BIBAD

Research Journal of Biological Sciences 9 (2): 16-19, 2016
ISSN: 1308-3961, E-ISSN: 1308-0261

In silico Screening and Validation of Immunogenic Viral RNA Sequences

Gökalp ÇELİK[1] Eda SÜER[2] Sadık ÇİĞDEM[1] Catherine Ann Moroski ERKUL[2] Burak YILMAZ[1] Yusuf ERKUL[1]*
[1]Sentegen Biotechnology, Bilkent Cyberpark, Ankara
[2]Sentebiolab Biotechnology, Bilkent Cyberpark, Ankara

*Corresponding Author:
E-mail: yusuf@kernalbio.com

Received: March 23, 2016
Accepted: June 31, 2016

Abstract

Identification of RNA sequences found in the genomes of viruses that induce intracellular innate immunity has important implications in the fields of immunology and mRNA based gene therapy. Such findings help deepen our understanding of the innate immune system and can help us devise effective mRNA based gene therapies, i.e. mRNA replacements [1]. Although in vitro studies performed thus far have sporadically identified several immunogenic RNA sequences, it is still not possible to predict which sequences would be immunogenic within the cell. Within eukaryotic cellular and endosomal membranes, there are several receptors that are responsible for recognizing viral genomes. Some of these receptors recognize DNA with unmethylated CpG sequences (TLR9) or double stranded RNA (TLR3), while others such as TLR7, TLR8, PKR, or OAS are activated by single stranded (foreign) RNA molecules [2].

Crystal structure analysis of TLR8 revealed that this receptor can be activated by the binding of uridine and a short oligoribonucleotide with a predicted length of 2-20 bases [3]. Differential recognition of the viral genome by these receptors, which spares self RNAs, suggests the presence of selective activation by sequences that are frequently found in the viral genomes but that are either absent or infrequent in the human genome. We therefore identified 12 nucleotide-long RNA sequences within the genomes of RNA viruses and scored them according to their frequencies using a brute-force algorithm. For validation, we compared the scores of known immunogenic sequences to those of non-immunogenic sequences. Consequently, we note that in silico screening can be an effective method for identifying immunogenic RNA sequences. Following in vitro validation of identified immunogenic sequences, we aim to develop a codon optimization algorithm to design non-immunogenic mRNAs for gene replacement therapies.

Keywords: Viral RNA, Immunogenicity, In silico screening, mRNA Vaccines, mRNA replacement

INTRODUCTION

In vertebrates, innate immunity provides a basic level of protection against pathogens by sensing foreign molecules or molecular patterns associated with invaders. Some of these receptors bind structural elements, such as bacterial cell wall lipids, flagellar proteins, or fungal cell wall polysaccharides while others recognize viral genome (for a general review, please see [4, 5]). Among the latter, human TLR7 and TLR8 have been known for their role in the recognition of single stranded RNA of viruses [6–9]. When activated, these receptors in turn stimulate various intracellular signalling pathways (e.g. NF-κB and IRF7 pathways) resulting in cytokine and interferon release, translation stalling, RNA degradation, and apoptosis [2]. Collectively, these molecular responses are intended to halt the viral replication cycle within the cell and alert neighboring cells against potential invaders.

Although the structure of TLR7 has not been definitively solved via crystallography yet, potential ligand bindings sites were identified via homology modelling [10]. The crystal structure of TLR8, however, was determined recently [3]. It revealed two seperate ligand binding domains: one binding uridine and another binding a short, 2-20 nucleotide long oligoribonucleotide (ORN). It was also shown that ligand binding at both domains was required to activate the receptor. These findings are in alignment with other groups' work on TLR8 ligands, which have consistently shown GU- and AU-rich ORN sequences to be activators of TLR8 [6, 11, 12].

TLR7 and TLR8 are known to recognize viral RNA, but not mammalian mRNA or tRNA [13, 14]. There are two major mechanisms that could underly this selective binding. The first one is the recognition of unmodified RNA, because mammalian mRNAs and tRNAs are chemically modified. Mammalian RNAs are known to contain over 110 modified nucleotides [15]. There are numerous studies demonstrating that in vitro transcribed (IVT) mRNA containing modified nucleotides such as pseudouridine, N1-methylpseudouridine, 2-thiouridine, 5-methylcytidine, and N6-methyladenosine does not activate TLR7 or TLR8 and, therefore, is not immunogenic [14, 16–20]. The second mechanism with which selective binding of TLRs to viral RNA could occur is the recognition of sequence patterns that are frequently found in viral RNAs. Viruses with ssRNA genomes, such as HIV and HCV, have sequences that are known to be immunogenic via TLR7 and/or TLR8 [6, 9, 12, 21]. Futhermore, it was recently discovered that codon optimized mRNA sequences, which were engineered to have a higher GC content (similar to that of the human genome), significantly reduced the immunogenicity of IVT mRNA [1]. Based on these results, we aimed to identify immunogenic ORN sequences via analyzing viral genomes to assist in the design of deliberately immunogenic or non-immunogenic RNAs. We chose 12 nucleotides as the optimal ORN length because, based on our preliminary analysis, the scores of 9-mers were not conducive in differentiating between immunogenic and non-immunogenic sequences and the frequency of 17-mers were too low for a consequential analysis. We therefore analyzed 12-mers in our training data set, and then generated a 12-mer list from available ORN sequences in our validation data set.

Immunogenic ORN sequences can function as RNA adjuvants to stimulate TLRs and induce a cytokine response [22, 23]. Such ORN adjuvants can be used in indications where small molecule TLR agonists have shown promise against pathogens or cancer, antivirals for the treatment of HIV, HBC, HCV, or influenza infections, or immune modulators in Alzheimers [24- 29]. Additionally, through the use

of a novel codon optimization algorithm, these immunogenic sequences can be avoided while creating synthetic mRNA sequences to enable well-tolerated and highly efficacious mRNA replacement therapies.

Here, we report the results of our in silico screen that ranked viral ORNs (12-mers) according to their frequencies in the ssRNA viral genome database and hence scored their immunogenic potential. We then proceeded to validate these scores, by comparing the frequencies of previously reported immunogenic ORNs to those of non-immunogenic ORNs [30]. This comparison yielded a statistically significant difference between the two groups, indicating that in silico screening can be a valid approach for the identification of potentially immunogenic sequences.

MATERIALS and METHODS

Identification of viral sequences

A total of 1923 sequences, which belong to single stranded RNA viruses and their positive and negative, unclassified and unassigned strands were identified and 467 of these sequences were downloaded using a human host filter from the NCBI viral genome database (http://www.ncbi.nlm.nih.gov/genomes/GenomesGroup.cgi?taxid=10239).

Validation

For validation, single-stranded RNA sequences with previously identified immunogenicities (positive and negative) were used. From the ssRNA data set of Chaudhary et al. [30], 2875 sequences with positive immunogenicity (immunogenic) and 5372 sequences with negative immunogenicity (non-immunogenic) were obtained.

Algorithm

Viral and control sequences downloaded in FASTA format were subjected to a brute-force string match algorithm which allowed 9, 12 and 17-base long oligoribonucleotides with unique sequences to be scanned and listed with their repetition number in a .txt file. The use of a brute-force string match algorithm was chosen due to its moderate computing power and memory which suits the low number of viral sequences provided. Moreover, this approach eliminated the risk of overlooking any sequences which could have occurred by using greedy or heuristic algorithms.

Due to insufficient diversity of 9-base long sequences (absence of virus-only sequences) and small number of repetitions of 17-base long sequences (with a maximum of 3 or 4 repetitions), these were unsuitable for statistical analysis and resulted in the evaluation of 12-base long sequences. These sequences are suitable for analysis due to their sufficient number of repetitions and the presence of 12-base long virus-only sequences.

Pseudocode:

```
Algorithm BruteForceStringMatch(T[0...n-1],
P[0...m-1])
for i ← 0 to n-m do
j ← 0
while j < m and P[j] = T[i+j] do
j++
if j = m then return i
return -1
```

Statistical analysis

The frequency scores of previously identified 12-mers according to their presence in the immunogenic set and non-immunogenic set from Chaudhary et al. [30] were compared in SPSS software using Mann-Whitney test. The level of significance (alpha value) was set to 0.05 and the p values less than this were considered significant.

RESULTS and DISCUSSION

The frequency values (scores) of 12-mer ssRNAs in virus genomes were identified. The most frequent 12-mer (TTTTTTTTTTTT/UUUUUUUUUUUU) has a score of 195. In order to evaluate whether or not this pool of sequences contains immunogenic sequences, validation was performed using positive (immunogenic) and negative (non-immunogenic) short ssRNA data. Positive and negative ssRNA 12-mers were listed and scored according to their viral frequencies (Table 1). Upon comparing these scores with the scores of ssRNA sequences that are considered immunogenic and non-immunogenic using Mann-Whitney test, immunogenic sequences were observed to have a higher score ($p<0.0004$). The results showed that, as expected, 12-mer ssRNAs found in the virus genome are immunogenic (Figure 1a and 1b), and the method we used in our study is therefore a valid method to identify the immunogenicity of a given RNA sequence.

Table 1. Viral 12-mers and their scores assigned according to their frequency in the viral genome. The first two sequences (in red) are taken as outliers in the SPSS analysis.

Sequence	Score
TTTTTTTTTTTT	195
AAAAAAAAAAAA	136
AAAAACTTAGGA	81
TTTAAGAAAAAA	65
AAGAAAAACTTA	58
AAAAAACTTAGG	56
TAAGAAAAACTT	56
TTAAGAAAAAAC	50
ATTAAGAAAAAC	48
AGAAAAACTTAG	48
GAAAAACTTAGG	46
GGTTAGAGGAGA	41
GTTAGAGGAGAC	40
TTAGAGGAGACC	39
TAGAGGAGACCC	39
TGGTACATGTGG	39
TTAAGAAAAACT	38
CCTAGGATCCAC	37
TAGGATCCACTG	37
GATCCACTGTGC	37
ATCCACTGTGCG	37
AGGATCCACTGT	37
GCCTAGGATCCA	37
GGATCCACTGTG	37

ATTAAGAAAAAA	36
TTAAGAAAAAAT	36
CTAGGATCCACT	36
TGAAAAAAACAT	36
TATTAAGAAAAA	35
AAAACTTAGGAT	35
CAACATGATGGG	35
ACAACATGATGG	35
TACAACATGATG	34
GATTAAGAAAAA	34

Mann-Whitney Test

Ranks

	Grup	N	Mean Rank	Sum of Ranks
Puan	1.00	2875	4370.60	12565478.50
	2.00	5372	3992.02	21445149.50
	Total	8247		

Test Statistics[a]

	Puan
Mann-Whitney U	7013271.500
Wilcoxon W	21445149.50
Z	-9.103
Asymp. Sig. (2-tailed)	.000

a. Grouping Variable: Grup

Mann-Whitney Test

Ranks

	Grup	N	Mean Rank	Sum of Ranks
Puan	1.00	2873	4367.90	12548985.50
	2.00	5372	3992.02	21445149.50
	Total	8245		

Test Statistics[a]

	Puan
Mann-Whitney U	7013271.500
Wilcoxon W	21445149.50
Z	-9.041
Asymp. Sig. (2-tailed)	.000

a. Grouping Variable: Grup

Figure 1. Results of the Mann-Whitney test to identify immunogenicity of the viral 12-mers. Group 1.00 denotes sequences with positive immunogenicity and Group 2.00 denotes sequences with negative immunogenicity. A) Results of analysis including the two outlier sequences. B) Results of analysis excluding the two outlier sequences. In the results of A and B, a $p<0.0004$ shows that the results are significant in both cases and that the viral 12-mers are immunogenic.

Cellular administration of unmodified synthetic RNA results in the stimulation of cellular innate immunity [2]. This is because unmodified RNA, such as viral RNA, is considered "foreign" and is confiscated as soon as possible, before the cell's machinery can use it to produce protein. Although it has been shown that cellular administration of in vitro synthesized modified RNA (with pseudouridine and methylcytidine) results in a significantly abolished immune response compared to that of unmodified RNA, there is always room for improvement [14, 16, 31, 32]. Additional modifications, such as optimization of 5' or 3' UTRs, are still required to attain an increased half-life. Thess et al. [1] have optimized the sequence of RNA instead of using modified nucleotides to prevent immunogenicity and have shown that codon optimization is a valid method to evade innate immunity and an alternative to using modified bases. Codon optimization is often used to maximize protein expression by increasing the translation efficiency of the gene of interest. This technique may involve transformation of a given oligonucleotide sequence of one species to the oligonucleotide sequence of another species (i.e. mouse sequence to human sequence or vice versa). Another strategy is to simply change the codons to overcome codon bias and to replace a low frequency codon with a higher frequency codon so that, while the sequence changes, it still encodes the same protein. With the data to prove that codon optimization in synthetic RNA can enhance half-life and translation efficiency, the same principle can be applied to the designated RNA sequence. With synthetic mRNA therapeutics on the rise, identifying the immunogenicity of a given RNA sequence would be of great use because sequences with positive immunogenicity can be transformed into non-immunogenic sequences via codon optimization. Development of an algorithm that employs codon optimization to design non-immunogenic RNA is the final target of this study. Synthesis of RNA with low immunogenicity will lead to the more effective development of RNA-based vaccine and therapeutics.

The comparison method we have designed proves to be a valid checkpoint before designing and synthesizing the RNA of interest. As a future study, the immunogenicity of these sequences must be validated in vitro. Therefore, production of these sequences as synthetic RNA and administration to a cellular system such as monocyte-derived dendritic cells (MDDC) is required.

REFERENCES

[1] A. Thess, S. Grund, B. L. Mui, M. J. Hope, P. Baumhof, M. Fotin-Mleczek, and T. Schlake. 2015. Sequence-engineered mRNA Without Chemical Nucleoside Modifications Enables an Effective Protein Therapy in Large Animals. Molecular Therapy. 23(9):1456-1464.

[2] U. Sahin, K. Karikó, and Ö. Türeci. 2014. mRNA-based therapeutics — developing a new class of drugs. Nature Reviews Drug Discovery. 13(10):759-780.

[3] H. Tanji, U. Ohto, T. Shibata, M. Taoka, Y. Yamauchi, T. Isobe, K. Miyake, and T. Shimizu. 2015. Toll-like recept or 8 senses degradation products of single-stranded RNA. Nature Structural Molecular Biology. 22(2):109-115.

[4] S. G. Hansen, J. C. Ford, M. S. Lewis, A. B. Ventura, C. M. Hughes, L. Coyne-Johnson, N. Whizin, K. Oswald, R. Shoemaker, T. Swanson, A. W. Legasse, M. J. Chiuchiolo, C. L. Parks, M. K. Axthelm, J. a Nelson, M. a Jarvis, M. Piatak, J. D. Lifson, and L. J. Picker. 2011. Profound early control of highly pathogenic SIV by an effector memory T-cell vaccine. Nature. 473(7348):523-527.

[5] J. K. Dowling and A. Mansell. 2016. Toll-like receptors: the swiss army knife of immunity and vaccine development. Clinical & Translational Immunology. 5(5):e85.

[6] F. Heil, H. Hemmi, H. Hochrein, F. Ampenberger, C. Kirschning, S. Akira, G. Lipford, H. Wagner, and S. Bauer.

2004. Species-specific recognition of single-stranded RNA via toll-like receptor 7 and 8., Science. 303(5663):1526-1529.

[7] S. S. Diebold, T. Kaisho, H. Hemmi, S. Akira, and C. Reis e Sousa. 2004. Innate antiviral responses by means of TLR7-mediated recognition of single-stranded RNA. Science. 303(5663):1529-1531.

[8] M. Kane, L. K. Case, C. Wang, L. Yurkovetskiy, S. Dikiy, T. V Golovkina, O. Adachi, et al. 2011. Innate immune sensing of retroviral infection via Toll-like receptor 7 occurs upon viral entry. Immunity. 35(1):135-145.

[9] N. Lee, W. N. Moss, T. A. Yario, J. A. Steitz, A. Arvey, I. Tempera, K. Tsai, et al. 2015. EBV noncoding RNA binds nascent RNA to drive host PAX5 to viral DNA. Cell. 160(4):607-618.

[10] T. Wei, J. Gong, F. Jamitzky, W. M. Heckl, R. W. Stark, and S. C. Rössle. 2009. Homology modeling of human Toll-like receptors TLR7, 8, and 9 ligand-binding domains. Protein Science. 18(8):1684-1691.

[11] A. Forsbach, J.-G. Nemorin, C. Montino, C. Müller, U. Samulowitz, A. P. Vicari, M. Jurk, G. K. Mutwiri, A. M. Krieg, G. B. Lipford, and J. Vollmer. 2008. Identification of RNA sequence motifs stimulating sequence-specific TLR8-dependent immune responses. Journal of Immunology. 180(6):3729-3738.

[12] Y. Zhang, M. El-Far, F. P. Dupuy, M. S. Abdel-Hakeem, Z. He, F. A. Procopio, Y. Shi, E. K. Haddad, P. Ancuta, R.-P. Sekaly, and E. A. Said. 2016. HCV RNA Activates APCs via TLR7/TLR8 While Virus Selectively Stimulates Macrophages Without Inducing Antiviral Responses. Scientific Reports. 6:29447-29459.

[13] G. K. Koski, K. Kariko, S. Xu, D. Weissman, P. A. Cohen, and B. J. Czerniecki. 2004. Cutting Edge: Innate Immune System Discriminates between RNA Containing Bacterial versus Eukaryotic Structural Features That Prime for High-Level IL-12 Secretion by Dendritic Cells. Journal of Immunology. 172(7):3989-3993.

[14] K. Karikó, M. Buckstein, H. Ni, and D. Weissman. 2005. Suppression of RNA recognition by Toll-like receptors: The impact of nucleoside modification and the evolutionary origin of RNA. Immunity. 23(2):165-175.

[15] S. Li and C. E. Mason. 2014. The pivotal regulatory landscape of RNA modifications. Annual Review of Genomics and Human Genetics. 15:127-150.

[16] K. Karikó, H. Muramatsu, F. A. Welsh, J. Ludwig, H. Kato, S. Akira, and D. Weissman. 2008. Incorporation of pseudouridine into mRNA yields superior nonimmunogenic vector with increased translational capacity and biological stability. Molecular Therapy. 16(11):1833-1840.

[17] M. S. D. Kormann, G. Hasenpusch, M. K. Aneja, G. Nica, A. W. Flemmer, S. Herber-jonat, M. Huppmann, L. E. Mays, M. Illenyi, A. Schams, M. Griese, I. Bittmann, R. Handgretinger, D. Hartl, J. Rosenecker, and C. Rudolph. 2011. Expression of therapeutic proteins after delivery of chemically modified mRNA in mice. Nature Biotechnology. 29(2):154-157.

[18] Y. Wang, Y. Li, J. I. Toth, M. D. Petroski, Z. Zhang, and J. C. Zhao. 2014. N6-methyladenosine modification destabilizes developmental regulators in embryonic stem cells. Nature Cell Biology. 16(2):191-198.

[19] O. Andries, S. Mc Cafferty, S. C. De Smedt, R. Weiss, N. N. Sanders, and T. Kitada. 2015. N1-methylpseudouridine-incorporated mRNA outperforms pseudouridine-incorporated mRNA by providing enhanced protein expression and reduced immunogenicity in mammalian cell lines and mice. Journal of Controlled Release. 217:337-344.

[20] X. Li, X. Xiong, K. Wang, L. Wang, X. Shu, S. Ma, and C. Yi. 2016. Transcriptome-wide mapping reveals reversible and dynamic N1-methyladenosine methylome. Nature Chemical Biology. 12(5):311-316.

[21] K. W. Cohen, A.-S. Dugast, G. Alter, M. J. McElrath, and L. Stamatatos. 2015. HIV-1 single-stranded RNA induces CXCL13 secretion in human monocytes via TLR7 activation and plasmacytoid dendritic cell-derived type I IFN. Journal of Immunology. 194(6):2769-2775.

[22] D. N. Nguyen, S. C.-Y. Chen, J. Lu, M. Goldberg, P. Kim, A. Sprague, T. Novobrantseva, J. Sherman, S. Shulga-Morskaya, A. de Fougerolles, J. Chen, R. Langer, and D. G. Anderson. 2009. Drug delivery-mediated control of RNA immunostimulation. Molecular Therapy. 17(9):1555-1562.

[23] C. Bourquin, L. Schmidt, A.-L. Lanz, B. Storch, C. Wurzenberger, D. Anz, N. Sandholzer, R. Mocikat, M. Berger, H. Poeck, G. Hartmann, V. Hornung, and S. Endres. 2009. Immunostimulatory RNA oligonucleotides induce an effective antitumoral NK cell response through the TLR7. Journal of Immunology, 183(10):6078-6086.

[24] H. M. Jones, P. L. S. Chan, P. H. van der Graaf, and R. Webster. 2012. Use of modelling and simulation techniques to support decision making on the progression of PF-04878691, a TLR7 agonist being developed for hepatitis C. British Journal of Clinical Pharmacology. 73(1):77-92.

[25] B. G. Zhao, J. P. Vasilakos, D. Tross, D. Smirnov, and D. M. Klinman. 2014. Combination therapy targeting toll like receptors 7, 8 and 9 eliminates large established tumors. Journal for Immunotherapy of Cancer. 2:12-21.

[26] M. E. Gambuzza, V. Sofo, F. M. Salmeri, L. Soraci, S. Marino, and P. Bramanti. 2014. Toll-like receptors in Alzheimer's disease: a therapeutic perspective. CNS & Neurological Disorders Drug Targets. 13(9):1542-1558.

[27] M. A. Moody, S. Santra, N. A. Vandergrift, L. L. Sutherland, T. C. Gurley, M. S. Drinker, A. A. Allen, et al. 2014. Toll-like receptor 7/8 (TLR7/8) and TLR9 agonists cooperate to enhance HIV-1 envelope antibody responses in rhesus macaques. Journal of Virology. 88(6):3329-3339.

[28] I. Rebbapragada, G. Birkus, J. Perry, W. Xing, H. Kwon, and S. Pflanz. 2016. Molecular Determinants of GS-9620-Dependent TLR7 Activation. PLoS One. 11(1):e0146835.

[29] S. J. Dovedi, A. L. Adlard, Y. Ota, M. Murata, E. Sugaru, E. Koga-Yamakawa, K. Eguchi, et al. 2016. Intravenous administration of the selective toll-like receptor 7 agonist DSR-29133 leads to anti-tumor efficacy in murine solid tumor models which can be potentiated by combination with fractionated radiotherapy. Oncotarget. 7(13):17035-17046.

[30] K. Chaudhary, G. Nagpal, S. K. Dhanda, and G. P. S. Raghava. 2016. Prediction of Immunomodulatory potential of an RNA sequence for designing non-toxic siRNAs and RNA-based vaccine adjuvants. Scientific Reports. 6:20678-20688.

[31] K. Karikó, H. Muramatsu, J. Ludwig, and D. Weissman. 2011. Generating the optimal mRNA for therapy: HPLC purification eliminates immune activation and improves translation of nucleoside-modified, protein-encoding mRNA. Nucleic Acids Research. 39(21):e142.

[32] K. Karikó, H. Muramatsu, J. M. Keller, and D. Weissman. 2012. Increased erythropoiesis in mice injected with submicrogram quantities of pseudouridine-containing mRNA encoding erythropoietin. Molecular Therapy. 20(5):948-953.