(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
06 September 2024 (06.09.2024)

WIPO | PCT

(10) International Publication Number
WO 2024/180048 A2

(51) International Patent Classification:
Not classified

(21) International Application Number:
PCT/EP2024/054929

(22) International Filing Date:
27 February 2024 (27.02.2024)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
2302887.1 28 February 2023 (28.02.2023) GB

(71) Applicant: **RIBOPRO B.V.** [NL/NL]; Kloosterstraat 9, 5349AB Oss (NL).

(72) Inventor: **VAN ASBECK, Alexander**; Kloosterstraat 9, Building RK, 5349AB Oss (NL).

(74) Agent: **HUTTER, Anton** et al.; Venner Shipley LLP, The Surrey Research Park, 5 Stirling House, Stirling Road, Guildford Surrey GU2 7RF (GB).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CV, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, MG, MK, MN, MU, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, CV, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SC, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, ME, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report (Rule 48.2(g))*

(54) Title: RNA

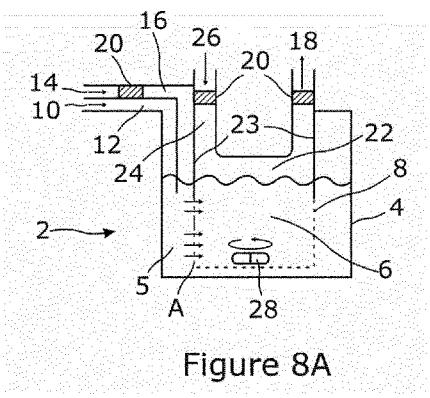

Figure 8A

(57) Abstract: The invention relates to ribonucleic acid (RNA), and to methods for synthesizing and co-transcriptionally capping RNA. The invention is especially concerned with methods for enhanced co-transcriptional capping of mRNA using cap analogues. The invention also relates to apparatus for performing such methods, and to devices that enable the improved synthesis of in vitro transcribed mRNA, and more specifically to the addition of a 5' cap structure to in vitro transcribed mRNA.

WO 2024/180048 A2

RNA

The invention relates to ribonucleic acid (RNA), and particularly, although not exclusively, to methods for synthesizing and co-transcriptionally capping RNA. The invention is especially concerned with methods for enhanced co-transcriptional capping of mRNA using cap analogues. The invention also relates to apparatus for performing such methods, and to devices that enable the improved synthesis of *in vitro* transcribed mRNA, and more specifically to the addition of a 5′ cap structure to *in vitro* transcribed mRNA.

All living organisms make use of DNA to store their genetic information. Most of these genes code for specific proteins which control all cellular functions, including metabolic processes, cellular signalling to allow cells to respond to their environment, the structure of cells and ultimately the organism, and also intercellular communication through cell-cell contacts and hormones. Key to translating the genetic information from the level of DNA into proteins is a molecule called messenger RNA (mRNA). Each gene is first transcribed into an mRNA which is then translated through ribosomes into a protein. The key elements of this process are very similar across all kingdoms of life from bacteria to fungi to plants to animals. Following transcription of the mRNA, so-called posttranscriptional modifications occur that further modify the primary mRNA structure, however, these posttranscriptional processes and/or their functional implications differ between kingdoms of life.

Messenger RNA is a polymer of ribonucleic acids with a backbone formed from the pentose sugar ribose and phosphate. The nucleobases adenine, guanine, cytosine and uridine are attached to these sugars via the 1′C-atom. The resulting nucleosides are coupled to each other via phosphate molecules that form esters with the hydroxy group of the 3′C-atom of one ribose and of the hydroxy group of the 5′C-atom of another ribose. The 5′- and 3′C-atoms give the mRNA molecule a directionality. During transcription, the mRNA is synthesized from 5′ to 3′ along the DNA template. Also, translation through the ribosome occurs from 5′ to 3′. One unit of the mRNA polymer consisting of nucleoside and phosphate is called a nucleotide.

As one of the structural modifications, for all eukaryotic mRNAs, a stretch of adenine-nucleotides is attached to the 3′-end of the mRNA molecule. This poly-A tail plays a key role for controlling the lifetime and expression of the mRNA as the poly-A tail is bound through proteins that protect the mRNA from degradation by

RNA exonucleases. Poly-A tails also occur in prokaryotes, however, in this case rather destabilize than stabilize mRNA.

A further important modification of the mRNA is the addition of a so-called 5′ cap structure, which is a chemical modification of the 5′-end of the mRNA. The 5′ cap is required to initiate assembly of ribosomes at the mRNA and thus protein expression from the mRNA. Bacteria and eukaryotes differ in the structure of the cap and these structural differences also form the basis for the recognition of foreign mRNA through the innate immune system of higher organisms as explained, below. Importantly, in eukaryotes, the 5′ cap is added in the nucleus which means that all cytosolic mRNA that does not carry a proper 5′ cap is most likely of foreign (pathogen) origin and so potentially poses a threat.

Next to mediating initiation of translation through ribosome binding, the 5′ cap also protects the mRNA from 5′-to-3′ exonucleases, and prevents the activation of 5′ phosphate triggering of RIG-I, which mediates an potent intracellular anti-viral response.

As 5′ cap-initiated translation is the major mechanism for initiation of translation. Without a 5′ cap present, the majority of mRNA molecules would not be active due to a lack of ribosome engagement and subsequent scanning and translation of the mRNA. Only internal ribosomal entry sites can substitute for this property of the 5′ cap. However, mRNAs generated by *in vitro* transcription (IVT) are typically generated for translation initiated by 5′ cap binding. In the absence of a 5′ cap, the 5′ end is readily available for 5′-to-3′ exonucleases, such as the human exonuclease Xrn1, which digests mRNA that is de-capped by Dcp2, or that is uncapped. When uncapped, mRNA molecules typically have 5′ terminal triphosphate groups, which mediate binding of RIG-I (retinoic acid-inducible gene I) and similar proteins, resulting in the activation of the IFN-I response by this pattern Recognition Receptors. Under pathophysiological conditions, this response serves, as a main function, to limit viral spreading, initiate and promote the innate immune response, and activate the adaptive immune response for long-term, extensive viral control. The viral spreading is mainly limited by the production of pro-inflammatory cytokines and limitation of the expression of viral proteins.

As a consequence, introduction of uncapped mRNA into a cell will typically yield low to no substantial protein expression from that transfected mRNA, trigger pro-

inflammatory cytokines that cause inflammation, and depending on the amount transfected may cause cell-death and/or (global) inhibition of protein translation. Thus, it is of paramount importance to ensure high capping efficiency (*i.e.* that most, or substantially all, mRNA molecules in a pharmaceutical mRNA preparation are capped) during the mRNA production process.

Finally, eukaryotic mRNAs are mostly formed as so-called pre-mRNAs which consist of exons and introns. Only the exons code for proteins, thus during splicing, which also happens in the nucleus, the introns are removed. Furthermore, individual nucleobases can be chemically modified through enzymatic machineries which is a further means to distinguish foreign from self mRNA. Only fully matured mRNA leaves the nucleus for expression in the cytosol.

Considering that (i) mRNA can code for virtually any protein and that (ii) the presence of mRNA in the cytosol is transient, thereby not posing the risk of permanent changes of the genome, and also that (iii) mRNA, unlike DNA, does not have to enter the nucleus to show expression, thus can also be applied to non-dividing, differentiated cells, synthetic mRNA introduced into cells may be considered a highly interesting modality to yield cellular protein expression. In 1990, it was first demonstrated that injection of naked mRNA yielded an immune response against the protein that the mRNA encoded for, thus demonstrating that mRNAs could be applied *in vivo* as a message for protein production (Wolff JA, et al. Science. 1990;247:1465–1468). Two years later, it was also demonstrated that *in vivo* delivery of mRNA could yield expression of a bioactive protein (Jirikowski GF, et al. Science. 1992;255:996–998).

However, it took more than twenty years until the full potential of mRNA as a highly valuable molecular modality in medicine and biotechnology was more broadly recognized. This reservation was a result of the fact that mRNA was broadly recognized as a biologically very unstable molecule that is rapidly degraded by RNases which are omnipresent in the body and also on the skin and thus complicate handling of mRNA. Furthermore, because of its large molecular weight and negative charge, delivery vehicles had to be developed that complex with the mRNA and thereby protect the mRNA from degradation and mediate cellular entry and cytosolic delivery. Finally, at about the same time that exploration of mRNA as a therapeutic modality started, also mechanisms of innate immunity were discovered that showed that exogenously added mRNA can act as a strong immunostimulatory

molecule that causes inflammatory reactions and shuts down cellular protein biosynthesis as explained above.

It, therefore, became clear that *in vitro* transcription of mRNA was more challenging than initially thought, as only mRNA that mimics as closely as possible endogenous mRNA can avoid recognition through innate immunity receptors and thus yield high protein expression without side effects. Therefore, innovations that improve the *in vitro* synthesis of mRNA, including synthesis yields, and also structural changes that reduce innate immunity are of enormous economic value, as exemplified by the patents for inclusion of chemically modified bases (PCT/US06/32372 – RNA containing modified nucleosides and methods of use thereof) and of modified 5′ cap structures (US7074596B2 – Synthesis and use of anti-reverse mRNA cap analogues).

Following the break-through success of the mRNA-based SARS-CoV2 vaccines, Comirnaty (marketed by Pfizer/BioNTech) and SpikeVax (marketed by Moderna), which were made possible through the above-mentioned insights into mRNA biology, RNA (in particular, mRNA) is considered to be one of the fastest growing molecular modalities in therapy development and biotechnology with a multi-billion dollar market, the ultimate scope of which is still difficult to predict but which will, to a major extent, depend on costs-of-goods and safety. Costs-of-goods directly relate to methods of manufacturing, and safety to the avoidance of side products which also includes incompletely 5′-capped mRNAs. In particular, for applications in medicine that require more frequent applications of larger amounts of mRNA as for enzyme replacement therapy where the mRNA codes for proteins that a patient is missing due to a genetic defect, both factors will be crucial for the long-term success of therapeutic mRNA.

There is, therefore, a need to provide improved methods for *in vitro* transcription of RNA that enable the *in vitro* transcription of mRNA with higher 5′ capping yields, at lower costs of goods than the current state of the art. Considering the enormous economic significance of *in vitro* transcribed mRNA in medicine and biotechnology, those skilled in the art will recognize the enormous economic potential of the invention described herein. In addition, there is also a need for novel devices that enable an improved synthesis of *in vitro* transcribed mRNA, and more specifically to the addition of a 5′cap structure to *in vitro* transcribed mRNA, via methods. The resultant mRNA products should bear a higher degree of 5′ cap structures with less

consumption of expensive reagents and simplified purification procedures, which lowers the cost of *in vitro* transcription of mRNA and causes less activation of innate immune responses by 5′ uncapped mRNA which in turn yields higher protein expression and causes less unwanted side effects of the *in vitro* transcribed mRNA when used in *in vitro* or *in vivo* applications.

The inventors set out to improve *in vitro* transcription of RNA with higher 5′ capping yields. As discussed above, the 5′ cap is an important feature of the mRNA molecule, but unfortunately also one of the most expensive components to add, regardless of the method of introduction. Generally, there are two methods of introducing a 5′ cap on a nascent mRNA molecule. There include:

(i) post-transcriptional capping via enzymes like Vaccinia capping enzyme or Fausto capping enzyme, which are typically employed in a secondary step subsequent to the *in vitro* transcription of the mRNA due to incompatibilities in buffer requirements (salts) between the different enzymes. In addition, some methyl donor molecules such as S-adenosylmethionine (SAM) are unstable and may thereby limit the effective duration of the IVT reaction. These enzymes modify the 5′-end of the nascent RNA, recruit and couple via a series of intermediate steps a GTP and methylate the inversely coupled (5′-5′coupling) GTP to form the mature cap. Under the correct conditions and with an uncomplicated 5′end (limited or weak secondary structure formation), the aforementioned enzymes are typically highly efficient and the cap is exclusively added in the correct orientation; and

(ii) co-transcriptional capping via cap-analogues, including mCap (TriLink Biotechnologies, Inc), anti-reverse cap analogue (ARCA – ThermoFisher Scientific), and trinucleotide cap-analogues. In this process, cap-analogues that mimic the 5′ cap (*i.e.,* dinucleotides with a 5′-5′ linkage or a trinucleotide with a 5′-5′ linkage) are mixed into the IVT reaction and incorporated by the RNA polymerase as the first nucleotide instead of the competing nucleotide (*e.g.,* rGTP in the case of ARCA). The use of mCAP results in around 50% of the cap being incorporated in the wrong orientation, resulting in 50% of inactive mRNA molecules. Anti-reverse cap analogue (ARCA) has modifications on the sugar backbone that prevent the incorporation in the wrong orientation. Similarly, tri-nucleotide cap-analogues are also exclusively incorporated in the correct orientation.

For each of the above, the underlying principles of incorporation and competition with the initiating nucleotide are the same. Typically, in batch *in vitro* transcription reactions involving co-transcriptional capping, a fixed ARCA:GTP (or other cap-anlogue:competing nucleotide(s)) ratio is set at the start of the reaction, alongside a fixed amount of several other reagents, including nucleoside triphosphates (NTPs), magnesium, RNA polymerase, deoxyribonucleic acid (DNA), and optionally pyrophosphatase. During the reaction, the NTPs are consumed, with the nucleotide that is first depleted limiting the yield of the reaction. To achieve a reasonable capping efficiency, the nucleotide that is competing with the cap-analogue (e.g., rGTP competes with ARCA to be incorporated as first nucleotide of the mRNA) and which would produce upon incorporation an uncapped mRNA, is limited in concentration. Alternatively, the cap analogue could be added at elevated concentrations, but this strategy is limited by cost and the solubility of the cap-analogue.

As a consequence, a balance must be struck between the mRNA yield and the capping efficiency achieved by the IVT reaction. The most frequently used protocols employ a 4:1 ratio of Cap-analogue (*e.g.,* ARCA) over competing nucleotide (*e.g.,* rGTP), resulting in a theoretical capping efficiency of around 80% if the tendency for incorporation as the first nucleotide is equal between both molecules. In such protocols, the amount of limiting nucleotide (e.g., rGTP) is limited to $1/4^{th}$ of the concentration of the other nucleotides (ATP, CTP, UTP and/or their chemically modified counterparts), limiting the yield to around $1/4^{th}$.

Based on the inventors' calculations, therefore, it can be determined that when performing IVT reactions involving co-transcriptional capping via a cap-analogue, the consumptions of the cap-analogue is a fraction of the consumption of the (competing) nucleotide. In fact, for a given length of mRNA, the relative consumption of the cap-analogue versus the nucleotides can be expressed as 1/length of the mRNA, or more precisely the consumption of the cap-analogue versus the competing nucleotide can be expressed as 1/(length of the mRNA * the fraction of the competing nucleotide in said mRNA). The majority of (therapeutically relevant) mRNA products are at least 500nt long, resulting in a ratio of at least 1/500, up to 1/20,000 for the longest constructs. As such, it can be reasoned that the concentration of the cap-analogue during a batch-reaction remains relatively constant and the unconsumed cap-analogue constitutes a significant expense and waste.

To remedy this, fed-batch reactions from the prior art add feeding solutions containing fresh amounts of nucleotides that are being consumed during the reaction, thereby restoring the original concentration of such nucleotides. Such feeding solutions typically contain a mixture of all of the consumed nucleotides, optionally in a relative ratio adjusted to the empirically (WO2020185811) or calculated relative consumption by the reaction.

In order to achieve high capping efficiencies during the co-transcriptionally capped IVT reaction, a high cap analogue to competing nucleotide ratio must be maintained to prevent the formation of undesired uncapped RNA contaminants being formed by the incorporation of the competing nucleotide. The theoretical capping efficiency at any given moment of the reaction can be expressed as: capping efficiency = $100*([\text{cap-analogue}]_t / [\text{sum of the competing nucleotide(s)}]_t)$ wherein t = a specific moment in time, assuming equal or similar incorporation rates of such the cap-analogue and competing nucleotide(s). As stated before, in a batch reaction (see Table 1), the cap-analogue/competing nucleotide (e.g., ARCA/rGTP) ratio increases over the duration of the reaction due to the higher consumption of the competing nucleotide compared of the cap-analogue. Assuming that the competing nucleotide is the yield-limiting nucleotide and falls to 0mM, and assuming equal or similar incorporation rates, the capping efficiency can be expressed as: capping efficiency = $100*([\text{cap-analogue}]_{start} / ([\text{sum of the competing nucleotide(s)}]_{start} / 2))$.

Table 1 – Capping efficiency and ratio of components for a batch reaction

Desired capping efficiency	Starting ratio cap-analogue/ competing nucleotide/other NTPs in batch reaction	Theoretical yield compared to batch reaction with equal amounts of NTPs
80% - 90%	4:1:4	25%
90% - 95%	9:1:9	11%
95% - 97.5%	20:1:20	5%
99% - 99.5%	99:1:99	1%

In a fed-batch reaction (see Table 2), however, the capping efficiency will not only depend on the ratio of the starting concentration of cap-analogue and competing

nucleotide(s), but also on the timing, concentration and ratio of each in the feeding solution.

Table 2 – Capping efficiency and ratio of components for a fed batch reaction

Desired capping efficiency	Average ratio cap-analogue/ competing nucleotide/other NTPs in fed-batch reaction	Theoretical yield compared to batch reaction with equal amounts of NTPs
80%	4:1:4	25% per feeding
90%	9:1:9	11% per feeding
95%	20:1:20	5% per feeding
99%	99:1:99	1% per feeding

From Tables 1 and 2, it can be appreciated that extreme high cap-analogue to competing nucleotide(s) ratios are required to obtain mRNA with a high degree of capping. This can often only be achieved by using relatively low concentrations of the competing nucleotide (for example, when 10mM ARCA is used, to achieve >95% capping, an average concentration of 500uM of rGTP must be used). In general, the lower the concentration of the rate-limiting nucleotide, the slower the reaction speed. These principles remain valid with the use of tri-nucleotide cap-analogues with improved incorporation in the mRNA, as well as the use of engineered RNA polymerase variants that show an increased preference for cap-analogue incorporation over competing nucleotide incorporation on the first positions of the nascent mRNA, albeit that the ratios will be more favourable.

It can also be appreciated that a high number of feedings are required to obtain the same yield of RNA as compared to a comparable reaction with equal amounts of each NTP. The combination of a slower reaction speed and a higher number of feedings significantly prolong the duration of the reaction. Therefore, it is highly beneficial to maximize the transcription rate. In general, more frequent feedings to replenish the rate-limiting nucleotide(s), while maintaining an average rate of addition to the rate of consumption result in the highest reaction rate at a given desired capping efficiency as shown in theoretical modelling below. In view of this, the inventors have realised that continuous feeding is, therefore, a desired feature of any fed-batch system aiming to achieve a high capping efficiency.

Referring to Figure 6, in the case of high-frequency or continuous feeding of concentrated material (to avoid diluting the reaction too much), fine-tuned fluid control of small volumes is required. Depending on the total reaction volume and the associated tolerated feeding volume this may be challenging. For example, for small scale reactions (for example 20µl, typically producing 100-200µg of mRNA in a batch reaction without co-transcriptional capping and each NTP at 5-10mM), the scale of the fluid becomes prohibitively small to exert sufficient precision, as shown in Table 3 below. For the purposes of this application, the practical limits of microfluidic feed of NTPs are within the range of 0.4µl/min to >100ml/min. For the purposes of this application, the practical limits of macrofluidic feed of NTPs is within the range of >60µl/min to >100L/min.

Table 3 – Capping reaction yield achieved by continuous feeding of reagents

Reaction volume / yield	Concentration of feed stock	Feed amount per NTP	Duration of reaction	Continuous feed rate	
20µl – 200µg	100mM	2µl	8h	0.25µl/h	4.2nl/min
	10mM	20µl	8h	2.5µl/h	42nl/min
	1mM	200µl	8h	25µl/h	0.42µl/min
100µl – 1mg	100mM	20µl	8h	1.25µl/h	21nl/min
	10mM	200µl	8h	12.5µl/h	0.21µl/min
	1mM	2ml	8h	125µl/h	2.1µl/min
1ml – 10mg	100mM	100µl	8h	12.5µl/h	0.21µl/min
	10mM	1ml	8h	125µl/h	2.1µl/min
	1mM	10ml	8h	1.25ml/h	21µl/min
10ml – 100mg	100mM	1ml	8h	125µl/h	2.1µl/min
	10mM	10ml	8h	1.25ml/h	21µl/min
	1mM	100ml	8h	12.5ml/h	210µl/min
100ml – 1gr	100mM	10ml	8h	1.25ml/h	21µl/min
	10mM	100ml	8h	12.5ml/h	210µl/min
	1mM	1L	8h	125ml/h	2.1ml/min

Besides difficulty in fluid control, dilution of the IVT reaction by repeated or continuous administration of feeding solution presents another hurdle. Whereas industrial batch IVT reactions utilize 5-10mM of each NTP (with the exception of the competing nucleotide(s)) to drive high-speed, high-efficiency reactions all present

at the start of the reaction. Industrial batch-fed IVT reactions on the other hand supply the same total amount of NTPs as the batch reactions distributed over multiple feedings or a continuous feeding. When fed from a concentrated 100mM feed stock solution, each nucleotide encompasses a total of 5-10% of the total reaction volume. When supplying more than 1 NTP, a multiple of this amount is added to the reaction, resulting in a potentially undesired dilution of the remaining reaction components, including RNAP, DNA, etc. with as much as 40%. When using less concentrated feed stocks, the issue of dilution is compounded.

Finally, (continuous) fed-batch reactions that feed concentrated feed solutions rely on (continuous) vigorous mixing of the solution to distribute the components from the feed solution throughout the reaction volume. Vigorous mixing, especially for prolonged periods of time, causes damage to the proteins, lowering the productivity of the system. Less vigorous mixing results in a prolonged time of unmixed or incompletely mixed solutions.

Despite recent advances, current protocols for co-transcriptional IVT are unable to achieve sufficiently high capping efficiency in combination with high reaction yield and/or high consumption of the expensive cap-analogue components. Solutions relying on current fed-batch protocols are unable to achieve the necessary precise control over the feed rate to maintain constant very low levels of the competing nucleotide required to maintain optimal reaction speed at a given (extreme) cap-analogue-to-competing nucleotide ratio.

Therefore, there is an urgent need for improved methods and associated apparatuses to obtain a combination of high yield and high(er) capping efficiency for the economic production of high quality mRNA.

The inventors have devised a novel IVT method, specifically for co-transcriptional capping of RNA with cap-analogues in which they select for high cap-analogue-to-competing nucleotide ratios (for therapeutically relevant capping efficiency). However, as a consequence, the yield of such reactions is too low to be practically or economical. They have, therefore, provided a fed-batch protocol to address the yields. However, a problem with using a fed-batch approach is that feeding from concentrated stock to prevent overly diluting the reaction results in such low volumes that adequate fluid control is challenging, often leading to large fluctuations in the reaction conditions. Conversely, feeding from diluted stock

solutions results in an overly diluted reaction. Therefore, the method of the invention relies on using diluted stock to obtain sufficiently precise control coupled with simultaneous/subsequent excess fluid removal to correct for the dilution. In addition, the inventors use a system of highly distributed addition of the feed solution to the reaction volume to prevent local, temporal deviations in the overall reaction composition.

Accordingly, in a first aspect of the invention, there is provided a method of preparing a capped ribonucleic acid (RNA) molecule, wherein the method comprises carrying out a fed-batch transcription reaction in the presence of a cap analogue in a substantially constant volume, thereby resulting in a substantially capped RNA molecule.

In a second aspect, there is provided a capped ribonucleic acid (RNA) molecule obtained, or obtainable by, the method of the first aspect.

In a third aspect, there is provided a ribonucleic acid (RNA) production apparatus for preparing capped RNA, the apparatus comprising a reaction chamber in which a transcription reaction occurs in the presence of a cap analogue, and a feed means for feeding a feed solution comprising one or more reagent required for producing capped RNA, wherein the apparatus is configured, in use, to feed the reagents, via the feed means, to the reaction chamber in a fed-batch mode, and carry out transcription in the reaction chamber in a substantially constant volume, to thereby produce a substantially capped RNA molecule.

In yet another aspect, there is provided a method of preparing a capped ribonucleic acid (RNA) molecule, wherein the method comprises carrying out a fed-batch transcription reaction in the presence of a cap analogue, thereby resulting in a substantially capped RNA molecule.

Advantageously, the methods and apparatus of the invention achieve surprisingly higher 5′ capping yields than batch or standard fed-batch methods, and at lower costs of goods than the current state of the art. The resultant capped RNA molecules of the second aspect of the invention bear a higher degree of 5′ cap structures with less consumption of expensive reagents and simplified purification procedures, which considerably lowers the cost of *in vitro* transcription of RNA. Furthermore, the increased level of 5′ capping yields higher protein expression and

causes less unwanted side effects of the *in vitro* transcribed mRNA when used in *in vitro* or *in vivo* applications. For example, they cause less activation of the innate immune responses by 5′ uncapped RNA when administered to a subject, for example as a vaccine.

The expression "substantially constant volume" can mean that the volume of the transcription reaction is maintained, and is substantially fixed or the same, when at steady-state, i.e. the volume does not increase or decrease over the course of the reaction, and so can be described as being volume-balanced. This is in contrast to prior art methods using fed-batch approaches in which the volume of the reaction increases over the course of the reaction.

However, it will be appreciated that there may be very small fluctuations in volume in the method of the invention. Preferably, any increase or decrease in reaction volume is less than 5%, 4% or 4%. More preferably, any increase or decrease in reaction volume is less than 3%, 2% or 1%. Even more preferably, any increase or decrease in reaction volume is less than 0.5%, 0.4% or 0.3%. Most preferably, any increase or decrease in reaction volume is less than 0.2%, 0.1% or 0.05%.

The method of the invention may be *in vitro* or *ex vivo*. However, most preferably the method is carried out *in vitro.* Preferably, the method comprises *in vitro* transcription (IVT).

The RNA may be single-stranded or double-stranded. The RNA may be coding or non-coding. The RNA may be selected from a group of RNA molecules consisting of: messenger RNA (mRNA), micro RNA (miRNA); interference RNA (RNAi); short interfering RNA (siRNA); short hairpin RNA (shRNA); anti-sense RNA; RNA aptamers; self-amplifying RNA (saRNA); coding RNA; non-coding RNA; and circular RNA.

Preferably, the RNA comprises mRNA. The mRNA may comprise the basic elements of the cap, a 5′ UTR, a 3′ UTR, an IRES, a coding sequence, and a poly(A) tail of variable length.

The RNA molecule (which is preferably mRNA) may be at least at least 50 bases in length, at least 60 bases in length, at least 75 bases in length, at least 100 bases in length, at least 200 bases in length, at least 300 bases in length, at least 400 bases in length, at least 500 bases in length, at least 600 bases in length, at least 700 bases in

length, at least 800 bases in length, or at least 900 bases in length. The skilled person would appreciate that when the RNA is double-stranded, for example double-stranded RNA, "bases in length" will refer to the length of base pairs.

The RNA molecule, most preferably mRNA, may be at least 1000 bases in length, at least 2000 bases in length, at least 3000 bases in length, at least 4000 bases in length, at least 5000 bases in length, at least 6000 bases in length, at least 7000 bases in length, at least 8000 bases in length, at least 9000 bases in length, at least 10,000 bases in length, at least 11,000 bases in length or at least 12000 bases in length.

The method preferably comprises contacting: (i) a template nucleic acid sequence, (ii) an RNA polymerase, (iii) a plurality of nucleotide triphosphates (NTPs), and (iv) a cap analogue, wherein the RNA polymerase transcribes the template nucleic acid with the plurality of NTPs and the cap analogue, to thereby form the capped RNA molecule, preferably in a single one-pot reaction.

The term "fed-batch" will be well-known to the skilled person, and can mean a process whereby one or more substrates of the reaction (i.e. the NTPs, cap analogue, and/or optionally a buffer and/or Mg2+ ions) are fed to the reactor during the reaction and in which the product (i.e. the capped RNA molecule) remains in the reactor until the end of the run. This is in contrast to a "batch" reaction in which all of the substrates are provided in the reactor at the start of a reaction, which is then allowed to proceed to completion. Feb-batch is also different to a "continuous" reaction in which the substrates are continuously fed into the reactor, and the product is continuously removed from the reactor.

The template nucleic acid sequence is preferably transcribed by the RNA polymerase to produce the RNA molecule. Preferably, therefore, the template nucleic acid sequence comprises DNA. Preferably, the DNA comprises a promoter suitable for the RNA polymerase, wherein the promoter is suitably located upstream of the nucleic acid sequence templating for the desired RNA molecule. The template nucleic acid may be made synthetically, for example by PCR, or doggybone DNA. The template nucleic acid may comprise a vector, and is preferably a plasmid. The template nucleic acid may comprise a restriction site or other suitable linearization site to allow for run-off transcription by the RNA polymerase.

Preferably, the method comprises the use of an RNA polymerase, which may be selected from a group consisting of: T7; T3; SP6; KP34; Syn5; or other DNA-dependent RNA polymerase; or a mutated variant of any of these RNA polymerases. Each of these RNA polymerases is able to perform the synthesis of the capped RNA according to the method of the invention. Preferably, the method comprises the use of T7 RNA polymerase or a variant thereof. In an embodiment of the invention, more than 1 RNA polymerase is used simultaneously in the method (preferably the IVT reaction) to transcribe multiple RNA sequences from multiple DNA templates or the same RNA sequence as multiple RNA molecules with different properties.

Preferably, the plurality of nucleotide triphosphates (NTPs) are selected from a group consisting of ATP, GTP, CTP and/or UTP, or modified variants thereof, including pseudoUTP, N1-methyl-UTP, m5CTP, m6ATP, m1ATP, Inosine triphosphate, hm5CTP, m1GTP, m7GTP, or m6AmTP. NTPs are the building blocks of RNA. Preferably, the method comprises the use of substantially equal proportions of each of ATP, CTP and/or UTP, and an amount of GTP that is adjusted to the amount of cap analogue in relation to the desired capping efficiency. However, in some embodiments, the method may comprise different ratios between each of ATP, GTP, CTP and/or UTP. This will depend on the sequence of the template nucleic acid and resultant RNA molecule to be transcribed therefrom.

Preferably, the cap analogue is selected from a group of cap analogues consisting of: an mCap (m7G(5')ppp(5')G, TriLink Biotechnologies, Inc); an anti-reverse cap analogue (a version of mCap modified by 3’ OH methylation on the m7G) ARCA – ThermoFisher Scientific); and a trinucleotide cap-analogue (e.g., Cleancap – TriLink Biotechnologies); and modifications and/or combinations thereof. The skilled person will readily appreciate the different types of modifications that can be made to these cap analogues.

It will be appreciated that the cap analogue mimics the 5’ cap (*i.e.,* dinucleotides with a 5’-5’ linkage, or a trinucleotide with a 5’-5’ linkage, wherein the 5’-5’ linked nucleoside is Guanosine, methylated at the 7’ position or variants thereof) and is comprised in the reaction mixture, and incorporated by the RNA polymerase as the first nucleotide of the nascent RNA instead of a corresponding competing nucleotide.

For example, in an embodiment in which the cap analogue is ARCA, then the competing nucleotide is rGTP. In addition, in an embodiment in which the cap analogue is mCap, the competing nucleotide is rGTP. In embodiment in which the cap analogue is a trinucleotide cap-analogue, the competing nucleotide is (i) rGTP, if the cap-analogue binding sequence is GG, (ii) rATP, if the cap-analogue binding sequence is AG, wherein the cap-analogue binding sequence is defined as the nucleotides that bind to the template nucleic acid sequence (preferably DNA) and are the first two incorporated nucleotides.

Preferably, the method comprises contacting the template nucleic acid sequence, the RNA polymerase, the plurality of nucleotide triphosphates (NTPs), and the cap analogue in the presence of magnesium ions. Preferably, the concentration of magnesium ions is 1-5mM greater than the total concentration of NTPs present and/or fed to the reaction up to that moment. For example, a fed-batch reaction containing 10mM of each NTP may preferably comprise >40mM Mg2+, a fed-batch reaction initially containing 5mM of each NTP may initially comprise >20mM Mg2+, and upon feeding with an additional 5mM of each NTP, an additional 20mM of Mg2+. Since excess Mg2+ can increase the tendency of the RNA polymerase to produce dsRNA via RNA-templated-RNA synthesis, the excess Mg2+ should be minimized during the entire reaction. Therefore, it is preferred to add any extra Mg2+ required following addition of NTPs during or immediately after such feeding, instead of providing excess Mg2+ at the start of the reaction. As a general formula, the Mg2+ concentration may preferably be: ((SUM of the total amount of each NTP and cap-analogue provided / volume of the reaction volume) + 1-5mM of Mg2+ in the reaction volume).

Preferably, the method comprises contacting the reagents with a buffering agent. In a preferred embodiment, the buffering reagent is Tris (tris(hydroxymethyl)aminomethane) with a suitable counter-ion, preferably hydrochloride. The concentration of buffering agent may be preferably 10-50 mM of Tris-HCl set at pH 7.8-7.9 (at 25°C), more preferably 20-40 mM of Tris-HCl set at pH 7.8-7.9 (at 25°C), and most preferably 40 mM of Tris-HCl set at pH 7.8-7.9 (at 25°C).

Preferably, the method comprises contacting the reaction reagents with base. Preferably, the base as used in prolonged and frequently-fed reactions, wherein the buffering capacity of the initially provided buffering agent is exceeded. The

concentration of base may be preferably equal or exceeding the amount of any acid added to, or generated during the reaction. In a preferred embodiment, however, no base is needed because the buffering capacity of the initially provided buffering agent is sufficient to maintain the desired pH of the reaction ± 0.1 units.

Preferably, the method comprises contacting the reagents with a reducing agent. The reducing agent may be selected from a group consisting of Dithiothreitol (DTT), dithioerythritol (DTE), beta-mercaptoethanol, and Tris (2-carboxyethyl) phosphine (TCEP). Preferably, however, the reducing agent is DTT. The concentration of the reducing agent may be 0.5-5 mM, more preferably 1-3 mM, and even more preferably 1 mM. In IVT reactions that are prolonged incubated and exposed to intermittent or continuous air-flow (*e.g.,* to evaporate excess fluid, as discussed herein), a higher reducing agent concentration may be required to off-set the increased oxidation (and inactivation) of reducing agent and/or reaction components. In such cases, the amount of active, unoxidized reducing agent is preferably maintained at 1mM by the addition of additional reducing agent. Alternatively to DTT, dithioerythritol (DTE), beta-mercaptoethanol, or Tris (2-carboxyethyl) phosphine (TCEP) may be used as a reducing agent at similar concentrations.

Preferably, the method comprises contacting the reagents with spermidine. The concentration of spermidine may be 0.1-10 mM, more preferably 1-5 mM, and even more preferably 2 mM. In a preferred embodiment, the concentration of spermidine is relative to the amount of DNA and/or RNA polymerase in the reaction.

The NTPs and cap analogue (and optionally buffer and Mg2+) are preferably contained within a feed solution. Advantageously, and preferably, the RNA is synthesized and co-transcriptionally capped with a high degree of correct capping when the feed solution is supplied in a more dilute form compared to a more concentrated form. Therefore, preferably the method comprises feeding the reagents in a dilute feeding solution. Preferably, the dilute feed solution may comprise a concentration of feeding components such that the total number of feedings add >25% of the initial reaction volume, thereby resulting in >25% dilution of the components present in the initial reaction volume, but not or at lower concentration in the feed solution, if no fluid removal is or would be applied.

In a preferred embodiment, the dilute feed solution may comprise a concentration of feeding components such that the total number of feedings add >20% of the initial reaction volume, more preferably >40% of the initial volume, and more preferably add >100%, of the initial reaction volume, if no fluid removal is or would be applied. Alternatively, the dilute feed solution is defined as a feed solution containing 1 or more components >2 times less concentrated, more preferably >4 times less concentrated, more preferably >10 times less concentrated, than their respective solubility limit and/or technical manufacturability limit. For example, GTP is reported to have a maximum solubility of around 100mg/ml (169.74mM), thus stock solution of 100mM are routinely the highest concentrated stock solutions available. Feeding from 100mM stock, results typically in 5-10% dilution of the initial reaction volume (total 5-10mM of each NTP) without removal of excess fluid. Therefore, 2x diluted corresponds to 50mM and 10-20% dilution of the reaction (total 5-10mM of each NTP), 4x diluted corresponds to 25mM and 20-40% dilution of the reaction (total 5-10mM of each NTP), 10x diluted corresponds to 10mM and 50-100% dilution of the reaction (total 5-10mM of each NTP).

In a preferred embodiment, the diluted feed solution is defined by the dilution of the cap-analogue competing nucleotide compared to the most concentrated stock solution available for said competing nucleotide, and is at least 2x, more preferably 5x, more preferably 10x, optionally more than 25x, and optionally more than 100x. Preferably, not all components of the feed solution are diluted or diluted equally, but added in the amount required to maintain stable reaction conditions related to such component.

In one preferred embodiment, the method comprises feeding (or replenishing) the competing nucleotide (*e.g.,* rGTP) that is competing with the nucleotide-cap analogue (*e.g.,* ARCA) at a rate that is suitable to maintain the ratio between the competing nucleotide (*e.g.,* rGTP) and the cap analogue (*e.g.,* ARCA) between pre-determined minimum and maximum thresholds. The feeding (or replenishing) may be continuous or intermittent. The minimum threshold of nucleotide-cap analogue:competing nucleotide (e.g. ARCA:rGTP, or another cap analogue, and/or another competing nucleotide) may be determined by the desired capping efficiency, whereas the maximum threshold of nucleotide-cap analogue:competing nucleotide (e.g. ARCA:rGTP, or another cap analogue, and/or another competing nucleotide) may be determined by the minimum desired reaction speed and/or (final) yield of the reaction.

In a preferred embodiment, the method comprises feeding only the competing nucleotide (*e.g.,* rGTP in the case of using ARCA as cap analogue), instead of a mixture of the nucleotides (which would be being consumed). The feeding may be continuous or intermittent.

Preferably, the method comprises continuously feeding only the competing nucleotide, less preferred the same number of feedings as the non-competing to competing nucleotide ratio (e.g. ratio = 9:1, then 9x feeding) with a volume that is the feed stock concentration * the amount of the competing nucleotide before consumption.

Preferably, the method comprises >=4x feedings with >=10% of reaction volume each of a matched concentration of feed stock, every 1-4 hours. More preferably, the method comprises >=10x feedings with >=10% of reaction volume each of a matched concentration of feed stock, every 2-4 hours. Even more preferably, the method comprises >=20x feedings with >= 5% of reaction volume each of a matched concentration of feed stock, every 3-4 hours.

In such an embodiment, the maximal desired concentration of each of the non-competing nucleotides is preferably added to the reaction at start of the reaction, thereby enabling the highest transcription rate possible, and the transcription rate is only limited by the cap analogue competing nucleotide provided at a concentration substantially lower than the non-competing nucleotide and/or cap analogue.

In one preferred embodiment, the composition of the reaction at the start of the reaction comprises 5mM, more preferably 6mM, and more preferably 7mM of each of rATP, rCTP, rUTP, or their derivatives (*e.g.,* pseudoUTP, N1-methyl-UTP, m5CTP, m6ATP, m1ATP, Inosine triphosphate, hm5CTP, m1GTP, m7GTP, or m6AmTP) or mixes thereof. In another embodiment, the composition of the reaction at the start of the reaction comprises 8mM, more preferably 9mM, and even more preferably 10mM of each of rATP, rCTP, rUTP, or their derivatives (*e.g.,* pseudoUTP, N1-methyl-UTP, m5CTP, m6ATP, m1ATP, Inosine triphosphate, hm5CTP, m1GTP, m7GTP, or m6AmTP) or mixes thereof.

In such an embodiment, the starting and/or average concentration of the competing nucleotide (*e.g.,* preferably rGTP when ARCA is the cap analogue), is selected as a fraction of the cap analogue concentration sufficient to achieve the desired capping efficiency preferably 1:9, more preferably 1:10, and more preferably 1:20.

It should be appreciated that 1:9 would equate to a 90% capping efficiency and is what the inventors consider to be the lowest capping efficiency that one should consider useful. Moreover, 90% capping efficiency is a significantly higher capping efficiency compared to the commonly used 80% when using a 4:1 ratio. 1:10 capping efficiency is also more preferred due to the quality improvement, and 1:20 would result in >95% capping efficiency and a low concentration of 0.5mM GTP at start of the reaction. Concentrations below 0.5mM of any nucleotide significantly suffer from a reduction in the reaction speed.

Preferably, the starting and/or average concentration of the competing nucleotide (e.g., rGTP when ARCA is the cap-analogue used) is selected as a fraction of the cap-analogue concentration sufficient to achieve the desired capping efficiency preferably 1:50, more preferably 1:75, and even more preferably 1:100.

Therefore, advantageously, and preferably, the method and apparatus comprises a volume-balanced (i.e. a substantially fixed steady-state volume) fed-batch *in vitro* transcription (IVT) method, that allows for high yield transcription of RNA, with a high capping efficiency via co-transcriptional capping with a high cap-analogue to competing nucleotide ratio.

In a preferred embodiment, the feed solution comprises the competing nucleotide(s) and, optionally, counterions of the nucleotide(s), in water which is RNAse-free. Preferably, the water does not comprise any buffer or salt components. In another embodiment, the feed solution comprises competing nucleotide(s) in water (*e.g.,* RNAse-free) without the addition of buffer or salt components, other than any counterions of the nucleotide(s) and a matched amount of Mg2+ with its appropriate counterions (*e.g.,* chloride, acetate).

In one embodiment, the cap analogue (*e.g.,* ARCA, mCAP or trinucleotide cap analogue etc.) starting concentration is sufficiently high to allow for a competing nucleotide (*e.g.,* rGTP) concentration sufficient to maintain a suitable reaction rate,

preferably >5mM, more preferably >8mM, and more preferably >10mM. In such embodiment, a suitable transcription rate is defined as >1% of the desired yield per hour, more preferably >5% of the desired yield per hour, more preferably >10% of the desired yield per hour.

In a preferred embodiment, the method comprises reducing the volume of the transcription reaction during or after feeding to compensate for the additional volume of the feed solution added thereto. Preferably, the volume reductions maintain the desired volume and/or concentration of the transcription reaction, and increase or maintain reaction speed. Such volume reductions may be carried out post-feeding or substantially continuously. Thus, preferably, the method comprises feeding only the competing nucleotide, while reducing the volume, more preferably at a high cap analogue to competing nucleotide ratio.

Thus, in a preferred embodiment, the feeding of the competing nucleotide and/or any other reaction component (such as, other nucleotides, base, buffer, salts (NaCl or TrisCl), Mg2+, etc.) is performed to allow fluids to be simultaneously, or subsequently, added and removed. Preferably, the addition and removal of fluid is matched as to not change the volume of the reaction, and thus the concentration of the non-fed reaction components (*e.g.,* RNA polymerase and/or template DNA), beyond predefined parameters. In a preferred embodiment, the addition and removal of fluid is matched to maintain a substantially identical volume of the reaction. Steady state is preferred. Furthermore, the removal of the fluid is performed in such a way that only water is removed, and not any substantial amount of dissolved reaction components or products.

In one preferred embodiment, excess fluid (which is preferably water) is removed from the reaction by means of evaporation, thereby keeping the volume of the reaction substantially constant.

Water evaporation from a water surface is mainly dependent on water temperature, air temperature, air humidity and air velocity above the water surface. The amount of evaporated water per unit of time can be expressed as:

$$gh = \Theta A (xs - x)$$

wherein:

gh = amount of evaporated water per hour (kg/h)

Θ = (*25 + 19 v*) = evaporation coefficient (kg/m2h)

v = velocity of air above the water surface (m/s)

A = water surface area (m2)

xs = maximum humidity ratio of saturated air at the same temperature as the water surface (kg/kg) (kg H2O in kg Dry Air); and

x = humidity ratio air (kg/kg) (kg H2O in kg Dry Air).

For the purposes of this invention, the practical limits of evaporative excess water removal are considered to be: ~0.25µl/min/cm2 (0.5m/s air velocity, 20°C air temperature, 37°C water temperature, 50% humidity) to ~15µl/min/cm2 (10m/s air velocity, 20°C air temperature, 37°C water temperature, 30% humidity).

Table 4 – Evaporation parameters

Reaction volume / yield	Max surface area (microfluidic) device (A) – cm2	Surface area in cubic device (B) – cm2	Max. continuous evaporation rate in (A, left) and (B, right) – µl/min	
20µl – 200µg	0.2	0.073	3	1
100µl – 1mg	0.33	0.22	5	3.3
1ml – 10mg	3.3	1	50	15
10ml – 100mg	33.3	4.62	500	69.3
100ml – 1gr	333.3	22	5000	330

As illustrated in Figure 7, the feed rate of the feed solution strongly depends, following a linear relationship, on the concentration of the stock solution used for feeding the IVT reaction with fresh NTPs, preferably GTP when using mCap and/or ARCA as cap-analogue and the total volume of the reaction. In the grey shading, Figure 7A shows the range of feed rates that are feasible with microfluidics; all volume-feed stock concentration combinations that overlap this grey area are theoretically achievable, assuming an 8h continuous feeding protocol. As is evident from the figure, small volumes, as typically used for screening-purposes, would need impractical feeding rates if concentrated volumes were to be used. Instead,

using a diluted feeding stock, as discussed above, manageable feed rates can be achieved.

Similarly, following the minimal and maximal continuous evaporation rates from Table 4 for each volume, Figure 7B shows that using a cubic (all dimensions roughly equal) or surface optimized reaction container (minimal height fluid layer, with maximal surface area), evaporation rates that match or exceed the feed rates from Figure 7A can be achieved, assuming a maximal evaporation rate of 15µl/min/cm2. The grey marked area indicates all feed-rate/feed stock concentration combinations that can be compensated by evaporation from a device having a cubic design. In dark grey (Figure 7B), there is shown the feed-rate/feed stock concentration combinations that require either a evaporation rate of >15µl/min/cm2 or a device wherein the surface area of the reaction volume that is in contact with the airflow leading to evaporation is enlarged to enable enhanced evaporation. In Figure 7C, the results of Figure 7A and 7B are combined, furthermore considering the mixing kinetics, showing that feed stock concentrations of in between 1mM and 10mM provide the idealized conditions for rapid and complete mixing, sufficient control over the feed-rate (through microfluidics devices) and matched evaporation, for volumes <1ml. For volumes >1ml, the ideal feeding stock concentration is around 10mM for rapid and complete mixing, as well as precise control over feed rate and matched evaporation.

In a preferred embodiment, therefore, evaporation of the excess water is achieved by the application of a controlled gas flow over the reaction mixture. The gas may be nitrogen or air, for example. However, air is preferred. Preferably, the magnitude of the gas flow is precisely controlled. Preferably, the gas flow is in direct contact with the surface of the reaction volume, and preferably not in contact with the feed solution flow. To avoid contamination, the gas flow is preferably sterilized, more preferably filtered. For example, the gas (preferably air) flow may be sterilised by passing it through a HEPA filter.

To increase the effectiveness of the evaporation, the speed of the gas flow may be at least 0.5m/s, at least 2m/s, at least 3.5m/s, at least 5m/s, at least 10m/s, or >10m/s. However, it will be appreciated that care should be taken to avoid foaming and splashing of the reaction mixture.

In another embodiment, as an alternative, or complementary method of increasing

the effectiveness of the evaporation, the humidity of the gas (preferably air) may be decreased prior to contacting the gas flow with the surface of the reaction mixture. In a preferred embodiment, the humidity is below 50%, more preferably below 40%, and even more preferably below 30%. The modification of the humidity may be achieved by any means known to the skilled person.

In some embodiments, to increase the effectiveness of the evaporation, the temperature of the gas flow (preferably air) may be between 4°C and 50°C. In a preferred embodiment, the gas temperature is maintained at the same temperature as the reaction mixture to avoid cooling the fluid. In another embodiment, the gas temperature is maintained at room temperature (e.g., 19-23°C).

In another embodiment of the invention, the removal of excess fluid (preferably water) is achieved via osmosis by contacting a sufficiently large surface of the reaction volume with a semi-permeable or water-selective membrane, such as a PDMS-membrane, a dialysis membrane or a molecular sieve. Preferably, the membrane has a pore size which is smaller than the protein, DNA and RNA components of the reaction, but larger (and therefore allowing the passage of the components) than NTPs, buffering components, salts, reductive agents and/or spermidine.

In such an embodiment, the driving force for water removal is a substantially larger osmolality on the receiving side of the membrane compared to the IVT reaction composition. Such increased osmolality may be achieved by the presence of high concentrations of salt (*e.g.,* NaCl), polymeric substances, and other water-interacting molecules. Transfer of the osmotic substance through the membrane is preferably avoided by the membrane selectivity to avoid changing the reaction composition.

Preferably, the method comprises feeding pressurised air (which may be optionally filtered, e.g. with a HEPA filter) into the reaction chamber in order to provide sufficient pressure to drive a predictable and/or sufficiently large volume of feed solution through the semi-permeable membrane.

In a preferred embodiment of the invention, the method of removal of excess water is (continuously) adjusted and/or fine-tuned via a PID (proportional integral

derivative controller) loop responding to measurements of the fluid volume and/or osmolality of the IVT reaction volume.

In a preferred embodiment of the invention, the method comprises measuring the reaction fluid level, preferably by optically measuring the fluid height. Such optical measurements may be performed in a substantially vertical direction, thereby detecting the reflection on the fluid surface. Such measurements can be achieved by, for example, measuring the angle-of-return via infrared measurements. Alternatively, the optical measurement may be performed in a horizontal direction, either detecting the breaking of the optical path by the surface of the fluid, or the changed absorption by the fluid. A variety of wavelengths are suitable for measuring fluid levels.

In another embodiment, the reaction fluid level may be measured by an electronic conductive probe. Such a probe may be in inserted into the container containing the IVT reaction. A conductive probe may function either by providing a signal once the reaction fluid makes direct contact with the probe that is positioned directly above the desired fluid level, or by providing a proportional signal by being partially or fully immersed in the reaction fluid.

In another embodiment, the reaction fluid level may be measured by continuously weighing the entire system and subtracting the empty weight of the reaction container, as well as the feeding solution not yet incorporated in the reaction volume. The remaining weight directly corresponds to the volume of the IVT reaction. Excess weight compared to the theoretical or starting weight of the solution is determined to be excess fluid, and calculations can be carried out accordingly.

In another embodiment, the reaction fluid level may be measured by acoustic distance sensing, using a probe that is substantially above the fluid surface, thereby detecting the reflection of the sound waves on the fluid surface. A variety of wavelengths are suitable for measuring fluid levels.

In another embodiment, the method may comprise calculating excess fluid from the measurement of the osmolality of the solution and comparing this to the desired osmolality of the solution at that stage of the IVT reaction. Such a method relies on a measurable difference in osmolality in the feed solution compared to the

osmolality of the IVT reaction in certain stage(s) of the IVT reaction. Methods for measuring the osmolality are known to those ordinarily skilled in the art.

In one embodiment of the invention, the method comprises determining the feed rate of the feed solution based on empirical evidence or a theoretical calculation of the consumption rate of the competing nucleotide (and cap-analogue) by a known amount of RNA polymerase with known activity in equal or similar reaction compositions, additionally accounting for the composition of the RNA sequence, preferably mRNA. In a preferred embodiment, the feed rate is started at the start of the reaction and substantially matches the consumption rate to maintain the cap-analogue-to-competing nucleotide ratio. Alternatively, the feeding is started after a certain amount of time has elapsed.

In another embodiment, the method comprises adjusting (preferably, continuously) the feed rate of the feed solution is based on real-time or intermittent measurements of the reaction substrates (*e.g.,* one or more nucleotides, preferably the competing nucleotide) and/or the reaction products (preferably the mRNA).

In a preferred embodiment of the invention, the method comprises feeding (or replenishing), to the reaction, the nucleotide (*e.g.* rGTP) that is competing with the nucleotide-cap analogue (*e.g.* ARCA) by feeding the competing nucleotide (e.g., rGTP) to the IVT reaction mixture, preferably from an external reservoir. The feeding (or replenishing) may be continuous or intermittent. In such a case, the method preferably comprises continuously or intermittently mixing the reaction reservoir to resolve local differences in the concentration of the competing nucleotide (e.g. rGTP concentration), and thus local differences in the ratio between the cap analogue and the competing nucleotide (e.g. ARCA:rGTP ratio). Such mixing is preferably performed in a way not resulting in disruption of protein/enzyme function due to shear stress, or disrupt the IVT reaction in another way.

In a preferred embodiment of the invention, the method comprises feeding the feed solution to the reaction in a distributed manner, thereby preventing the formation of temporal high local concentrations of the feed solution, and optionally sub-optimal cap-analogue-to-competing nucleotide ratios. Such distributed feeding preferably comprises releasing the feed solution at space apart locations within the reaction volume, either simultaneously or sequentially. In a preferred embodiment,

and advantageously, distributed feeding results in a minimization of the fluid stream/droplet/solid particle compared to the surrounding transcription reaction fluid, to thereby enable faster diffusion, mixing and/or dissolving, as well as minimize the local, temporal deviation of the desired reaction conditions, specifically the cap-analogue-to-competing nucleotide ratio.

Preferably, the method comprises feeding, to the reaction, either intermittently or continuously, the competing nucleotide that is competing with the nucleotide-cap analogue, and/or any other reaction components, through a semi-permeable membrane, wherein the semi-permeable membrane comprises a plurality of axially, radially and/or longitudinally spaced apart microscopic openings and/or pores through which the feed solution is passed.

In a preferred embodiment of the invention, the method comprises feeding, to the reaction, the competing nucleotide (*e.g.* rGTP) that is competing with the nucleotide-cap analogue (*e.g.* ARCA), and/or any other reaction components, through a microfluidic channel. Preferably, the microfluidic channel comprises a plurality of axially, radially and/or longitudinally spaced apart microfluidic openings through which the feed solution (preferably the competing nucleotide) may be passed. The reaction container is preferably intermittently or continuously mixed to distribute such freshly added nucleotide throughout the IVT reaction mixture. The feeding may be continuous or intermittent. In a preferred embodiment, the reaction container is part of, or leak-free affixed to, the microfluidic channel feeding the nucleotide.

In another embodiment of the invention, the method comprises feeding, to the reaction, the competing nucleotide (*e.g.* rGTP) that is competing with the nucleotide-cap analogue (*e.g.* ARCA), and/or any other reaction components, by means of one or more needles, pipette tips, robotic probes, serological pipettes, Pasteur pipettes, or (an)other suitable probe(s). The feeding may be continuous or intermittent.

Preferably, the one or more needle, pipette tip, robotic probe, serological pipette, Pasteur pipette, or (an)other suitable probe comprises a plurality of axially, radially and/or longitudinally spaced apart microfluidic openings through which feed solution (preferably the competing nucleotide) may be passed. The one or more needle, pipette tip, robotic probe, serological pipette, Pasteur pipette, or (an)other

suitable probe may be operated by direct displacement (*e.g.* a plunger of solid material), pumped by an immiscible fluid, pumped by air pressure (*e.g.* by a (robotic) pipette, or (continuous pressure source)), magnets, or gravity. The probes may hold part of, or the entire volume of the competing nucleotide (*e.g.*, rGTP) or mixture of the competing nucleotide (e.g., rGTP) and cap-analogue (e.g., ARCA). Alternatively, or in addition, the probes may be connected directly, via tubing, via pipes, via microfluidic channels, or via any other suitable means to an external reservoir holding the nucleotide to be fed to the reaction (*e.g.,* rGTP or the mixture of rGTP and ARCA).

The or each probe and/or connected reservoir may be cooled (for preservation), and/or (pre-)heated (for the addition of pre-warmed solution). Preferably, the or each probe is pre-heated and the storage reservoir is cooled such that the nucleotide/nucleotide mix has the highest stability during storage until immediately prior to addition to the reaction mix. Advantageously, pre-heating the solution containing the competing nucleotide and/or ARCA prevents temperature fluctuations of the reaction mixture and more stable synthesis conditions.

In one embodiment of the invention, the method comprises feeding, to the reaction, the competing nucleotide (*e.g.* rGTP) that is competing with the nucleotide-cap analogue (*e.g.* ARCA) by means of one or more needles, pipette tips, robotic probes, serological pipettes, Pasteur pipettes, or another suitable probes, wherein such probe contains a plurality of independent doses of said nucleotide.

In a preferred embodiment, the independent doses are separated by air gaps or another suitable immiscible medium. In such embodiments, the probe may remain in contact with the reaction vessel and does not need to be refilled between doses, allowing the maintenance of a closed system and limit the chance of introduction of contamination. The feeding may be continuous or intermittent.

In one embodiment of the invention, the method comprises feeding, to the reaction, the competing nucleotide, and/or any other reaction components, over the surface of the reaction volume by spraying, nebulizing or any other means of creating substantially small droplets. Preferably, droplets are distributed substantially evenly over the fluid comprising the reaction volume. The spray or nebulized fluid may be applied to part of the entire surface of the reaction volume.

In a preferred embodiment, the spray or nebulized fluid is to a portion of the reaction volume surface that is separate from (i.e. spaced apart from) the reaction volume surface that is exposed to air flow for the purpose of evaporating excess fluid.

In one embodiment of the invention, the method comprises feeding, to a reaction, the competing nucleotide (*e.g.* rGTP) that is competing with the nucleotide-cap analogue (*e.g.* ARCA) dissolved in a suitable liquid medium (*e.g.* a buffer, (RNAse-free) (demineralized) water, or another aqueous solution) or as solid (*e.g.,* dry powder), optionally bound to a carrier. When added as a solid, preferably a fine dry powder, it is dissolved directly into the IVT reaction mixture. The rate of addition is such that the solid dissolves right away and limited so that no high local concentrations of the nucleotide is formed. The feeding may be continuous or intermittent.

In another embodiment of the invention, the method comprises feeding, to a reaction, the competing nucleotide (*e.g.* rGTP) that is competing with the nucleotide-cap analogue (*e.g.* ARCA), and/or any other reaction components, by dissolving, bursting, swelling, degrading, exchanging, or otherwise releasing from a carrier that is added at the start of the reaction/incubation, and/or during the reaction, and/or intermittently fed, and/or continuously fed to the reaction. Hence, the feeding may be continuous or intermittent. Certain carriers may display a single or multiple types of release mechanism (*e.g.* swelling and degrading). Such carriers may take the form of a microbubble, liposome, solid, salt, polymer, matrix, hydrogel, or other suitable form of carrier that allows slow-release and/or controlled release of the nucleotide of interest to the solution.

The process or reaction leading to the release of the nucleotide of interest may be autocatalytic, temperature induced, induced by a certain pH or change in pH, catalysed by an external factor (*e.g.,* light of a specific wavelength), or in a preferred embodiment catalysed by any IVT product or metabolite.

In one embodiment of the invention, the method comprises feeding, to the reaction, the competing nucleotide in combination with other (rate- or yield-limiting) nucleotides.

In one embodiment of the invention, the method comprises feeding, to the reaction, the competing nucleotide in combination with a matched amount of

magnesium salt (*e.g.,* chloride or acetate) to maintain a steady amount of free magnesium ions in the solution.

In another embodiment of the invention, method comprises feeding, to the reaction, the competing nucleotide is added in an amount of liquid that matches the evaporation from the reaction since the last feeding/start of the IVT reaction.

In yet another embodiment of the invention, the method comprises feeding, to the reaction, competing nucleotide is added in a reaction buffer with a concentration that maintains the concentration of buffer and ions of total reaction at steady/idealized levels.

In one embodiment of the invention, both the nucleotide (*e.g.* rGTP) competing with the cap analogue (*e.g.* ARCA), as well as the cap analogue itself may be present at substantially low concentrations, resulting in the depletion or reduction in concentration of both. In such cases, the method comprises feeding, to the reaction, both the competing nucleotide and the cap analogue in a suitable relative ratio and rate. In a preferred embodiment, the relative ratio of competing nucleotide and cap analogue in the solution fed to the reaction is substantially similar to the relative consumption of each component during the transcription of a given RNA sequence (*e.g.* 1 ARCA to 250 rGTP for an mRNA sequence containing 250 G). The rate of supplying the nucleotide (as a mixture) is dictated by the reaction speed of the IVT reaction and the concentration of the nucleotide (mixture) in the feeding solution.

In another embodiment, the feed solution preferably contains, in addition to said competing nucleotide, one or more components of the IVT reaction, such as the buffer and/or one or more of the other NTPs. By feeding such reaction components in addition to the rate-limiting nucleotide, the reaction conditions may be maintained within pre-determined parameters even if other components of the IVT mixture are diluted by the feed solution.

In a preferred embodiment of the invention, the method comprises dissolving the cap-analogue-competing nucleotide (*e.g.* GTP) in (un)buffered RNAse-free water. Preferably, the method comprises feeding dissolved competing nucleotide directly to the IVT reaction at a concentration and rate suitable to maintain the rate-limiting nucleotide in a desired ratio to the cap-analogue. The concentration of the rate-

limiting nucleotide (*e.g.* rGTP) in the feed solution is preferably such that the volume of feed solution added to the IVT reaction matches the rate of evaporation from the IVT reaction in a given reaction container, preventing any substantial dilution or concentration of the IVT mixture. The evaporation rate from the IVT mixture may be controlled or modified by changing the shape, volume or material of the reaction container, as well as modifying the volume of air contacted with the surface of the solution, environmental temperature (including air temperature), reaction temperature, relative humidity of the air or gas contacting the fluid surface, presence of moist absorbing solids and other methods known to those ordinary skilled in the art.

In one embodiment of the invention, the method comprises producing multiple sequences of RNA in a single IVT reaction. The multiple sequences of RNA may comprise modified and unmodified RNA, preferably mRNA. Preferably, for each (modified) (m)RNA sequence, a separate cap analogue is used. For example ARCA with first binding nucleotide G on the first RNA sequence, and a cap-analogue complementary to T on the second RNA sequence, and so on. In such use cases, the methods of the invention may be adapted to the feeding of 2 rate-limiting nucleotides simultaneously or independently.

The apparatus of the third aspect is preferably used to perform the method of the first aspect.

Therefore, in one embodiment, the reaction chamber comprises a first section into which the feed solution is fed, and a second section in which the transcription reaction occurs, wherein the first and second section are separated by a semi-permeable membrane. Preferably, the semi-permeable membrane is a dialysis membrane. Preferably, the semi-permeable membrane comprises a pore size which is smaller than the protein, DNA and RNA components of the reaction, but larger than, and therefore allowing the passage therethrough of NTPs, buffering components, salts, reductive agents and/or spermidine.

In one embodiment (as shown in Figure 8A), the first section is an outer volume of the reaction chamber, and the second section is an inner volume of the reaction chamber at least partially enclosed by the semi-permeable membrane. In another embodiment (as shown in Figure 8B), however, the first section is an inner volume

of the reaction chamber, and the second section is an outer volume of the reaction chamber at least partially enclosed by the semi-permeable membrane.

Preferably, the feed means is configured to feed the feed solution to the first section of the reaction chamber, thereby resulting in contact with one side of the semi-permeable membrane, wherein positive pressure created by feeding fluid drives the feed solution and its contained components through the semi-permeable membrane. It will be appreciated that the semi-permeable membrane provides a distributed manner of mixing in the feeding solution.

In one embodiment, the apparatus may comprise a gas inlet configured to feed pressurized gas into the first section of the reaction chamber, thereby providing additional pressure on the feed solution. The gas may be air or nitrogen. Air is preferred. The gas inlet may comprise a filter, such as HEPA filter to prevent contamination. Advantageously, by providing a continuous pressure to the feed solution by the pressurized gas, it is thereby forced through the membrane at a rate greater than the diffusion rate, preventing the transfer of transcription reaction components from the second section into the first section, thereby producing a uni-directional flow.

The apparatus preferably comprises a gas (preferably air) outlet configured to release excess pressure in the reaction chamber. Advantageously, this allows the feed solution to be continually passed from the first section to the second section.

In order to keep the volume of the transcription reaction substantially constant at steady state, such that the ratios of the various reagents (especially the cap analogue : competing nucleotide ratio) are kept constant, excess fluid is removed from the reaction by evaporation. Therefore, preferably the reaction chamber comprises an evaporation zone which is disposed above the section in which the transcription reaction occurs, and in which fluid (preferably water) evaporates to compensate for the additional volume provided by the feed, to thereby keep the reaction volume substantially constant. Preferably, the apparatus comprises a gas inlet configured to feed gas into the evaporation zone to facilitate evaporation of the fluid. The gas may be air or nitrogen. Air is preferred. The gas inlet may comprise a filter, such as HEPA filter to prevent contamination. Preferably, the gas inlet is in fluid communication with gas outlet.

In another embodiment (as shown in Figure 8C), the feed means preferably comprises a probe which comprises a distal outlet through which the feed solution is passed. Preferably, the outlet is arranged to extend into the reaction liquid in the reaction chamber. The probe may be a needle, a pipette tip, a Pasteur or serological pipette, or any other suitable probe for the addition of feed solution. The apparatus may comprise a magnetic or physical stirrer, which may be optionally connected to the probe.

In yet another embodiment (as shown in Figure 8D), the probe comprises a plurality of axially, radially and/or longitudinally spaced apart outlets through which the feed solution is passed, preferably into the liquid.

In yet a further embodiment (as shown in Figure 8E), the probe comprises a spraying device or nebulizer which is configured to produce droplets of the feed solution in the reaction chamber. Preferably, the spraying device or nebulizer is mounted above the surface of the reaction liquid. Preferably, a substantial part (preferably all) of sprayed droplets combine with the reaction liquid.

Preferably, the apparatus comprises a mixer, for example a stirrer, which is configured to mix the feed solution in the second section. Alternatively, the apparatus may comprise a shaking platform on which the reaction chamber is placed to stir the reaction volume.

All of the features described herein (including any accompanying claims, abstract and drawings), and/or all of the steps of any method or process so disclosed, may be combined with any of the above aspects in any combination, except combinations where at least some of such features and/or steps are mutually exclusive.

For a better understanding of the invention, and to show how embodiments of the same may be carried into effect, reference will now be made, by way of example, to the accompanying Figures, in which:-

Figure 1A shows the yield and nano-luciferase reporter mRNA activity produced in IVT reactions with varying nucleotide compositions (10mM of each NTP, 10mM of CTP/ATP/UTP and 2.5mM GTP, or 2.5mM of each NTP) and DNA templates containing differentiated transcription start sites (TSS) (GGG…, GGA…, GAA…). The

effect of nucleotide composition of the IVT reaction and the effect of the TSS on the yield of the IVT reaction over time, measured by means of extracting, purifying, and measuring the absorbance of purified samples taken at indicated timepoints. Each aliquot was 20µl. **Figure 1B** shows the luciferase activity of the reporter mRNA produced by the co-transcriptional IVT reaction of (A).

Figure 2 shows the yield of (m)RNA of a fed-batch embodiment of the method of the invention, wherein all non-competing nucleotides were provided at the start of the reaction at 10mM, and the competing nucleotide (GTP) at 2.5mM, followed by re-feedings with the equivalent of 2.5mM (final concentration) of GTP every 2h to increase the yield of the reaction, while maintaining a high cap-analogue (ARCA) to competing nucleotide (GTP) ratio. 20µl samples were taken after 15, 30, 45, 60, 90, 120, 180, 240, 300, 360, and 480min, quenched with EDTA, silica-column purified and the concentration was measured via UV-absorption.

Figure 3 demonstrates the effect of increasing the cap-analogue (ARCA) to competing nucleotide (GTP) ratio when using a DNA template with a GGG TSS and 10mM of ATP, CTP, UTP and ARCA, as well as a variable amount of GTP at the start of the reaction, on the yield and mRNA activity. After 2h incubation at 37°C, the 20µl reactions were terminated by quenching with EDTA and purified by silica-column before measuring the yield with UV-absorbance at 230/260/280nm. **Figure 3A** shows the effect of ARCA-GTP ratio on the yield of the reaction after 2h incubation. **Figure 3B** shows the effect of ARCA-GTP ratio on the reporter mRNA activity after transfection on HeLa cells.

Figure 4 shows the effect of the concentration of the feeding solution containing 1-100mM of GTP on the yield and activity of the mRNA from a fed-batch IVT reaction run for 30h total, with 3h intervals between feedings, starting with a 20:1 ARCA-to-GTP ratio. **Figure 4A** shows the yield of the reaction, compared with the theoretical yield from the reaction. **Figure 4B** shows the mRNA activity of the various samples prepared with various feeding solution concentrations.

Figure 5 shows the yield and mRNA activity of reactions fed 9 times with a diluted feeding solution containing 5mM GTP in a closed vessel, a vessel opened to the ambient environment, and a vessel connected to a dry N2 gas stream, set to evaporate a similar volume as added during the feeding over a period of 3h. **Figure**

5A shows the yield. **Figure 5B** shows the Luciferase mRNA activity 24h after transfection on HeLa cells.

Figure 6 shows three embodiments for feeding the feeding solution containing nucleotides, *i.e.,* **Figure 6A**: a fed-batch reaction with a single feeding (of competing nucleotide) producing large fluctuations in the cap-analogue to competing nucleotide ratio and thus in the capping efficiency; **Figure 6B**: similar to A, a fed-batch reaction, however, with multiple feeds of competing nucleotide, producing smaller fluctuations in cap-analogue to competing nucleotide ratio, and thus also smaller fluctuations in capping efficiency. Interestingly, theoretical modelling shows that the average reaction speed is faster with more feedings, allowing >2 feedings for each of the 2x larger feedings in A, while maintaining the same capping efficiency; and **FigureC**: continuous feeding with the competing nucleotide at the rate of consumption of such competing nucleotide, producing a steady cap-analogue to competing nucleotide ratio and thus a steady capping efficiency, as the higher productivity of the system. **Figure 6D** shows the productivity of the reaction using A, B or C feeding methods, showing that the average productivity per unit of time of the IVT reaction increases with an increased frequency of feedings, as long as the feedings feed the competing nucleotide at the rate of consumption. Despite being on average faster, the more frequent feedings are initially slower, due to the initially higher transcription rate related to the increased levels of the rate-limiting competing nucleotide. These results only apply to conditions wherein the competing nucleotide is on average and/or continuously present at rate-limiting concentrations. Once the competing nucleotide is present above rate-limiting concentrations, the difference in productivity of each method (A,B, or C) is negligible.

Figure 7 shows the results of theoretical modelling of the feed-rate, evaporation rate and mixing kinetics for a variety of total reaction volumes and feed stock concentrations. **Figure 7A**: shows the feed rates required to maintain the desired capping efficiency during a continuous feeding of 8h for each of the volumes. The grey area shows the range of feed rates that can practically be achieved with microfluidics, according to Table 3. **Figure 7B**: shows the feed rates required to maintain the desired capping efficiency during a continuous feeding of 8h for each of the volumes similar to A. The grey area shows the range of feed rates that can be compensated for by evaporation in a cubic device, assuming a maximum evaporation rate of 15µl/min/cm2. The dark grey area shows the feed rates that

require a device that optimizes the surface area of the reaction to be able to compensate the flow rate with evaporation, assuming a maximum evaporation of 15µl/min/cm2. **Figure 7C**: shows the feed rates required to maintain the desired capping efficiency during a continuous feeding of 8h for each of the volumes similar to A. The information from 7A, and 7B is combined with estimates of the mixing kinetics extrapolated from Figure 4B to display the most suitable feed rate, feed stock, total volume combination wherein a high quality mRNA can be produced via co-transcriptional capping IVT, with a high yield and a constant volume.

Figure 8 shows six different embodiments of reactor design for performing the methods of the invention.

Examples

The inventors set out devise a novel method for *in vitro* transcription (IVT) of mRNA that enable the *in vitro* transcription of mRNA with higher 5' capping yields, at lower costs of goods than the current state of the art. Considering the enormous economic significance of *in vitro* transcribed mRNA in medicine and biotechnology, those skilled in the art will recognize the enormous economic potential of the present invention. Moreover, the present invention relates to devices that enable an improved synthesis of *in vitro* transcribed mRNA, and more specifically to the addition of a 5'cap structure to *in vitro* transcribed mRNA, via said protocols. The resultant mRNA products bear a higher degree of 5'cap structures with less consumption of expensive reagents and simplified purification procedures, which lowers the cost of *in vitro* transcription of mRNA and causes less activation of innate immune responses by 5' uncapped mRNA which in turn yields higher protein expression and causes less unwanted side effects of the *in vitro* transcribed mRNA when used in *in vitro* or *in vivo* applications.

Example 1 – Batch IVT reactions to determine the influence of nucleotide composition on the productivity and mRNA quality

200ng of each PCR-generated DNA templates containing a 5' T7-promoter, and 3' encoded poly-A-tail of 150nt was added to a 200µl IVT reaction. The DNA templates were identical, except for the first 3 nucleotides after the T7 promoter, the so-called transcription start site (TSS):-

Template A: T7_promoter-GGG-remainder of 5'UTR-CDS-3'UTR-poly-A-tail,
Template B: T7_promoter-GGA-remainder of 5'UTR-CDS-3'UTR-poly-A-tail,
Template C: T7_promoter-GAA-remainder of 5'UTR-CDS-3'UTR-poly-A-tail.

Three variants of the batch IVT reaction were prepared by mixing 2µl M11 T7RNAP (RiboPro), 2µl of Hiscribe reaction buffer (NEB), 1µl RNAse-inhibitor (NEB), 1 µl IPP (NEB), and variable amounts of ATP, CTP, UTP and GTP from 100mM stock (NEB) to achieve reaction composition 1, containing 10mM of each NTP, reaction composition 2, containing 10mM of UTP, CTP and ATP, as well as 2.5mM GTP, and reaction composition 3, containing 2.5mM of each NTP. All reactions contained 10mM ARCA from a 50mM stock solution (Jena Bioscience).

Directly after mixing, the IVT reactions were incubated for 3h at 37°C in closed PCR tubes, allowing minimal to no evaporation. Samples of 20µl were taken at indicated timepoints and immediately mixed with 80µl of 25mM EDTA (Sigma-Aldrich) to quench the reaction. The samples were subsequently purified by standard silica column (Monarch kit, NEB) purification according to manufacturer's protocol and double eluted with RNAse-free water (Invitrogen) before measuring 230/260/280nm absorption on the iD3 platereader (Molecular Devices) in a 384-well plate with transparent bottom and black walls (Greiner).

Purified mRNA was subsequently mixed with PBS and pre-diluted Lipofectamine MessengerMax (Invitrogen) according to manufacturer's protocol. 100ng of purified mRNA was hereby transfected in 80% confluent HeLa cells in DMEM/F12 medium + 10% FCS. After 24h incubation, medium was collected and assayed with the Secreted Nanoluciferase assay (Promega) according to manufacturer's protocol in a black 384-well plate (Greiner) on the iD3 platereader (Molecular Devices) collecting all wavelengths with medium sensitivity settings. Samples of cells treated with only Lipofectamine and buffer were used for background subtraction.

The results, as shown in Figure 1A, clearly show that a higher total amount of NTPs results in a higher yield, with a near perfect correlation between the amount of the most limiting NTP and the mRNA yield. IVT reaction containing 10mM of each NTP can theoretically achieve >220µg yield, however, in this assay the maximum yield was roughly 180-195µg, probably due to the limited incubation time and potential losses during purification. IVT reactions containing either all NTPs or only GTP at 2.5mM showed a maximum of roughly 50µg yield. Interestingly, despite producing a similar final yield, reactions containing 10mM of UTP, CTP and ATP and 2.5mM of GTP showed a higher yield at shorter timepoints, especially <90min, than the IVT reactions containing 2.5mM of each NTP, and reach saturation earlier. No

differences in the final yield were found between different DNA templates harbouring different TSS.

Surprisingly, the activity of the produced mRNA trended upwards for all reactions, as shown in Figure 1B, suggesting a higher quality of mRNA being produced near the end of the co-transcriptional IVT reaction. This effect became more pronounced for samples with a 2.5mM of competing nucleotide (rGTP; competing with ARCA for incorporation as the first nucleotide in the reaction), suggesting a higher co-transcriptional capping with ARCA near the end of the reaction. In any case, a lower concentration of rGTP delivered a pronounced effect on the activity of the mRNA. Interestingly, this effect was correlated with the TSS used; GGG and GGA start-sites showed similar mRNA activity for all respective reaction conditions, however, for the GAA template, especially the reaction containing 10mM of UTP, ATP and CTP, and 2.5mM of GTP showed a lower activity than the reaction containing 2.5mM of each NTP. A possible explanation is the T7 RNAP polymerase may skip the first nucleotide of the DNA template and incorporate the second templated nucleotide. For the GGG and GGA containing templates still either a GTP or ARCA is incorporated, according to the then-current ratio between both. However, in the GAA template, the more concentrated ATP could be substituted for GTP or ARCA, despite being less favoured for transcription initiation by the enzyme.

These results suggest that reactions containing a lower amount of cap-analogue competing nucleotide produces a higher capping efficiency and a higher quality mRNA.

Example 2 - Fed-batch IVT reactions to increase the yield of the reaction

IVT reactions were prepared as in Example 1, except that only DNA template A (GGG TSS) was used and the volume increased to 250µl by scaling the components proportionally. A reaction containing 10mM of each NTP and 10mM of ARCA was used as control for reactions containing 10mM of CTP, UTP and ATP, and 2.5mM GTP. Immediately after preparation, the IVT reactions were incubated at 37°C and samples of 20µl were removed at indicated timepoints and processed similar to Example 1. After 2h incubation, samples C, D and E received an additional dose (fed batch) equivalent to 2.5mM (final concentration) of GTP, which was gently mixed into the reaction by pipetting and returned to incubation at 37°C. After an additional 2h incubation, samples D and E received an additional dose (fed batch) equivalent to 2.5mM (final concentration) GTP, and were further incubated at 37°C.

After an additional 2h incubation, sample E received a final dose (fed batch) of equivalent to 2.5mM (final concentration) GTP, and was incubated once more for 2h. The reaction time was equal for all samples; 8h in total.

The results shown in Figure 2 clearly show the benefit of the fed-batch method, increasing both the productive phase and overall yield of the reaction. Each addition of GTP increases the yield of the IVT reaction. With up to 3 feedings of GTP (dose equivalent to 2.5mM final concentration), the total amount of GTP added is the same as added at the start to the control reaction, and the total yield approaches the yield of the control reaction. Interestingly, for each subsequent dose of GTP, the additive amount of (m)RNA generated is slightly lower than the amount produced under the previous amount and the transcription rate is slightly reduced, producing ultimately 10% less in sample E compared to sample A.

Example 3 - High cap-analogue-to-competing nucleotide ratio increases mRNA activity but decreases yield

Batch IVT reactions were prepared as in Example 1, except that only DNA template A (GGG TSS) was used and the volume was decreased to 20µl by scaling the components proportionally. All reactions contained 10mM of CTP, UTP, ATP and ARCA, and variable amount of GTP. Immediately after preparation, the IVT reactions were incubated at 37°C and after 2h processed similar to Example 1.

The results, as displayed in Figure 3A, show a dose-dependent reduction in yield from the IVT reaction, which deviates from theoretical predictions, the higher the ARCA to GTP ratio, suggesting slower transcription when lowering the GTP concentration. This may be compensated by longer incubation times. Interestingly, also a dose-dependent increase in mRNA activity was observed, with higher ACRA:GTP ratio producing a more active mRNA. Surprisingly, the increase in activity exceeded the increase in theoretical capping efficiency.

Example 4 - Fed-batch reactions from concentrated and diluted stock

Based on the high activity of the resultant mRNA from IVT reactions with elevated ARCA:GTP ratio, for subsequent experiments, a 20:1 ratio of ARCA over GTP was chosen.

IVT reactions were prepared as in Example 1, except that only DNA template A (GGG TSS) was used and the volume was decreased to 100µl by scaling the

components proportionally. All reactions contained 10mM of CTP, UTP and ATP, and 0.5mM GTP. Immediately after preparation, the IVT reactions were incubated at 37°C. Sample A received every 3h a dose (fed batch) of GTP equivalent to 0.5mM final concentration from a stock solution of 100mM (0.5µl), Sample B received every 3h a dose of GTP equivalent to 0.5mM final concentration from a stock of 10mM (5µl), Sample C received every 3h a dose of GTP equivalent to 0.5mM final concentration from a stock of 1mM (50µl). After addition, the added volume was gently mixed into the reaction by pipetting, which was subsequently returned to incubation at 37°C. To avoid changing the salt concentration of the IVT reaction, the diluted stock solutions were prepared with 1x IVT reaction buffer. After 9 doses, the reaction was terminated by addition of an excess of EDTA and processed similar to Example 1.

The addition of 0.5µl of 100mM GTP by hand proved to be challenging. However, as is shown in Figure 4A, using a more dilute feeding solution reduces the yield of the reaction, presumably due to dilution of the reaction components. Surprisingly, as shown in Figure 4B, the activity of the resultant mRNA was higher for IVT reactions fed with mildly diluted feed solution.

Example 5 – Fed-batch reaction from diluted stock with compensatory water evaporation

Based on the results of Example 4, the inventors next explored whether a combination of dilute feeding solution and compensatory water evaporation in order to maintain a fixed reaction volume would result in an optimized yield and activity of the resultant mRNA. Such a protocol would allow for continuous automated feeding without (significantly) altering the concentration of the reaction via dilution, as the volume would remain essentially fixed at steady-state.

IVT reactions were prepared as in Example 4, Sample A (control) received every 3h a dose of GTP equivalent to 0.5mM final concentration from a stock solution of 100mM (0.5µl) in a closed tube, Sample B (control) received every 3h a dose of GTP equivalent to 0.5mM final concentration from a stock of 5mM (10µl) in a closed tube, Sample C received every 3h a dose of GTP equivalent to 0.5mM final concentration from a stock of 5mM (10µl) in an opened tube, Sample D received every 3h a dose of GTP equivalent to 0.5mM final concentration from a stock of 5mM (10µl) in an open tube into which a stream of dry N2 gas was blown at a rate

that was found in pilot experiments to evaporate around 3.3µl/h, thereby maintaining a steady volume.

Interestingly, as shown in Figure 5A, feeding from a concentrated stock (100mM – Sample C) or a diluted stock (5mM – Sample D) resulted in very similar yields when excess fluid was allowed to evaporate. Furthermore, surprisingly, as shown in Figure 5B, mRNA activity resulting from the IVT reaction using the diluted stock (Sample D – 5mM) was increased compared to IVT reactions fed from the more concentrated stock (Sample C – 100mM).

Example 6 – IVT reactor designs for achieving fixed volume, fed-batch IVT

Referring to Figure 8A, there is shown a first embodiment of a reactor 2 for performing the IVT method according to the invention. The reactor 2 includes a reaction container 4 consisting of an outer chamber 5, and an inner reaction volume 6 in which the IVT reaction occurs, enclosed by a semi-permeable membrane 8 (preferably a dialysis membrane 8), with a pore size smaller than the protein, DNA and RNA components of the reaction, but larger (and therefore allowing the passage of the components) than NTPs, buffering components, salts, reductive agents and/or spermidine.

At reaction setup, a feed solution 10 comprising the IVT reaction feed components is fed along a feed conduit 12 to the outer chamber 5 of the reaction container 4 resulting in contact with the outside of the semi-permeable membrane 8 that encloses the inner reaction volume 6. The positive pressure created by feeding fluid 10 into an otherwise closed system, drives the feeding solution 10 and its contained components through the semi-permeable membrane 8 (in the direction of arrows A) from the outer chamber 5 to enlarge the volume of the inner reaction volume 6.

In case insufficient pressure is generated to drive a predictable and/or sufficiently large volume of feed solution 10 through the semi-permeable membrane 8 by the feeding of the feed solution 10 itself, additional pressure can be applied to the feeding solution 10 by feeding pressurized air 14 via air inlet conduit 16 into the outer chamber 5. To prevent contamination, the conduit 16 providing the additional air pressure is optionally outfitted with a HEPA filter 20. The use of a semi-permeable membrane 8 provides a distributed manner of mixing in the feeding solution 10. By providing a continuous pressure to the feeding solution 10, the feeding solution 10 can be pressed through the membrane 8 (from the outer

chamber 5 into the inner reaction volume 6) at a rate greater than the diffusion rate, preventing the transfer of IVT reaction components from reaction volume 6 into the feeding solution 10, effectively creating a uni-directional flow. Excess pressure is released from the system via air outlet 18 that connects to the top of the reaction container 4, effectively allowing the feeding fluid 10 to be pumped from the outer chamber 5 of reactor to the inner reaction volume 6 in the reaction container 4.

In order to keep the volume of the IVT reaction substantially constant at steady state, such that the ratios of the various reagents (especially the cap analogue : competing nucleotide ratio) are kept constant, excess fluid is removed from the reaction by evaporation. Thus, the reaction reactor 2 includes an evaporation zone 22, which is a pocket of air disposed above the inner reaction volume 6 in which the IVT reaction occurs, and contained within solid walls 23, at least part of which is optionally provided by air inlet 16. The evaporation zone 22 is fed with fresh (optionally conditioned and/or filtered) air 24 via an air conduit 26, and is also connected to air outlet 18. To ensure optimal mixing of the feed solution 10 within the inner reaction volume 6, the reaction container 4 is optionally fitted with a magnetic stirrer 28. Alternatively, the entire reactor 2 can be placed on a shaking platform (not shown), optionally moving in a "figure-8" or other mixing pattern, effectively stirring the reaction volume 6 along the inside of the semi-permeable membrane 8 (and the inner reaction volume 6).

Referring now to Figure 8B, there is shown a second embodiment of the reactor 2. In this embodiment, the arrangement of the inner reaction volume 6, and the outer chamber 5 containing the feeding solution 10 in contact with the semi-permeable membrane 8, from Figure 8A, have been inverted. Accordingly, in this embodiment of the reactor 2, an inner chamber 30, limited by the semi-permeable membrane 8, contains the feeding solution 10, provided via feed conduit 12. The feeding solution 10 is then pushed outwards (in the direction of arrows B) via excess pressure through the semi-permeable membrane 8 into an outer container 32 containing an reaction volume 34.

To enable controlled evaporation in the evaporation zone 22 to keep the reaction volume substantially constant, the air inlet conduit 16 providing conditioned air 14 is connected to the reaction container 32 such that the conditioned air 14 contacts the upper surface of the reaction volume 34, circulates around the inner chamber

30 containing the feeding solution 10, towards the air outlet conduit 18. The advantage of the second embodiment of the reactor 2 shown in Figure 8B compared to the first embodiment shown in Figure 8A is that the shape of the inner chamber 30 enclosed by the semi-permeable membrane 8 is maintained by the pressure applied on the semi-permeable membrane 8, without any additional structures.

Referring now to Figure 8C, there is shown a third embodiment of the reactor 2. In this embodiment, a single reaction container 36 includes a single chamber (in contrast to the embodiments shown in Figure 8A and Figure 8B), which contains the reaction volume 38 and the added feeding solution 10. The feeding solution 10 is provided to the closed system via probe 40, which in some embodiments may be a needle, a pipette tip, a Pasteur or serological pipette, or any other suitable probe for the addition of feeding solution 10. A magnetic or physical (optionally connected to the probe 40) stirrer 28 may be introduced to the reaction container 36 to enable mixing of the reaction volume with the feeding solution. Alternatively, the apparatus of Figure 8C can be placed on a shaking platform to mix the contents of the device. Similar to Figure 8A and 8B, the embodiment of Figure 8C contains an air inlet conduit 26 to allow addition of (optionally conditioned) air 24 to the reaction container 36, such that the air 14 interacts with the surface of reaction volume 38, and is removed via air outlet 18. Each is optionally outfitted with a HEPA filter 20 to prevent contamination. The evaporation zone 22 is clearly shown disposed in the space above the reaction volume 38 and ensures that excess fluid is removed via evaporation.

Referring to Figure 8D, there is shown a fourth embodiment of the reactor 2. It follows the same design principles as in Figure 8C, with the exception that the probe 40, via which feeding solution 10 is fed into the reaction volume 38, is modified to feed the feeding solution 10 in a distributed manner, preferably via a multitude of axially, radially and/or longitudinally spaced apart microfluidic openings 42. Mixing of the IVT reagents and evaporation to keep the reaction volume constant and at steady state is performed similarly to the embodiment of Figure 8C.

Referring to Figure 8E, there is shown a fifth embodiment of the reactor 2. This embodiment follows the same design as the embodiments shown in Figure 8C and 8D, except that the feeding probe 40 is a spraying device or nebulizer 44 mounted at a suitable location above the surface of the reaction volume 6, such that a

substantial (preferably all) part of the sprayed droplets 46 merge with the reaction volume 38, providing another method of distributed addition of the feeding solution 10.

Lastly referring to Figure 8F, there is shown a sixth embodiment of the reactor 2. In this embodiment, conduits 10 and 16 are shown being separate from one another rather than being joined together as shown in Figure 8A.

It is understood that the dimensions of each of the components of the apparatuses 2 of the invention can be adjusted, scaled or otherwise modified to fit the required volumes, feeding, mixing and evaporation rates. Furthermore, the relative dimensions may be modified to provide for devices with an optimized volume to surface ratio, or otherwise advantageous shape. Finally, it is understood to those skilled in the art that conduits may be connected to the device in a permanently or temporary affixed manner, via direct ligation, screwing, pressure clamping, force fitting or any other method known to those skilled in the art. Furthermore, the conduits feeding and removing fluids and/or air may be connected as a bundle or each separately to the container. Aspects of the device may be symmetrical, but may, in contrast to drawings, also be asymmetrical.

Conclusions

The inventors have devised a novel IVT method, specifically for co-transcriptional capping of RNA with cap-analogues in which they select for very high cap-analogue-to-competing nucleotide ratios (for therapeutically relevant capping efficiency). However, as a consequence, the yield of such reactions is very low to be practically or economical. They have therefore provided a fed-batch protocol to address the problems with yields. However, one issue with this is that feeding from a concentrated stock to prevent overly diluting the reaction can result in such low volumes that adequate fluid control is challenging, leading to large fluctuations in the reaction conditions. Conversely, feeding from diluted stock solutions results in an overly diluted reaction. Therefore, a preferred embodiment of the method of the invention relies on using diluted stock to obtain sufficiently precise control coupled with simultaneous/subsequent excess fluid removal to correct for the dilution. In addition, the inventors use a system of highly distributed addition of the feed solution to the reaction volume to prevent local, temporal deviations in the overall reaction composition.

Claims

1. A method of preparing a capped ribonucleic acid (RNA) molecule, wherein the method comprises carrying out a fed-batch transcription reaction in the presence of a cap analogue in a substantially constant volume, thereby resulting in a substantially capped RNA molecule.

2. The method according to claim 1, wherein the method comprises *in vitro* transcription (IVT).

3. The method according to either claim 1 or claim 2, wherein the RNA is selected from a group of RNA molecules consisting of: messenger RNA (mRNA), micro RNA (miRNA); interference RNA (RNAi); short interfering RNA (siRNA); short hairpin RNA (shRNA); anti-sense RNA; RNA aptamers; self-amplifying RNA (saRNA); coding RNA; non-coding RNA; and circular RNA.

4. The method according to any preceding claim, wherein the RNA comprises mRNA.

5. The method according to any preceding claim, wherein the method comprises contacting:

(i) a template nucleic acid sequence, optionally DNA,

(ii) an RNA polymerase, optionally selected from a group consisting of: T7; T3; SP6; KP34; Syn5; or other DNA-dependent RNA polymerase; or a mutated variant of any of these RNA polymerases,

(iii) a plurality of nucleotide triphosphates (NTPs), optionally selected from a group consisting of ATP, GTP, CTP and/or UTP, or modified variants thereof, including pseudoUTP, N1-methyl-UTP, m5CTP, m6ATP, m1ATP, Inosine triphosphate, hm5CTP, m1GTP, m7GTP, and m6AmTP, and

(iv) a cap analogue,

wherein the RNA polymerase transcribes the template nucleic acid with the plurality of NTPs and the cap analogue, to thereby form the capped RNA molecule, preferably in a single one-pot reaction.

6. The method according to any preceding claim, wherein the cap analogue is selected from a group of cap analogues consisting of: an mCap (m7G(5')ppp(5')G, TriLink Biotechnologies, Inc); an anti-reverse cap analogue (a version of mCap

modified by 3' OH methylation on the m7G) ARCA – ThermoFisher Scientific); and a trinucleotide cap-analogue (e.g., Cleancap – TriLink Biotechnologies); and modifications and/or combinations thereof.

7. The method according to any preceding claim, wherein the cap analogue mimics the 5' cap, and is comprised in the reaction mixture, and incorporated by RNA polymerase as the first nucleotide of the nascent RNA instead of a corresponding competing nucleotide.

8. The method according to any one of claims 5-7, wherein the method comprises contacting the template nucleic acid sequence, the RNA polymerase, the plurality of nucleotide triphosphates (NTPs), and the cap analogue in the presence of magnesium ions, wherein the concentration of magnesium ions is 1-5mM greater than the total concentration of NTPs present and/or fed to the reaction up to that moment, optionally wherein the method comprises adding any extra Mg2+ required following addition of NTPs during or immediately after such feeding, instead of providing excess Mg2+ at the start of the reaction.

9. The method according to any preceding claim, wherein the method comprises contacting the reagents with a buffering agent, wherein the concentration of buffering agent may be preferably 10-50 mM of Tris-HCl set at pH 7.8-7.9 (at 25°C).

10. The method according to any preceding claim, wherein the method comprises contacting the reaction reagents with base, wherein the concentration of base is equal or exceeding the amount of any acid added to, or generated during the reaction.

11. The method according to any preceding claim, wherein the method comprises contacting the reagents with a reducing agent, wherein the reducing agent is selected from a group consisting of Dithiothreitol (DTT), dithioerythritol (DTE), beta-mercaptoethanol, and Tris (2-carboxyethyl) phosphine (TCEP), optionally wherein the concentration of the reducing agent is 0.5-5 mM, or 1-3 mM, or about 1 mM.

12. The method according to any preceding claim, wherein the method comprises contacting the reagents with spermidine, wherein the concentration of spermidine is 0.1-10 mM, or 1-5 mM, or 2 mM.

13. The method according to any preceding claim, wherein the method comprises feeding the reagents in a dilute feeding solution, wherein the dilute feed solution comprises a concentration of feeding components such that the total number of feedings add >25% of the initial reaction volume, thereby resulting in >25% dilution of the components present in the initial reaction volume, but not or at a lower concentration in the feed solution, if no fluid removal is or would be applied.

14. The method according to claim 13, wherein the dilute feed solution comprises a concentration of feeding components such that the total number of feedings add >20% of the initial reaction volume, or add >40% of the initial volume, or add >100%, of the initial reaction volume, if no fluid removal is or would be applied.

15. The method according to claim 13 or claim 14, wherein the dilute feed solution is defined as a feed solution containing 1 or more components >2 times less concentrated, or >4 times less concentrated, or >10 times less concentrated, than their respective solubility limit and/or technical manufacturability limit.

16. The method according to any one of claims 13-15, wherein the diluted feed solution is defined by the dilution of the cap-analogue competing nucleotide compared to the most concentrated stock solution available for said competing nucleotide, and is at least 2x, 5x, 10x, optionally more than 25x, or more than 100x.

17. The method according to any preceding claim, wherein the method comprises feeding or replenishing, either continuously or intermittently, a competing nucleotide that is competing with the nucleotide-cap analogue at a rate that is suitable to maintain the ratio between the competing nucleotide and the cap analogue between pre-determined minimum and maximum thresholds, wherein the minimum threshold of nucleotide-cap analogue:competing nucleotide is determined by the desired capping efficiency, whereas the maximum threshold of nucleotide-

cap analogue:competing nucleotide is determined by the minimum desired reaction speed and/or yield of the reaction.

18. The method according to any preceding claim, wherein the method comprises feeding, either continuously or intermittently, only the competing nucleotide, instead of a mixture of the nucleotides.

19. The method according to claim 18, wherein the maximal desired concentration of each of the non-competing nucleotides is added to the reaction at start of the reaction, thereby enabling the highest transcription rate possible, and the transcription rate is only limited by the cap-analogue competing nucleotide provided at a concentration substantially lower than the non-competing nucleotide and/or cap-analogue.

20. The method according to any preceding claim, wherein the composition of the reaction at the start of the reaction comprises 5mM, 6mM, or 7mM of each of rATP, rCTP, rUTP, or their derivatives, optionally, pseudoUTP, N1-methyl-UTP, m5CTP, m6ATP, m1ATP, Inosine triphosphate, hm5CTP, m1GTP, m7GTP, m6AmTP, or a mixture thereof.

21. The method according to any preceding claim, wherein the composition of the reaction at the start of the reaction comprises 8mM, or 9mM, or 10mM of each of rATP, rCTP, rUTP, or their derivatives, optionally, pseudoUTP, N1-methyl-UTP, m5CTP, m6ATP, m1ATP, Inosine triphosphate, hm5CTP, m1GTP, m7GTP, m6AmTP, or a mixture thereof.

22. The method according to any preceding claim, wherein the starting and/or average concentration of the competing nucleotide is selected as a fraction of the cap analogue concentration sufficient to achieve the desired capping efficiency: (i) 1:9, or 1:10, or 1:20; or (ii) 1:50, or 1:75, or 1:100.

23. The method according to any preceding claim, wherein the feed solution comprises the competing nucleotide(s) and, optionally, counterions of the nucleotide(s), in water which is RNAse-free.

24. The method according to any preceding claim, wherein the cap analogue starting concentration is sufficiently high to allow for a competing nucleotide

concentration sufficient to maintain a suitable reaction rate, wherein the cap analogue starting concentration is >5mM, or >8mM, or >10mM.

25. The method according to any preceding claim, wherein the method comprises reducing the volume of the transcription reaction during and after feeding to compensate for the additional volume of the feed solution added thereto.

26. The method according to any preceding claim, wherein the method comprises feeding only the competing nucleotide, while reducing the volume, optionally at a high cap analogue to competing nucleotide ratio.

27. The method according to any preceding claim, wherein the feeding of the competing nucleotide and/or any other reaction component is performed to allow fluids to be simultaneously, or subsequently, added and removed.

28. The method according to claim 27, wherein the addition and removal of fluid is matched as to not change the volume of the reaction, and thus the concentration of the non-fed reaction components, beyond predefined parameters.

29. The method according to either claim 27 or claim 28, wherein the addition and removal of fluid is matched to maintain a substantially identical volume of the reaction, preferably wherein the removal of the fluid is performed in such a way that only water is removed, and not any substantial amount of dissolved reaction components or products.

30. The method according to any one of claims 25-29, wherein excess fluid is removed from the reaction by means of evaporation, thereby keeping the volume of the reaction substantially constant.

31. The method according to claim 30, wherein evaporation of the excess fluid is achieved by the application of a controlled gas, preferably air, flow over the reaction mixture, wherein the airflow is in direct contact with the surface of the reaction volume, and not in contact with the feed solution flow.

32. The method according to claim 31, wherein the magnitude of the gas, preferably air, flow is controlled, and wherein the speed of the gas flow is at least 0.5m/s, at least 2m/s, at least 3.5m/s, at least 5m/s, at least 10m/s, or >10m/s.

33. The method according to any one of claims 31 or 32, wherein the air humidity is decreased prior to contacting the air flow with the surface of the reaction mixture, optionally wherein the air humidity is below 50%, or below 40%, or below 30%.

34. The method according to any one of claims 31-33, wherein the gas, preferably air, temperature is maintained at the same temperature as the reaction mixture to avoid cooling the fluid.

35. The method according to any one of claims 25-34, wherein the removal of excess fluid is achieved via osmosis by contacting a sufficiently large surface of the reaction volume with a semi-permeable or water-selective membrane, wherein the membrane has a pore size which is smaller than the protein, DNA and RNA components of the reaction, but larger than NTPs, buffering components, salts, reductive agents and/or spermidine.

36. The method according to claim 35, wherein the method comprises feeding pressurised air into the reaction chamber in order to provide sufficient pressure to drive a predictable and/or sufficiently large volume of feed solution through the semi-permeable membrane.

37. The method according to any one of claims 25-36, wherein the method of removal of excess fluid is adjusted and/or fine-tuned via a PID loop responding to measurements of the fluid volume and/or osmolality of the IVT reaction volume.

38. The method according to any preceding claim, wherein the method comprises measuring the reaction fluid level, preferably by optically measuring the fluid height, wherein such optical measurements are performed in: (i) a substantially vertical direction, thereby detecting the reflection on the fluid surface; or (ii) a substantially horizontal direction, either detecting the breaking of the optical path by the surface of the fluid, or the changed absorption by the fluid.

39. The method according to claim 38, wherein the reaction fluid level is measured by an electronic conductive probe, or by continuously weighing the entire

system and subtracting the empty weight of the reaction container, as well as the feeding solution not yet incorporated in the reaction volume.

40. The method according to any preceding claim, wherein the method comprises calculating excess fluid from the measurement of the osmolality of the solution and comparing this to the desired osmolality of the solution at that stage of the IVT reaction.

41. The method according to any preceding claim, wherein the method comprises determining the feed rate of the feed solution based on empirical evidence or a theoretical calculation of the consumption rate of the competing nucleotide and cap-analogue by a known amount of RNA polymerase with known activity in equal or similar reaction compositions, additionally accounting for the composition of the RNA sequence, preferably mRNA, optionally wherein the feed rate is started at the start of the reaction and substantially matches the consumption rate to maintain the cap-analogue-to-competing nucleotide ratio.

42. The method according to any preceding claim, wherein the method comprises adjusting the feed rate of the feed solution based on real-time or intermittent measurements of the reaction substrates and/or the reaction products.

43. The method according to any preceding claim, wherein the method comprises continuously or intermittently mixing the reaction reservoir to resolve local differences in the concentration of the competing nucleotide, and thus local differences in the ratio between the cap analogue and the competing nucleotide.

44. The method according to any preceding claim, wherein the method comprises feeding the feed solution to the reaction in a distributed manner, thereby preventing the formation of temporal high local concentrations of the feed solution, and optionally sub-optimal cap-analogue-to-competing nucleotide ratios, wherein distributed feeding comprises releasing the feed solution at space apart locations within the reaction volume, either simultaneously or sequentially.

45. The method according to claim 44, wherein the method comprises feeding, to the reaction, either intermittently or continuously, the competing nucleotide that is competing with the nucleotide-cap analogue, and/or any other reaction components, through a semi-permeable membrane, wherein the semi-permeable

membrane comprises a plurality of axially, radially and/or longitudinally spaced apart microscopic openings and/or pores through which the feed solution is passed.

46. The method according to claim 44, wherein the method comprises feeding, to the reaction, either intermittently or continuously, the competing nucleotide that is competing with the nucleotide-cap analogue, and/or any other reaction components, through a microfluidic channel, wherein the microfluidic channel comprises a plurality of axially, radially and/or longitudinally spaced apart microfluidic openings through which the feed solution is passed.

47. The method according to claim 44, wherein the method comprises feeding, to the reaction, either intermittently or continuously, the competing nucleotide that is competing with the nucleotide-cap analogue, and/or any other reaction components, by means of one or more needles, pipette tips, robotic probes, serological pipettes, Pasteur pipettes, or (an)other suitable probe(s).

48. The method according to any preceding claim, wherein the method comprises feeding, to the reaction, the competing nucleotide, and/or other reaction components, over the surface of the reaction volume by spraying, nebulizing or any other means of creating substantially small droplets, optionally wherein droplets are distributed substantially evenly over the fluid comprising the reaction volume.

49. The method according to claim 47, wherein the spray or nebulized fluid is applied to part of the entire surface of the reaction volume, or wherein the spray or nebulized fluid is applied to a portion of the reaction volume surface that is separate from the reaction volume surface that is exposed to air flow for the purpose of evaporating excess fluid.

50. The method according to any preceding claim, wherein the method comprises feeding, to a reaction, the competing nucleotide that is competing with the nucleotide-cap analogue, and/or other reaction components, by dissolving, bursting, swelling, degrading, exchanging, or otherwise releasing from a carrier that is added at the start of the reaction/incubation, and/or during the reaction, and/or intermittently fed, and/or continuously fed to the reaction.

51. A capped ribonucleic acid (RNA) molecule obtained, or obtainable by, the method according to any one of claims 1-50.

52. A ribonucleic acid (RNA) production apparatus for preparing capped RNA, the apparatus comprising a reaction chamber in which a transcription reaction occurs in the presence of a cap analogue, and a feed means for feeding a feed solution comprising one or more reagent required for producing capped RNA, wherein the apparatus is configured, in use, to feed the reagents, via the feed means, to the reaction chamber in a fed-batch mode, and carry out transcription in the reaction chamber in a substantially constant volume, to thereby produce a substantially capped RNA molecule.

53. The apparatus according to claim 51, wherein the apparatus is used to perform the method according to any one of claims 1-50.

54. The apparatus according to either claim 52 or 53, wherein the reaction chamber comprises a first section into which the feed solution is fed, and a second section in which the transcription reaction occurs, wherein the first and second section are separated by a semi-permeable membrane, preferably wherein the semi-permeable membrane is a dialysis membrane.

55. The apparatus according to claim 54, wherein the semi-permeable membrane comprises a pore size which is smaller than the protein, DNA and RNA components of the reaction, but larger than, and therefore allowing the passage therethrough of NTPs, buffering components, salts, reductive agents and/or spermidine.

56. The apparatus according to either claim 54 or 55, wherein the feed means is configured to feed the feed solution to the first section of the reaction chamber, thereby resulting in contact with one side of the semi-permeable membrane, wherein positive pressure created by feeding fluid drives the feed solution and its contained components through the semi-permeable membrane.

57. The apparatus according to any one of claims 54-56, wherein the apparatus comprises a gas, preferably air, inlet configured to feed pressurized gas into the first section of the reaction chamber, providing additional pressure on the feed solution.

58. The apparatus according to any one of claims 52-57, wherein the apparatus comprises a gas, preferably air, outlet configured to release excess pressure in the reaction chamber.

59. The apparatus according to any one of claims 52-58, wherein the reaction chamber comprises an evaporation zone which is disposed above the section in which the transcription reaction occurs, and in which fluid evaporates to compensate for the additional volume provided by the feed, to thereby keep the reaction volume substantially constant.

60. The apparatus according to claim 59, wherein the apparatus comprises a gas, preferably air, inlet configured to feed gas, preferably air, into the evaporation zone to facilitate evaporation of the fluid, optionally wherein the gas inlet is in fluid communication with the gas outlet.

61. The apparatus according to any one of claims 52-60, wherein the feed means comprises a probe which comprises a distal outlet through which the feed solution is passed, optionally wherein the probe is a needle, a pipette tip, a Pasteur or serological pipette, or any other suitable probe for the addition of feed solution.

62. The apparatus according to claim 61, wherein the outlet is arranged to extend into the reaction liquid in the reaction chamber.

63. The apparatus according to claim 61 or 62, wherein the probe comprises a plurality of axially, radially and/or longitudinally spaced apart outlets through which the feed solution is passed, preferably into the liquid.

64. The apparatus according to any one of claims 61-63, wherein the probe comprises a spraying device or nebulizer which is configured to produce droplets of the feed solution in the reaction chamber, optionally wherein the spraying device or nebulizer is mounted above the surface of the reaction liquid.

Figure 1A

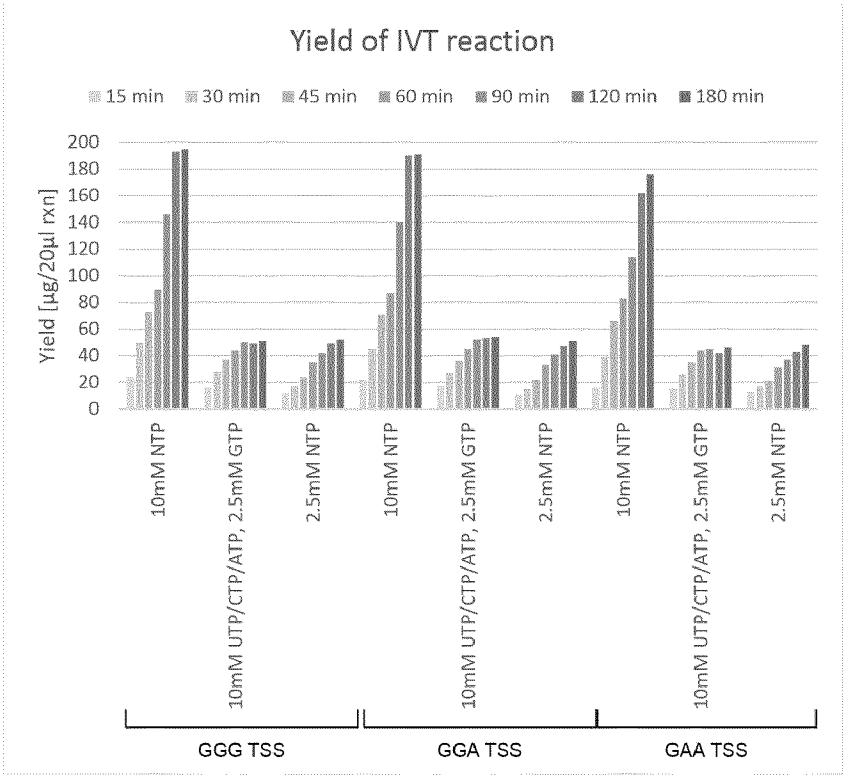

Figure 1B

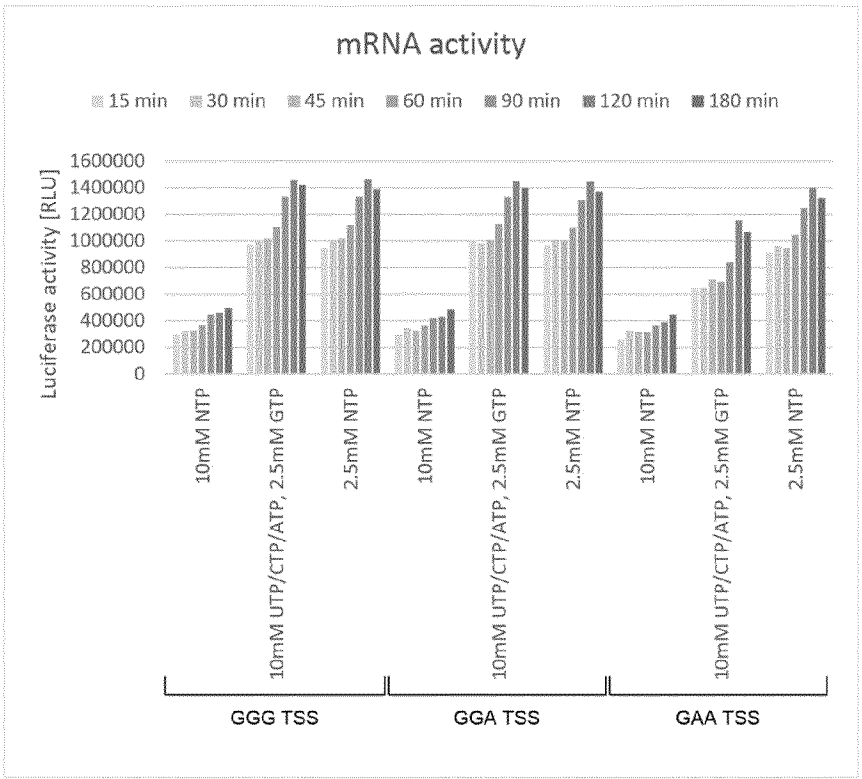

Figure 2

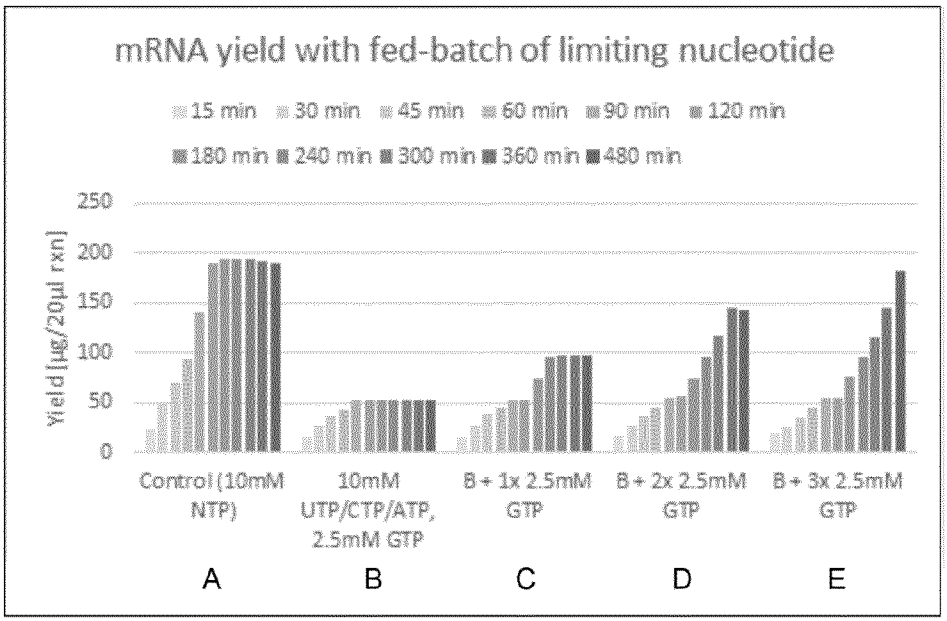
mRNA yield with fed-batch of limiting nucleotide
15 min
30 min
45 min
60 min
90 min
120 min
180 min
240 min
300 min
360 min
480 min
Yield [µg/20µl rxn]
250
200
150
100
50
0
Control (10mM NTP)
10mM UTP/CTP/ATP, 2.5mM GTP
B + 1x 2.5mM GTP
B + 2x 2.5mM GTP
B + 3x 2.5mM GTP
A
B
C
D
E

Figure 3A

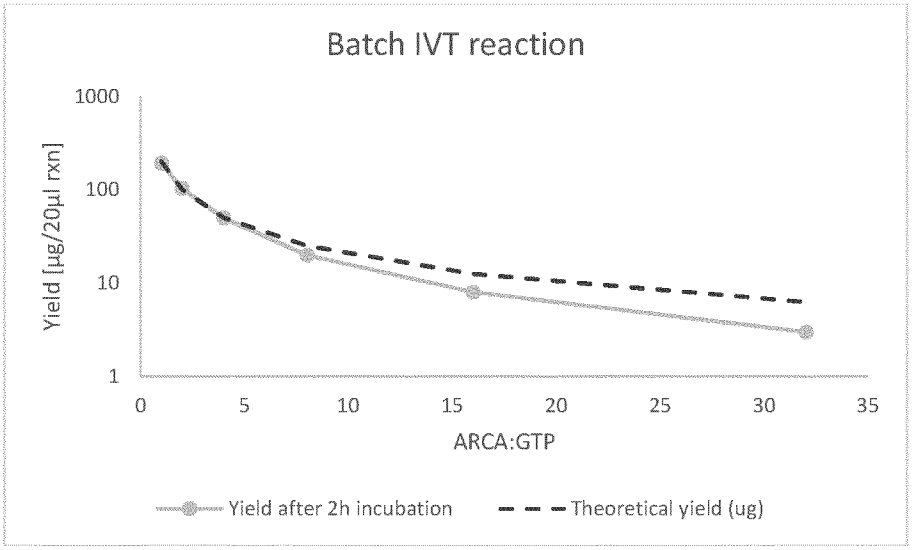

Figure 3B

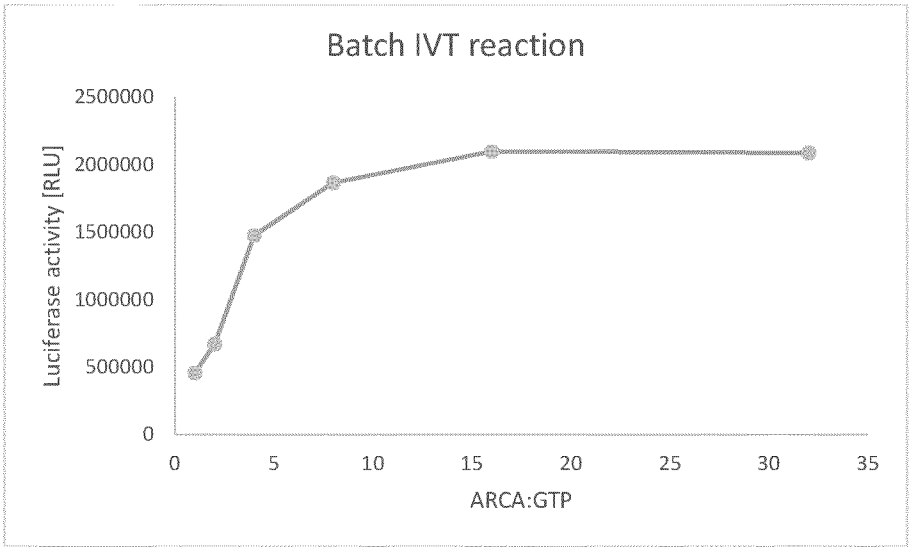

Figure 4A

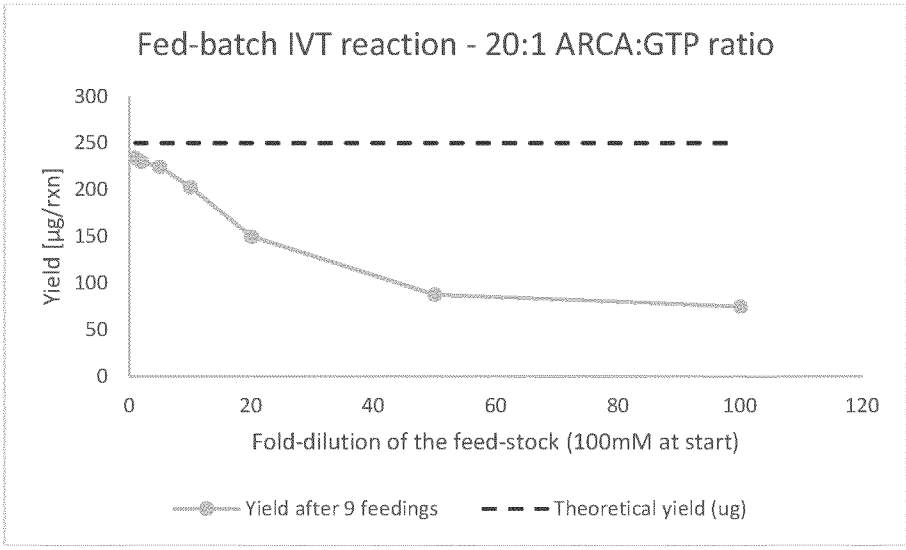

Figure 4B

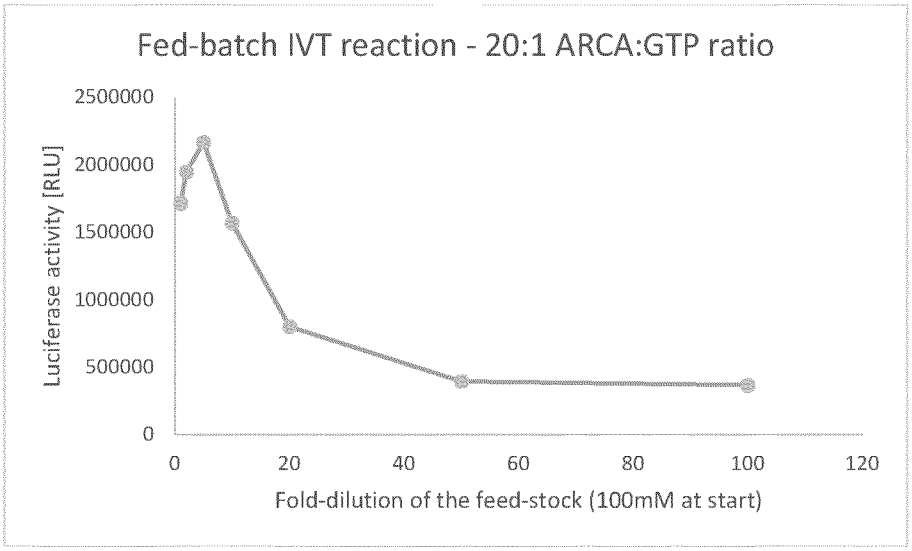

Figure 5A

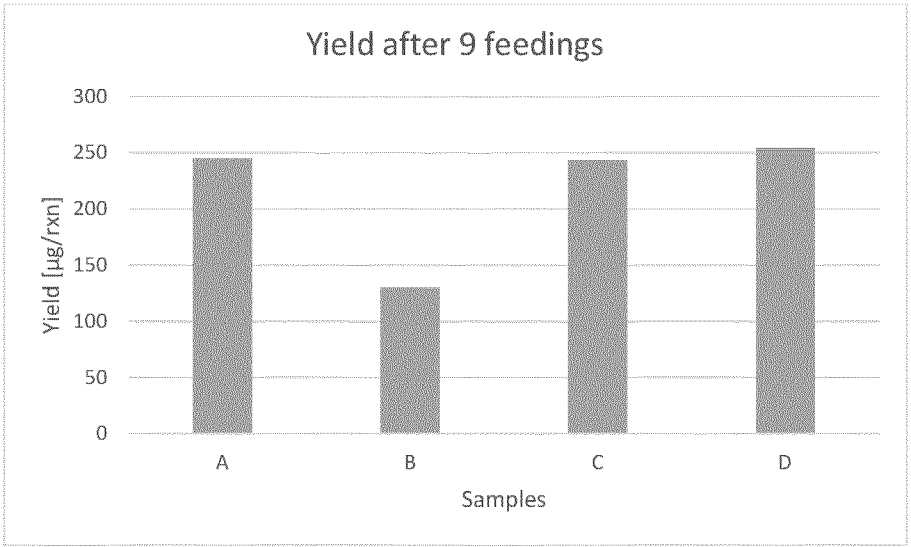

Figure 5B

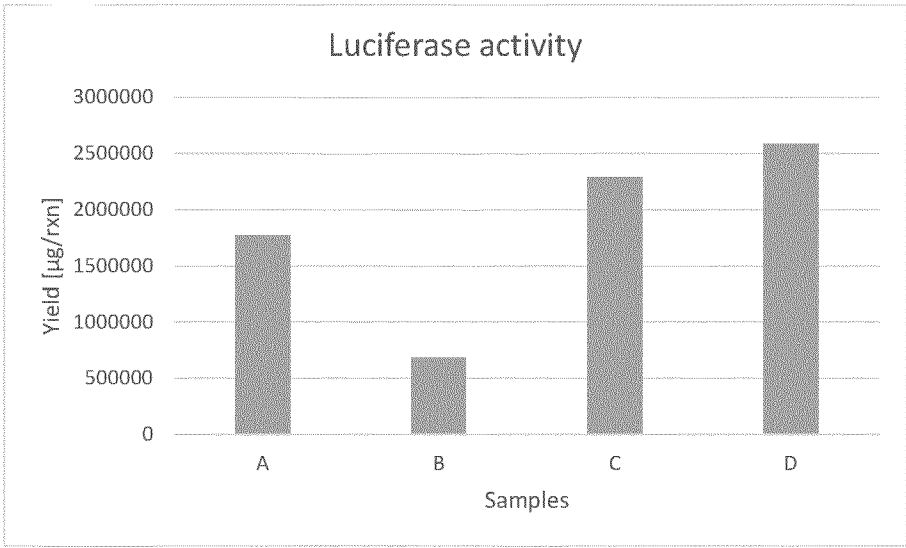

Figure 6

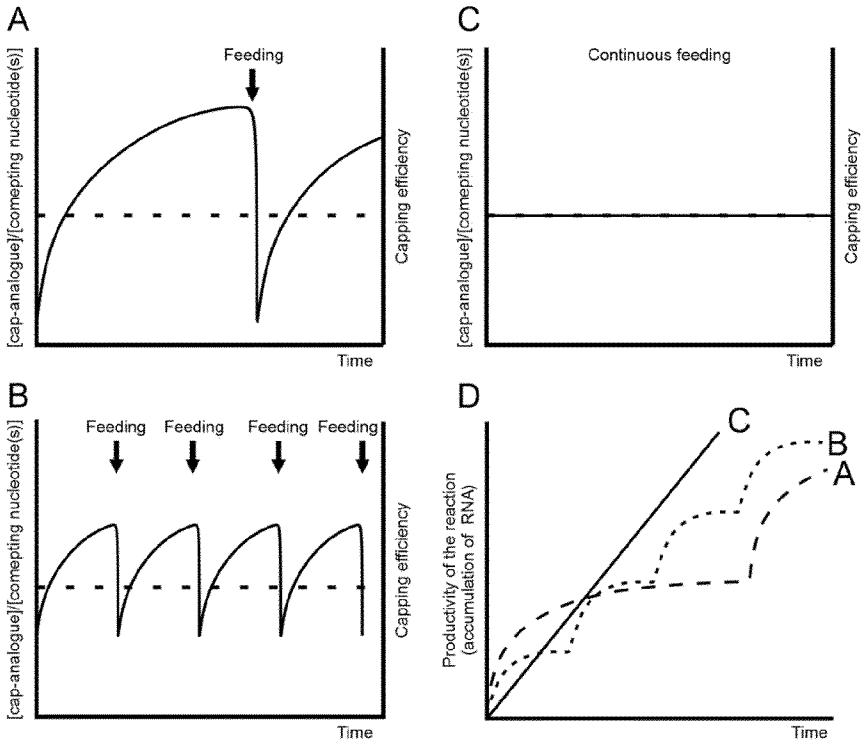
A
Feeding
[cap-analogue]/[competing nucleotide(s)]
Capping efficiency
Time
C
Continuous feeding
[cap-analogue]/[competing nucleotide(s)]
Capping efficiency
Time
B
Feeding
Feeding
Feeding
Feeding
[cap-analogue]/[competing nucleotide(s)]
Capping efficiency
Time
D
C
B
A
Productivity of the reaction (accumulation of RNA)
Time

Figure 7

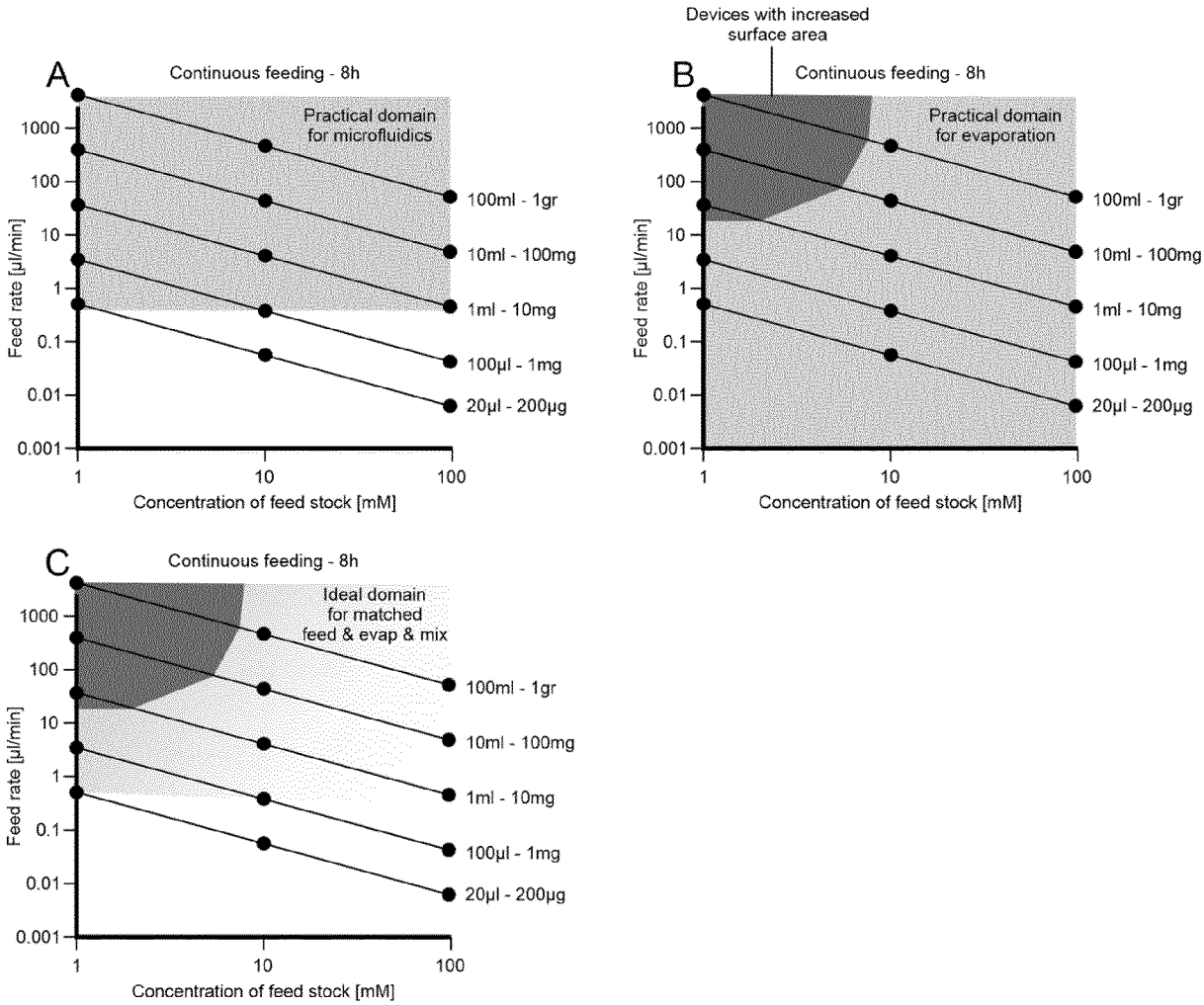
A
Continuous feeding - 8h
Practical domain for microfluidics
Feed rate [µl/min]
1000
100
10
1
0.1
0.01
0.001
1
10
100
Concentration of feed stock [mM]
100ml - 1gr
10ml - 100mg
1ml - 10mg
100µl - 1mg
20µl - 200µg
B
Devices with increased surface area
Continuous feeding - 8h
Practical domain for evaporation
C
Continuous feeding - 8h
Ideal domain for matched feed & evap & mix

Figure 8

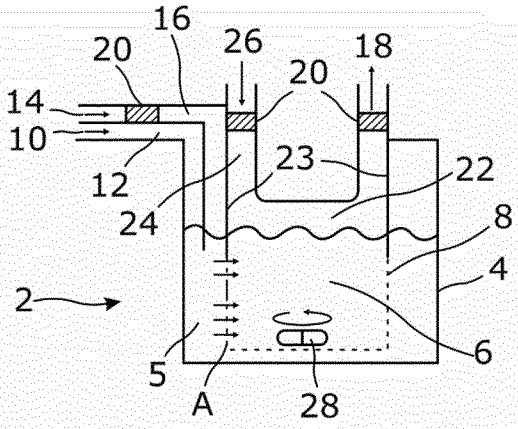

Figure 8A

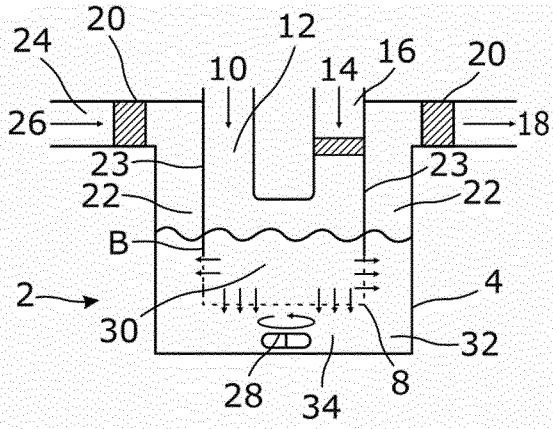

Figure 8B

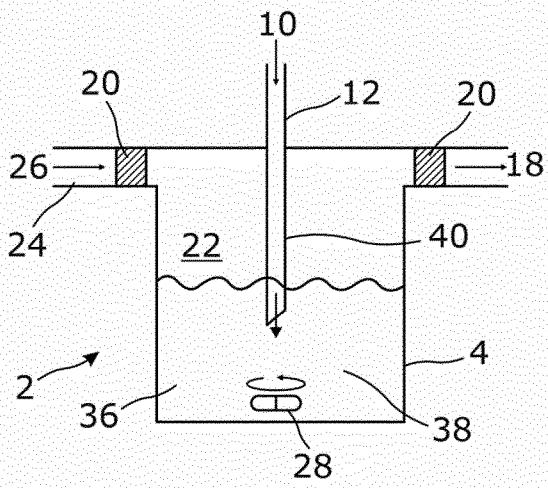

Figure 8C

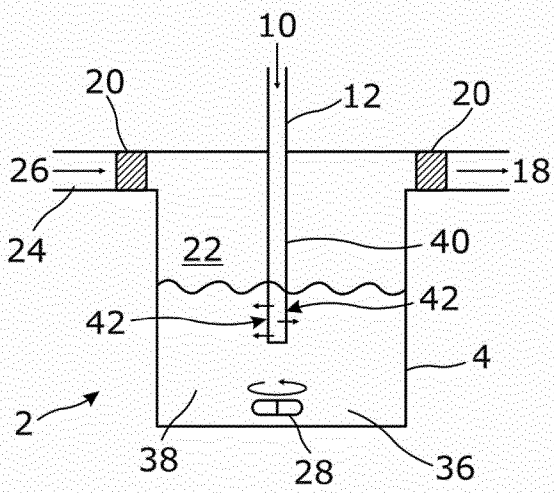

Figure 8D

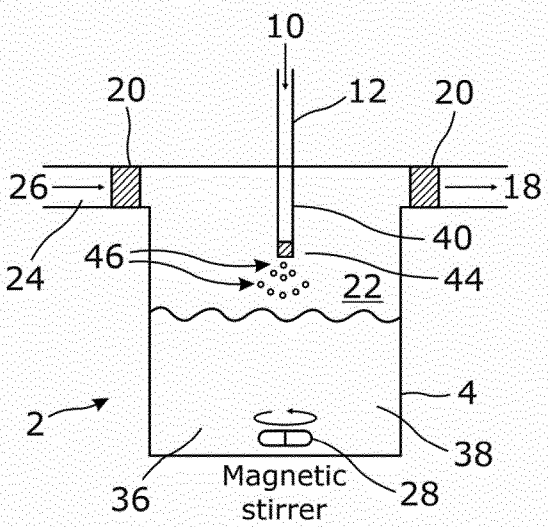

Figure 8E

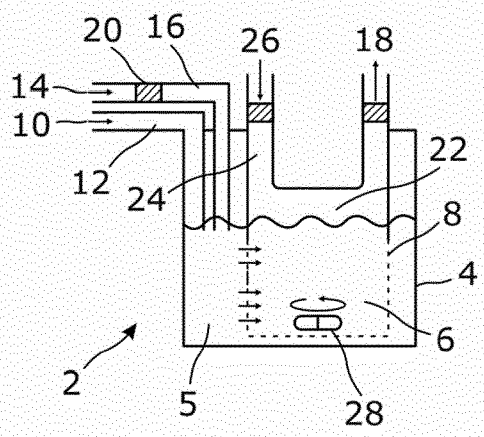

Figure 8F