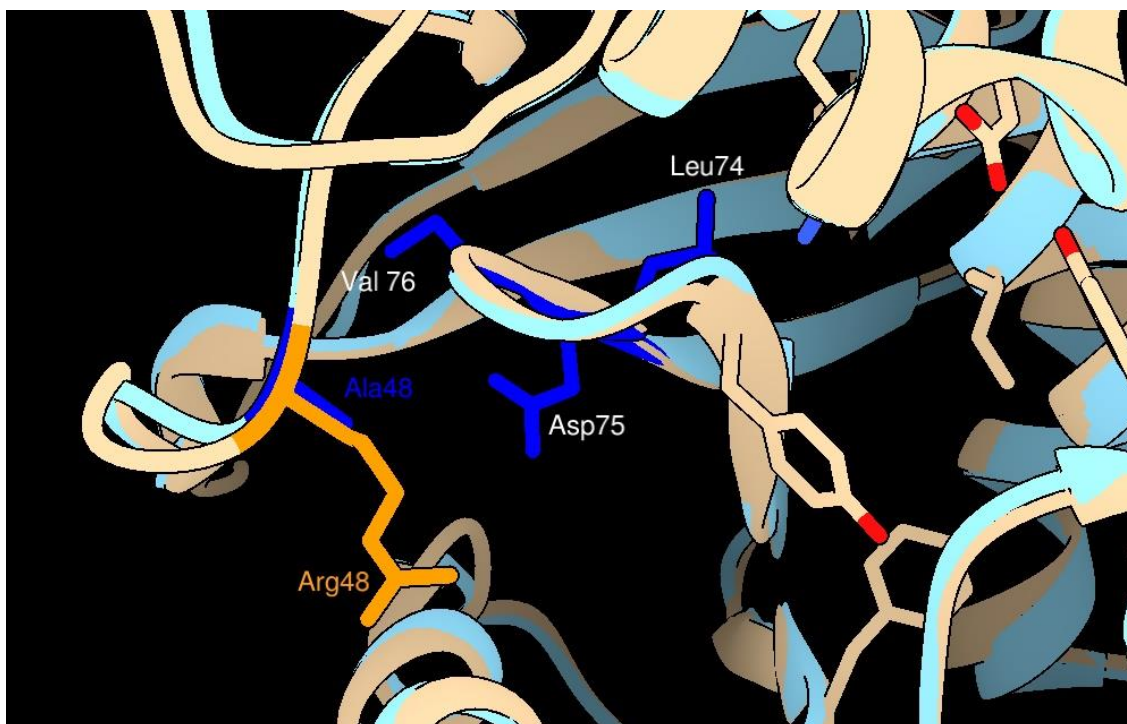
Figure 5 – Detail of the structural overlap shown in Figure 4; on the right (in blue), the three amino acids of the LDV triad are observed, while on the left, the single mutation that replaced Arg48 (in orange) in the wild type RTA with an Ala (in blue) to generate the R48A mutant can be seen

In both substitutions, there is a significant reduction in the size of the side chain, as well as a total loss of charge (in the case of R48A) or polarity (in the case of N97A).

From what can be gathered from experimental evidence, the mutants do not cause VLS, however, the reason is not hypothesized in the reference article by Smallshaw *et al.* (2003). Two hypotheses can, however, be formulated regarding this improvement:

(1) the mutation eliminated some direct interaction between the LDV triad and the mutated amino acids, inhibiting the effect of the triad in some way;
(2) the mutation caused a variation in folding such as to destabilize the activity of the triad.

Regarding hypothesis (1), in the comparison images shown above, there is a distance between the amino acids of the triad and the substituted amino acids of just under 10 Å, which is enough for the establishment of significant direct interactions between the LDV triad and the substituted amino acids. At this point, it is necessary to report the fact that the positions of the mutations were purposely chosen to be far from the catalytic site but proximal to the LDV region. Similarly, hypothesis (2) can be considered, stating that variations in the side chains introduced by the substitutions may have altered the pattern of interactions within the protein to the point of changing the conformation of the native state, destabilizing the activity of the LDV triad, without excessively destabilizing the catalytic site. From these observations, it is possible to deduce that both hypotheses could be true. Moreover, it is important to notice that one hypothesis does not necessarily exclude the other.

For a more significant verification of what the two hypotheses claim, substitutions that are different from alanine in the two mutants N97A and R48A could be considered. In particular, amino acids that maintain charge neutrality but have side chains of different lengths could be substituted, in order to introduce different bulks and generate a series of alternative mutants. By doing so, it would be possible to study the effect of uncharged side chains of various lengths on the LDV triad and the catalytic site.

To put this into practice, in order to generate some exemplary alternative mutants, three neutral amino acids with side chains of increasing bulk (valine, leucine, and phenylalanine) were inserted into the two original mutants R48A and N97A, replacing alanine. The substitution was carried out on the models of the mutants generated by Phyre[2], using the editing tool of UCSF Chimera, which allows the substitution of individual amino acids, calculating the best orientations through the use of the Dunbrack rotamer library. This library is capable of providing a series of potential orientations of the side chains accompanied by their positional likelihood (understood as a probability value in the $[0,1]_{\mathbb{R}}$ interval). Among the various possible orientations, the most probable ones were selected. The six alternative mutants derived from such substitutions are as follows (*Figure 8*):

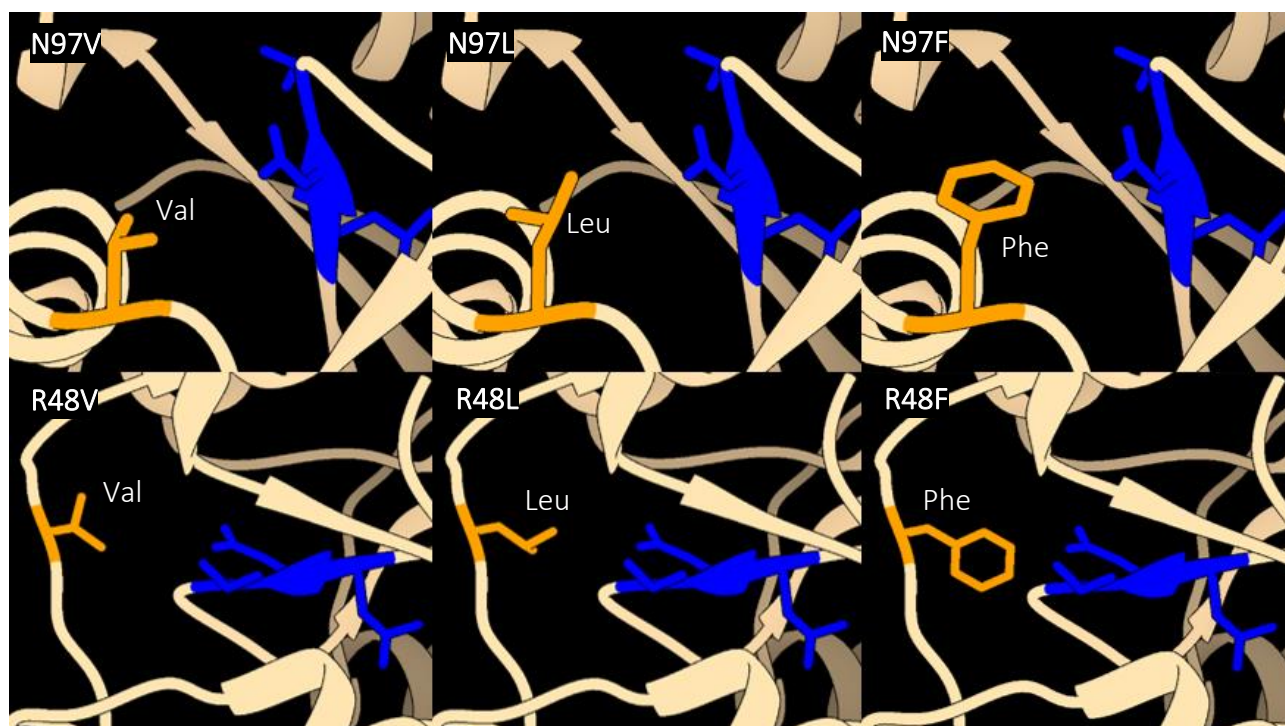| Mutation	Probability of the chosen orientation for the side chain
N97V	0,542
N97L	0,532
N97F	0,782
R48V	0,471
R48L	0,787
R48F	0,217

Figure 6 – The top panel shows three examples of substitutions of three alternative uncharged amino acids to alanine in the N97A mutant. In blue, the LDV triad is represented in each image, while in orange, the various mutations are represented (in order from left to right): Val, Leu, and Phe; the three images at the bottom show three examples of alternative substitutions in the R48A mutant, in which, instead of Ala, Val, Leu, and Phe have been inserted (in order from left to right)

From the alternative mutants visible in *Figure 8*, it would be possible to perform *in silico* molecular dynamics experiments, in order to assess the effects of the bulk of neutral side chains of different sizes on the catalytic activity (docking with the substrate) and on the activity of the LDV triad (docking with the targets responsible for the onset of VLS). Although the ideal for a molecular dynamics analysis would be an experimental model with the highest possible resolution, the high number of potential mutants to be used would require impractical times. In such cases, modeling through fold recognition comes into play. Therefore, before proceeding with a molecular dynamics project based on *in silico* generated models, it may be appropriate to have at least a general idea of the actual quality of such models. For this purpose, it is possible to use online tools created to evaluate the quality of protein models. In this case, two famous tools are used: ANOLEA (Atomic Non-Local Environment Assessment) (http://melolab.org/anolea/), useful for estimating the stability of individual amino acids, and QMEAN by Swiss-Model (https://swissmodel.expasy.org/qmean/), which provides scores and estimates on the general quality of the model. The analysis with these tools is carried out, as an example, on the original mutants R48A and N97A.

From the analysis on ANOLEA, it is possible to obtain a log file from which it is possible to extrapolate a graph that shows the energy values (in kT, i.e., the product of temperature by the Boltzmann constant) of the various amino acids (*Plots 1* and *2*).
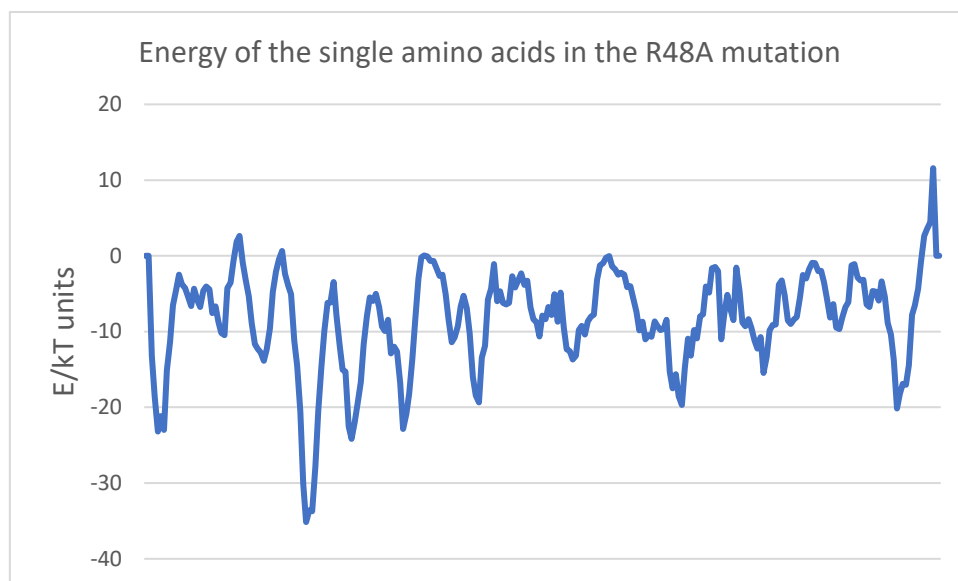

Plot 1 – Plot illustrating the energy values for the model of the R48A mutant generated with Phyre[2] (the total non-local energy of the protein (E/kT units) is equal to -2146, while the non-local normalized energy Z-score is equal to -0.14); numerical data obtained with ANOLEA
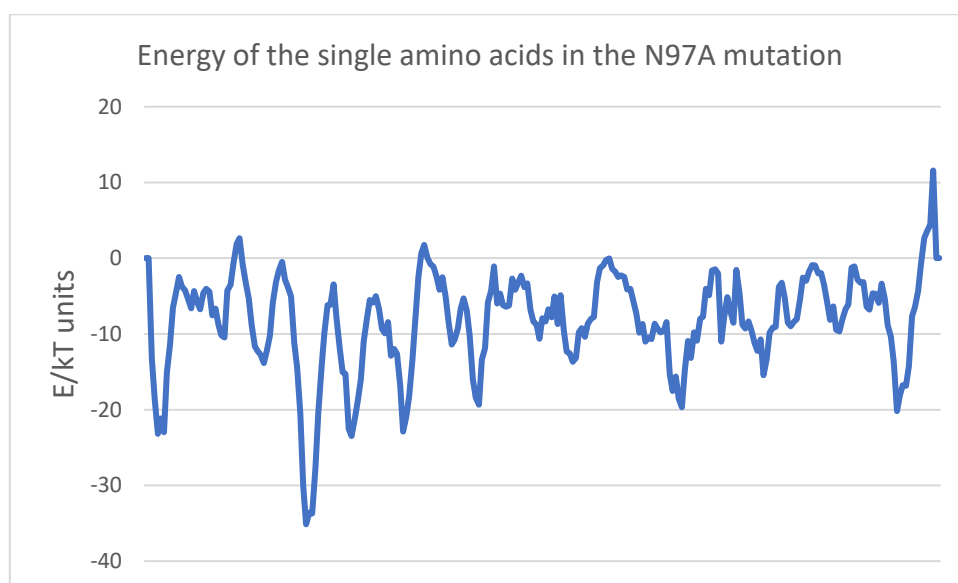

Plot 2 – Plot illustrating the energy values for the model of the N97A mutant generated with Phyre[2] (the total non-local energy of the protein (E/kT units) is equal to -2147, while the non-local normalized energy Z-score is equal to -0.11); numerical data obtained with ANOLEA

Generally, the energy values obtained with the ANOLEA analysis predict good models in terms of energetic stability.

A subsequent analysis conducted through QMEAN, allowed for the assessment of the absolute and local quality of the models through various estimates. The results of the analysis revealed the following data (*Plots 3* and *4* and *Figures 6* and *7*).
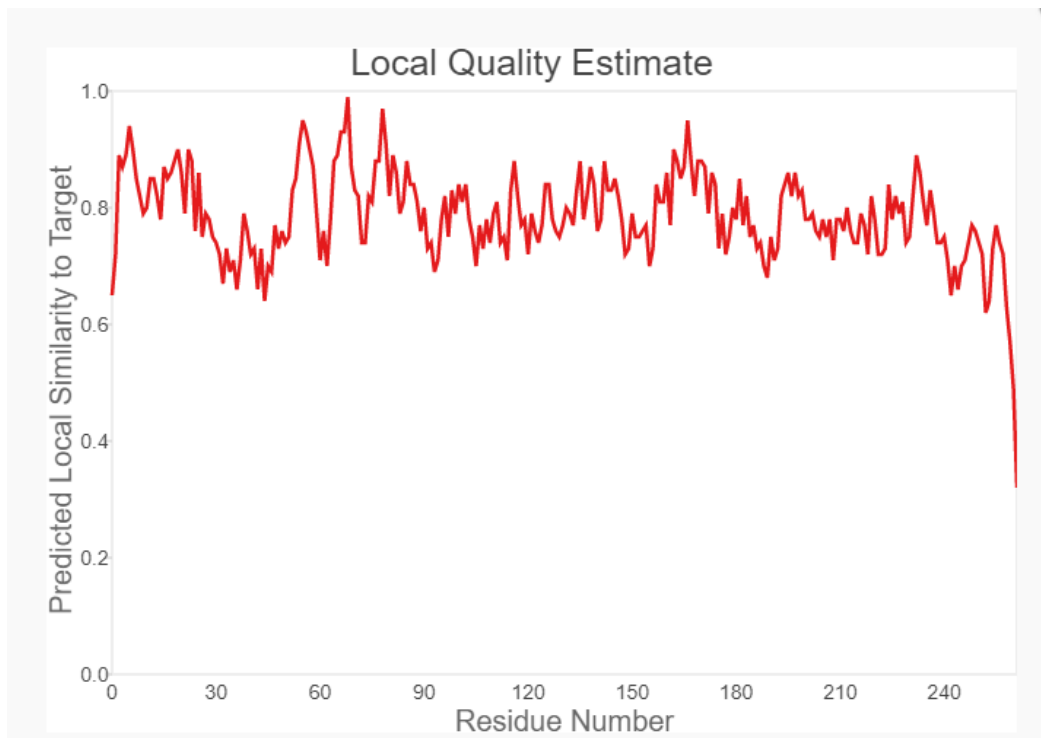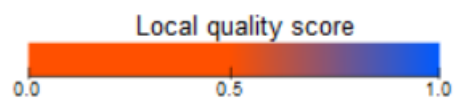
Plot 3 – Plot estimating the quality for each single amino acid in the R48A mutant. The estimates are close to values of 1.0, indicating a good predicted quality for the model
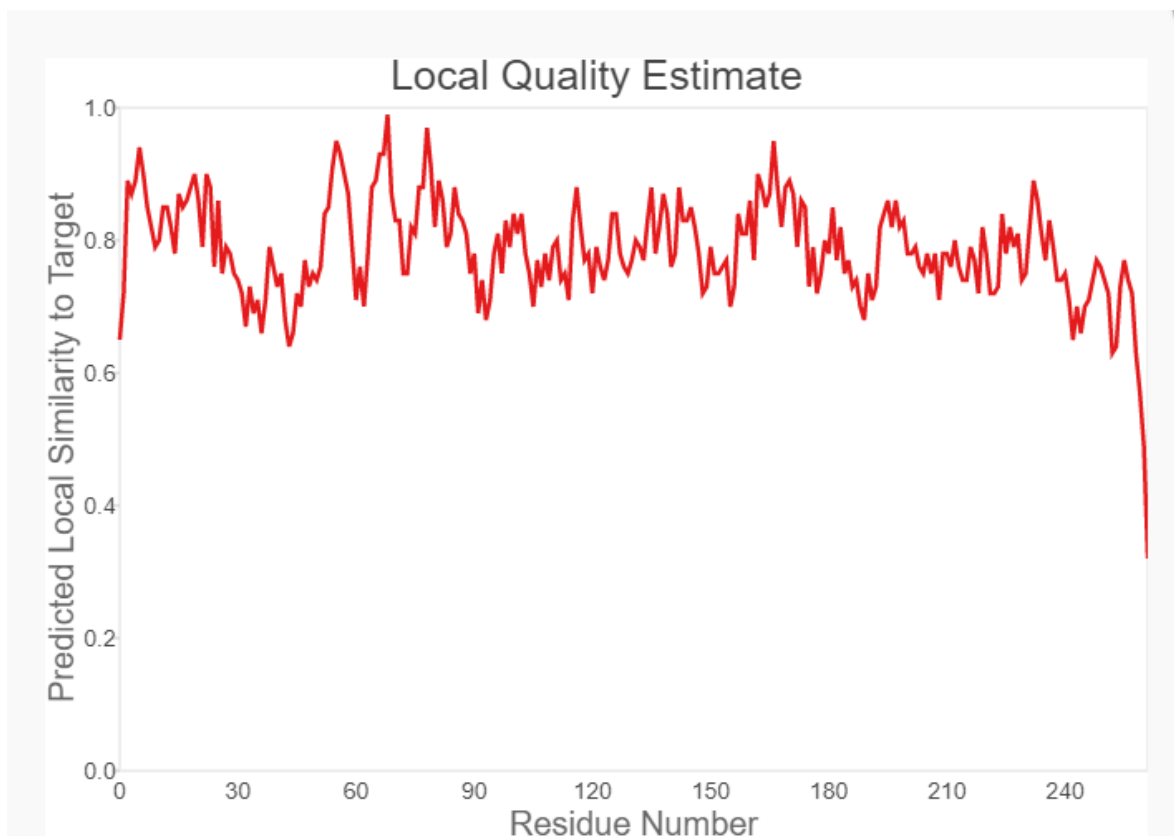
Plot 4 – Plot estimating the quality for each single amino acid in the N97A mutant. In this case, as well, the estimates are close to values of 1.0, indicating good predicted quality for the model.

QMEAN4 Value: -0.30 ⊘

❷ Sequence colored by local quality:

Local quality score

0.0   0.5   1.0

A:	QYPIINFTTAGATVQSYTNFIRAVRGRLTTGADVRHEIPVLPNAVGLPIN	50
A:	QRFILVELSNHAELSVTLALDVTNAYVVGYRAGNSAYFFHPDNQEDAEAI	100
A:	THLFTDVQNRYTFAFGGNYDRLEQLAGNLRENIELGNGPLEEAISALYYY	150
A:	STGGTQLPTLARSFIICIQMISEAARFQYIEGEMRTRIRYNRRSAPDPSV	200
A:	ITLENSWGRLSTAIQESNQGAFASPIQLQRRNGSKFSVYDVSILIPIIAL	250
A:	MVYRCAPPPSSQF	263

Figure 7 – Local analysis of the R48A mutant; it is possible to notice high stability values in this case

QMEAN4 Value: -0.30 ⊙

⊙ Sequence colored by local quality:

Local quality score

0.0   0.5   1.0

A: QYPIINFTTAGATVQSYTNFIRAVRGRLTTGADVRHEIPVLPNRVGLPIN   50
A: QRFILVELSNHAELSVTLALDVTNAYVVGYRAGNSAYFFHPDAQEDAEAI   100
A: THLFTDVQNRYTFAFGGNYDRLEQLAGNLRENIELGNGPLEEAISALYYY   150
A: STGGTQLPTLARSFIICIQMISEAARFQYIEGEMRTRIRYNRRSAPDPSV   200
A: ITLENSWGRLSTAIQESNQGAFASPIQLQRRNGSKFSVYDVSILIPIIAL   250
A: MVYRCAPPPSSQF   263

Figure 8 - Local analysis of the N97A mutant; even in this case, it is possible to notice high stability values

These evaluative results provide good quantitative estimates on the actual quality of the models generated for fold recognition, allowing greater confidence in their use for potential molecular dynamics studies. Similarly, the estimates can be repeated for the alternative mutants shown in *Figure 8*. The next step before initiating a molecular dynamics analysis could be optimization. Although the two models are of good stereochemical quality and possess a certain stability, it might be useful to optimize them by correcting the less stable amino acid positions and those with higher energy. Useful softwares for this purpose could be both SCRWL4 (http://dunbrack.fccc.edu/scwrl4/index.php) or ModRefiner (https://zhanglab.ccmb.med.umich.edu/ModRefiner/), however, they will not be used here, as it is not actually possible to access molecular dynamics tools in this context.

Expression of the Studied Mutants in the W3110 strain of *Escherichia coli*

After completing the molecular dynamics experimental phases, it is possible to select the most promising mutants from the various ones tested, in order to start the experimental phase on a limited number of mutants, thus reducing costs and time of experiments.

Potential mutants selected for expression in *E. coli* can be used for subsequent testing phases. In particular, in this case, an ideal vector for optimized expression in *E. coli* of the recombinant RTA chain could be the pKK223-3 plasmid (Smallshaw *et al.*, 2003) (*Figure 9*). It contains indeed the strong inducible *tac* promoter and ensures resistance to ampicillin up to $100\ \mu g/ml$. Through the AddGene database (https://www.addgene.org/), it is possible to find the plasmid card with all the information necessary for the expression experiment. The protein-coding sequence obtained from the ENA database (after adding the mutations) needs to be optimized for expression in *E. coli* using a software like JCat (http://www.jcat.de/), which is capable of replacing the codons of a nucleotide sequence to conform to the codon usage of a specific selected organism.

According to the plasmid vector pKK223-3 card available on the EMBL (European Molecular Biology Laboratory) database (https://www.embl.de/), the plasmid is to be used within LacIq strains, for this reason, in the optimization phase using JCat, the W3110 strain of *E. coli* is selected. The coding sequence of the entire wild typericin optimized by JCat is as follows:

ATGAAACCGGGTGGTAACACCATCGTTATCTGGATGTACGCTGTTGCTACCTGGCTGTG
CTTCGGTTCTACCTCTGGCTGGAGCTTCACCCTGGAAGACAACAACATCTTCCCGAAACA
GTACCCGATCATCAACTTCACCACCGCTGGTGCTACCGTTCAGTCTTACACCAACTTCATC
CGTGCTGTTCGTGGTCGTCTGACCACCGGTGCTGACGTTCGTCACGAAATCCCGGTTCT
GCCGAACCGTGTTGGTCTGCCGATCAACCAGCGTTTCATCCTGGTTGAACTGTCTAACC
ACGCTGAACTGTCTGTTACCCTGGCTCTGGACGTTACCAACGCTTACGTTGTTGGTTACC
GTGCTGGTAACTCTGCTTACTTCTTCCACCCGGACAACCAGGAAGACGCTGAAGCTATC
ACCCACCTGTTCACCGACGTTCAGAACCGTTACACCTTCGCTTTCGGTGGTAACTACGAC
CGTCTGGAACAGCTGGCTGGTAACCTGCGTGAAAACATCGAACTGGGTAACGGTCCGC
TGGAAGAAGCTATCTCTGCTCTGTACTACTACTCTACCGGTGGTACCCAGCTGCCGACCC
TGGCGCGTAGCTTCATCATCTGCATACAGATGATCTCTGAAGCTGCTCGTTTCCAGTACA
TCGAAGGTGAAATGCGTACCCGTATCCGTTACAACCGTCGTTCTGCTCCGGACCCGTCT
GTTATCACCCTGGAAAACTCTTGGGGTCGTCTGTCTACCGCTATCCAGGAATCTAACCAG
GGTGCTTTCGCTTCTCCGATCCAGCTGCAGCGTCGTAACGGTTCTAAATTCTCTGTTTAC
GACGTTTCTATCCTGATCCCGATCATAGCGCTGATGGTATACAGATGCGCTCCGCCGCC
GTCTTCTCAGTTCTCTCTGCTGATCCGTCCGGTTGTTCCGAACTTCAACGCTGACGTTTG
CATGGACCCGGAACCGATCGTTCGTATCGTTGGTCGTAACGGTCTGTGCGTTGACGTTC
GTGACGGCAGGTTCCACAACGGTAATGCTATCCAGCTGTGGCCGTGCAAATCTAACACC
GACGCTAACCAGCTGTGGACCCTGAAACGTGACAACACCATCCGTTCTAACGGTAAATG
CCTGACCACCTACGGTTACTCTCCGGGTGTTTACGTTATGATCTACGACTGCAACACCGC
TGCTACCGACGCTACCCGTTGGCAGATCTGGGACAACGGTACCATCATCAACCCGCGTT
CTTCTCTGGTTCTGGCTGCTACCTCTGGTAACTCTGGTACCACCCTGACCGTTCAGACCA
ACATCTACGCTGTTTCTCAGGGTTGGCTGCCGACCAACAACACCCAGCCGTTCGTTACCA
CCATCGTTGGTCTGTACGGTCTGTGCCTGCAGGCTAACTCTGGTCAGGTTTGGATCGAA
GACTGCTCTTCTGAAAAAGCTGAACAGCAGTGGGCTCTGTACGCTGACGGTTCTATCCG
TCCGCAGCAGAACCGTGACAACTGCCTGACCTCTGACTCTAACATCCGTGAAACCGTTGT
TAAAATCCTGTCTTGCGGTCCGGCTTCTTCTGGTCAGCGTTGGATGTTCAAAAACGACG
GTACCATCCTGAACCTGTACTCTGGTCTGGTTCTGGACGTTCGTGCTTCTGACCCGTCTC
TGAAACAGATCATCCTGTACCCGCTGCACGGTGACCCGAACCAGATCTGGCTGCCGCTG
TTCTAA

Legenda

	- signal peptide
	- RTA chain
	- linker peptide
	- RTB chain
	- stop codon (ochre)

JCat returns a Codon Adaptation Index (CAI) value of **0.1573** for the original sequence, while it returns a CAI value of **0.9499** for the optimized sequence. A quick verification using BLAST allows for a local alignment between the original and the optimized sequence. Through the comparison, it is learned that JCat has replaced 454 nucleotides, so the two sequences have a 74% identity. A quick translation check with the ExPASy translation tool and subsequent BLAST alignment of the original amino acid sequence with the one derived from the optimized nucleotide sequence allows for verification that the two amino

acid sequences match 100%. Therefore, the nucleotide sequence has changed, but the amino acid sequence is completely conserved.

From the optimized sequence above, only the region responsible for encoding the RTA chain, including the related signal peptide, is selected, thus obtaining the ORF (i.e., the Open Reading Frame) to be integrated into the pKK223-3 plasmid vector (906 nucleotides):

ATGAAACCGGGTGGTAACACCATCGTTATCTGGATGTACGCTGTTGCTACCTGGCTGTG
CTTCGGTTCTACCTCTGGCTGGAGCTTCACCCTGGAAGACAACAACATCTTCCCGAAACA
GTACCCGATCATCAACTTCACCACCGCTGGTGCTACCGTTCAGTCTTACACCAACTTCATC
CGTGCTGTTCGTGGTCGTCTGACCACCGGTGCTGACGTTCGTCACGAAATCCCGGTTCT
GCCGAAC**MMM**GTTGGTCTGCCGATCAACCAGCGTTTCATCCTGGTTGAACTGTCTAAC
CACGCTGAACTGTCTGTTACCCTGGCTCTGGACGTTACCAACGCTTACGTTGTTGGTTAC
CGTGCTGGTAACTCTGCTTACTTCTTCCACCCGGAC**NNN**CAGGAAGACGCTGAAGCTAT
CACCCACCTGTTCACCGACGTTCAGAACCGTTACACCTTCGCTTTCGGTGGTAACTACGA
CCGTCTGGAACAGCTGGCTGGTAACCTGCGTGAAAACATCGAACTGGGTAACGGTCCG
CTGGAAGAAGCTATCTCTGCTCTGTACTACTACTCTACCGGTGGTACCCAGCTGCCGACC
CTGGCGCGTAGCTTCATCATCTGCATACAGATGATCTCTGAAGCTGCTCGTTTCCAGTAC
ATCGAAGGTGAAATGCGTACCCGTATCCGTTACAACCGTCGTTCTGCTCCGGACCCGTC
TGTTATCACCCTGGAAAACTCTTGGGGTCGTCTGTCTACCGCTATCCAGGAATCTAACCA
GGGTGCTTTCGCTTCTCCGATCCAGCTGCAGCGTCGTAACGGTTCTAAATTCTCTGTTTA
CGACGTTTCTATCCTGATCCCGATCATAGCGCTGATGGTATACAGATGCGCTCCGCCGC
CGTCTTCTCAGTTC

The two highlighted codons above are those to be replaced with the various mutations of interest obtained from the most promising mutants in the molecular dynamics tests. The codon MMM corresponds to the codon of the RTA chains mutated at position 48, while the codon NNN corresponds to the codon of the RTA chains mutated at position 97.
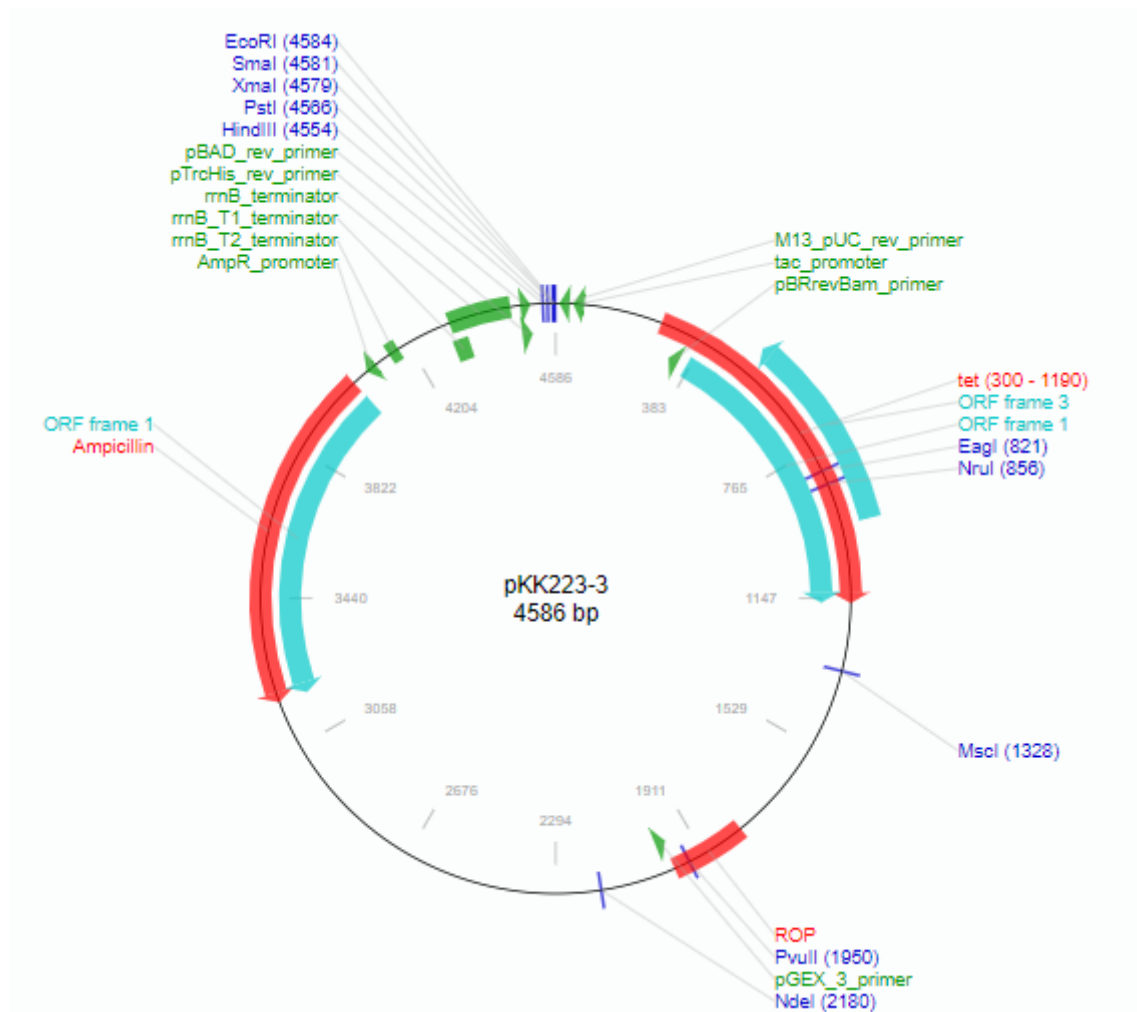
Figure 9 – Circular map of the pKK223-3 vector highlighting sites of various interest

It is necessary to choose an appropriate restriction site in the plasmid to integrate the sequence. Therefore, EcoRI (located at position 4584) is chosen. By using the online tool NEBcutter V2.0 (http://nc2.neb.com/NEBcutter2/), it is possible to verify that the EcoRI restriction site is not present within the sequence to be integrated into the plasmid (EcoRI is indeed listed among the zero cutters, a list that contains all RE sites not present in the submitted sequence). Therefore, it is possible to integrate the sequence using EcoRI as the RE sequence, adding the latter upstream and downstream of the CDS.

Below is the final sequence including the restriction site (highlighted in dark orange in the nucleotide sequence shown below), that will allow the translocation into the plasmid, avoiding problems due to frame shifts and without unintended cuts in the sequence:

<mark>CATATG</mark>ATGAAACCGGGTGGTAACACCATCGTTATCTGGATGTACGCTGTTGCTACCTG
GCTGTGCTTCGGTTCTACCTCTGGCTGGAGCTTCACCCTGGAAGACAACAACATCTTCCC
GAAACAGTACCCGATCATCAACTTCACCACCGCTGGTGCTACCGTTCAGTCTTACACCAA
CTTCATCCGTGCTGTTCGTGGTCGTCTGACCACCGGTGCTGACGTTCGTCACGAAATCCC
GGTTCTGCCGAAC<mark>MMM</mark>GTTGGTCTGCCGATCAACCAGCGTTTCATCCTGGTTGAACTG
TCTAACCACGCTGAACTGTCTGTTACCCTGGCTCTGGACGTTACCAACGCTTACGTTGTT
GGTTACCGTGCTGGTAACTCTGCTTACTTCTTCCACCCGGAC<mark>NNN</mark>CAGGAAGACGCTG
AAGCTATCACCCACCTGTTCACCGACGTTCAGAACCGTTACACCTTCGCTTTCGGTGGTA
ACTACGACCGTCTGGAACAGCTGGCTGGTAACCTGCGTGAAAACATCGAACTGGGTAAC
GGTCCGCTGGAAGAAGCTATCTCTGCTCTGTACTACTACTCTACCGGTGGTACCCAGCT
GCCGACCCTGGCGCGTAGCTTCATCATCTGCATACAGATGATCTCTGAAGCTGCTCGTTT
CCAGTACATCGAAGGTGAAATGCGTACCCGTATCCGTTACAACCGTCGTTCTGCTCCGG
ACCCGTCTGTTATCACCCTGGAAAACTCTTGGGGTCGTCTGTCTACCGCTATCCAGGAAT
CTAACCAGGGTGCTTTCGCTTCTCCGATCCAGCTGCAGCGTCGTAACGGTTCTAAATTCT
CTGTTTACGACGTTTCTATCCTGATCCCGATCATAGCGCTGATGGTATACAGATGCGCTC
CGCCGCCGTCTTCTCAGTTC<mark>GTATAC</mark>

The expression project is intended for practical application in a wet lab, so as to verify the accuracy of the predictions on the mutants and to finally evaluate any actual improvements in catalytic activity with the ultimate goal of preventing VLS occurrence during anticancer treatments.

Reference

Akbari, B., Farajnia, S., Ahdi Khosroshahi, S., Safari, F., Yousefi, M., Dariushnejad, H., & Rahbarnia, L. (2017). Immunotoxins in cancer therapy: Review and update. International reviews of immunology, 36(4), 207-219.

Baluna, R., Coleman, E., Jones, C., Ghetie, V., & Vitetta, E. S. (2000). The effect of a monoclonal antibody coupled to ricin A chain-derived peptides on endothelial cells in vitro: insights into toxin-mediated vascular damage. Experimental cell research, 258(2), 417-424.

Endo, Y., & Tsurugi, K. (1988). The RNA N-glycosidase activity of ricin A-chain. The characteristics of the enzymatic activity of ricin A-chain with ribosomes and with rRNA. Journal of Biological Chemistry, 263(18), 8735-8739.

Lord, M. J., Jolliffe, N. A., Marsden, C. J., Pateman, C. S., Smith, D. C., Spooner, R. A., Watson, P. D. & Roberts, L. M. (2003). Ricin. Toxicological reviews, 22(1), 53-64.

Magnusson, S., Kjeken, R., & Berg, T. (1993). Characterization of two distinct pathways of endocytosis of ricin by rat liver endothelial cells. Experimental cell research, 205(1), 118-125.

Molecular graphics and analyses performed with UCSF Chimera, developed by the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco, with support from NIH P41-GM103311.

Olsnes, S., FERNANDEZ-PUENTES, C., Carrasco, L., & Vazquez, D. (1975). Ribosome Inactivation by the Toxic Lectins Abrin and Ricin: Kinetics of the Enzymic Activity of the Toxin A-Chains. European journal of biochemistry, 60(1), 281-288.

Smallshaw, J. E., Ghetie, V., Rizo, J., Fulmer, J. R., Trahan, L. L., Ghetie, M. A., & Vitetta, E. S. (2003). Genetic engineering of an immunotoxin to eliminate pulmonary vascular leak in mice. Nature biotechnology, 21(4), 387.

Sperti, S., Montanaro, L., Mattioli, A., & Stirpe, F. (1973). Inhibition by ricin of protein synthesis in vitro: 60S ribosomal subunit as the target of the toxin. Biochemical Journal, 136(3), 813-815.

UCSF Chimera--a visualization system for exploratory research and analysis. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. J Comput Chem. 2004 Oct;25(13):1605-12.

Weston, S. A., Tucker, A. D., Thatcher, D. R., Derbyshire, D. J., & Pauptit, R. A. (1994). X-ray structure of recombinant ricin A-chain at 1.8 Å resolution. Journal of molecular biology, 244(4), 410-422.