White paper

Identification and Quantification of Components of In-Vitro Transcription (IVT) Reaction in a Mixture Using Process Raman

Authors

Chiara Bruckmann[1], Paolo La Rocca[1], Maria Colombo[2], Nimesh Khadka[3]

[1] Pharma Service Group, Thermo Fisher Scientific, Monza, Italy

[2] Pharma Service Group, Thermo Fisher Scientific, Gosselies, Belgium

[3] Analytical Instrument Group, Thermo Fisher Scientific, Tewksbury, Massachusetts, USA

Significance

Therapeutics and vaccines based on messenger ribonucleic acid (mRNA) are creating new opportunities in drug manufacturing. These compounds are produced using *in vitro transcription* (IVT), where a template DNA is transcribed in a mRNA molecule using an enzymatic reaction. Reagents needed in IVT reaction include the ribonucleotide triphosphates (rNTPs, including ATP, GTP, CTP, and UTP) as building blocks that are assembled to form the mRNA. These might also consist of modified ribonucleotide triphosphates (e.g., N1-Methyl pseudoUTP, 2-Thio-UTP, 5-Methyl-CTP) and capping analogues (e.g., Anti-Reverse Cap Analog, or ARCA). The efficiency of the IVT reaction depends on the concentration of these reactants. Optimizing the concentration of these reagents in the reaction is essential to achieving a high yield of mRNA and avoiding any undesirable byproducts. Thus, real-time monitoring of the IVT reaction for the mRNA production would provide actionable information that allows fine-tuning of these parameters. In this feasibility study, we evaluated process Raman as an in-line tool for real-time identification and quantification of IVT components in a mixture (Figure 1).

Introduction

The IVT reaction is a cell-free enzymatic process used for synthesizing mRNA.[1] It is the process used by research laboratories and biopharmaceutical industries to manufacture therapeutic mRNA. Notable examples include the SAR-CoV-2 mRNA vaccine and the recently FDA-approved respiratory syncytial virus (RSV) mRNA vaccine,[2] along with some promising clinical trial mRNA therapeutics for cancer and cell and gene therapy.[3–5]

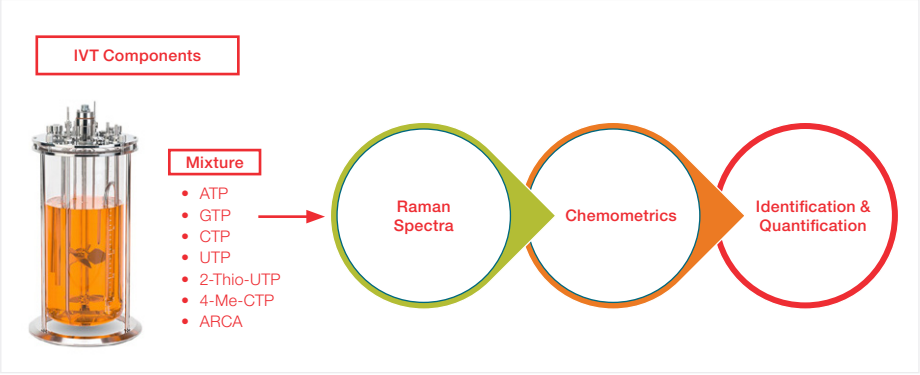


Figure 1.

thermo scientific

The IVT reaction comprises several key components: deoxyribonucleic acid (DNA) template, RNA polymerase, ribonucleoside triphosphates (rNTPs), eventually modified ribonucleotides and capping analogs, pyrophosphatase, RNase inhibitor, spermidine, and buffer components. The template DNA's information is decoded through the enzymatic action of RNA polymerase, and complementary rNTPs are assembled via phosphodiester bonds to produce mRNA. Modified nucleotides are often used to synthesize the mRNA to enhance translation and avoid unwanted innate immune responses for *in vivo* applications. The yield and quality of mRNA produced by the IVT reaction depend upon critical process parameters (CPPs), such as the concentration of rNTPs, modified rNTPs, capping agents, and the enzymatic and buffer components.[6]

Process Raman spectroscopy is gaining popularity in the biopharmaceutical industry as a process analytical technology (PAT) for monitoring and controlling various bioprocesses.[7] These processes include upstream cell culturing in bioreactors, downstream purifications, enzymatic transformations, and fill-and-finish operations. The success of process Raman spectroscopy in bioprocess applications stems from its intrinsic properties of high analyte specificity, minimal water interference, few overtones or combination bands (thus allowing straightforward spectral interpretation), fast and reliable in-line operation, and non-destructive probing of biomolecules in their native state.[8,9]

In this proof-of-concept (PoC) study, we aimed to leverage the inherent benefits of Raman spectroscopy and demonstrate the feasibility of using the process Raman to identify and quantify the critical components of the IVT reaction in a mixture. The results presented here set the foundation for using process Raman as the in-line PAT solution for real-time monitoring, optimizing, and controlling of IVT reactions.

Experimental details

Identification of the individual IVT components: nucleotide triphosphates and capping analogues

The IVT reaction comprises a mixture of components summarized in Table 1. We obtained Raman spectra of the following compounds: rNTPs (ATP, GTP, CTP, and UTP, Thermo Fisher Scientific, Vilnius, Lithuania); two modified nucleotides (2-Thio-UTP and 5-Methyl-CTP, Jena Biosciences, Jena, Germany), and anti-reverse cap analogues (ARCA, Jena Bioscences, Jena, Germany). Data were collected using individual component solutions at fixed concentrations to identify each analyte based on unique Raman signatures. All the samples were prepared at a concentration of 3 mM in tris-acetate buffer (25 mM, pH 7.9) in a glass vial with a total volume of 1 mL. The Raman spectra were collected using the Thermo Scientific™ MarqMetrix™ All-In-One Process Raman Analyzer integrated with a micro-immersible probe. The micro immersible probe allows the acquisition of the Raman spectra from a droplet or micro volumes of sample. All Raman spectra were acquired using the acquisition settings of power 450 mW, integration time 3000 ms, and total averages of 3. Each spectrum was preprocessed as follows: i) Automatic Whittaker Smoothing (AWS) Filter (p=0.001, λ=10,000), ii) standard normal variate (SNV), and iii) mean centering. Principal component analysis (PCA) was performed on the preprocessed data. All chemometric works were performed using SOLO 9.3.1 (2024). Eigenvector Research. Inc. Manson, WA USA 98831.

As shown in Table 1, ATP, GTP, UTP, CTP, 2-Thio-UTP, 5-Methyl-CTP, and ARCA are consumed during the IVT reaction with concomitance formation of mRNA. Thus, these components are CPPs that should be monitored to optimize or control the IVT reaction.[10]

Reagent	IVT starting concentration	IVT final concentration	It is essential to be monitored
Tris Ac pH 7.9	25 mM	25 mM	No
MgAc	20-40 mM	20-40 mM	No
DTT	1-10 mM	1-10 mM	No
Spermidine	2 mM	2 mM	No
Adenosine Triphosphate (ATP)	5-10 mM	Reduction	**Yes**
Guanine Triphosphate (GTP)	5-10 mM	Reduction	**Yes**
Uracil Triphosphate (UTP)	5-10 mM	Reduction	**Yes**
Cytosine Triphosphate (CTP)	5-10 mM	Reduction	**Yes**
Modified ribonucleotides (rNTPs) (e.g. 2-Thio-UTP. 5-Methyl-CTP)	5-10 mM	Reduction	**Yes**
Capping Reagent (e.g. ARCA)	4-10 mM	Reduction	**Yes**
Pyrophosphatase	0.001 U/uL	0.001 U/uL	No
RNase inhibitor	0.5-1 U/uL	0.5-1 U/uL	No
Polymerase T7	150-250 U/ug Template	150-250 U/ug Template	No
DNA Template	25 ng/uL	25 ng/uL	No
mRNA	-	1-5 ug/uL	**Yes**
IVT by-product: dsRNA	-	1-10 ng/uL	*Desirable*

Table 1. IVT components and key parameters of interest for real-time monitoring.

Quantification of the components in a mixture using the DoE model

We evaluated the effectiveness of process Raman in differentiating and quantifying rNTPs, two modified rNPTs and the capping analogue ARCA in a mixture. The training data for the mixture were created using the design space determined by the uniform design algorithm (Table 2). The uniform design is one of the statistical methods used in the design of experiments. The input variables are uniformly distributed along the design space in this method. There are several benefits of the uniform design: it helps in producing sample sets that represent the design space well; it doesn't make any prior assumptions on the model type (linear or non-linear); and it allows for the largest possible levels for each factor among all experimental designs.[11,12] The concentration for each component ranged from 0 to 10 mM, which is typical for IVT reactions. The Raman spectra were acquired using the acquisition setting as detailed above.

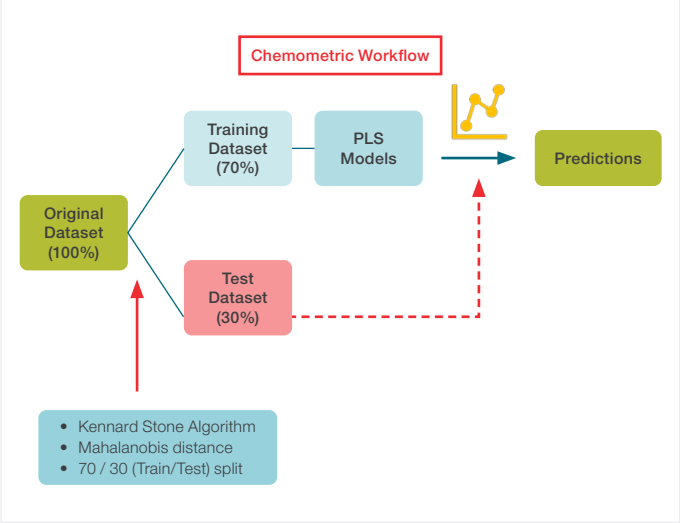
Sample	ATP (mM)	GTP (mM)	CTP (mM)	UTP (mM)	5-Me-CTP (mM)	2-Thio-UTP (mM)	ARCA (mM)
1	8.71	7.24	9.70	6.14	2.80	3.85	8.33
2	5.59	7.02	5.87	1.26	3.85	8.46	1.60
3	7.99	3.28	5.17	7.56	4.16	9.19	9.53
4	9.49	4.97	7.25	3.64	5.31	1.61	7.03
5	4.28	4.85	2.74	9.76	6.13	0.80	6.06
6	7.57	3.88	0.28	6.76	1.59	4.08	1.92
7	2.44	7.56	9.46	5.81	6.95	9.46	3.06
8	4.69	1.95	9.19	5.21	9.27	1.04	1.11
9	2.19	1.06	1.56	1.43	6.80	2.91	6.71
10	3.32	3.03	6.42	9.20	7.66	6.07	1.36
11	6.23	5.81	1.04	7.27	9.40	8.57	8.56
12	4.50	5.23	4.12	0.45	0.21	5.48	5.01
13	4.99	2.52	8.00	3.03	1.32	9.69	8.06
14	0.88	1.64	3.59	1.85	3.09	8.13	7.54
15	1.60	3.54	4.61	8.22	3.37	1.99	4.50
16	3.90	2.76	0.54	5.62	4.47	0.27	7.87
17	3.08	0.22	2.50	4.95	3.65	5.00	2.52
18	9.71	2.26	2.20	2.71	8.68	3.61	4.65
19	0.60	6.75	6.62	3.29	2.14	1.22	9.20
20	5.79	0.49	7.52	8.40	8.11	5.29	8.93
21	9.24	9.41	3.31	8.67	4.67	4.50	0.32
22	1.88	4.58	8.86	7.87	2.47	7.59	2.22
23	8.98	5.47	1.38	3.85	1.93	7.82	3.88
24	6.39	9.74	7.83	7.14	7.30	2.66	5.13
25	3.59	7.91	1.95	9.01	0.53	6.36	7.20
26	1.30	8.80	5.53	6.50	1.06	2.34	3.35
27	5.21	8.18	3.81	4.71	8.41	1.92	9.74
28	7.10	1.29	6.08	9.54	5.61	7.30	4.17
29	8.09	6.13	4.96	0.28	6.51	0.48	2.64
30	6.76	9.22	8.74	0.98	4.94	5.89	6.49
31	8.40	8.90	4.45	4.12	7.88	8.96	5.61
32	7.27	0.84	8.30	2.38	0.75	3.12	0.81
33	2.78	8.45	0.75	2.18	5.85	6.66	0.60
34	0.28	4.16	7.07	0.72	8.95	4.69	5.83
35	1.05	6.41	3.01	4.45	9.74	7.02	3.60

Table 2. Design of experiment (DoE) for developing a chemometric model for quantifying each component in a mixture.

The acquired Raman spectra and associated reference values were split into a training set (70%) and a testing set (30%) using the Kennard Stone algorithm (Schematics 1). The PLS quantitative models were developed for each component using the 480 to 3240 cm^{-1} spectral region. Each spectrum was preprocessed as follows: i) normalization to water band using infinity norm calculated in the region 3080 to 3230 cm^{-1}; ii) SavGol filter (1st derivative, order 2, window size 13); and iii) mean centering. To prevent the overfitting of the calibration model, internal cross-validation (CV) was performed using the Venetian blind cross-validation strategy (number of data split = 7 and number of samples per blind = 1). Initially, twenty different PLS models for each analyte were developed using latent variables (LVs) 1 through 20. The root mean square error of calibration (RMSEC) and cross-validation (RMSECV) were calculated for each model. Finally, the optimum number of LVs was selected to develop each PLS model. The selection criteria were defined such that the RMSECV did not improve significantly by adding more LVs, and the ratio of RMSEC to RMSECV was close to 1. After selecting the appropriate PLS model for each component, the initially segregated test set was used to validate the model performance, as shown in Schematics 1. The root means square error of prediction (RMSEP), R^2 prediction, and prediction bias were calculated to evaluate model performance.

Feasibility of identification of mRNA within the mixture

Next, we acquired the Raman spectra of mRNA, DTT, spermidine, DNA template, pyrophosphatase, RNA polymerase, and magnesium acetate using the same acquisition parameters described above. These spectra were overlayed with the mean of the spectra of the DoE samples after applying preprocessing as i) normalization to water band using infinity norm calculated in the region 3080 to 3230 cm^{-1}, and ii) SavGol filter (2nd derivative, order = 2, window size = 13). Spectral analysis was performed to identify the unique Raman signature of mRNA.



Schematics 1. Showing workflow for chemometric analysis.

Results

The Raman spectra of ATP, CTP, GTP, UTP, ARCA, 2-Me-CTP, 5-Thio-UTP, DNA template, pyrophosphatase, RNA polymerase, DTT, and spermidine were acquired in the tris-acetate buffer. In the first analysis phase, the Raman spectra of ATP, CTP, GTP, UTP, 2-Me-CTP, 5-Thio-UTP, and ARCA were considered. Figure 3A shows distinct spectral features of ATP, CTP, GTP, UTP, 2-Me-CTP, 5-Thio-UTP, and ARCA. The differences in the spectral features were visually distinct, but they must be mathematically expressed for quantitative or quantitative analysis. One of the approaches commonly applied in chemometrics is principal component analysis (PCA). PCA is a widely used method in data analysis that effectively reduces the dimensionality of a dataset while retaining the essential discriminatory information. PCA captures the underlying variations and patterns present in the spectral features by transforming the original spectral data into a set of orthogonal variables known as principal components. The training data are projected into the principal components that give PCA scores, which serve as input responses for classification (also regression in principal component regression (PCR)) based on their spectral characteristics. The decomposition of the Raman spectra of Figure 3A using the PCA is shown in Figure 3B. Each class had distinctive scores in the PCA space, which can be used for molecular identification or classification.

In the second phase of experiments, the Raman spectra of template DNA (at 25ng/µL), pyrophosphatase (0.001 U/µL), and T7 RNA polymerase (200 U/ug template) were acquired in the tris-acetate buffer at the concentration relevant to the IVT process. However, the Raman signals of these analytes were below the instrument's detection limit and, hence, not considered for further analysis. On the other hand, DTT (10 mM) and spermidine (2 mM) have distinct Raman signatures. They were included in the PCA analysis, as shown in Figure 3A. The spectra were preprocessed to facilitate visual perception (removing baseline using AWS, SNV normalization, and mean centering). The mean centering operation calculates the mean spectra of all data and subtracts it from individual spectra to produce differential spectra. As shown in Figure 4A, the differential spectrum of each component was visually distinct, leading to the different PCA scores (Figure 4B). The results shown in Figures 3 and 4 highlight that the unique molecular signature of ATP, CTP, GTP, UTP, ARCA, 2-Me-CTP, and 5-Thio-UTP, DTT, and spermidine in the Raman spectra can be leveraged for identification and quantification of these components.

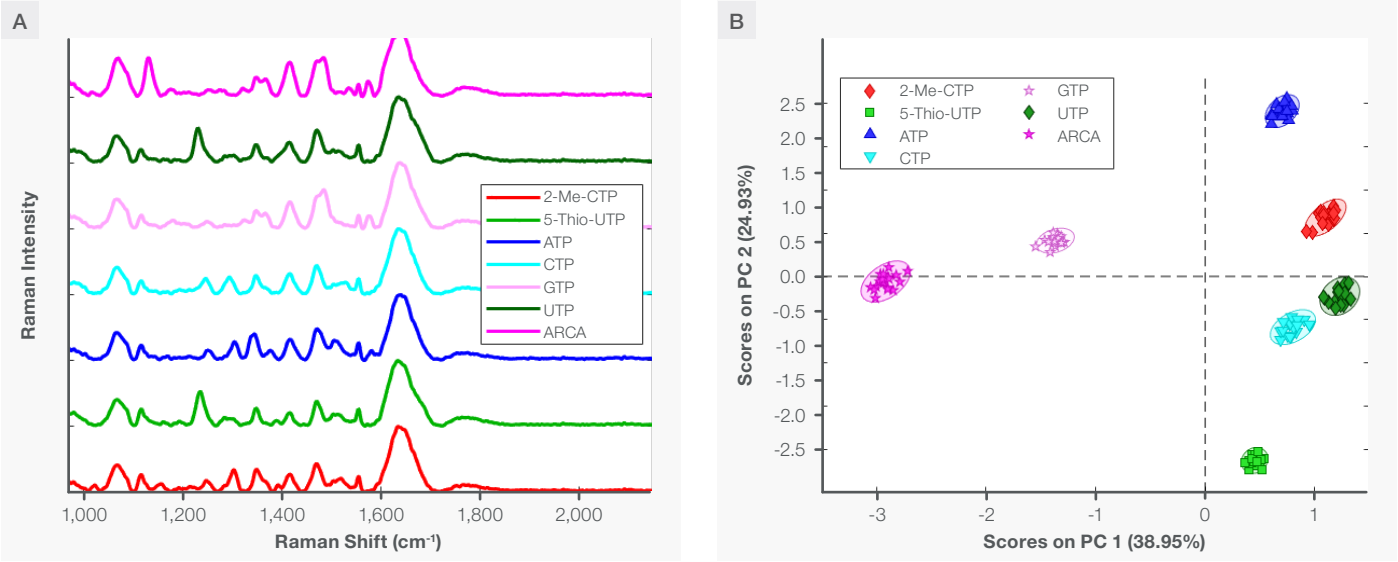


Figure 3. Plot **A** shows the distinctive Raman features of ATP, CTP, GTP, UTP, ARCA, 2-Me-CTP, and 5-Thio-UTP in the tris-acetate buffer. The baseline from each spectrum was removed and normalized using SNV. Plot **B** shows the clustering of molecules in the different PCA space. The ellipsoid on each class shows 95% confidence boundaries.

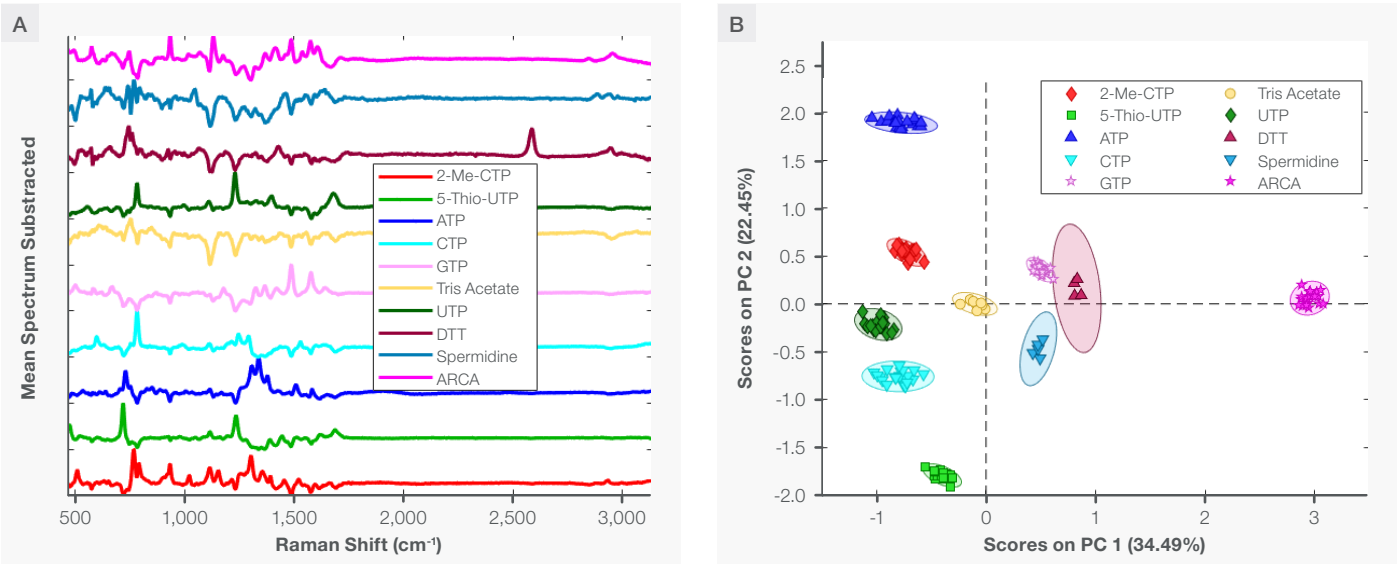


Figure 4. Plot **A** shows the spectra after subtraction of mean spectrum from analyte spectrum. The differential spectra provide visual inspection to spectral differences. Plot **B** shows the clustering of each class in the PCA space.

After establishing Raman to identify the critical components of the IVT reactions, we evaluated the potential of process Raman to quantify key components in a mixture. We focused on ATP, CTP, GTP, UTP, ARCA, 2-Me-CTP, and 5-Thio-UTP. They are the building blocks for the in vitro mRNA synthesis and thus are CPPs. For this study, the mixture of ATP, CTP, GTP, UTP, ARCA, 2-Me-CTP, and 5-Thio-UTP was prepared at the concentrations defined by the design space of the uniform design (UD) algorithm (Table 2). The Raman spectra for each DoE sample were acquired. Figure 5 shows the Raman spectra of the DoE samples after the baseline removal.

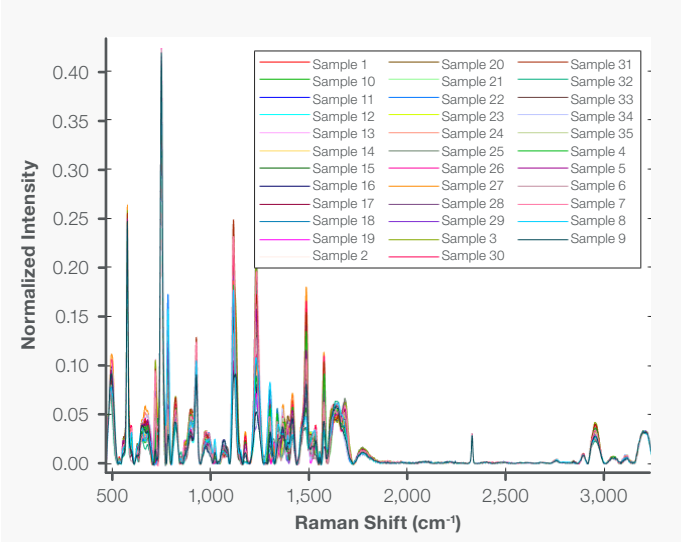


Figure 5. Spectral overlay of Raman data collected on the DoE samples as listed on Table 2.

The spectral information was correlated to the reference concentrations of each component in the mixture using a partial least squares (PLS) regression algorithm. A separate PLS model was developed for each analyte, as shown in Figure 6. As explained above, the number of latent variables for each PLS model was selected based on RMSECV. The developed PLS models were applied to the test samples, as shown in Schematic 1. The RMSEP was calculated to assess the performance of the model. In the correlation plots shown in Figure 6 (A to G), the grey dots are training data points, while the red diamonds are test data. The model statistics are summarized in Table 3.

The PLS models were developed using 4 or 5 LVs models, as shown in Table 3. The low values of RMSECV, RMSEP, cross-validation (CV) bias, and prediction bias, along with high R^2 for CV and prediction indicate that the models are linear, accurate, and reliable within the 0 to 10 mM concentration range. The validity of the PLS models was further substantiated by calculating the VIP score. The VIP score plots for the PLS models are shown in Figure 6. The threshold for the VIP score was set to 1, represented by the dotted red line. The model considers Raman shifts with VIP scores higher than one important. As illustrated in Figure 7, the VIP scores for PLS models differed for different analytes, indicating that the prediction was based on model specificity rather than nonspecific correlation among the analytes. The dominant Raman shifts in the VIP score plot for each PLS model agreed with the molecular vibration of the specific analyte. The details of the VIP score will be discussed in future work after reinforcing the models with the process data.

Analyte	No. of LVs	RMSEC (mM)	RMSECV (mM)	Bias CV (mM)	R^2 CV	RMSEP (mM)	Bias Prediction (mM)	R^2 Prediction
ATP	5	0.172	0.263	-0.034	0.991	0.222	-0.104	0.996
GTP	5	0.128	0.329	-0.021	0.987	0.245	0.078	0.994
CTP	5	0.141	0.249	0.002	0.992	0.216	0.052	0.996
5-Met-CTP	4	0.224	0.360	0.050	0.987	0.159	-0.057	0.998
UTP	5	0.259	0.389	0.047	0.979	0.439	0.045	0.983
2-Thio-UTP	4	0.176	0.256	-0.025	0.993	0.306	-0.088	0.986
ARCA	5	0.123	0.211	-0.013	0.995	0.350	0.010	0.982

Table 3. Model Statistics for selected analytes listed in Table 1.

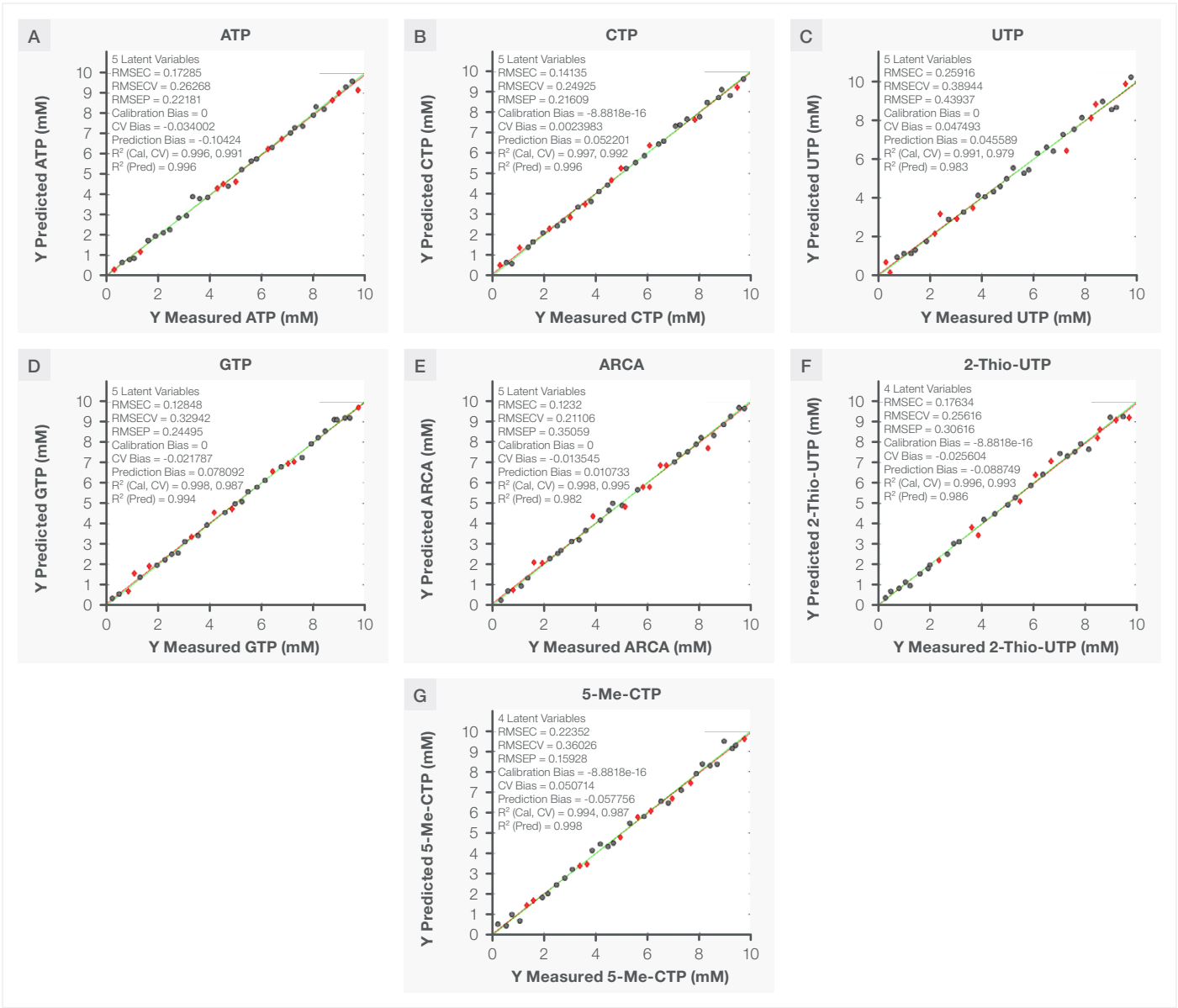


Figure 6. Showing the correlation plots of measured vs predicted for different analytes. The **grey** dots are training data while the **red** diamonds are test samples.

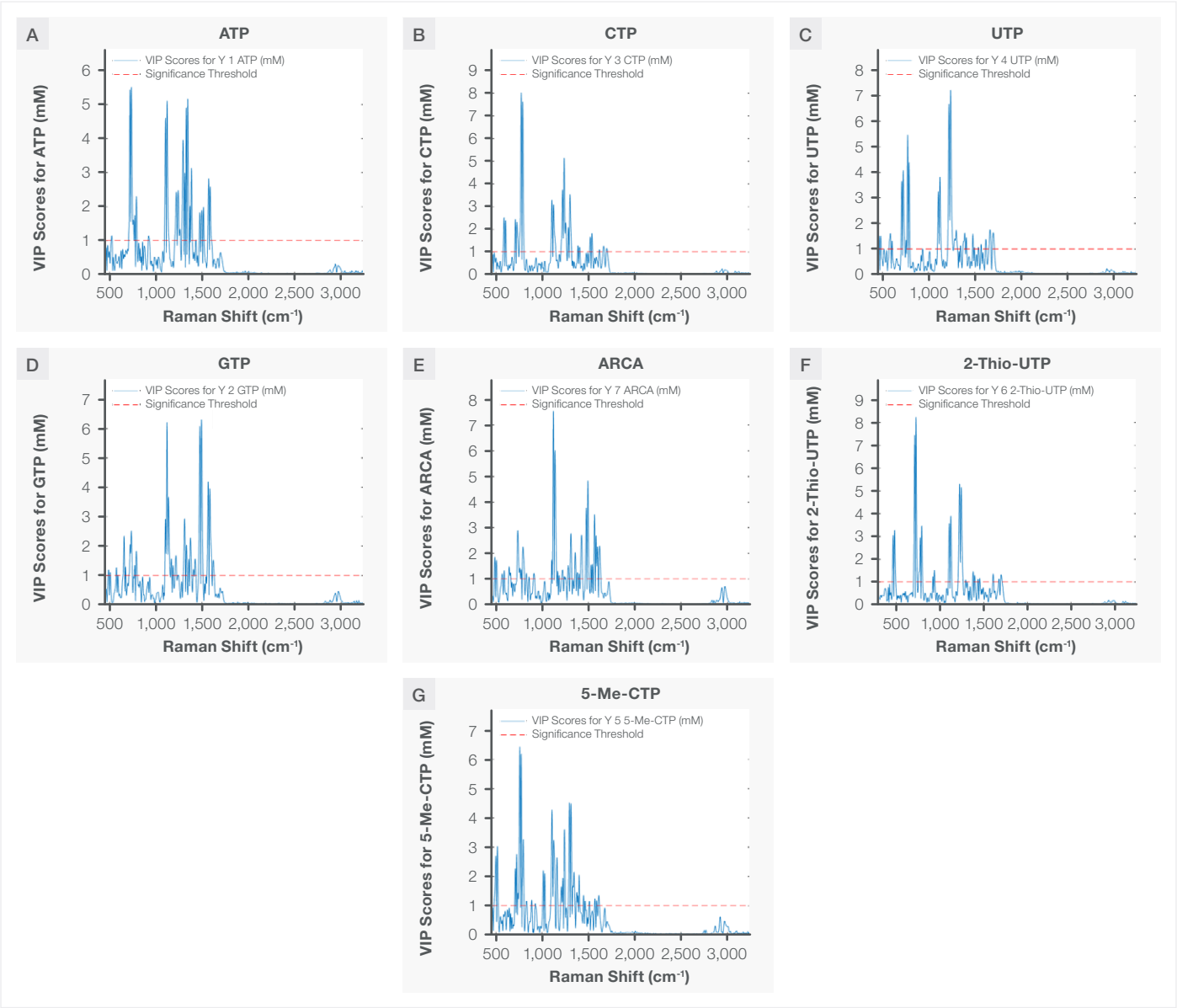


Figure 7. Showing the VIP score plot for the PLS models used in this study. The red dotted line (VIP score =1) represents the threshold. Any Raman shifts with VIP score more than 1 are considered important for the model. The VIP score plots (A to G) are different for different analytes indicating model specificity.

Feasibility of Monitoring Synthesized mRNA

The synthesized mRNA comprises rNTPs, ARCA, 2-Thio-UTP, and 5-Me-CTP. The identification and quantification of mRNA in the mixture depends on the distinctive Raman signature of mRNA. We collected the Raman spectra of mRNA at a concentration of 2 mg/mL in the tris acetate buffer and overlayed them with the spectra shown in Figures 4 and 5. The selected concentration represents a typical yield of mRNA from the IVT reaction. As shown in Figure 8, the O-P-O vibration mode of mRNA at ~ 814 cm^{-1} was unique compared to the other IVT components. The distinct spectrum of mRNA compared to the spectra of single nucleotides suggests the feasibility of Raman application for monitoring mRNA synthesis in the IVT reaction.

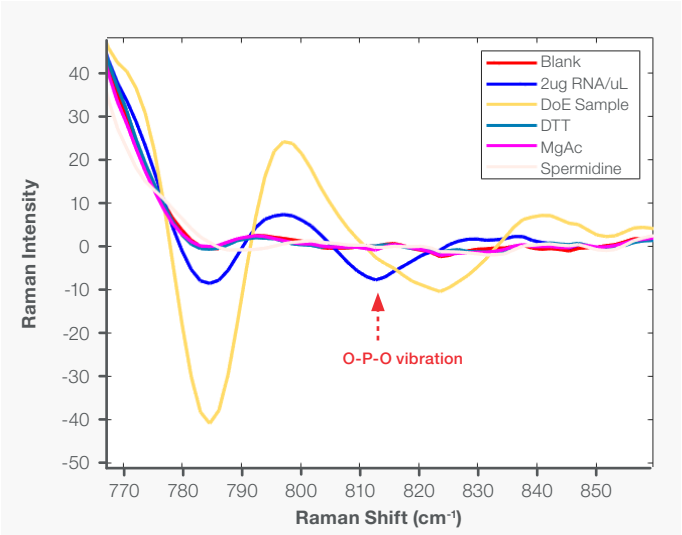


Figure 8. Shown is the spectral overlay of blank (red; tris acetate buffer), RNA (blue), average of DoE sample of Table 2 (gold), DTT (teal), Mg-acetate (magenta), and spermidine (light peach). The Raman band at ~814 cm^{-1} of RNA is unique compared to other IVT components.

Conclusion

The IVT reaction utilizes multiple and expensive components. Under certain conditions, it could also lead to unwanted byproduct synthesis,[13] which is detrimental to the quality of the final mRNA product. Monitoring, optimizing, and controlling IVT reactions may lead to better yield and quality of the mRNA produced and save the costs of unused IVT components. Raman spectroscopy is an analytical technology with high molecular specificity that does not require sample preparation and is amenable to an aqueous environment such as the IVT reaction buffer. In-line Process Raman might represent a PAT solution for real-time monitoring of the IVT reaction. This proof-of-concept study shows that Raman spectroscopy can identify and quantify key components of the IVT reaction in a mixture.

In future work, we intend to collect the real-time Raman data of the IVT process and further validate the existing model. Our final aim is to implement Raman spectroscopy as an in-line monitoring method for IVT reactions.

References

1. Kwon. H.; Kim. M.; Seo. Y.; Moon. Y. S.; Lee. H. J.; Lee. K.; Lee. H. Emergence of Synthetic mRNA: *In Vitro* Synthesis of mRNA and Its Applications in Regenerative Medicine. *Biomaterials* **2018**. *156*. 172–193. https://doi.org/10.1016/j.biomaterials.2017.11.034.
2. Mullard. A. FDA Approves mRNA-Based RSV Vaccine. *Nat. Rev. Drug Discov.* **2024**. https://doi.org/10.1038/d41573-024-00095-3.
3. Beck. J. D.; Reidenbach. D.; Salomon. N.; Sahin. U.; Türeci. Ö.; Vormehr. M.; Kranz. L. M. mRNA Therapeutics in Cancer Immunotherapy. *Mol. Cancer* **2021**. *20* (1). 69. https://doi.org/10.1186/s12943-021-01348-0.
4. Liu. C.; Shi. Q.; Huang. X.; Koo. S.; Kong. N.; Tao. W. mRNA-Based Cancer Therapeutics. *Nat. Rev. Cancer* **2023**. *23* (8). 526–543. https://doi.org/10.1038/s41568-023-00586-2.
5. Yamamoto. A.; Kormann. M.; Rosenecker. J.; Rudolph. C. Current Prospects for mRNA Gene Delivery. *Eur. J. Pharm. Biopharm.* **2009**. *71* (3). 484–489. https://doi.org/10.1016/j.ejpb.2008.09.016.
6. Abu-Absi. N. R.; Kenty. B. M.; Cuellar. M. E.; Borys. M. C.; Sakhamuri. S.; Strachan. D. J.; Hausladen. M. C.; Li. Z. J. Real-Time Monitoring of Multiple Parameters in Mammalian Cell Culture Bioreactors Using an In-Line Raman Spectroscopy Probe. *Biotechnol. Bioeng.* **2011**. *108* (5). 1215–1221. https://doi.org/10.1002/bit.23023.
7. Esmonde-White. K. A.; Cuellar. M.; Lewis. I. R. The Role of Raman Spectroscopy in Biopharmaceuticals from Development to Manufacturing. *Anal. Bioanal. Chem.* **2022**. *414* (2). 969–991. https://doi.org/10.1007/s00216-021-03727-4.
8. Esmonde-White. K. A.; Cuellar. M.; Uerpmann. C.; Lenain. B.; Lewis. I. R. Raman Spectroscopy as a Process Analytical Technology for Pharmaceutical Manufacturing and Bioprocessing. *Anal. Bioanal. Chem.* **2017**. *409* (3). 637–649. https://doi.org/10.1007/s00216-016-9824-1.
9. Buckley. K.; Ryder. A. G. Applications of Raman Spectroscopy in Biopharmaceutical Manufacturing: A Short Review. *Appl. Spectrosc.* **2017**. *71* (6). 1085–1116. https://doi.org/10.1177/0003702817703270.
10. Lee. K. H.; Song. J.; Kim. S.; Han. S. R.; Lee. S.-W. Real-Time Monitoring Strategies for Optimization of in Vitro Transcription and Quality Control of RNA. *Front. Mol. Biosci.* **2023**. *10*. https://doi.org/10.3389/fmolb.2023.1229246.
11. Fang. K.-T.; Lin. D. K. J.; Winker. P.; Zhang. Y. Uniform Design: Theory and Application. *Technometrics* **2000**. *42* (3). 237–248. https://doi.org/10.1080/00401706.2000.10486045.
12. Zhang. L.; Liang. Y.-Z.; Jiang. J.-H.; Yu. R.-Q.; Fang. K.-T. Uniform Design Applied to Nonlinear Multivariate Calibration by ANN. *Anal. Chim. Acta* **1998**. *370* (1). 65–77. https://doi.org/10.1016/S0003-2670(98)00256-6.
13. Stover. N. M.; Ganko. K.; Braatz. R. D. Mechanistic Modeling of in Vitro Transcription Incorporating Effects of Magnesium Pyrophosphate Crystallization. *Biotechnol. Bioeng. n/a* (n/a). https://doi.org/10.1002/bit.28699.