Applicant:
CureVac AG, Germany

RNA Sequence Adaptation

Introduction

The present invention is directed to a method for modifying the retention time of RNA on a chromatographic column. The present invention also concerns a method for purifying RNA from a mixture of at least two RNA species. Furthermore, the present invention relates to a method for co-purifying at least two RNA species from a mixture of at least two RNA species. In particular, the present invention provides a method for harmonizing the numbers of A and/or U nucleotides in at least two RNA species. The present invention is also directed to RNA obtainable by said methods, a composition comprising said RNA or a vaccine comprising said RNA and methods for producing such RNA and compositions. Further, the invention concerns a kit, particularly a kit of parts, comprising the RNA, composition or vaccine. The invention is further directed to a method of treating or preventing a disorder or a disease, first and second medical uses of the RNA, composition and vaccine. Moreover, the present invention concerns a method for providing an adapted RNA sequence or an adapted RNA mixture.

Ribonucleic acid (RNA)-based therapeutics can be used in immunotherapy, gene therapy and genetic vaccination. They can provide highly specific and individual treatment options for the therapy of a large variety of diseases.

For certain medical treatments and applications, it is desired to apply an RNA mixture comprising different RNA molecule species. Examples of such treatments based on an RNA mixture include the application of polyvalent RNA mixtures that provide protection against several serotypes of a pathogen (e.g. hemagglutinin (HA) from multiple serotypes of Influenza A and B virus); RNA mixtures that provide different antigens from one pathogen (e.g. different antigens from Influenza, such as HA, nucleoprotein (NP), neuraminidase (NA) etc.); RNA mixtures that provide protection against several isoforms or variants of a cancer antigen (e.g. prostate specific antigen (PSA) in the context of prostate carcinoma); RNA mixtures that provide different epitopes of an antigen; RNA mixtures that contain a cancer specific and/or patient specific mixture of cancer antigens (expressed antigens or mutated antigens); RNA mixtures that encode a variety of antibodies (e.g., antibodies that are targeted against different epitopes of one or more proteins), or any other therapeutically active RNA mixture (e.g., encoding different isoforms of an enzyme for molecular therapy, different therapeutic proteins for treatment of an indication wherein several proteins have to be supplemented).

A significant step forward in the field of RNA-based therapeutics was achieved by the establishment of an RNA manufacturing process that has been approved by regulatory authorities, implementing various quality controls on DNA level and RNA level as described in detail in WO2016/180430A1. A key element of said process is the preparative purification of the RNA product via RP-HPLC (as described in WO 2008/077592A1).

So far, RNA production methods are, however, only suitable to produce one single specific RNA molecule species at a time, such as an RNA molecule encoding one specific therapeutic target. Each RNA molecule species has to be

produced in a separate production process. For the production of an RNA mixture comprising different RNA molecule species, separate production processes have to be performed, and the separately produced RNA molecule species have to be mixed to generate an RNA mixture. Accordingly, the conventional production of an RNA mixture (e.g. polyvalent vaccine) containing several RNA molecule species (e.g. RNA coding for different antigens, e.g. Influenza virus antigens) is laborious, costly, and time consuming since it requires several runs for DNA template production, RNA production und HPLC purification of the RNA product. Apart from economic reasons, especially in the context of pandemic scenarios, an acceleration of the production of RNA mixtures (e.g. polyvalent RNA vaccine) may be highly advantageous and of major importance for public health. Therefore, it is desirable to produce and purify RNA mixtures simultaneously, ideally in only one production and purification process.

Simultaneous production, purification, and analysis of RNA mixtures are of particular importance for the manufacturing of multivalent/polyvalent vaccines, particularly a multivalent/polyvalent influenza RNA vaccine. Advantageously, the manufacturing of such a multivalent/polyvalent influenza RNA vaccine should allow for a fast exchange of (one or many) antigens (e.g. for seasonal adaptations of the vaccine; pandemic scenario) without changing the conditions for key processes of production (e.g., purification, analysis etc.).

Recently, a procedure for the simultaneous production of different RNA molecule species in one production process has been developed, as described in WO20171090134, aiming to economize and accelerate the production of RNA mixture based therapeutics.

Simultaneous production of different RNA molecule species as e.g. described in WO20171090134 would save time, labor costs, production costs, and production capacities (e.g., space, equipment) in the manufacturing of RNA mixture based therapeutics. However, a severe problem is associated with the use of RP-HPLC for simultaneous preparative purification of the obtained RNA mixture (herein also referred to as "co-purification"). Different RNA molecule species in an RNA mixture commonly elute at different time points (due to different retention time in HPLC), which renders co-purification via HPLC technically impossible in most of the cases (illustrated in Figure 1A and B). For example, co-purification via HPLC would only be possible in cases, where all different RNA molecule species comprised in the RNA mixture elute at essentially the same time point leading to an overlay of the different RNA product peaks facilitating separation of clean full-length RNA product mixture (peak fraction) and separation from impurities (illustrated in Figure 1C and D).

Accordingly, the purification of an RNA-mixture, such as an RNA-based therapeutic comprising multiple RNA species (e.g. a multivalent RNA vaccine), is not feasible with the current methodology known in the art. In particular, the co-purification of individual RNA species in an RNA mixture is not feasible. However, such a co-purification is highly desirable, for example in the context of the development of a multivalent/polyvalent influenza RNA vaccine platform that is capable of being adapted to seasonal changes in the antigen composition. In such a situation, rapid exchanges of antigens (e.g. influenza hemagglutinin (HA) and/or neuraminidase (NA)) are required. The methods available in the field do not allow for the co-purification of different RNA species from an RNA mixture, in particular via HPLC, since different species of RNA molecules encoding different antigens would typically elute at different time points.

For similar reasons, also simultaneous HPLC-based quality control of RNA mixtures or (multivalent) vaccine (either produced in one production process or individually produced and mixed) is not feasible when using prior art techniques. For the analysis of therapeutics comprising mixtures of RNAs, it is required that the different components (i.e. different RNA molecule species) of the drug product can be characterized in terms of presence, integrity, ratio and quantity (quality control parameter). Such quality controls may be implemented during or following production of the RNA mixture, and/or as a batch release quality control. However, as described above, different species of RNA molecules in an RNA mixture typically elute at different time points, which strongly impedes the simultaneous assessment of the quality/integrity of the individual RNA molecule species/or the whole RNA mixture via HPLC ("co-analysis") (illustrated in Figure 1A and B). A reliable determination of the integrity of the RNA mixture via HPLC would only be possible in cases where the clear discrimination of the peak areas for each RNA molecule species is possible (simultaneous determination of the RNA quality for each RNA molecule species; see Figure 1E) and/or where different RNA species in the RNA mixture elute at essentially the same time point, leading to an essentially complete overlay of the HPLC peaks (determination of the RNA quality of the whole RNA mixture; see Figure 1C and D).

Accordingly, also the analysis of an RNA mixture, such as an RNA therapeutic comprising multiple RNA species (e.g. a multivalent Influenza RNA vaccine), is not feasible with the current methodology provided in the art. In particular, co-analysis of RNA species in an RNA mixture is not possible. For example, co-analysis, e.g. via HPLC, is not applicable in the case of a multivalent RNA vaccine platform, where rapid exchanges of antigens are required (e.g. for seasonal vaccines, such as an influenza vaccine), since different species of RNA molecules each encoding different antigens (such as Influenza HA and/or NA) elute at different time points, typically leading to partially overlapping HPLC peaks.

In summary, the technical difficulties described above strongly hinder the (co-)purification and the (co-)analysis of RNA mixtures. In particular, RP-HPLC, which represents a key element for the purification and the analysis of RNA in general and particularly of GMP-grade RNA, cannot be employed in the (co-)purification and (co-)analysis of individual RNA species in RNA mixtures. Furthermore, the described problems have to be solved to facilitate the development of a multivalent/polyvalent RNA manufacturing process for producing fast-adjustable RNA vaccines, e.g. influenza vaccines

Summarizing the above, there remains an unmet need for a system that allows for purification and/or analysis of RNA mixtures.

Therefore, it is the object of the underlying invention to provide a system for purification and/or analysis of RNA, in particular of an RNA mixture. It is further a preferred object of the present invention to provide a system for (co-)purification and/or (co-)analysis of individual RNA species in an RNA mixture comprising multiple RNA species.

This object is solved by the claimed subject matter.

Description of the invention

The present application is filed together with a sequence listing in electronic format, which is part of the description of the present application. The information contained in the electronic format of the sequence listing filed together with this application is incorporated herein by reference in its entirety. Where reference is made herein to a 'SEQ ID NO:' the corresponding nucleic acid sequence or amino acid sequence in the sequence listing having the respective identifier is referred to, unless stated otherwise.

For the sake of clarity and readability the following definitions are provided. Any technical feature mentioned for these definitions may be read on each and every embodiment of the invention. Additional definitions and explanations may be specifically provided in the context of these embodiments.

Definitions:

Adaptive immune response: The term "adaptive immune response" as used herein will be recognized and understood by the person of ordinary skill in the art, and is, for example, intended to refer to an antigen-specific response of the immune system. Antigen specificity allows for the generation of responses that are tailored to specific pathogens or pathogen-infected cells. The ability to mount these tailored responses is usually maintained in the body by "memory cells" (B-cells). In the context of the invention, the antigen is provided by the RNA coding region encoding at least one antigenic peptide, or protein.

Adaptive immune system: The adaptive immune system is essentially dedicated to eliminate or prevent pathogenic growth. It typically regulates the adaptive immune response by providing the vertebrate immune system with the ability to recognize and remember specific pathogens (to generate immunity), and to mount stronger attacks each time the pathogen is encountered.

Adjuvant/adjuvant component: An adjuvant or an adjuvant component in the broadest sense is typically a pharmacological and/or immunological agent that may enhance the effect of other agents, such as a drug or vaccine. It is to be interpreted in a broad sense and refers to a broad spectrum of substances. Typically, these substances are able to increase the immunogenicity of antigens. For example, adjuvants may be recognized by the innate immune systems and, e.g., may elicit an innate immune response. "Adjuvants" typically do not elicit an adaptive immune response. Insofar, "adjuvants" do not qualify as antigens. Their mode of action is distinct from the effects triggered by antigens resulting in an adaptive immune response.

Antigen: In the context of the present invention "antigen" refers typically to a substance which may be recognized by the immune system, preferably by the adaptive immune system, and is capable of triggering an antigen-specific immune response, e.g. by formation of antibodies and/or antigen-specific T cells as part of an adaptive immune response. Typically, an antigen may be or may comprise a peptide or protein which may be presented by the MHC to T-cells. In the sense of the present invention an antigen may be the product of translation of a provided RNA as defined herein. In this context, also fragments, variants and derivatives of peptides and proteins comprising at least one epitope are understood as antigens. In the context of the present invention, tumor antigens and pathogenic antigens as defined herein are particularly preferred.

Bicistronic RNA, multicistronic RNA: A bicistronic or multicistronic RNA is typically an RNA, preferably an mRNA, that typically may have two (bicistronic) or more (multicistronic) coding regions. A coding region in this context is a sequence of codons that is translatable into a peptide or protein.

Carrier / polymeric carrier: A carrier in the context of the invention may typically be a compound that facilitates transport and/or complexation of another compound (cargo). A polymeric carrier is typically a carrier that is formed of a polymer. A carrier may be associated to its cargo by covalent or non-covalent interaction. A carrier may transport nucleic acids, e.g. RNA or DNA, to the target cells. The carrier may – for some embodiments – be a cationic component.

Cationic component: The term "cationic component" typically refers to a charged molecule, which is positively charged (cation) at a pH value typically from 1 to 9, preferably at a pH value of or below 9 (e.g. from 5 to 9), of or below 8 (e.g. from 5 to 8), of or below 7 (e.g. from 5 to 7), most preferably at a physiological pH, e.g. from 7.3 to 7.4. Accordingly, a cationic component may be any positively charged compound or polymer, preferably a cationic peptide or protein which is positively charged under physiological conditions, particularly under physiological conditions in vivo. A "cationic peptide or protein" may contain at least one positively charged amino acid, or more than one positively charged amino acid, e.g. selected from Arg, His, Lys or Orn. Accordingly, "polycationic" components are also within the scope exhibiting more than one positive charge under the conditions given.

Coding region: A coding region, in the context of the invention, is typically a sequence of several nucleotide triplets, which may be translated into a peptide or protein. A coding region preferably contains a start codon, i.e. a combination of three subsequent nucleotides coding usually for the amino acid methionine (ATG), at its 5'-end and a subsequent region which usually exhibits a length which is a multiple of 3 nucleotides. A coding region is preferably terminated by a stop-codon (e.g., TAA, TAG, TGA). Typically, this is the only stop-codon of the coding region. Thus, a coding region in the context of the present invention is preferably a nucleotide sequence, consisting of a number of nucleotides that may be divided by three, which starts with a start codon (e.g. ATG) and which preferably terminates with a stop codon (e.g., TAA, TGA, or TAG). The coding region may be isolated or it may be incorporated in a longer nucleic acid sequence, for example in a vector or an mRNA. In the context of the present invention, a coding region may also be termed "protein coding region", "coding sequence", "CDS", "open reading frame" or "ORF".

Fragment of a sequence: A fragment of a sequence may typically be a shorter portion of a full-length sequence of e.g. a nucleic acid molecule or an amino acid sequence. Accordingly, a fragment, typically, comprises or consists of a sequence that is identical to the corresponding stretch within the full-length sequence. A preferred fragment of a sequence in the context of the present invention, comprises or consists of a continuous stretch of entities, such as nucleotides or amino acids corresponding to a continuous stretch of entities in the molecule the fragment is derived from, which represents at least 5%, 10%, 20%, preferably at least 30%, more preferably at least 40%, more preferably at least 50%, even more preferably at least 60%, even more preferably at least 70%, and most preferably at least 80% of the total (i.e. full-length) molecule from which the fragment is derived. Preferably, a fragment of a sequence as used herein is at least 5%, 10%, 20%, 30%, 40%, 50%, 60%, preferably at least 70%,

more preferably at least 80%, even more preferably at least 90%, even more preferably at least 95%, most preferably at least 99%, identical to a sequence, from which it is derived. Preferably, a fragment as used herein has the same biological function or specific activity compared to the full-length molecule.

Heterologous sequence: Two sequences are typically understood to be 'heterologous' if they are not derivable from the same gene or in the same allele. I.e., although heterologous sequences may be derivable from the same organism, they naturally (in nature) do not occur in the same nucleic acid molecule, such as in the same mRNA.

Homolog (of a nucleic acid sequence/amino acid sequence): The term "homolog" typically refers to a sequence of the same or of another species that is related, but preferably not identical, to a reference sequence. The term "homolog" encompasses orthologs as well as paralogs. In this context, "orthologs" are proteins encoded by genes in different species that evolved from a common ancestral gene by speciation. Orthologs often retain the same function(s) in the course of evolution. Thus, functions may be lost or gained when comparing a pair of orthologs. "Paralogs" are genes produced via gene duplication within a genome. Paralogs may also evolve new functions or eventually become pseudogenes. In the context of the present invention, a homolog of a nucleic acid sequence or of an amino acid sequence is preferably at least 5%, 10%, 20%, 30%, 40%, 50%, 60%, more preferably at least 70%, even more preferably at least 80%, even more preferably at least 90%, even more preferably at least 95%, most preferably at least 99%, identical to a reference sequence. It is further preferred that a "homolog" as used herein consists of a continuous stretch of entities, such as nucleotides or amino acid residues, corresponding to a continuous stretch of entities in the reference molecule, which represents at least 5%, 10%, 20%, preferably at least 30%, more preferably at least 40%, more preferably at least 50%, even more preferably at least 60%, even more preferably at least 70%, and most preferably at least 80% of the total (i.e. full-length) reference molecule.

Immunogen: In the context of the present invention an immunogen may be typically understood to be a compound that is able to stimulate an immune response. Preferably, an immunogen is a peptide, polypeptide, or protein. In a particularly preferred embodiment, an immunogen in the sense of the present invention is the product of translation of a provided nucleic acid molecule, preferably an RNA molecule as defined herein. Typically, an immunogen elicits at least an adaptive immune response.

Immune response: An immune response may typically be a specific reaction of the adaptive immune system to a particular antigen (so called specific or adaptive immune response) or an unspecific reaction of the innate immune system (so called unspecific or innate immune response), or a combination thereof.

Immune system: The immune system may protect organisms from infection. If a pathogen succeeds in passing a physical barrier of an organism and enters this organism, the innate immune system provides an immediate, but non-specific response. If pathogens evade this innate response, vertebrates possess a second layer of protection, the adaptive immune system. Here, the immune system adapts its response during an infection to improve its recognition of the pathogen. This improved response is then retained after the pathogen has been eliminated, in the form of an immunological memory, and allows the adaptive immune system to mount faster and stronger attacks each time this pathogen is encountered. According to this, the immune system comprises the innate and the adaptive immune system. Each of these two parts typically contains so called humoral and cellular components.

Nucleic acid sequence / amino acid sequence: The sequence of a nucleic acid molecule is typically understood to be the particular and individual order, i.e. the succession of its nucleotides. The sequence of a protein or peptide is typically understood to be the order, i.e. the succession of its amino acid residues.

Peptide: A peptide or polypeptide is typically a polymer of amino acid monomers, linked by peptide bonds. A peptide typically contains less than 50 monomer units. Nevertheless, the term peptide is not a disclaimer for molecules having more than 50 monomer units. Long peptides are also called polypeptides, typically having between 50 and 600 monomeric units. The term 'polypeptide' as used herein, however, is typically not limited by the length of the molecule it refers to. In the context of the present invention, the term 'polypeptide' may also be used with respect to peptides comprising less than 50 (e.g. 10) amino acids or peptides comprising even more than 600 amino acids.

Polyvalent/polyvalent vaccine: A polyvalent vaccine (also referred to as 'multivalent vaccine') typically contains antigens (or fragments or variants thereof) from more than one strain of a virus, or different antigens (or fragments or variants thereof) of the same virus, or any combination thereof. The term "polyvalent vaccine" describes that this vaccine has more than one valence. In the context of the invention, a polyvalent vaccine (e.g. Influenza vaccine, Norovirus vaccine) preferably comprises a vaccine comprising nucleic acids encoding antigenic peptides or proteins derived from several (genetically) different viruses or comprising nucleic acids encoding different antigens from (genetically) the same virus, or a combination thereof. In a preferred embodiment, a polyvalent vaccine comprises 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100 or even more different RNAs or RNA species, preferably as described herein, each encoding at least one different antigenic peptide or protein. Methods to produce polyvalent mRNA vaccines are disclosed in the PCT application WO20171090134.

Therapeutically effective amount: A therapeutically effective amount in the context of the invention is typically understood to be an amount that is sufficient to induce a pharmaceutical effect, such as an immune response, altering a pathological level of an expressed peptide or protein, or substituting a lacking gene product, e.g., in case of a pathological situation.

Protein: A protein typically comprises one or more peptides or polypeptides. A protein is typically folded into 3-dimensional form, which may be required for the protein to exert its biological function.

RNA, mRNA: RNA is the usual abbreviation for ribonucleic acid. It is a nucleic acid molecule, i.e. a polymer consisting of nucleotides. These nucleotides are usually adenosine-monophosphate, uridine-monophosphate, guanosine-monophosphate and cytidine-monophosphate monomers which are connected to each other along a so-called backbone. The backbone is formed by phosphodiester bonds between the sugar, i.e. ribose, of a first and a phosphate moiety of a second, adjacent monomer. The specific succession of the monomers is called the RNA-sequence. Usually RNA may be obtainable by transcription of a DNA-sequence, e.g., inside a cell. In eukaryotic cells, transcription is typically performed inside the nucleus or the mitochondria. Typically, transcription of DNA

usually results in the so-called premature RNA which has to be processed into so-called messenger-RNA, usually abbreviated as mRNA. Processing of the premature RNA, e.g. in eukaryotic organisms, comprises a variety of different posttranscriptional-modifications such as splicing, 5'-capping, polyadenylation, export from the nucleus or the mitochondria and the like. The sum of these processes is also called maturation of RNA. The mature messenger RNA usually provides the nucleotide sequence that may be translated into an amino-acid sequence of a particular peptide or protein. Typically, a mature mRNA comprises a 5'-cap, a 5'-UTR, a coding region, a 3'-UTR and a poly(A) sequence. Aside from messenger RNA, several non-coding types of RNA exist which may be involved in regulation of transcription and/or translation.

Sequence identity, identity (of a sequence): Two or more sequences are identical if they exhibit the same length and order of nucleotides or amino acids. The percentage of identity typically describes the extent, to which two sequences are identical, i.e. it typically describes the percentage of nucleotides that correspond in their sequence position with identical nucleotides of a reference sequence. In order to determine the degree of identity, the sequences to be compared are considered to exhibit the same length, i.e. the length of the longest sequence of the sequences to be compared. This means that a first sequence consisting of 8 nucleotides is 80% identical to a second sequence consisting of 10 nucleotides comprising the first sequence. Hence, in the context of the present invention, identity of sequences preferably relates to the percentage of nucleotides of a sequence which have the same position in two or more sequences having the same length. Therefore, e.g. a position of a first sequence may be compared with the corresponding position of the second sequence. If a position in the first sequence is occupied by the same component (residue) as is the case at a position in the second sequence, the two sequences are identical at this position. If this is not the case, the sequences differ at this position. If insertions occur in the second sequence in comparison to the first sequence, gaps can be inserted into the first sequence to allow a further alignment. If deletions occur in the second sequence in comparison to the first sequence, gaps can be inserted into the second sequence to allow a further alignment. The percentage to which two sequences are identical is then a function of the number of identical positions divided by the total number of positions including those positions which are only occupied in one sequence. The percentage to which two sequences are identical can be determined using a mathematical algorithm. A preferred, but not limiting, example of a mathematical algorithm which can be used is the algorithm of Karlin et al. (1993), PNAS USA, 90:5873-5877 or Altschul et al. (1997), Nucleic Acids Res., 25:3389-3402. Such an algorithm is integrated in the BLAST program. Sequences which are identical to the sequences of the present invention to a certain extent can be identified by this program.

Transfection: The term "transfection" refers to the introduction of nucleic acid molecules, such as DNA or RNA (e.g. mRNA) molecules, into cells, preferably into eukaryotic cells. In the context of the present invention, the term "transfection" encompasses any method known to the skilled person for introducing nucleic acid molecules into cells, preferably into eukaryotic cells, such as into mammalian cells. Such methods encompass, for example, electroporation, lipofection, e.g. based on cationic lipids and/or liposomes, calcium phosphate precipitation, nanoparticle based transfection, virus based transfection, or transfection based on cationic polymers, such as DEAE-dextran or polyethylenimine etc. Preferably, the introduction is non-viral.

Vaccine: A vaccine is typically understood to be a prophylactic or therapeutic material providing at least one antigen, preferably an immunogen. The antigen or immunogen may be derived from any material that is suitable for

vaccination. For example, the antigen or immunogen may be derived from a pathogen, such as from bacteria or virus particles etc., or from a tumor or cancerous tissue. The antigen or immunogen stimulates the body's adaptive immune system to provide an adaptive immune response.

Vector: The term "vector" refers to a nucleic acid molecule, preferably to an artificial nucleic acid molecule. A vector in the context of the present invention is suitable for incorporating or harboring a desired nucleic acid sequence, such as a nucleic acid sequence comprising a coding region. Such vectors may be storage vectors, expression vectors, cloning vectors, transfer vectors etc. A storage vector is a vector which allows the convenient storage of a nucleic acid molecule, for example, of an mRNA molecule. Thus, the vector may comprise a sequence corresponding, e.g., to a desired mRNA sequence or a part thereof, such as a sequence corresponding to the coding region and the 3'-UTR and/or the 5'-UTR of an mRNA. An expression vector may be used for production of expression products such as RNA, e.g. mRNA, or peptides, polypeptides or proteins. For example, an expression vector may comprise sequences needed for transcription of a sequence stretch of the vector, such as a promoter sequence, e.g. an RNA polymerase promoter sequence. A cloning vector is typically a vector that contains a cloning site, which may be used to incorporate nucleic acid sequences into the vector. A cloning vector may be, e.g., a plasmid vector or a bacteriophage vector. A transfer vector may be a vector which is suitable for transferring nucleic acid molecules into cells or organisms, for example, viral vectors. A vector in the context of the present invention may be, e.g., an RNA vector or a DNA vector. Preferably, a vector is a DNA molecule. Preferably, a vector in the sense of the present application comprises a cloning site, a selection marker, such as an antibiotic resistance factor, and a sequence suitable for multiplication of the vector, such as an origin of replication. Preferably, a vector in the context of the present application is a plasmid vector.

Vehicle: A vehicle is typically understood to be a material that is suitable for storing, transporting, and/or administering a compound, such as a pharmaceutically active compound. For example, it may be a physiologically acceptable liquid which is suitable for storing, transporting, and/or administering a pharmaceutically active compound.

Variant of a sequence: A variant of a nucleic acid sequence or an amino acid sequence typically differs from the original sequence in one or more residues, such as one or more substituted, inserted and/or deleted nucleotides or amino acid residues. Preferably, these variants have the same biological function or specific activity compared to the full-length molecule. In the context of the present invention, a variant of a nucleic acid sequence or of an amino acid sequence is preferably at least 5%, 10%, 20%, 30%, 40%, 50%, 60%, more preferably at least 70%, even more preferably at least 80%, even more preferably at least 90%, even more preferably at least 95%, most preferably at least 99%, identical to a reference sequence. It is further preferred that a "variant" as used herein comprises or consists of a continuous stretch of entities, such as nucleotides or amino acid residues, corresponding to a continuous stretch of entities in the reference molecule, which represents at least 5%, 10%, 20%, 30%, 40%, 50%, 60%, preferably at least 70%, more preferably at least 80%, even more preferably at least 90%, even more preferably at least 95%, most preferably at least 99%, of the total (i.e. full-length) reference molecule.

Method for modifying the retention time of an RNA on a chromatographic column:

One aspect of the present invention concerns a method for modifying the retention time of RNA on a chromatographic column. In particular, the invention provides a method for purifying RNA, such as chromatographic purification of RNA from a mixture of RNA species. By modifying the retention time of RNA, the inventive method preferably allows for physical separation of said RNA.

According to preferred embodiments, the invention relates to a method for modifying the retention time of an RNA on a chromatographic column, wherein the method comprises a step of adapting the RNA (sequence) by altering the number of adenine (A) and/or uracil (U) nucleotides in the RNA sequence with respect to the number of A and/or U nucleotides in the original RNA sequence.

The inventors surprisingly found that the total number of A and U nucleotides in an RNA influences the retention time of an RNA on a chromatographic column in a chromatographic process, such as high pressure (high pressure) liquid chromatography (HPLC), more preferably reversed-phase HPLC.

As used herein, the term 'retention time' typically refers to the amount of time a compound, such as RNA, spends on a chromatographic column, after it has been injected. If two compounds in a sample, e.g. two RNA species in an RNA mixture, have different retention times, each compound will spend a different amount of time on the chromatographic column. Retention times are usually quoted in units of seconds or minutes.

In this context, it was also found by the inventors that the separation factor of a given chromatographic method can be modulated by altering the number of adenine (A) and/or uracil (U) nucleotides in an RNA sequence. An increased difference between at least two RNA species in the number of A and/or U nucleotides was found to be correlated with an increased separation factor (cf. Figure 16). The inventive method may therefore be employed for increasing the separation factor of a chromatographic method by modifying the number of adenine (A) and/or uracil (U) nucleotides in an RNA sequence. A decreased difference between at least two RNA species in the number of A and/or U nucleotides is correlated with a decreased separation factor. The inventive method may therefore be employed for decreasing the separation factor of a chromatographic method by modifying the number of adenine (A) and/or uracil (U) nucleotides in an RNA sequence.

As used herein, the term "RNA species" typically relates to a plurality of RNA molecules or (at least one RNA molecule) having the same nucleotide sequence.

Preferably, the retention time of an RNA species is the time required for said RNA species to migrate, or elute, from the column, measured from the instant the sample is injected into the mobile phase stream to the point at which the peak maximum occurs. Alternatively, the retention time may also be determined by measuring the time from the instant the sample is injected to the appearance of the first unretained RNA at the outlet. In preferred embodiments, the dependence of the retention time on a given flow rate is removed by calculating the corresponding retention volume, which is typically calculated as the retention time multiplied by the volumetric flow rate of the mobile phase. In preferred embodiments, the retention volume may thus be modified by the method according to the invention.

The retention time of RNA is modified by the inventive method, wherein the retention time is preferably increased or decreased. In particular, the retention time of an original RNA (sequence) is modified by altering the number of A and/or U nucleotides in said original RNA (sequence), thereby obtaining an adapted RNA (sequence) having a modified retention time. In this context, the term 'original RNA (sequence)' may refer to any RNA (sequence), which retention time may be modified by the method as described herein. Hence, an 'original RNA (sequence)' may be a wild type RNA sequence. Furthermore, also a modified RNA sequence, preferably an optimized RNA sequence, more preferably as described herein (e.g. an RNA sequence, which has been modified (optimized) with respect to its G/C content), may be used as an 'original RNA (sequence)' in the meaning of the present invention. The term 'adapted RNA (sequence)' as used herein preferably refers to an RNA (sequence), which is derived from the original RNA (sequence), from which it differs by the total number of A and/or U nucleotides. In other words, an adapted RNA (sequence) may be obtained by the inventive method, in particular by altering the number of A and/or U nucleotides in an original RNA (sequence) and thus preferably modifying the RNA's retention time. In a preferred embodiment, an adapted RNA (sequence) is obtained by the inventive method, which differs from the original RNA (sequence) only in the number of A and/or U nucleotides or in the corresponding sequence changes at the respective nucleotide positions.

The retention time of an original RNA is thus typically modified by the inventive method, wherein the retention time of the adapted RNA is preferably modified, more preferably increased or decreased, with respect to the retention time of the original RNA.

As used herein, the term 'RNA (sequence)' may refer to an RNA as an individual compound or to an RNA species, preferably as described herein, or to the nucleotide sequence of a given RNA or RNA species. In the context of the present invention, the term 'RNA species' typically relates to an RNA defined by a certain nucleotide sequence. A plurality of RNA molecules having the same nucleotide sequence may also be referred to herein as 'RNA species'.

In particular, the term 'RNA species' as used herein may refer to a first RNA species, such as a first plurality of RNA molecules sharing the same nucleotide sequence (or to a population of RNA molecules sharing the same nucleotide sequence), wherein said RNA species is preferably present in a mixture with a second (or further) RNA species, such as with a second (or further) plurality of RNA molecules sharing a nucleotide sequence, which is distinct from the nucleotide sequence of the first RNA species (or with a second (or further) population of RNA molecules sharing a nucleotide sequence, which is distinct from the nucleotide sequence of the first RNA species).

The term 'RNA' as used herein may refer to any type of RNA without limitation. Preferably, the term RNA refers to a molecule or to a molecule species selected from the group consisting of long-chain RNA, coding RNA, non-coding RNA, single stranded RNA (ssRNA), double stranded RNA (dsRNA), linear RNA (linRNA), circular RNA (circRNA), messenger RNA (mRNA), RNA oligonucleotides, small interfering RNA (siRNA), small hairpin RNA (shRNA), antisense RNA (asRNA), CRISPR/Cas9 guide RNAs, riboswitches, immunostimulating RNA (isRNA), ribozymes, aptamers, ribosomal RNA (rRNA), transfer RNA (tRNA), viral RNA (vRNA), retroviral RNA or replicon RNA, small nuclear RNA (snRNA), small nucleolar RNA (snoRNA), microRNA (miRNA), circular RNA (circRNA), and a Piwi-interacting RNA (piRNA). Preferably, the RNA is a coding RNA. According to a particularly preferred embodiment, the method is

used for modifying the retention time of a coding RNA, preferably an mRNA, more preferably as described herein. According to preferred embodiments, the coding RNA, preferably the mRNA, is a long-chain RNA as described herein.

The method according to the invention is preferably used for modifying the retention time of a single-stranded RNA (ssRNA) or a double-stranded RNA (dsRNA), more preferably of an ssRNA. As used herein, the term 'single-stranded RNA' or 'ssRNA' typically refers to an RNA molecule consisting of a single RNA strand, which may optionally comprise (e.g. based on the conditions), secondary structure elements (e.g. hairpins, stem-loops, etc.). In the context of the present invention, the term 'double-stranded RNA' or 'dsRNA' typically refers to an RNA molecule comprising two RNA strands, which preferably form a (hetero)duplex.

Preferably, RNA as used herein comprises more than 30 nucleotides. More preferably, the RNA is not selected from siRNA, small hairpin RNA, microRNA or small nuclear RNA (snRNA). In a particularly preferred embodiment, the RNA as used herein is not siRNA.

According to a preferred embodiment, the method is used for modifying the retention time of a long-chain RNA. The term 'long-chain RNA' as used herein herein typically refers to an RNA molecule, preferably as described herein, which preferably comprises at least 30 nucleotides. Alternatively, a long-chain RNA may comprise at least 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750 or at least 800 nucleotides. A long-chain RNA molecule may further comprise at least 100 nucleotides, even more preferably at least 200 nucleotides. A long-chain RNA, in the context of the present invention, further preferably comprises from 30 to 50.000 nucleotides, from 30 to 20000 nucleotides, from 100 to 20000 nucleotides, from 200 to 20000 nucleotides, from 200 to 15000 nucleotides, from 500 to 20000 nucleotides, from 800 to 20000 nucleotides, from 800 to 5000 nucleotides or from 800 to 2000 nucleotides. The term 'long-chain RNA' as used herein is not limited to a certain type of RNA, but merely refers to the number of nucleotides comprised in said RNA. For example, the long-chain RNA may be a coding RNA, preferably an mRNA, more preferably as described herein.

According to preferred embodiments of the invention, the number of A and/or U nucleotides in the original RNA sequence is determined. In the context of the present invention, the phrase 'number of A and/or U nucleotides' may refer to the total number of adenine (A) nucleotides, to the total number of uracil (U) nucleotides or, more preferably, to the total number of A and U nucleotides. The number of A and/or U nucleotides is preferably increased or decreased to a target number, which may be a known number or a number, which has been pre-determined, preferably as described herein. That target number is preferably selected according to the respective purpose. The modification of the retention time is typically the greater the larger the difference is between the target number and the original number of A and/or U nucleotides in the original RNA sequence.

Accordingly, the inventive method comprises a step of determining the number of A and/or U nucleotides in the original RNA sequence and a step of increasing or decreasing said number of A and/or U nucleotides in order to obtain the adapted RNA sequence.

As used herein, the term 'adenine nucleotide' or 'A nucleotide' typically refers to an adenine nucleotide or to an analogue thereof, such as a chemically modified A nucleotide, preferably as described herein. The term 'uracil nucleotide' or 'U nucleotide' may refer to an uracil nucleotide or to an analogue thereof, such as a chemically modified U nucleotide, preferably as described herein. In this context, the phrase 'number of A and/or U nucleotides' may also refer to the total number of adenine (A) nucleotides and adenine nucleotide analogues, to the total number of uracil (U) nucleotides and uracil nucleotide analogues, or, more preferably, to the total number of A nucleotides, adenine nucleotide analogues, U nucleotides and U nucleotide analogues. In a similar manner, the terms 'guanine nucleotide' or 'G nucleotide' may refer to a guanine nucleotide or to an analogue thereof, such as a chemically modified G nucleotide, preferably as described herein. The term 'cytosine nucleotide' or 'C nucleotide' may refer to a cytosine nucleotide or to an analogue thereof, such as a chemically modified C nucleotide, preferably as described herein.

Wherever reference is made herein to an A, U, G, or C nucleotide, the respective nucleotide analogues, preferably as described herein, more preferably a chemically modified nucleotide, is also comprised by that term.

Where reference is made herein to an 'uracil nucleotide', an 'U nucleotide' or an analogue of an uracil nucleotide, it is further understood that these terms may also relate to a thymine nucleotide in the DNA sequence corresponding to an RNA sequence (e.g. in a DNA vector encoding an RNA as described herein).

The length of the original RNA is preferably not modified so that the total number of nucleotides in the original RNA and the adapted RNA is essentially the same. Alternatively, the length of the original RNA may also be increased or decreased, e.g. by addition or deletion of nucleotides.

In preferred embodiments, G and/or C nucleotides in the original RNA sequence are replaced with A and/or U nucleotides thereby altering, particularly increasing the number of A and/or U nucleotides in the RNA sequence with respect to the number of A and/or U nucleotides in the original RNA sequence.

According to some embodiments, the inventive method comprises a step of replacing G and/or C nucleotides in the original RNA sequence with A and/or U nucleotides.

Alternatively, A and/or U nucleotides in the original RNA sequence may be replaced with G and/or C nucleotides thereby altering, particularly decreasing the number of A and/or U nucleotides in the RNA sequence with respect to the number of A and/or U nucleotides in the original RNA sequence.

According to some embodiments, the inventive method comprises a step of replacing A and/or U nucleotides in the original RNA sequence with G and/or C nucleotides.

According to some embodiments, the method may also comprise a step of introducing additional nucleotides, preferably G and/or C nucleotides, more preferably A and/or U nucleotides, into an RNA sequence. Therein, nucleotides are preferably inserted into a non-coding region of the original RNA sequence.

In preferred embodiments, the invention concerns a method for modifying the retention time of an RNA comprising at least one coding region. In these embodiments, the number of A and/or U nucleotides is preferably altered within a coding region, preferably as described herein. For example, G and/or C nucleotides in a coding region are replaced with A and/or U nucleotides in order to increase the number of A and/or U nucleotides in the coding region. Alternatively, A and/or U nucleotides in a coding region may be replaced with G and/or C nucleotides in order to decrease the number of A and/or U nucleotides in the coding region.

In a particularly preferred embodiment, the number of A and/or U nucleotides in at least one coding region in an RNA is altered without modifying the amino acid sequence encoded by the coding sequence of the original RNA (sequence). The degeneracy of the genetic code allows for exchanges of codons in the coding region, which do not alter the encoded amino acid sequence. The method according to the invention thus preferably comprises exchanging at least one codon in a coding region of the RNA with an alternative codon encoding the same amino acid, wherein the alternative codon is preferably characterized by a number of A and/or U nucleotides that is different from the number of A and/or U numbers in the original codon. Reference is made to the description herein of the inventive method for providing an adapted RNA sequence, where the selection of suitable codon exchanges is explained in more detail.

The number of A and/or U nucleotides is preferably increased or decreased to such an extent that the retention time is sufficiently modified for a given purpose. Depending on the purpose, the number of A and/or U nucleotides may be adapted to a pre-determined target number. For example, the number of A and/or U nucleotides in a given (original) RNA sequence may be adapted to the number or essentially the same number of A and/or U nucleotides in the sequence of another RNA species, which is also present in a mixture and which is to be co-purified with the RNA to be adapted. RNA species having essentially the same number of A and/or U nucleotides are typically characterized by essentially the same retention time on a chromatographic column, such as an HPLC column. According these RNA species elute from the column in the same fraction (peak). Alternatively, the target number may be selected such that the difference from the number of A and/or U nucleotides in another RNA species in a mixture is sufficient for the adapted RNA to elute in a separate fraction (peak) (which allows for determining the integrity of each RNA species).

In preferred embodiments, the number of A and/or U nucleotides in an RNA sequence is altered so that the number of A and/or U nucleotides in the adapted RNA sequence differs by at least 1, preferably by at least 2, more preferably by at least 3, even more preferably by at least 4, even more preferably by at least 5, even more preferably by at least 10, most preferably by at least 15, from the number of A and/or U nucleotides in the original RNA sequence. Alternatively, the number of A and/or U nucleotides in the adapted RNA sequence may differ by at least 5, preferably by at least 10, more preferably by at least 20, even more preferably by at least 30, even more preferably by at least 40, even more preferably by at least 50, most preferably by at least 60, from the number of A and/or U nucleotides in the original RNA sequence. Alternatively, the number of A and/or U nucleotides in the adapted RNA sequence may differ by at least 60, preferably by at least 80, more preferably by at least 100, even more preferably by at least 120, even more preferably by at least 140, even more preferably by at least 160, most preferably by at least 200, from the number of A and/or U nucleotides in the original RNA sequence.

According to preferred embodiments, the method is used for increasing the retention time of the adapted RNA with respect to the retention time of the original RNA, wherein the number of A and/or U nucleotides in the adapted RNA sequence is increased by at least 1, preferably by at least 2, more preferably by at least 3, even more preferably by at least 4, even more preferably by at least 5, even more preferably by at least 10, most preferably by at least 15, with respect to the number of A and/or U nucleotides in the original RNA sequence. Alternatively, the number of A and/or U nucleotides in the adapted RNA sequence may be increased by at least 5, preferably by at least 10, more preferably by at least 20, even more preferably by at least 30, even more preferably by at least 40, even more preferably by at least 50, most preferably by at least 60, with respect to the number of A and/or U nucleotides in the original RNA sequence. Alternatively, the number of A and/or U nucleotides in the adapted RNA sequence may be increased by at least 60, preferably by at least 80, more preferably by at least 100, even more preferably by at least 120, even more preferably by at least 140, even more preferably by at least 160, most preferably by at least 200, with respect to the number of A and/or U nucleotides in the original RNA sequence.

The method may also be used for decreasing the retention time of the adapted RNA with respect to the retention time of the original RNA, wherein the number of A and/or U nucleotides in the adapted RNA sequence is decreased by at least 1, preferably by at least 2, more preferably by at least 3, even more preferably by at least 4, even more preferably by at least 5, even more preferably by at least 10, most preferably by at least 15, with respect to the number of A and/or U nucleotides in the original RNA sequence. Alternatively, the number of A and/or U nucleotides in the adapted RNA sequence may be decreased by at least 5, preferably by at least 10, more preferably by at least 20, even more preferably by at least 30, even more preferably by at least 40, even more preferably by at least 50, most preferably by at least 60, with respect to the number of A and/or U nucleotides in the original RNA sequence. Alternatively, the number of A and/or U nucleotides in the adapted RNA sequence may be decreased by at least 60, preferably by at least 80, more preferably by at least 100, even more preferably by at least 120, even more preferably by at least 140, even more preferably by at least 160, most preferably by at least 200, with respect to the number of A and/or U nucleotides in the original RNA sequence.

In some embodiments, the method comprises a step of adapting the original RNA sequence by altering the number of A and/or U nucleotides in such a manner that the ratio of the number of A nucleotides to the number of U nucleotides is from 0.2 to 5, preferably from 0.5 to 3, more preferably from 1 to 3, even more preferably from 1 to 2.5, even more preferably from 1.2 to 2, even more preferably from 1.4 to 2, even more preferably from 1.5 to 2, most preferably from 1.6 to 2.

According to some embodiments, the retention time of RNA on a chromatographic column is modified, wherein altering the number of A nucleotides in the RNA sequence has a stronger impact on the retention time of the RNA than altering the number of U nucleotides of the RNA sequence. Preferably, the change in retention time achieved by altering the number of A nucleotides is about 1.2, 1.3, 1.4, 1.5, 1.6, 1.7, 1.8, 1.9 or 2 times, preferably 1.5 times, stronger than the change in retention time achieved by altering the same number of U nucleotides.

In some embodiments, the ratio of the number of A nucleotides to the number of U nucleotides in the RNA sequence is increased (without changing the total number of A and U nucleotides) with respect to the ratio of the number of

A nucleotides to the number of U nucleotides in the original RNA sequence, thereby increasing the retention time of the adapted RNA.

Alternatively, the ratio of the number of A nucleotides to the number of U nucleotides in the RNA sequence is decreased (without changing the total number of A and U nucleotides) with respect to the ratio of the number of A nucleotides to the number of U nucleotides in the original RNA sequence, thereby decreasing the retention time of the adapted RNA.

Each embodiment of the inventive method, which comprises a step of altering the number of A and/or U nucleotides in the RNA sequence in any manner that leads to a desired modification of the retention time (decrease or increase) may be used as individual step or in any combination thereof.

Accordingly, as defined herein, an increase of the retention time of the RNA may preferably be obtained by replacing G and/or C nucleotides in the original RNA sequence, preferably in the coding region, with A and/or U nucleotides and/or by introducing additional A and/or U nucleotides into the original RNA sequence, preferably into an UTR region, and/or by increasing the ratio of the number of A nucleotides to the number of U nucleotides (without changing the total number of A and U nucleotides) in the RNA sequence with respect to the ratio of the number of A nucleotides to the number of U nucleotides in the original RNA sequence, wherein all of these adaptations are suitably applied without modifying the amino acid sequence encoded by the coding region of the original RNA (sequence).

Accordingly, as defined herein, a decrease of the retention time of the RNA may be obtained by replacing A and/or U nucleotides in the original RNA sequence, preferably in the coding region, with G and/or C nucleotides and/or by introducing additional G and/or C nucleotides into the original RNA sequence, preferably into an UTR region, and/or by decreasing the ratio of the number of A nucleotides to the number of U nucleotides (without changing the total number of A and U nucleotides) in the RNA sequence with respect to the ratio of the number of A nucleotides to the number of U nucleotides in the original RNA sequence, wherein all of these adaptations are suitably applied without modifying the amino acid sequence encoded by the coding region of the original RNA (sequence).

In preferred embodiments, the method comprises a step of adapting the original RNA sequence by altering the number of A and/or U nucleotides in the RNA sequence as described herein with respect to a method for purifying at least one RNA species from a mixture and/or as described herein with respect to a method for co-purifying at least two RNA species from a mixture.

The method according to the invention is preferably provided as a method for modifying the retention time of RNA on a chromatographic column, wherein the chromatographic column is an HPLC column. In this context, the method is preferably used for modifying the retention time of RNA under conditions suitable for analytical or preparative purification of RNA. Chromatographic procedures and experimental conditions suitable for purifying RNA on an analytical or on a preparative scale, in particular by RP-HPLC, are known in the art and were described, for example, in WO 2008/077592, which is hereby incorporated by reference in its entirety.

In preferred embodiments, the method is used for modifying the retention time of RNA on a chromatographic column in a reversed-phase HPLC (RP-HPLC). Preferably, the RP-HPLC is as described in WO 2008/077592. According to a preferred embodiment, the chromatographic column comprises a reversed phase as a stationary phase. In reversed-phase chromatography, a nonpolar compound is preferably used as the stationary phase and a polar solvent, such as an aqueous solution of e.g. acetonitrile and/or methanol, is used as the mobile phase for elution.

The column, which is used as stationary phase, may be provided in bead form or as a monolithic column, which is preferably a polymerized 'block', i.e. a block which fills a substantial part of the chromatography column. Irrespective of its precise nature, the polymeric stationary phase, in particular if used for preparative purification, is preferably porous in its nature, which means that the beads or the block are characterized by pores. In other embodiments, the reversed phase is a non-porous reversed phase.

In a preferred embodiment, in particular if used for preparative purification, a porous reversed phase material is provided with a particle size of 8.0 µm to 50 µm, in particular 8.0 to 30, still more preferably 8.0 to 25 µm. The reversed phase material may be present in the form of small beads. The method according to the invention may be performed particularly favourably with a porous reversed phase with this particle size, optionally in bead form, wherein particularly good separation results are obtained for preparative purification.

In a preferred embodiment, the reversed phase has a pore size of 1000 Å to 5000 Å, in particular a pore size of 1000 Å to 4000 Å, more preferably 1500 Å to 4000 Å, 2000 Å to 4000 Å or 2500 Å to 4000 Å. Particularly preferred pore sizes for the reversed phases are 1000 Å to 2000 Å, more preferably 1000 Å to 1500 Å and most preferably 1000 Å to 1200 Å or 3500-4500 Å. In other embodiments, the reversed phase is a porous reversed phase with undefined pore size.

A pore size of 1000 Å to 5000 Å, in particular a pore size of 1000 Å to 4000 Å, more preferably 1500 Å to 4000 Å, 2000 Å to 4000 Å or 2500 Å to 4000 Å may be suitable to separate an RNA from other components of a mixture, the RNA preferably having a size as mentioned herein with respect to the long-chain RNA, e.g. of up to about 15000 nucleotides (as single stranded RNA molecule) or base pairs (as double stranded RNA molecule), in particular 100 to 10000, more preferably 500 to 10000 nucleotides or base pairs, even more preferably 800 to 5000 nucleotides or base pairs and even more preferably 800 to 2000 nucleotides or base pairs. However, the pore size of the reversed phase may also be selected in dependence of the size of the RNA to be separated, i.e. a larger pore size may be selected, if larger RNA molecules are to be separated and smaller pore sizes may be selected, if smaller RNA molecules may be selected. This is due to the effect, that the retention of the RNA molecules and the separation not only depends on the interaction of the (reversed) phase but also on the possibility of molecules to get inside the pores of the matrix and thus provide a further retention effect. Without being limited thereto, e.g. a pore size for the reversed phase of about 2000 Å to about 5000 Å, more preferably of about 2500 to about 4000, most preferably of about 3500 to about 4500 Å, may thus be used to separate larger RNA molecules, e.g. RNA molecules of 100 to 10000, more preferably 500 to 10000 nucleotides or base pairs, even more preferably 800 to 5000 nucleotides or base pairs and even more preferably 800 to 2000 nucleotides or base pairs. Alternatively, without being limited thereto, a pore size of for the reversed phases of about 1000 Å to about 2500 Å, more preferably of

about 1000 Å to about 2000 Å, and most preferably of about 1000 Å to 1200 Å may be used to separate smaller RNA molecules, e.g. RNA molecules of about 30-1000, 50-1000 or 100-1000 or 20-200, 20-100, 20-50 or 20-30 nucleotides may also be separated in this way.

In general, any material known to be used as reverse phase stationary phase, in particular any polymeric material may be used for chromatographic method in the context of the present invention. The stationary phase may be composed of organic and/or inorganic material. Examples for polymers to be used for the present invention are (non-alkylated or alkylated) polystyrenes, (non-alkylated or alkylated) polystyrenedivinylbenzenes, silica gel, silica gel modified with non-polar residues, particularly silica gel modified with alkyl containing residues, more preferably with butyl-, octyl and/or octadecyl containing residues, silica gel modified with phenylic residues, polymethacrylates, etc. or other materials suitable e.g. for gel chromatography or other chromatographic methods as mentioned above, such as dextran, including e.g. Sephadex® and Sephacryl® materials, agarose, dextran/agarose mixtures, polyacrylamide, etc..

Preferably, the chromatographic column comprises a material selected from the group consisting of polystyrene, a non-alkylated polystyrene, an alkylated polystyrene, a polystyrenedivinylbenzene, a non-alkylated polystyrenedivinylbenzene, an alkylated polystyrenedivinylbenzene, a silica gel, a silica gel modified with non-polar residues, a silica gel modified with alkyl containing residues, selected from butyl-, octyl and/or octadecyl containing residues, a silica gel modified with phenylic residues, or a polymethacrylate.

In a particularly preferred embodiment, the chromatographic column comprises a material selected from the group consisting of a polystyrene polymer, a non-alkylated polystyrene polymer, an alkylated polystyrene polymer, a non-alkylated polystyrenedivinylbenzene polymer, an alkylated polystyrenedivinylbenzene, a silica gel, a silica gel modified with non-polar residues, particularly silica gel modified with alkyl containing residues, more preferably with butyl-, octyl and/or octadecyl containing residues, porous silica gel modified with phenylic residues, polymethacrylates. All these materials may be porous, preferably as described herein, or non-porous.

Stationary phases with polystyrenedivinylbenzene are known per se. The per se known polystyrenedivinyl-benzenes already used for HPLC methods, which are commercially obtainable, may be used for the chromatographic method in the context of the invention.

In preferred embodiments, a non-alkylated porous polystyrenedivinylbenzene is particularly preferred, which, without being limited thereto, may have in particular a particle size of 8.0 ± 1.5 µm, in particular 8.0 ± 0.5 µm, and a pore size of 1000- 1500 Å, in particular 1000-1200 Å or 3500-4500 Å.

In further embodiments, an alkylated, macro porous monolithic polystyrenedivinylbenzene is particularly preferred, which, without being limited thereto, may have a pore size distribution of about 0.1um–10µm, particularly about of 1um-10µm, more particularly of about 1um-6µm.

In further embodiments a porous silica gel is used. The porous silica gel may be prepared from tetraethoxysilane and bis(triethoxysilyl)ethane used in a 4:1 molar ratio and the porous silica gel is modified with an octadecyl carbon

chain. Such porous silica gel is commercially available, e.g. XBRIDGE™ OST C18 from Waters or AQUITY UPLC OST C18 from Waters.

The silica gel may have a particle size of 0.5 to 5 µm, preferably of 0.7 to 4 µm, more preferably of 1 to 3 µm, even more preferably of 1.5 to 2 µm and most preferably of 1.7 µm. The pore size of the porous silica gel may be 50 to 300Å, preferably 70 to 250Å, more preferably 100 to 200Å, even more preferably 120 to 170Å and most preferably it is 135Å.

In a preferred embodiment, a mixture of an aqueous solvent and an organic solvent is used in HPLC as the mobile phase for eluting the RNA. It is favourable for a buffer to be used as the aqueous solvent which has in particular a pH of 6.0-8.0, for example of about 7, for example 7.0. Preferably the buffer is triethylammonium acetate, more preferably a 0.02 M to 0.5 M, in particular 0.08 M to 0.12 M, even more preferably an about 0.1 M triethylammonium acetate buffer. In a preferred embodiment, the organic solvent which is used in the mobile phase is acetonitrile, methanol, ethanol, 1-propanol, 2-propanol and acetone or a mixture thereof, very particularly preferably acetonitrile. In a particularly preferred embodiment, the mobile phase is a mixture of 0.1 M triethylammonium acetate, pH 7, and acetonitrile.

Any one of the individual steps or features (or a combination of such steps or features) described herein with respect to the method for modifying the retention time of an RNA may also be applied with respect to any one of the other aspects of the present invention as described herein, in particular to the method for purifying at least one RNA species from a mixture of at least two RNA species, the method for co-purifying at least two RNA species from a mixture of at least two RNA species, the method for harmonizing the numbers of A and/or U nucleotides in the sequences of at least two RNA species, or the method for providing an adapted RNA as described herein.

Method for purifying at least one RNA species from a mixture:

In a further aspect, the present invention concerns a method for purifying at least one RNA species from a mixture of at least two RNA species. This method is preferably characterized by any one of the method steps or any one of the features (or a combination thereof) described herein with regard to other aspects of the invention, in particular with regard to the method for modifying the retention time of an RNA or the method for co-purifying at least two RNA species from a mixture comprising at least two RNA species.

In preferred embodiments, the method for purifying at least one RNA species from a mixture of at least two RNA species comprises:

a) a step of adapting the sequence of the at least one RNA species by altering the number of A and/or U nucleotides with respect to the number of A and/or U nucleotides in the original sequence; and

b) a chromatographic step, wherein the adapted RNA species is separated from at least one other RNA species.

In the context of the present invention, the term 'purifying' is understood to mean that a desired RNA (species) in a sample is separated and/or isolated from another compound, which is further present in a mixture, typically a solution. These other compounds may be, for example, impurities or other RNA species, from which the desired

RNA (species) is to be separated. Thus, after HPLC purification the RNA is present in a purer form than in the originally introduced RNA-containing sample prior to HPLC purification. Undesired constituents of RNA-containing samples which therefore need to be separated may in particular be impurities derived from an in vitro transcription reaction, such as degraded RNA fragments or fragments, which have arisen as a result of premature termination of transcription, or also excessively long transcripts if plasmids are not completely linearized. In addition, impurities such as enzymes, for example RNases and polymerases, and nucleotides may be separated.

Using the methods according to the invention, RNA is purified which has a higher purity after purification than the starting material. It is desirable in this respect for the degree of purity to be as close as possible to 100%. A degree of purity of more than 70%, in particular 80%, very particularly 90% and most favourably 99% or more may be achieved in this way.

According to a preferred embodiment, the method for purifying at least one RNA species from a mixture of at least two RNA species comprises in step b) an HPLC process, preferably an RP-HPLC process, as described herein and, in further detail, in WO 2008/077592.

According to preferred embodiments, the method for purifying at least one RNA species from a mixture of at least two RNA species is used as an analytical purification process, e.g. during quality control in RNA manufacture. Preferably, the method is used for determining the identity and/or integrity of at least one RNA species. For example, the sequence of one RNA, which may have the same or essentially the same retention time as another RNA species in the mixture and thus elutes in one fraction (peak) together with that other RNA species, may be adapted in order for the adapted RNA species to have a distinct retention time, which results in separation of the fractions (peaks), into which the respective RNA species elute (see also Figure 1E).

According to a preferred embodiment, the method for purifying at least one RNA species from a mixture of at least two RNA species further comprises a step of determining the identity and/or integrity of at least one RNA species.

The term "integrity" describes whether complete RNA molecules are present in a sample applied to the analytical purification process as described herein. If complete RNA molecules are present, the RNA molecule of the at least one RNA species in a mixture of at least two RNA species will have the expected length. If incomplete RNA molecules are present, the length of the RNA molecule will be different from the expected or known length. Low integrity or differences in the length of the RNA molecules could be due to, amongst others, degradation, cleavage, incorrect base pairing, lack of or incomplete capping, lack of or incomplete polyadenylation, or incomplete transcription. This will become evident in a broadening of the HPLC peak of the RNA during the analytical purification process as described herein.

In an analytical purification process, smaller amounts of RNA are typically subjected to an HPLC. A typical amount of RNA subjected to analytical purification process may be smaller than about 5µg, e.g. about 4µg or less, about 3µg or less, about 2µg or less, about 1µg or less, about 500ng or less, about 400ng or less, about 300ng or less, about 200ng or less, about 100ng or less, or about 50ng or less. Preferably, the amount of RNA subjected to an analytical purification process is in a range selected from about 1ng to about 5µg, from about 10ng to about 4µg,

from about 10ng to about 3µg, from about 10ng to about 2µg, from about 10ng to about 1µg, or from about 50ng to 5µg. More preferably, the amount of RNA subjected to an analytical purification process is at least about 10ng.

The method for purifying at least one RNA species from a mixture of at least two RNA species may also be used on a preparative scale. In particular, the method according to the invention may be employed in order to purify an adapted RNA sequence from a mixture, which has been pre-purified in at least one other purification step (e.g. by HPLC).

In a preparative purification process, larger amounts of RNA are typically subjected to an HPLC. A typical amount of RNA subjected to a preparative purification process may be larger than 5µg, e.g. about 10µg or more, about 100µg or more, about 1mg or more, about 10mg or more, about 100mg or more, about 1g or more, about 10g or more, about 100g or more, about 1kg or more, about 10kg or more. Preferably, the amount of RNA subjected to a preparative purification process is in a range selected from about 5µg to about 10kg, from about 5µg to about 1kg, from about 5µg to about 500g, from about 5µg to about 100g, from about 5µg to about 50g, or from about 5µg to about 10g.

Preferably, step a) of the method comprises a step of modifying the retention time of at least one RNA species from a mixture of at least two RNA species on a chromatographic column as described herein.

In certain embodiments, the retention time of the at least one RNA species to be purified (in its unmodified/non-adapted state, corresponding to the original RNA (sequence)) is essentially the same as the retention time of the RNA species, from which it is to be separated or the retention times are so close that the RNA species elute from a chromatographic column in the same fraction (peak) or in overlapping fractions (peaks) that do not allow for physical separation of the RNA species, and in addition, do not allow for determining the integrity of at least one RNA species.

In preferred embodiments, the number of A and/or U nucleotides in the sequence of the at least one RNA species to be purified is altered as described herein so that physical separation of the at least one RNA species from other components in a mixture is feasible by chromatography, preferably by HPLC, more preferably by RP-HPLC. The retention time of the at least one RNA species is preferably modified such that the RNA species elutes in a separate fraction (peak), which does not comprise at least one other compound (e.g. another RNA species or impurities such as abortive sequences etc.), from which the at least one RNA species is to be separated. The modified retention time of the at least one adapted RNA species is different from the retention time of at least one further RNA species in the mixture, from which it may preferably be physically separated, more preferably by HPLC.

According to a particularly preferred embodiment, step a) of the method comprises adapting the sequence of the at least one RNA species to be purified so that the number of A and/or U nucleotides in the adapted sequence of the at least one RNA species differs by at least 1, preferably by at least 2, more preferably by at least 3, even more preferably by at least 4, even more preferably by at least 5, even more preferably by at least 10, most preferably by at least 15, from the number of A and/or U nucleotides in the sequence of at least one other RNA species, which is present in the mixture and from which the adapted RNA species is to be separated. Alternatively, the number of

A and/or U nucleotides in the adapted sequence of the at least one RNA species may differ by at least 5, preferably by at least 10, more preferably by at least 20, even more preferably by at least 30, even more preferably by at least 40, even more preferably by at least 50, most preferably by at least 60, from the number of A and/or U nucleotides in the sequence of another RNA species in the mixture.

In preferred embodiments, e.g. an analytical purification process, step a) of the method comprises adapting the sequence of the at least one RNA species to be purified so that the number of A and/or U nucleotides in the adapted sequence of the at least one RNA species may differ by at least 10, preferably by at least 20, more preferably by at least 30, even more preferably by at least 40, even more preferably by at least 50, even more preferably by at least 60, most preferably by at least 70, from the number of A and/or U nucleotides in the sequence of another RNA species in the mixture.

In preferred embodiments, e.g. a preparative purification process, step a) of the method comprises adapting the sequence of the at least one RNA species to be purified so that the number of A and/or U nucleotides in the adapted sequence of the at least one RNA species may differ by at least 80, preferably by at least 100, more preferably by at least 120, even more preferably by at least 140, even more preferably by at least 180, even more preferably by at least 200, most preferably by at least 220, from the number of A and/or U nucleotides in the sequence of another RNA species in the mixture.

According to preferred embodiments, e.g. in an analytical purification process, the sequence of the at least one RNA species to be purified is adapted so that the number of A and/or U nucleotides in the adapted RNA differs from the respective number in the sequence of at least one other RNA species in the mixture, from which it is to be separated by a number, which is equal to or greater than the total number of nucleotides in the sequence of the at least one RNA to be purified multiplied by a factor of at least 0.05, at least 0.06, at least 0.07, at least 0.08, at least 0.09, at least 0.1 or at least 0.15. In a preferred embodiment, the number of A and/or U nucleotides in the adapted RNA differs from the respective number in the sequence of the at least one other RNA species in the mixture, from which it is to be separated by a number, which is equal to or greater than the total number of nucleotides in the sequence of the at least one RNA to be purified multiplied by a factor of 0.065. In these embodiments, the at least one RNA species to be purified preferably has the same length (i.e. comprises the same total amount of nucleotides) as the RNA species, from which it is to be separated. More preferably, the lengths of the RNA species differ by not more than about 20%, not more than about 10%, not more than about 5%, not more than about 4%, not more than about 3%, not more than about 2%, or not more than about 1%.

According to preferred embodiments, e.g. in an analytical purification process, the sequence of the at least one RNA species to be purified is adapted so that the number of A and/or U nucleotides in the adapted RNA differs from the respective number in the sequence of at least one other RNA species in the mixture, from which it is to be separated by a number, which is equal to or greater than the total number of nucleotides in the sequence of the at least one RNA to be purified multiplied by a factor of at least 0.01, at least 0.02, at least 0.03, at least 0.04, or at least 0.05. In these embodiments, the at least one RNA species to be purified preferably has the same length (i.e. comprises the same total amount of nucleotides) as the RNA species, from which it is to be separated. More preferably, the lengths of the RNA species differ by not more than about 20%, not more than about 10%, not more

than about 5%, not more than about 4%, not more than about 3%, not more than about 2%, or not more than about 1%.

According to preferred embodiments, e.g. in a preparative purification, the sequence of the at least one RNA species to be purified is adapted so that the number of A and/or U nucleotides in the adapted RNA differs from the respective number in the sequence of at least one other RNA species in the mixture, from which it is to be separated by a number, which is equal to or greater than the total number of nucleotides in the sequence of the at least one RNA to be purified multiplied by a factor of at least 0.20, at least 0.22, at least 0.24, at least 0.26, at least 0.28, at least 0.30 or at least 0.35. In these embodiments, the at least one RNA species to be purified preferably has the same length (i.e. comprises the same total amount of nucleotides) as the RNA species, from which it is to be separated. More preferably, the lengths of the RNA species differ by not more than about 20%, not more than about 10%, not more than about 5%, not more than about 4%, not more than about 3%, not more than about 2%, or not more than about 1%.

Method for co-purifying at least two RNA species from a mixture:

According to a further aspect, the present invention provides a method for co-purifying at least two RNA species from a mixture comprising at least two RNA species, wherein the method comprises:

a) a step of adapting the sequence of at least one RNA species by altering the number of A and/or U nucleotides in the RNA with respect to the number of A and/or U nucleotides in the original RNA sequence; and

b) a chromatographic step, wherein the at least one RNA species having an adapted sequence and at least one further RNA species are co-purified.

This method preferably comprises any one of the method steps or any one of the features (or a combination thereof) described herein with regard to other aspects of the invention, in particular with regard to the method for modifying the retention time of an RNA or with regard to the method for purifying at least one RNA species from a mixture of at least two RNA species.

The method is preferably used for co-purifying at least two RNA species from a mixture comprising at least two RNA species, wherein the at least two RNA species (in their unmodified/non-adapted state) are characterized by distinct retention times, which result in elution of the at least two RNA species in separate fractions (peaks), which do not allow for co-purification.

In the method for co-purifying at least two RNA species from a mixture comprising at least two RNA species, the number of A and/or U nucleotides in an original RNA sequence is preferably altered in such a manner that the retention time of the RNA is essentially identical to or at least sufficiently close to the retention time of another compound (such as another RNA species), which is further present in a mixture together with the adapted RNA, for the adapted RNA and the other compound (e.g. another RNA species) to be co-purified by a chromatographic procedure, preferably by HPLC, more preferably by RP-HPLC as described herein. The adapted RNA and the other compound (e.g. another RNA species) preferably elute from the chromatographic column together in one fraction (or appear in the chromatogram as one single peak, respectively). The adaptation of the number of A and/or U nucleotides in an RNA (sequence) in order to obtain an adapted RNA having essentially the same retention time as

another RNA species may also be referred to herein as '(peak) harmonization' or 'harmonization of the retention times'.

More preferably, the number of A and/or U nucleotides in an original RNA sequence is preferably altered in such a manner that the retention time of the RNA (species) differs by not more than about 20%, preferably not more than about 10%, not more than about 5%, not more than about 4%, not more than about 3%, not more than about 2%, not more than about 1%, not more than 0.5% or not more than 0.1%, from the retention time of another compound (such as another RNA species), which is further present in a mixture together with the adapted RNA.

In a further aspect, the present invention thus also provides a method for harmonizing the numbers of A and/or U nucleotides in the sequences at least two RNA species, the method comprising adapting the sequence of at least one RNA species by altering the number of A and/or U nucleotides in the RNA sequence with respect to the number of A and/or U nucleotides in the original RNA sequence. Said method preferably applies any one of the steps or any one of the features (or a combination thereof) described herein with respect to the method for co-purifying at least two RNA species.

Hence, the retention times of harmonized RNA species as used herein preferably differ by not more than about 20%, more preferably by not more than about 10%, by not more than about 5%, by not more than about 4%, by not more than about 3%, by not more than about 2%, by not more than about 1%, by not more than 0.5% or by not more than 0.1%. Even more preferably, harmonized RNA species are characterized by essentially identical retention times.

In a further aspect, the present invention provides a method for synthesis of a mixture comprising at least two harmonized RNA species, preferably as described herein, the method comprising

a) a step comprising harmonizing the numbers of A and/or U nucleotides in the sequences of at least two RNA species, wherein the sequence of at least one RNA species is adapted by altering the number of A and/or U nucleotides in the RNA sequence with respect to the number of A and/or U nucleotides in the original RNA sequence; and

b) a step of synthesis of the at least two harmonized RNA species.

Step a) of this method preferably comprises adaptation of the RNA sequence of at least one of the RNA sequences by altering the number of A and/or U nucleotides in the RNA sequence with respect to the number of A and/or U nucleotides in the original RNA. In this respect and also with regard to step b), reference is made to the description of the inventive method for modifying the retention time of RNA on a chromatographic column, the inventive method for purifying at least one RNA from a mixture of at least two RNA species or the inventive method for co-purifying at least two RNA species from a mixture comprising at least two RNA species. Any one of the features described herein with respect to one or more of said methods preferably also apply to the inventive method for synthesis of a mixture comprising at least two harmonized RNA species.

In some embodiments, step a) of the method for synthesis of a mixture comprising at least two harmonized RNA species may comprise:

i) optionally, determining the total number of nucleotides in an original RNA sequence;

ii) determining the number of A and/or U nucleotides in the original RNA sequence;

iii) determining the codons in the original RNA sequence that can be replaced with at least one alternative codon without changing the encoded amino acid sequence;

iv) adjusting the number of A and/or U nucleotides in the RNA sequence to a pre-set number of A and/or U nucleotides by replacing at least one original codon with an alternative codon, wherein the alternative codon encodes the same amino acid as the original codon and is further characterized in a higher content of A and/or U nucleotides.

Step b) of the method for synthesis of a mixture comprising at least two harmonized RNA species preferably comprises the separate synthesis of the at least two harmonized RNA species. In order to obtain a mixture of harmonized RNA species, at least two harmonized RNA species are preferably mixed together. Alternatively, step b) comprises synthesis of the at least two harmonized RNA species in one batch. The at least two harmonized RNA species may be synthesized by any method known in the art, e.g. by chemical synthesis. Preferably, the harmonized RNA species are synthesized by RNA in vitro transcription.

The term "mixture" as used in this context preferably refers to a solid, semi-solid or liquid mixture. Preferably, the mixture comprising at least two harmonized RNA species may be a liquid, more preferably an aqueous solution. Further suitable solvents, buffers and excipients are known in the art and are also further described herein, in particular with respect to the inventive composition. In preferred embodiments, the mixture comprising at least two harmonized RNA species, which is obtained by the inventive method for synthesis of a mixture comprising at least two harmonized RNA species, is a composition that is preferably characterized by any one of the features described with respect to the composition (comprising at least two RNA species) as described herein.

According to a preferred embodiment, the method for co-purifying at least two RNA species (or the method for harmonizing the number of A and/or U nucleotides in the sequences of at least two RNA species or the method for synthesis of a mixture comprising at least two harmonized RNA species) comprises a step of determining a target number of A and/or U nucleotides, to which the number of A and/or U nucleotides in the at least two RNA species to be co-purified (to be harmonized) is adapted. For example, if a mixture contains a multitude of RNA species, which are characterized by essentially the same retention time, and two or more RNA species are to be separated from the remaining RNA species (i.e. co-purified), then a target number of A and/or U nucleotides is determined, preferably as described herein, which results in the RNA species to be co-purified (or harmonized) having essentially the same retention time, thus eluting from a chromatographic column within the same fraction (peak), allowing for (co-)separation of these RNA species from the undesired RNA species in the mixture.

In order for two (or more) RNA species to elute from a chromatographic column, such as a (reversed-phase) HPLC column, within the same fraction (peak), thus allowing for co-purification of the two (or more) RNA species, preferably in an analytical scale purification, the numbers of A and/or U nucleotides in the two (or more) RNA species preferably differ from each other by not more than 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 30 or 35. In preferred embodiments, the numbers of A and/or U nucleotides in the RNA species preferably differ from each other by not more than 2, more preferably by not more than 3, even more preferably by not more than 4, even more preferably by not more than 5, most preferably by not more than 6. In

a further preferred embodiment, the numbers of A and/or U nucleotides in the RNA species differ from each other by not more than 10, preferably by not more than 15, more preferably by not more than 20, even more preferably by not more than 25, most preferably by not more than 30. It is particularly preferred, that the numbers of A and/or U nucleotides in the two or more RNA species are identical.

For preparative scale purification, in order for two (or more) RNA species to elute from a chromatographic column, such as a (reversed-phase) HPLC column, within the same fraction (peak), thus allowing for co-purification of the two (or more) RNA species, the numbers of A and/or U nucleotides in the two (or more) RNA species preferably differ from each other by not more than 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 110, 120, 130, 140, 150. In preferred embodiments, the numbers of A and/or U nucleotides in the RNA species preferably differ from each other by not more than 2, more preferably by not more than 3, even more preferably by not more than 4, even more preferably by not more than 5, most preferably by not more than 6. In a further preferred embodiment, the numbers of A and/or U nucleotides in the RNA species differ from each other by not more than 10, preferably by not more than 20, more preferably by not more than 30, even more preferably by not more than 40, most preferably by not more than 50. It is particularly preferred, that the numbers of A and/or U nucleotides in the two or more RNA species are identical.

According to a preferred embodiment, the numbers of A and/or U nucleotides in the RNA species, which are to be co-purified and which preferably have the same length, differ by a number, which is equal to or lower than the total number of nucleotides in the RNA multiplied by a factor equal to or lower than 0.008, equal to or lower than 0.0075, equal to or lower than 0.007, equal to or lower than 0.0065, equal to or lower than 0.006, equal to or lower than 0.0055, equal to or lower than 0.005, equal to or lower than 0.0045, equal to or lower than 0.004, equal to or lower than 0.0035 or equal or lower than 0.003. More preferably, the numbers of A and/or U nucleotides in the RNA species, which are to be co-purified and which preferably have the same length, differ by a number, which is equal to or lower than the total number of nucleotides in the RNA multiplied by a factor equal to or lower than 0.0065, preferably equal to or lower than 0.0063.

According to preferred embodiments, the numbers of A and/or U nucleotides in the RNA species, which are to be co-purified and which preferably have the same length, differ by a number, which is equal to or lower than the total number of nucleotides in the RNA multiplied by a factor equal to or lower than 0.030, equal to or lower than 0.025, equal to or lower than 0.020, equal to or lower than 0.015, equal to or lower than 0.014, equal to or lower than 0.013, equal to or lower than 0.012, equal to or lower than 0.011, equal to or lower than 0.010, equal to or lower than 0.009 or equal or lower than 0.008.

Modified RNA:

According preferred embodiments of the invention, the RNA (species) as described herein, may be in the form of modified RNA, wherein any modification, as defined herein, may be introduced into the RNA. A modification as defined herein preferably leads to an artificial RNA as described herein, which is further stabilized.

According to one embodiment, the RNA, may thus be provided as a 'stabilized RNA', preferably as a 'stabilized mRNA', that is to say as an RNA that is essentially resistant to in vivo degradation (e.g. by an exo- or endo-nuclease). Such stabilization can be effected, for example, by a modified phosphate backbone of an artificial RNA. A backbone modification in connection with the present invention is a modification in which phosphates of the backbone of the nucleotides contained in the RNA are chemically modified. Nucleotides that may be preferably used in this connection contain e.g. a phosphorothioate-modified phosphate backbone, preferably at least one of the phosphate oxygens contained in the phosphate backbone being replaced by a sulfur atom. Stabilized RNAs may further include, for example: non-ionic phosphate analogues, such as, for example, alkyl and aryl phosphonates, in which the charged phosphonate oxygen is replaced by an alkyl or aryl group, or phosphodiesters and alkylphosphotriesters, in which the charged oxygen residue is present in alkylated form. Such backbone modifications typically include, without implying any limitation, modifications from the group consisting of methylphosphonates, phosphoramidates and phosphorothioates (e.g. cytidine-5'-O-(1-thiophosphate)).

In the following, specific modifications are described, each of which is preferably capable of 'stabilizing' the RNA, preferably an mRNA, as described herein.

Chemical modifications:

The terms 'nucleic acid modification' or 'RNA modification' as used herein may refer to chemical modifications comprising backbone modifications as well as sugar modifications or base modifications.

In this context, a modified RNA as defined herein may contain nucleotide analogues/modifications, e.g. backbone modifications, sugar modifications or base modifications. A backbone modification in connection with the present invention is a modification, in which phosphates of the backbone of the nucleotides contained in the RNA as defined herein are chemically modified. A sugar modification in connection with the present invention is a chemical modification of the sugar of the nucleotides of the RNA as defined herein. Furthermore, a base modification in connection with the present invention is a chemical modification of the base moiety of the nucleotides of the artificial RNA. In this context, nucleotide analogues or modifications are preferably selected from nucleotide analogues, which are applicable for transcription and/or translation.

Sugar Modifications:

The modified nucleosides and nucleotides, which may be incorporated into a modified RNA as described herein, can be modified in the sugar moiety. For example, the 2' hydroxyl group (OH) can be modified or replaced with a number of different "oxy" or "deoxy" substituents. Examples of "oxy" -2' hydroxyl group modifications include, but are not limited to, alkoxy or aryloxy (-OR, e.g., R = H, alkyl, cycloalkyl, aryl, aralkyl, heteroaryl or sugar); polyethyleneglycols (PEG), $-O(CH_2CH_2O)_nCH_2CH_2OR$; "locked" nucleic acids (LNA) in which the 2' hydroxyl is connected, e.g., by a methylene bridge, to the 4' carbon of the same ribose sugar; and amino groups (-O-amino, wherein the amino group, e.g., NRR, can be alkylamino, dialkylamino, heterocyclyl, arylamino, diarylamino, heteroarylamino, or diheteroaryl amino, ethylene diamine, polyamino) or aminoalkoxy.

"Deoxy" modifications include hydrogen, amino (e.g. NH_2; alkylamino, dialkylamino, heterocyclyl, arylamino, diaryl amino, heteroaryl amino, diheteroaryl amino, or amino acid); or the amino group can be attached to the sugar through a linker, wherein the linker comprises one or more of the atoms C, N, and O.

The sugar group can also contain one or more carbons that possess the opposite stereochemical configuration than that of the corresponding carbon in ribose. Thus, an RNA can include nucleotides containing, for instance, arabinose as the sugar.

Backbone Modifications:

The phosphate backbone may further be modified in the modified nucleosides and nucleotides, which may be incorporated into a modified RNA as described herein. The phosphate groups of the backbone can be modified by replacing one or more of the oxygen atoms with a different substituent. Further, the modified nucleosides and nucleotides can include the full replacement of an unmodified phosphate moiety with a modified phosphate as described herein. Examples of modified phosphate groups include, but are not limited to, phosphorothioate, phosphoroselenates, borano phosphates, borano phosphate esters, hydrogen phosphonates, phosphoroamidates, alkyl or aryl phosphonates and phosphotriesters. Phosphorodithioates have both non-linking oxygens replaced by sulfur. The phosphate linker can also be modified by the replacement of the linking oxygen with nitrogen (bridged phosphoroamidates), sulfur (bridged phosphorothioates) and carbon (bridged methylene-phosphonates).

Base Modifications:

The modified nucleosides and nucleotides, which may be incorporated into a modified nucleic acid, preferably an mRNA as described herein can further be modified in the nucleobase moiety. Examples of nucleobases found in a nucleic acid such as RNA include, but are not limited to, adenine, guanine, cytosine and uracil. For example, the nucleosides and nucleotides described herein can be chemically modified on the major groove face. In some embodiments, the major groove chemical modifications can include an amino group, a thiol group, an alkyl group, or a halo group.

In particularly preferred embodiments of the present invention, the nucleotide analogues/modifications are selected from base modifications, which are preferably selected from 2-amino-6-chloropurineriboside-5'-triphosphate, 2-Aminopurine-riboside-5'-triphosphate; 2-aminoadenosine-5'-triphosphate, 2'-Amino-2'-deoxycytidine-triphosphate, 2-thiocytidine-5'-triphosphate, 2-thiouridine-5'-triphosphate, 2'-Fluorothymidine-5'-triphosphate, 2'-O-Methyl inosine-5'-triphosphate 4-thiouridine-5'-triphosphate, 5-aminoallylcytidine-5'-triphosphate, 5-aminoallyluridine-5'-triphosphate, 5-bromocytidine-5'-triphosphate, 5-bromouridine-5'-triphosphate, 5-Bromo-2'-deoxycytidine-5'-triphosphate, 5-Bromo-2'-deoxyuridine-5'-triphosphate, 5-iodocytidine-5'-triphosphate, 5-Iodo-2'-deoxycytidine-5'-triphosphate, 5-iodouridine-5'-triphosphate, 5-Iodo-2'-deoxyuridine-5'-triphosphate, 5-methylcytidine-5'-triphosphate, 5-methyluridine-5'-triphosphate, 5-Propynyl-2'-deoxycytidine-5'-triphosphate, 5-Propynyl-2'-deoxyuridine-5'-triphosphate, 6-azacytidine-5'-triphosphate, 6-azauridine-5'-triphosphate, 6-chloropurineriboside-5'-triphosphate, 7-deazaadenosine-5'-triphosphate, 7-deazaguanosine-5'-triphosphate, 8-azaadenosine-5'-triphosphate, 8-azidoadenosine-5'-triphosphate, benzimidazole-riboside-5'-triphosphate, N1-methyladenosine-5'-triphosphate, N1-methylguanosine-5'-triphosphate, N6-methyladenosine-5'-triphosphate, O6-methylguanosine-5'-triphosphate, pseudouridine-5'-triphosphate, or puromycin-5'-triphosphate, xanthosine-5'-triphosphate. Particular

preference is given to nucleotides for base modifications selected from the group of base-modified nucleotides consisting of 5-methylcytidine-5'-triphosphate, 7-deazaguanosine-5'-triphosphate, 5-bromocytidine-5'-triphosphate, and pseudouridine-5'-triphosphate.

In some embodiments, modified nucleosides include pyridin-4-one ribonucleoside, 5-aza-uridine, 2-thio-5-aza-uridine, 2-thiouridine, 4-thio-pseudouridine, 2-thio-pseudouridine, 5-hydroxyuridine, 3-methyluridine, 5-carboxymethyl-uridine, 1-carboxymethyl-pseudouridine, 5-propynyl-uridine, 1-propynyl-pseudouridine, 5-taurinomethyluridine, 1-taurinomethyl-pseudouridine, 5-taurinomethyl-2-thio-uridine, 1-taurinomethyl-4-thio-uridine, 5-methyl-uridine, 1-methyl-pseudouridine, 4-thio-1-methyl-pseudouridine, 2-thio-1-methyl-pseudouridine, 1-methyl-1-deaza-pseudouridine, 2-thio-1-methyl-1-deaza-pseudouridine, dihydrouridine, dihydropseudouridine, 2-thio-dihydrouridine, 2-thio-dihydropseudouridine, 2-methoxyuridine, 2-methoxy-4-thio-uridine, 4-methoxy-pseudouridine, and 4-methoxy-2-thio-pseudouridine.

In some embodiments, modified nucleosides include 5-aza-cytidine, pseudoisocytidine, 3-methyl-cytidine, N4-acetylcytidine, 5-formylcytidine, N4-methylcytidine, 5-hydroxymethylcytidine, 1-methyl-pseudoisocytidine, pyrrolo-cytidine, pyrrolo-pseudoisocytidine, 2-thio-cytidine, 2-thio-5-methyl-cytidine, 4-thio-pseudoisocytidine, 4-thio-1-methyl-pseudoisocytidine, 4-thio-1-methyl-1-deaza-pseudoisocytidine, 1-methyl-1-deaza-pseudoisocytidine, zebularine, 5-aza-zebularine, 5-methyl-zebularine, 5-aza-2-thio-zebularine, 2-thio-zebularine, 2-methoxy-cytidine, 2-methoxy-5-methyl-cytidine, 4-methoxy-pseudoisocytidine, and 4-methoxy-1-methyl-pseudoisocytidine.

In other embodiments, modified nucleosides include 2-aminopurine, 2, 6-diaminopurine, 7-deaza-adenine, 7-deaza-8-aza-adenine, 7-deaza-2-aminopurine, 7-deaza-8-aza-2-aminopurine, 7-deaza-2,6-diaminopurine, 7-deaza-8-aza-2,6-diaminopurine, 1-methyladenosine, N6-methyladenosine, N6-isopentenyladenosine, N6-(cis-hydroxyisopentenyl)adenosine, 2-methylthio-N6-(cis-hydroxyisopentenyl) adenosine, N6-glycinylcarbamoyladenosine, N6-threonylcarbamoyladenosine, 2-methylthio-N6-threonyl carbamoyladenosine, N6,N6-dimethyladenosine, 7-methyladenine, 2-methylthio-adenine, and 2-methoxy-adenine.

In other embodiments, modified nucleosides include inosine, 1-methyl-inosine, wyosine, wybutosine, 7-deaza-guanosine, 7-deaza-8-aza-guanosine, 6-thio-guanosine, 6-thio-7-deaza-guanosine, 6-thio-7-deaza-8-aza-guanosine, 7-methyl-guanosine, 6-thio-7-methyl-guanosine, 7-methylinosine, 6-methoxy-guanosine, 1-methylguanosine, N2-methylguanosine, N2,N2-dimethylguanosine, 8-oxo-guanosine, 7-methyl-8-oxo-guanosine, 1-methyl-6-thio-guanosine, N2-methyl-6-thio-guanosine, and N2,N2-dimethyl-6-thio-guanosine.

In some embodiments, the nucleotide can be modified on the major groove face and can include replacing hydrogen on C-5 of uracil with a methyl group or a halo group.

In specific embodiments, a modified nucleoside is 5'-O-(1-thiophosphate)-adenosine, 5'-O-(1-thiophosphate)-cytidine, 5'-O-(1-thiophosphate)-guanosine, 5'-O-(1-thiophosphate)-uridine or 5'-O-(1-thiophosphate)-pseudouridine.

In further specific embodiments, a modified RNA may comprise nucleoside modifications selected from 6-aza-cytidine, 2-thio-cytidine, alpha-thio-cytidine, Pseudo-iso-cytidine, 5-aminoallyl-uridine, 5-iodo-uridine, N1-methyl-pseudouridine, 5,6-dihydrouridine, alpha-thio-uridine, 4-thio-uridine, 6-aza-uridine, 5-hydroxy-uridine, deoxy-thymidine, 5-methyl-uridine, Pyrrolo-cytidine, inosine, alpha-thio-guanosine, 6-methyl-guanosine, 5-methyl-cytdine, 8-oxo-guanosine, 7-deaza-guanosine, N1-methyl-adenosine, 2-amino-6-Chloro-purine, N6-methyl-2-amino-purine, Pseudo-iso-cytidine, 6-Chloro-purine, N6-methyl-adenosine, alpha-thio-adenosine, 8-azido-adenosine, 7-deaza-adenosine.

In some embodiments, the RNA as used herein comprises at least one coding region as defined herein, wherein the coding region comprises at least one modified uridine nucleoside, preferably selected from N(1)-methylpseudouridine (m1ψ), pseudouridine (ψ), 5-methoxyuridine. Alternatively or in addition, the RNA as used herein comprises at least one coding region as defined herein, wherein the coding region comprises at least one modified cytosine nucleoside, preferably 5-methylcytosine.

Lipid modification:
According to a further embodiment, a modified RNA as defined herein can contain a lipid modification. Such a lipid-modified RNA as defined herein typically further comprises at least one linker covalently linked with that RNA, and at least one lipid covalently linked with the respective linker. Alternatively, the lipid-modified RNA comprises at least one RNA as defined herein and at least one (bifunctional) lipid covalently linked (without a linker) with that RNA. According to a third alternative, the lipid-modified RNA comprises an RNA molecule as defined herein, at least one linker covalently linked with that RNA, and at least one lipid covalently linked with the respective linker, and also at least one (bifunctional) lipid covalently linked (without a linker) with that RNA. In this context, it is particularly preferred that the lipid modification is present at the terminal ends of a linear RNA.

<u>Optimized coding sequences of the original RNA:</u>
The term "optimized coding sequence" relates to coding sequences that differ in at least one codon (triplets of nucleotides coding for one amino acid) compared to the corresponding wild type coding sequence. Preferably, a codon "optimized coding sequence" in the context of the invention may show improved resistance to in vivo degradation and/or improved stability in vivo, and/or improved translatability in vivo. Codon optimization in the broadest sense make use of the degeneracy of the genetic code wherein multiple codons may encode the same amino acid and may be used interchangeably to optimize/modify the coding sequence for in vivo applications (e.g. vaccination).

In particularly preferred embodiments, the at least one RNA or at least one RNA species is an mRNA comprising a coding region, wherein the coding region of the original RNA sequence comprises an optimized coding sequence, wherein the optimized coding sequence is selected from C maximized coding sequence, G/C increased coding sequence, G/C optimized coding sequence, human codon usage optimized coding sequence, CAI maximized coding sequence, or any combination thereof.

According to certain embodiments, the RNA, particularly the coding sequence of the original RNA may be optimized, wherein the C content of the at least one coding sequence may be increased, preferably maximized, compared to

the C content of the corresponding wild type coding sequence (herein referred to as "C maximized coding sequence"). The amino acid sequence encoded by the C maximized coding sequence of the nucleic acid sequence is preferably not modified as compared to the amino acid sequence encoded by the respective wild type nucleic acid coding sequence. The generation of a Cytosine optimized RNA as described above may suitably be carried out using a C maximization method according to WO2015/062738. In this context, the disclosure of WO2015/062738 is included herewith by reference.

According to some embodiments, the RNA, particularly the coding sequence of the original RNA may be optimized, wherein the G/C content of the coding sequence may be increased compared to the G/C content of the corresponding wild type coding sequence (herein referred to as "G/C increased coding sequence"). The term "G/C increased coding sequence" relates to RNA, preferably the coding sequence of the original RNA of the invention that comprises an increased number of guanosine and/or cytosine nucleotides as compared to the corresponding wild type nucleic acid sequence. Such an increased number may be generated by substitution of codons containing adenosine or thymidine nucleotides by codons containing guanosine or cytosine nucleotides. If the enriched G/C content occurs in a coding sequence of DNA or RNA, it makes use of the degeneracy of the genetic code. In particular, in case of RNA, sequences having an increased G (guanosine)/C (cytosine) content are more stable than sequences having an increased A (adenosine)/U (uracil) content. The amino acid sequence encoded by the G/C content modified coding sequence of the nucleic acid sequence is preferably not modified as compared to the amino acid sequence encoded by the respective wild type nucleic acid coding sequence. Preferably, the G/C content of the original RNA coding sequence of the present invention is increased by at least 10%, preferably by at least 20%, more preferably by at least 30%, most preferably by at least 40% compared to the G/C content of the coding sequence of the corresponding wild type nucleic coding sequence.

According to preferred embodiments, the RNA, particularly the coding sequence of the original RNA may be optimized, wherein the G/C content of the coding sequence may be optimized compared to the G/C content of the corresponding wild type coding sequence (herein referred to as "G/C optimized coding sequence"). "Optimized" in that context refers to a coding sequence wherein the G/C content is preferably increased to the essentially highest possible G/C content. The amino acid sequence encoded by the G/C content optimized coding sequence of the nucleic acid sequence is preferably not modified as compared to the amino acid sequence encoded by the respective wild type nucleic acid coding sequence. The generation of a G/C optimized RNA sequences as described above may suitably be carried out using a G/C optimization method explained in WO2002/098443. In this context, the disclosure of WO2002/098443 is included in its full scope in the present invention.

In particularly preferred embodiments, the at least one RNA or at least one RNA species is an mRNA comprising a coding region, wherein the at least one coding region of the original RNA sequence comprises a nucleic acid sequence, which is G/C optimized.

According to preferred embodiments, the RNA, particularly the coding sequence of the original RNA may be optimized, wherein the codons coding sequence may be optimized to the human codon usage (herein referred to as "human codon usage optimized coding sequence"). Codons encoding the same amino acid occur at different frequencies in a subject, e.g. a human. Accordingly, the coding sequence of the RNA as defined herein is preferably

optimized such that the frequency of the codons encoding the same amino acid corresponds to the naturally occurring frequency of that codon according to the human codon usage. For example, in the case of the amino acid Alanine (Ala), the wild type coding sequence is preferably adapted in a way that the codon "GCC" is used with a frequency of 0.40, the codon "GCT" is used with a frequency of 0.28, the codon "GCA" is used with a frequency of 0.22 and the codon "GCG" is used with a frequency of 0.10 etc. Accordingly, such a procedure (as exemplified for Ala) is applied for each amino acid encoded by the coding sequence of the RNA, preferably the original RNA.

According to preferred embodiments, the RNA, particularly the coding sequence of the original RNA may be optimized, wherein the codon adaptation index (CAI) may be increased or preferably maximised (herein referred to as "CAI maximized coding sequence"). Accordingly, it is preferred that all codons of the wild type sequence that are relatively rare in the cell (e.g. a human) are exchanged for a respective codon that is frequent in the cell, wherein the frequent codon encodes the same amino acid as the relatively rare codon. Suitably, the most frequent codons are used for each encoded amino acid. Suitably, the RNA, preferably the original RNA of the present invention comprises at least one coding sequence, wherein the codon adaptation index (CAI) of the at least one coding sequence is at least 0.5, at least 0.8, at least 0.9 or at least 0.95. Most preferably, the codon adaptation index (CAI) of the at least one coding sequence is 1. For example, in the case of the amino acid alanine (Ala) present in the amino acid sequence encoded by the at least one coding sequence of the nucleic acid sequence according to the invention, the wild type coding sequence is adapted in a way that the most frequent human codon "GCC" is always used for said amino acid. Accordingly, such a procedure (as exemplified for Ala) is applied for each amino acid encoded by the coding sequence of the RNA, preferably the original RNA to obtain CAI maximized coding sequences.

5'-cap structure

According to another preferred embodiment of the invention, the RNA as used herein may be modified by the addition of a 5'-cap structure, which preferably stabilizes the RNA as described herein.

The term "5'-cap structure" as used herein will be recognized and understood by the person of ordinary skill in the art, and is for example intended to refer to a modified nucleotide (cap analogue), particularly a guanine nucleotide, added to the 5' end of an RNA molecule, e.g. an mRNA molecule. Preferably, the 5'-cap is added using a 5'-5'-triphosphate linkage (also named m7GpppN). Further examples of 5'-cap structures include glyceryl, inverted deoxy abasic residue (moiety), 4',5' methylene nucleotide, 1-(beta-D-erythrofuranosyl) nucleotide, 4'-thio nucleotide, carbocyclic nucleotide, 1,5-anhydrohexitol nucleotide, L-nucleotides, alpha-nucleotide, modified base nucleotide, threo-pentofuranosyl nucleotide, acyclic 3',4'-seco nucleotide, acyclic 3,4-dihydroxybutyl nucleotide, acyclic 3,5 dihydroxypentyl nucleotide, 3'-3'-inverted nucleotide moiety, 3'-3'-inverted abasic moiety, 3'-2'-inverted nucleotide moiety, 3'-2'-inverted abasic moiety, 1,4-butanediol phosphate, 3'-phosphoramidate, hexylphosphate, aminohexyl phosphate, 3'-phosphate, 3'phosphorothioate, phosphorodithioate, or bridging or non-bridging methylphosphonate moiety. Further modified 5'-cap structures which may be used in the context of the present invention are cap1 (additional methylation of the ribose of the adjacent nucleotide of m7GpppN), cap2 (additional methylation of the ribose of the 2nd nucleotide downstream of the m7GpppN), cap3 (additional methylation of the ribose of the 3rd nucleotide downstream of the m7GpppN), cap4 (additional methylation of the ribose of the 4th nucleotide downstream of the m7GpppN), ARCA (anti-reverse cap analogue), modified ARCA (e.g. phosphothioate modified

ARCA), inosine, N1-methyl-guanosine, 2'-fluoro-guanosine, 7-deaza-guanosine, 8-oxo-guanosine, 2-amino-guanosine, LNA-guanosine, and 2-azido-guanosine.

A 5' cap (cap0 or cap1) structure may also be formed in chemical RNA synthesis or RNA in vitro transcription (co-transcriptional capping) using cap analogues.

The term "cap analogue" as used herein will be recognized and understood by the person of ordinary skill in the art, and is for example intended to refer to a non-polymerizable di-nucleotide that has cap functionality in that it facilitates translation or localization, and/or prevents degradation of a nucleic acid molecule, particularly of an RNA molecule, when incorporated at the 5' end of the nucleic acid molecule. Non-polymerizable means that the cap analogue will be incorporated only at the 5' terminus because it does not have a 5' triphosphate and therefore cannot be extended in the 3' direction by a template-dependent polymerase, particularly, by template-dependent RNA polymerase. Examples of cap analogues include, but are not limited to, a chemical structure selected from the group consisting of m7GpppG, m7GpppA, m7GpppC; unmethylated cap analogues (e.g., GpppG); dimethylated cap analogue (e.g., m2,7GpppG), trimethylated cap analogue (e.g., m2,2,7GpppG), dimethylated symmetrical cap analogues (e.g., m7Gpppm7G), or anti reverse cap analogues (e.g., ARCA; m7,2'OmeGpppG, m7,2'dGpppG, m7,3'OmeGpppG, m7,3'dGpppG and their tetraphosphate derivatives). Further cap analogues have been described previously (WO2008/016473, WO2008/157688, WO2009/149253, WO2011/015347, and WO2013/059475). Further suitable cap analogons in that context are described in WO2017/066793, WO2017/066781, WO2017/066791, WO2017/066789, WO2017/066782, WO2017/066797, wherein the disclosures referring to cap analogues are incorporated herewith by reference.

In a preferred embodiment, the 5'-cap structure is added co-transcriptionally using cap-analogues, preferably as defined herein, in an RNA in vitro transcription reaction as described herein. Preferred cap-analogues in the context of the invention are m7G(5')ppp(5')G (m7G) or 3'-O-Me-m7G(5')ppp(5')G. Further preferred cap-analogues in the context of the invention are m7G(5')ppp(5')(2'OMeA)pG or m7G(5')ppp(5')(2'OMeG)pG to co-transcriptionally generate cap1 structures.

In another embodiment, the 5'-cap structure is added via enzymatic capping using capping enzymes (e.g. vaccinia virus capping enzymes, commercially available capping kits) to generate cap0 or cap1 or cap2 structures. In other embodiments, the 5'-cap structure (cap0, cap1) is added via enzymatic capping using immobilized capping enzymes, e.g. in a capping reactor (WO2016/193226).

m7GpppN (cap0), cap1 and cap2 are 5'-cap structure naturally occurring in RNA transcribed by polymerase II and is therefore not considered as modification comprised in a modified RNA in this context. Accordingly, a modified RNA sequence of the present invention may comprise a cap0, cap1 or cap2, but additionally the modified RNA sequence may comprise at least one further modification as defined herein.

Poly(A) sequence

According to another preferred embodiment of the invention, the RNA as used herein may contain a poly(A) sequence or a poly(A) tail.

The terms "poly(A) sequence", "poly(A) tail" or "3'-poly(A) tail" are recognized and understood by the person of ordinary skill in the art, and are for example intended to be a sequence of adenosine nucleotides, typically located at the 3'-end of an RNA, of up to about 400 adenosine nucleotides. In the context of the present invention, a poly(A) sequence may be located within an mRNA or any other nucleic acid molecule, such as, e.g., in a vector, for example, in a vector serving as template for the generation of an RNA, preferably an mRNA, e.g., by transcription of the vector.

In a preferred embodiment, the poly(A) sequence, suitable located at the 3' terminus, is typically about 10 to 200 adenosine nucleotides, preferably about 10 to 100 adenosine nucleotides, more preferably about 40 to 80 adenosine nucleotides or even more preferably about 50 to 70 adenosine nucleotides. Preferably, the poly(A) sequence in the RNA as used herein is derived from a DNA template by RNA in vitro transcription.

In other embodiments, the poly(A) sequence is obtained in vitro by common methods of chemical synthesis without being necessarily transcribed from a DNA template.

In further embodiments, poly(A) sequences are generated by enzymatic polyadenylation of the RNA (after RNA in vitro transcription) using commercially available polyadenylation kits and corresponding protocols known in the art, or alternatively, by using immobilized poly(A) polymerases, e.g. in a polyadenylation reactor (as described in WO2016174271).

Alternatively, the RNA as used herein may comprise a polyadenylation signal. A polyadenylation signal typically comprises a hexamer consisting of adenine and uracil/thymine nucleotides, preferably the hexamer sequence AAUAAA. Other sequences, preferably hexamer sequences, are also conceivable. Polyadenylation typically occurs during processing of a pre-mRNA (also called premature-mRNA). In this context, a consensus polyadenylation signal is preferred comprising the NN(U/T)ANA consensus sequence. In a particularly preferred embodiment, the polyadenylation signal comprises one of the following sequences: AA(U/T)AAA or A(U/T)(U/T)AAA (wherein uridine is usually present in RNA and thymidine is usually present in DNA).

In some embodiments, the RNA as used herein may contain a poly(A) sequence derived from a vector and at least one additional poly(A) sequence generated by enzymatic polyadenylation, e.g. as described in WO2016/091391.

Poly(C) sequence

According to a further preferred embodiment, the RNA as used herein may contain a poly(C) sequence. The term "poly(C) sequence" has to be understood as a long sequence of cytosine nucleotides, typically about 10 to about 200 cytosine nucleotides, preferably about 10 to about 100 cytosine nucleotides, more preferably about 10 to about 70 cytosine nucleotides or even more, preferably about 20 to about 50, or even about 20 to about 30 cytosine nucleotides. A poly(C) sequence may preferably be located 3' of the coding sequence comprised by a nucleic acid. Preferably, the poly(C) sequence in the RNA as used herein is derived from a DNA template by RNA in vitro transcription.

UTRs

In a preferred embodiment, the at least one RNA or the at least one RNA species comprises a 3'-UTR and/or a 5'-UTR. Preferably, a 3'-UTR or a 5'-UTR as used herein comprises one or more 3'-UTR elements or 5'-UTR elements, respectively. Accordingly, the RNA as used herein may comprise at least one 5'-UTR element and/or at least one 3'-UTR element. In this context, an UTR element comprises or consists of a nucleic acid sequence, which is derived from a 5'-UTR or 3'-UTR of any naturally occurring gene or which is derived from a fragment, a homolog or a variant of the 5'-UTR or 3'-UTR of a gene. Preferably, the 5'-UTR or 3'-UTR element used in the context of the present invention is heterologous to the at least one coding sequence of the RNA as used herein. 5'-UTR or 3'-UTR elements are preferably derived from naturally occurring genes. In other embodiments, synthetically engineered 5'-UTR or 3'-UTR elements may be used in the context of the present invention.

The term "3'-untranslated region (3'-UTR)" will be recognized and understood by the person of ordinary skill in the art, and are for example intended to refer to a part of a nucleic acid molecule, which is located 3' (i.e. "downstream") of a coding sequence and which is typically not translated into protein. Usually, a 3'-UTR is the part of an mRNA which is located between the coding sequence (cds) and the poly(A) sequence of the mRNA. In the context of the invention, the term 3'-UTR may also comprise elements, which are not encoded in the DNA template, from which RNA is transcribed, but which are added after transcription during maturation, e.g. a poly(A) sequence.

The term "5'-untranslated region (5'-UTR)" will be recognized and understood by the person of ordinary skill in the art, and are for example intended to refer to a part of a nucleic acid molecule, which is located 5' (i.e. "upstream") of a coding sequence and which is not translated into protein. A 5'-UTR is typically understood to be a particular section of messenger RNA (mRNA), which is located 5' of the coding sequence of the mRNA. Typically, the 5'-UTR starts with the transcriptional start site and ends one nucleotide before the start codon of the coding sequence. Preferably, the 5'-UTRs have a length of more than 20, 30, 40 or 50 nucleotides. The 5'-UTR may comprise elements for controlling gene expression, also called regulatory elements. Such regulatory elements may be, for example, ribosomal binding sites. The 5'-UTR may be post-transcriptionally modified, for example by addition of a 5'-cap.

3'-UTR elements

In a preferred embodiment, the RNA as used herein may comprise at least one 3'-UTR element, which is typically located within the 3'-UTR of the RNA as described herein.

Preferably, the 3'-UTRs in the context of the invention are heterologous to the coding sequence. More preferably, the RNA as used herein, preferably the 3'UTR of the RNA as used herein, comprises at least one heterologous 3'-UTR element.

Preferably, the at least one 3'-UTR element comprises or consists of a nucleic acid sequence derived from the 3'-UTR of a chordate gene, preferably a vertebrate gene, more preferably a mammalian gene, most preferably a human gene, or from a variant of the 3'-UTR of a chordate gene, preferably a vertebrate gene, more preferably a mammalian gene, most preferably a human gene.

Preferably, the RNA as used herein comprises a 3'-UTR element, preferably as described herein, which may be derivable from a gene that relates to RNA with an enhanced half-life (that provides a stable RNA). Preferably, the 3' UTR element is a nucleic acid sequence derived from a 3' UTR of a gene, which preferably encodes a stable RNA, or from a homolog, a fragment or a variant of said gene

In a particularly preferred embodiment, the 3'-UTR element comprises or consists of a nucleic acid sequence, which is derived from a 3'-UTR of a gene selected from the group consisting of an albumin gene, an alpha-globin gene, a beta-globin gene, a tyrosine hydroxylase gene, a lipoxygenase gene, and a collagen alpha gene, such as a collagen alpha 1(I) gene, or from a variant of a 3'-UTR of a gene selected from the group consisting of an albumin gene, an alpha-globin gene, a beta-globin gene, a tyrosine hydroxylase gene, a lipoxygenase gene, and a collagen alpha gene, such as a collagen alpha 1(I) gene according to SEQ ID NOs: 1369-1393 of the patent application WO2013/143700, or from a homolog, a fragment or a variant thereof. Particularly preferred are 3'-UTR sequences according to SEQ ID NO: 1369 (Human albumin 3'-UTR) and SEQ ID NO: 1376 (albumin7 3'-UTR) of the patent application WO2013/143700. In this context, the disclosure of WO2013/143700 is incorporated herein by reference. Alternatively, the RNA as used herein preferably comprises a 3'-UTR element, which comprises or consists of a nucleic acid sequence selected from SEQ ID NOs: 30572, 30574, 30576, 30578, 30580, 30582 or 30584 or a fragment or variant thereof.

In other embodiments, the RNA as used herein comprises a 3'-UTR element according to any one of SEQ ID NOs: 10 to 205 of WO2015/101414 and SEQ ID NO: 1 and 2 of WO2015/101415. In this context, the disclosures of WO2015/101414 and WO2015/101415 are incorporated herewith by reference.

In some embodiments, the RNA used herein comprises a 3'-UTR element, which may be any 3'-UTR element described in WO2016/107877. In this context, the disclosure of WO2016/107877 relating to 3'-UTR elements/sequences is herewith incorporated by reference. Particularly preferred 3'-UTR elements are SEQ ID NOs: 1 to 24 and SEQ ID NOs: 49 to 318 of the patent application WO2016/107877, or fragments or variants of these sequences. In this context, it is particularly preferred that the 3'-UTR element of the RNA sequence according to the present invention comprises or consists of a corresponding RNA sequence of the nucleic acid sequence according SEQ ID NOs: 1 to 24 and SEQ ID NOs: 49 to 318 of the patent application WO2016/107877.

In certain embodiments, the RNA as defined herein comprises a 3'-UTR element, which may be any 3'-UTR element as described in WO2017/036580. In this context, the disclosure of WO2017/036580 relating to 3'-UTR elements/sequences is herewith incorporated by reference. Particularly preferred 3'-UTR elements are nucleic acid sequences according to SEQ ID NOs: 152 to 204 of the patent application WO2017/036580, or fragments or variants of these sequences. In this context, it is particularly preferred that the 3'-UTR element of the RNA sequence according to the present invention comprises or consists of a corresponding RNA sequence of the nucleic acid sequence according SEQ ID NOs: 152 to 204 of the patent application WO2017/036580.

Further, particulraly suitable 3'-UTRs are disclosed in WO2018172556. The disclosure of WO2018172556, in particular relating to GNAS, CASP1, PSMB3, ALB, COX6B1, NDUFA1 and RPS9, is included herewith by reference. Preferred 3'-UTR elements are derived from GNAS, CASP1, PSMB3, ALB, COX6B1, NDUFA1 or RPS9. Particularly

preferred 3'-UTR elements are nucleic acid sequences according to SEQ ID NOs: 23 to 36 as disclosed in WO2018172556, or fragments or variants of these sequences.

5'-UTR elements

In a preferred embodiment, the RNA as used herein may comprise at least one 5'-UTR element, which is typically located within the 5'-UTR of the RNA as described herein.

Preferably, the 5'-UTRs in the context of the invention are heterologous to the coding sequence. More preferably, the RNA as used herein, preferably the 5'UTR of the RNA as used herein, comprises at least one heterologous 5'-UTR element.

In preferred embodiments, the RNA as used herein comprises at least one 5'-UTR element. Preferably, the at least one 5'-UTR element comprises or consists of a nucleic acid sequence derived from the 5'-UTR of a chordate gene, preferably a vertebrate gene, more preferably a mammalian gene, most preferably a human gene, or from a variant of the 5'-UTR of a chordate gene, preferably a vertebrate gene, more preferably a mammalian gene, most preferably a human gene.

In certain embodiments, the RNA as used herein comprises at least one 5'-UTR element, which may be any 5'-UTR element as described in the patent application WO2013/143700. Suitably, the at least one 5'-UTR element is derived from a nucleic acid sequence selected from the group consisting of SEQ ID NOs: 1-1363, SEQ ID NO: 1395, SEQ ID NO: 1421 and SEQ ID NO: 1422 of the patent application WO2013/143700. In this context, the disclosure of WO2013/143700 is incorporated herewith by reference.

In preferred embodiments, the RNA as used herein comprises at least one heterologous 5'-UTR element, which comprises or consists of a nucleic acid sequence, which is derived from the 5'-UTR of a TOP gene, preferably from a corresponding RNA sequence, or a homolog, a fragment, or a variant thereof, preferably lacking the 5'TOP motif. More preferably, the at least one heterologous 5'-UTR element comprises or consists of a nucleic acid sequence, which is derived from a 5'-UTR of a TOP gene encoding a ribosomal protein, preferably from a corresponding RNA sequence, or a homolog, a fragment or a variant of said nucleic acid sequence, preferably lacking the 5'TOP motif

In preferred embodiments, the at least one 5'-UTR element comprises or consists of a nucleic acid sequence, which is derived from a 5'-UTR of a TOP gene encoding a ribosomal protein or from a variant of a 5'-UTR of a TOP gene encoding a ribosomal protein. For example, the 5'-UTR element comprises or consists of a nucleic acid sequence, which is derived from a 5'-UTR of a nucleic acid sequence according to any of SEQ ID NOs: 67, 170, 193, 244, 259, 554, 650, 675, 700, 721, 913, 1016, 1063, 1120, 1138, and 1284-1360 of the patent application WO2013/143700, a corresponding RNA sequence, a homolog thereof, or a variant thereof as described herein, preferably lacking the 5'-TOP motif.

In some embodiments, the at least one 5'-UTR element, preferably a heterologous 5'-UTR element, comprises or consists of a nucleic acid sequence, which is derived from a 5'-UTR of a TOP gene encoding a ribosomal Large protein (RPL), preferably from RPL32 or RPL35A, or from a gene selected from the group consisting of HSD17B4,

ATP5A1, AIG1, COXC6 or ABCB7(MDR), or from a homolog, a fragment or variant of any one of these genes, preferably lacking the 5'TOP motif.

Preferably, the 5'-UTR element comprises or consists of a nucleic acid sequence, which is derived from a 5'-UTR of a TOP gene encoding a ribosomal Large protein (RPL) or from a homolog or variant of a 5'-UTR of a TOP gene encoding a ribosomal Large protein (RPL). For example, the 5'-UTR element comprises or consists of a nucleic acid sequence, which is derived from a 5'-UTR of a nucleic acid sequence according to any of SEQ ID NOs: 67, 259, 1284-1318, 1344, 1346, 1348-1354, 1357, 1358, 1421 and 1422 of the patent application WO2013/143700, a corresponding RNA sequence, a homolog thereof, or a variant thereof as described herein, preferably lacking the 5'-TOP motif.

In a preferred embodiment, the 5'-UTR element comprises or consists of a nucleic acid sequence, according SEQ ID NO: 1368 of the patent application WO2013/143700, or preferably to a corresponding RNA sequence, or fragments or variants thereof. In another preferred embodiment, the 5'-UTR element comprises or consists of a nucleic acid sequence according to SEQ ID NOs: 1412-1420 of the patent application WO2013/143700, or a corresponding RNA sequence, or fragments or variants thereof.

Alternatively, the 5'-UTR element preferably comprises or consists of a nucleic acid sequence according to SEQ ID NO: 30570 as disclosed herein, or a fragment or variant thereof.

In some embodiments, the RNA as used herein comprises a 5'-UTR element, which may be any 5'-UTR element described in WO2016/107877. In this context, the disclosure of WO2016/107877 relating to 3'-UTR elements/sequences is herewith incorporated by reference. Particularly preferred 5'-UTR elements are nucleic acid sequences according to SEQ ID NOs: 25 to 30 and SEQ ID NOs: 319 to 382 of the patent application WO2016/107877, or fragments or variants of these sequences. In this context, it is particularly preferred that the 5'-UTR element of the RNA sequence according to the present invention comprises or consists of a corresponding RNA sequence of the nucleic acid sequence according SEQ ID NOs: 25 to 30 and SEQ ID NOs: 319 to 382 of the patent application WO2016/107877.

In some embodiments, the RNA as used herein comprises a 5'-UTR element, which may be any 5'-UTR element as described in WO2017/036580. In this context, the disclosure of WO2017/036580 relating to 3'-UTR elements/sequences is herewith incorporated by reference. Particularly preferred 5'-UTR elements are nucleic acid sequences according to SEQ ID NOs: 1 to 151 of the patent application WO2017/036580, or fragments or variants of these sequences. In this context, it is particularly preferred that the 5'-UTR element of the RNA sequence according to the present invention comprises or consists of a corresponding RNA sequence of the nucleic acid sequence according to SEQ ID NOs: 1 to 151 of the patent application WO2017/036580.

Further, particulraly suitable 5'-UTRs are disclosed in WO2018172556. The disclosure of WO2018172556, in particular relating to SLC7A3, ATP5A1, RPL32, HSD17B4, NOSIP, ASAH1, RPL31, TUBB4B, UBQLN2, MP68 and NDUFA4, is included herewith by reference. Preferred 3'-UTR elements are derived from SLC7A3, ATP5A1, RPL32, HSD17B4, NOSIP, ASAH1, RPL31, TUBB4B, UBQLN2, MP68 or NDUFA4. Particularly preferred 3'-UTR elements are

nucleic acid sequences according to SEQ ID NOs: 1 to 22 as disclosed in WO2018172556, or fragments or variants of these sequences.

Preferably, the at least one 5'-UTR element as defined herein and the at least one 3'-UTR element as defined herein act synergistically to increase protein production from the at least one RNA sequence as described above.

Histone stem-loop

In a particularly preferred embodiment, the RNA as used herein comprises a histone stem-loop sequence/structure. The term "histone stem-loop" as used herein will be recognized and understood by the person of ordinary skill in the art, and are for example intended to refer to nucleic acid sequences that are predominantly found in histone histone mRNAs. Exemplary histone stem-loop sequences are described in Lopez et al. (Davila Lopez, M., & Samuelsson, T. (2008), RNA, 14(1)). The stem-loops in histone pre-mRNAs are typically followed by a purine-rich sequence known as the histone downstream element (HDE). These pre-mRNAs are processed in the nucleus by a single endonucleolytic cleavage approximately 5 nucleotides downstream of the stem-loop, catalyzed by the U7 snRNP through base pairing of the U7 snRNA with the HDE.

Such histone stem-loop sequences are preferably selected from histone stem-loop sequences as disclosed in WO2012/019780, the disclosure relating to histone stem-loop sequences/structures incorporated herewith by reference.

A histone stem-loop sequence suitable to be used within the present invention is preferably derived from formulae (I) or (II) of the patent application WO2012/019780. According to a further preferred embodiment the RNA as defined herein may comprise at least one histone stem-loop sequence derived from at least one of the specific formulae (Ia) or (IIa) of the patent application WO2012/019780.

A particular preferred histone stem-loop sequence is the nucleic acid sequence according to SEQ ID NO: 30586 disclosed herein, or the sequence CAAAGGCTCTTTTCAGAGCCACCA (according to SEQ ID NO: 43 of the patent application WO2015024667) or more preferably the corresponding RNA sequence CAAAGGCUCUUUUCAGAGCCACCA (according to SEQ ID NO: 44 of the patent application WO2015024667).

Any of the above modifications may be applied to the RNA (or RNA species) in the context of the present invention and may be, if suitable or necessary, be combined with each other in any combination, provided, these combinations of modifications do not interfere with each other in the respective nucleic acid sequence. A person skilled in the art will be able to take his choice accordingly.

Accordingly, the RNA as used herein may comprise a 5'-UTR and/or a 3'-UTR preferably comprising at least one histone stem-loop. The 3'-UTR of the RNA as used herein preferably comprises also a poly(A) and/or a poly(C) sequence, preferably as defined herein. The single elements of the 3' UTR may occur therein in any order from 5' to 3' along the sequence of the RNA sequence of the present invention. In addition, further elements as described herein, may also be contained, such as a stabilizing sequence as defined herein (e.g. derived from the UTR of a

globin gene), IRES sequences, miRNA binding sites etc. Each of the elements may also be repeated in the RNA as used herein at least once (particularly in di- or multicistronic constructs), preferably twice or more.

In a preferred embodiment, the RNA (RNA species) as used herein is monocistronic, bicistronic or multicistronic. Preferably, the RNA comprises at least two coding regions and at least one IRES sequence.

Accordingly, in a preferred embodiment, the RNA as used herein comprises, preferably in 5'- to 3'-direction:

a) a 5'-cap structure (cap0, cap1, cap2), preferably m7GpppN (cap0);

b) optionally, a 5'-UTR element as defined herein;

c) at least one coding sequence;

d) a 3'-UTR element as defined herein;

e) optionally, a poly(A)sequence, preferably comprising 10 to 200, 10 to 100, 40 to 80 or 50 to 70 adenine nucleotides, more preferably comprising 64 adenine nucleotides;

f) optionally, a poly(C)sequence, preferably consisting of 10 to 200, 10 to 100, 20 to 70, 20 to 60 or 10 to 40 cytosine nucleotides, more preferably comprising 30 cytosine nucleotides;

g) optionally, a histone stem-loop;

h) optionally, an additional poly(A) sequence;

or a fragment or variant of any of these nucleic acid sequences.

In certain embodiments, the RNA (or RNA species) as used herein comprises, preferably in 5' to 3' direction, the following elements:

a) a 5'-cap structure, preferably m7GpppN,

b) at least one coding region;

c) a 3'-UTR element comprising a nucleic acid sequence, which is derived from an alpha-globin gene, preferably comprising the nucleic acid sequence according to SEQ ID NO: 30578, or a homolog, a fragment or a variant of any one of these nucleic acid sequences,

d) a poly(A)sequence, preferably comprising 10 to 200, 10 to 100, 40 to 80 or 50 to 70 adenine nucleotides, more preferably comprising 64 adenine nucleotides,

e) a poly(C)sequence, preferably consisting of 10 to 200, 10 to 100, 20 to 70, 20 to 60 or 10 to 40 cytosine nucleotides, more preferably comprising 30 cytosine nucleotides, and

f) a histone stem-loop, preferably comprising the nucleic acid sequence according to SEQ ID NO. 30586, or a fragment or variant of any one of these nucleic acid sequences.

According to some preferred embodiments, the RNA (or RNA species) as used herein comprises, preferably in 5' to 3' direction, the following elements:

a) a 5'-cap structure, preferably m7GpppN,

b) a 5'-UTR element, which comprises a nucleic acid sequence, which is derived from the 5'-UTR of a TOP gene, preferably comprising a nucleic acid sequence according to SEQ ID NO: 30570, or a homolog, a fragment or a variant of any one of these nucleic acid sequences,

c) at least one coding region,

d) a 3'-UTR element comprising a nucleic acid sequence, which is derived from an albumin gene, preferably comprising the corresponding RNA sequence of the nucleic acid sequence according to SEQ ID NO: 30582 or 30584, or a homolog, a fragment or a variant of any one of these nucleic acid sequences,

e) a poly(A)sequence, preferably comprising 10 to 200, 10 to 100, 40 to 80 or 50 to 70 adenine nucleotides, more preferably comprising 64 adenine nucleotides,

f) a poly(C)sequence, preferably consisting of 10 to 200, 10 to 100, 20 to 70, 20 to 60 or 10 to 40 cytosine nucleotides, more preferably comprising 30 cytosine nucleotides, and

g) a histone stem-loop, preferably comprising the nucleic acid sequence according to SEQ ID NO. 30586, or a fragment or variant of any one of these nucleic acid sequences.

The at least one coding region (or coding sequence) of the RNA (or RNA species) that is adapted according to the methods disclosed herein, preferably encodes a peptide or a protein.

In preferred embodiments, the at least one coding region encodes an antigen derived from a pathogen causing an infectious disease, or a fragment or variant of such an antigen. More preferably, the at least one coding region encodes a viral antigen, a bacterial antigen, a fungal antigen or a protozoan antigen, or a fragment or variant of any one of these antigens.

In this context, further preferred are antigens derived from pathogens selected from the group consisting of Acinetobacter baumannii, Anaplasma genus, Anaplasma phagocytophilum, Ancylostoma braziliense, Ancylostoma duodenale, Arcanobacterium haemolyticum, Ascaris lumbricoides, Aspergillus genus, Astroviridae, Babesia genus, Bacillus anthracis, Bacillus cereus, Bartonella henselae, BK virus, Blastocystis hominis, Blastomyces dermatitidis, Bordetella pertussis, Borrelia burgdorferi, Borrelia genus, Borrelia spp, Brucella genus, Brugia malayi, Bunyaviridae family, Burkholderia cepacia and other Burkholderia species, Burkholderia mallei, Burkholderia pseudomallei, Caliciviridae family, Campylobacter genus, Candida albicans, Candida spp, Chlamydia trachomatis, Chlamydophila pneumoniae, Chlamydophila psittaci, CJD prion, Clonorchis sinensis, Clostridium botulinum, Clostridium difficile, Clostridium perfringens, Clostridium perfringens, Clostridium spp, Clostridium tetani, Coccidioides spp, coronaviruses, Corynebacterium diphtheriae, Coxiella burnetii, Crimean-Congo hemorrhagic fever virus, Cryptococcus neoformans, Cryptosporidium genus, Cytomegalovirus (CMV), Dengue viruses (DEN-1, DEN-2, DEN-3 and DEN-4), Dientamoeba fragilis, Ebolavirus (EBOV), Echinococcus genus, Ehrlichia chaffeensis, Ehrlichia ewingii, Ehrlichia genus, Entamoeba histolytica, Enterococcus genus, Enterovirus genus, Enteroviruses, mainly Coxsackie A virus and Enterovirus 71 (EV71), Epidermophyton spp, Epstein-Barr Virus (EBV), Escherichia coli O157:H7, O111 and O104:H4, Fasciola hepatica and Fasciola gigantica, FFI prion, Filarioidea superfamily, Flaviviruses, Francisella tularensis, Fusobacterium genus, Geotrichum candidum, Giardia intestinalis, Gnathostoma spp, GSS prion, Guanarito virus, Haemophilus ducreyi, Haemophilus influenzae, Helicobacter pylori, Henipavirus (Hendra virus Nipah virus), Hepatitis A Virus, Hepatitis B Virus (HBV), Hepatitis C Virus (HCV), Hepatitis D Virus, Hepatitis E Virus, Herpes simplex virus 1 and 2 (HSV-1 and HSV-2), Histoplasma capsulatum, HIV (Human immunodeficiency virus), Hortaea werneckii, Human bocavirus (HBoV), Human herpesvirus 6 (HHV-6) and Human herpesvirus 7 (HHV-7), Human metapneumovirus (hMPV), Human papillomavirus (HPV), Human parainfluenza viruses (HPIV), Japanese encephalitis virus, JC virus, Junin virus, Kingella kingae, Klebsiella granulomatis, Kuru prion, Lassa virus, Legionella pneumophila, Leishmania genus, Leptospira genus, Listeria monocytogenes, Lymphocytic choriomeningitis virus

(LCMV), Machupo virus, Malassezia spp, Marburg virus, Measles virus, Metagonimus yokagawai, Microsporidia phylum, Molluscum contagiosum virus (MCV), Mumps virus, Mycobacterium leprae and Mycobacterium lepromatosis, Mycobacterium tuberculosis, Mycobacterium ulcerans, Mycoplasma pneumoniae, Naegleria fowleri, Necator americanus, Neisseria gonorrhoeae, Neisseria meningitidis, Norovirus, Nocardia asteroides, Nocardia spp, Onchocerca volvulus, Orientia tsutsugamushi, Orthomyxoviridae family (Influenza), Paracoccidioides brasiliensis, Paragonimus spp, Paragonimus westermani, Parvovirus B19, Pasteurella genus, Plasmodium genus, Pneumocystis jirovecii, Poliovirus, Rabies virus, Respiratory syncytial virus (RSV), Rhinovirus, rhinoviruses, Rickettsia akari, Rickettsia genus, Rickettsia prowazekii, Rickettsia rickettsii, Rickettsia typhi, Rift Valley fever virus, Rotavirus, Rubella virus, Sabia virus, Salmonella genus, Sarcoptes scabiei, SARS coronavirus, Schistosoma genus, Shigella genus, Sin Nombre virus, Hantavirus, Sporothrix schenckii, Staphylococcus genus, Staphylococcus genus, Streptococcus agalactiae, Streptococcus pneumoniae, Streptococcus pyogenes, Strongyloides stercoralis, Taenia genus, Taenia solium, Tick-borne encephalitis virus (TBEV), Toxocara canis or Toxocara cati, Toxoplasma gondii, Treponema pallidum, Trichinella spiralis, Trichomonas vaginalis, Trichophyton spp, Trichuris trichiura, Trypanosoma brucei, Trypanosoma cruzi, Ureaplasma urealyticum, Varicella zoster virus (VZV), Varicella zoster virus (VZV), Variola major or Variola minor, vCJD prion, Venezuelan equine encephalitis virus, Vibrio cholerae, West Nile virus, Western equine encephalitis virus, Wuchereria bancrofti, Yersinia enterocolitica, Yersinia pestis, and Yersinia pseudotuberculosis, preferably the pathogen is selected from the group consisting of influenza virus, respiratory syncytial virus (RSV), Herpes simplex virus (HSV), human Papilloma virus (HPV), Human immunodeficiency virus (HIV), Plasmodium, Staphylococcus aureus, Dengue virus, Chlamydia trachomatis, Cytomegalovirus (CMV), Hepatitis B virus (HBV), Mycobacterium tuberculosis, Rabies virus, Rotavirus and Yellow Fever Virus, Zika virus.

According to a particularly preferred embodiment, the at least one coding region encodes an antigen, or a fragment or variant thereof, derived from an influenza virus, more preferably from an influenza A, an influenza B, or an influenza C virus (strain) or a variant of any of these.

In this context it, is particularly preferred that the at least one coding region of the RNA used herein encodes at least one antigen, or a fragment or variant thereof, selected from the group consisting of hemagglutinin (HA), neuraminidase (NA), nucleoprotein (NP), matrix protein 1 (M1), matrix protein 2 (M2), non-structural protein 1 (NS1), non-structural protein 2 (NS2), nuclear export protein (NEP), polymerase acidic protein (PA), polymerase basic protein PB1, PB1-F2, or polymerase basic protein 2 (PB2) of an influenza virus or a variant thereof. In some embodiments, the at least one coding region of the RNA used herein encodes at least one antigenic peptide or protein, or a fragment or variant thereof, which is derived from any one of the influenza virus proteins mentioned herein.

In a preferred embodiment, the at least one coding region of the RNA used herein encodes at least one antigen selected from hemagglutinin (HA) and/or neuraminidase (NA) of an influenza virus, or a fragment or variant thereof. More preferably, the at least one coding region of the RNA used herein encodes at least one antigenic peptide or protein derived from hemagglutinin (HA) and/or neuraminidase (NA) of an influenza virus, or a fragment or variant thereof. In this context, the hemagglutinin (HA) and the neuraminidase (NA) may be chosen from the same influenza virus or from different influenza viruses (or different influenza virus strains, respectively).

Influenza antigens and antigenic peptides or proteins derived from influenza virus are described in international patent applications PCT/EP2016/075862 and PCT/EP2017/060663, which are hereby incorporated by reference in their entirety.

In preferred embodiments, the RNA (species) as used herein, preferably the original RNA (species) that is to be adapted according to the methods disclosed herein or the adapted RNA (species), comprises at least one coding region, which encodes a peptide or protein comprising or consisting of an amino acid sequence according to any one of SEQ ID NOs: SEQ ID NOs: 1-30504, 213713, 213738, 213739, 213787, 213792, 213797, 213802, 213996-214023, 214100-214127, 214212-214239, 214316-214343, 214420-214447, 214524-214551, 214628-214655, 214732-214759, 214836-214863, 214940-214967, 215044, 215049-215076, 215161, 215166-215193, 215278, 215283-215310, 215395, 215400-215427, 215512, 215517-215544 as described in PCT/EP2017/060663, or a fragment or variant of any one of these amino acid sequences.

According to some embodiments, the RNA (species) used herein, preferably the original RNA (species) that is to be adapted, comprises at least one coding region comprising or consisting of a nucleic acid sequence according to any one of SEQ ID NOs: 30505-213528, 213529-213557, 213740-213746, 213788, 213789, 213793, 213794, 213798, 213799, 213803, 213804, 214024-214051, 214128-214155, 214240-214267, 214344-214371, 214448-214475, 214552-214579, 214656-214683, 214760-214787, 214864-214891, 214968-214995, 215045, 215046, 215077-215104, 215162, 215163, 215194-215221, 215279, 215280, 215311-215338, 215396, 215397, 215428-215455, 215513, 215514, 215545-215572, 215629, 215632, 215638-215835, 215892, 215836-215889 as described in PCT/EP2017/060663, or a fragment or variant of any one of these nucleic acid sequences.

According to a further preferred embodiment, the RNA (species) used herein comprises at least one coding region encoding an antigen, or a fragment or variant thereof, derived from a Norovirus, preferably selected from a GI.1 to GI.17 Norovirus, GII.1 to GII.24 Norovirus, GIII.1 to GIII.4 Norovirus, GIV.1 to GIV.4 Norovirus and GV.1 to GV.4 Norovirus more preferably, from a Norovirus selected from the group consisting of GI.1 Norovirus and GII.4 Norovirus or a variant of any of these.

In this context it, is particularly preferred that the at least one coding region of the RNA used herein encodes at least one antigen, or a fragment or variant thereof, selected from the group consisting of Norovirus non-structural proteins NS1/NS2, NS3, NS4, NS5, NS6, NS7, Norovirus capsid protein VP1 and Norovirus capsid protein VP2 or a variant thereof. In some embodiments, the at least one coding region of the RNA used herein encodes at least one antigenic peptide or protein, or a fragment or variant thereof, which is derived from any one of the Norovirus proteins mentioned herein.

In a preferred embodiment, the at least one coding region of the RNA used herein encodes at least one antigen selected from Norovirus capsid protein VP1 or Norovirus capsid protein VP2 or a fragment or variant thereof. In this context, the Norovirus capsid protein VP1 and/or Norovirus capsid protein VP2 may be chosen from the same Norovirus or from different Noroviruses (or different Norovirus strains, respectively). Norovirus antigens and antigenic peptides or proteins derived from Norovirus are also described in international patent application PCT/EP2017/060673, which is hereby incorporated by reference in its entirety.

In preferred embodiments, the RNA (species) as used herein, preferably the original RNA (species) that is to be adapted according to the methods disclosed herein or the adapted RNA (species), comprises at least one coding region, which encodes a peptide or protein (derived from a Norovirus) comprising or consisting of an amino acid sequence according to any one of SEQ ID NOs: 1-4410 as described in PCT/EP2017/060673, or a fragment or variant of any one of these amino acid sequences.

According to some embodiments, the RNA (species) used herein, preferably the original RNA (species) that is to be adapted, comprises at least one coding region comprising or consisting of a nucleic acid sequence according to any one of SEQ ID NOs: 4411-39690, 39713-39746 as described in PCT/EP2017/060673, or a fragment or variant of any one of these nucleic acid sequences.

According to some embodiments, the RNA (species) used herein, preferably the original RNA (species) that is to be adapted or the adapted RNA (species), comprises at least one coding region encoding a tumor antigen. Suitably, the tumor antigen is selected from the list consisting of 5T4, 707-AP, 9D7, AFP, AlbZIP HPG1, alpha-5-beta-1-integrin, alpha-5-beta-6-integrin, alpha-actinin-4/m, alpha-methylacyl-coenzyme A racemase, ART-4, ARTC1/m, B7H4, BAGE-1, BCL-2, bcr/abl, beta-catenin/m, BING-4, BRCA1/m, BRCA2/m, CA 15-3/CA 27-29, CA 19-9, CA72-4, CA125, calreticulin, CAMEL, CASP-8/m, cathepsin B, cathepsin L, CD1, CD20, CD22, CD25, CDE30, CD33, CD4, CD52, CD55, CD56, CD80, CDC27/m, CDK4/m, CDKN2A/m, CEA, CLCA2, CML28, CML66, COA-1/m, coactosin-like protein, collage XXIII, COX-2, CT- 9/BRD6, Cten, cyclin B1, cyclin D1, cyp-B, CYPB1, DAM-10, DAM-6, DEK-CAN, EFTUD2/m, EGFR, ELF2/m, EMMPRIN, EpCam, EphA2, EphA3, ErbB3, ETV6-AML1, EZH2, FGF-5, FN, Frau-1, G250, GAGE-1, GAGE-2, GAGE-3, GAGE-4, GAGE-5, GAGE-6, GAGE7b, GAGE-8, GDEP, GnT-V, gpl OO, GPC3, GPNMB/m, HAGE, HAST-2, hepsin, Her2/neu, HERV-K-MEL, HLA-A*0201-R1 71, HLA-A1 1 /m, HLA-A2/m, HNE, homeobox NKX3.1, HOM-TES-14/SCP-1, HOM-TES-85, HPV-E6, HPV-E7, HSP70-2M, HST-2, hTERT, iCE, IGF-1 R, IL-13Ra2, IL-2R, IL-5, immature laminin receptor, kallikrein-2, kallikrein-4, Ki67, KIAA0205, KIAA0205/m, KK-LC-1, K-Ras/m, LAGE-A1, LDLR-FUT, MAGE-A1, MAGE-A2, MAGE-A3, MAGE-A4, MAGE-A6, MAGE-A9, MAGE-A10, MAGE-A12, MAGE-B1, MAGE-B2, MAGE-B3, MAGE-B4, MAGE-B5, MAGE-B6, MAGE-B10, MAGE-B16, MAGE-B17, MAGE-C1, MAGE-C2, MAGE-C3, MAGE-D1, MAGE-D2, MAGE-D4, MAGE- El, MAGE-E2, MAGE-F1, MAGE-H1, MAGEL2, mammaglobin A, MART-1/melan-A, MART-2, MART-2/m, matrix protein 22, MC1 R, M-CSF, MEI/m, mesothelin, MG50/PXDN, MMP1 1, MN/CA IX-antigen, MRP-3, MUC-1, MUC-2, MUM-1/m, MUM-2/m, MUM-3/m, myosin class I/m, NA88-A, N-acetylglucosaminyltransferase-V, Neo-PAP, Neo-PAP/m, NFYC/m, NGEP, NMP22, NPM/ALK, N-Ras/m, NSE, NY-ESO-B, NY-ESO-1, OA1, OFA- iLRP, OGT, OGT/m, OS-9, OS-9/m, osteocalcin, osteopontin, pi 5, p190 minor bcr-abl, p53, p53/m, PAGE-4, PAI-1 , PAI-2, PAP, PART-1, PATE, PDEF, Pim-1 -Kinase, Pin-1, Pml/PARalpha, POTE, PRAME, PRDX5/m, prostein, proteinase-3, PSA, PSCA, PSGR, PSM, PSMA, PTPRK m, RAGE-1, RBAF600/m, RHAMM/CD1 68, RU1, RU2, S-100, SAGE, SART-1, SART-2, SART-3, SCC, SIRT2/m, Sp17, SSX-1, SSX-2/HOM-MEL-40, SSX-4, STAMP-1, STEAP-1, survivin, survivin-2B, SYT-SSX-1, SYT-SSX-2, TA-90, TAG-72, TARP, TEL-AML1 , TGFbeta, TGFbetaRII, TGM-4, TPI/m, TRAG-3, TRG, TRP-1, TRP-2/6b, TRP/INT2, TRP-p8, tyrosinase, UPA, VEGFR1, VEGFR-2/FLK-1, and WT1.

In another particularly preferred embodiment, the RNA (species) used herein, preferably the original RNA (species) that is to be adapted or the adapted RNA (species), comprises at least one coding region comprising or consisting

of a nucleic acid sequence encoding at least one therapeutic peptide or protein which can be used inter alia in the treatment of e.g. metabolic or endocrine disorders. These and other proteins are understood to be therapeutic, as they are meant to treat the subject by replacing its defective endogenous production of a functional protein in sufficient amounts. Accordingly, such therapeutic proteins are typically mammalian, in particular human proteins.

Preferably, the therapeutic peptide or protein is selected from or derived from any therapeutic peptide or protein which is used or can be used for medical treatment (e.g. protein replacement), antibodies, T cell receptors, gene-editing proteins (e.g. Cas9).

In another particularly preferred embodiment, the RNA (species) as used herein, preferably the original RNA (species) that is to be adapted or the adapted RNA (species), encodes at least one therapeutic antibody selected from the group consisting of antibodies which are used inter alia for the treatment of cancer or tumor diseases, immune disorders, infectious diseases, Alzheimer's disease, asthma, and antibodies which are used for the treatment of diverse disorders, e.g. osteoporosis, tooth decay, idiopathic pulmonary fibrosis, focal segmental glomerulosclerosis, pain, muscular dystrophy, and Neovascular age-related macular degeneration.

In another particularly preferred embodiment the RNA (species) used herein, preferably the original RNA (species) that is to be adapted or the adapted RNA (species), comprises at least one coding region comprising or consisting of a nucleic acid sequence encoding at least one allergen associated with allergy or an allergic disease (allergens or allergenic antigens) and/or at least one autoimmune self-antigens (autoantigens).

Adapted RNA and vectors:

In a further aspect, the present invention relates to the RNA that is obtainable by the methods according to the invention. More specifically, the invention provides an adapted RNA (or an RNA (species) with an adapted sequence, respectively), which is obtainable by the method for modifying the retention time of an RNA on a chromatographic column, by the method for purifying at least one RNA species from a mixture of at least two RNA species, by the method for co-purifying at least two RNA species from a mixture comprising at least two RNA species, by the method for harmonizing the numbers of A and/or U nucleotides in the sequences of at least two RNA species, or by the method for providing an adapted RNA as described herein.

The (adapted) RNA according to the invention is preferably characterized in that its sequence was adapted by altering the number of A and/or U nucleotides with respect to the number of A and/or U nucleotides in the original RNA sequence. The original RNA sequence is preferably an RNA sequence as described herein, preferably the corresponding wild-type RNA sequence, more preferably a corresponding RNA sequence that was modified as described herein, e.g. G/C content modified or codon optimized.

In preferred embodiments, the adapted RNA according to the invention is characterized by any one of the features described herein, in particular with respect to the inventive methods.

In certain embodiments, the adapted RNA according to the invention is characterized by a ratio of the number of A nucleotides to the number of U nucleotides in the RNA sequence, which is in the range from 0.2 to 5, preferably

from 0.5 to 3, more preferably from 1 to 3, even more preferably from 1 to 2.5, even more preferably from 1.2 to 2, even more preferably from 1.4 to 2, even more preferably from 1.5 to 2, most preferably from 1.6 to 2.

According to a particularly preferred embodiment, the RNA according to the invention comprises at least one coding region comprising or consisting of a nucleic acid sequence according to any one of SEQ ID NO: 26 to 14079, 14080 to 16264, 16265 to 28640, 28641 to 30568, or a fragment or variant of any one of these nucleic acid sequences.

In a further aspect, the present invention concerns a vector comprising the adapted RNA sequence according to the invention. In preferred embodiments, the vector according to the invention is a DNA vector comprising a nucleic acid sequence corresponding to the adapted RNA sequence according to the invention, or a fragment or variant thereof. Alternatively, the vector according to the invention is a DNA vector comprising a nucleic acid sequence encoding the adapted RNA sequence according to the invention, or a fragment or variant thereof. Preferably, the vector according to the invention is a plasmid vector or a viral vector.

RNA production

The RNA according to the present invention may be prepared using any method known in the art, including synthetic methods such as e.g. solid phase RNA synthesis, as well as in vitro methods, such as RNA in vitro transcription reactions. Particularly preferred is RNA in vitro transcription.

The terms "RNA in vitro transcription" or "in vitro transcription" relate to a process, wherein RNA is synthesized in a cell-free system (in vitro). DNA, particularly plasmid DNA (or PCR product), is typically used as template for the generation of RNA transcripts. RNA may be obtained by DNA-dependent in vitro transcription of an appropriate DNA template, which according to the present invention is preferably a linearized plasmid DNA template. The promoter for controlling in vitro transcription can be any promoter for any DNA-dependent RNA polymerase. Particular examples of DNA-dependent RNA polymerases are the T7, T3, and SP6 RNA polymerases. A DNA template for in vitro RNA transcription may be obtained by cloning of a nucleic acid, in particular cDNA corresponding to the respective RNA to be in vitro transcribed, and introducing it into an appropriate vector for in vitro transcription, for example into plasmid DNA. In a preferred embodiment of the present invention the DNA template is linearized with a suitable restriction enzyme, before it is transcribed in vitro. The cDNA may be obtained by reverse transcription of mRNA or chemical synthesis. Moreover, the DNA template for in vitro RNA synthesis may also be obtained by gene synthesis.

Reagents used in RNA in vitro transcription typically include: a DNA template (linearized plasmid DNA or PCR product) with a promoter sequence that has a high binding affinity for its respective RNA polymerase such as bacteriophage-encoded RNA polymerases (T7, T3, SP6, or Syn5); ribonucleoside triphosphates (NTPs) for the four bases (adenine, cytosine, guanine and uracil); optionally, a cap analogue as defined herein (e.g. m7G(5')ppp(5')G (m7G)); optionally, further modified nucleotides as defined herein; a DNA-dependent RNA polymerase capable of binding to the promoter sequence within the DNA template (e.g. T7, T3, SP6, or Syn5 RNA polymerase); optionally, a ribonuclease (RNase) inhibitor to inactivate any contaminating RNase; optionally, a pyrophosphatase to degrade pyrophosphate, which may inhibit transcription; MgCl2, which supplies Mg2+ ions as a co-factor for the polymerase;

a buffer (Tris or HEPES) to maintain a suitable pH value, which can also contain antioxidants (e.g. DTT), and/or polyamines such as spermidine at optimal concentrations, or a buffer system as disclosed in WO2017/109161.

In embodiments, the nucleotide mixture used in RNA in vitro transcription may additionally contain modified nucleotides as defined herein. In embodiments, the nucleotide mixture (i.e. the fraction of each nucleotide in the mixture) may be optimized for the given RNA sequence, preferably as described WO2015/188933.

In embodiment where more than one different RNA as defined herein has to be produced, e.g. where 2, 3, 4, 5, 6, 7, 8, 9, 10 or even more different RNA molecules have to be produced, procedures as described in WO2017/109134 may be suitably used.

In a preferred embodiment, RNA production is performed under current good manufacturing practice (GMP), implementing various quality control steps on DNA and RNA level, according to WO2016/180430. The obtained RNA products are preferably purified using PureMessenger® (CureVac, Tübingen, Germany; RP-HPLC according to WO2008/077592) and/or tangential flow filtration (as described in WO2016/193206). In a preferred embodiment, the RNA, particularly the purified RNA is lyophilized according to WO2016/165831 or WO2011/069586 to yield a temperature stable RNA as defined herein. The RNA of the invention, particularly the purified RNA may also be lyophilized using spray-drying or spray-freeze drying according to WO2016/184575 or WO2016184576 to yield a temperature stable RNA as defined herein.

Method for providing an adapted RNA sequence

According to a further aspect, the invention also provides a method for providing an adapted RNA sequence, wherein the method preferably comprises the steps of

a) optionally, determining the total number of nucleotides in an original RNA sequence;

b) determining the number of A and/or U nucleotides in the original RNA sequence;

c) determining the codons in the original RNA sequence that can be replaced with at least one alternative codon without changing the encoded amino acid sequence; and

d) adjusting the number of A and/or U nucleotides in the RNA sequence to a pre-set number of A and/or U nucleotides by replacing at least one original codon with an alternative codon, wherein the alternative codon encodes the same amino acid as the original codon and is further characterized in a higher content of A and/or U nucleotides.

In preferred embodiments, step b) of the method further comprises determining the number of G and/or C nucleotides in the original RNA sequence.

Preferably, the method for providing an adapted RNA according to the invention further comprises a step e) which comprises producing the RNA. Preferably, the RNA is produced by any suitable method known in the art or by a method as described herein under the section 'RNA production'. More preferably, the method involves a step of in vitro transcription, preferably as described herein. In a particularly preferred embodiment, the method involves in vitro transcription and a step of purifying the obtained RNA in a chromatographic process, preferably an HPLC process, more preferably an RP-HPLC process, preferably as described herein.

In the following, the concept of the inventive method for providing an adapted RNA is further illustrated by describing a preferred embodiment (see also Example 6 herein). The basic principle of sequence adaptation, particularly of AU adaptation by changing the codon usage, is exemplarily also illustrated in Figure 2.

An automated *in silico* method (algorithm) may be used to set the number of any nucleotide in an RNA sequence to a defined value, without altering the amino acid sequence (step d) of the method). In the context of the invention, the automated *in silico* method may used for sequence adaptation, in particular adaptation of the number of A and/or U nucleotides (AU count)) of RNA sequences in order to allow harmonization of RNA mixtures for HPLC co-analysis and/or HPLC co-purification.

Table 1 below summarizes the codon changes (for each genetically encoded amino acid) in the coding sequence that may be applied to increase the number of A and/or U nucleotides of an RNA sequence, without changing the encoded amino acid sequence. Table 2 summarizes the codon changes (for each genetically encoded amino acid) in the coding sequence that may be applied to decrease the number of A and/or U nucleotides of an RNA sequence, without changing the encoded amino acid sequence.

Table 1: Codon changes that allow for increase in the number of A and/or U nucleotides

Amino acid	**codon**	**Codon for AU increase**	**Change in AU count**	**Amino acid**	**codon**	**Codon for AU increase**	**Change in AU count**
Ala	GCG	GCA/GCU	+1	Pro	CCG	CCU/CCA	+1
Ala	GCA	n.a.		Pro	CCA	n.a.	
Ala	GCU	n.a.		Pro	CCU	n.a.	
Ala	GCC	GCA/GCU	+1	Pro	CCC	CCU/CCA	+1
Cys	UGU	n.a.		Gln	CAG	CAA	+1
Cys	UGC	UGU	+1	Gln	CAA	n.a.	
Asp	GAU	n.a.		Arg	AGG	AGA	+1
Asp	GAC	GAU	+1	Arg	AGA	n.a.	
Glu	GAG	GAA	+1	Arg	CGG	CGU/CGA	+1
Glu	GAA	n.a.		Arg	CGG	AGA	+2
Phe	UUU	n.a.		Arg	CGA	AGA	+1
Phe	UUC	UUU	+1	Arg	CGU	AGA	+1
Gly	GGG	GGA/GGU	+1	Arg	CGC	CGU/CGA	+1
Gly	GGA	n.a.		Arg	CGC	AGA	+2
Gly	GGU	n.a.		Ser	AGU	n.a.	
Gly	GGC	GGA/GGU	+1	Ser	AGC	AGU/UCA/UCU	+1
His	CAU	n.a.		Ser	UCG	AGU/UCA/UCU	+1
His	CAC	CAU	+1	Ser	UCA	n.a.	
Ile	AUA	n.a.		Ser	UCU	n.a.	

Amino acid	codon	Codon for AU increase	Change in AU count	Amino acid	codon	Codon for AU increase	Change in AU count
Ile	AUU	n.a.		Ser	UCC	AGU/UCA/UCU	+1
Ile	AUC	AUA/AUU	+1	Thr	ACG	ACA/ACU	+1
Lys	AAG	AAA	+1	Thr	ACA	n.a.	
Lys	AAA	n.a.		Thr	ACU	n.a.	
Leu	UUG	UUA	+1	Thr	ACC	ACA/ACU	+1
Leu	UUA	n.a.		Val	GUG	GUA/GUU	+1
Leu	CUG	UUG/CUA/CUU	+1	Val	GUA	n.a.	
Leu	CUG	UUA	+2	Val	GUU	n.a.	
Leu	CUA	UUA	+1	Val	GUC	GUA/GUU	+1
Leu	CUU	UUA	+1	Trp	UGG	n.a.	
Leu	CUC	UUG/CUA/CUU	+1	Tyr	UAU	n.a.	
Leu	CUC	UUA	+2	Tyr	UAC	UAU	+1
Met	AUG	n.a.		Stop	UGA*	UAA	+1
Asn	AAU	n.a.		Stop	UAG	UAA	+1
Asn	AAC	n.a.		Stop	UAA	n.a.	--

n.a. = not applicable

Table 2: Codon changes that allow for decrease in the number of A and/or U nucleotides

Amino acid	codon	Codon for AU decrease	Change in AU count	Amino acid	codon	Codon for AU decrease	Change in AU count
Ala	GCG	n.a.		Pro	CCG	CCC	-1
Ala	GCA	GCG/GCC	-1	Pro	CCA	CCG	-1
Ala	GCU	GCG/GCC	-1	Pro	CCU	CCG	-1
Ala	GCC	n.a.		Pro	CCA	CCC	-2
Cys	UGU	UGC	-1	Pro	CCU	CCC	-2
Cys	UGC	n.a.		Pro	CCC	n.a.	
Asp	GAU	GAC	-1	Gln	CAG	n.a.	
Asp	GAC	n.a.		Gln	CAA	CAG	-1
Glu	GAG	n.a.		Arg	AGG	CGC/CGG	-1
Glu	GAA	GAG	-1	Arg	AGA	AGG/CGA/CGU	-1
Phe	UUU	UUC	-1	Arg	AGA	CGC/CGG	-2
Phe	UUC	n.a.		Arg	CGG	n.a.	
Gly	GGG	n.a.		Arg	CGA	CGC/CGG	-1
Gly	GGA	GGG/GGC	-1	Arg	CGU	CGC/CGG	-1
Gly	GGU	GGG/GGC	-1	Arg	CGC	n.a.	
Gly	GGC	n.a.		Ser	AGU	n.a.	

Amino acid	codon	Codon for AU decrease	Change in AU count	Amino acid	codon	Codon for AU decrease	Change in AU count
His	CAU	CAC	-1	Ser	AGC	n.a.	
His	CAC	n.a.		Ser	UCG	n.a.	
Ile	AUA	AUC	-1	Ser	UCA	UCG/AGC	-1
Ile	AUU	AUC	-1	Ser	UCU	UCG/AGC	-1
Ile	AUC	n.a.		Ser	UCC	n.a.	
Lys	AAG	n.a.		Thr	ACG	n.a.	
Lys	AAA	AAG	-1	Thr	ACA	ACG/ACC	-1
Leu	UUG	CUG/CUC	-1	Thr	ACU	ACG/ACC	-1
Leu	UUA	UUG/CUA/CUU	-1	Thr	ACC	n.a.	
Leu	UUA	CUG/CUC	-2	Val	GUG	n.a.	
Leu	CUG	n.a.		Val	GUA	GUG/GUC	-1
Leu	CUA	CUG/CUC	-1	Val	GUU	GUG/GUC	-1
Leu	CUU	CUG/CUC	-1	Val	GUC	n.a.	
Leu	CUC	n.a.		Trp	UGG	n.a.	
Met	AUG	n.a.		Tyr	UAU	UAC	-1
Asn	AAU	n.a.		Tyr	UAC	n.a.	
Asn	AAC	n.a.		Stop	UGA	n.a.	
				Stop	UAG	n.a.	
				Stop	UAA	UGA/UAG	-1

n.a. = not applicable

In the initial phase of the method, a matrix for each codon comprised in the RNA sequence to be adapted is preferably created, identifying possible codon exchanges, which do not result in a change of the encoded amino acid (herein referred to as "exchange matrix"). An exemplary "exchange matrix" is shown in Formula (I).

Formula (I):

$$\left.\begin{matrix} A & 1 \\ C & 1 \\ G & 1 \\ T & 1 \end{matrix}\right\} \text{CGA}$$
$$* \quad 4$$

Formula (I) shows that for codon "CGA" a change to an alternative codon offers the option of increasing the number of A nucleotides by 1 (e.g.: CGA -> AGA), offers the option of increasing the number of C nucleotides by 1 (e.g. CGA ->CGC), offers the option of increasing the number of G nucleotides by 1 (e.g. CGA -> CGG), and offers the option of increasing the number of T nucleotides by 1 (e.g. CGA -> CGT).

Exchange matrices are preferably generated for each individual codon in the RNA sequence to be adapted. Using said exchange matrices, the potential maximum number of the respective nucleotides (A and T(U) count,

respectively) in each codon may be determined (without changing the amino acid sequence). Accordingly, all codons of the sequence, which is to be adapted, are analyzed with respect to potential codon exchanges by step-wise iteration, wherein in each iteration step the corresponding codon is analysed using the respective exchange matrix (as outlined above) for potential nucleotide changes. For example, if no changes are theoretically possible in the respective codon, e.g. as in the case of "ATG" or "TGG", the corresponding exchange matrix as exemplarily shown in Formula (II) was used (* of exchange matrix = 0).

Formula (II):

$$\left.\begin{matrix} A & 0 \\ C & 0 \\ G & 0 \\ T & 0 \\ * & 0 \end{matrix}\right\} \text{ATG}$$

Formula (II) shows that for codon "ATG" a change to an alternative codon offers no option of increasing the number of A nucleotides, C nucleotides, G nucleotides or T nucleotides (as there are no alternative codons for ATG (Met)).

In cases where exchanges according to the respective exchange matrix (* > 0) are theoretically possible, the codon is further analysed if the change can be implemented under the premise that e.g. only codons that offer the option of increasing the number of A and/or T(U) nucleotides are adapted. Therefore the intersection between the target nucleotides (e.g. A and/or T(U)) and the nucleotides that potentially generate a positive result (that is, A and/or T(U) change; see e.g. Formula (I)) in the current exchange matrix are constructed. As a result, each codon may be categorized and grouped into one of three categories:

Category 1 (category "favourable"):
Potential codon exchanges allowing an increase in only one target nucleotide (in the present example A or T(U)). For example, the codon "GAC" (Asp) can be changed to "GAU" (Asp) in order to increase the number of A and T(U) nucleotides. No further analysis is required since that modification does not have any further impact (besides the one mentioned above) on the number of A and T(U) nucleotides.

Category 2 (category "possible"):
Potential codon exchanges allowing the increase in both target nucleotides (in the present example A and T(U)). For example, codon "GCA" can be changed to "GCU", which would increase the T(U) count but at the same time decrease the A count. Accordingly, further analysis would be required with respect to codons belonging to this the category in order to decide, whether the number of one of the two target nucleotides (T(U)) in this example should be increased at the expense of a reduction of the number of the other target nucleotide (A).

Category 3 (category "impossible"):
Codons in the RNA sequence, for which no alternative codons exist (* = 0). Examples for this category 3 are ATG (Met; start codon) or "UGG" (Trp).

All codons of the original sequence are preferably categorized in that manner. It may then preferably be calculated how many potential nucleotide exchanges have been identified for all target nucleotides (A, T(U)).

In preferred embodiments, the method comprises adapting the RNA sequence using only codons from category 1 as defined herein. In some embodiments, codons from category 2 may also be used (for example, as soon as all codons from category 1 have been used), thus offering additional adaptation possibilities for A and T(U) counts. If category 2 is required in order to achieve the desired nucleotide counts, calculation of the following ratio may be used for identifying the exchange nucleotide (nucleotide A or T(U)):

$$\frac{c_i}{x_i - p_i}$$

wherein i represents the corresponding target, c_i is the count of possible adaptation positions of i in category 2, x_i is the desired threshold for i, p_i the count for the already changed identified adaptation positions. All calculated ratios are preferably ranked and starting from lowest to highest ratio, the changes from category 2 are applied, until the desired threshold has been reached or until all the possible exchanges from category 2 have been performed. This procedure is preferably carried out iteratively for any RNA sequence, where the target nucleotide count, e.g. the target number of A and/or U nucleotides, cannot be achieved by only using exchanges according to category 1.

In cases where the desired target nucleotide count cannot be achieved (as all alternative codons from category 1 and 2 have been used, which means that no further changes are possible), an adapted sequence is preferably generated that is matching the target nucleotide count as close as possible.

According to some embodiments of the method, one of the following improvements may be implemented therein:

1. Besides the basic equal distribution, which is used in the embodiment described above and which is based on the exchange possibilities, other distribution models may also be envisaged, such as normal distributed, first occurrences distribution, last occurrences distribution or random-based distribution. Alternatively, the mean of the possible changes or median of the possible changes may be determined and all exchanges may be arranged around these values.
2. The exchange matrix may further contain additional information about the codon for the target sequence (e.g. codon usage etc.). This creates a further criteria for the question of whether a codon exchange is desirable or not, facilitating adaptation to a specific codon usage or a different nucleotide ratio in the target sequence.
3. A third category may further be implemented by sequences or motifs, which should be avoided by an exchange (e.g. a recognition motif of a restriction enzyme, promotor sequences or sequences building not desired secondary structures, etc.).
4. Automated binning of input sequences may be performed, based on their length and the occurrence of the desired target nucleotides in order to identify optimal nucleotide counts for A and/or U adaptation.

Any one of the steps or combination of steps described with respect to the method for providing an adapted RNA according to the invention may preferably be applied to the other methods described herein, in particular to the

method for modifying the retention time of an RNA on a chromatographic column, to the method for purifying at least one RNA species from a mixture of at least two RNA species, to the method for co-purifying at least two RNA species from a mixture comprising at least two RNA species, or to the method for harmonizing the numbers of A and/or U nucleotides in the sequences of at least two RNA species.

Composition and vaccine

In a further aspect, the present invention provides a composition comprising at least one adapted RNA sequence as described herein and, optionally, a pharmaceutically acceptable carrier. The inventive composition comprising the at least one adapted RNA as described herein is preferably a pharmaceutical composition or a vaccine as described herein. According to a preferred embodiment, the inventive composition is a polyvalent vaccine, preferably as described herein.

In preferred embodiments, the composition according to the invention comprises an RNA obtainable by a method as described herein, in particular by the method for modifying the retention time of an RNA on a chromatographic column, the method for purifying at least one RNA species from a mixture of at least two RNA species, the method for co-purifying at least two RNA species from a mixture comprising at least two RNA species, the method for harmonizing the numbers of A and/or U nucleotides in the sequences of at least two RNA species, or the method for providing an adapted RNA as described herein.

According to a preferred embodiment, the composition according to the invention comprises at least two RNA species, wherein the sequence of at least one RNA species is adapted by altering the number of A and/or U nucleotides in the sequence of the at least one RNA species with respect to the number of A and/or U nucleotides in the original RNA sequence. More preferably, the composition comprises at least two RNA species, wherein at least one RNA species is an adapted RNA as obtained by any one of the methods according to the invention, preferably by the method for modifying the retention time of an RNA on a chromatographic column, the method for purifying at least one RNA species from a mixture of at least two RNA species, the method for co-purifying at least two RNA species from a mixture comprising at least two RNA species, the method for harmonizing the numbers of A and/or U nucleotides in the sequences of at least two RNA species, or the method for providing an adapted RNA as described herein. More preferably, the numbers of A and/or U nucleotides in the sequences of the at least two RNA species comprised in the inventive composition are harmonized as described herein.

In preferred embodiments, the composition according to the invention comprises at least two RNA species, e.g., at least 3, at least 4, at least 5, at least 6, at least 7, at least 8, at least 9, or at least 10 RNA species, wherein the sequence of at least one RNA species is adapted by altering the number of A and/or U nucleotides in the sequence of the at least one RNA species with respect to the number of A and/or U nucleotides in the original RNA sequence, wherein the numbers of A and/or U nucleotides in the two (or more) RNA species preferably differ from each other by not more than 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 110, 120, 130, 140, 150. In preferred embodiments, the numbers of A and/or U nucleotides in the RNA species preferably differ from each other by not more than 2, more preferably by not more than 3, even more preferably by not more than 4, even more preferably by not more than 5, most preferably by not more than 6. In a further preferred embodiment, the numbers of A and/or U nucleotides in the

RNA species differ from each other by not more than 10, preferably by not more than 20, more preferably by not more than 30, even more preferably by not more than 40, most preferably by not more than 50. It is particularly preferred, that the numbers of A and/or U nucleotides in the two or more RNA species are identical.

Preferably, the inventive composition comprises or consists of at least one adapted RNA (species) as described herein and a pharmaceutically acceptable carrier. The expression "pharmaceutically acceptable carrier" as used herein preferably includes the liquid or non-liquid basis of the inventive composition, which is preferably a pharmaceutical composition or a vaccine. If the inventive composition is provided in liquid form, the carrier will preferably be water, typically pyrogen-free water; isotonic saline or buffered (aqueous) solutions, e.g. phosphate, citrate etc. buffered solutions. Water or preferably a buffer, more preferably an aqueous buffer, may be used, containing a sodium salt, preferably at least 50 mM of a sodium salt, a calcium salt, preferably at least 0.01 mM of a calcium salt, and optionally a potassium salt, preferably at least 3 mM of a potassium salt. According to a preferred embodiment, the sodium, calcium and, optionally, potassium salts may occur in the form of their halogenides, e.g. chlorides, iodides, or bromides, in the form of their hydroxides, carbonates, hydrogen carbonates, or sulfates, etc. Without being limited thereto, examples of sodium salts include e.g. NaCl, NaI, NaBr, Na_2CO_3, $NaHCO_3$, Na_2SO_4, examples of the optional potassium salts include e.g. KCl, KI, KBr, K_2CO_3, $KHCO_3$, K_2SO_4, and examples of calcium salts include e.g. $CaCl_2$, CaI_2, $CaBr_2$, $CaCO_3$, $CaSO_4$, $Ca(OH)_2$. Furthermore, organic anions of the aforementioned cations may be contained in the buffer.

Furthermore, one or more compatible solid or liquid fillers or diluents or encapsulating compounds may be used as well, which are suitable for administration to a person. The term "compatible" as used herein means that the constituents of the inventive composition are capable of being mixed with the at least one (adapted) RNA (species) of the composition, in such a manner that no interaction occurs, which would substantially reduce the biological activity or the pharmaceutical effectiveness of the inventive composition under typical use conditions. Pharmaceutically acceptable carriers, fillers and diluents must, of course, have sufficiently high purity and sufficiently low toxicity to make them suitable for administration to a person to be treated. Some examples of compounds which can be used as pharmaceutically acceptable carriers, fillers or constituents thereof are sugars, such as, for example, lactose, glucose, trehalose and sucrose; starches, such as, for example, corn starch or potato starch; dextrose; cellulose and its derivatives, such as, for example, sodium carboxymethylcellulose, ethylcellulose, cellulose acetate; powdered tragacanth; malt; gelatin; tallow; solid glidants, such as, for example, stearic acid, magnesium stearate; calcium sulfate; vegetable oils, such as, for example, groundnut oil, cottonseed oil, sesame oil, olive oil, corn oil and oil from theobroma; polyols, such as, for example, polypropylene glycol, glycerol, sorbitol, mannitol and polyethylene glycol; alginic acid.

Further additives, which may be included in the composition are emulsifiers, such as, for example, Tween; wetting agents, such as, for example, sodium lauryl sulfate; colouring agents; taste-imparting agents, pharmaceutical carriers; tablet-forming agents; stabilizers; antioxidants; preservatives.

In a preferred embodiment the (adapted) RNA as defined herein, comprised in the composition, the pharmaceutical composition, the vaccine as defined herein, is complexed or at least partially complexed or associated with one or more cationic or polycationic compound preferably with cationic or polycationic polymer, cationic or polycationic

polysaccharide, cationic or polycationic lipid, cationic or polycationic protein, cationic or polycationic peptide, or any combinations thereof.

The term "cationic or polycationic compound" as used herein will be recognized and understood by the person of ordinary skill in the art, and are for example intended to refer to a charged molecule, which is positively charged at a pH value ranging from about 1 to 9, at a pH value ranging from about 3 to 8, at a pH value ranging from about 4 to 8, at a pH value ranging from about 5 to 8, more preferably at a pH value ranging from about 6 to 8, even more preferably at a pH value ranging from about 7 to 8, most preferably at a physiological pH, e.g. ranging from about 7.2 to about 7.5. Accordingly, a cationic component, e.g. a cationic peptide, cationic protein, cationic polymer, cationic polysaccharide, cationic lipid may be any positively charged compound or polymer which is positively charged under physiological conditions. A "cationic or polycationic peptide or protein" may contain at least one positively charged amino acid, or more than one positively charged amino acid, e.g. selected from Arg, His, Lys or Orn. Accordingly, "polycationic" components are also within the scope exhibiting more than one positive charge under the given conditions.

Cationic or polycationic compounds, being particularly preferred in this context may be selected from the following list of cationic or polycationic peptides or proteins of fragments thereof: protamine, nucleoline, spermine or spermidine, or other cationic peptides or proteins, such as poly-L-lysine (PLL), poly-arginine, basic polypeptides, cell penetrating peptides (CPPs), including HIV-binding peptides, HIV-1 Tat (HIV), Tat-derived peptides, Penetratin, VP22 derived or analog peptides, HSV VP22 (Herpes simplex), MAP, KALA or protein transduction domains (PTDs), PpT620, prolin-rich peptides, arginine-rich peptides, lysine-rich peptides, MPG-peptide(s), Pep-1, L-oligomers, Calcitonin peptide(s), Antennapedia-derived peptides, pAntp, pIsl, FGF, Lactoferrin, Transportan, Buforin-2, Bac715-24, SynB, SynB(1), pVEC, hCT-derived peptides, SAP, or histones. More preferably, the nucleic acid as defined herein, preferably the mRNA as defined herein, is complexed with one or more polycations, preferably with protamine or oligofectamine, most preferably with protamine. In this context protamine is particularly preferred.

Further preferred cationic or polycationic compounds, which can be used as transfection or complexation agent may include cationic polysaccharides, for example chitosan, polybrene etc.; cationic lipids, e.g. DOTMA, DMRIE, di-C14-amidine, DOTIM, SAINT, DC-Chol, BGTC, CTAP, DOPC, DODAP, DOPE: Dioleyl phosphatidylethanol-amine, DOSPA, DODAB, DOIC, DMEPC, DOGS, DIMRI, DOTAP, DC-6-14, CLIP1, CLIP6, CLIP9, oligofectamine; or cationic or polycationic polymers, e.g. modified polyaminoacids, such as beta-aminoacid-polymers or reversed polyamides, etc., modified polyethylenes, such as PVP etc., modified acrylates, such as pDMAEMA etc., modified amidoamines such as pAMAM etc., modified polybetaaminoester (PBAE), such as diamine end modified 1,4 butanediol diacrylate-co-5-amino-1-pentanol polymers, etc., dendrimers, such as polypropylamine dendrimers or pAMAM based dendrimers, etc., polyimine(s), such as PEI, poly(propyleneimine), etc., polyallylamine, sugar backbone based polymers, such as cyclodextrin based polymers, dextran based polymers, etc., silan backbone based polymers, such as PMOXA-PDMS copolymers, etc., blockpolymers consisting of a combination of one or more cationic blocks (e.g. selected from a cationic polymer as mentioned above) and of one or more hydrophilic or hydrophobic blocks (e.g. polyethyleneglycole); etc.

In embodiments, the composition or vaccine comprises at least one (adapted)RNA as described herein, which is complexed with one or more polycationic compounds and/or a polymeric carrier, and at least one free RNA, wherein the at least one complexed RNA is preferably identical to the at least one RNA according to the present invention.

The term "polymeric carrier" as used herein will be recognized and understood by the person of ordinary skill in the art, and are for example intended to refer to a compound that facilitates transport and/or complexation of another compound (cargo). A polymeric carrier is typically a carrier that is formed of a polymer. A polymeric carrier may be associated to its cargo (nucleic acid, RNA) by covalent or non-covalent interaction

In this context it is particularly preferred that the at least one RNA of the inventive composition is complexed at least partially with a cationic or polycationic compound and/or a polymeric carrier, preferably cationic proteins or peptides. In this context, the disclosure of WO2010/037539 and WO2012/113513 is incorporated herewith by reference. Partially means that only a part of the artificial nucleic acid is complexed with a cationic compound and that the rest of the artificial nucleic acid is (comprised in the inventive pharmaceutical composition, immunogenic composition) in uncomplexed form ("free").

Further preferred cationic or polycationic proteins or peptides may be derived from formula (Arg)l;(Lys)m;(His)n;(Orn)o;(Xaa)x of the patent application WO2009/030481, the disclosure of WO2009/030481 relating thereto incorporated herewith by reference.

According to a preferred embodiment, the composition of the present invention comprises the RNA as defined herein, and a polymeric carrier. A polymeric carrier used according to the invention might be a polymeric carrier formed by disulfide-crosslinked cationic components. The disulfide-crosslinked cationic components may be the same or different from each other. The polymeric carrier can also contain further components. It is also particularly preferred that the polymeric carrier used according to the present invention comprises mixtures of cationic peptides, proteins or polymers and optionally further components as defined herein, which are crosslinked by disulfide bonds as described herein.

In this context, polymeric carriers according to formula {(Arg)l;(Lys)m;(His)n;(Orn)o;(Xaa')x(Cys)y} and formula Cys,{(Arg)l;(Lys)m;(His)n;(Orn)o;(Xaa)x}Cys_2 of the patent application WO2012/013326 are preferred, the disclosure of WO2012/013326 relating thereto incorporated herewith by reference.

In a further particular embodiment, the polymeric carrier which may be used to complex the RNA as defined herein or any further nucleic acid comprised in the (pharmaceutical) composition or vaccine according to the invention may be derived from a polymeric carrier molecule according formula (L-P^1-S-[S-P^2-S]$_n$-S-P^3-L) of the patent application WO2011/026641, the disclosure of WO2011/026641 relating thereto incorporated herewith by reference.

In other embodiments, the composition, which is preferably a composition, a pharmaceutical composition or a vaccine, comprises at least one artificial nucleic acid as described herein, wherein the at least one artificial nucleic acid is complexed or associated with polymeric carriers and, optionally, with at least one lipid component as

described in the PCT applications PCT/EP2016/06322, PCT/EP2016/063227, PCT/EP2016/063229, PCT/EP2016/063226. In this context, the disclosures of PCT/EP2016/06322, PCT/EP2016/063227, PCT/EP2016/063229, PCT/EP2016/063226 is herewith incorporated by reference.

In preferred embodiments, the polymeric carrier compound is formed by, or comprises or consists of the peptide elements CysArg12Cys or CysArg12 or TrpArg12Cys. In particularly preferred embodiments, the polymeric carrier compound consists of a ($R_{12}C$)-($R_{12}C$) dimer, a ($WR_{12}C$)-($WR_{12}C$) dimer, or a (CR_{12})-($CR_{12}C$)-(CR_{12}) trimer, wherein the individual peptide elements in the dimer (e.g. (WR12C)), or the trimer (e.g. (CR12)), are connected via -SH groups.

In embodiments, where the complexed (adapted) RNA is complexed with cationic or polycationic peptides or proteins as the carrier compound, the nitrogen/phosphate ratio of the complexed RNA ranges from about 0.1 to about 20, or from about 0.2 to about 15, or from about 2 to about 15, or from about 2 to about 12, wherein the N/P ratio is defined as the mole ratio of the nitrogen atoms of the basic groups of the cationic peptide or polymer to the phosphate groups of the nucleic acid, preferably the RNA.

Accordingly, the composition as defined herein, comprising at least one RNA as defined herein, wherein the N/P ratio of the at least one artificial nucleic acid, preferably the RNA as defined herein, to the one or more cationic or polycationic compound as defined herein, preferably protamine, is in the range of about 0.1 to 20, including a range of about 0.3 to 4, of about 0.5 to 2, of about 0.7 to 2 and of about 0.7 to 1.5.

Suitably, the at least one RNA as defined herein, is complexed with one or more cationic or polycationic compounds as defined herein, in a weight ratio selected from a range of about 6:1 (w/w) to about 0.25:1 (w/w), more preferably from about 5:1 (w/w) to about 0.5:1 (w/w), even more preferably of about 4:1 (w/w) to about 1:1 (w/w) or of about 3:1 (w/w) to about 1:1 (w/w), and most preferably a ratio of about 3:1 (w/w) to about 2:1 (w/w) of nucleic acid to cationic or polycationic compound and/or with a polymeric carrier; or optionally in a nitrogen/phosphate ratio of nucleic acid to cationic or polycationic component and/or polymeric carrier in the range of about 0.1-10, preferably in a range of about 0.3-4 or 0.3-1, and most preferably in a range of about 0.5-1 or 0.7-1, and even most preferably in a range of about 0.3-0.9 or 0.5-0.9.

In this context it is particularly preferred that the at least one RNA as defined herein is complexed with protamine.

Suitably, the complexed RNA is complexed with protamine by addition of protamine-trehalose solution to the RNA sample at an RNA to protamine weight to weight ratio (w/w) of 2:1.

In preferred embodiments, the composition, the pharmaceutical composition, the vaccine as defined herein comprises the at least one RNA as defined herein which is complexed with one or more cationic or polycationic compounds (e.g. protamine), and at least one free RNA.

In preferred embodiments, the at least one complexed RNA (e.g. protamine complexed RNA) is identical to the at least one free RNA.

Preferably, the molar ratio of the RNA of the adjuvant component (e.g. protamine-complexed RNA) to the free nucleic acid, particularly the free RNA may be selected from a molar ratio of about 0.001:1 to about 1:0.001, including a ratio of about 1:1.

Preferably the ratio of complexed RNA of the adjuvant component (e.g. protamine-complexed RNA), to free nucleic acid, particularly the free RNA, may be selected from a range of about 5:1 (w/w) to about 1:10 (w/w), more preferably from a range of about 4:1 (w/w) to about 1:8 (w/w), even more preferably from a range of about 3:1 (w/w) to about 1:5 (w/w) or 1:3 (w/w), wherein the ratio is most preferably about 1:1 (w/w).

With respect to a composition comprising an adjuvant component (e.g. protamine-complexed RNA) and a free RNA component as defined herein, the disclosure of WO2009/144230 is incorporated herewith by reference. Such a composition comprising an adjuvant component as defined herein (e.g. protamine complexed RNA) and a free RNA component as defined herein may be generated using means and methods as disclosed in WO2016/165825.

In preferred embodiments, the composition, which is preferably a pharmaceutical composition, a vaccine, comprises at least one (adapted)RNA as described herein, wherein the at least one RNA is complexed or associated with one or more lipids (e.g. cationic lipids and/or neutral lipids), thereby forming liposomes, lipid nanoparticles (LNPs), lipoplexes, and/or nanoliposomes.

In the context of the present invention, the term "lipid nanoparticle", also referred to as "LNP", is not restricted to any particular morphology, and include any morphology generated when a cationic lipid and optionally one or more further lipids are combined, e.g. in an aqueous environment and/or in the presence of an RNA. For example, a liposome, a lipid complex, a lipoplex and the like are within the scope of a lipid nanoparticle (LNP).

LNPs typically comprise a cationic lipid and one or more excipient selected from neutral lipids, charged lipids, steroids and polymer conjugated lipids (e.g. PEGylated lipid). The RNA may be encapsulated in the lipid portion of the LNP or an aqueous space enveloped by some or the entire lipid portion of the LNP. The RNA or a portion thereof may also be associated and complexed with the LNP. An LNP may comprise any lipid capable of forming a particle to which the nucleic acids are attached, or in which the one or more nucleic acids are encapsulated. Preferably, the LNP comprising nucleic acids comprises one or more cationic lipids, and one or more stabilizing lipids. Stabilizing lipids include neutral lipids and PEGylated lipids.

In one embodiment, the LNP consists essentially of (i) at least one cationic lipid; (ii) a neutral lipid; (iii) a sterol, e.g. , cholesterol; and (iv) a PEG-lipid, e.g. PEG-DMG or PEG-cDMA, in a molar ratio of about 20-60% cationic lipid: 5-25% neutral lipid: 25-55% sterol; 0.5-15% PEG-lipid.

In that context, a preferred sterol is cholesterol. The sterol can be about 10 mol % to about 60 mol % or about 25 mol % to about 40 mol % of the lipid particle. In one embodiment, the sterol is about 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, or about 60 mol % of the total lipid present in the lipid particle. In another embodiment, the LNPs include from about 5% to about 50% on a molar basis of the sterol, e.g., about 15% to about 45%, about 20% to about 40%, about 48%, about 40%, about 38.5%, about 35%, about 34.4%, about 31.5% or about 31% on a molar basis (based upon 100% total moles of lipid in the lipid nanoparticle).

The cationic lipid of an LNP may be cationisable, i.e. it becomes protonated as the pH is lowered below the pK of the ionizable group of the lipid, but is progressively more neutral at higher pH values. At pH values below the pK, the lipid is then able to associate with negatively charged nucleic acids. In certain embodiments, the cationic lipid comprises a zwitterionic lipid that assumes a positive charge on pH decrease.

The LNP may comprise any further cationic or cationisable lipid, i.e. any of a number of lipid species which carry a net positive charge at a selective pH, such as physiological pH.

Such lipids include, but are not limited to, N,N-dioleyl-N,N-dimethylammonium chloride (DODAC); N-(2,3-dioleyloxy)propyl)-N,N,N-trimethylammonium chloride (DOTMA); N,N-distearyl-N,N-dimethylammonium bromide (DDAB); N-(2,3dioleoyloxy)propyl)-N,N,N-trimethylammonium chloride (DOTAP); 3-(N-(N',N'dimethylaminoethane)-carbamoyl)cholesterol (DC-Chol), N-(1-(2,3-dioleoyloxy)propyl)N-2-(sperminecarboxamido)ethyl)-N,N-dimethylammonium trifluoracetate (DOSPA), dioctadecylamidoglycyl carboxyspermine (DOGS), 1,2-dioleoyl-3-dimethylammonium propane (DODAP), N,N-dimethyl-2,3-dioleoyloxy)propylamine (DODMA), and N-(1,2dimyristyloxyprop-3-yl)-N,N-dimethyl-N-hydroxyethyl ammonium bromide (DMRIE).

Additionally, a number of commercial preparations of cationic lipids are available which can be used in the present invention. These include, for example, LIPOFECTIN® (commercially available cationic liposomes comprising DOTMA and 1,2-dioleoyl-sn-3phosphoethanolamine (DOPE), from GIBCO/BRL, Grand Island, N.Y.); LIPOFECTAMINE® (commercially available cationic liposomes comprising N-(1-(2,3dioleyloxy)propyl)-N-(2-(sperminecarboxamido)ethyl)-N,N-dimethylammonium trifluoroacetate (DOSPA) and (DOPE), from GIBCO/BRL); and TRANSFECTAM® (commercially available cationic lipids comprising dioctadecylamidoglycyl carboxyspermine (DOGS) in ethanol from Promega Corp., Madison, Wis.). The following lipids are cationic and have a positive charge at below physiological pH: DODAP, DODMA, DMDMA, 1,2-dilinoleyloxy-N,N-dimethylaminopropane (DLinDMA), 1,2-dilinolenyloxy-N,N-dimethylaminopropane (DLenDMA).

The further cationic lipid may also be an amino lipid. Representative amino lipids include, but are not limited to, 1,2-dilinoleyoxy-3-(dimethylamino)acetoxypropane (DLin-DAC), 1,2-dilinoleyoxy-3morpholinopropane (DLin-MA), 1,2-dilinoleoyl-3-dimethylaminopropane (DLinDAP), 1,2-dilinoleylthio-3-dimethylaminopropane (DLin-S-DMA), 1-linoleoyl-2-linoleyloxy-3dimethylaminopropane (DLin-2-DMAP), 1,2-dilinoleyloxy-3-trimethylaminopropane chloride salt (DLin-TMA.Cl), 1,2-dilinoleoyl-3-trimethylaminopropane chloride salt (DLin-TAP.Cl), 1,2-dilinoleyloxy-3-(N-methylpiperazino)propane (DLin-MPZ), 3-(N,Ndilinoleylamino)-1,2-propanediol (DLinAP), 3-(N,N-dioleylamino)-1,2-propanediol (DOAP), 1,2-dilinoleyloxo-3-(2-N,N-dimethylamino)ethoxypropane (DLin-EG-DMA), and 2,2-dilinoleyl-4-dimethylaminomethyl-[1,3]-dioxolane (DLin-K-DMA), 2,2- dilinoleyl-4-(2-dimethylaminoethyl)-[l,3]-dioxolane (DLin-KC2-DMA); dilinoleyl-methyl-4-dimethylaminobutyrate (DLin-MC3-DMA); MC3 (US20100324120).

Other suitable (cationic) lipids are disclosed in WO2009/086558, WO2009/127060, WO2010/048536, WO2010/054406, WO2010/088537, WO2010/129709, WO2011/153493, US2011/0256175, US2012/0128760, US2012/0027803, and US8158601. In that context, the disclosures of WO2009/086558, WO2009/127060, WO2010/048536, WO2010/054406, WO2010/088537, WO2010/129709, WO2011/153493, US2011/0256175, US2012/0128760, US2012/0027803, and US8158601 are incorporated herewith by reference.

In a particularly preferred embodiment the LNP comprises a cationic lipid with the formula (III) according to the patent application PCT/EP2016/075929. In this context, the disclosure of PCT/EP2016/075929 relating to cationic lipids is incorporated herewith by reference.

The amount of the permanently cationic lipid or lipidoid may be selected taking the amount of the nucleic acid cargo into account. In one embodiment, these amounts are selected such as to result in an N/P ratio of the nanoparticle(s) or of the composition in the range from about 0.1 to about 20. In this context, the N/P ratio is defined as the mole ratio of the nitrogen atoms ("N") of the basic nitrogen-containing groups of the lipid or lipidoid to the phosphate groups ("P") of the RNA which is used as cargo. The N/P ratio may be calculated on the basis that, for example, 1µg RNA typically contains about 3nmol phosphate residues, provided that the RNA exhibits a statistical distribution of bases. The "N"-value of the lipid or lipidoid may be calculated on the basis of its molecular weight and the relative content of permanently cationic and - if present - cationisable groups.

In certain embodiments, the LNP comprises one or more additional lipids which stabilize the formation of particles during their formation.

Suitable stabilizing lipids include neutral lipids and anionic lipids. The term "neutral lipid" refers to any one of a number of lipid species that exist in either an uncharged or neutral zwitterionic form at physiological pH. Representative neutral lipids include diacylphosphatidylcholines, diacylphosphatidylethanolamines, ceramides, sphingomyelins, dihydro sphingomyelins, cephalins, and cerebrosides.

Exemplary neutral lipids include, for example, distearoylphosphatidylcholine (DSPC), dioleoylphosphatidylcholine (DOPC), dipalmitoylphosphatidylcholine (DPPC), dioleoylphosphatidylglycerol (DOPG), dipalmitoylphosphatidylglycerol (DPPG), dioleoyl-phosphatidylethanolamine (DOPE), palmitoyloleoylphosphatidylcholine (POPC), palmitoyloleoyl-phosphatidylethanolamine (POPE) and dioleoyl-phosphatidylethanolamine 4-(N-maleimidomethyl)-cyclohexane-1carboxylate (DOPE-mal), dipalmitoyl phosphatidyl ethanolamine (DPPE), dimyristoylphosphoethanolamine (DMPE), distearoyl-phosphatidylethanolamine (DSPE), 16-O-monomethyl PE, 16-O-dimethyl PE, 18-1-trans PE, 1-stearioyl-2-oleoylphosphatidyethanol amine (SOPE), and 1,2-dielaidoyl-sn-glycero-3-phophoethanolamine (transDOPE). In one embodiment, the neutral lipid is 1,2-distearoyl-sn-glycero-3phosphocholine (DSPC).

In some embodiments, the LNPs comprise a neutral lipid selected from DSPC, DPPC, DMPC, DOPC, POPC, DOPE and SM. In various embodiments, the molar ratio of the cationic lipid to the neutral lipid ranges from about 2:1 to about 8:1.

LNP in vivo characteristics and behavior can be modified by addition of a hydrophilic polymer coating, e.g. polyethylene glycol (PEG), to the LNP surface to confer steric stabilization. Furthermore, LNPs can be used for specific targeting by attaching ligands (e.g. antibodies, peptides, and carbohydrates) to its surface or to the terminal end of the attached PEG chains (e.g. via PEGylated lipids).

In some embodiments, the LNPs comprise a polymer conjugated lipid. The term "polymer conjugated lipid" refers to a molecule comprising both a lipid portion and a polymer portion. An example of a polymer conjugated lipid is a PEGylated lipid. The term "PEGylated lipid" refers to a molecule comprising both a lipid portion and a polyethylene

glycol portion. PEGylated lipids are known in the art and include 1-(monomethoxy-polyethyleneglycol)-2,3-dimyristoylglycerol (PEG-s- DMG) and the like.

In certain embodiments, the LNP comprises an additional, stabilizing-lipid which is a polyethylene glycol-lipid (PEGylated lipid). Suitable polyethylene glycol-lipids include PEG-modified phosphatidylethanolamine, PEG-modified phosphatidic acid, PEG-modified ceramides (e.g. PEG-CerC14 or PEG-CerC20), PEG-modified dialkylamines, PEG-modified diacylglycerols, PEG-modified dialkylglycerols. Representative polyethylene glycol-lipids include PEG-c-DOMG, PEG-c-DMA, and PEG-s-DMG. In one embodiment, the polyethylene glycol-lipid is N-[(methoxy poly(ethylene glycol)$_{2000}$)carbamyl]-1,2-dimyristyloxlpropyl-3-amine (PEG-c-DMA). In one embodiment, the polyethylene glycol-lipid is PEG-c-DOMG). In other embodiments, the LNPs comprise a PEGylated diacylglycerol (PEG-DAG) such as 1-(monomethoxy-polyethyleneglycol)-2,3-dimyristoylglycerol (PEG-DMG), a PEGylated phosphatidylethanoloamine (PEG-PE), a PEG succinate diacylglycerol (PEG-S-DAG) such as 4-O-(2',3'-di(tetradecanoyloxy)propyl-1-O-(ω-methoxy(polyethoxy)ethyl)butanedioate (PEG-S-DMG), a PEGylated ceramide (PEG-cer), or a PEG dialkoxypropylcarbamate such as ω-methoxy(polyethoxy)ethyl-N-(2,3di(tetradecanoxy)propyl)carbamate or 2,3-di(tetradecanoxy)propyl-N-(ω-methoxy(polyethoxy)ethyl) carbamate. In various embodiments, the molar ratio of the cationic lipid to the PEGylated lipid ranges from about 100:1 to about 25:1.

The total amount of nucleic acid, particularly the RNA in the lipid nanoparticles varies and may be defined depending on the e.g. RNA to total lipid w/w ratio. In one embodiment of the invention the RNA to total lipid ratio is less than 0.06 w/w, preferably between 0.03 w/w and 0.04 w/w.

The composition, pharmaceutical composition, or vaccine comprising at least one (adapted) RNA according to the invention may be provided in liquid and or in dry (e.g. lyophylized) form. In a preferred embodiment, the RNA according to the invention or the composition is provided in lyophilized form. The RNA and the composition thus provide a possibility to store (irrespective of the ambient temperature and also without cooling) an RNA and a composition suitable for vaccination against, for example, an infectious disease, such as an infection with influenza virus or norovirus. Preferably, the at least one lyophilized RNA or composition is reconstituted in a suitable buffer, advantageously based on an aqueous carrier, e.g. Ringer-Lactate solution, prior to use, such as administration to a subject.

In a preferred embodiment, the composition is a pharmaceutical composition or a vaccine, which typically comprises a safe and effective amount of at least one (adapted) RNA as defined herein. As used herein, "safe and effective amount" means an amount of the RNA of the composition or vaccine as defined above, that is sufficient to significantly induce an immune response against, for example, an infectious disease as described herein, such as an infection with influenza virus. At the same time, however, a "safe and effective amount" is small enough to avoid serious side effects that is to say to permit a sensible relationship between advantage and risk. The determination of these limits typically lies within the scope of sensible medical judgment. In relation to the vaccine or composition, the expression "safe and effective amount" preferably means an amount of the RNA that is suitable for stimulating the adaptive immune system in such a manner that no excessive or damaging immune reactions are achieved but, preferably, also no such immune reactions below a measurable level. Such a "safe and effective amount" of the RNA of the composition or vaccine as defined above may furthermore be selected, for example in dependence of

the type of RNA, e.g. monocistronic, bi- or even multicistronic mRNA, since a bi- or even multicistronic mRNA may lead to a significantly higher expression of the encoded polypeptide(s) than use of an equal amount of a monocistronic mRNA. A "safe and effective amount" of the RNA of the composition or vaccine as defined above may furthermore vary in connection with the particular objective of the treatment and also with the age and physical condition of the patient to be treated, and similar factors, within the knowledge and experience of the accompanying doctor. The vaccine or composition according to the invention can be used according to the invention for human and also for veterinary medical purposes, as a pharmaceutical composition or as a vaccine.

In a preferred embodiment, the RNA of the composition, vaccine or kit of parts according to the invention is provided in lyophilized form. Preferably, the lyophilized RNA is reconstituted in a suitable buffer, advantageously based on an aqueous carrier, prior to administration, e.g. Ringer-Lactate solution, which is preferred, Ringer solution, a phosphate buffer solution.

According to a preferred embodiment, the buffer suitable for injection may be used as a carrier in the inventive vaccine or composition or for resuspending the inventive vaccine or the inventive composition. Such a buffer suitable for injection may be, for example, the liquid or non-liquid basis/carrier as described herein. Ringer-Lactate solution is particularly preferred as a liquid basis.

The choice of a pharmaceutically acceptable carrier is determined, in principle, by the manner, in which the vaccine or the composition is administered. The vaccine or composition can be administered, for example, systemically or locally. Routes for systemic administration in general include, for example, transdermal, oral, parenteral routes, including subcutaneous, intravenous, intramuscular, intraarterial, intradermal and intraperitoneal injections and/or intranasal administration routes. Routes for local administration in general include, for example, topical administration routes but also intradermal, transdermal, subcutaneous, or intramuscular injections or intralesional, intracranial, intrapulmonal, intracardial, and sublingual injections. More preferably, the vaccine or the composition may be administered by an intradermal, subcutaneous, or intramuscular route, preferably by injection, which may be needle-free and/or needle injection. Compositions/vaccines are therefore preferably formulated in liquid or solid form. The suitable amount of the vaccine or composition to be administered can be determined by routine experiments with animal models. Such models include, without implying any limitation, rabbit, sheep, mouse, rat, dog and non-human primate models. Preferred unit dose forms for injection include sterile solutions of water, physiological saline or mixtures thereof. The pH of such solutions should be adjusted to about 7.4. Suitable carriers for injection include hydrogels, devices for controlled or delayed release, polylactic acid and collagen matrices. Suitable pharmaceutically acceptable carriers for topical application include those which are suitable for use in lotions, creams, gels and the like. If the vaccine is to be administered perorally, tablets, capsules and the like are the preferred unit dose form. The pharmaceutically acceptable carriers for the preparation of unit dose forms which can be used for oral administration are well known in the prior art. The choice thereof will depend on secondary considerations such as taste, costs and storability, which are not critical for the purposes of the present invention, and can be made without difficulty by a person skilled in the art.

According to another embodiment, the (pharmaceutical) composition or the vaccine may comprise an adjuvant. An adjuvant may be used, for example, in order to enhance the immunostimulatory properties of the vaccine or

composition. In this context, an adjuvant may be understood as any compound, which is suitable to support administration and delivery of the vaccine or composition according to the invention. Furthermore, such an adjuvant may, without being bound thereto, initiate or increase an immune response of the innate immune system, i.e. a non-specific immune response. In other words, when administered, the vaccine or composition according to the invention typically initiates an adaptive immune response due to the at least one antigenic peptide or protein contained in the vaccine or composition. Additionally, the vaccine or composition according to the invention may generate an (supportive) innate immune response due to addition of an adjuvant as defined herein to the vaccine or composition according to the invention.

Such an adjuvant may be selected from any adjuvant known to a skilled person and suitable for the present case, i.e. supporting the induction of an immune response in a mammal. Preferably, the adjuvant may be selected from the group consisting of, without being limited thereto, TDM, MDP, muramyl dipeptide, pluronics, alum solution, aluminium hydroxide, ADJUMER™ (polyphosphazene); aluminium phosphate gel; glucans from algae; algammulin; aluminium hydroxide gel (alum); highly protein-adsorbing aluminium hydroxide gel; low viscosity aluminium hydroxide gel; AF or SPT (emulsion of squalane (5%), Tween 80 (0.2%), Pluronic L121 (1.25%), phosphate-buffered saline, pH 7.4); AVRIDINE™ (propanediamine); BAY R1005™ ((N-(2-deoxy-2-L-leucylamino-b-D-glucopyranosyl)-N-octadecyl-dodecanoyl-amide hydroacetate); CALCITRIOL™ (1-alpha,25-dihydroxy-vitamin D3); calcium phosphate gel; CAP™ (calcium phosphate nanoparticles); cholera holotoxin, cholera-toxin-A1-protein-A-D-fragment fusion protein, sub-unit B of the cholera toxin; CRL 1005 (block copolymer P1205); cytokine-containing liposomes; DDA (dimethyldioctadecylammonium bromide); DHEA (dehydroepiandrosterone); DMPC (dimyristoylphosphatidylcholine); DMPG (dimyristoylphosphatidylglycerol); DOC/alum complex (deoxycholic acid sodium salt); Freund's complete adjuvant; Freund's incomplete adjuvant; gamma inulin; Gerbu adjuvant (mixture of: i) N-acetylglucosaminyl-(P1-4)-N-acetylmuramyl-L-alanyl-D-glutamine (GMDP), ii) dimethyldioctadecylammonium chloride (DDA), iii) zinc-L-proline salt complex (ZnPro-8); GM-CSF); GMDP (N-acetylglucosaminyl-(b1-4)-N-acetylmuramyl-L-alanyl-D-isoglutamine); imiquimod (1-(2-methypropyl)-1H-imidazo[4,5-c]quinoline-4-amine); ImmTher™ (N-acetylglucosaminyl-N-acetylmuramyl-L-Ala-D-isoGlu-L-Ala-glycerol dipalmitate); DRVs (immunoliposomes prepared from dehydration-rehydration vesicles); interferon-gamma; interleukin-1beta; interleukin-2; interleukin-7; interleukin-12; ISCOMS™; ISCOPREP 7.0.3.™; liposomes; LOXORIBINE™ (7-allyl-8-oxoguanosine); LT oral adjuvant (E.coli labile enterotoxin-protoxin); microspheres and microparticles of any composition; MF59TM; (squalene-water emulsion); MONTANIDE ISA 51™ (purified incomplete Freund's adjuvant); MONTANIDE ISA 720™ (metabolisable oil adjuvant); MPL™ (3-Q-desacyl-4'-monophosphoryl lipid A); MTP-PE and MTP-PE liposomes ((N-acetyl-L-alanyl-D-isoglutaminyl-L-alanine-2-(1,2-dipalmitoyl-sn-glycero-3-(hydroxyphosphoryloxy))-ethylamide, monosodium salt); MURAMETIDE™ (Nac-Mur-L-Ala-D-Gln-OCH3); MURAPALMITINE™ and D-MURAPALMITINE™ (Nac-Mur-L-Thr-D-isoGIn-sn-glyceroldipalmitoyl); NAGO (neuraminidase-galactose oxidase); nanospheres or nanoparticles of any composition; NISVs (non-ionic surfactant vesicles); PLEURAN™ (beta-glucan); PLGA, PGA and PLA (homo- and co-polymers of lactic acid and glycolic acid; microspheres/nanospheres); PLURONIC L121™; PMMA (polymethyl methacrylate); PODDS™ (proteinoid microspheres); polyethylene carbamate derivatives; poly-rA: poly-rU (polyadenylic acid-polyuridylic acid complex); polysorbate 80 (Tween 80); protein cochleates (Avanti Polar Lipids, Inc., Alabaster, AL); STIMULON™ (QS-21); Quil-A (Quil-A saponin); S-28463 (4-amino-otec-dimethyl-2-ethoxymethyl-1H-imidazo[4,5 c]quinoline-1-ethanol); SAF-1™ ("Syntex adjuvant formulation"); Sendai proteoliposomes and Sendai-containing lipid matrices; Span-85

(sorbitan trioleate); Specol (emulsion of Marcol 52, Span 85 and Tween 85); squalene or Robane® (2,6,10,15,19,23-hexamethyltetracosan and 2,6,10,15,19,23-hexamethyl-2,6,10,14,18,22-tetracosahexane); stearyltyrosine (octadecyltyrosine hydrochloride); Theramid® (N-acetylglucosaminyl-N-acetylmuramyl-L-Ala-D-isoGlu-L-Ala-dipalmitoxypropylamide); Theronyl-MDP (Termurtide™ or [thr 1]-MDP; N-acetylmuramyl-L-threonyl-D-isoglutamine); Ty particles (Ty-VLPs or virus-like particles); Walter-Reed liposomes (liposomes containing lipid A adsorbed on aluminium hydroxide), and lipopeptides, including Pam3Cys, in particular aluminium salts, such as Adju-phos, Alhydrogel, Rehydragel; emulsions, including CFA, SAF, IFA, MF59, Provax, TiterMax, Montanide, Vaxfectin; copolymers, including Optivax (CRL1005), L121, Poloaxmer4010), etc.; liposomes, including Stealth, cochleates, including BIORAL; plant derived adjuvants, including QS21, Quil A, Iscomatrix, ISCOM; adjuvants suitable for costimulation including Tomatine, biopolymers, including PLG, PMM, Inulin; microbe derived adjuvants, including Romurtide, DETOX, MPL, CWS, Mannose, CpG nucleic acid sequences, CpG7909, ligands of human TLR 1-10, ligands of murine TLR 1-13, ISS-1018, IC31, Imidazoquinolines, Ampligen, Ribi529, IMOxine, IRIVs, VLPs, cholera toxin, heat-labile toxin, Pam3Cys, Flagellin, GPI anchor, LNFPIII/Lewis X, antimicrobial peptides, UC-1V150, RSV fusion protein, cdiGMP; and adjuvants suitable as antagonists including CGRP neuropeptide.

Further adjuvants that may be suitably used are also provided in WO2016/203025. With respect to suitable adjuvants that may be comprised in order to enhance the immunostimulatory properties of the composition according to the invention, the adjuvants according to Claim 7 and/or Claim 17 of WO2016/203025, and the disclosure relating thereto, are included herewith by reference.

In certain embodiments, the composition, the pharmaceutical composition, the immunogenic composition as defined herein may comprise at least one adjuvant, wherein the at least one adjuvant may be an nucleic acid adjuvant having the formula GlXmGn or nucleic acid adjuvant having the formula ClXmCn as disclosed in WO2008014979 and WO2009095226 respectively, the disclosure relating thereto incorporated herein by reference.

The vaccine or composition can additionally contain one or more auxiliary substances in order to further increase the immunogenicity. A synergistic action of the RNA of the composition or vaccine as defined herein and of an auxiliary substance, which may be optionally be co-formulated (or separately formulated) with the vaccine or composition as described above, is preferably achieved thereby. Depending on the various types of auxiliary substances, various mechanisms can come into consideration in this respect. For example, compounds that permit the maturation of dendritic cells (DCs), for example lipopolysaccharides, TNF-alpha or CD40 ligand form a first class of suitable auxiliary substances. In general, it is possible to use as auxiliary substance any agent that influences the immune system in the manner of a "danger signal" (LPS, GP96, etc.) or cytokines, such as GM-CFS, which allow an immune response produced by the immune-stimulating adjuvant according to the invention to be enhanced and/or influenced in a targeted manner. Particularly preferred auxiliary substances are cytokines, such as monokines, lymphokines, interleukins or chemokines, that – additional to induction of the adaptive immune response by the encoded at least one antigen - promote the innate immune response, such as IL-1, IL-2, IL-3, IL-4, IL-5, IL-6, IL-7, IL-8, IL-9, IL-10, IL-12, IL-13, IL-14, IL-15, IL-16, IL-17, IL-18, IL-19, IL-20, IL-21, IL-22, IL-23, IL-24, IL-25, IL-26, IL-27, IL-28, IL-29, IL-30, IL-31, IL-32, IL-33, INF-alpha, IFN-beta, INF-gamma, GM-CSF, G-CSF, M-CSF, LT-beta or TNF-alpha, growth factors, such as hGH. Preferably, such immunogenicity increasing

agents or compounds are provided separately (not co-formulated with the inventive vaccine or composition) and administered individually.

The vaccine or composition can also additionally contain any further compound, which is known to be immune-stimulating due to its binding affinity (as ligands) to human Toll-like receptors TLR1, TLR2, TLR3, TLR4, TLR5, TLR6, TLR7, TLR8, TLR9, TLR10, or due to its binding affinity (as ligands) to murine Toll-like receptors TLR1, TLR2, TLR3, TLR4, TLR5, TLR6, TLR7, TLR8, TLR9, TLR10, TLR11, TLR12 or TLR13.

Another class of compounds, which may be added to the vaccine or composition in this context, may be CpG nucleic acids, in particular CpG-RNA or CpG-DNA. A CpG-RNA or CpG-DNA can be a single-stranded CpG-DNA (ss CpG-DNA), a double-stranded CpG-DNA (dsDNA), a single-stranded CpG-RNA (ss CpG-RNA) or a double-stranded CpG-RNA (ds CpG-RNA). The CpG nucleic acid is preferably in the form of CpG-RNA, more preferably in the form of single-stranded CpG-RNA (ss CpG-RNA). The CpG nucleic acid preferably contains at least one or more (mitogenic) cytosine/guanine dinucleotide sequence(s) (CpG motif(s)). According to a first preferred alternative, at least one CpG motif contained in these sequences, that is to say the C (cytosine) and the G (guanine) of the CpG motif, is unmethylated. All further cytosines or guanines optionally contained in these sequences can be either methylated or unmethylated. According to a further preferred alternative, however, the C (cytosine) and the G (guanine) of the CpG motif can also be present in methylated form.

Preferably, the above compounds are formulated and administered separately from the above composition or vaccine (of the invention) containing the RNA according to the invention.

According to a preferred embodiment, the composition according to the invention comprises at least two RNA species, wherein the sequence of at least one RNA species is adapted by altering the number of A and/or U nucleotides in the sequence of the at least one RNA species with respect to the number of A and/or U nucleotides in the original RNA sequence, wherein at least one RNA species comprises at least one coding region encoding at least one peptide or protein derived from a Norovirus, or a fragment or a variant of said peptide or protein. Preferably, the at least one adapted RNA in the inventive composition comprises at least one coding region encoding a peptide or protein derived from a Norovirus, or a fragment or a variant of said peptide or protein. More preferably, the at least one coding region encodes a peptide or protein derived from a Norovirus as described herein, or a fragment or variant thereof. In some embodiments, the composition of the present invention is a Norovirus vaccine, more preferably a polyvalent Norovirus vaccine.

Suitable Norovirus antigens and antigenic peptides or proteins derived from Norovirus are described in international patent application PCT/EP2017/060673, which is hereby incorporated by reference in its entirety.

In preferred embodiments, the RNA (species) as used herein, preferably the original RNA (species) that is to be adapted or the adapted RNA (species), comprises at least one coding region, which encodes a peptide or protein (derived from a Norovirus) comprising or consisting of an amino acid sequence according to any one of SEQ ID NOs: 1-4410 as described in PCT/EP2017/060673, or a fragment or variant of any one of these amino acid sequences.

According to some embodiments, the RNA (species) used herein, preferably the original RNA (species) that is to be adapted, comprises at least one coding region comprising or consisting of a nucleic acid sequence according to any one of SEQ ID NOs: 4411-39690, 39713-39746 as described in PCT/EP2017/060673, or a fragment or variant of any one of these nucleic acid sequences.

In a further preferred embodiment, the composition according to the invention is a polyvalent Norovirus vaccine, which comprises at least two RNA species, wherein the sequence of at least one RNA species is adapted by altering the number of A and/or U nucleotides in the sequence of the at least one RNA species with respect to the number of A and/or U nucleotides in the original RNA sequence, wherein at least one RNA species comprises at least one coding region encoding at least one peptide or protein derived from an Norovirus, or a fragment or a variant of said peptide or protein, and wherein at least one RNA species comprises at least one coding region encoding a different peptide or protein derived from a Norovirus, or a fragment or a variant of said peptide or protein.

In a particularly preferred embodiment, the composition according to the invention is a polyvalent Norovirus vaccine, which comprises

a) at least one, or a plurality, or at least more than one RNA species comprising at least one coding region comprising or consisting of at least one nucleic acid sequence according to any one of SEQ ID NOs: 4411-39690, 39713-39746 as described in PCT/EP2017/060673, or a fragment or variant of any one of these nucleic acid sequences;

and

b) at least one, or a plurality, or at least more than one adapted RNA species comprising at least one coding region encoding at least one peptide or protein derived from an Norovirus, or a fragment or a variant of said peptide or protein, the at least one coding region preferably comprising or consisting of at least one adapted nucleic acid sequence encoding an amino acid sequence according to any one of SEQ ID NOs: 1-4410 as described in PCT/EP2017/060673, or a fragment or variant of any one of these nucleic acid sequences.

According to a particularly preferred embodiment, the composition according to the invention comprises at least two RNA species, wherein the sequence of at least one RNA species is adapted by altering the number of A and/or U nucleotides in the sequence of the at least one RNA species with respect to the number of A and/or U nucleotides in the original RNA sequence, wherein at least one RNA species comprises at least one coding region encoding at least one peptide or protein derived from an influenza virus, or a fragment or a variant of said peptide or protein. Preferably, the at least one adapted RNA in the inventive composition comprises at least one coding region encoding a peptide or protein derived from an influenza virus, or a fragment or a variant of said peptide or protein. More preferably, the at least one coding region encodes a peptide or protein derived from an influenza virus as described herein, or a fragment or variant thereof. In some embodiments, the composition of the present invention is an influenza vaccine, more preferably a polyvalent influenza vaccine.

In a preferred embodiment, the composition comprises at least two RNA species as described herein, wherein at least one RNA species comprises at least one coding region encoding at least one antigen selected from hemagglutinin (HA) and/or neuraminidase (NA) of an influenza virus, or a fragment or variant thereof. More

preferably, the at least one coding region of the RNA encodes at least one antigenic peptide or protein derived from hemagglutinin (HA) and/or neuraminidase (NA) of an influenza virus, or a fragment or variant thereof. In this context, the hemagglutinin (HA) and the neuraminidase (NA) may be chosen from the same influenza virus or from different influenza viruses (or different influenza virus strains, respectively).

Suitable influenza antigens and antigenic peptides or proteins derived from influenza virus are described in international patent applications PCT/EP2016/075862 and PCT/EP2017/060663, which are herein incorporated by reference in their entirety.

Preferably, the at least one RNA (species), preferably the original RNA (species) that is to be adapted or the adapted RNA (species), in the inventive composition comprises at least one coding region, which encodes a peptide or protein comprising or consisting of an amino acid sequence according to any one of SEQ ID NOs: 1-30504, 213713, 213738, 213739, 213787, 213792, 213797, 213802, 213996-214023, 214100-214127, 214212-214239, 214316-214343, 214420-214447, 214524-214551, 214628-214655, 214732-214759, 214836-214863, 214940-214967, 215044, 215049-215076, 215161, 215166-215193, 215278, 215283-215310, 215395, 215400-215427, 215512, 215517-215544 as described in PCT/EP2017/060663, or a fragment or variant of any one of these amino acid sequences.

According to some embodiments, the at least one RNA species, preferably the original RNA (species) that is to be adapted, in the inventive composition comprises at least one coding region comprising or consisting of a nucleic acid sequence according to any one of SEQ ID NOs: 30505-213528, 213529-213557, 213740-213746, 213788, 213789, 213793, 213794, 213798, 213799, 213803, 213804, 214024-214051, 214128-214155, 214240-214267, 214344-214371, 214448-214475, 214552-214579, 214656-214683, 214760-214787, 214864-214891, 214968-214995, 215045, 215046, 215077-215104, 215162, 215163, 215194-215221, 215279, 215280, 215311-215338, 215396, 215397, 215428-215455, 215513, 215514, 215545-215572, 215629, 215632, 215638-215835, 215892, 215836-215889 as described in PCT/EP2017/060663, or a fragment or variant of any one of these nucleic acid sequences.

According to a particularly preferred embodiment, the at least one RNA species, preferably an adapted RNA species, in the inventive composition comprises at least one coding region comprising or consisting of a nucleic acid sequence according to any one of SEQ ID NO: 26 to 14079, 14080 to 16264, 16265 to 28640, 28641 to 30568 as disclosed herein, or a fragment or variant of any one of these nucleic acid sequences.

In a further preferred embodiment, the composition according to the invention is a polyvalent influenza vaccine, which comprises at least two RNA species, wherein the sequence of at least one RNA species is adapted by altering the number of A and/or U nucleotides in the sequence of the at least one RNA species with respect to the number of A and/or U nucleotides in the original RNA sequence, wherein at least one RNA species comprises at least one coding region encoding at least one peptide or protein derived from an influenza virus, or a fragment or a variant of said peptide or protein, and wherein at least one RNA species comprises at least one coding region encoding a different peptide or protein derived from an influenza virus, or a fragment or a variant of said peptide or protein. More preferably, the composition comprises at least one RNA species encoding a peptide or protein derived from

influenza HA, or a fragment or variant thereof, and at least one other RNA species encoding a peptide or protein derived from influenza NA, or a fragment or variant thereof, wherein at least one of said RNA species is an adapted RNA as described herein.

In a particularly preferred embodiment, the composition according to the invention is a polyvalent influenza vaccine, which comprises

a) at least one, or a plurality, or at least more than one RNA species comprising at least one coding region comprising or consisting of

at least one nucleic acid sequence according to any one of SEQ ID NOs: 30505-213528, 213529-213557, 213740-213746, 213788, 213789, 213793, 213794, 213798, 213799, 213803, 213804, 214024-214051, 214128-214155, 214240-214267, 214344-214371, 214448-214475, 214552-214579, 214656-214683, 214760-214787, 214864-214891, 214968-214995, 215045, 215046, 215077-215104, 215162, 215163, 215194-215221, 215279, 215280, 215311-215338, 215396, 215397, 215428-215455, 215513, 215514, 215545-215572, 215629, 215632, 215638-215835, 215892, 215836-215889 as described in PCT/EP2017/060663, or a fragment or variant of any one of these nucleic acid sequences;

and

b) at least one, or a plurality, or at least more than one adapted RNA species comprising at least one coding region encoding at least one peptide or protein derived from an influenza virus, or a fragment or a variant of said peptide or protein, the at least one coding region preferably comprising or consisting of at least one nucleic acid sequence according to any one of SEQ ID NO: 26 to 14079, 14080 to 16264, 16265 to 28640, 28641 to 30568 as disclosed herein, or a fragment or variant of any one of these nucleic acid sequences.

Method for producing an RNA composition

A further aspect of the present invention concerns a method for producing an RNA composition as described herein. The method is preferably for producing an RNA composition, preferably a (polyvalent) vaccine, comprising at least two RNA species, wherein the sequence of at least one RNA species has been adapted by altering the number of A and/or U nucleotides in the sequence of the at least one adapted RNA species, preferably as described herein, more preferably as described herein with respect to the method for modifying the retention time of an RNA on a chromatographic column, to the method for purifying at least one RNA species from a mixture of at least two RNA species, to the method for co-purifying at least two RNA species from a mixture comprising at least two RNA species, to the method for harmonizing the numbers of A and/or U nucleotides in the sequences of at least two RNA species, to the method fro providing an adapted RNA or to the composition or vaccine as described herein.

In a preferred embodiment, the method provides a system, which allows for fast and efficient production of a composition comprising at least two RNA species, which are selected from a pool of a RNA species. Starting out from a pool of RNA species/sequences, the at least two RNA species of the composition can quickly be exchanged by the methods laid out herein. The concept is also illustrated by Fig. 3 and Example 7. For example, in the case of a polyvalent influenza vaccine, from an AU adapted RNA sequence pool, RNA species encoding antigens, e.g. HA and/or NA antigens, can easily be exchanged ("+": addition of new RNA species to the RNA mixture; "-": removal of RNA species from the mixture) for e.g. seasonal influenza vaccine production (see Fig. 3; A: season A; B: season B; C: season C). In that manner, the method allows for quick seasonal adaptation of the influenza vaccine. The

outlined concept is also suitable for other fast-adapting pathogens, e.g. Norovirus, and may be used to quickly adapt RNA based Norovirus vaccines.

The method may thus preferably involve the steps of adapting the sequence of at least one RNA sequence, which is comprised in the composition. Methods for producing RNA are known in the art and described herein.

Kits

According to another aspect, the present invention also provides kits, particularly kits of parts, comprising at least one (adapted) RNA (species) according to the invention, the composition comprising at least one (adapted) RNA (species) according to the invention, optionally a liquid vehicle for solubilising and optionally technical instructions with information on the administration and dosage of the (adapted) RNA (species) as described herein. Preferably, the (adapted) RNA (species) as described herein or the composition comprising at least one (adapted) RNA (species) according to the invention is provided in a separate part of the kit, wherein the (adapted) RNA (species) as described herein or the composition comprising at least one (adapted) RNA (species) according to the invention are preferably lyophilised. More preferably, the kit further contains as a part a vehicle for solubilising the RNA as described herein, the composition comprising at least one RNA (species) according to the invention, the vehicle preferably being Ringer-lactate solution. Any of the above kits may be used in a treatment or prophylaxis as defined herein. More preferably, any of the above kits may be used as a vaccine, preferably a vaccine against infection with an infectious disease, such as an influenza virus infection or Norovirus infection.

(Medical) Use and application:

The present invention furthermore provides several applications and uses of the (adapted) RNA (species) according to the invention, the composition/vaccine comprising at least one (adapted) RNA (species) according to the invention or of kits comprising same. In particular, the (pharmaceutical) composition or the vaccine may be used for human and also for veterinary medical purposes, preferably for human medical purposes, as a pharmaceutical composition in general or as a vaccine.

Consequently, in a further aspect, the present invention is directed to the first medical use of the (adapted) RNA (species) according to the invention, the composition/vaccine comprising at least one (adapted) RNA (species) according to the invention or the kit or kit of parts as defined herein as a medicament. In particular, the invention provides the use of at least one (adapted) RNA (species) as defined herein, or a fragment or variant thereof as described herein for the preparation of a medicament.

According to another aspect, the present invention is directed to the second medical use of the (adapted) RNA (species) according to the invention, the composition comprising at least one (adapted) RNA (species) according to the invention or the kit or kit of parts as described herein for the treatment of an infection with an influenza virus. In particular, the present invention provides the at least one (adapted) RNA (species) as described herein to be used for the preparation of a medicament, wherein the (adapted) RNA (species) as described herein is preferably formulated together with a pharmaceutically acceptable vehicle and an optionally additional adjuvant and an optionally additional further component as defined herein.

Further, the present invention is directed to the second medical use of the (adapted) RNA (species) according to the invention, the composition comprising at least one (adapted) RNA (species) according to the invention or the kit or kit of parts as described herein for the treatment of an infection with a Norovirus. In particular, the present invention provides the at least one (adapted) RNA (species) as described herein to be used for the preparation of a medicament, wherein the (adapted) RNA (species) as described herein is preferably formulated together with a pharmaceutically acceptable vehicle and an optionally additional adjuvant and an optionally additional further component as defined herein.

The composition or the vaccine comprising at least one (adapted) RNA (species) according to the invention can be administered, for example, systemically or locally. Routes for systemic administration in general include, for example, transdermal, oral, parenteral routes, including subcutaneous, intravenous, intramuscular, intraarterial, intradermal and intraperitoneal injections and/or intranasal administration routes. Routes for local administration in general include, for example, topical administration routes but also intradermal, transdermal, subcutaneous, or intramuscular injections or intralesional, intracranial, intrapulmonal, intracardial, and sublingual injections. More preferably, the vaccine may be administered by an intradermal, subcutaneous, or intramuscular route.

In a further aspect the invention provides a method of treating or preventing a disorder, wherein the disorder is preferably an infection with influenza virus or Norovirus, or a disorder related to an infection with influenza virus or Norovirus, wherein the method comprises administering to a subject in need thereof the (adapted) RNA (species) according to the invention, the inventive composition comprising at least one (adapted) RNA (species) according to the invention, or the inventive kit or kit of parts.

In particular, such a method may preferably comprise the steps of:

a) providing the (adapted) RNA (species) according to the invention, the vaccine/composition comprising at least one (adapted) RNA (species) according to the invention, or the inventive kit or kit of parts described herein; and

b) applying or administering the (adapted) RNA (species) according to the invention, the vaccine/composition comprising at least one (adapted) RNA (species) according to the invention, or the kit or kit of parts described herein to a tissue or an organism.

Brief description of the drawings

Figure 1: illustrates the technical problem associated with HPLC co-purification and/or co-analysis of RNA mixtures. Schematic drawings of HPLC histograms (RNA mixture comprising different RNA molecule species) are shown.

Figure 1A and 1B: RNA molecule species of an RNA mixture differ in their retention times. Impurities such as abortive sequences cannot be separated from each other. In addition, as the histograms partially overlap, quality attributes, such as integrity of the individual species, cannot be determined. Figure 1A shows the separate histograms for each RNA species in the mixture.

Figure 1C and 1D: In rare cases, three RNA species of an RNA mixture may have similar retention times, and impurities such as abortive sequences could be separated from each other. In addition, as the peaks entirely overlap, quality attributes, such as integrity of the whole RNA mixture, could be determined. Figure 1C shows the separate histograms for each RNA species in the mixture. Figure 1E: RNA species have different retention times. As the histograms are entirely separated from each other, quality attributes, such as integrity of individual RNA species of the RNA mixture, can be determined.

Figure 2: illustrates the basic principle of adapting a given nucleic acid sequence encoding a short protein (amino acid sequence: AWHPVAC) to either increase the AU count or decrease the AU count. Figure 2A: As initial step of the adaptation method/algorithm, all codons of the nucleic acid sequence are categorized and potential exchanges with alternative codons are allocated for each individual codon (codon changes that allow increase or decrease of AU count are listed in Table 1 and Table 2). In Figure 2A, codons labeled with an asterisk (*) may allow for an increase in AU count if the respective codon is changed accordingly; codons labeled with a hash (#) may allow for a decrease in AU count if the respective codon is changed accordingly; codons labeled with a cross ("x") do not lead to a change in AU count. Figure 2B shows the adaptation of the input nucleic acid sequence to an AU count increased target sequence. All four codons, which can potentially be replaced by alternative codons, are changed to the codon with a larger AU count (codons highlighted). Figure 2C shows the adaptation of the input nucleic acid sequence to an AU decreased target sequence. All three codons, which can potentially be replaced by alternative codons, are changed to the codon with a lower AU count (codons highlighted).

Figure 3: illustrates the modular design principle of the inventive multivalent/polyvalent mRNA vaccine platform (e.g. influenza), where RNA species can be exchanged rapidly without changing the manufacturing conditions (HPLC purification and/or HPLC analysis). From an AU adapted RNA sequence pool, RNA species encoding antigens, e.g. HA and/or NA antigens, can easily be exchanged ("+": addition of new RNA species to the RNA mixture; "-": removal of RNA species from the mixture) for e.g. seasonal influenza vaccine production (A: season A; B: season B; C: season C). The general concept can also be used for e.g. Norovirus antigens.

Figure 4: shows that the number of adenine nucleotides in an RNA sequence correlates with HPLC retention times. RNA species (1–4) encoding firefly luciferase with varying polyA sizes were generated and individually analyzed via HPLC. HPLC chromatograms were superimposed. 1=A25; 2=A35; 3=A50; 4=A64. A detailed description of the experiment is provided in Example 1.

Figure 5: shows that adaptation of the number of adenine nucleotides in RNA sequences enables harmonization of HPLC retention time. Figure 5A: Adenine adapted RNA sequences encoding HA-B Brisbane are shown (superimposed). Figure 5B: Adenine adapted RNA sequences encoding HA-B Phuket are shown (superimposed). (1) Non-adapted sequence; (2) 9 adenines introduced in

cds; (3) 9 adenine stretch added in the UTR or (4) 9 adenines added in polyA. A detailed description of the experiment is provided in Example 2.

Figure 6: shows that adaptation of the adenine count enables harmonization of HPLC retention time of RNA sequences encoding HA-A and RNA sequences encoding HA-B. RNA sequences encoding HA-B were adapted to match the A count of HA-A RNA sequences by increasing the A count accordingly. The asterisk indicates that the HPLC peaks completely overlap (declining slope of the peak determines retention time). Of note: For the purpose of Fig. 6, a distinction of individual chromatograms is not required. A detailed description of the experiment is provided in Example 2.

Figure 7: shows that HPLC is a particularly suitable method for co-analysis of an RNA mixture. RNA Mixtures of intact and degraded RNA at different ratios were analyzed via HPLC. Figure 7A: Overlay of HPLC chromatograms of different RNA mixtures showing the amount of intact RNA. Figure 7B: Overlay of HPLC chromatograms of different RNA mixtures showing the amount of degraded RNA. Of note: For the purpose of Fig. 7, a distinction of individual chromatograms is not required. A detailed description of the experiment is provided in Example 3.

Figure 8: shows that an adenine adapted RNA mixture (encoding three different NA antigens) generates one discrete HPLC peak, suitable for co-analysis and co-purification. Figure 8A: HPLC histograms for individual non-adapted RNA sequences (1-3) and the resulting non-adapted RNA mixture (4) are shown. Figure 8B: HPLC histogram for individual sequence adapted RNA sequences (1-3) and the resulting harmonized RNA mixture (4) are shown. A detailed description of the experiment is provided in Example 4.

Figure 9: shows a further illustration of the technical problem. RNA species encoding HA and NA of Influenza virus A and B have partially overlapping HPLC chromatograms due to different AU counts, illustrating the problem in the art that co-purification and/or co-analysis of such an RNA mixture (comprising all seven RNA species) would be technically impossible. 1= RNA species encoding neuraminidase of Influenza virus B (NA-B, Brisbane); 2= two RNA species encoding neuraminidase of Influenza virus A (NA-A, Hongkong, California); 3= two RNA species encoding hemagglutinin of Influenza virus B (HA-B, Brisbane, Phuket); two RNA molecule species encoding hemagglutinin of Influenza virus A (HA-A, Hongkong, California). A detailed description of the experiment is provided in Example 5.

Figure 10: shows that adenine and/or uracil count correlates with HPLC retention times.

Figure 10A: Total number/count of A and/or U and the content (%) of AU of different HA and NA RNA species plotted against the RNA retention time on HPLC is shown. The total number/count of A and/or U correlates with HPLC retention time, whereas the content of AU does not correlate with the HPLC retention time.

Figure 10B: Total number/count of G and/or C and the content (%) of GC of different HA and NA RNA species plotted against the RNA retention time on HPLC is shown. Both guanine/cytosine (G/C) count and content (%) do not correlate with HPLC retention times. A detailed description of the experiment is provided in Example 5.

Figure 11: illustrates the sequence adaptation strategy for an RNA mixture comprising three different RNA molecule species. Two product peaks have to be shifted by 17 AU or 40 AU (shifted peaks: dashed lines) to obtain a sequence adapted RNA mixture, where each of the three components can be co-analyzed on HPLC (AU count difference for the adapted sequences: 70 AU). A detailed description of the experiment is provided in Example 6.

Figure 12: overlay of HPLC chromatograms of constructs with different AU count on a monolithic ethylvinylbenzene-divinylbenzene copolymer. Flow rates and gradient profiles indicated. First peak corresponds to a construct with AU count 824, second peak corresponds to a construct with AU count 894, third peak corresponds to a construct with AU count 1137, fourth peak corresponds to a construct with AU count 1139, last peak corresponds to a construct with AU count 1424. A detailed description of the experiment is provided in Example 8.

Figure 13: overlay of HPLC chromatograms of constructs with different AU count on a particulate poly(styrene)-divinylbenzene (PVD) column. Flow rates and gradient profiles indicated. First peak corresponds to a construct with AU count 824, second peak corresponds to a construct with AU count 894, third peak corresponds to a construct with AU count 1137, fourth peak corresponds to a construct with AU count 1139, last peak corresponds to a construct with AU count 1424. A detailed description of the experiment is provided in Example 8.

Figure 14: overlay of HPLC chromatograms of constructs with different AU count on a Silica-based C4 column. Flow rates and gradient profiles indicated. First peak corresponds to a construct with AU count 824, second peak corresponds to a construct with AU count 894, third peak corresponds to a construct with AU count 1137, fourth peak corresponds to a construct with AU count 1139, last peak corresponds to a construct with AU count 1424. A detailed description of the experiment is provided in Example 8.

Figure 15: overlay of HPLC chromatograms of constructs with different AU count on a PLPR-S column. Flow rates and gradient profiles indicated. First peak corresponds to a construct with AU count 824, second peak corresponds to a construct with AU count 894, third peak corresponds to a construct with AU count 1137, fourth peak corresponds to a construct with AU count 1139, last peak corresponds to a construct with AU count 1424. A detailed description of the experiment is provided in Example 8.

Figure 16: plot of separation factor alpha against AU count difference for representative HPLC runs of Example 8. Monolithic: Values and respective trend line for monolithic ethylvinylbenzene-divinylbenzene copolymer with flowrate 0.5, cf. Figure 12; PVD: Values and respective trend line for particulate poly(styrene)-divinylbenzene (PVD) column with flowrate 0.3, cf. Figure 13; C4-F1: Values and respective trend line for silica-based C4 column with flowrate 1.0, cf. Figure 14; PLPR-S: Values and respective trend line for PLPR-S column with flowrate 1.0, cf. Figure 15. A detailed description of the experiment is provided in Example 8.

Examples

The Examples shown in the following are merely illustrative and shall describe the present invention in a further way. These Examples shall not be construed to limit the present invention thereto.

Table 3: Materials used

U3000 UHPLC-System	Thermo Scientific
HPLC column poly(styrene-divinylbenzen) matrix)	Thermo Scientific
WFI	Fresenius Kabi, Ampuwa
Acetonitril (MS-grade)	Fisher Scientific
0.1 M TEAA in WFI (Eluent A)	
25% ACN in 0.1 M TEAA (Eluent B)	

Example 1: Examination of the correlation between homopolymer stretches of nucleotides on HPLC retention times

The inventors surprisingly found that not the size of an RNA, but the total number of adenine nucleotides (A nucleotides) and/or uracil nucleotides (U nucleotides) of an RNA is influencing HPLC retention times. Further details are provided in the following.

1.1. Preparation of DNAs encoding firefly luciferase including varying stretches of adenines:

The DNA sequence encoding firefly luciferase protein was introduced into a modified pUC19 derived vector backbone to comprise a 5'-UTR derived from the 32L4 ribosomal protein (32L4 TOP 5'-UTR) and a 3'-UTR derived from albumin, a histone-stem-loop structure, and stretches of varying numbers of adenine nucleotides (also referred to in the following as 'polyA stretch' or 'A homopolymer') at the 3'-terminal end and. The complete RNA sequences are provided in the sequence listing (see Table 4 below).

Table 4: Constructs used in the experiment

Encoded protein	Length of polyA stretch	SEQ ID NO:
luciferase	A25	1
	A35	2
	A50	3
	A64	4

DNA plasmids were linearized using EcoRI and transcribed *in vitro* using DNA dependent T7 RNA polymerase in the presence of a nucleotide mixture and cap analog under suitable buffer conditions. The obtained individual RNA products were purified using PureMessenger® as described in WO 2008/077592 A1 and subsequently analyzed using HPLC.

1.2 Determination of HPLC retention times:

Individual RNA samples were diluted to 0.1 g/L using water for injection (WFI). 10µl of the diluted RNA sample were injected into the HPLC column (monolithic poly(styrene-divinylbenzen) matrix). The RP HPLC analysis was performed using the following conditions:

Gradient 1: Buffer A (0.1 M TEAA (pH 7.0)); Buffer B (0.1 M TEAA (pH 7.0) containing 25% acetonitrile). Starting at 30% buffer B the gradient extended to 32% buffer B in 2 minutes, followed by an extension to 55% buffer B over 15 minutes at a flow rate of 1 ml/min (adapted from WO 2008/077592). Chromatograms were recorded at a wavelength of 260 nm.

In order to examine an eventual correlation between the presence and the extent of nucleotide homopolymer stretches on the one hand and the HPLC retention time on the other hand, chromatograms of each HPLC analysis run were superimposed. Figure 4 shows the superposition of HPLC runs of RNAs with varying stretches of adenine nucleotides.

Results

As shown in Figure 4, the analysis of the HPLC retention time of different RNA molecule species differing only in the length of A nucleotide homopolymer stretches (i.e. in the total number of A nucleotides) show a clear correlation of the total number of A nucleotides in the RNA sequences and HPLC retention time. Longer A homopolymers led to an increase in retention time, suggesting that the observed effect on HPLC retention time is caused by the increased number of A nucleotides (+25 adenine nucleotides; + 35 adenine nucleotides; + 50 adenine nucleotides; + 64 adenine nucleotides).

Notably, changes in the total number of cytosine nucleotides (C nucleotides) did not have an influence on HPLC retention time (not shown). As only the number of A nucleotides and not the number of C nucleotides influences HPLC retention times, an effect merely caused by elongation of the RNA molecule species can be ruled out.

Example 2: Harmonization of HPLC retention times of different HA RNA sequences for co-purification and/or co-analysis by adaptation of the adenine count

The inventors surprisingly found that the adaptation the total number of A nucleotides in two or more different RNA species (e.g. RNA molecules comprising different sequences encoding influenza HA-B) is suitable to harmonize the HPLC retention times, so that co-purification and/or co-analysis becomes feasible. Further details are provided in the following.

2.1. Adaptation of the total number of A nucleotides:

As the previous examples show a correlation between the number of A nucleotides in an RNA sequence and the respective HPLC retention time, RNA sequences encoding HA antigens (four different RNA sequences encoding influenza HA) were adapted so that they comprised (essentially) the same number of A nucleotides. The sequence adaptation was performed in such a way that the encoded amino acid sequence was unchanged, either by exploiting the degeneracy of the genetic code (compare with Table 1 and Table 2) or by introducing an adenine stretch into the polyA tail or the UTR of the RNA molecule species.

The goal was to adapt the sequences in a way to facilitate co-purification and/or co-analysis of an RNA mixture comprising different HA RNA molecule species by obtaining a complete overlay of the four chromatograms (harmonization) in HPLC, which is a prerequisite for a cost-effective and fast production of an influenza vaccine based on an mRNA mixture (e.g., for the development of a multivalent/polyvalent influenza RNA vaccine platform, cf. Figure 3).

In order to harmonize the retention times of all RNA molecule species encoding different HA antigens (HA-A and HA-B), GC-optimized DNA sequences encoding different HA proteins of Influenza B were adapted by increasing the number of A nucleotides by adapting the coding sequence (via codon exchange), by elongating the poly A sequence, or by introducing additional A nucleotides into the UTR region (see Table 5 below). The adaptation was performed by increasing the total number of A nucleotides in the HA-B sequences by 9 in order to shift the total number of A nucleotides in the HA-B sequences closer to the number of A nucleotides in the HA-A sequences. DNA constructs and RNA prepared as explained in Example 1.

Table 5: HA-constructs used in the experiment

Encoded Antigen	Mode of adaptation	A count of RNA*	AU count of RNA**	SEQ ID NO:
HA-B Brisbane	Not adapted	467	723	5
	9 A nucleotides introduced into cds by codon exchange	476	732	6
	9 A stretch introduced into poly A tail	476	732	7
	9 A stretch introduced into the UTR	476	734	8
HA-B Phuket	Not adapted	458	717	9
	9 A nucleotides introduced into cds by codon exchange	467	726	10
	9 A stretch in poly A tail	467	726	11
	9 A stretch introduced into the UTR	467	728	12
HA-A California	Not adapted	476	737	13
HA-A Hongkong	Not adapted	481	729	14

**A-count of RNA*: total number of A nucleotides in the respective RNA*
***AU-count of RNA*: total number of A and U nucleotides in the respective RNA*

2.2. Effect of the total number of A nucleotides on HPLC retention time:

HPLC sample preparation and HPLC analysis were performed as described Example 1. In order to examine the effect of the number of A nucleotides on HPLC retention time, the chromatograms of each RNA species were superimposed and analyzed.

Figure 5A shows four superimposed chromatograms for RNA molecule species encoding HA-B/Brisbane (one non-adapted sequence (1) and three adapted sequences (2, 3 and 4, respectively)). Figure 5B shows four superimposed chromatograms for the RNA molecule species encoding HA-B/Phuket (one non-adapted sequence (1) and three adapted sequences (2, 3 and 4, respectively)). Figure 6 shows superimposed chromatograms for adapted RNA molecule species (9 Adenines introduced into cds by codon exchange) encoding HA-B/Brisbane, adapted RNA molecule species encoding HA-B/Phuket and two non-adapted RNA molecule species encoding HA-A (HA-A California; HA-A Hongkong).

Results:

The results show that the adaptation of the number of A nucleotides in the RNA sequences (see Figure 5A and Figure 5B) by addition of 9 adenines led to a shift in HPLC retention time. As indicated in Figure 5A and Figure 5B,

the effect of an A-stretch, either introduced into the UTR or introduced into the polyA tail had a slightly stronger effect on the retention time.

Figure 6 shows that HA-B sequences were successfully adapted, and that HA-A peaks and HA-B peaks are harmonized. This adaptation of the sequences, which results in completely overlapping HPLC peaks, allows for co-analysis of the individual RNA species in the mixture and for simultaneous determination of the integrity of the RNAs in the mixture. Moreover, harmonization of HPLC retention times allows for a simultaneous RNA purification (co-purification).

Of note, as analyzed and explained in further detail in Example 5, the surprisingly precise overlap of the HA-A sequences and adapted HA-B sequences as observed in Figure 6 can also be explained by the closely matching number of A nucleotides and U nucleotides (AU count) of the respective RNA sequences (HA-B Brisbane: AU count 732; HA-B Phuket: AU count 726; HA-A California: AU count 737; HA-A Hongkong: AU count: 729; see Table 4).

Example 3: Evaluation of suitability of HPLC for co-analysis of RNA mixture

The inventors showed that HPLC is a particularly suitable method for co-analysis of an RNA mixture. Further details are provided in the following.

3.1. Preparation of test RNA:

RNA for testing the HPLC system was generated according to Example 1.

3.2. Directed degradation of RNA and preparation of RNA mixtures of different integrities:

RNA samples were degraded at 90°C for 140 minutes. Subsequently, intact RNA and degraded RNA were mixed in different ratios of intact RNA: degraded RNA (90:10, 80:20, 70:30, 60:40, 50:50, 40:60, 30:70, 20:80, and 10:90) and respective RNA mixtures of varying integrities were applied to analytic HPLC. Analytic HPLC was performed as described in Example 1. For analysis, HPLC runs of the different RNA mixtures were superimposed. The results are shown in Figure 7.

Results:

As Figure 7 shows, RNA integrity of an RNA mixture can be determined in a scenario where the RNA mixture has the same retention time (harmonized RNA peak; peak of the individual RNA components, in that case integer RNA and degraded RNA, are completely overlapping). Accordingly, the analytic system is suitable for the analysis of a (polyvalent) RNA mixture according to the present invention.

Example 4: Harmonization of HPLC retention times of NA RNA sequences for co-purification and/or co-analysis by adaptation of the number of A nucleotides

The inventors surprisingly found that an RNA mixture (encoding three different NA antigens) comprising RNA species with an adapted number of A nucleotides generates one harmonized HPLC peak, suitable for co-analysis and co-purification. Further details are provided in the following.

4.1. Adaptation of the number of A nucleotides in DNA encoding NA proteins of several Influenza strains:

The goal was to adapt NA RNA sequences in a way to facilitate co-purification and/or co-analysis of an RNA mixture of different NA RNA molecule species by obtaining a complete overlay of the three chromatograms, which is a prerequisite for a cost-effective and fast production of an RNA-mixture based influenza vaccine (e.g. for the development of a multivalent influenza RNA vaccine, cf. Figure 3)

In order to harmonize the retention time of all RNA molecule species encoding different NA antigens, GC-optimized DNA sequences encoding different NA proteins of Influenza were adapted by decreasing the number of A nucleotides by altering the coding sequence (codon exchange; see Table 6 below). The adaptation was performed in order to decrease the number of A nucleotides in RNA encoding NA H3N2 and mRNA encoding NA H1N1 to essentially match the number of A nucleotides in RNA encoding NA Influenza B.

DNA constructs and RNA were prepared as explained in Example 1.

Table 6: NA-constructs used in the experiment

Encoded Antigen	Mode of adaptation	SEQ ID NO:
NA Influenza B (Brisbane)	Not adapted	15
NA H3N2 (Hongkong)	Not adapted	16
	17 A removed from cds by codon exchange	17
NA H1N1 (California)	Not adapted	18
	16 A removed from cds by codon exchange	19

4.2. Effect of the number of A nucleotides on HPLC retention time:

HPLC sample preparation and HPLC analysis were performed as described in Example 1.

In order to examine the effect of the number of A nucleotides on HPLC retention time, the chromatograms of non-adapted RNA species were superimposed and analyzed. In addition, non-adapted RNA molecule species were mixed (100 ng each), applied as a mixture, and analyzed by HPLC (see Figure 8A). In addition, adapted NA H3N2 (Hongkong) RNA, adapted NA H3N2 (California) RNA, and NA Influenza B (Brisbane) RNA were mixed (100 ng each) and applied as a mixture, and analyzed by HPLC (see Figure 8B). Figure 8A shows superimposed chromatograms for non-adapted RNA molecule species encoding NA (1, 2, 3) next to the chromatogram of the corresponding RNA mixture (4). Figure 8B shows the chromatogram of the harmonized NA RNA mixture (4).

Results

The results show that the adaptation of the number of A nucleotides in the individual RNA sequences of an RNA mixture leads to adaptation of the retention time of the RNA mixture and a discrete HPLC peak (see Figure 8B), which allows for co-analysis and co-purification (even though the individual HPLC peaks show a slight variation). In

contrast, a non-adapted RNA mixture generates a broad, non-discrete HPLC double-peak (see Figure 8B) that is not suitable for co-analysis and co-purification.

Of note, the adaptation (reduction) of the number of A nucleotides in SEQ ID NOs: 17 and 19 was performed by changing serine codon AGC to codon UCC, which led to a decrease in A count and to an increase in U count (AU count was therefore stable; ratio of A:U was decreased), suggesting that the observed slight variation in the HPLC chromatograms of the individually analyzed adapted sequences was caused by a shift in the A:U ratio. Accordingly, adaptation of the A:U ratio can also be used for sequence adaptations according to the invention.

Example 5: Examination of the influence of nucleotides on HPLC retention time

As shown in the previous examples, the adaptation of the number of A nucleotides in RNA sequences allows for harmonization of HPLC chromatograms, which is a requirement for co-analysis and/or co-purification. The inventors further found that the number of A and/or U nucleotides correlates with HPLC retention time. That finding provides even more options for adapting an RNA sequence and to harmonize HPLC chromatograms of RNA mixtures. Further details are provided in the following.

5.1. Preparation of DNA encoding HA proteins of several Influenza strains:

DNA sequences encoding different haemagglutinin (HA) and neuraminidase (NA) proteins, two glycoproteins found on the surface of influenza viruses (Influenza A and Influenza B), were generated, and RNA was produced as described in Example 1.

Table 7: HA-constructs used in the experiment:

Encoded antigen	SEQ ID NO:
HA-A California	13
HA-A Hongkong	14
HA-B Brisbane	5
HA-B Phuket	9
NA H1N1 (California)	18
NA H3N2 (Hongkong)	16
NA Influenza B (Brisbane)	15

5.2. Correlation between the total number of a nucleotide and/or the relative content of a nucleotide and the HPLC retention time:

HPLC sample preparation and HPLC analysis were performed as described Example 1.

In a first step, the individually produced RNA constructs (RNA species) encoding HA and NA antigens were separately analyzed on HPLC. The superimposed HPLC chromatograms are shown in Figure 9. The superimposition of the chromatograms of the seven different RNA species showed that all chromatograms partially overlap, which

makes both, co-purification and co-analysis via HPLC technically impossible (compare illustration of problem in the art, Figure 1).

For a better understanding of the impact of the nucleotide sequence on HPLC retention time, the correlation between the nucleotide count (A,U,G, and C) and nucleotide content for each RNA molecule species and HPLC retention time was examined.

Figure 10A shows the correlation between number and content (AU%) of A and U nucleotides of different HA and NA RNA species and their respective HPLC retention times. Figure 10B shows the correlation between number and content (GC%) of guanine (G) and cytosine (C) nucleotides of different HA and NA RNA species and their respective HPLC retention times.

Results:

Figure 10A shows a clear correlation of the number of A and/or U nucleotides with the respective retention time. Such a correlation was not found for the content of A and U nucleotides (AU %). Figure 10B shows HPLC retention times are neither influenced by the number of G and/or C nucleotides, nor by the content of G and/or C nucleotides (GC %).

Notably, the correlation between the number of A nucleotides and the retention time is stronger than the correlation between the number of U nucleotides and the retention time; In line with that, the results of Example 4 also suggested that the effect of A nucleotides on retention time is stronger than the effect of U nucleotides on retention time.

Overall, the number of A and U nucleotides shows the best correlation and will allow for the most precise way for adapting RNA sequences to harmonize RNA mixtures for co-analysis and co-purification.

Example 6: Development of an automated nucleotide adaptation method (algorithm)

The inventors developed an automated *in silico* method (algorithm) to set the number of any nucleotide in an RNA sequence to a certain defined value, without altering the amino acid sequence. In the context of the invention, the automated *in silico* method was used for sequence adaptation (adaptation of the number of A and/or U nucleotides (AU count)) of RNA sequences to allow harmonization of RNA mixtures for HPLC co-analysis and/or HPLC co-purification. Further details are provided in the following.

6.1 Sequence analysis and definition of target AU count:

The objective of the experiment was to generate RNA sequences for an adapted RNA mixture (comprising three different RNA molecule species encoding antibodies) suitable for co-analysis using HPLC. The AU count has to be adapted in all RNA molecule such that their respective HPLC chromatograms are completely separated (difference in the AU count of at least 70), allowing for co-analysis of their integrity.

Three antibody sequences (SEQ ID NOs: 20-22) were selected and GC optimized DNA (SEQ ID NOs: 23-25) sequences were generated (essentially according to Example 1). Nucleotide numbers were determined for the

respective GC optimized sequences (product 1, product 2, product 3; see Table 8 below) to be able to define optimal numbers of A and U (T) nucleotides for HPLC co-analysis.

Table 8: Nucleotide numbers for GC optimized constructs:

product	Length	A count	T (U) count	AT (AU) count	SEQ ID NO:
1	81	19	13	32	23
2	258	59	26	85	24
3	429	77	55	132	25

To adapt the RNA molecule species comprised in the RNA mixture for HPLC co-analysis, the target AU counts for product 2 and product 3 were set to the following values, allowing integrity analysis on HPLC when analyzed as an RNA mixture:

Table 9: Adaptation strategy for co-analysis of the RNA mixture:

product	AT (AU) count (non-adapted)	Change in AU count	Target AT (AU) count
1	32	0	32
2	85	+17	102
3	132	+40	172

As indicated in Table 9, the target AU count for each product RNA was set in such a manner that the AU counts of the three RNA sequences differ by at least 70 nucleotides (strategy illustrated in Figure 11).

6.2 AU sequence adaptation method:

In the following, the sequence adaptation method is exemplarily described for product 2 (+17 AU) (SEQ ID NO: 24). As the number of A nucleotides in the sequence was larger than the number of T (U) nucleotides, the adaptation values were set to +8A and +9T(U) in order to maintain the distribution of A and U nucleotides in the resulting AU adapted sequence.

In the initial phase of the method (algorithm), a matrix for each codon comprised in the sequence was created, identifying possible changes (herein referred to as "exchange matrix"). An exemplary "exchange matrix" is shown in Formula (I).

Formula (I):

$$\left.\begin{matrix} A & 1 \\ C & 1 \\ G & 1 \\ T & 1 \\ * & 4 \end{matrix}\right\} \text{CGA}$$

Formula (I) shows that for codon "CGA" a change to an alternative codon offers the option of increasing the number of A nucleotides by 1 (e.g.: CGA -> AGA), offers the option of increasing the number of C nucleotides by 1 (e.g. CGA ->CGC), offers the option of increasing the number of G nucleotides by 1 (e.g. CGA -> CGG), and offers the option of increasing the number of T nucleotides by 1 (e.g. CGA -> CGT).

Exchange matrices were generated for each individual codon in the sequence. Using said exchange matrices, the potential maximum number of the respective nucleotides (A and T(U) count, respectively) in each codon was determined (without changing the amino acid sequence). Accordingly, all 63 codons of the sequence were analyzed, and the potential alternative codons were assembled in a table structure as shown in Table 10 by way of example.

Table 10: Exemplary table of alternative codons allowing for a change in the number of a nucleotide

Codons	**Alternative codons**
CGA	CGT, CGC, CGG, AGA, AGG
GAT	GAC
GAC	GAT
ATG	no alternative codon
...	...

Next, the sequence according to SEQ ID NO: 24 was iteratively divided into separate codons and stored in table format, which resulted in a list as exemplarily shown in Table 11 (positions 1, 2, 3, 4 ... 86 and 87 of the sequence are indicated).

Table 11: Codon list of SEQ ID NO 24:

Codon position	**Codon**
1	ATG
2	AGC
3	ATC
4	ATC
...	...
86	GAG
87	AGC

Next, the list of codons (see Table 11) was analysed for possible codon changes by step-wise iteration, wherein in each iteration step the corresponding codon was analysed using the respective exchange matrix (as outlined above) for potential nucleotide changes. For example, if no changes were theoretically possible in the respective codon, e.g. as in the case of "ATG" or "TGG", the corresponding exchange matrix as exemplarily shown in Formula (II) was used (* of exchange matrix = 0).

Formula (II):

$$\left.\begin{matrix} A & 0 \\ C & 0 \\ G & 0 \\ T & 0 \\ * & 0 \end{matrix}\right\} \text{ATG}$$

Formula (II) shows that for codon "ATG" a change to an alternative codon offers no option of increasing the number of A nucleotides, C nucleotides, G nucleotides or T nucleotides (as there are no alternative codons for ATG (Met)).

In cases where changes according to the respective exchange matrix (* > 0) were theoretically possible, the codon was further analysed if these changes can be implemented under the premise that e.g. only codons that offer the option of increasing the number of A and/or T(U) nucleotides were adapted. Therefore the intersection between the target nucleotides (e.g. A and/or T(U)) and the nucleotides that potentially generate a positive result (that is, A and/or T(U) change; see e.g. Formula (I)) in the current exchange matrix was constructed. As a result, each codon was categorized and grouped in three categories:

Category 1 (category "favourable"):
Potential codon exchanges allowing an increase in only one target nucleotide (in the present example A or T(U)). For example, the codon "GAC" (Asp) can be changed to "GAU" (Asp) in order to increase the number of A and T(U) nucleotides. No further analysis is required since that modification does not have any further impact (besides the one mentioned above) on the number of A and T(U) nucleotides.

Category 2 (category "possible"):
Potential codon exchanges allowing the increase in both target nucleotides (in the present example A and T(U)). For example, codon "GCA" can be changed to "GCU", which would increase the T(U) count but at the same time decrease the A count. Accordingly, further analysis would be required with respect to codons belonging to this the category in order to decide, whether the number of one of the two target nucleotides (T(U)) in this example should be increased at the expense of a reduction of the number of the other target nucleotide (A).

Category 3 (category "impossible"):
Codons in the RNA sequence, for which no alternative codons exist (* = 0). Examples for this category 3 are ATG (Met; start codon) or "UGG" (Trp).

All codons of the original sequence were categorized in that manner. After this step, there were three categories with a total of 87 entries (for 87 codons present in SEQ ID NO: 24). For the next step, category 3 was no longer considered, as a codon change will not influence the target nucleotide count (A, T(U)).

Next, it was calculated how many potential nucleotide changes have been identified for all target nucleotides (A, T(U)). For the SEQ ID NO: 24 the possible nucleotide changes are listed (category 1, category 2; see Table 12).

Table 12: Nucleotide changes that can potentially be applied to SEQ ID NO 24

Nucleotide	Potential changes	Category
A	22	1
T	15	1
A,T	47	2

Accordingly, 22 favourable changes (see category 1 / "favourable" as explained above) were identified for A and 15 favourable changes were identified for T(U). As the adaptation values were set to +8A and +9T(U) the adaptation at codon positions were all taken from category 1. Table 13 summarizes the introduced codon exchanges that were equally distributed across the sequence.

Table 13: Codon exchanges introduced into SEQ ID NO 24

Position A	Exchange	A increase	Position T	Exchange	T(U) increase
30	AAG -> AAA	+1	3	AGC -> AGT	+1
48	AAG -> AAA	+1	12	AGC -> AGT	+1
78	AAG -> AAA	+1	63	GAC -> GAT	+1
141	CAG -> CAA	+1	90	AGC -> AGT	+1
165	GAG -> GAA	+1	102	AGC -> AGT	+1
213	GAG -> GAA	+1	117	AAC -> AAT	+1
234	GAG -> GAA	+1	138	AGC -> AGT	+1
252	GAG -> GAA	+1	150	AAC -> AAT	+1
			180	AGC -> AGT	+1

These exchanges resulted in the following adapted sequence according to SEQ ID NO: 25.

In the above described example, potential nucleotide exchanges from category 2 were not implemented. However, in scenarios where e.g. T(U) counts larger than 15 are needed, codons from category 2 are used as soon as all codons from category 1 have been used, in order to obtain additional adaptation possibilities for A and T(U) counts. If category 2 is required in order to achieve the desired nucleotide counts, calculation of the following ratio will identify the exchange nucleotide (nucleotide A or T(U)):

$$\frac{c_i}{x_i - p_i}$$

wherein *i* represents the corresponding target, c_i is the count of possible adaptation positions of *i* in category 2, x_i is the desired threshold for *i*, p_i the count for the already changed identified adaptation positions. All calculated ratios are ranked and starting from lowest to highest ratio, the changes from category 2 are applied, until the desired threshold has been reached or until all the possible exchanges from category 2 have been performed. This procedure is carried out iteratively for all targets, where the desired numbers cannot be achieved by only using exchanges according to category 1.

For example, category 2 contains 47 codons (see Table 12), which could potentially be exchanged in order to increase the A or T count. Accordingly, further changes are implemented from category 2 until the desired threshold has been reached or until all the codons from category 2 have been used as well. For SEQ ID NO: 24, a change of the T(U) count to e.g. 20 would result in additional adaptations by using the following alternative codons from category 2 (additional codon exchanges from category 1 are not shown): 6=ATT, 9=ATT, 21=CTT, 144=CCT, 183=TCT.

In cases where the desired target nucleotide count cannot be achieved (as all alternative codons from category 1 and 2 have been used, which means that no further changes are possible), an adapted sequence is generated that is matching the target nucleotide count as close as possible.

In order to further optimize the above described method (algorithm), the following improvements are implemented:

1. The basic equal distribution, which was used in the experiments described above, is based on the exchange possibilities. Other distribution models may also be envisaged, such as normal distribution, first occurrences distribution, last occurrences distribution or random-based distribution. Alternatively, the mean of the possible changes or median of the possible changes may be determined and all exchanges may be arranged around these values.
2. The exchange matrix contains additional information about the codon for the target sequence (e.g. codon usage etc.). This creates a further criteria for the question of whether a codon exchange is desirable or not, facilitating adaptation to a specific codon usage or a different nucleotide ratio in the target sequence.
3. Implementation of a third category by sequences or motifs which should be avoided by an exchange (e.g. a recognition motif of a restriction enzyme, promotor sequences or sequences building not desired secondary structures, etc.).
4. Automated binning of input sequences, based on their length and the occurrence of the desired target nucleotides in order to identify optimal nucleotide counts for A and/or U adaptation.

Example 7: Generation and use of a polyvalent influenza virus RNA platform for fast-adjustable influenza vaccine production

A pool of GC optimized coding sequences encoding HA antigens were AU adapted to a count of 612 AU (360 A and 252 T) resulting in an AU adapted HA sequence pool (SEQ ID NO: 26 to 16263). A pool of GC optimized coding sequences encoding NA antigens were AU adapted to an AU count of 488 AU (271A and 217U) resulting in an AU adapted NA sequence pool (SEQ ID NOs: 16264-30567). AU adaptations were performed according to the invention, essentially as described in Example 6. The adaptation allows co-purification of RNA mixtures comprising adapted HA RNA sequences and co-purification of RNA mixtures comprising adapted NA sequences. Moreover the adaptation

allows co-analysis of an RNA mixture comprising adapted HA and NA RNA species as the RNA sequences encoding HA (AU count 612) and the RNA sequences encoding NA (AU count 488) generate separated peaks (AU count difference: 124), suitable for analysis of integrity using HPLC.

HA RNA mixtures are produced according to procedures as disclosed in the PCT application WO2017/109134 using GC optimized AT adapted DNA templates (generated as described in Example 1). In short, a DNA construct mixture (each of which comprising different HA coding sequences and a T7 promotor) is used as a template for simultaneous RNA *in vitro* transcription to generate a mixture of HA mRNA constructs. Subsequently, the obtained harmonized RNA mixture is used for co-purification using RP-HPLC.

In a parallel reaction, NA RNA mixtures are produced according to procedures as disclosed in the PCT application WO 2017/109134 using GC optimized AT adapted DNA templates. In short, a DNA construct mixture (each of which comprising different NA coding sequences and a T7 promotor) is used as a template for simultaneous RNA in vitro transcription to generate a mixture of HA mRNA constructs. Subsequently, the obtained harmonized RNA mixture is used for co-purification using RP-HPLC.

The purified mRNA mixture encoding HA antigens and the purified mRNA mixture encoding NA antigens are mixed to generate a HA/NA RNA mixture. The integrity of the mixture (that is of the NA peak and the HA peak) is co-analyzed via HPLC as described herein.

Advantageously, the AU adaptation of HA sequences and NA sequences in order to harmonize chromatographic peaks for HPLC based co-purification and co-analysis according to the invention facilitates the production of mRNA-based multivalent influenza vaccines, which may be quickly adapted to demand, e.g. in seasonal influenza vaccine design or in a pandemic scenario (compare with Figure 3).

Example 8: Suitability of the method on various reverse phase HPLC matrices

The inventors found that modification of the retention time of an RNA via adaptation of A and/or U count is not restricted to a certain reverse phase column chemistry.

To test whether a modification in retention time via an adaptation of A and/or U count can also be observed on other reverse phase columns (in Example 1 – 7, a monolithic poly(styrene-divinylbenzene)matrix has been used), the following columns were tested:

- monolithic ethylvinylbenzene-divinylbenzene copolymer (ThermoFisher Scientific) (see Figure 13)
- particulate poly(styrene)-divinylbenzene column (ThermoFisher Scientific) (see Figure 14)
- Silica-based C4 column (ThermoFisher Scientific) (see Figure 15)
- PLRPS, non-alkylated porous poly(styrene-divinylbenzene)matrix(see Figure 16)

RNA encoding yellow fever virus antigens was generated. The constrcuts encode the same antigen (YFV(17D)-prME), comprise the same UTR elements, and have the same size. The AU count for each constrcust was changed via coding-sequence adaptation. The constructs are listed in Table 14.

Table 14: AU adapted YFV constructs

SEQ ID NO	Antigen	RNA size	A	U	G	C	AU	GC
30588	prME	2311	474	350	544	943	**824**	1487
30589	prME	2311	587	307	704	713	**894**	1417
30590	prME	2311	633	504	627	547	**1137**	1174
30591	prME	2311	788	351	555	617	**1139**	1172
30592	prME	2311	886	538	516	371	**1424**	887

To evaluate the effect of AU adaptation on retention time, 500ug of each construt was subjected individually to the respective column. In addition, for each column two different flow-rates were tested. Results are shown in Figures 12-15.

In addition, the separation factor alpha was determined for each construct on each column tested. In chromatography, the separation factor alpha expresses the ratio of retention times of two compounds. Accordingly, a separation factor of value of larger 1.0 means that separation of two compounds occurred. In the present analysis, SEQ ID NO: 30592 (with AU 1424) was taken as a reference for separation factor calculation. The calculated separation factors are shon in Table 15. The obtained separation factor values were plotted against the AU count difference of the constructs (SEQ ID NO: 30592, with AU 1424 was taken as a reference). The diagram is shown in Figure 16.

Table 15: AU adapted YFV constructs

SEQ ID NO	**RNA size**	**AU count**	AU difference	Alpha monolithic	Alpha PVD	Alpha C4	Alpha PLPR-S
30592	2311	1424	0	1,00	1,00	1,00	1,00
30591	2311	1139	285	1,07	1,05	1,05	1,08
30590	2311	1137	287	1,09	1,07	1,06	1,11
30589	2311	894	530	1,15	1,12	1,10	1,18
30588	2311	824	600	1,19	1,15	1,12	1,24

Conclusion:

As shown in Figure 13, the modulation of retention time works on a monolithic ethylvinylbenzene-divinylbenzene-copolymer. All four constructs elute at different retention times and could be clearly separated. In the present case, the separation of construct with AU824 from the construct with AU894 was strong enough to allow further analysis of the peaks.

As shown in Figure 14, the modulation of retention time works on a particulate poly(styrene)-divinylbenzene column. All four constructs elute at different retention times and could be clearly separated. In the present case,

the separation of construct with AU824 from the construct with AU894 was strong enough to allow further analysis of the peaks.

As shown in Figure 15, the modulation of retention time works on a Silica-based C4 column column. All four constructs elute at different retention times and could be clearly separated. In the present case, the separation of construct with AU824 from the construct with AU1139 was strong enough to allow further analysis of the peaks.

Figure 16 summarizes the inventive concept of the present invention. An adaption of the AU count of the different RNA constructs (in other words: increasing the difference in AU count between the constructs) led to a modification in retention time, thereby allowing a separation on HPLC. Notably, that effect was observed irrespective of the column matrix used.

Of course, a harmonization of AU counts (in other words: decreasing the difference in AU count between the constructs) would also lead to a modification in retention time, thereby allowing co-purification on HPLC.

Claims

1. Method for modifying the retention time of RNA on a chromatographic column,
wherein the method comprises a step of adapting the RNA sequence by altering the number of A and/or U nucleotides in the RNA sequence with respect to the number of A and/or U nucleotides in the original RNA sequence.

2. The method according to claim 1, wherein the method comprises a step of determining the number of A and/or U nucleotides in the original RNA sequence and a step of increasing or decreasing said number of A and/or U nucleotides in order to obtain the adapted RNA sequence.

3. The method according to claim 1 or 2, wherein the method comprises a step of replacing G and/or C nucleotides in the original RNA sequence with A and/or U nucleotides.

4. The method according to any one of the preceding claims, wherein the method comprises a step of replacing A and/or U nucleotides in the original RNA sequence with G and/or C nucleotides.

5. The method according to any one of the preceding claims, wherein the method comprises a step of introducing additional nucleotides, preferably G and/or C nucleotides, more preferably A and/or U nucleotides, into a non-coding region of the original RNA sequence.

6. The method according to any one of the preceding claims, wherein the RNA comprises a coding region and wherein the number of A and/or U nucleotides in the RNA sequence is altered within the coding region.

7. The method according to claim 6, wherein the amino acid sequence encoded by the coding region of the adapted RNA sequence is not modified with respect to the amino acid sequence encoded by the coding region of the original RNA sequence.

8. The method according to any one of the preceding claims, wherein the retention time of the adapted RNA is increased by increasing the number of A and/or U nucleotides by at least 2, preferably by at least 3, more preferably by at least 4, even more preferably by at least 5, most preferably by at least 10, with respect to the number of A and/or U nucleotides in the original RNA sequence.

9. The method according to any one of claims 1 to 7, wherein the retention time of the adapted RNA is decreased by decreasing the number of A and/or U nucleotides by at least 2, preferably by at least 3, more preferably by at least 4, even more preferably by at least 5, most preferably by at least 10, with respect to the number of A and/or U nucleotides in the original RNA sequence.

10. The method according to any one of the preceding claims, wherein the chromatographic column is an HPLC column.

11. The method according to claim 10, wherein the chromatographic column comprises a reversed phase as a stationary phase.

12. The method according to any one of the preceding claims, wherein the chromatographic column comprises a material, preferably a porous material, selected from the group consisting of polystyrene, a non-alkylated polystyrene, an alkylated polystyrene, a polystyrenedivinylbenzene, a non-alkylated polystyrenedivinylbenzene, an alkylated polystyrenedivinylbenzene, a silica gel, a silica gel modified with non-polar residues, a silica gel modified with alkyl containing residues, selected from butyl-, octyl and/or octadecyl containing residues, a silica gel modified with phenylic residues, or a polymethacrylate.

13. The method according to any one of the preceding claims, wherein the chromatographic column is a column formed by beads or a monolithic column.

14. Method for purifying at least one RNA species from a mixture of at least two RNA species, wherein the method comprises:
a) a step of adapting the sequence of the at least one RNA species by altering the number of A and/or U nucleotides with respect to the number of A and/or U nucleotides in the original sequence; and
b) a chromatographic step, wherein the adapted RNA species is separated from at least one other RNA species.

15. The method according to claim 14, wherein step a) comprises modifying the retention time of the RNA on a chromatographic column as defined in any one of claims 1 to 13.

16. The method according to claim 14 or 15, wherein the number of A and/or U nucleotides in the adapted RNA species differs from the number of A and/or U nucleotides in the at least one other RNA species by at least 2, preferably by at least 3, more preferably by at least 4, even more preferably by at least 5, most preferably by at least 10.

17. The method according to any one of claims 14 to 16, wherein the number of A and/or U nucleotides in the adapted RNA species differs from the number of A and/or U nucleotides in the at least one other RNA species by at least 20, preferably by at least 30, more preferably by at least 40, even more preferably by at least 50, most preferably by at least 60.

18. The method according to any one of claims 14 to 17, which further comprises a step of determining the identity and/or integrity of at least one RNA species.

19. Method for co-purifying at least two RNA species from a mixture comprising at least two RNA species, wherein the method comprises:
a) a step of adapting the sequence of at least one RNA species by altering the number of A and/or U nucleotides in the RNA sequence with respect to the number of A and/or U nucleotides in the original RNA sequence; and

b) a chromatographic step, wherein the at least one RNA species having an adapted sequence and at least one further RNA species are co-purified.

20. The method according to claim 19, wherein step a) comprises determining a target number of A and/or U nucleotides, to which the number of A and/or U nucleotides in the at least two RNA species to be co-purified is adapted.

21. The method according to claim 19 or 20, wherein the numbers of A and/or U nucleotides in the sequences each of the at least two RNA species to be co-purified differ by not more than 20, more preferably by not more than 30, even more preferably by not more than 40, most preferably by not more than 50.

22. The method according to any one of claims 19 to 21, wherein the numbers of A and/or U nucleotides in the sequences each of the at least two RNA species to be co-purified differ by not more than 2, preferably by not more than 3, more preferably by not more than 4, even more preferably by not more than 5, most preferably by not more than 10.

23. The method according to any one of claims 18 to 22, wherein the at least two RNA species to be co-purified elute within the same fraction.

24. Method for harmonizing the numbers of A and/or U nucleotides in the sequences of at least two RNA species, the method comprising adapting the sequence of at least one RNA species by altering the number of A and/or U nucleotides in the RNA sequence with respect to the number of A and/or U nucleotides in the original RNA sequence.

25. The method according to claim 24 further comprising any one of the features defined in claims 1 to 23.

26. The method according to claim 24 or 25, wherein the numbers of A and/or U nucleotides in the sequences of each of the at least two harmonized RNA species differ from each other by not more than 20, more preferably by not more than 30, even more preferably by not more than 40, most preferably by not more than 50.

27. The method according to any one of claims 24 to 26, wherein the numbers of A and/or U nucleotides in the sequences of the at least two harmonized RNA species differ from each other by not more than 2, preferably by not more than 3, more preferably by not more than 4, even more preferably by not more than 5, most preferably by not more than 10.

28. The method according to any one of claims 24 to 27, wherein the method comprises adapting the sequence of at least two RNA species by altering the number of A and/or U nucleotides in the RNA sequences with respect to the number of A and/or U nucleotides in the original RNA sequences.

29. Method for synthesis of a mixture comprising at least two harmonized RNA species, the method comprising

a) a step comprising harmonizing the numbers of A and/or U nucleotides in the sequences of at least two RNA species, wherein the sequence of at least one RNA species is adapted by altering the number of A and/or U nucleotides in the RNA sequence with respect to the number of A and/or U nucleotides in the original RNA sequence; and

b) a step of synthesis of the at least two harmonized RNA species.

30. The method according to claim 29, wherein step b) comprises the separate synthesis of the at least two harmonized RNA species.

31. The method according to claim 30, wherein step b) comprises mixing the at least two harmonized RNA species.

32. The method according to claim 29, wherein step b) comprises the synthesis of the at least two harmonized RNA species in one batch.

33. The method according to any one of claims 29 to 32, wherein step b) comprises an in vitro transcription step.

34. The method according to any one of claims 29 to 33, which is further characterized by any one of the features defined in any one of claims 1 to 28.

35. The method according to any one of claims 1 to 13, 14 to 18, 19 to 23, 24 to 28 or 29 to 34, wherein at least one RNA or RNA species comprises at least 30, preferably at least 50, more preferably at least 100, even more preferably at least 200 nucleotides, most preferably at least 500 nucleotides.

36. The method according to any one of claims 1 to 13, 14 to 18, 19 to 23, 24 to 28, 29 to 34 or 35, wherein at least one RNA or RNA species is monocistronic, bicistronic or multicistronic.

37. The method according to any one of claims 1 to 13, 14 to 18, 19 to 23, 24 to 28, 29 to 34, 35 or 36, wherein at least one RNA or RNA species is a coding RNA, preferably an mRNA.

38. The method according to any one of claims 1 to 13, 14 to 18, 19 to 23, 24 to 28, 29 to 34 or 35 to 37, wherein at least one RNA or RNA species comprises a 5'-cap structure.

39. The method according to any one of claims 1 to 13, 14 to 18, 19 to 23, 24 to 28, 29 to 34 or 35 to 38, wherein at least one RNA or at least one RNA species is an mRNA comprising a coding region, wherein the coding region of the original RNA sequence comprises an optimized coding sequence,
wherein the optimized coding sequence is selected from C maximized coding sequence, G/C increased coding sequence, G/C optimized coding sequence, human codon usage optimized coding sequence, CAI maximized coding sequence, or any combination thereof, and

wherein the encoded amino acid sequence is preferably not modified compared to the amino acid sequence encoded by the corresponding wild-type mRNA.

40. The method according to any one of claims 1 to 13, 14 to 18, 19 to 23, 24 to 28, 29 to 34 or 35 to 39, wherein at least one RNA or at least one RNA species is an mRNA comprising a coding region, wherein the at least one coding region comprises a nucleic acid sequence, which is G/C optimized.

41. The method according to any one of claims 1 to 13, 14 to 18, 19 to 23, 24 to 28, 29 to 34 or 35 to 40, wherein at least one RNA or at least one RNA species comprises at least one histone stem-loop.

42. The method according to any one of claims 1 to 13, 14 to 18, 19 to 23, 24 to 28, 29 to 34 or 35 to 41, wherein at least one RNA or at least one RNA species comprises a 3'-UTR.

43. The method according to claim 42, wherein the 3'-UTR comprises a poly(A) sequence and/or a poly(C) sequence.

44. The method according to claim 42 or 43, wherein the 3'-UTR comprises at least one heterologous 3'-UTR element.

45. The method according to claim 44, wherein the at least one heterologous 3'-UTR element comprises a nucleic acid sequence derived from a 3'-UTR of a gene selected from the group consisting of an albumin gene, an alpha-globin gene, a beta-globin gene, a tyrosine hydroxylase gene, a lipoxygenase gene, and a collagen alpha gene, or a homolog, a fragment or a variant of said nucleic acid sequence.

46. The method according to any one of claims 1 to 13, 14 to 18, 19 to 23, 24 to 28, 29 to 34 or 35 to 45, wherein at least one RNA or at least one RNA species comprises a 5'-UTR.

47. The method according to claim 46, wherein the 5'-UTR comprises at least one heterologous 5'-UTR element.

48. The method according to claim 47, wherein the at least one heterologous 5'-UTR element comprises a nucleic acid sequence, which is derived from the 5'-UTR of a TOP gene, preferably from a corresponding RNA sequence, or a homolog, a fragment, or a variant thereof, preferably lacking the 5'TOP motif.

49. The method according to claim 47 or 48, wherein the at least one heterologous 5'-UTR element comprises a nucleic acid sequence, which is derived from a 5'-UTR of a TOP gene encoding a ribosomal protein, preferably from a corresponding RNA sequence, or a homolog, a fragment or a variant of said nucleic acid sequence, preferably lacking the 5'TOP motif.

50. The method according to any one of claims 47 to 49, wherein the at least one heterologous 5'-UTR element comprises a nucleic acid sequence, which is derived from a 5'-UTR of a TOP gene encoding a ribosomal Large protein (RPL), preferably from RPL32 or RPL35A, or from a gene selected from the group consisting

of HSD17B4, ATP5A1, AIG1, COXC6 or ABCB7(MDR), or from a homolog, a fragment or variant of any one of these genes, preferably lacking the 5'TOP motif.

51. The method according to any one of claims 1 to 13, 14 to 18, 19 to 23, 24 to 28 or 29 to 50, wherein at least one RNA or at least one RNA species comprises, preferably in 5' to 3' direction, the following elements:
a) a 5'-cap structure
b) optionally, a 5'-UTR element,
c) at least one coding region;
d) a 3'-UTR element,
e) optionally, a poly(A) sequence, preferably comprising 10 to 200, 10 to 100, 40 to 80 or 50 to 70 adenine nucleotides,
f) optionally, a poly(C) sequence, preferably comprising 10 to 200, 10 to 100, 20 to 70, 20 to 60 or 10 to 40 cytosine nucleotides,
g) optionally, a histone stem-loop,
h) optionally, an additional poly(A) sequence, or a fragment or variant of any one of these nucleic acid sequences.

52. The method according to any one of claims 1 to 13, 14 to 18, 19 to 23, 24 to 28 or 29 to 51, wherein at least one RNA or at least one RNA species comprises at least one coding region and wherein the coding region preferably comprises a nucleic acid sequence according to any one of SEQ ID NO: 26 to 14079, 14080 to 16264, 16265 to 28640, 28641 to 30568, or a fragment or variant of any one of these nucleic acid sequences.

53. RNA, preferably an adapted RNA, obtainable by a method according to any one of claims 1 to 13, 14 to 18, 19 to 23, 24 to 28, 29 to 34 or 35 to 52.

54. Vector comprising a nucleic acid sequence corresponding to the RNA according to claim 53 or encoding the RNA according to claim 53.

55. Composition comprising at least two RNA species, wherein the sequence of at least one RNA species is adapted by altering the number of A and/or U nucleotides in the sequence of the at least one RNA species with respect to the number of A and/or U nucleotides in the original RNA sequence.

56. The composition according to claim 55, wherein the sequence of at least one RNA species is adapted by a method comprising any one of the features defined in any one of claims 1 to 13, 14 to 18, 19 to 23, 24 to 28, 29 to 34, or 35 to 52.

57. The composition according to claim 55 or 56, wherein the numbers of A and/or U nucleotides in the sequences of at least two RNA species are harmonized, preferably by using the method according to any one of claims 24 to 28.

58. The composition according to any one of claims 55 to 57, wherein the composition comprises at least one RNA species comprising at least one coding region encoding a peptide or protein derived from Influenza virus, or a fragment or a variant of said peptide or protein.

59. The composition according to any one of claims 55 to 58, wherein at least one RNA species encodes Influenza virus hemagglutinin (HA) and/or Influenza virus neuraminidase (NA), or a fragment or variant thereof.

60. The composition according to any one of claims 55 to 59, wherein the composition optionally further comprises a pharmaceutically acceptable carrier.

61. The composition according to any one of claims 55 to 60, wherein the RNA of at least one RNA species is complexed or at least partially complexed or associated with one ore more cationic or polycationic compound, which is preferably selected from cationic or polycationic polymer, cationic or polycationic polysaccharide, cationic or polycationic lipid, cationic or polycationic protein, cationic or polycationic peptide, or any combination thereof.

62. The composition according to any one of claims 55 to 61, wherein the RNA of at least one RNA species is complexed or associated with a lipid, preferably a cationic or neutral lipid, thereby preferably forming liposomes, lipoplexes, lipid nanoparticles, or nanoliposomes.

63. Vaccine comprising the RNA according to claim 53 or the composition according to any one of claims 55 to 62.

64. The vaccine according to claim 63, which is a polyvalent vaccine.

65. The composition according to any one of claims 55 to 62 or the vaccine according to claim 63 or 64, which further comprises an adjuvant.

66. The vaccine according to any one of claims 63 to 65, which is an Influenza virus vaccine, preferably a polyvalent Influenza vaccine.

67. Kit or kit of parts comprising the RNA according to claim 53, the composition according to any one of claims 55 to 62, or the vaccine according to any one of claims 63 to 66, optionally a liquid vehicle for solubilising, and optionally technical instructions providing information on administration and dosage of the components.

68. The composition according to any one of claims 55 to 62, the vaccine according to any one of claims 63 to 66, or the kit or kit of parts according to claim 67 for use as a medicament.

69. Method for providing an adapted RNA, the method comprising the steps of
    a) optionally, determining the total number of nucleotides in an original RNA sequence;

b) determining the number of A and/or U nucleotides in the original RNA sequence;

c) determining the codons in the original RNA sequence that can be replaced with at least one alternative codon without changing the encoded amino acid sequence;

d) adjusting the number of A and/or U nucleotides in the RNA sequence to a pre-set number of A and/or U nucleotides by replacing at least one original codon with an alternative codon, wherein the alternative codon encodes the same amino acid as the original codon and is further characterized in a higher content of A and/or U nucleotides; and

e) optionally, producing the RNA, preferably via in vitro transcription.