# Toxicity Estimation of Components of our mRNA-LNP:

## Toxicity Estimation of our Anti-IL8:

We utilise our amino acid sequence to estimate the toxicity of our anti-IL8 protein:

MPLLLLLPLLWAGALAEVQLLESGGGLVQPGGSLRLSCAASGFTFSYYGM

GWVRQAPGKGLEWVSGISYSGSGTYYADSVKGRFTISRDNSKNTLYLQMN

SLRAEDTAVYYCARDYVGNLDYWGQGTLVTVSSGGGGSGGGGSGGGGSDI

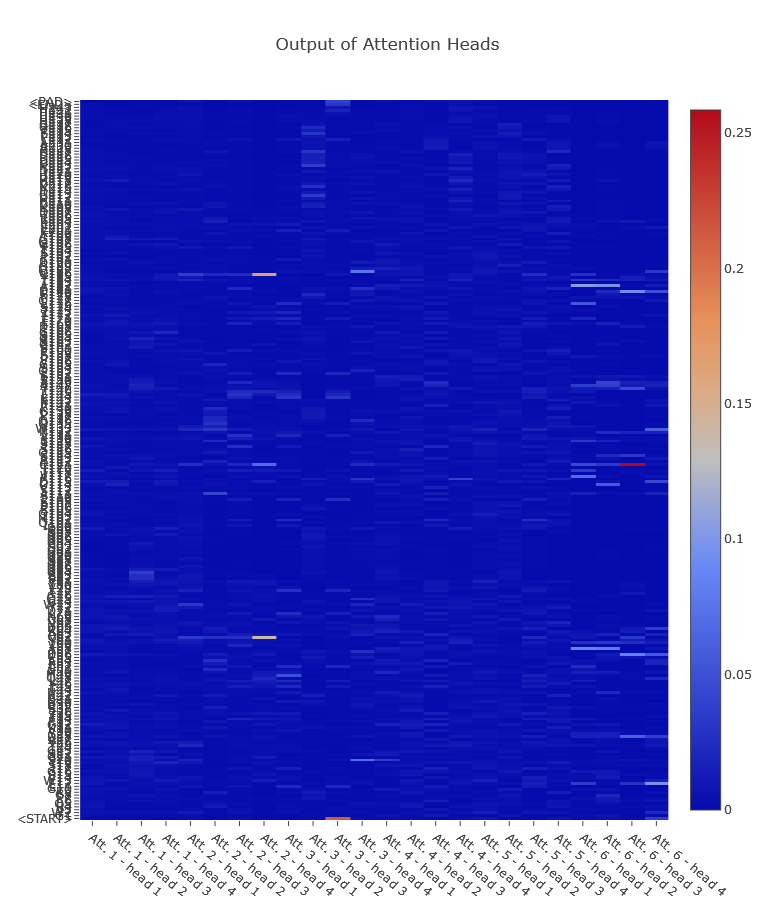
QMTQSPSSLSASVGDRVTITCRASQSISSYLNWYQQKPGKAPKLLIYAAS

SLQSGVPSRFSGSGSGTDFTLTISSLQPEDFATYYCQQSDTPSTFGQGTK

LEIKRTDYKDHDGDYKDHDIDYKDDDDKAAALPETGGHHHHHH

To check for toxicity, we utilize CSM-Toxin, a web server for predicting protein toxicity, developed by Biosig Lab. CSM-Toxin provides a comprehensive suite for rapid identification of toxin proteins. Having uploaded our AA sequence to the server, the following result was obtained:

|  |  |  |
| --- | --- | --- |
| ID | Sequence | Prediction |
| PLLLLLPLLWAGALAEVQLLESGGGLVQPGGSLRLSCAASGFTFSYYGM | GWVRQAPGKGLEWVSGISYSGSGTYYADSVKGRFTISRDNSKNTLYLQMNSLRAEDTAVYYCARDYVGNLDYWGQGTLVTVSSGGGGSGGGGSGGGGSDIQMTQSPSSLSASVGDRVTITCRASQSISSYLNWYQQKPGKAPKLLIYAASSLQSGVPSRFSGSGSGTDFTLTISSLQPEDFATYYCQQSDTPSTFGQGTKLEIKRTDYKDHDGDYKDHDIDYKDDDDKAAALPETGGHHHHHH | Non-toxic |

A detailed CSV was also provided as output, along with the following graph:

Where a higher value indicates more toxicity of a given attention head. Clearly, our protein had close to no toxicity.

## How it works:

This study used UniProt release 2022\_04 data to identify toxic and non-toxic proteins. The model was trained using CD-HIT version 4.8.1, discarding sequences with non-standard residue codes. The final dataset contains 2475 toxic sequences and 214,740 non-toxic sequences, with a toxic to non-toxic ratio of approximately 90. The predictive model was built using raw amino acid sequences without additional features extracted or generated.

The curated sequences were divided into two groups: 203 toxic and 2337 non-toxic sequences uploaded after July 2021, used in a blind test set for CSM-Toxin and ToxinPred2, and 236 positive and 21,294 negative sequences for cross-validation. The remaining data was split into five parts for training and validation, with average performance metrics.

ProteinBERT was used as a base of the model. ProteinBERT, a model inspired by BERT, treats amino acids as words and protein sequences as sentences. It uses attention mechanisms to capture complex connections between amino acids. ProteinBERT was pre-trained using the Masked Language Model technique, capturing connections between amino acids and their surroundings. The model was pre-trained on over 100 million protein sequences from the UniProt database. The focus was on Global Ontology for binary prediction (toxic/non-toxic) based on the entire sequence.

## Toxicity of various components of our lipid.

Our lipid composition is as follows:

* DSPC: 790.15 g/mol
* ALC-0315: 766.29 g/mol
* Cholesterol: 386.65 g/mol
* PEG: 2941.642 g/mol

To determine toxicity of individual components, we utilised the Toxicity Estimation Software Tool (TEST). Out of the available endpoints present in the software, we used

* **The Oral Rat LD50 Endpoint:** This indicates a concentration of a given chemical that leads to the death of 50% rats in a given population when administered orally.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Query** | **SmilesRan** | **Exp\_Value: -Log10(mol/kg)** | **Pred\_Value: -Log10(mol/kg)** | **Exp\_Value: mg/kg** | **Pred\_Value: mg/kg** |
| PEG | [H]OCCO | 1.12 | 1.89 | 4698.42 | 802.86 |
| DSPC | O=C(OCC(OC(=O)CCCCCCCCCCCCCCCCC)COP(=O)([O-])OCC[N+](C)(C)C)CCCCCCCCCCCCCCCCC | N/A | N/A | N/A | N/A |
| Cholesterol | [H]C12CC=C3CC(O)CCC3(C)C2([H])CCC4(C)C([H])(CCC14[H])C(C)CCCC(C)C | N/A | 2.55 | N/A | 1078.24 |
| ALC-0315 | O=C(OCCCCCCN(CCCCO)CCCCCCOC(=O)C(CCCCCC)CCCCCCCC)C(CCCCCC)CCCCCCCC | N/A | N/A | N/A | N/A |

We further observe that neither experimental nor predicted values of toxic concentrations were present for DSPC AND ALC-0315. Hence, we further utilise the software’s ability to provide results for similar chemicals which generated the following data:

## Predictions for the test chemical and for the most similar chemicals in the training set for DSPC:

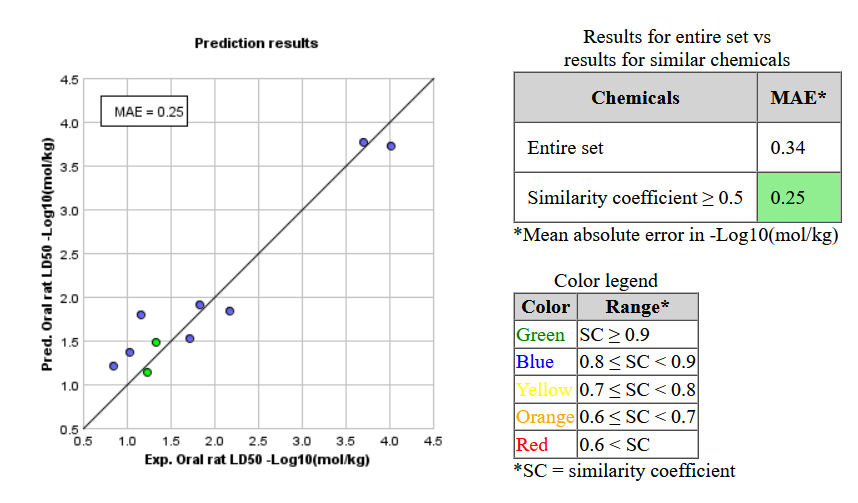
A graph with numbers and a chart with numbers and a chart with numbers

Description automatically generated with medium confidence

A screenshot of a graph

Description automatically generated

## Predictions for the test chemical and for the most similar chemicals in the training set for ALC-0315:



A chart with different colored squares

Description automatically generated

* **Ames Mutagenicity:** This checks if the compound is positive for mutagenicity if it induces revertant colony growth in ant strain of *Salmonella typhimurium.*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Query** | **SmilesRan** | **Exp\_Value: -Log10(mol/kg)** | **Pred\_Value: -Log10(mol/kg)** | **Exp\_Result:** | **Pred\_Result: mg/kg** |
| DSPC | [O=C(OCC(OC(=O)CCCCCCCCCCCCCCCCC)COP(=O)([O-])OCC[N+](C)(C)C)CCCCCCCCCCCCCCCCC | N/A | -0.01 | N/A | Mutagenicity Negative |
| Cholesterol | [H]C12CC=C3CC(O)CCC3(C)C2([H])CCC4(C)C([H])(CCC14[H])C(C)CCCC(C)C | N/A | 0.2 | N/A | Mutagenicity Negative |
| PEG | [[H]OCCO | 0.00 | -0.03 | Mutagenicity Negative | Mutagenicity Negative |
| ALC-0315 | O=C(OCCCCCCN(CCCCO)CCCCCCOC(=O)C(CCCCCC)CCCCCCCC)C(CCCCCC)CCCCCCCC | N/A | 0.03 | N/A | Mutagenicity Negative |

We observe that all the compounds in our batch are found to be Mutagenicity Negative.

* **Developmental Toxicity:** This checks whether or not a chemical causes developmental toxicity effects to humans or animals.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Query** | **SmilesRan** | **Exp\_Value: -Log10(mol/kg)** | **Pred\_Value: -Log10(mol/kg)** | **Exp\_Result:** | **Pred\_Result: mg/kg** |
| DSPC | O=C(OCC(OC(=O)CCCCCCCCCCCCCCCCC)COP(=O)([O-])OCC[N+](C)(C)C)CCCCCCCCCCCCCCCCC | N/A | N/A | N/A | N/A |
| Cholesterol | H]C12CC=C3CC(O)CCC3(C)C2([H])CCC4(C)C([H])(CCC14[H])C(C)CCCC(C)C | N/A | 0.97 | N/A | Developmental Toxicant |
| PEG | [H]OCCO | N/A | 0.28 | N/A | Developmental NON-Toxicant |
| ALC-0315 | O=C(OCCCCCCN(CCCCO)CCCCCCOC(=O)C(CCCCCC)CCCCCCCC)C(CCCCCC)CCCCCCCC | N/A | N/A | N/A | N/A |

We observe that some of the above compounds in the batch can be toxic to animals like Cholesterol.

In single chemical mode, one can generate predicted environmental transformation products by checking “Run CTS” which can help us create suitable environment like hydrolysis, abiotic reduction and, human metabolism. .These tests have been made to consider “human metabolism” in animals cells. If likely transformation products are generated, the toxicity will be displayed.

This test is especially useful as we are using the mRNA on animal cells, so this will be able to tell us to what extent Cholesterol and other compounds can be used so as to avoid toxicity.

## Issues with the method:

Clearly, while TEST works well for some of the chemicals present in our LNP, we observe that the first major issue is the metric of testing toxicity of chemicals. LD50 focuses on oral administration of given component in a rat population, whereas our LNP is aimed to administered intravenously and intraperitoneally. Further, for DSPC, the predictions for similar chemicals and chemicals in training data set vary largely. Even for mutagenicity and developmental toxicity, we can expect experimental data to vary from the predicted values.

Further, another obvious issue is the fact that individual components and assembled LNPs with mRNA have variations in toxicity that are difficult to predict by the analysis of individual components only. We also were unable to find other reliable *in silico* methods to determine immunogenicity of our mRNA-LNP, while leaning over existing literature.

## References:

* Morozov V, Rodrigues CHM, Ascher DB. CSM-Toxin: A Web-Server for Predicting Protein Toxicity. Pharmaceutics. 2023 Jan 28;15(2):431. doi: 10.3390/pharmaceutics15020431. PMID: 36839752; PMCID: PMC9966851.
* Gadaleta, D., Vukovic;, K., Toma, C., Lavado, G. J., Karmaus, A. L., Mansouri, K., Kleinstreuer, N. C., Benfenati, E., & Roncaglioni, A. (2019, August 30). *Sar and QSAR modeling of a large collection of LD50 rat acute oral toxicity data - journal of Cheminformatics*. BioMed Central. https://jcheminf.biomedcentral.com/articles/10.1186/s13321-019-0383-2
* https://biopharma.labcorp.com/catalog/crop-chemical/us-epa-toxicity-estimation-software-suite-test.html
* https://www.epa.gov/chemical-research/toxicity-estimation-software-tool-test