

UNIVERSIDADE FEDERAL DE JUIZ DE FORA
INSTITUTO DE CIÊNCIAS EXATAS
DEPARTAMENTO DE QUÍMICA
PROGRAMA DE PÓS-GRADUAÇÃO EM QUÍMICA

**Desenvolvimento e otimização de estratégias analíticas visando o diagnóstico de covid-19
em amostras de urina**

Olívia Brito de Oliveira Moreira

Juiz de Fora

2024

Olivia Brito de Oliveira Moreira

Desenvolvimento e otimização de estratégias analíticas visando o diagnóstico de covid-19 em amostras de urina

Tese apresentada ao Programa de Pós-Graduação em Química, área de concentração: Química Analítica, da Universidade Federal de Juiz de Fora, como requisito à obtenção do título de Doutora em Química.

Orientador: Prof. Dr. Marcone Augusto Leal de Oliveira

Juiz de Fora

2024

Ficha catalográfica elaborada através do programa de geração automática da Biblioteca
Universitária da UFJF,

com os dados fornecidos pelo(a) autor(a)

Moreira, Olívia Brito de Oliveira.

Desenvolvimento e otimização de estratégias analíticas visando o diagnóstico de covid-19 em amostras de urina / Olívia Brito de Oliveira Moreira. -- 2024.

146 f.

Orientador: Marccone Augusto Leal de Oliveira

Tese (doutorado) - Universidade Federal de Juiz de Fora, Instituto de Ciências Exatas. Programa de Pós-Graduação em Química, 2024.

1. Metabolômica clínica. 2. COVID-19. 3. Eletroforese Capilar. 4. Espectrometria de massas. I. Oliveira, Marccone Augusto Leal de, orient. II. Título.

Olívia Brito de Oliveira Moreira

Desenvolvimento e otimização de estratégias analíticas visando o diagnóstico de COVID-19 em amostras de urina

Tese apresentada ao Programa de Pós-Graduação em Química da Universidade Federal de Juiz de Fora como requisito parcial à obtenção do título de Doutora em Química. Área de concentração: Química.

Aprovada em 20 de março de 2024.

BANCA EXAMINADORA

Prof. Dr. Marcone Augusto Leal de Oliveira - Orientador
Universidade Federal de Juiz de Fora

Profa. Dra. Gisele André Baptista Canuto
Universidade Federal da Bahia

Profa. Dra. Quezia Bezerra Cass
Universidade Federal de São Carlos

Dra. Karina Helena Moraes Cardozo
Grupo Fleury

Prof. Dr. Emanuel Carrilho
Universidade de São Paulo

Juiz de Fora, 26/03/2024.



Documento assinado eletronicamente por **GISELE ANDRE BAPTISTA CANUTO, Usuário Externo**, em 26/03/2024, às 09:04, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Quezia Bezerra Cass, Usuário Externo**, em 26/03/2024, às 15:38, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Emanuel Carrilho, Usuário Externo**, em 26/03/2024, às 19:50, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Marcone Augusto Leal de Oliveira, Professor(a)**, em 26/03/2024, às 20:13, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Karina Helena Moraes Cardozo, Usuário Externo**, em 27/03/2024, às 10:21, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



A autenticidade deste documento pode ser conferida no Portal do SEI-Ufjf (www2.ufjf.br/SEI) através do ícone Conferência de Documentos, informando o código verificador **1762714** e o código CRC **DFE774CD**.

AGRADECIMENTOS

Ao meu orientador Marcone Augusto Leal de Oliveira pela oportunidade, apoio e amizade durante a realização dos projetos de mestrado e doutorado.

A todos os colegas do Grupo de Química Analítica e Quimiometria (GQAQ) pela parceria e amizade durante todos os anos de Pós-Graduação.

Aos inúmeros colaboradores que participaram da realização desse projeto: pesquisadores, técnicos e professores do Instituto de Biologia da UFJF, da Faculdade de Farmácia da UFJF, do Instituto de Química da USP, do Instituto de Química da Unicamp, do Instituto de Química da UFABC e da Faculdade de Enfermagem da UFJF. A toda equipe do Lemos Laboratórios de Análises Clínicas e da Secretaria de Saúde da Prefeitura de Juiz de Fora, aos voluntários que tiveram a iniciativa de participar junto à nossa pesquisa a partir da doação de amostras e à equipe de estudantes do GQAQ que participaram do projeto COVID-19.

Aos órgãos de fomento: Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES); Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPQ) 424032/2018-0, 303867/2020-5, e bolsa estudantil: 142502/2020-0); Finep -CT-INFRA 01/2013-REF 0633/13; RQ-MG (CEX.RED- 00010-14); Instituto Nacional de Ciência e Tecnologia Bioanalítica (INCTBio) (FAPESP 2014/50867-3, CNPq 465389/2014-7, 401256/2020-0); Fundação de Amparo à Pesquisa do Estado de Minas Gerais” (FAPEMIG) (APQ-00585-21).

Aos secretários do programa de Pós-Graduação em Química e do Instituto de Ciências Exatas.

Aos professores, funcionários e técnicos do Departamento de Química do Instituto de Ciências Exatas da UFJF pela disponibilidade.

Aos membros da banca avaliadora de qualificação e defesa pelas contribuições.

À Universidade Federal de Juiz de Fora.

Por fim, agradeço a todos amigos, familiares e parceiros do Grupo de Química Analítica e Quimiometria e demais dos programas de Pós-Graduação e Graduação que me acompanharam durante a realização deste trabalho.

“[...] Knowledge belongs to humanity, and is the torch which illuminates the world.”

(Louis Pasteur)

RESUMO

COVID-19 (*Coronavirus Disease*) está incluída no grupo de doenças com importante ameaça à saúde pública, principalmente em países tropicais e com grande percentual de população de baixa renda, como é o caso do Brasil. A COVID-19 foi a causa da recente pandemia global que manteve o estado de emergência por aproximadamente 3 anos. A doença é uma infecção viral cujas manifestações sintomáticas se assemelham a outras doenças, como as doenças causadas por Arbovírus e Influenza, apesar de necessitarem de tratamentos distintos. Neste contexto, o teste diagnóstico laboratorial assertivo e acessível à população acometida se faz necessário, apesar de ainda ser limitado, sendo a investigação de novas estratégias analíticas aplicáveis ao diagnóstico destas doenças a proposta deste trabalho. Otimização envolvendo coleta, preparo e armazenamento de amostras biológicas foram inicialmente estudadas, seguido do desenvolvimento de métodos analíticos diversos utilizando uma abordagem instrumental multiplataforma em associação com ciência de dados para o estudo investigativo de metabólitos presentes na urina de pacientes acometidos com a infecção em foco em comparação com indivíduos saudáveis, sendo gerados conjuntos de dados variados e extensos. Foram otimizados e validados métodos segmentados de análise de ácidos orgânicos, aminoácidos e creatinina por eletroforese capilar. e abrangentes por cromatografia líquida acoplada a espectrometria de massas de alta resolução e espectroscopia no infravermelho utilizando ferramentas computacionais e planejamento de experimentos voltado para busca de biomarcadores. Resultados qualitativos obtidos por eletroforese capilar com detecção por ultravioleta aplicados a grupos numerosos de amostras de urina com precisão de 1.22% e por espectroscopia de infravermelho foram aplicados para construção de modelos de predição por aprendizado de máquina para a COVID-19 com exatidão de 77.9 % e 81.0 %, respectivamente, úteis como rotinas de triagem analítica. Análises quantitativas de creatinina em urina por eletroforese capilar com detecção UV utilizando calibração por padrão interno com protocolo de injeção automatizada foram validadas, obtendo precisão analítica de 2.43 %. O uso de eletroforese capilar associado a espectrometria de massa também foi avaliado. A aplicabilidade dos métodos em amostras diversas e em outras frentes analíticas são também discutidas ao longo do texto.

Palavras-chave: metabolômica; COVID-19; urina; eletroforese capilar; espectrometria de massas.

ABSTRACT

COVID-19 (*Coronavirus Disease*) is now a part of a group of diseases that represents an important threat to public health, mainly in tropical countries and with a large percentage of low-income populations, as is the case from Brazil. COVID-19 was the cause of the recent global pandemic that maintained a state of emergency for approximately 3 years. The disease is a viral infection caused by the new SARS-CoV-2 that shares symptomatic manifestations with other diseases, such as the ones caused by Arboviroses and Influenza for example, despite requiring different treatments. In this context, assertive laboratory diagnosis tests that are accessible to the affected population is necessary. Therefore, the investigation of new analytical strategies applicable to the diagnosis of COVID-19 is the proposal of this work. Optimizations involving collection, preparation and storage of biological samples were initially studied, followed by the development of a series of analytical methods using a multiplatform instrumental approach in association with data science for the investigative study of metabolites present in urine of patients affected by COVID-19 infection in comparison with healthy individuals, generating diverse and comprehensive data sets. Segmented methods for analyzing organic acids, amino acids and creatinine by capillary electrophoresis were optimized and validated., as well as comprehensive approaches using liquid chromatography coupled to high-resolution mass spectrometry and infrared spectroscopy using computational tools and experimental design aimed at searching for biomarkers. Qualitative results obtained by capillary electrophoresis with ultraviolet detection applied to numerous groups of urine samples with an accuracy of 1.22% and by infrared spectroscopy were applied to build machine learning prediction models for COVID-19 with an accuracy of 77.9% and 81.0%, respectively, useful as analytical screening routines. Quantitative analysis of creatinine in urine by capillary electrophoresis with UV detection using internal standard calibration with automated injection protocol were validated, obtaining analytical precision of 2.43%. The use of capillary electrophoresis associated with mass spectrometry was also evaluated. The applicability of the methods to different samples and other analytical fronts are also discussed throughout the text.

Keywords: metabolomics, COVID-19; urine; capillary electrophoresis; mass spectrometry.

LISTA DE ILUSTRAÇÕES

- FIGURA 1 - Resumo esquemático das etapas do trabalho 32
- FIGURA 2 – Representação ilustrativa do processo inicial de armazenamento das amostras 36
- FIGURA 3 - Representação ilustrativa do preparo das amostras de urina a partir de réplicas autênticas 37
- FIGURA 4 - Curva de mobilidade efetiva vs pH exemplificando o comportamento de alguns ácidos orgânicos dicarboxílicos e hidroxiácidos 48
- FIGURA 5 - Eletroferogramas de QC de urina (A) com condicionamento convencional: NaOH 1 mol L⁻¹ (60 s), água (60 s) e BGE (60 s) e (B) com condicionamento otimizado: NaOH 1 mol L⁻¹ (60 s), água (30 s), HCl 1 mol L⁻¹ (30 s), água (30 s) e BGE (120 s). 49
- FIGURA 6 - Eletroferogramas resultantes da análise da mistura de padrões (em preto), QC de urina (em azul). (1) Tartarato; (2) Malato; (3) Lactato; (4) Succinato; *Cloroeto e outros ânions rápidos. 51
- FIGURA 7 - Exemplo de eletroferogramas resultantes da análise de ácidos orgânicos dicarboxílicos e hidroxiácidos de 20 urinas diferentes 53
- FIGURA 8 – Esquema representativo das etapas referentes ao processamento de dados obtidos por CZE-UV a partir de ferramentas de aprendizado de máquina. 56
- FIGURA 9 – Resultados do pré-processamento dos eletroferogramas. 57
- FIGURA 10 – Outliers 57
- FIGURA 11- Eletroferogramas resultantes da análise da mistura de padrões (em preto), QC de saliva (em laranja). (1) Tartarato; (2) Malato; (3) Lactato; (4) Succinato. 63
- FIGURA 12 – Curva de mobilidade efetiva vs pH exemplificando o comportamento de alguns ácidos orgânicos dicarboxílicos e hidroxiácidos contendo possíveis interferentes do sinal em t = 9 min (em linhas coloridas). 64
- FIGURA 13 – Eletroferograma simulado contendo interferentes do tartarato (em azul) comparado a uma amostra QC de saliva (em laranja). (1) Oxalato; (2) Malonato; (3) Tartarato; (4) Fumarato; (5) Fosfato e (6) Citrato. 65
- FIGURA 14 – Eletroferogramas resultantes da análise individual de algumas amostras aleatórias de saliva. 67
- FIGURA 15 – Box-Plot da distribuição de concentração de tartarato salivar (em termos de massa por volume de saliva) entre homens e mulheres. 67

- FIGURA 16 - Curva de mobilidade efetiva vs pH exemplificando o comportamento de alguns ácidos orgânicos monocarboxílicos de cadeia curta 69
- FIGURA 17 - Eletroferogramas resultantes da análise da mistura de padrões (em preto), QC de saliva (em laranja) e QC de urina (em azul). (1) Formiato; (2) Acetato; (3) Propionato; (4) Butirato; (5) Valerato. 70
- FIGURA 18 – Representação do planejamento em estrela para duas variáveis.71FIGURA 19– Eletroferogramas resultantes dos testes de otimização do eletrólito para o método dos ácidos orgânicos monocarboxílicos. 73
- FIGURA 20 – Simulação de analitos do método para ácidos monocarboxílicos.74FIGURA 21 - Curva de mobilidade efetiva da creatinina, imidazol e aminoácidos essenciais (em escala de cinza). 81
- Figura 22 – Estudo do efeito de matriz para análise quantitativa de creatinina. (A) Eletroferogramas resultantes da análise de (I) matriz sintética; (II) matriz adicionada de padrão de creatinina; (III) QC de urina. (B) Espectros UV dos picos de creatinina em solução padrão (em lilás) e em urina (em azul). 83
- FIGURA 23 - Representação ilustrativa do sistema de injeção com adição de padrão interno “on-column”. 83
- FIGURA 24 - Eletroferogramas da aplicação do método em (I) simulação no software Peakmaster®; (II) padrões de creatinina e alguns aminoácidos; (III) QC de urina; (IV) QC de soro sanguíneo e (V) matriz artificial de soro sanguíneo. 85
- FIGURA 25 - Box-Plot da distribuição de concentração de creatinina em urina entre indivíduos saudáveis e com histórico de COVID-19 e Dengue. 88
- FIGURA 26 - Eletroferogramas do perfil de aminoácidos de 10 amostras de urina aleatórias analisadas por CZE-UV. 89
- FIGURA 27 - Eletroferograma resultante da análise de aminoácidos em saliva por CZE-UV. 90
- FIGURA 28 - Resultados MRM das análises de aminoácidos realizadas por CZE-MS/MS 92
- FIGURA 29 - Eletroferogramas totais da análise da mistura de padrões de aminoácidos por CZE-MS/MS na concentração de $10 \mu\text{mol L}^{-1}$. 94
- FIGURA 30 - Espectro de massa do tipo *profile* de um QC de urina obtido a partir do método por CZE-MS. 95

FIGURA 31 - Representação do planejamento fatorial 3^3 tipo Box-Behnken com fatores codificados.	98
FIGURA 32 Cromatogramas resultantes da aplicação do planejamento de experimentos	102
FIGURA 33 - Superfície de resposta da interação entre a voltagem do fragmentador (X_3) e Nozzle voltage (X_1) considerando o número de moléculas geradas pela aplicação de cada combinação de fatores	105
FIGURA 34 - Cromatogramas de pico base extraídos para uma amostra controle de urina a partir do método otimizado no modo ESI + (A) e ESI – (B).	107
Figura 35 - TIC e BPC resultante da aplicação do método em outras matrizes biológicas.	108
FIGURA 36 - Esquema representativo das etapas referentes ao processamento de dados obtidos por CZE-UV a partir de ferramentas de aprendizado de máquina	111
FIGURA 37 - Espectros vibracionais por infravermelho médio de amostras de urina.	113
FIGURA 38 - Resultado comparativo da aplicação de pré-processamento aos dados de FTIR. (A) Sem tratamento; (B) SNV; (C) MSC.	116

LISTA DE TABELAS

TABELA 1– Características do conjunto amostral para estudos relacionados à COVID-19	38
Tabela 2 - Características do conjunto amostral para estudos relacionados às arboviroses combinadas com COVID-19	40
TABELA 3 Parâmetros instrumentais dos métodos analíticos para ácidos orgânicos desenvolvidos por CZE-UV	45
TABELA 4 - Equações para cálculo de falta de ajuste e significância da regressão	47
TABELA 5 - Regressão e parâmetros estatísticos utilizados para verificação da qualidade de dados	54
TABELA 6 - Avaliação comparativa de desempenho de todos os modelos classificatórios por CZE-UV.	60
TABELA 7 - Matriz de confusão combinada dos modelos de aprendizado de máquina não lineares associados aos dados de CZE-UV.	61
TABELA 8 – Fatores de variação do planejamento estrela.	72
TABELA 9 - Parâmetros experimentais dos métodos desenvolvidos por CZE-UV para aminoácidos e creatinina.	77
TABELA 10 - Parâmetro de fragmentação das massas de transição de cada analito.	78
Tabela 11 - Pureza de Pico calculada para a creatinina urinária.	82
TABELA 12 - Dados comparativos de reprodutibilidade a partir de injeção convencional e “on-column” em termos das áreas e tempo de migração de Creatinina urinária (UCrn) e Imidazol (IS, 20 mmol L ⁻¹).	84
TABELA 13 - Modelos de regressão linear por adição de padrão interno e parâmetros de performance analítica para doseamento de creatinina em urina e soro sanguíneo expressos em valores de média e desvio padrão. RSD em parênteses.'método otimizado por CZE-UV expressos como média de cada triplicata e seus respectivos RSD.	91

TABELA 15 - Modelos de regressão linear por adição de padrão interno para aminoácidos em urina por CZE-ESI-QQQ.	93
TABELA 16 - Níveis associados a cada experimento descrito pelo planejamento Box-Behnken.	99
TABELA 17 - Informações extraídas dos cromatogramas resultantes da aplicação do planejamento factorial	103
TABELA 18 - Coeficientes resultantes do modelo quadrático	104
TABELA 19 -Análise de variância associadas aos ajustes do modelo quadrático	105
TABELA 20 - Avaliação comparativa de desempenho de todos os modelos classificatórios por FTIR.	114
TABELA 21 - Matriz de confusão combinada dos modelos de aprendizado de máquina não lineares associados os dados de FTIR	115

LISTA DE ABREVIATURAS E SIGLAS

3,5-DNB	Ácido 3,5-Dinitrobenzóico
ANN	<i>Artificial Neural Network</i>
ANOVA	<i>Analysis of Variance</i>
ANVISA	Agência Nacional de Vigilância Sanitária
ATR	<i>Attenuated Total Reflectance,</i>
AU	<i>Arbitrary Units</i>
BGE	<i>Background Electrolyte</i>
BPC	<i>Base Peak Chromatogram</i>
CDC	<i>Centers for Disease Control and Prevention</i>
CE-MS	<i>Capillary Electrophoresis with Mass Spectrometry</i>
CEP	Comitê de Ética em Pesquisa
CE-UV	<i>Capillary Electrophoresis with UV Detection</i>
CHIKV	Vírus da Chikungunya
CoV	<i>Coronavírus</i>
COVID-19	<i>Coronavirus Disease 2019</i>
CTAB	Brometo de Cetiltrimetilamônio
CZE	<i>Capillary Zone Electrophoresis</i>
DAD	<i>Diode Array Detector</i>
DENV	Vírus da Dengue
DLN	<i>Deep Learning</i>
DOE	<i>Design of Experiments</i>
EIC	<i>Extracted Ion Chromatogram</i>
ELISA	<i>Enzyme Linked Immuno Sorbent Assay</i>
EOF	<i>Electroosmotic Flow</i>
EPI	Equipamentos de Proteção Individual
ESI	<i>Electrospray Ionization,</i>
FDA	<i>Food and Drug Administration</i>
FTIR	<i>Fourier-Transform Infrared Spectroscopy</i>
GC	Grupo Controle
GL	Graus de Liberdade

GT	Grupo Teste
HPLC-MS	<i>High Performance Liquid Chromatography with Mass Spectrometry</i>
ICH	<i>International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use</i>
IR	<i>Infrared</i>
IS	<i>Internal Standard</i>
LAMP	<i>Loop-Mediated Isothermal Amplification</i>
LOC	<i>Lab-on-a-Chip</i>
MCC	<i>Matthew's Correlation Coefficient</i>
MERS-CoV	<i>Middle East Respiratory Syndrome Coronavirus</i>
MIR	<i>Mid-Infrared Spectroscopy</i>
ML	<i>Machine Learning</i>
MRM	<i>Multiple Reaction Monitoring,</i>
MS/MS	<i>Tandem Mass Spectrometry</i>
MSC	<i>Multiplicative Scatter Correction</i>
OLS	<i>Ordinary Least Squares,</i>
OMS	Organização Mundial da Saúde
OPLS-DA	<i>Orthogonal Partial Least Square Discriminant Analysis</i>
PCA	<i>Principal Component Analysis</i>
PHEIC	<i>Public Health Emergency of International Concern</i>
PLS-DA	<i>Partial Least-Squares Discriminant Analysis</i>
POC	<i>Point-of-Care</i>
QC	<i>Quality Control,</i>
QQQ	Espectrômetro de Massas Triplo-Quadrupolo
QTOF	<i>Quadrupole Time-of-Flight</i>
RDC	Resolução da Diretoria Colegiada
RF	<i>Random Forest</i>
RSD	<i>Relative Standard Deviation</i>
RSM	<i>Response Surface Methodology</i>
RT-PCR	<i>Reverse Transcription Polymerase Chain Reaction</i>
SARS-COV-2	<i>Severe Acute Respiratory Syndrome Coronavirus 2</i>

SCFA	<i>Short-Chain Fatty Acids,</i>
Scrn	Creatinina Sérica
SEIP	<i>Short-End Injection Programm</i>
SNV	<i>Standard Normal Variation,</i>
SUS	Sistema Único de Saúde
TCLE	Termo de Consentimento Livre e Esclarecido
TIC	<i>Total Ion Chromatogram</i>
TRIS	Tris-(Hidroximetil)-Aminometano
UBS	Unidades Básicas de Saúde
Ucrn	Creatinina Urinária
UV	Ultravioleta
ZIKV	Vírus da Zika

LISTA DE SÍMBOLOS

C_x	concentração do analito
F_{calc}	valor calculado para teste f
F_{tab}	valor tabelado para teste f
H	altura de pico
m	quantidade de níveis independentes
MQ_{ep}	média quadrática de erro puro
MQ_{faj}	média quadrática da falta de ajuste
MQ_R	média quadrática de regressão
MQ_r	média quadrática dos resíduos
n_i	o número de repetições no nível
p	número de parâmetros do modelo
r	coeficiente de correlação
S_b	desvio padrão da linha base
SQ_{ep}	soma quadrática de erro puro
SQ_{faj}	soma quadrática da falta de ajuste
SQ_R	soma quadrática de regressão
SQ_r	soma quadrática dos resíduos
\bar{y}	média dos valores observados
\hat{y}_i	valores previstos de y
\bar{y}_i	média dos valores observados em cada nível
y_{ij}	valores observados
β	coeficiente
β_0	termo constante
β_i	termo de primeira ordem
β_{ij}	termo quadrático
β_{ij}	termo de interação
v	grau de liberdade
α	distribuição iônica

μ_{eff}

mobilidade efetiva

μ_j

mobilidade eletroforética

SUMÁRIO

1 INTRODUÇÃO	22
1.1 MEDICINA DIAGNÓSTICA	23
1.2 METABOLÔMICA	24
1.3 COVID-19	25
1.3.1 Testes diagnósticos	27
1.4 JUSTIFICATIVAS DO PROJETO	28
2 OBJETIVOS	31
3 ESTRATÉGIAS DE AÇÃO	32
4 AMOSTRAS	34
4.1 CEP COVID-19	34
4.2 COLETA E ARMAZENAMENTO DAS AMOSTRAS DE URINA	35
4.2.1 Preparo de amostras	36
4.2.2 Especificações do conjunto amostral	38
4.3 AMOSTRAS DE SALIVA E SORO SANGUÍNEO	39
4.4 AMOSTRAS COM HISTÓRICO DE ARBOVIROSES	40
5 METABOLÔMICA ALVO POR ELETROFORESE CAPILAR	42
5.1 ÁCIDOS ORGÂNICOS	43
5.1.2 Instrumentação e métodos	44
5.1.3 Construção de curvas de mobilidade	45
5.1.4 Análise de variância e verificação de pressupostos estatísticos	46
5.1.5. Perfil de ácidos orgânicos dicarboxílicos e hidroxíácidos em urina	47
5.1.6 Perfil de ácidos orgânicos dicarboxílicos e hidroxíácidos em saliva	62
5.1.7 Perfil de ácidos orgânicos monocarboxílicos de cadeia curta	68
5.2 AMINOÁCIDOS E CREATININA	75
5.2.1 Reagentes e soluções	76

5.2.3 Instrumentação e métodos por CZE-MS	77
5.2.4 Construção de modelos de regressão e parâmetros de desempenho analítico	79
5.2.5 Perfil de aminoácidos e doseamento de creatinina por CZE-UV.....	80
5.2.4 Análise de aminoácidos em urina por CZE-MS	92
6 METABOLÔMICA GLOBAL POR ESPECTROMETRIA DE MASSAS DE ALTA RESOLUÇÃO.....	97
6.1 DESENVOLVIMENTO E OTIMIZAÇÃO DE MÉTODO ANALÍTICO POR HPLC-ESI-QTOF	97
6.1.1 Reagentes e soluções	99
6.1.2 Instrumentação e métodos	100
6.1.3 Experimental.....	100
6.1.4 Processamento de dados.....	101
6.1.5 Resultados e discussões	101
6.2 APLICAÇÃO DO MÉTODO	106
6.3 CONSIDERAÇÕES GERAIS.....	108
7 APRENDIZADO DE MÁQUINA ASSOCIADO À ESPECTROSCOPIA VIBRACIONAL POR INFRAVERMELHO.	110
7.1.1 Instrumentação e aquisição de dados	111
7.1.2 Processamento de dados.....	112
7.1.3 Resultados e discussões	112
8 CONCLUSÃO.....	118
REFERÊNCIAS	120
APÊNDICE I.....	132
APÊNDICE II.....	134
APÊNDICE III.....	137

1 INTRODUÇÃO

A medicina diagnóstica é uma prática baseada no processo investigativo relacionado à saúde de um indivíduo que busca por assistência médica ao identificar manifestações fisiológicas incomuns. Diagnósticos precisos vem sendo o foco da medicina moderna que une conhecimentos multidisciplinares, atendimento personalizado, evidências individuais e coletivas somadas às investigações a nível molecular, o que demanda por testes diagnósticos mais assertivos, específicos e acessíveis à população. Neste contexto, a química analítica e as rotinas científicas baseadas no estudo do metaboloma humano são uma importante estratégia para o desenvolvimento e aprimoramento de testes diagnósticos de doenças em geral. O presente trabalho foi planejado originalmente como um estudo aprofundado do perfil metabólico de pacientes acometidos individualmente com Dengue, Zika e Chikungunya a partir da análise abrangente de amostras de urina e saliva destes pacientes, na intenção de propor uma nova abordagem que auxilie no diagnóstico de cada uma dessas doenças, sendo estas estratégias não invasivas, confiáveis e com custo reduzido. Ainda no início da realização das ações sugeridas neste trabalho, um projeto com objetivo semelhante, aplicado à COVID-19, foi incorporado ao trabalho. Similarmente, o foco desta proposta adicional é o estudo comparativo de metabólitos encontrados em urina de pacientes contaminados e não contaminados com o vírus causador da doença pandêmica. Realizado a partir do uso de instrumentação analítica diversa associado a ferramentas estatísticas, as ações do projeto foram direcionadas para o desenvolvimento de novas abordagens diagnósticas. Uma vez que as infecções supracitadas apesar de apresentarem particularidades se assemelham em alguns aspectos, principalmente nas manifestações sintomáticas iniciais cujo diagnóstico precoce pode ser decisivo, o estudo comparativo entre todas as doenças mencionadas também se faz pertinente. Destaca-se que, em vista da importância do estado emergencial ocasionado pela COVID-19, os esforços foram direcionados ao cumprimento deste projeto. Durante a realização do trabalho, algumas outras aplicações para os métodos analíticos desenvolvidos foram sendo investigadas, considerando a busca por aprimoramento de testes diagnósticos de outras doenças a partir das análises dos materiais biológicos supracitados e outros. Ao longo deste texto serão discutidos os objetivos, propostas, estratégias de ação, desdobramentos dos projetos, estudos já realizados e perspectivas para o trabalho.

1.1 MEDICINA DIAGNÓSTICA

O diagnóstico de qualquer doença é um conceito chave dos preceitos da assistência médica, sendo definido como o “conjunto preexistente de categorias acordado pela profissão médica para designar uma condição específica”. A partir do diagnóstico, o médico propõe uma explicação sobre as manifestações sintomáticas do paciente, que conseqüentemente, guiará as propostas de ações subsequentes (BALOGH; MILLER; BALL, 2016). De acordo com o Instituto de Medicina (*Institute of Medicine*, IOM) da Academia Nacional de Ciências dos Estados Unidos (*The National Academies of Science, Engineering and Medicine*), o diagnóstico de uma doença é um processo complexo que deve ser feito a partir de atividades colaborativas (BALOGH; MILLER; BALL, 2016; BELARD et al., 2017).

Uma vez que um indivíduo busca por assistência médica se inicia um processo investigativo contínuo e interativo centrado nas observações do paciente e baseado na coleta abrangente de informações. Histórico clínico, exames físicos, exames laboratoriais e consultas em especialistas de áreas diversas são os pilares do processo diagnóstico, que tem como objetivo a integração e a interpretação de informações voltadas para a proposição de hipóteses que culminem em uma lista de potenciais diagnósticos, ou em um diagnóstico definitivo. O processo segue até que as abordagens de tratamento tenham sido definidas, de modo que o paciente tenha seu problema de saúde resolvido (BALOGH; MILLER; BALL, 2016).

São considerados erros de diagnóstico, a falha em estabelecer uma explicação precisa para o problema do paciente em tempo hábil, a não aplicação de testes indicados ou uso de testes e terapias inadequadas, além da falha nas ações sobre os resultados e monitoramento de testes (BALOGH; MILLER; BALL, 2016; BELARD et al., 2017). Neste sentido, a medicina de precisão, apesar de não ser um conceito novo, vem se tornando uma tendência na área. Baseada no atendimento personalizado e com foco no diagnóstico mais assertivo, nesta prática são considerados tanto os dados individuais e características populacionais, além das informações clínicas multidisciplinares, estilo de vida, genética, fatores ambientais e estudo de biomarcadores (KHOURY; IADEMARCO; RILEY, 2016; KÖNIG et al., 2017; MONTAGU; PRINCENTON, 2015; TEBANI et al., 2016).

Considerando os testes diagnósticos especificamente, foi a partir do desenvolvimento da química, fisiologia, fisiopatologia e dos microscópios no século XIX que houve um grande progresso na medicina diagnóstica (MARTINS, 2014). Atualmente, a química analítica se insere nesse cenário como uma ferramenta técnico-científico para o desenvolvimento de novos testes, sobretudo com a rápida evolução das ciências ômicas (AHMED, 2020; MCCOLL et al., 2019; TEBANI et al., 2016).

Entende-se por ciências ômicas a integração de conhecimentos multidisciplinares voltados para a caracterização abrangente dos constituintes de um organismo. O termo “ômica” do inglês “*omics*” é relacionado ao conceito inicial de genoma e se refere a todas as diferentes tecnologias utilizadas para atingir um entendimento a nível molecular do objeto em estudo (CANUTO et al., 2018; SAFADY, 2021). Nas últimas duas décadas houve um crescimento expressivo na área, justamente devido ao seu potencial em adquirir conjuntos de dados extensos que são traduzidos em informações de grande utilidade médica. Com ela, o desenvolvimento da biotecnologia, das instrumentações analíticas e da ciência de dados emergiu ao que alguns autores chamam de nova era da medicina integrativa e/ou medicina de precisão, visto que investigações avançadas a nível molecular podem hoje ser aplicadas em análises de rotinas para estudo de doenças diversas (OLIVIER et al., 2019; PLAZA; GARCÍA-GALBIS; MARTINEZ-ESPINOSA, 2017).

1.2 METABOLÔMICA

Como parte da “cascata ômica”, termo do inglês “*omics cascade*”, a metabolômica é a ciência utilizada com o objetivo de investigar o funcionamento de organismos e suas alterações metabólicas a partir de abordagens imparciais de alta eficiência (CANUTO et al., 2018; PATTI; YANES; SIUZDAK, 2012; TEBANI et al., 2016). A metabolômica em particular é pautada no estudo aprofundado do metabolismo de espécies em geral frente às perturbações externas e internas a nível molecular, considerando espécies de até 1500 Da. Na prática, são estudados comparativamente os conjuntos de metabólitos (metaboloma) de um grupo de indivíduos sujeito a qualquer tipo de alteração fisiológica sob investigação, frente ao metaboloma de outro grupo, ausente da alteração selecionada, sendo estes chamados de grupo teste e controle, respectivamente (BYERS et al., 2019; CANUTO et al., 2018).

Do ponto de vista analítico, as pesquisas inseridas no conceito da metabolômica são realizadas a partir de um fluxograma de trabalho que foca em análises abrangentes (metabolômica global, *untargeted*) ou em grupos classes específicas de metabólitos (metabolômica alvo, *targeted*). De acordo com os preceitos da metabolômica, a aplicação de instrumentação analítica diversa, tais como, cromatografias, técnicas espectroscópicas e de eletromigração, espectrometria de massas e ressonância magnética nuclear é recomendada a fim de reunir informações de forma abrangente e complementar visando um resultado completo e representativo da realidade da amostra (CANUTO et al., 2018). Aliado à instrumentação estão, processamentos de dados, análises estatísticas diversas e conhecimentos biológicos associados à interpretação de resultados, além de cuidados com coleta, preparo e armazenamento de amostras biológicas, junto com outras ações inseridas nas fases pré-analíticas do estudo. Em geral, o processo pré-analítico e analítico e o tipo de amostra que será utilizado são pensados a partir da definição do propósito do estudo (CANUTO et al., 2018; PATTI; YANES; SIUZDAK, 2012).

Um dos objetivos comuns dos estudos envolvendo a metabolômica, como o deste trabalho, é a descoberta de biomarcadores ou outras características que diferenciam estes grupos entre si, de modo que a pesquisa possa evoluir até a elaboração de dispositivos e/ou rotinas laboratoriais que atuem como um teste que auxilie no diagnóstico da anomalia em foco. Neste trabalho, os objetos principais de estudo são a Dengue, Zika, Chikungunya e a recente COVID-19, considerando o perfil metabólico diferencial entre indivíduos contaminados com cada uma dessas doenças comparado a indivíduos saudáveis.

1.3 COVID-19

A COVID-19 (*Coronavirus Disease 2019*) causada pelo SARS-CoV-2 (*Severe Acute Respiratory Syndrome Coronavirus 2*), compartilha de alguns aspectos com as arboviroses em foco neste trabalho. Descoberta no final de 2019 em Wuhan (China), a doença foi declarada pela OMS como emergência global entre o período de março de 2020 a maio de 2023. Os coronavírus (CoV) são uma classe de vírus diversa da família *Coronaviridae* encontrados em aves, mamíferos e outras espécies de hospedeiros. Em geral, são agentes causadores de doenças respiratórias e intestinais.

Em 2002, um grande surto de pneumonia com características atípicas foi notificado em Hong Kong, causado pelo o que viria ser identificado como o coronavírus

da síndrome respiratória aguda grave (SARS-CoV), originado em morcegos e transmitidos para humanos por meio de intermediários como algumas espécies de cães. Em 2012, um novo coronavírus foi isolado na Arábia Saudita, sendo posteriormente chamado de MERS-CoV (*Middle East Respiratory Syndrome Coronavirus*) e, por fim, o betacoronavírus SARS-CoV-2 sendo isolado em 2019 já com grande incidência global e rápida transmissão entre humanos. Ao longo do período de emergência global, algumas cepas variantes do SARS-CoV-2, foram identificadas ao gerar novos picos de crescimento no número de casos, como as variantes Delta e Ômicron (LAMERS; HAAGMANS, 2022; SUN et al., 2020; UZUNIAN, 2020).

A doença é transmitida por: (a) contato direto com indivíduo infectado; (b) exposição a gotículas respiratórias na qual o vírus esteja contido; ou (c) gotículas menores dispersas no ar (aerossol) a uma distância estimada de um raio de 1 metro. Ou seja, a infecção é dada a partir da exposição de mucosas com gotículas contaminadas com vírus. Além disso, há registros que o SARS-CoV-2 sobrevive, ainda que por tempo limitado, em superfícies. Dessa forma, a recomendação da OMS e órgãos de saúde pública durante o auge da contaminação da doença foi o isolamento social, higienização imediata das mãos mediante contato a superfícies potencialmente contaminadas e uso de máscaras faciais para mitigar o contágio. Apesar da fase aguda da emergência já ter se encerrado, algumas informações como ciclo de transmissão, período de transmissibilidade, suscetibilidade e outras informações ainda estão sendo estudados pela comunidade médico-científica (BRASIL, 2021; SUN et al., 2020; WHO, 2021).

Em relação aos sintomas, quando presentes os mais comuns incluem febre, tosse, coriza, irritação na garganta, congestão nasal, fadiga, dores no corpo e enxaqueca, diarreia, perda de olfato e paladar em casos leves e moderados. Em casos graves, é comum o aparecimento de desconforto respiratório com saturação de oxigênio menor que 95 % e a taquipneia (respiração acelerada) em crianças. Com frequência a COVID-19 foi associada a manifestações neurológicas. A doença evolui para um caso crítico se apresentar sepse e insuficiência respiratória grave, levando a uma falência múltipla dos órgãos (BRASIL, 2023; CDC, 2024). Em geral, as manifestações sintomáticas em crianças são mais leves do que em adultos e estima-se que cerca de 80 % da população acometida com a doença no Brasil apresentaram sintomas leves e moderados, 15 % desenvolveram sintomas graves e 5 % evoluíram para uma forma crítica da doença (BRASIL. MINISTÉRIO DA SAÚDE., 2023; PARK et al., 2020). O total acumulado

mundial de casos, atualizado em 7 de janeiro de 2024 pela OMS, está em cerca de 774 milhões com 7 milhões de óbitos reportados. No Brasil, somam-se 37,5 milhões de casos com 702 mil óbitos registrados (WHO, 2024b).

Em vista da importância do estado pandêmico e rápida evolução da doença, os esforços da comunidade médico-científica foram voltados para o desenvolvimento de testes diagnósticos, medicamentos, atendimentos de emergência e vacinas. Em julho de 2020 a primeira vacina foi introduzida e atualmente 13.5 bilhões de doses já foram administradas, o que representa 67 % da população mundial com a primeira série de imunização completa e 32 % com pelo menos uma dose de reforço (WHO, 2023d).

1.3.1 Testes diagnósticos

O “padrão-ouro” de testagem adaptado para identificação do SARS-CoV-2 é o, RT-PCR, que consiste na detecção de RNA viral mediada pela análise de material coletado do nariz e garganta de paciente a partir de 3 dias do início dos sintomas (DUTTA et al., 2022; OLIVEIRA et al., 2022; YÜCE; FILIZTEKIN; ÖZKAYA, 2021). Conforme introduzido nas seções anteriores, o teste RT-PCR baseia-se na extração do RNA viral do material coletado da nasofaringe do paciente com auxílio dos chamados *swabs*, a amostra então é submetida ao processo de transcrição reversa em cDNA (DNA complementar) seguido de amplificação, resultando no valor da carga viral da amostra (CDC, 2023a; YÜCE; FILIZTEKIN; ÖZKAYA, 2021) Apesar de ser o exame laboratorial de primeira escolha, de acordo com Rabaan e colaboradores (2021), há certa discordância da comunidade médica sobre qual é de fato a carga viral que melhor representa um diagnóstico positivo ou negativo para a doença (RABAAN et al., 2021). Conforme descrevem alguns autores especializados, o teste RT-PCR apresenta algumas limitações quanto à demanda de pessoal qualificado e de um laboratório de biossegurança de nível 2 no caso da COVID-19 devido suas características de contágio, tudo isso resultando em um exame oneroso para o paciente (DUTTA et al., 2022).

Os testes de antígeno, popularmente chamados de testes rápidos, fornecem resultados de 15 a 30 minutos com custo reduzido, incluindo os autotestes agora disponíveis para consumidores em farmácias. Esses testes são realizados por meio de ensaios de fluxo lateral, onde a amostra de material da nasofaringe reage com uma solução para que o vírus libere seus antígenos. Após colocar a solução sobre um dispositivo de papel, esses antígenos migrarão pelo dispositivo à medida que reagem com os anticorpos

específicos do Sars-Cov-2, obtendo uma resposta qualitativa. De acordo com os Centros de Controle e Prevenção de Doenças (*Centers for Disease Control and Prevention, CDC*), os resultados positivos de um teste de antígeno são muito confiáveis, porém, em comparação com testes baseados em ácido nucleico como RT-PCR, são menos sensíveis para detectar partículas virais e a auto coleta de material pode ser ineficaz, portanto, um único resultado negativo desse tipo de teste não descarta infecção (CDC, 2023a, 2023b; DUTTA et al., 2022).

Teste sorológico por técnicas que envolvem ELISA ou amplificação isotérmica de ácidos nucleicos (*Loop-mediated isothermal amplification, LAMP*) são exemplos de outras opções já incluídas nos portfólios de exames clínicos (BÖGER et al., 2021; OLIVEIRA et al., 2022). Os novos testes disponibilizados em farmácias recentemente também incluem exames de COVID-19 (antígenos, *check-up* de anticorpos pós covid, teste de imunidade e teste molecular) (CASERTA; BAIENSE; ANDRADE, 2023).

Algumas outras alternativas são discutidas em revisões recentes e incluem a fabricação de sensores, biossensores, sondas baseadas em espectroscopias, nanomateriais e outros (BÖGER et al., 2021; FILCHAKOVA et al., 2022). Nessas publicações são ponderados prós e contras de cada teste, que envolvem em sua maioria uma discussão sobre relação entre o custo do exame para o paciente e a confiabilidade do resultado, além dos maiores cuidados técnicos para realização de testes laboratoriais no caso da COVID-19. Autores da área médica defendem ainda que o exame laboratorial por si só não pode ser definitivo, devendo ser associado a exames radiológicos, por exemplo (WALLER et al., 2020) uma vez que o RT-PCR ou os imunoenaios pode ser influenciado por fatores como coleta de material, momento ideal para testagem e a variação genética recorrente do Sars-CoV-2 (TAHAMTAN; ARDEBILI, 2020).

1.4 JUSTIFICATIVAS DO PROJETO

A partir de todas as considerações feitas nas seções anteriores, nota-se que há a necessidade de inovações nas estratégias diagnósticas para doenças infectocontagiosas. Visto que apresentam sintomas semelhantes e complicações tão características, o diagnóstico realizado de forma precisa nos primeiros dias de manifestações sintomáticas é um ponto chave para que o tratamento dessas doenças seja funcional. Com foco nas doenças causadas por vírus que compartilham de sequências genéticas semelhantes, os testes baseados na detecção de proteínas virais, outros particulados de vírus e antígenos

demandam por uma maior especificidade. Além disso, um cenário na qual fosse possível identificar qual infecção viral, independente de qual seja, o paciente contraiu utilizando um dispositivo de teste único seria o ideal. Por se tratar de doenças que afetam principalmente a população de baixa renda, todos esses aspectos devem considerar o desenvolvimento de novos métodos que sejam acessíveis, com possibilidade de distribuição em massa.

. Desse modo, o foco deste trabalho foi investigar a possibilidade do desenvolvimento de novas estratégias de diagnósticos que considerem a identificação de metabólitos em preferências à particulados virais pensando em uma coleta não invasiva que proteja o paciente e o responsável por examinar o material. Uma vez que a carga viral de SARS-CoV-2 em saliva é bem significativa, sendo inclusive uma via de transmissão (YOON et al., 2020) e que toda fase analítica se realiza em laboratório de química com nível de biossegurança restrito (NB-2) (PENNA et al., 2010; TANG; MIRANDA, 2019), somado ao fato que as coletas envolveriam maior exposição do paciente, optou-se por não usar amostras salivares de pacientes diagnosticados com COVID-19. Sobretudo, a urina como amostra biológica foi selecionada prioritariamente devido a quantidade de informações que pode ser obtida após análises. De forma complementar, a saliva também foi considerada para estudos de desenvolvimento de método visando outras aplicações.

Neste projeto, não apenas o fluxograma clássico da metabolômica foi explorado. Análises feitas a partir de instrumentações simplificada como infravermelho e eletroforese capilar com detecção UV, ambas associadas a ferramentas estatísticas, também têm sido estudadas com foco na observação de grupos de compostos segmentados como, por exemplo, ácidos orgânicos e aminoácidos. Por fim, uma série de métodos analíticos elaborados a partir da análise de urina e saliva de grupos controle e teste, usando plataformas analíticas diversas para investigação de possíveis grupo de biomarcadores e/ou perfil de compostos que possam ser relacionados a uma impressão digital das doenças em questão, e que futuramente, possam evoluir a um novo exame laboratorial oferecido a indivíduos com suspeita de algumas dessas doenças.

Ainda que o foco deste trabalho seja o diagnóstico, por se tratar dos estudos de metaboloma acredita-se que informações obtidas durante o desenvolvimento do projeto sobre a doença e como elas agem no corpo humano possam ser úteis do ponto de vista clínico. Ademais, os protocolos analíticos implementados ao longo do trabalho poderão se estender a outros estudos investigativos e classificatórios não apenas na área clínica,

como também setores da área ambiental, farmacêutica, alimentícia e outras. Neste contexto, algumas produções incluindo artigos de revisão e de aplicação relacionadas aos temas do projeto foram publicadas em revistas científicas especializadas e estão listadas no Apêndice I.

2 OBJETIVOS

Elaboração de estudos voltados para o desenvolvimento de novas proposições bioanalíticas para auxiliar no diagnóstico rápido e assertivo da COVID-19 a partir da investigação do metaboloma humano acessados a partir de ferramentas analíticas abrangentes. Especificamente, o trabalho é focado em:

(a) realização de estudo de metabólitos como ferramenta inicial para determinação de potenciais biomarcadores moleculares relacionados à COVID-19, utilizando Espectroscopias e Espectrometria de Massas associadas a técnicas de separação;

(b) utilização de ferramentas quimiométricas e de aprendizado de máquina associados à instrumentação analítica diversa para desenvolvimento de rotinas preditivas que diferenciem grupos acometidos COVID-19 e grupos de indivíduos não infectados;

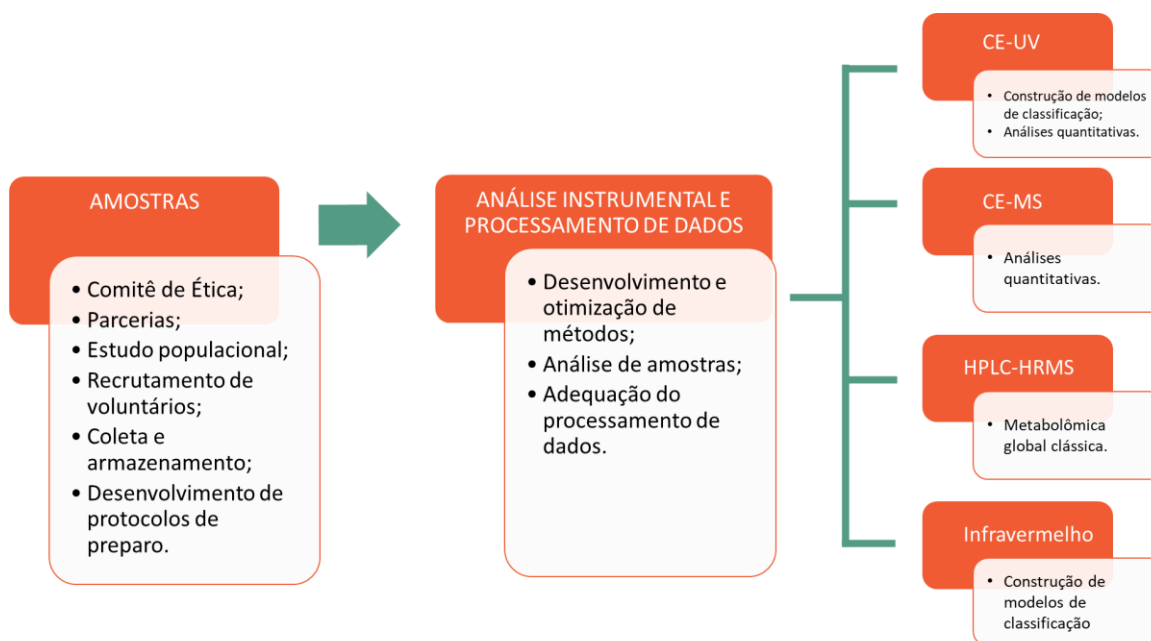
(c) avaliação dos potenciais grupos de biomarcadores moleculares associados à doença e desenvolvimento, otimização e validação métodos analíticos quantitativos aplicáveis ao diagnóstico;

(d). aplicação dos métodos otimizados e validados à análise de fluidos biológicos.

3 ESTRATÉGIAS DE AÇÃO

A fim de cumprir os objetivos estabelecidos, inicialmente foram elaboradas propostas de parcerias e estudos em colaboração com especialistas de áreas diversas para construção de protocolos de trabalho envolvendo coleta, armazenamento e preparo de amostras. De posse das amostras de fluidos biológicos humanos, um trabalho investigativo extenso foi elaborado considerando: (a) aplicação do fluxograma clássico da metaboloma global a partir do uso de espectrometria de massas de altas resolução; (b) desenvolvimento de métodos analíticos considerando instrumentação simples para análise de grupos segmentados de moléculas a fim de estudar possíveis classes de biomarcadores; (c) obtenção de conjuntos de dados espectrais a fim de associá-los com aprendizado de máquina profundo para construção de modelos preditivos; (d) a avaliação crítica das instrumentações analíticas e ferramentas estatísticas envolvidas de modo a inferir sobre os métodos mais adequados para cada aplicação; (e) estudar a possibilidade de aplicação dos métodos analíticos otimizados em outras frentes e inferir sobre como o desenvolvimento do projeto pode contribuir para pesquisas futuras. Neste sentido, um resumo esquemático dos processos analíticos elaborados neste trabalho é representado na Figura 1.

FIGURA 1 - Resumo esquemático das etapas do trabalho



Fonte: Elaborado pela autora (2024).

O texto encontra-se dividido em capítulos de acordo com a técnica analítica e analito, quando pertinente, na qual cada estudo foi fundamentada. No capítulo 4 é disposto um descritivo completo do preparo e manuseio das amostras, bem como as etapas burocráticas envolvendo pesquisa com humanos. Em seguida, as etapas de desenvolvimento analítico, aplicação de método, resultados e discussões associadas a cada etapa foram divididos em três capítulos: o capítulo 5 envolve os experimentos por eletroforese capilar, das quais foram subdivididos em análises de ácidos orgânicos mono- e di- carboxílicos, creatinina e aminoácidos; o capítulo 6 descreve as ações e resultados do estudo a partir da metabolômica global por espectrometria de massas de alta resolução; finalizando a discussões sobre o desenvolvimento do trabalho, o capítulo 7 inclui os resultados da associação entre o aprendizado de máquina e a espectroscopia vibracional por infravermelho.

Conforme descrito na seção introdutória, o projeto original desta tese envolve o estudo do metaboloma humano a partir da análise de urina e saliva de indivíduos infectados com Dengue, Zika e Chikungunya. Desse modo, ainda que os estudos deste trabalho especificamente não tenham sido concluídos, algumas ações e estudos preliminares foram realizados e estão dispostos ao final do documento na forma de APÊNDICE.

4 AMOSTRAS

A coleta de amostras de fluidos biológicos humanos voltados para pesquisa acadêmica requer autorização de um Comitê de Ética (CEP) para que seja devidamente executado. Para este trabalho, foi submetido ao CEP pertinente o projeto de título “Técnicas de eletromigração e cromatográficas aplicadas ao estudo de metaboloma da urina para o diagnóstico da COVID-19”.

4.1 CEP COVID-19

As amostras de urina relacionadas a COVID-19 foram coletadas sob autorização do Comitê de Ética do Hospital Universitário da Universidade Federal de Juiz de Fora (CEP HU-UFJF) sob os protocolos de autorização 4.473.404; 4.566.092; 5.039.371, referentes respectivamente: à autorização de coleta de amostras para realização do projeto; à participação do Lemos Laboratórios de Análises Clínicas como parceiro; ao aumento do n amostral em vista da alta oferta de amostras. Inicialmente, foi definido um quantitativo total de coletas de 500 amostras, sendo posteriormente protocolado um adendo de 500 amostras adicionais. Para autorização de participação de candidatos na pesquisa, os voluntários foram orientados a preencher e assinar um Termo de Consentimento Livre e Esclarecido (TCLE) somado à uma ficha de anamnese, documentos estes aprovados pelo CEP HU-UFJF.

Todas as coletas foram realizadas em parceria com Lemos Laboratórios de Análises Clínicas sediado na cidade de Juiz de Fora, MG, que contém unidades de coleta em diversos bairros da cidade. A pesquisa foi oferecida a indivíduos maiores de 18 anos com suspeita de terem contraído a doença e que se submeteram ao teste de RT-PCR. Os testes de detecção do vírus SARS-CoV-2 foram realizados por RT-PCR utilizando o TaqPath™ COVID-19 CE-IVD RT-PCR KiT da Thermo Fisher Scientific (Palo Alto, CA, EUA) que detecta o vírus nas amostras de esfregaço (*swab*) nasofaríngeo com sensibilidade mínima de 10 cópias genômicas. Os resultados foram processados utilizando o software interpretativo COVID-19 da Thermo Fisher Scientific (Palo Alto, CA, EUA) (THERMOFISHER SCIENTIFIC, 2021).

O recrutamento de voluntários foi realizado por agentes de saúde funcionários do Lemos Laboratórios que, uma vez cientes das especificidades do projeto, abordaram os candidatos informando-os brevemente sobre o trabalho e solicitando a leitura,

preenchimento e assinatura dos documentos supracitados. Para aqueles que aceitaram participar, uma cópia do TCLE contendo informações sobre o projeto foi fornecida. Além da documentação e da amostra não foram solicitados aos voluntários nenhum tipo de preparo e condições específicas de quadro clínico para a doação do material. Todas as coletas aconteceram em períodos randômicos do dia.

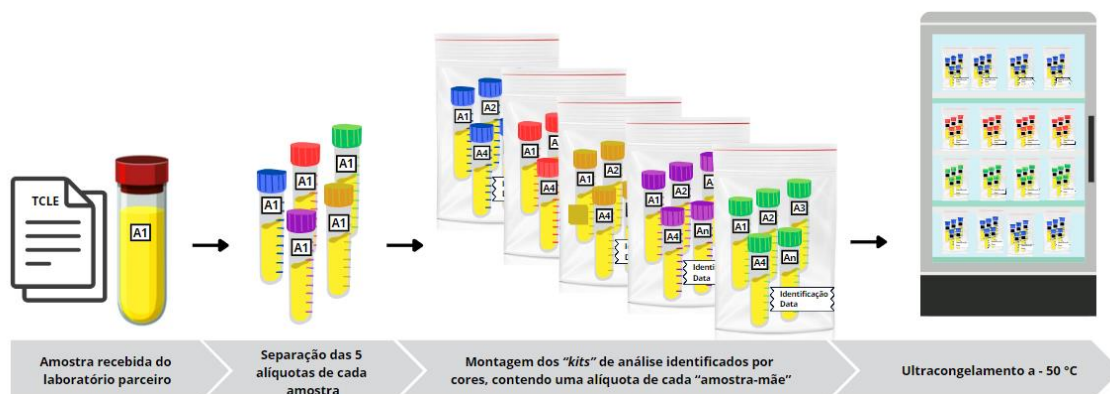
4.2 COLETA E ARMAZENAMENTO DAS AMOSTRAS DE URINA

Cada voluntário coletou o material em tubos de centrífuga de 50 mL armazenando-os em sacos plásticos do tipo *zip lock*. Uma vez entregue o material aos funcionários do laboratório parceiro, a amostra passou por processo de triagem na qual foi adicionada 10 % (v/v) de uma solução contendo acetona e metanol (60:40 %, v/v) para inativação de eventuais particulados virais (WIDERA et al., 2021). Em seguida, as amostras foram congeladas e enviadas para o laboratório do Grupo de Química Analítica e Quimiometria, GQAQ (Departamento de Química, Instituto de Ciências Exatas, UFJF) via transporte específico para amostras biológicas. A temperatura foi monitorada durante todo o processo de coleta, triagem, transporte e armazenamento. Ao todo foram coletadas 686 amostras válidas (considerando todas as entregas com documentação em conformidade com o exigido pelo Comitê de Ética) durante julho de 2021 e janeiro de 2022. Nesse período o Brasil estava enfrentando a segunda onda da pandemia como reação da variante delta e o período de vacinação já havia se iniciado (DUTTA, 2022; WHO, 2023d).

Uma vez recebidas as amostras, de cada uma das “amostras-mãe” homogeneizadas, 5 alíquotas foram separadas e etiquetadas para congelamento em freezer a $-50\text{ }^{\circ}\text{C}$. Dessa forma, foram montados 5 *kits* de análise contendo uma alíquota das 686 “amostras-mãe” coletadas (Figura 2). Cada uma dessas alíquotas foi denominada posteriormente de “amostras-filhas”. O objetivo desse modelo de armazenamento é inicialmente evitar ciclos de descongelamentos, ou seja, evitar que fosse necessário o descongelamento de todas as amostras a todo momento em que análises fossem realizadas. Dessa forma, evitamos não apenas a redução da durabilidade e o comprometimento da integridade das amostras, bem como, com manuseio da amostra apenas uma vez evita-se que estas sejam contaminadas com substâncias indesejadas. As

amostras foram recebidas uma vez por semana durante o período citado e submetidas ao protocolo de armazenamento descrito no dia do recebimento.

FIGURA 2 – Representação ilustrativa do processo inicial de armazenamento das amostras



Fonte: Elaborado pela autora (2024).

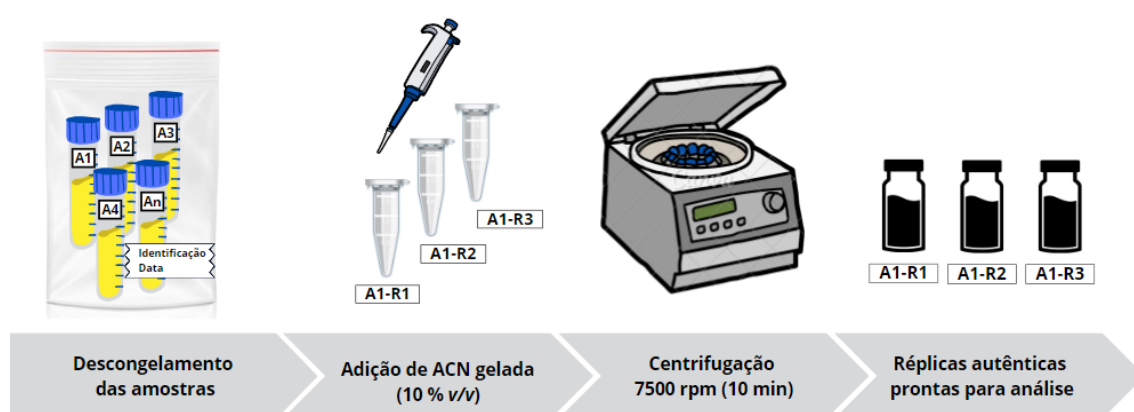
Durante todo o manuseio dessas amostras a equipe envolvida fez o uso de equipamentos de proteção individual (EPIs) contendo avental descartável, protetor facial, máscara, sapato emborrachado hospitalar e luvas descartáveis, considerando a importante infectividade da COVID-19. Todo material residual de qualquer uma dessas e outras etapas posteriores de preparo e análise de amostras foi inutilizado e descartado em *containers* específicos para material biológico infectante, sendo coletado pelo setor de sustentabilidade da UFJF e levado para incineração e/ou descontaminação por autoclave.

4.2.1 Preparo de amostras

As amostras selecionadas para análise foram retiradas do ultrafreezer, descongeladas em temperatura ambiente em cerca de 20 minutos e homogêneas individualmente. Todos os experimentos foram feitos a partir da análise de três alíquotas autênticas extraídas de uma das "amostras-filhas" descongeladas, ou seja, três alíquotas individuais de uma mesma amostra foram submetidas independentemente ao processo completo de preparo (Figura 3). Assim, alíquotas de 2 mL de cada uma das 100 amostras descongeladas foram individualmente transferidas para tubos de centrífuga de 15 mL, na qual foram adicionados 200 µL de acetonitrila gelada para completar o processo de precipitação de proteínas. Em seguida, as réplicas foram centrifugadas a 7500 rpm (44.100 G) por 10 minutos e o sobrenadante transferido para o frasco de análise. À medida

que os processos foram concluídos, todas as amostras foram mantidas em geladeira a 4°C. É recomendado que amostras de urina não devem permanecer mais de 2 horas sem refrigeração, logo, as amostras só foram retiradas da geladeira durante o manuseio, a ser realizado de forma mais eficiente e rápida o possível. Todas as amostras biológicas também foram reservadas em geladeira destinada para tal, a fim de evitar contaminações cruzadas.

FIGURA 3 - Representação ilustrativa do preparo das amostras de urina a partir de réplicas autênticas



Fonte: Elaborado pela autora (2024).

De cada bloco de réplicas (R1; R2; R3) foram retiradas alíquotas de 20 μ L, que foram misturadas em um *pool* para produção da amostra de controle de qualidade analítico (*Quality Control*, QC). Considerando também o preparo de três réplicas autênticas, o que resultou em 3 soluções estoques de QC. O QC foi utilizado durante todos processos de otimização de método e em intervalos de sequências analíticas extensas para verificação de precisão e outros parâmetros associados à qualidade do conjunto de dados. Esse protocolo analítico foi utilizado para análises de urina em: eletroforese capilar com detecção UV (*Capillary Electrophoresis with UV detection*, CE-UV) e detecção por espectrometria de massas (*Capillary Electrophoresis with Mass Spectrometry*, CE-MS), cromatografia líquida de alta eficiência associado à espectrometria de massas (*High Performance Liquid Chromatography with Mass Spectrometry*, HPLC-MS) e espectroscopia vibracional na região do infravermelho.

4.2.2 Especificações do conjunto amostral

Para os experimentos envolvendo análises de urina do projeto COVID-19 foram selecionadas 100 amostras, sendo 38 positivas e 62 negativas para SARS-CoV-2 de acordo com o resultado do RT-PCR. Sendo assim, o grupo teste (GT) é constituído por amostras com resultado positivo para o vírus e o grupo controle (GC) por amostras negativas. O resultado dos testes de RT-PCR e qualquer outra informação relacionada ao paciente só foram acessados após finalização da fase analítica, de modo que os estudos fossem desenvolvidos de acordo com protocolo duplo-cego em que doadores e analistas não sabem o resultado do teste.

As 100 amostras selecionadas representam um grupo diverso de indivíduos na qual a única diferenciação entre grupos teste e controle depende do resultado de RT-PCR para COVID-19. A partir do preenchimento do questionário de anamnese, algumas informações dos pacientes foram obtidas. Em geral, o grupo divide-se entre homens e mulheres de 21 a 77 anos, sintomáticos e assintomáticos, com ou sem fatores de risco para a doença. As informações completas estão dispostas na Tabela 1.

TABELA 1 – Características do conjunto amostral para estudos relacionados à COVID-19

Variáveis	Grupo Teste (n = 38)	Grupo controle (n = 62)
Resultado RT-PCR	Positivo	Negativo
Idade, média ($Q_{\min} - Q_{\max}$)	42 (21 – 72)	41 (21 -77)
Sexo masculino, <i>n</i>	16	38
Sexo feminino, <i>n</i>	22	24
Fumante, <i>n</i> (%)	1 (2.6)	6 (9.6)
Sintomático ^a , <i>n</i> (%)	30 (78.9)	28 (45.6)
Vacinados, <i>n</i> (%)	30 (78.9)	60 (96.7)
Comorbidades ^a , <i>n</i> (%)	10 (26.3)	10 (16.6)
Sintomatologia^a, <i>n</i> (%)^b		
Tosse	21 (70.0)	16 (57.1)
Coriza	19 (63.3)	18 (64.2)
Dor de cabeça	9 (30.0)	16 (57.1)

Dor	14 (46.7)	13 (46.4)
Febre	9 (30.0)	3 (10.7)
Perda de olfato	2 (6.7)	1 (3.5)
Diarréia/náusea	3 (10.0)	3 (10.7)
Irritação na garganta	4 (13.3)	0 (0.0)
Desconforto respiratório	0 (0.0)	1 (3.5)
Comorbidades^a, <i>n</i>		
Hipertensão	5	3
Diabetes	2	2
Asma	0	1
Anomalias respiratórias	1	1
Doenças cardiovasculares	1	0
Outros	1	3

Q_{\min} = valor mínimo Q_{\max} = Valor máximo; ^aDados autodeclarados; ^b Percentual de cada item entre o total de indivíduos que declararam manifestações sintomáticas.

Fonte: Elaborado pela autora (2024).

4.3 AMOSTRAS DE SALIVA E SORO SANGUÍNEO

Algumas amostras de saliva e soro sanguíneo (considerado para estudos adicionais) foram doadas pelo Lemos Laboratórios de Análises Clínicas para que testes piloto voltados para que o desenvolvimento de metodologias analíticas fosse realizado. Essas amostras foram doadas a partir de alíquotas residuais não processadas de exames de rotina do laboratório, sem que nenhuma informação dos pacientes fosse divulgada.

60 amostras de salivas foram cedidas, já processadas para remoção de proteínas a partir de protocolo não divulgado por motivos comerciais. Para análises em eletroforese capilar, nenhum preparo adicional foi realizado, por HPLC-MS as amostras foram submetidas ao mesmo procedimento descrito na seção 4.3.1 com certas adaptações devido ao volume disponível reduzido. Em tubos de centrifuga de 2 mL, alíquotas de 1 mL de cada amostra de saliva foram separadas e adicionadas de 100 μ L de uma solução contendo acetona e metanol (60:40 %, *v/v*) e 100 μ L de acetonitrila gelada. Em seguida, as réplicas foram centrifugadas a 3000 rpm (44.100 G) por 15 minutos e o sobrenadante transferido para o frasco de análise, mantidos em refrigeração a 4 °C até o momento da análise.

Em alguns estudos desenvolvidos ao longo desse projeto, a análise de outros fluidos biológicos se tornou pertinente, desse modo, 10 amostras de soro sanguíneo foram também cedidas sem nenhum preparo além da centrifugação inicial responsável para separar o sangue total em suas partes. Assim, em ambos equipamentos (CE-UV e HPLC-MS) as amostras foram submetidas ao procedimento de preparo adaptado descrito anteriormente.

4.4 AMOSTRAS COM HISTÓRICO DE ARBOVIROSES

Durante o processo de coleta das amostras sob autorização do CEP COVID-19 descrito na seção 4.3, os voluntários preencheram uma ficha de anamnese na qual informações de histórico quanto à infecção por arboviroses foram inseridas pelos mesmos. 64 amostras com histórico de arboviroses declarado foram coletadas (4 Zika, 7 Chikungunya e 53 Dengue). Foram somadas a esse grupo, 38 amostras positivas para COVID-19 e amostras negativas para todas essas doenças, totalizando 205 amostras independentes. Em geral, o grupo divide-se entre homens e mulheres de 19 a 75 anos, com ou sem doenças pré-existentes. As informações completas estão dispostas na Tabela 2.

TABELA 2 - Características do conjunto amostral para estudos relacionados às arboviroses combinadas com COVID-19

Variáveis	ZIKV	CHIKV	DENV	COVID	NEG.
Idade, média	34	34	40	42	39
($Q_{\min} - Q_{\max}$)	(20 – 44)	(27 -66)	(22 – 66)	(22 – 72)	(19 -75)
Sexo masculino, <i>n</i>	2	6	30	16	60
Sexo feminino, <i>n</i>	2	1	23	22	48
Fumante, <i>n</i> (%)	0 (0)	0 (0)	5 (9.4)	1 (2.6)	10 (9.2)
Doenças pré-existentes, <i>n</i>					
Hipertensão	1	0	7	2	12
Diabetes	0	0	3	2	7
Asma	0	0	0	0	0
Anomalias respiratórias	0	3	1	0	3

Doenças cardiovasculares	0	0	1	1	1
Outros	0	0	0	1	3

Fonte: Elaborado pela autora (2024).

Sobre o preparo, em tubos de centrífuga de 2 mL alíquotas de 1 mL de cada amostra foram separadas e adicionadas de 100 µL de acetonitrila gelada para precipitação de proteínas. Em seguida, as amostras foram centrifugadas a 3000 rpm (44.100 G) por 15 minutos e o sobrenadante transferido para o frasco de análise, mantidos em refrigeração a 4 °C até o momento da análise. Essas amostras foram analisadas por espectroscopia vibracional na região do infravermelho.

4.5 MATRIZES SINTÉTICAS

Em alguns estudos envolvendo avaliação de seletividade e efeitos de matriz, algumas matrizes artificiais de urina e soro sanguíneo foram preparadas a partir da adaptação de método descrito por (GODOY; EBERLIN; SIMIONATO, 2020). Para urina, foi preparada uma solução aquosa contendo 0,75 g de NaCl, 1,82 g de KCl, 1,8 g de ureia, 0,48 g de Na₂HPO₄ e 5 mg de albumina. Usualmente é também incluída creatinina na solução, porém, como é um dos analitos investigados ao longo do trabalho, não foi adicionada. Para o soro sanguíneo, uma solução salina tamponada com fosfato foi preparada misturando 0,34 g de KH₂PO₄ e 0,35 g Na₂HPO₄ em 25 mL de soro fisiológico e albumina a 2% (*m/v*). A solução foi ajustada em pH 7,4 com NaOH 1 mol L⁻¹. Depois de prontas, amostras das matrizes artificiais foram preparadas para análise conforme descrito nas seções 4.3.1 e 4.5, respectivamente.

5 METABOLÔMICA ALVO POR ELETROFORESE CAPILAR

Conforme comentado nas seções introdutórias, o metabolismo humano é o total de reações bioquímicas que acontecem nas células a fim de obter energia para que aquele organismo se mantenha vivo. Todo ser vivo aproveita nutrientes e moléculas como carboidratos, proteínas e gorduras para sobreviver. Estas e outras moléculas participam então de rotas metabólicas que apesar de serem influenciadas por idade, gênero, dieta, hábitos, fatores ambientais, infecções, doenças, e outros fatores, os produtos intermediários e finais dessas rotas metabólicas podem nos indicar como esse organismo está reagindo frente a perturbações externas e internas, por exemplo. É comum que quando infectado por agentes patogênicos, o organismo reorganize rotas metabólicas em detrimento de agir no combate a essa infecção como uma resposta imunológica, conseqüentemente, o estudo desse rearranjo metabólico pode levar às respostas de interesse clínico. Dentro desses produtos, aminoácidos e ácidos orgânicos são grupos que compreendem diversas moléculas essenciais que se destacam entre potenciais classes de biomarcadores (DA POIAN; EL-BACHA; LUZ, 2010; NAVA; RAJA, 2022). A urina inclusive, escolhida prioritariamente neste projeto, sendo uma solução aquosa resultante da filtração do sangue nos rins, é uma matriz na qual grande quantidade e variedade desses compostos hidrossolúveis estão contidos, o que favorece o estudo de tais metabólitos.

A eletroforese capilar é uma técnica de separação que opera em meio líquido na qual a separação dos analitos se dá por diferenças na migração dos compostos mediante aplicação de voltagem. A técnica oferece uma perspectiva interessante frente às cromatografias clássicas não associadas a meios de detecção universal no que diz respeito à seletividade, parâmetro importante ao se tratar de amostras biológicas quando o interesse é a realização de análises segmentadas de grupo de moléculas pré-selecionadas. Em suma, a eletroforese capilar mesmo associada à detecção ultravioleta permite que métodos sejam otimizados de forma que apenas um grupo de moléculas específicas sejam registrados nos dados de saída. Neste sentido, alguns métodos analíticos envolvendo a análise *targeted* de ácidos orgânicos, aminoácidos e compostos majoritários da urina foram desenvolvidos utilizando a CZE-UV a fim de obter informações qualitativas e quantitativas, quando possível, das amostras de fluidos biológicos. De modo complementar, a associação de eletroforese capilar com a espectrometria de massas

sequencial (*tandem mass spectrometry*, MS/MS) a partir de um equipamento triplo quadrupolo também foi explorada para testes quantitativos de metabólitos.

5.1 ÁCIDOS ORGÂNICOS

Ácidos orgânicos são conhecidas estruturas que se apresentam no corpo humano como produtos intermediários e finais de rotas metabólicas diversas. Em geral, estruturas com cadeias carbônicas curtas com um, dois ou mais grupos carboxila, alguns hidroxiácidos ou outras ramificações também estão comumente contidas. Estas espécies são os principais metabólitos urinários que derivam do processo oxidativo do metabolismo dos ácidos graxos e geralmente estão associados à fisiologia do jejum ou a patologias diretas como deficiência de enzimas catalisadoras de ácidos di-, hidroxi-carboxílicos e acidurias (JIN; HOPPEL; TSERNGS, 1992; TSERNG; JIN, 1991; VAIDYANATHAN; NARAYANAN; VASUDEVAN, 2011). Porém, de acordo com a literatura médica, quando um corpo humano é infectado, há um padrão de alterações no organismo devido à necessidade de combater a infecção que induz uma resposta de fase aguda levando a uma influência no metabolismo lipídico, como consequência, alterando o perfil de ácidos orgânicos gerados como metabólitos finais (BEISEL, 1975; KHOVIDHUNKIT et al., 2004). Dessa forma, por serem importantes biomarcadores, exames quantitativos de ácidos orgânicos específicos já vêm sendo oferecidos em laboratórios de análises clínicas especializadas.

Na intenção de obter um perfil simplificado de ácidos orgânicos em amostras biológicas, dois métodos analíticos foram desenvolvidos e otimizados de acordo com o fluido biológico de interesse. O primeiro método voltado à análise de ácidos dicarboxílicos, como succínico, tartárico, málico, fumárico, malônico e alguns hidroxiácidos como os ácidos láctico e glicólico, por exemplo. Um segundo método foi proposto para avaliação de ácidos monocarboxílicos de cadeia curta (*Short-chain Fatty Acids*, SCFA) e semelhantes. Essa distinção dos métodos é devida a algumas diferenças estruturais e físico-químicas dos analitos que faz com que cada um desses grupos seja favorecido por uma faixa de pH distinta e por componentes de eletrólitos específicos.

Todo o processo de desenvolvimento e otimização de método por CZE-UV considerando análise de ácidos orgânicos e as demais realizadas neste trabalho, foi baseado em referências pertinentes e testes experimentais associados com simulações de

resultados realizadas no software Peakmaster®, especializado em predizer resultados gerados por sistemas de eletromigração.

5.1.1 Reagentes e soluções

Todas as soluções foram preparadas em água deionizada por sistema de Osmose Reversa da Quimis (São Paulo, SP). Substâncias de referência, tais como ácidos málico, tartárico, succínico e DL-láctico, fórmico, acético, propiônico, butírico e valérico foram adquiridos da Sigma Aldrich (Saint Louis, EUA) com certificação de pureza. Demais materiais utilizados: Hidróxido de sódio (NaOH) da Synth (São Paulo, SP), Tris-(hidroximetil)-aminometano (TRIS), brometo de cetiltrimetilamônio (CTAB), ácido 3,5-dinitrobenzóico da Sigma Aldrich (Saint Louis, EUA), ácido ftálico e ácido clorídrico (HCl) da Vetec (Rio de Janeiro, RJ). Soluções-estoque das substâncias de referência foram preparadas individualmente a 10 mmol L^{-1} em água deionizada e armazenadas em refrigeração a $4 \text{ }^{\circ}\text{C}$ até análise. Soluções-estoque do tampão e de outros componentes foram preparadas individualmente e diluídas em água deionizada para o preparo do eletrólito de corrida (*Background Electrolyte*, BGE), também armazenado em refrigeração até o uso.

5.1.2 Instrumentação e métodos

Os experimentos por eletroforese capilar foram realizados em equipamentos modelos 7100 da Agilent Technologies (Palo alto, USA) com detector de arranjo de diodos (*Diode Array Detector*, DAD) e controlado por softwares OpenLab ChemStation C.01.07 e ChemStation B.04.03. Parâmetros otimizados específicos de cada método estão descritos na Tabela 3. Capilar de sílica fundida revestido com poliimida (Série TSP) (comprimento total 68.5 cm, $75 \text{ }\mu\text{m}$ de diâmetro interno e $365 \text{ }\mu\text{m}$ de diâmetro externo), injeção hidrodinâmica e condicionamento são parâmetros comuns entre os métodos. O condicionamento inicial foi feito a partir da injeção contínua (*flush*) de NaOH 1 mol L^{-1} a aproximadamente 940 mbar por 40 minutos, seguido por água deionizada por 20 minutos e BGE por 20 minutos. Entre corridas, o capilar foi recondicionado com NaOH 1 mol L^{-1} (60 s), água deionizada (30 s), HCl 1 mol L^{-1} (30 s), segunda rodada de água deionizada (30 s) e eletrólito (120 s).

TABELA 3 - Parâmetros instrumentais dos métodos analíticos para ácidos orgânicos desenvolvidos por CZE-UV

Parâmetros otimizados	MÉTODO A	MÉTODO B
	Dicarboxílicos e hidroxiácidos	Monocarboxílicos
Eletrólito	10 mmol L ⁻¹ Tris, 11 mmol L ⁻¹ ácido ftálico e 0.5 mmol L ⁻¹ CTAB (pH 3.9).	15 mmol L ⁻¹ Tris, 12 mmol L ⁻¹ ácido 3,5-dinitrobenzóico e 0.5 mmol L ⁻¹ CTAB (pH 7.5).
Temperatura (°C)	25 °C	30 °C
Comprimento de onda	240 nm	240 nm
Comprimento efetivo do capilar	60 cm	60 cm
Voltagem	- 10 kV	- 20 kV
Corrente gerada	5 uA	11 uA
Tempo de corrida	18 min	10 min
Programa de injeção	25 mbar x 2 s	20 mbar x 2 s

Fonte: Elaborado pela autora (2024)

5.1.3 Construção de curvas de mobilidade.

Curvas de mobilidade efetiva (μ_{eff}) vs pH dos analitos foram construídas a partir do somatório do produto da distribuição iônica (α) desse analito associado à escala de pH, com a sua respectiva mobilidade eletroforética (μ_j), de acordo com a equação (1):

$$(\mu_{eff})_i = \sum(\alpha_i \mu_j) \quad (1)$$

A fração de distribuição foi calculada a partir das equações de referência (2) e (3) com os valores da constante de ionização de cada espécie e valores de pH seguindo a escala tradicional de 1 a 14 em incrementos de 0.5. Para espécies com apenas um hidrogênio ionizável, temos:

$$\alpha_0 = \frac{[H_3O^+]}{[H_3O^+] + Ka} \quad (2)$$

$$\alpha_1 = \frac{[H_3O^+]}{[H_3O^+] + Ka} \quad (3)$$

Para cálculos relacionados a ácidos e/ou base polipróticos, devidas alterações nas equações (2) e (3) foram feitas, considerando o uso dos valores das demais constantes de ionização da espécie. Conforme descrito abaixo, para espécies com dois hidrogênios ionizáveis, temos:

$$\alpha_0 = \frac{[H_3O^+]^2}{[H_3O^+]^2 + [H_3O^+]Ka_1 + Ka_1Ka_2} \quad (4)$$

$$\alpha_1 = \frac{[H_3O^+]Ka_1}{[H_3O^+]^2 + [H_3O^+]Ka_1 + Ka_1Ka_2} \quad (5)$$

$$\alpha_2 = \frac{Ka_1Ka_2}{[H_3O^+]^2 + [H_3O^+]Ka_1 + Ka_1Ka_2} \quad (6)$$

Para espécie com três hidrogênios ionizáveis, temos:

$$\alpha_0 = \frac{[H_3O^+]^3}{[H_3O^+]^3 + [H_3O^+]^2Ka_1 + [H_3O^+]Ka_1Ka_2 + Ka_1Ka_2Ka_3} \quad (7)$$

$$\alpha_1 = \frac{[H_3O^+]^2Ka_1}{[H_3O^+]^3 + [H_3O^+]^2Ka_1 + [H_3O^+]Ka_1Ka_2 + Ka_1Ka_2Ka_3} \quad (8)$$

$$\alpha_2 = \frac{[H_3O^+]Ka_1Ka_2}{[H_3O^+]^3 + [H_3O^+]^2Ka_1 + [H_3O^+]Ka_1Ka_2 + Ka_1Ka_2Ka_3} \quad (9)$$

$$\alpha_3 = \frac{Ka_1Ka_2Ka_3}{[H_3O^+]^3 + [H_3O^+]^2Ka_1 + [H_3O^+]Ka_1Ka_2 + Ka_1Ka_2Ka_3} \quad (10)$$

Valores de pKa e de mobilidade eletroforética de cada espécie foram consultados no banco de dados gratuito do software Peakmaster®. Curvas de mobilidades foram construídas usando o software Microsoft Excel®

5.1.4 Análise de variância e verificação de pressupostos estatísticos.

Os valores de falta de ajuste e significância da regressão foram calculados a partir das equações de análise de variância (ANOVA, *Analysis of variance*) descritas na Tabela 4) na qual: n_i = o número de repetições no nível i ; m = quantidade de níveis independentes; p = número de parâmetros do modelo; \hat{y}_i = valores previstos; \bar{y} = média dos valores observados; y_{ij} = valores observados; \bar{y}_i = média dos valores observados em cada nível; GL = graus de liberdade. Em seguida os testes (equações 11 e 12) foram aplicados para obtenção dos valores de F_{calc} que são posteriormente comparados com os valores tabelados da distribuição F a um intervalo de confiança de 95 %. Testes de normalidade por Shapiro-Wilk e de homoscedasticidade por Bartlett foram calculados, com 95 % de

confiança, a partir das funções “*bartlett.test*” e “*shapiro.test*” inseridos em pacotes base do software RStudio® com linguagem de código “R”.

TABELA 4 - Equações para cálculo de falta de ajuste e significância da regressão

Fonte de variação	Soma Quadrática	GL	Média quadrática
Regressão	$SQ_R = \sum_i^m \sum_j^{n_i} (\hat{y}_i - \bar{y})^2$	$p - 1$	$MQ_R = \frac{SQ_R}{p - 1}$
Resíduos	$SQ_r = \sum_i^m \sum_j^{n_i} (y_{ij} - \hat{y}_i)^2$	$n - p$	$MQ_r = \frac{SQ_r}{n - p}$
Falta de ajuste	$SQ_{faj} = \sum_i^m \sum_j^{n_i} (\hat{y}_i - \bar{y}_i)^2$	$m - p$	$MQ_{faj} = \frac{SQ_{faj}}{m - p}$
Erro puro	$SQ_{ep} = \sum_i^m \sum_j^{n_i} (y_{ij} - \bar{y}_i)^2$	$n - m$	$MQ_{ep} = \frac{SQ_{ep}}{n - m}$

Fonte: (NETO; SCARMINIO; BRUNS, 2010)

$$\text{Teste da falta de ajuste: } F_{cal} = \frac{MQ_{faj}}{MQ_{ep}} \quad (11)$$

$$\text{Teste da significância de regressão: } F_{cal} = \frac{MQ_R}{MQ_r} \quad (12)$$

Valores de precisão analíticas foram obtidos a partir do cálculo do desvio padrão relativo (*Relative Standard Deviation*, RSD) de acordo com a equação 13:

$$RSD (\%) = \frac{\text{Desvio Padrão}}{\text{Média}} \times 100 \quad (13)$$

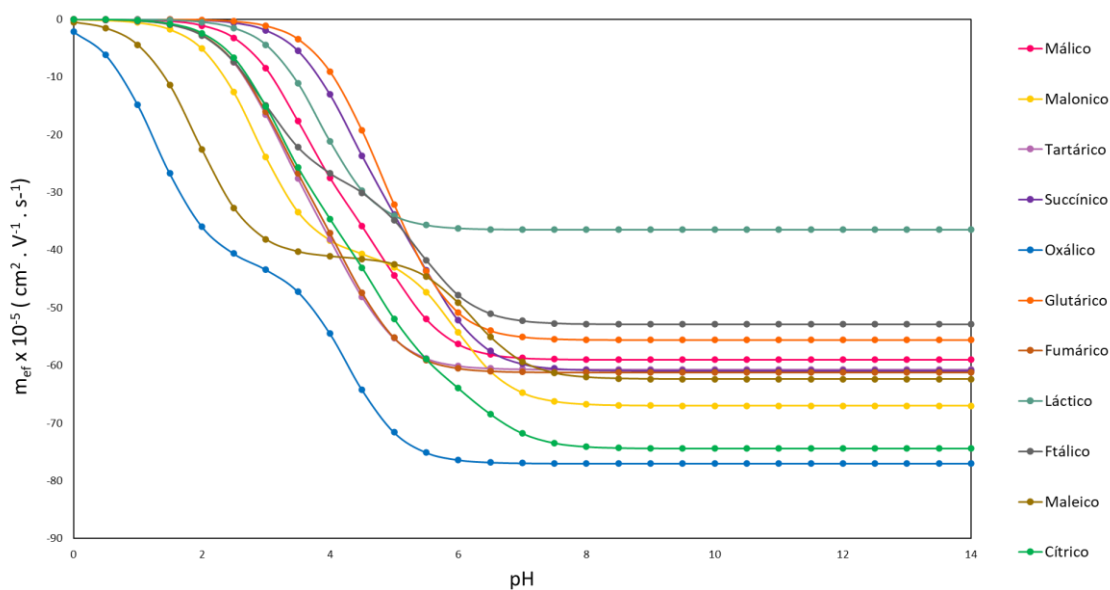
5.1.5. Perfil de ácidos orgânicos dicarboxílicos e hidroxiácidos em urina

O ponto de partida para o protocolo focado nos ácidos orgânicos dicarboxílicos foi um método publicado por (VAZ et al., 2012), que considera um eletrólito composto por TRIS, ácido ftálico e CTAB. Durante o processo de otimização, foram executados

experimentos submetendo o QC de urina e uma mistura de padrões contendo os ácidos tartárico, málico, succínico e láctico, que por apresentarem mobilidades próximas foram selecionados como um representativo do grupo dos possíveis analitos presentes.

Conforme comentado nas seções anteriores, a urina é um fluido corporal rico em compostos químicos livres, orgânicos e inorgânicos, possivelmente sendo o fluido com maior número de metabólitos. Portanto, não se espera uma separação física dos componentes suficiente para registrá-los no eletroferograma com boa resolução, pelo menos não quando o uso de detecção por ultravioleta está sendo considerado. A curva de mobilidade foi construída contendo alguns dos possíveis metabólitos cujo método por CZE-UV favorece a migração (Figura 4).

FIGURA 4 - Curva de mobilidade efetiva vs pH exemplificando o comportamento de alguns ácidos orgânicos dicarboxílicos e hidroxiácidos



Fonte: Elaborado pela autora (2024).

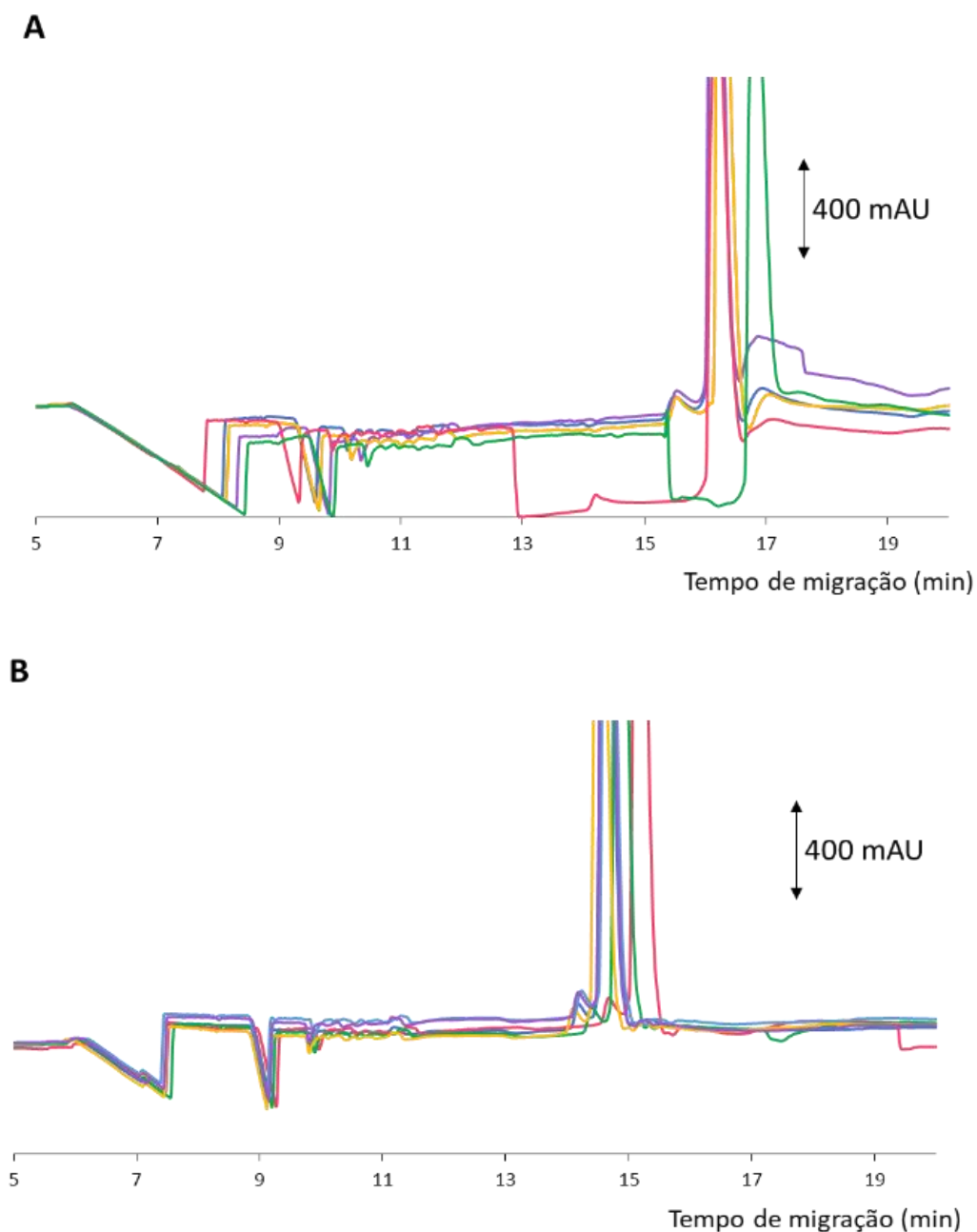
O primeiro parâmetro a ser otimizado foi o pH de trabalho. Seguindo a curva de mobilidade, a maior parte dos ácidos orgânicos adotados como referência começam a dissociação do hidrogênio da carboxila a partir de pH 3, aproximadamente. Após alguns testes iniciais envolvendo experimentos e simulações no Peakmaster®, o pH 3.9 foi ajustado a partir do tampão ácido ftálico/ftalato com TRISH^+ atuando como contra-íon. Em meios mais ácidos, as mobilidades iônicas dos analitos são menores. Por outro lado,

em meio acima de pH 4.5 aproximadamente, picos do sistema começam a ser registrados, também sendo mostrados em simulações.

Uma vez que os ácidos de interesse apresentam baixa absorvibilidade na região do ultravioleta, há necessidade de um cromóforo para mediar a detecção de forma indireta. Nesse caso, o ácido ftálico também atua como agente cromóforo, visto que tem mobilidade próxima aos analitos. Por fim, sabendo que os analitos estão na forma de ânions, optou-se por utilizar CTAB como inversor de fluxo com o sistema operando em polaridade invertida. Quando operado em polaridade normal, uma voltagem positiva é aplicada no lado da injeção da amostra na qual os cátions migram pelo capilar, uma vez que as paredes encobertas por grupos silanóis promovem o fluxo eletroosmótico (*electroosmotic flow*, EOF), responsável por conduzir os solutos em direção ao detector, polo negativo. Em caso de análise de ânions rápidos, todo o sistema deve ser invertido. O surfactante adicionado ao BGE interage com os grupos silanóis e demais componentes do BGE formando uma nova camada na parede do capilar que inverte a direção do EOF para o lado carregado positivamente. Assim, ao aplicar uma voltagem negativa, os ânions são carregados pelo EOF invertido em direção ao detector, agora polo positivo (TAVARES, 1997).

Demais condições (voltagem, parâmetros de injeção, temperatura, tamanho de capilar e comprimento de onda) foram também ajustadas a partir de testes experimentais. Concluindo a etapa de ajuste do método, foi necessária a otimização do condicionamento entre corridas a fim de melhorar a reprodutibilidade dos resultados (Figura 5). A limpeza apenas com eletrólito, ou soda e/ou água seguida de eletrólito não foi suficiente para manter o sistema acondicionado de forma constante. O uso de solventes orgânicos também foi testado, sem sucesso. A melhor opção foi o uso de ácido clorídrico a 1 mol L⁻¹ entre injeções de água, junto com aumento de tempo de *flush* do eletrólito ao final. De acordo com (DO NASCIMENTO et al., 2022), o uso de HCl previne a precipitação de compostos na parede do capilar, importante no caso de matrizes complexas como os fluidos biológicos. Desse modo, o condicionamento otimizado foi reproduzido em todos os métodos aplicados às amostras biológicas por CZE-UV descrito nas seções seguintes.

FIGURA 5 - Eletroferogramas de QC de urina (A) com condicionamento convencional: NaOH 1 mol L⁻¹ (60 s), água (60 s) e BGE (60 s) e (B) com condicionamento otimizado: NaOH 1 mol L⁻¹ (60 s), água (30 s), HCl 1 mol L⁻¹ (30 s), água (30 s) e BGE (120 s).

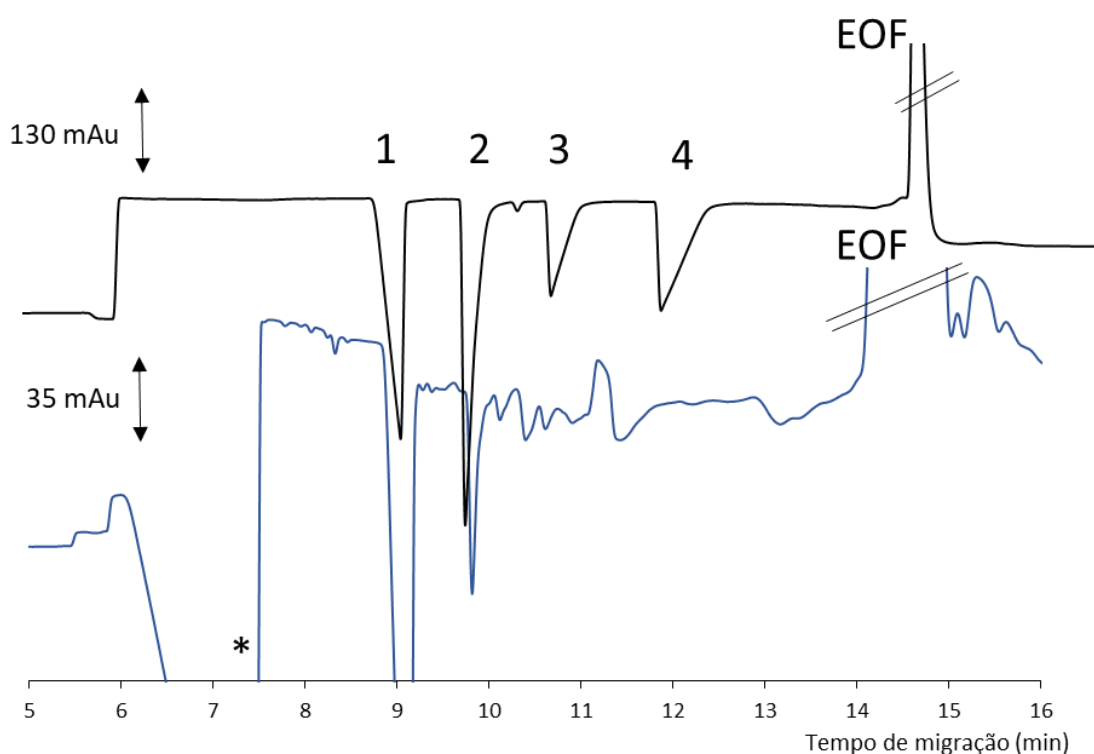


Fonte: Elaborado pela autora (2024).

Conforme esperado, os padrões utilizados como referência na otimização de métodos apresentaram ordem de migração condizente com o exposto anteriormente na

curva de mobilidade. A urina é uma solução que pode conter até milhares de compostos, o que justifica o perfil de eletroferograma obtido pela amostra controle que contém diversos ácidos orgânicos carboxílicos, hidroxiácidos e outras espécies de estruturas semelhantes entre o intervalo de migração de 8 a 12 minutos (Figura 6). O sinal de maior intensidade em 7 minutos é resultante da grande quantidade de cloreto e outros ânions mais rápidos, não necessariamente inorgânicos, visto que a mobilidade do oxalato, por exemplo, também se encontra nessa região.

FIGURA 6 - Eletroferogramas resultantes da análise da mistura de padrões (em preto), QC de urina (em azul). (1) Tartarato; (2) Malato; (3) Lactato; (4) Succinato; *Cloreto e outros ânions rápidos.



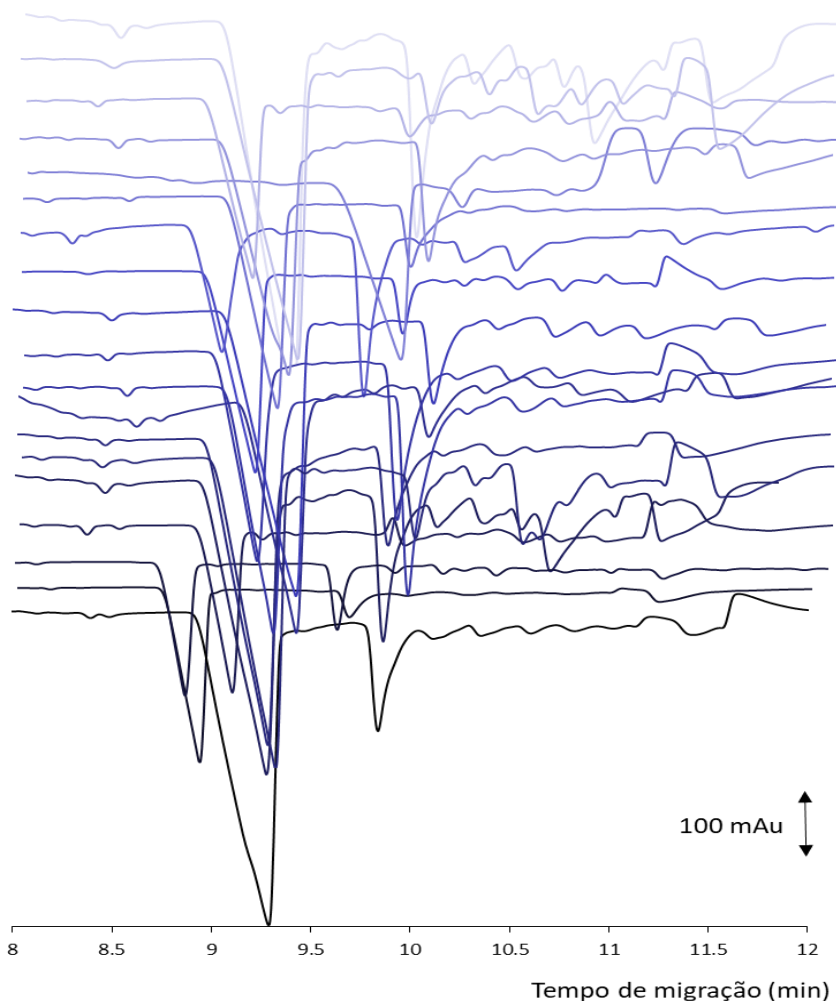
Fonte: Elaborado pela autora (2024).

O método otimizado foi aplicado às 100 amostras de urina pré-selecionadas, considerando análise de três réplicas autênticas de cada amostra. As sequências analíticas foram organizadas de forma que cada grupo de 100 amostras diferentes fossem analisadas de forma randômica intercaladas com: (a) curvas de calibração externa de quatro padrões

de ácidos orgânicos selecionados, a saber, tartárico, málico, succínico e láctico em cinco níveis de concentração (0.5; 2.0; 5.0; 7.0; 10.0 mmol L⁻¹); (b) QC; (c) água como branco.

Um bloco completo de análises foi finalizado em cerca de 80 horas, sendo que a cada 24 horas o capilar foi recondicionado conforme descrito na seção 5.1.2. Ao intervalo de cada batelada de 100 amostras, os eletrodos de platina e demais peças removíveis foram higienizadas com isopropanol em banho ultrassônico por 15 minutos, sendo novamente acopladas no equipamento e alinhadas. Esse procedimento de limpeza é feito para evitar que acúmulos de NaOH e outros reagentes que possam precipitar nas superfícies do equipamento comprometam a reprodutibilidade da análise. O processo analítico completo, considerando as 300 alíquotas e as corridas complementares, levou aproximadamente 10 dias. Como resultado, foram registrados eletroferogramas com diferentes perfis de ácidos orgânicos, exemplificado na Figura 7.

FIGURA 7 - Exemplo de eletroferogramas resultantes da análise de ácidos orgânicos dicarboxílicos e hidroxíácidos de 20 urinas diferentes



Fonte: Elaborado pela autora (2024).

Para avaliação de performance do método e verificação da reprodutibilidade dos dados brutos, curvas de calibração externa foram construídas a partir do método dos mínimos quadrados ordinários (*Ordinary Least Squares*, OLS). A linearidade foi verificada por coeficiente de correlação (r) e seguindo os preceitos estatísticos de aleatoriedade, falta de ajuste, independência e homocedasticidade. Os resultados (Tabela 5) são apresentados como valor médio de cada grupo de réplicas com seus respectivos desvios.

TABELA 5 - Regressão e parâmetros estatísticos utilizados para verificação da qualidade de dados

Analito	Coeficiente		<i>r</i>	Falta de ajuste	Significância da regressão
	linear	angular			
Tartarato	3.329 _(14,9)	57.16 _(2,21)	0.9867	Fcal = 1.145 < Ftab = 3.287	Fcal = 665.6 > Ftab = 4.413
	-14.28 _(21,6)	59.75 _(3,21)	0.9748	Fcal = 0.483 < Ftab = 3.287	Fcal = 344.5 > Ftab = 4.413
	-19.83 _(13,8)	64.75 _(2,06)	0.9910	Fcal = 1.021 < Ftab = 3.287	Fcal = 991.1 > Ftab = 4.413
Malato	6.444 _(17,04)	56.82 _(2,39)	0.9844	Fcal = 1.700 < Ftab = 3.287	Fcal = 564.2 > Ftab = 4.413
	-17.09 _(22,2)	59.78 _(3,12)	0.9764	Fcal = 0.551 < Ftab = 3.287	Fcal = 367.6 > Ftab = 4.413
	-16.46 _(12,8)	62.71 _(2,06)	0.9923	Fcal = 1.342 < Ftab = 3.287	Fcal = 1216 > Ftab = 4.413
Lactato	6.543 _(8,09)	26.82 _(1,34)	0.9788	Fcal = 3.171 < Ftab = 3.287	Fcal = 370.4 > Ftab = 4.413
	-3.482 _(11,4)	29.63 _(1,90)	0.9648	Fcal = 0.445 < Ftab = 3.287	Fcal = 238.2 > Ftab = 4.413
	0.912 _(6,02)	29.96 _(0,99)	0.9901	Fcal = 2.381 < Ftab = 3.287	Fcal = 754.6 > Ftab = 4.413
Succinato	4.133 _(12,1)	45.7 _(1,93)	0.9843	Fcal = 2.002 < Ftab = 3.287	Fcal = 560.2 > Ftab = 4.413
	-12.04 _(16,4)	47.82 _(2,61)	0.9742	Fcal = 0.582 < Ftab = 3.287	Fcal = 335.4 > Ftab = 4.413
	-14.48 _(11,6)	51.64 _(1,84)	0.9887	Fcal = 1.074 < Ftab = 3.287	Fcal = 783.8 > Ftab = 4.413

Falta de ajuste: Ftab = 3.287 ($v_1 = 3, v_2 = 15$); Significância da Regressão: Ftab = 4.413 ($v_1 = 1, v_2 = 18$).

Níveis de concentração ajustados: Tartarato (0.56; 2.25; 5.63; 7.88; 11.26) mmol L⁻¹; Malato (0.59; 2.38; 5.96; 8.35; 11.93) mmol L⁻¹; Lactato (0.50; 2.01; 5.04; 7.06; 11.08) mmol L⁻¹; Succinato (0.52; 2.10; 5.25; 7.35; 10.5) mmol L⁻¹.

Fonte: Elaborado pela autora (2024).

Todas as curvas de calibração cumpriram os requisitos de linearidade uma vez que os coeficientes de correlação são próximos a 1, os valores calculados de falta de ajuste estão abaixo do F_{tab} , bem como os valores para significância de regressão são maiores que o F_{tab} . Todos os dados são homocedásticos e estão dentro da normalidade. Uma vez que as análises correspondentes a esses dados foram obtidas de forma randômico, intercaladas com amostras e em intervalos de tempo extenso, os resultados sugerem que o sistema é constante na obtenção de dados, não há sinais de contaminação entre corridas, e é reprodutível, conferindo uma confiabilidade aos dados das amostras.

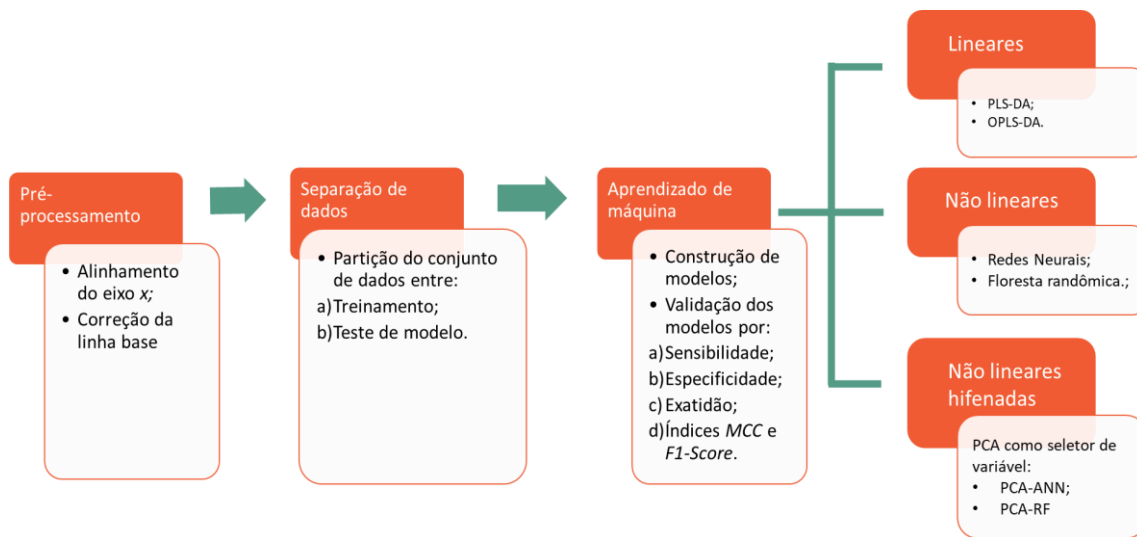
Adicionalmente, a precisão analítica foi determinada a partir de dados dos eletroferogramas de QC na qual foi quantificado o maior sinal registrado em termos de tartarato, obtendo-se os seguintes valores: $(12.78 \pm 0.51; 12.71 \pm 0.99; 12.48 \pm 0.52)$ mmol L⁻¹ com RSD de (4.01; 7.81; 4.20) % respectivamente, considerando cada um dos três blocos de análise. O valor global destas análises resultou em uma precisão média de

$12.66 \pm 0.15 \text{ mmol L}^{-1}$ de tartarato com 1.22 % de RSD. Para esses cálculos foram considerados 36 resultados de QC (12 de cada lote de 100 amostras).

Os resultados gerados registram um perfil de ácidos orgânicos de características específicas com diferentes nuances no perfil de cada amostra, havendo uma possibilidade de ser um grupo de marcadores capazes de diferenciar amostras contaminadas ou não, o que os torna um conjunto de dados promissor para criação de modelos classificatórios. Neste contexto, esse conjunto de dados foi inicialmente agrupado e processado utilizando ferramentas de calibração multivariada e aprendizado de máquina (*machine learning*), a dos eletroferograma em que está contida as informações de ácidos orgânicos exclusivamente (Figura 7), como uma impressão digital bioquímica. Diversos algoritmos de predição matemática, pré-processamento e associações foram testadas, discutidas a seguir. A etapa do trabalho que envolve rotinas estatísticas aprofundadas foi realizada em colaboração com outros estudantes do GQAQ.

Os eletroferogramas brutos foram extraídos individualmente em formato ".csv" usando o software Chemstation (Agilent Technologies) e agrupados posteriormente. Foi utilizado então como base de dados para de entrada para início do processamento quimiométrico uma planilha única contendo os valores de intensidade de sinal em unidades arbitrárias de cada amostra, na qual cada coluna representa uma amostra associada ao código de identificação e ao resultado de RT-PCR para COVID-19. As 100 amostras analisadas em triplicatas autênticas totalizaram um conjunto de 300 dados processados posteriormente como dados independentes, já com a intenção de fazer uma avaliação de performance dos modelos classificatórios construídos. O fluxograma completo do processo de construção de modelos de classificações entre grupos teste e controle está representado na Figura 8.

FIGURA 8 – Esquema representativo das etapas referentes ao processamento de dados obtidos por CZE-UV a partir de ferramentas de aprendizado de máquina.

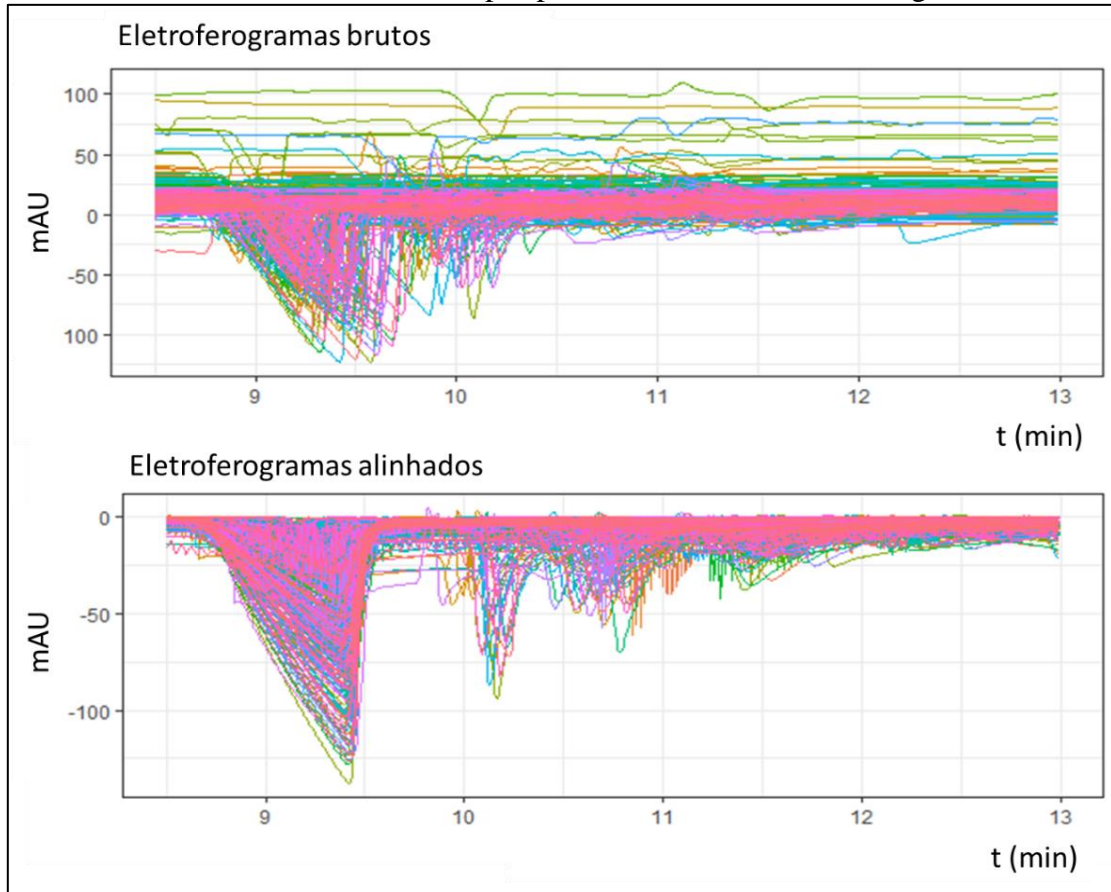


Fonte: Elaborado pela autora (2024).

Como pré-processamento, os eletroferogramas brutos foram alinhados em relação ao tempo de migração e a correção de linha base em referência a um eletroferograma médio utilizando o pacote “*ptw*” (Figura 9). Após pré-processamento, 4 outliers (eletroferogramas com perfil não condizente aos demais) foram desconsiderados (Figura 10), resultando em um conjunto de dados de 296 entradas trabalhados como amostras independentes.

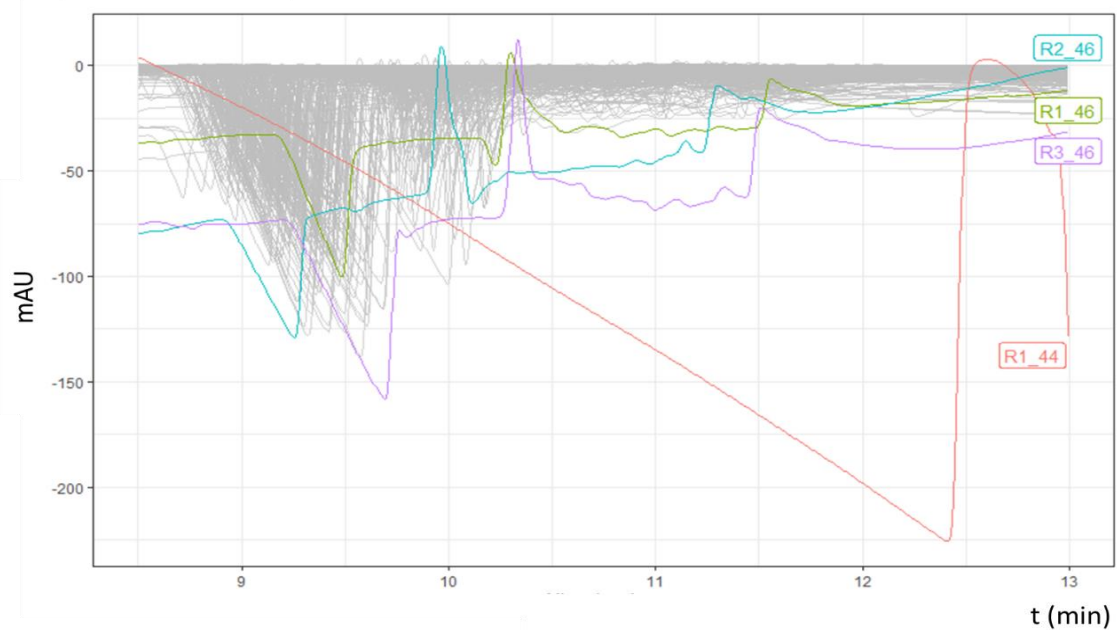
Em seguida, os dados pré-processados foram divididos em amostras-treino (80 %) e amostras-teste (20 %) para serem usados durante a modelagem dos dados e no teste desse modelo, respectivamente. Essa divisão foi feita aleatoriamente considerando todo o conjunto de amostras independentes, resultando em um bloco de amostras-treino contendo 237 dados de entrada (84 GT e 153 GC) e um bloco de amostras-teste de 59 dados (21 GT e 38 GC).

FIGURA 9 – Resultados do pré-processamento dos eletroferogramas.



Fonte: Elaborado pela autora (2024).

FIGURA 10 – *Outliers*



Fonte: Elaborado pela autora (2024).

Todos os modelos discutidos a seguir foram validados a partir da avaliação dos seguintes parâmetros: exatidão, especificidade, sensibilidade, MCC (*Matthew's Correlation Coefficient*) e F1-Score. A exatidão é a taxa de acerto do modelo, ou seja, o percentual das amostras classificadas corretamente. A sensibilidade representa a taxa de verdadeiros positivos e reflete a capacidade do modelo de identificar precisamente os casos com a condição alvo, sendo expressado pelo percentual de acertos dentro das amostras classificadas como positivas. Analogamente, a especificidade representa a taxa de verdadeiros negativos, medindo a capacidade do modelo de identificar corretamente os casos sem a condição alvo. O valor é expressado em termos do percentual de acertos dentro das amostras classificadas como negativas. Todos esses valores são calculados a partir das respostas da matriz de confusão resultante do teste do modelo aplicado aos 20 % do total de amostras pré-selecionadas randomicamente para tal. A matriz de confusões corresponde à tabela que contém o quantitativo de amostras classificadas como positivas e negativas comparado ao número de positivos e negativos reais.

Uma vez que as amostras desses estudos são classificadas como negativas e positivas para COVID-19, enfatiza-se fortemente que o algoritmo dentro da linguagem de programação *R* considera, no caso deste trabalho, não tendo o vírus (amostras atribuídas como “negativas” nos dados de entrada de acordo com o RT-PCR) como a classe positiva no resultado de saída usado para cálculos de desempenho supracitados. Esta especificação é verdadeira para todas as discussões estatísticas posteriores. Portanto, para evitar qualquer má interpretação, a sensibilidade reflete a verdadeira taxa de positivos (como sempre será), parâmetro calculado pelo algoritmo utilizando os dados fornecidos como “classe positiva”, ou no nosso caso, os valores negativos na matriz de confusão. Da mesma forma que, aquelas atribuídas como “positivas” nos dados de entrada (RT-PCR (+)) foram consideradas pelo algoritmo como a classe negativa, o que significa que este grupo que não possui a condição negativa considerada pela linguagem de programação *R*. Dito isto, a especificidade é o parâmetro calculado pelo algoritmo usando os dados fornecidos como “classe negativa”, ou no nosso caso, os valores positivos na matriz de confusão. Por fim, em resumo, os percentuais de precisão, sensibilidade e especificidade da forma que estão sendo expostos ao longo do texto estão representando a qualidade dos dados em relação a classificar positivos e negativos em relação aos testes de RT-PCR.

O MCC, ou coeficiente ϕ , é uma medida da qualidade de modelos de classificações binários que expressa o quão bem o método classifica as amostras quando comparado com resultados completamente aleatórios. É calculado a partir de uma equação única que utiliza todos os valores resultantes da matriz de confusão. Essa métrica varia de (-1) a (+1). (-1) indica que não há qualquer correlação entre os resultados, enquanto (0) indica que o método não é melhor do que uma classificação aleatória e (+1) significa que os recursos e os resultados finais estão completamente correlacionados, ou seja, quanto mais próximo de (+1), melhor. O *F1-Score* também é uma medida da qualidade de modelos binários, porém considera o poder de predição do método, cujo valor é obtido a partir de uma única equação que considera todos os valores da matriz de confusão classificados corretamente.

Métodos de regressão linear como a análise discriminante por mínimos quadrados parciais (*Partial Least-Squares Discriminant Analysis*, PLS-DA) e OPLS-DA (*Orthogonal Partial Least Square Discriminant Analysis*) foram inicialmente testados como modelos de classificação. Ambos são métodos de aprendizado supervisionado, ou seja, os dados de treinamento são associados a rótulos contendo informações sobre os dados na qual o modelo irá buscar uma relação direta entre a resposta instrumental (matriz X) e a informação fornecida ao sistema (matriz ou vetor Y) (GÉRON, 2019; MÁRCIA MIGUEL CASTRO FERREIRA, 2015). Frente ao PLS-DA, a OPLS-DA, chamada originalmente de projeções ortogonais à estrutura latente (*Orthogonalized Projection to Latent Structures*), foi desenvolvida como uma estratégia para lidar com matrizes contidas de grandes variações nos dados que não são diretamente correlacionados com a informações providas ao sistema, assim, o algoritmo particiona as informações relevantes em um conjunto de dados ortogonal (MARINI, 2020).

Em relação aos métodos não lineares, a Rede Neural Artificial (*Artificial Neural Network*, ANN) é inspirado na lógica por trás do processamento sensorial do cérebro humano. ANN é um método de aprendizado profundo criado pela simulação de uma rede de neurônios por meio de algoritmos que resulta em um modelo capaz de tomar decisões treinadas. Já a Floresta Randômica, (*Random Forest*, RF) é um método cujo algoritmo cria aleatoriamente várias árvores de decisão (método semelhante a fluxogramas que estabelecem conexões entre dados) e as combinam de forma a chegar em uma classificação final. Ambos métodos foram testados isolados ou associados com análise de componentes principais (*Principal Component Analysis*, PCA), utilizada como

ferramenta de seleção de variáveis, sendo denotados como PCA-ANN e PCA-RF, respectivamente.

Os parâmetros de desempenho dos modelos construídos estão listados na Tabela 6 e a matriz de confusão dos modelos não-lineares na Tabela 7. Em relação aos métodos lineares, apesar de apresentarem *F1-Scores* acima de 70 % indicando uma boa performance de acertos, o desempenho foi inferior em múltiplos parâmetros, incluindo MCC e especificidade, o que sugere algumas limitações de classificação equilibrados entre classes positivas e negativas, indicando que a correlação dos dados não se insere em um sistema linear, dessa forma sendo descartados como opções de classificação para esse conjunto de dados.

No geral, a PCA-ANN apresenta os maiores índices, como resultado de sua proficiência em classificar efetivamente instâncias positivas e negativas. A combinação de PCA para extração de variáveis significativas forneceu resultados melhores do que a ANN isolada na maior parte dos parâmetros de validação. A PCA-RF por outro lado, apresentou uma capacidade de classificação semelhante à RF isolada em vista dos valores obtidos, com exceção da especificidade que se mostrou superior com a RF sem associação com a PCA. Dessa forma, o modelo PCA-ANN surge como o método mais apropriado para tarefas de classificação binárias para esse conjunto de dados, beneficiando-se da sinergia entre a extração de características do PCA e as capacidades de aprendizagem não linear da ANN.

TABELA 6 - Avaliação comparativa de desempenho de todos os modelos classificatórios por CZE-UV.

Modelo	Exatidão	Sensibilidade	Especificidade	MCC	<i>F1-Score</i>
PLS-DA	62.71%	68.18%	46.67%	0.04	0.73
OPLS-DA	67.80%	69.39%	60.00%	0.23	0.78
ANN	72.9%	84.3%	59.3%	0.45	0.76
PCA-ANN	77.9%	79.1%	75.0%	0.52	0.84
RF	71.19%	69.81%	83.33%	0.33	0.81
PCA-RF	71.19%	74.42%	62.50%	0.34	0.78

Fonte: Própria autora (2024).

TABELA 7 - Matriz de confusão combinada dos modelos de aprendizado de máquina não lineares associados aos dados de CZE-UV.

Modelos não-lineares	Classe real	Classe predita	
		Negativa	Positiva
ANN	Negative	27	11
	Positive	5	16
RF	Negative	37	1
	Positive	16	5
PCA-ANN	Negative	34	4
	Positive	9	12
PCA-RF	Negative	32	6
	Positive	11	10

Fonte: Própria autora (2024).

Vale ressaltar que apesar de nenhum dos modelos ter conseguido ultrapassar o valor de 0,52 para MCC, os resultados indicam uma correlação clara entre a COVID-19 e o perfil de ácidos orgânicos em urina humana, numa realidade que entre todas as amostras a maior parte do perfil de ácidos orgânicos é a mesma, e também, as pequenas diferenças não podem ser extraídas diretamente apenas da ocorrência da doença, mas também de outros fatores como genética, alimentação, hidratação, hábitos regulares e muito mais.

Sobre a abordagem de leitura por CZE-UV em triplicatas autênticas, esta aumenta a robustez e generalização do modelo. Ao expor o modelo a variabilidades dentro de cada amostra durante o treinamento e o teste, melhorando sua capacidade de identificar dados novos e inéditos. Isto é particularmente valioso em química analítica, onde as amostras podem apresentar características únicas que são cruciais para uma classificação precisa. Além disso, a estratégia de medidas repetidas contribui para a captura mais eficaz de características específicas da amostra. Certas variações ou nuances podem ser específicas de amostras individuais, e a inclusão de eletroferogramas da mesma amostra em ambos os conjuntos de dados permite ao modelo discernir e aprender melhor esses recursos específicos da amostra.

A partir dos resultados obtidos desse estudo, conclui-se que rotinas envolvendo análise de urina associadas à ciência de dados são promissoras para estudos de triagem.

Além disso, esses resultados indicam que análises *targeted* envolvendo esse grupo de metabólitos são promissoras, cuja investigações aprofundadas podem gerar informações clínicas válidas para o melhor conhecimento da doença, não necessariamente apenas considerando o diagnóstico

Especificamente sobre a COVID-19, vários estudos têm indicado alterações no metabolismo de lipídios/ácidos graxos. Estudos de longo prazo com pacientes em recuperação de infecção por SARS também indicam desequilíbrios no metabolismo lipídico com perfis de soro sanguíneo diferenciais que incluem ácidos dicarboxílicos e hidroxiácidos como biomarcadores, o que corrobora com a interpretação biológica observada neste estudo (KOWALSKA et al., 2022; TANNER; ALFIERI, 2021; ZHAO et al., 2022).

Do ponto de vista analítico, técnicas de eletromigração em dimensão capilar associadas à detecção por ultravioleta são apresentadas neste trabalho como uma opção simplificada e automatizada onde dezenas de amostras podem ser lidas sem a necessidade de supervisão de um analista. Neste cenário, a CZE é apresentada como uma nova opção para obter dados semelhantes a espectros para ensaios classificatórios posteriores com as vantagens práticas comentadas na seção introdutória. Ainda que a utilização de CZE-UV como técnica de triagem em relação ao uso do resultado para decisões médicas ainda é distante, este protocolo analítico pode servir de triagem para estudos mais elaborados. Quando a metabolômica alvo está envolvida, o que vem à mente é a instrumentação multiplataforma que combina Espectrometria de Massas, Cromatografias e Ressonância Magnética Nuclear. Comparada com as técnicas baseadas em MS citadas, a CZE tem uma abordagem menos sofisticada, especialmente quando se pretende uma estratégia de rastreio. Neste contexto, embora o CZE-UV não tenha seletividade suficiente ou informações de impressão digital para identificação de moléculas dentro de matrizes biológicas complexas, pode ser útil como avaliação preliminar de amostras a serem analisadas posteriormente por um desses sistemas abrangentes.

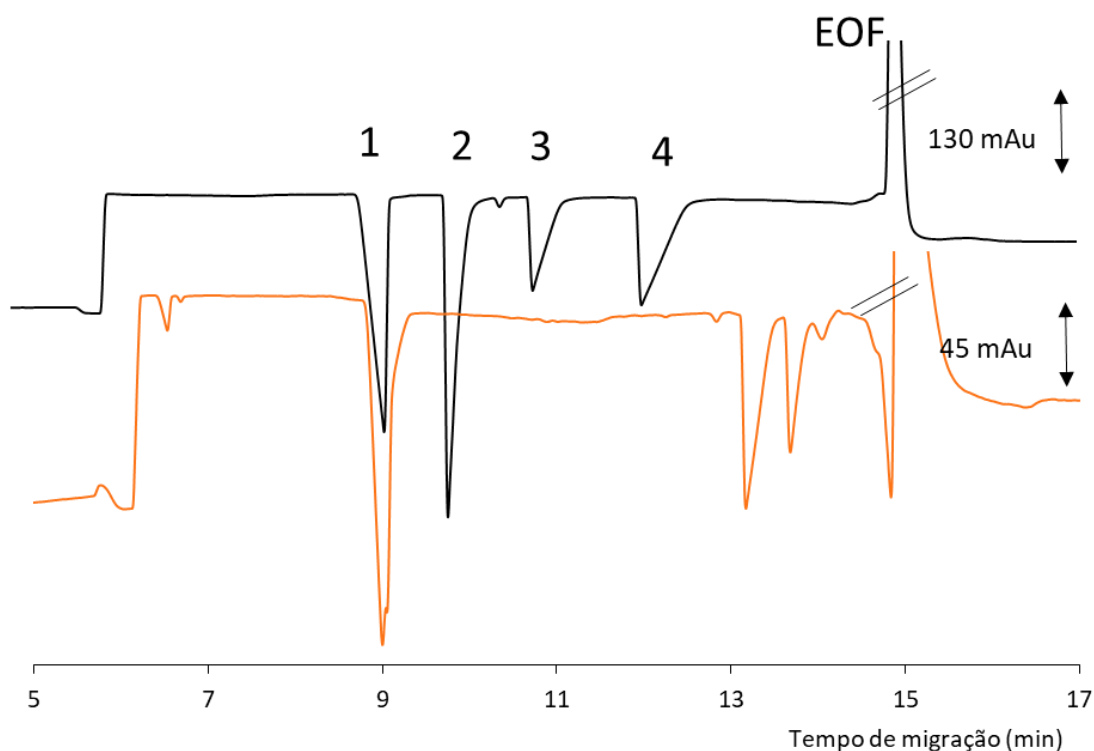
Os resultados referentes a este estudo estão atualmente em fase final de publicação em revista acadêmica da área de química analítica com indicativo Qualis Capes A1.

5.1.6 Perfil de ácidos orgânicos dicarboxílicos e hidroxiácidos em saliva

Uma amostra controle de saliva também foi analisada a partir do método otimizado a fim de testar sua aplicabilidade em outras matrizes. Um sinal intenso

condizente com o tempo de migração do tartarato foi registrado no eletroferograma (Figura 11). Ainda que outros ácidos com características estruturais próximas ao descrito anteriormente na curva de mobilidade (Figura 4) possam estar presentes na saliva, não estão em quantidades suficientes para que seja registrado, porém, a menor quantidade de metabólitos em saliva em comparação com a urina é esperada. Próximo a região do sinal do EOF são registrados sinais intensos correspondentes aos ácidos monocarboxílicos, espécies separadas com melhor resolução em método discutido posteriormente.

FIGURA 11- Eletroferogramas resultantes da análise da mistura de padrões (em preto), QC de saliva (em laranja). (1) Tartarato; (2) Malato; (3) Lactato; (4) Succinato.

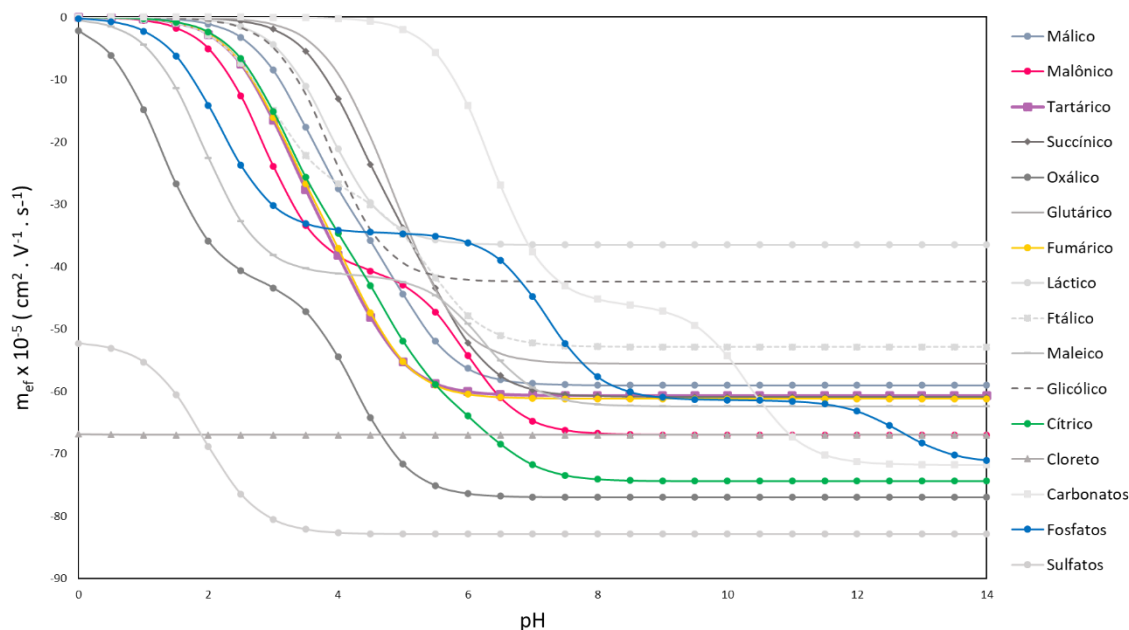


Fonte: Elaborado pela autora (2024).

Mediante a possibilidade de ser um método aplicável ao doseamento de metabólitos, a fim de identificar os picos registrados foi realizada uma busca por possíveis interferências considerando espécies de mobilidade próximas ao tartarato. Além dos demais ácidos orgânicos previstos inicialmente na curva de mobilidade favorecidos pelo método (Figura 4), alguns outros ácidos orgânicos somados a sais inorgânicos presentes na composição da saliva como fosfatos, carbonatos e cloreto podem ser favoráveis às

condições eletroforéticas desse sistema. A fim de avaliar possíveis comigrações, uma nova curva de mobilidade (Figura 12) contendo essas espécies foi construída a partir de dados tabelados providos na base do software PeakMaster®.

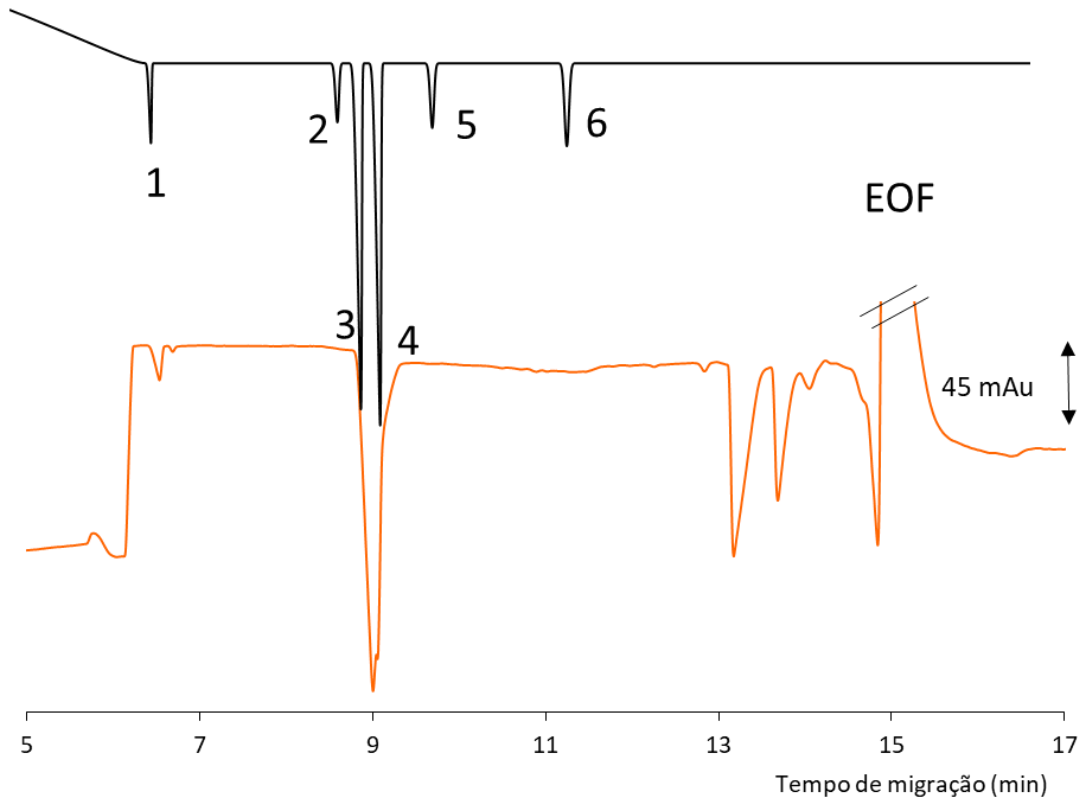
FIGURA 12 – Curva de mobilidade efetiva vs pH exemplificando o comportamento de alguns ácidos orgânicos dicarboxílicos e hidroxiácidos contendo possíveis interferentes do sinal em $t = 9$ min (em linhas coloridas).



Fonte: Elaborado pela autora (2024).

Daqueles que apresentam mobilidade crítica próximas ao tartarato, se encontram o íon dos ácidos malônico, cítrico, fumárico e íons inorgânicos fosfatos. De acordo com a simulação de eletroferograma que mimetiza as exatas condições experimentais desse método, os ácidos orgânicos citados apresentam um perfil de comigração com o tartarato (Figura 13). Fazendo uma comparação com o eletroferograma da amostra real, nota-se que o formato do pico de fato indica a presença de comigração de analitos, possivelmente tartarato e fumarato sendo registrados no mesmo tempo de migração.

FIGURA 13 – Eletroferograma simulado contendo interferentes do tartarato (em azul) comparado a uma amostra QC de saliva (em laranja). (1) Oxalato; (2) Malonato; (3) Tartarato; (4) Fumarato; (5) Fosfato e (6) Citrato.



Fonte: Elaborado pela autora (2024).

O ácido tartárico é comumente incluído como conservante, intensificador de sabor e acidulante em doces, sucos e vinhos, também sendo encontrado naturalmente em algumas plantas e frutas. O ácido fumárico é encontrado na natureza em cogumelos, carne fresca e na *Corydalis*, gênero botânico da família Fumariaceae, comercializada também em cápsulas e utilizada na medicina tradicional chinesa como terapia analgésica. Na indústria farmacêutica convencional, o ácido fumárico é ingrediente ativo de alguns medicamentos orientados para o tratamento de psoríases. Por fim, também é utilizado frequentemente como antioxidante em alguns segmentos da indústria de alimentos e bebidas.

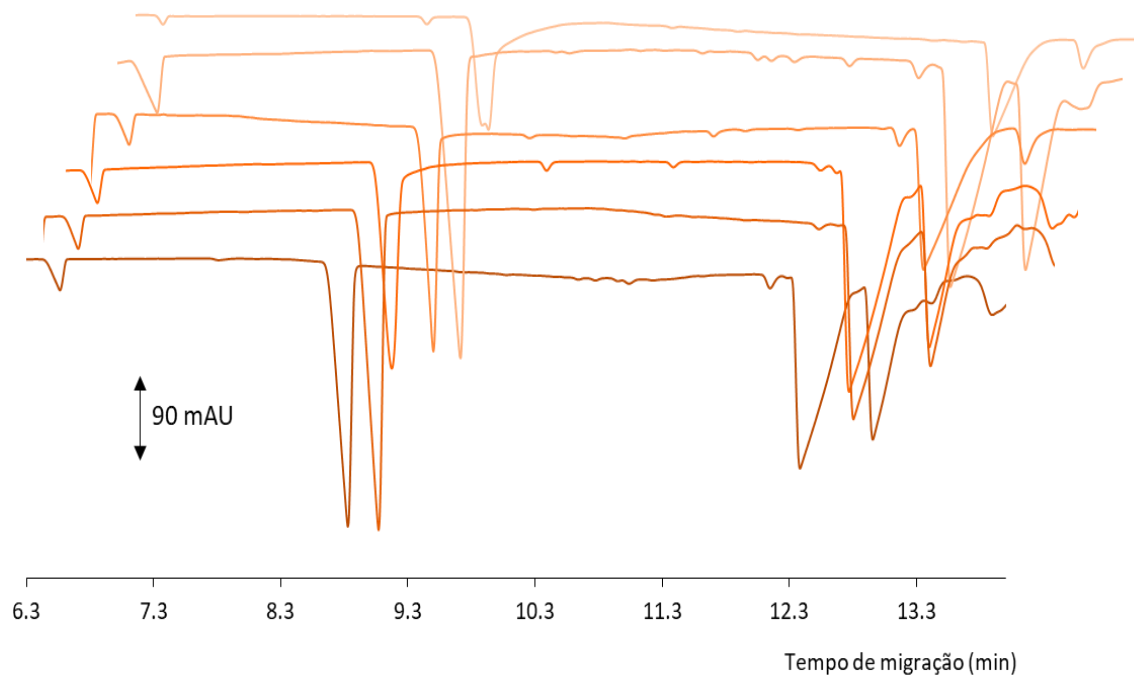
Enquanto metabólitos, o fumarato atua no ciclo de Krebs sendo convertido em malato pela enzima fumarase e já foi identificado como metabólito oncológico e está inserido no grupo de marcadores para monitoramento de anomalias no ciclo circadiano

(DALLMANN et al., 2012; HMDB, 2024a). O ácido tartárico é uma toxina muscular que inibe a produção de outros metabólitos importantes e apesar de ser um aditivo da indústria alimentícia a dose letal é relativamente baixa e um percentual do que é consumido (cerca de 20 %) é excretado pelo corpo ainda em sua estrutura inalterada. Como biomarcador em urina, já foi associado à atividade fúngica. Altos níveis de tartarato também já foram relacionados com autismo em crianças, no entanto, em adultos os níveis elevados podem ser consequência de fatores externos como consumo de vinhos e uvas (HMDB, 2024b). Assim sendo, ambos podem estar presentes em saliva, como metabólitos ou como molécula inalterada obtida por consumo desses alimentos.

O pico assinalado como (1) na Figura 13 foi identificado como oxalato. No organismo humano, o oxalato é o metabólito final sintetizado a partir do ácido glicólico ou do ácido ascórbico não aproveitado. A característica marcante dos oxalatos no corpo humano é a formação dos cálculos renais e dos cálculos acumulados nos dentes, conhecido como tártaro, também associados diretamente com a incidência de cáries (HMDB, 2024c; VINCENT et al., 2021; WAHL; KALLEE, 1994). Neste caso, o sinal aparenta estar separado de demais analitos, podendo ser uma possibilidade de método de doseamento, se relevante.

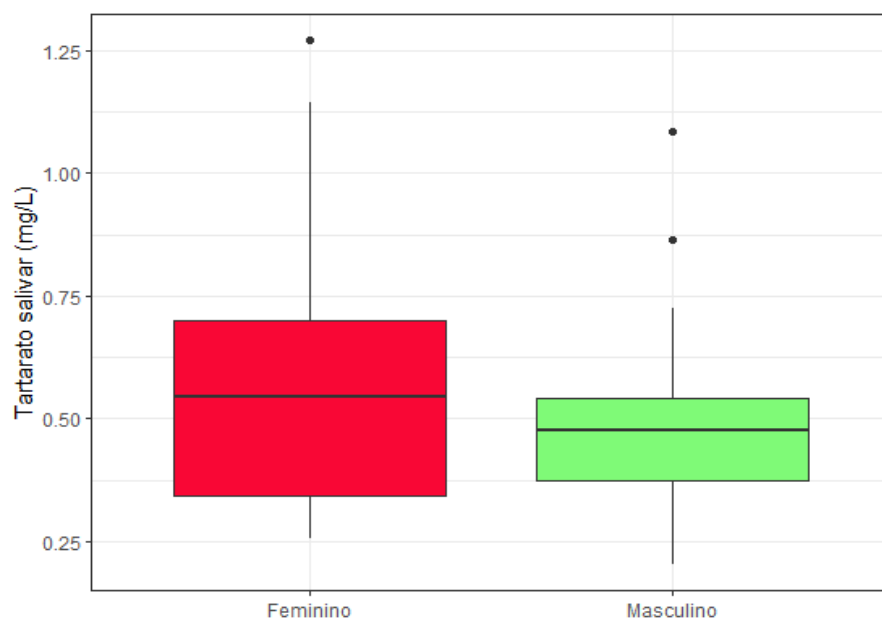
Os eletroferogramas individuais resultantes das 60 amostras analisadas indicam um perfil semelhante à amostra controle, sendo que em alguns casos há definição do pico em 9 min, ou seja, sem características de comigração, conforme exemplificado na Figura 14. Para estes, o tartarato foi quantificado a partir de curva de calibração externa ($y = 59.571 (\pm 1.57) x - 6.5856 (\pm 10.53)$; $r = 0.9972$; $F_{Faj} = 0.057$; $F_R = 1445$). Em média, obteve-se o valor de $0.54 (\pm 0.22) \text{ mg L}^{-1}$ expresso em massa de analito por volume de saliva, com uma variação na distribuição de valores entre homens e mulheres (Figura 15). No entanto, não foram encontrados na literatura registros do intervalo normal de tartarato salivar. Por fim, os eletroferogramas sugerem a presença de outros metabólitos na região entre 10 e 13 minutos, possivelmente ácidos com cadeias mais longas e com radicais que aumentem o raio da molécula.

FIGURA 14 – Eletroferogramas resultantes da análise individual de algumas amostras aleatórias de saliva.



Fonte: Elaborado pela autora (2024).

FIGURA 15 – *Box-Plot* da distribuição de concentração de tartarato salivar (em termos de massa por volume de saliva) entre homens e mulheres.



Fonte: Elaborado pela autora (2024).

Frente a essas observações, nota-se que os ácidos orgânicos citados aparentam ter importância biológica suficiente que poderia justificar a análise quantitativa desses metabólitos mediante estudos mais aprofundados. Sobre o método, com exceção do oxalato que aparenta estar separado com resolução adequada, para doseamento dos demais metabólitos adaptações no método seria necessária. Ainda que uma análise exploratória quantitativa para o tartarato tenha sido realizada, apenas com esses experimentos não há certeza sobre a identificação desse pico. Além disso, em vista da concentração baixa de metabólitos em saliva, práticas de pré-concentração são aplicáveis nesse caso. Ademais, de acordo com estudos completos do metaboloma salivar, ácidos orgânicos de cadeia curta e outras estruturas não favorecidas por esse método são registradas com maior frequência (DAME et al., 2015; GARDNER; CARPENTER; SO, 2020).

Em relação à aplicabilidade do método, em vista das considerações supracitadas, estudos adicionais devem ser realizados, como por exemplo, interferência da matriz e utilização de recursos para adição de padrões nas amostras voltados para a identificação de sinal. Para este trabalho especificamente, visto que não foram divulgadas informações dos pacientes exceto gênero e que não foram coletadas as amostras do projeto de arboviroses, um estudo desses metabólitos como biomarcadores para alguma condição específica não foi viável de elaboração, além disso, acredita-se que a resposta do método voltado para ácidos monocarboxílicos, discutido a seguir, é mais promissora para estudos do metaboloma salivar.

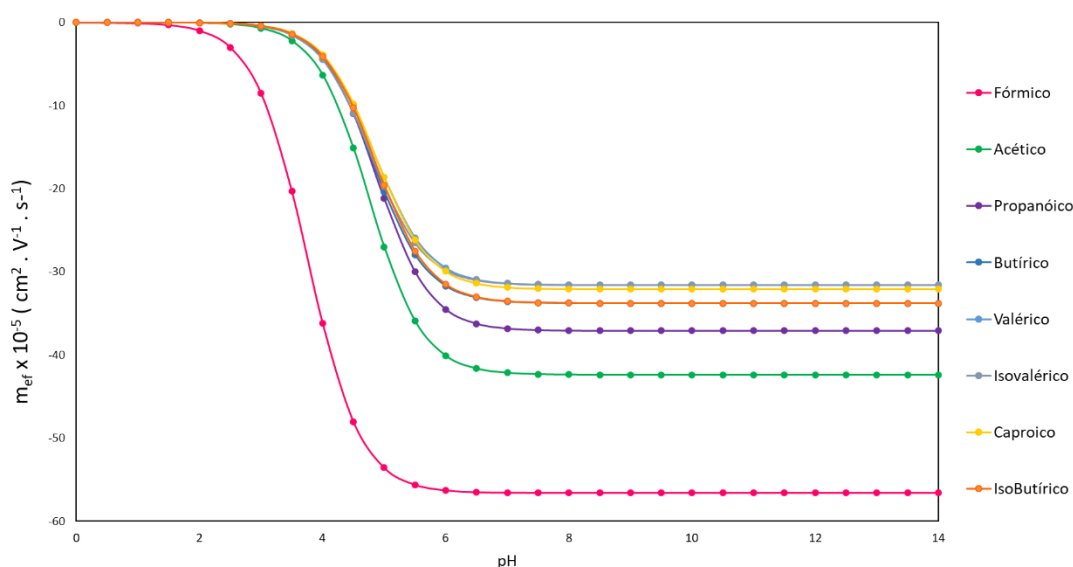
5.1.7 Perfil de ácidos orgânicos monocarboxílicos de cadeia curta

Os ácidos monocarboxílicos de cadeia curta (SCFA) também apresentam relevância bioquímica para estudos do metaboloma humano. Em geral, os SCFA são associados à saúde do sistema digestivo com implicações sistêmicas (KTSOYAN et al., 2011; O'RIORDAN et al., 2022). Além disso, estão listados como marcadores salivares relacionados a aspectos de saúde bucal também com implicações sistêmicas (DAME et al., 2015; GARDNER; CARPENTER; SO, 2020).

Considerando as características estruturais e físico-química dessas moléculas, novamente o passo inicial do desenvolvimento considera os valores de pKa desses compostos, valores estes maiores que os analitos em foco no método descrito anteriormente. A partir da curva de mobilidade efetiva desses compostos (Figura 16), em

faixas de pH acima de 4 tendendo à alcalinidade, os analitos começam a atingir valores maiores de mobilidade, o que justifica a necessidade de um segundo método. SCFA com cadeias alifáticas de 1 a 6 carbonos ainda são hidrofílicos, cadeias maiores começam a apresentar limitações quanto à solubilidade, sendo necessário o desenvolvimento de métodos em meio parcialmente não aquoso. Uma vez que o objetivo é o estudo de metabólitos em urina e saliva, o foco se mantém em analitos hidrossolúveis.

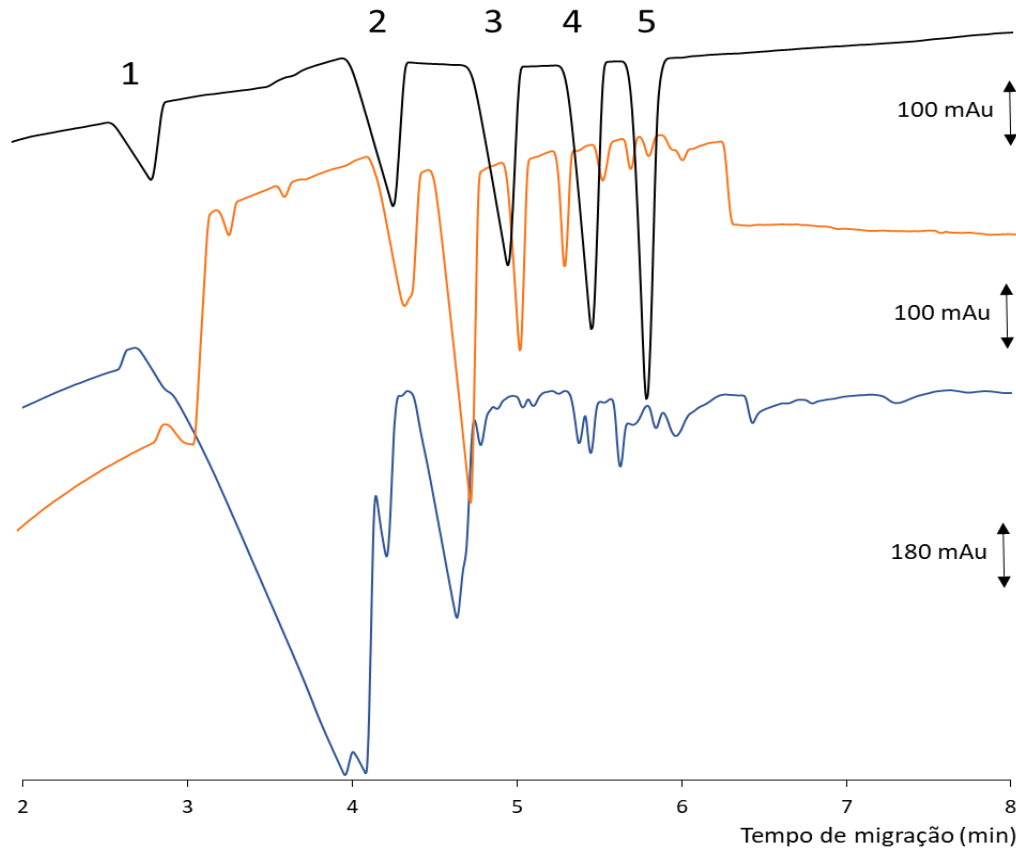
FIGURA 16 - Curva de mobilidade efetiva vs pH exemplificando o comportamento de alguns ácidos orgânicos monocarboxílicos de cadeia curta



Fonte: Elaborado pela autora (2024).

Após simulações utilizando o software Peakmaster®, condições promissoras foram encontradas em pH 7.5 no que diz respeito à resolução de picos e tempo de corrida. Foi selecionado o ácido 3,5-dinitrobenzóico (3,5-DNB) como cromóforo, uma vez que apresenta mobilidade próxima aos analitos. Considerando o pKa do TRIS (8.076), este também foi usado como base para composição do tampão TRIS/TRISH⁺, com a base conjugada do 3,5-DNB como contra íon. CTAB como inversor de fluxo também foi adicionado ao BGE. Ácidos fórmico, acético, propanoico, butírico e valérico foram usados como referência para otimização. Demais parâmetros instrumentais foram ajustados experimentalmente, mantendo tamanho de capilar e condicionamento igual ao método discutido anteriormente. O método foi novamente testado em amostras de saliva e urina (Figura 17).

FIGURA 17 - Eletroferogramas resultantes da análise da mistura de padrões (em preto), QC de saliva (em laranja) e QC de urina (em azul). (1) Formiato; (2) Acetato; (3) Propionato; (4) Butirato; (5) Valerato.



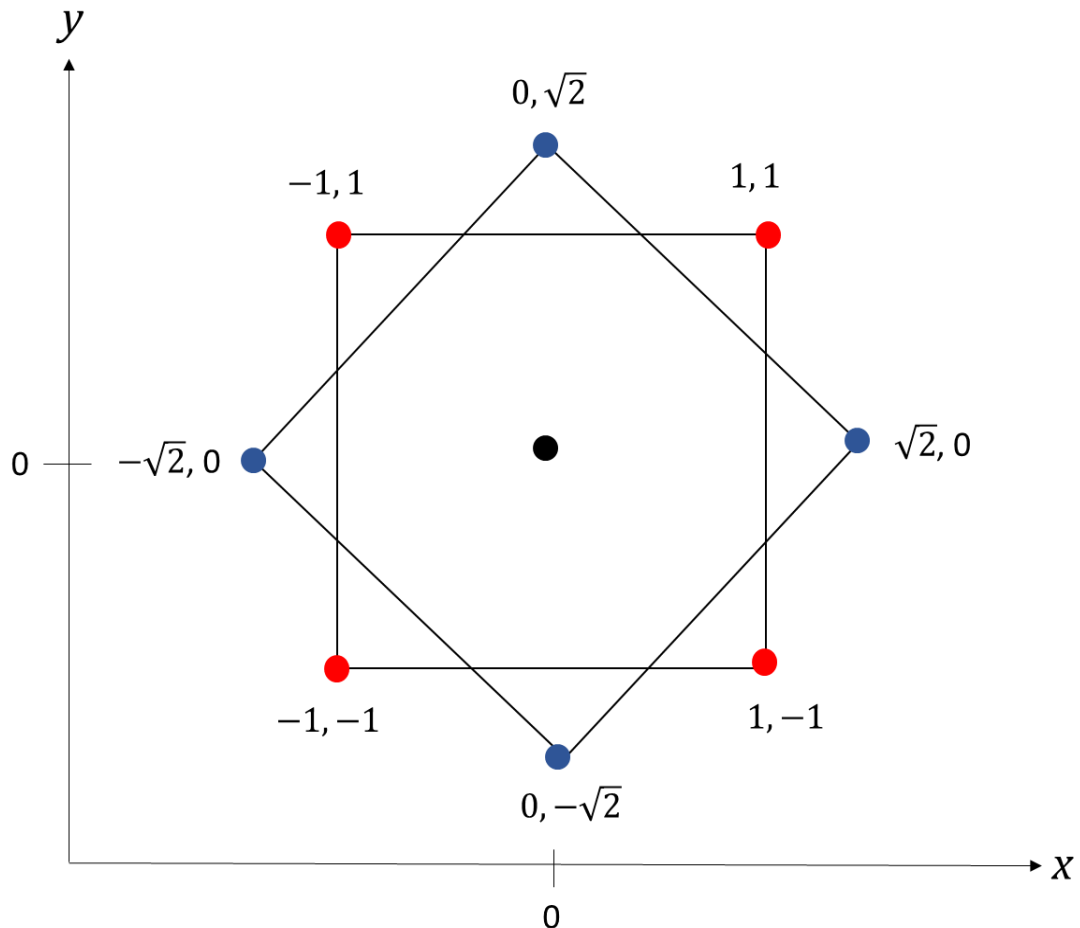
Fonte: Elaborado pela autora (2024)

Para uma matriz complexa como a urina, conforme esperado foi obtido um perfil geral de compostos favorecidos pelo método. O sinal mais intenso em 4 minutos sendo possivelmente consequência da alta concentração de íons inorgânicos rápidos. Por outro lado, para matriz salivar foi registrado um eletroferograma com picos definidos, novamente sendo uma possibilidade de método para doseamento de metabólitos.

Neste sentido, testes para identificação e melhoramento do método para saliva foram feitos. Na intenção de avaliar a influência da força iônica da resolução entre os picos, foi desenhado um planejamento de experimentos focados na concentração dos componentes do BGE, exceto o CTAB. Um planejamento estrela (Figura 18) foi selecionado para variação da proporção entre TRIS e 3,5-DNB (dois fatores) em valores próximos ao escolhido originalmente. Esse desenho corresponde a um planejamento

composto central na qual somada à parte fatorial tradicional representada pelo quadrado, os fatores também são variados utilizando o mesmo quadrado rotacionado em 45° em relação ao original unidos pelo mesmo ponto central (NETO; SCARMINIO; BRUNS, 2010).

FIGURA 18 – Representação do planejamento em estrela para duas variáveis.



Fonte: (NETO; SCARMINIO; BRUNS, 2010).

Ao todo são 11 ensaios, sendo 8 correspondentes a cada vértice (Figura 18) e a triplicata no ponto central. Os experimentos foram organizados conforme descrito na Tabela 8. Os fatores correspondem à concentração dos componentes do eletrólito exceto o CTAB mantido fixo na concentração original (0.5 mmol L^{-1}).

TABELA 8 – Fatores de variação do planejamento estrela.

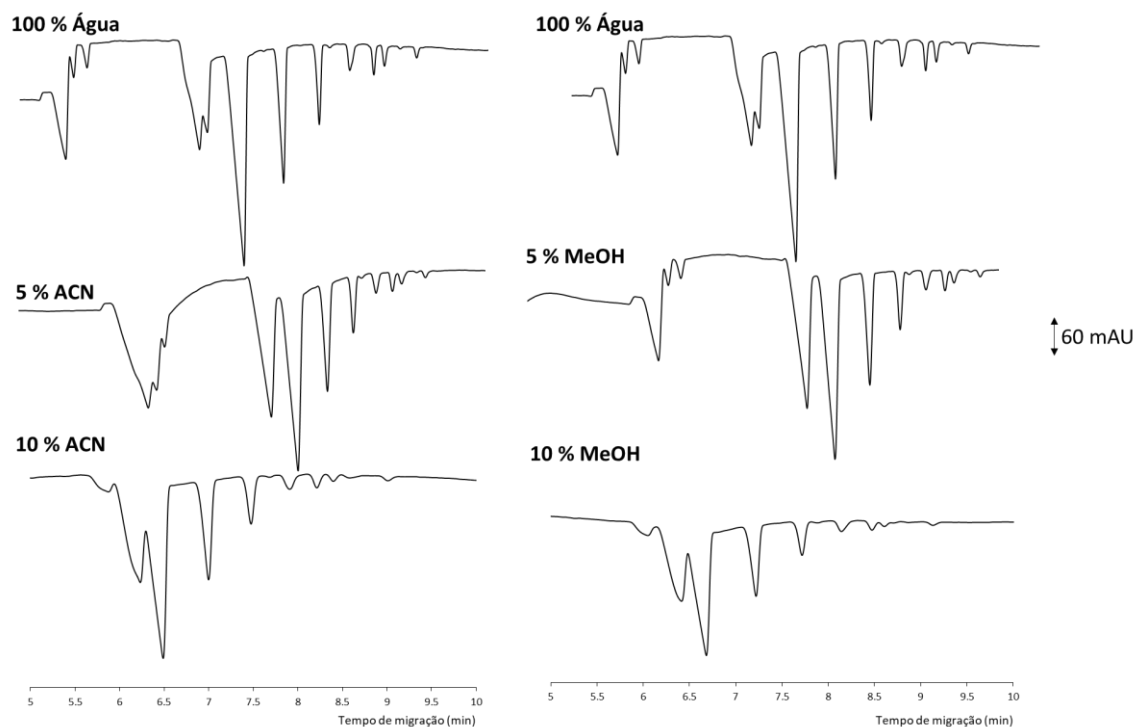
Ensaio	x_1	x_2	C_{TRIS} (mmol L ⁻¹)	$C_{\text{3,5-DNB}}$ (mmol L ⁻¹)
1	-1	-1	13	10
2	1	-1	17	10
3	-1	1	13	14
4	1	1	17	14
5	0	0	15	12
6	0	0	15	12
7	0	0	15	12
8	$-\sqrt{2}$	0	12.2	12
9	0	$\sqrt{2}$	15	14.8
10	$\sqrt{2}$	0	17.8	12
11	0	$-\sqrt{2}$	15	9.2

Fonte: (NETO; SCARMINIO; BRUNS, 2010)

Durante o preparo dos eletrólitos correspondentes aos ensaios descritos na Tabela 8, observou-se que nem todas as misturas permaneceram solúveis. Visto que o ácido 3,5-dinitrobenzóico só solubiliza em soluções aquosas na forma de sal, ou seja, após adição do TRIS para desprotonar o ácido carboxílico do 3,5-DNB, nem todas as misturas mantiveram uma proporção entre TRIS e 3,5-DNB suficiente para que o eletrólito fosse solúvel. daquelas que se mostraram aplicáveis, não foram gerados eletroferogramas de boa qualidade. Dessa forma, uma nova estratégia para melhorar a resolução entre picos foi adotada.

Foi testado experimentalmente o uso de solventes orgânicos a fim de modificar o raio de solvatação dos analitos e por consequência, promovendo melhor separação. Mantendo a concentração original de 15 mmol L⁻¹ de TRIS, 12 mmol L⁻¹ de 3.5-DNB e 0.5 mmol L⁻¹ de CTAB, foram adicionados separadamente acetonitrila e metanol à solução, ambos a 5 e 10 %. Os testes foram aplicados às amostras controle de saliva (Figura 19).

FIGURA 19– Eletroferogramas resultantes dos testes de otimização do eletrólito para o método dos ácidos orgânicos monocarboxílicos.

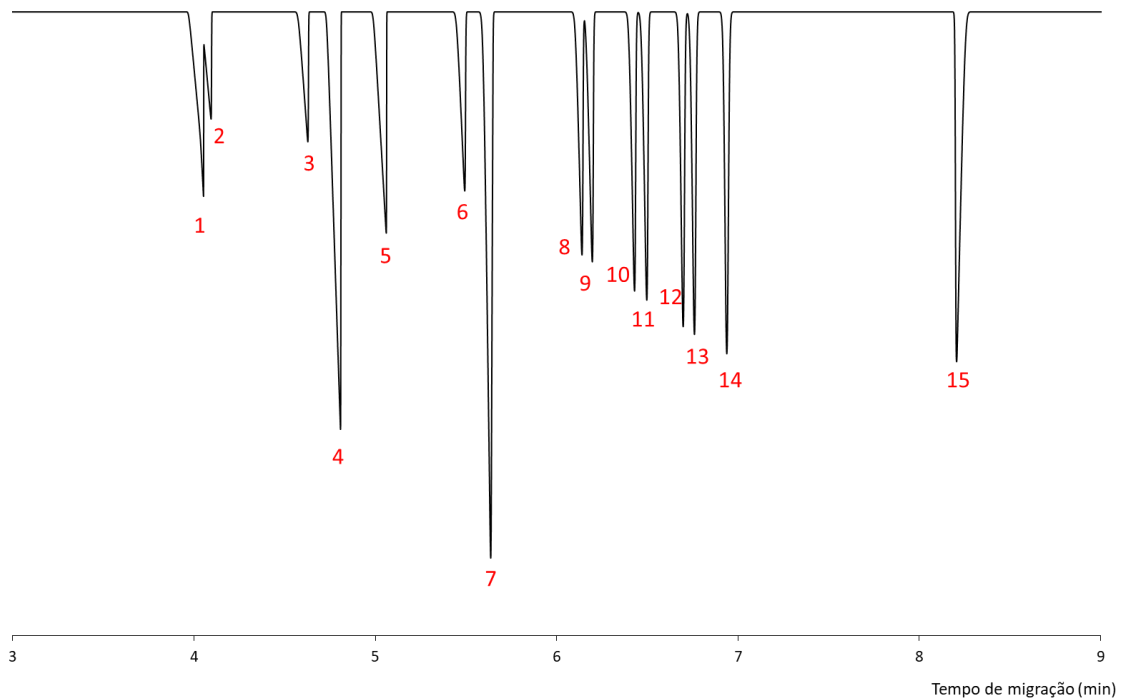


Fonte: Elaborado pela autora (2024).

Nota-se que o uso de 10 % de solvente orgânico prejudica a resolução dos picos em ambos os casos, provavelmente devido a desestabilização da camada responsável pela inversão de fluxo formada pelo CTAB. Percentuais menores de solvente mantiveram o perfil do eletroferograma com exceção do pico em torno de 7.6 minutos. Dessa forma, conclui-se que o uso do eletrólito original sem solvente orgânico ainda é a melhor opção.

Outros testes complementares envolvendo otimização de temperatura e voltagem, incluindo um novo planejamento de experimentos, poderia ser uma alternativa para melhorar o sinal característico de comigração, se estes se mostrarem relevantes para quantificação. A partir desse método, foi feita a simulação de analitos favoráveis a serem registrados por esse método (Figura 20). De modo oposto ao método discutido anteriormente, esse favorece a separação dos ácidos monocarboxílicos, porém ainda são registradas algumas outras estruturas de ácidos dicarboxílicos, hidroxíácidos e sais inorgânicos usualmente presentes em amostras de fluidos biológicos, sem separação entre eles.

FIGURA 20 – Simulação de analitos do método para ácidos monocarboxílicos.



1: Oxalato; 2: Cloreto; 3: Formiato; 4: Succinato + tartarato; 5: Fosfatos; 6: Carbonato; 7: Piruvato + glicolato; 8: Propionato; 9: Lactato; 10: Isobutirato; 11: Butirato; 12: Isovalerato; 13: valerato; 14: Caproato; 15: Caprilato.

Fonte: Elaborado pela autora (2024).

Em vista dos resultados discutidos, esse método mostra-se promissor para obtenção de perfil de metabólitos em fluidos biológicos seguido de processamento de dados voltados para modelos de predição, a exemplo do estudo elaborado para o método discutido anteriormente. Futuramente, pretende-se fazer a análises das amostras de urina do projeto COVID-19 por esse método e avaliar se os resultados de classificação serão melhores que o anterior. No entanto, ainda que algumas tentativas de identificação tenham sido priorizadas para a saliva, para análises quantitativas de quaisquer um dos metabólitos citados, existe a necessidade de adaptar o método para um sistema de massas de baixa resolução, ideal para rotinas de doseamento que exigem alta seletividade. Não foram realizados outros experimentos devido a problemas na infraestrutura que envolve o funcionamento dos equipamentos de eletroforese capilar.

5.2 AMINOÁCIDOS E CREATININA

Semelhante aos ácidos orgânicos, os aminoácidos são biomarcadores conhecidos, visto que são produtos e substratos de diversas rotas metabólicas, além de ser as estruturas fundamentais da síntese de proteínas, chamados de aminoácidos essenciais. Aminoácidos não proteínogênicos também desempenham papel importante no metabolismo de neurotransmissores e alguns hormônios, para citar algumas das diversas funções. No caso dessa classe de compostos, a gama de patologias clínicas a qual está associada é vasta, incluindo anomalias metabólicas, doenças cardiovasculares, autoimunes, oncológicas e neurológicas, por exemplo (GU et al., 2016; LING et al., 2023; MOREIRA et al., 2022).

São estruturas de baixa massa molecular e hidrossolúveis em sua maioria na qual grupos carboxila e amino são ligados a um mesmo carbono, podendo conter outros heteroátomos na cadeia. Desse modo, são moléculas facilmente suscetíveis à análise por eletroforese capilar. Seguindo o objetivo de obter um perfil desses compostos, um novo método por CZE-UV foi desenvolvido. Neste método, a análise de creatinina foi incluída, uma vez que também é um composto contido de um grupo amino com propriedades físico-químicas semelhantes aos aminoácidos, sendo possível uma análise simultânea desses compostos.

Para análises de urina em geral, a creatinina é o composto tradicionalmente utilizado como parâmetro de normalização, uma vez que está associado a atividade renal, ou seja, tendo a urina como uma solução de compostos de concentração desconhecida diluídas em água, o valor da creatinina urinária normatiza esse fator de diluição. Logo, a composição da urina como um todo em termos quantitativos está relacionada com os níveis de creatinina, tanto que, exames laboratoriais baseados na quantificação de analitos orgânicos diversos em urina são expressos em termos da razão de massa de analito por massa de creatinina.

Neste contexto, ainda que os aminoácidos sejam registrados no eletroferograma como um perfil, o método analítico por CZE-UV foi proposto para que a creatinina fosse separada dos demais com resolução suficiente para quantificação, na qual foi otimizado um protocolo de doseamento a partir de calibração externa com adição de padrão interno com programa de injeção automatizada via sistema de CE. Para explorar as análises quantitativas de aminoácidos em urina, um método por CZE-MS utilizando um

espectrômetro de massa do tipo triplo quadrupolo foi desenvolvido de modo complementar.

5.2.1 Reagentes e soluções

Todas as soluções foram preparadas em água deionizada por sistema de Osmose Reversa da Quimis (São Paulo, SP). Substâncias de referência, tais como creatinina, imidazol, Fenilalanina, Glicina, Lisina, Metionina, Histidina, Triptofano, Valina foram adquiridos da Sigma Aldrich (Saint Louis, EUA) com certificação de pureza. Demais materiais utilizados: Hidróxido de sódio (NaOH) da Synth (São Paulo, SP); Ácido fosfórico (H_3PO_4), dihidrogenofosfato de sódio (NaH_2PO_4), Ácido fórmico e Hidróxido de amônio (NH_4OH) da Sigma Aldrich (Saint Louis, EUA), Ácido clorídrico (HCl) da Vetec (Rio de Janeiro, RJ). Soluções-estoque das substâncias de referência foram preparadas individualmente a 10 mmol L^{-1} em água deionizada e armazenadas em refrigeração a $4 \text{ }^\circ\text{C}$ até análise. Soluções-estoque do tampão e de outros componentes foram preparadas individualmente e diluídas em água deionizada para o preparo do BGE, também armazenado em refrigeração até o uso.

5.2.2 Instrumentação e métodos por CZE-UV

Os experimentos por eletroforese capilar com detecção por ultravioleta foram realizados em equipamento modelo 7100 da Agilent Technologies (Palo alto, USA) com detector DAD e controlado por software OpenLab ChemStation C.01.07. Parâmetros otimizados específicos de cada método estão descritos na Tabela 9. Amostras e soluções de referência foram injetadas hidrodinamicamente em capilar de sílica fundida revestido com poliimida (Série TSP) (comprimento total 68.5 cm, $75 \mu\text{m}$ de diâmetro interno e $365 \mu\text{m}$ de diâmetro externo). O condicionamento inicial foi feito a partir da injeção contínua (*flush*) de NaOH com concentração de 1 mol L^{-1} a 940 mbar por 40 minutos, seguido por água deionizada por 20 minutos e BGE por 20 minutos. Conforme otimizado nos métodos anteriores, entre corridas, o capilar foi recondicionado com 1 mol L^{-1} NaOH (60 s), água deionizada (30 s), 1 mol L^{-1} HCl (30 s), segunda rodada de água deionizada (30 s) e eletrólito (120 s).

TABELA 9 - Parâmetros experimentais dos métodos desenvolvidos por CZE-UV para aminoácidos e creatinina.

Parâmetros otimizados	Método: Aminoácidos e creatinina
Eletrólito	80 mmol L ⁻¹ de tampão (H ₃ PO ₄ /NaH ₂ PO ₄) (pH 2.3).
Temperatura (°C)	25 °C
Comprimento de onda	200 nm
Comprimento efetivo do capilar	8.5 cm
Voltagem	- 15 kV
Corrente gerada	52 uA
Tempo de corrida	10 min
Programa de injeção	Amostra: -15 mbar x 4 s IS: -15 mbar x 4 s BGE: -10 mbar x 2 Espera: 2 s.

IS: *Internal Standard* (Padrão Interno).

Fonte: Elaborado pela autora (2024)

5.2.3 Instrumentação e métodos por CZE-MS

Os experimentos por eletroforese capilar com detecção por espectrometria de massas foram realizados em equipamento de CE modelo 7100 acoplado a espectrômetro de massas triplo-quadrupolo (QQQ) com fonte de ionização por eletrospray (*Electrospray Ionization*, ESI) e controlado por software MassHunter da Agilent Technologies B.08.00 (Palo alto, USA) disponível no Instituto de Química da Universidade de São Paulo. Foi utilizado capilar de sílica fundida revestido com poliimida (Série TSP) (comprimento total 75 cm, 50 µm de diâmetro interno e 365 µm de diâmetro externo) com controle de temperatura ajustado a 20 °C e condicionado inicialmente com injeção constante a 950 mbar de NH₄OH a 1 mol L⁻¹ por 40 minutos, água deionizada por 20 minutos e eletrólito por 20 minutos. Como eletrólito foi utilizado uma solução aquosa a 0.3 mol L⁻¹ de ácido fórmico e como líquido auxiliar (*Sheath Liquid*) uma solução de água e metanol (50:50

v/v) acrescentada de 5 mmol L⁻¹ de ácido fórmico, com fluxo constante de 0.6 mL min⁻¹. Entre corridas o capilar foi recondicionado com *flushs* de BGE 120 s. Parra corridas no modo *scan*, as amostras foram injetadas hidrodinamicamente a 20 mbar por 5s. Por monitoramento de reações múltiplas (*Multiple Reaction Monitoring*, MRM) utilizou-se a injeção em sequência de 20 mbar por 5 s de amostra seguido de padrão interno também 20 mbar por 5 s com 3 segundos de espera ao final.

Os dados foram coletados no modo positivo operado em (a) varredura completa (*scan*) com faixa de massa de 20 a 500 e voltagem de fragmentação 100 V; (b) MRM considerando massas moleculares de alguns aminoácidos fundamentais. Voltagens de fragmentação (V), colisão (V) e aceleração (V) foram individualmente selecionadas para cada analito investigado (Tabela 10). A fonte de ionização foi mantida a 200 °C, com fluxo de gás de secagem a 6 mL min⁻¹ e pressão de nebulização a 10 psi, com voltagem do capilar a 4500 V.

TABELA 10 - Parâmetro de fragmentação das massas de transição de cada analito.

Analito	Ion Precursor	Íon Produto	Fragmentação (V)	Colisão (V)	Acelerador (V)
Triptofano	204.9	188.0	100	6	3
	204.9	146	100	20	3
Fenilalanina	166.0	119.9	100	10	3
	166.0	102.9	100	37	3
Histidina	156.0	109.9	100	15	1
	156.0	82.9	100	31	1
Metionina	150.0	103.8	110	9	2
	150.0	55.9	100	15	3
Lisina	147.0	84.0	70	15	2
	147.0	55.9	120	50	2
Valina	118.0	72.0	100	10	1
	118.0	55.0	100	30	2
Glicina	76.0	48.8	90	10	3
	76.0	29.9	90	20	3
Creatinina	114.0	86.0	135	10	1
	114.0	43.9	120	35	2

Fonte: Elaborado pela autora (2024).

5.2.4 Construção de modelos de regressão e parâmetros de desempenho analítico

Análises quantitativas foram realizadas a partir de curvas de calibração por adição de padrão interno a partir de modelo OLS. Limites de detecção (LOD) e quantificação (LOQ) foram calculados a partir da abordagem sinal/ruído dos eletroferogramas de amostras controle. Inicialmente sendo necessário o cálculo da concentração do analito na amostra, a ser determinado de acordo com a equação 14:

$$C_x = \frac{S_1 \times C_s \times V_s}{(S_2 - S_1)V_x} \quad (14)$$

Em que S_1 é a área do pico do analito na amostra; S_2 é a área do pico do analito na amostra acrescida de padrão; C_s é a concentração da solução padrão; V é o volume adicionado de solução, sendo V_s e V_x os volumes de solução padrão e de amostra, respectivamente. Os valores encontrados para a concentração do analito na amostra (C_x) são inseridas nas equações 15 e 16 para a obtenção dos limites do método:

$$LOD = \frac{3 \times S_b \times C_x}{H_{MAX} - H_{MIN}} \quad (15) \quad LOQ = \frac{10 \times S_b \times C_x}{H_{MAX} - H_{MIN}} \quad (16)$$

S_b sendo o desvio padrão da linha base do eletroferograma da amostra controle. O fator S_b foi calculado utilizando ferramentas estatísticas do software Origin® a partir da seleção de uma região randômica do resultado de três análises de QC de urina. C_x é a concentração do analito na amostra, previamente calculado pela equação 13; denominadores das equações são a altura do pico. Os resultados de LOD e LOQ são expressos como a médias desses resultados e seu respectivo desvio padrão.

A exatidão foi calculada por teste de recuperação também em amostras de QC a partir da equação 17:

$$R (\%) = \frac{(S_2 - S_1)}{S_3} \times 100 \quad (17)$$

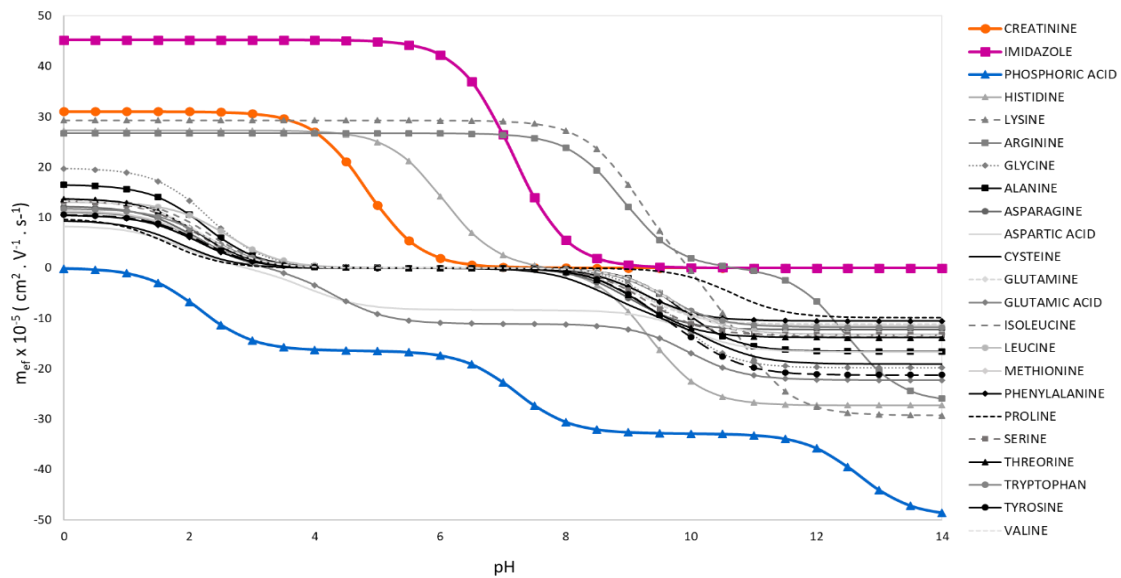
Em que S_1 é a área do pico do analito na amostra; S_2 é a área do pico do analito na amostra acrescida de padrão; S_3 a área do padrão.

5.2.5 Perfil de aminoácidos e doseamento de creatinina por CZE-UV

A determinação de creatinina por CZE-UV não é novidade, assim, o método foi otimizado a partir de estudos já publicados somado aos testes experimentais e simulações. A creatinina é facilmente ionizável na forma de cátion e apresenta em sua estrutura grupos cromóforos importantes, características comuns a alguns aminoácidos, como os 20 aminoácidos essenciais conhecidos. A partir de todas essas considerações, foi desenvolvido um método por detecção direta capaz de, simultaneamente, mediar a quantificação de creatinina e registrar o perfil de aminoácidos presente na amostra.

Lisina, histidina e arginina foram as principais preocupações em relação à comigração com creatinina (Figura 21), o que foi resolvido a partir do uso de solução tampão considerando a primeira ionização do ácido fosfórico (pH 2.16) na concentração de 80 mmol^{-1} . Outras concentrações (40 mmol^{-1} ; 120 mmol^{-1}) e outros componentes de BGE, como o TRIS em meio acidificado com HCL, foram testados. Apesar de atingirem o mesmo pH, não foram bem sucedidos em separar esses analitos críticos. Demais condições experimentais foram ajustadas experimentalmente. Além disso, uma vez que a velocidade de migração do imidazol é maior que os demais analitos e que não foram encontrados sinais correspondentes a ele nas amostras, ele foi então selecionado como padrão interno.

FIGURA 21 - Curva de mobilidade efetiva da creatinina, imidazol e aminoácidos essenciais (em escala de cinza).



Fonte:(MOREIRA et al., 2023a)

Na intenção de diminuir o tempo de corrida, foi testado o programa de injeção pelo lado mais próximo ao detector (*Short-end Injection Programm, SEIP*), na qual as amostras são injetadas pelo *outlet* a partir da aplicação de uma pressão negativa e da inversão da voltagem, ainda que o sistema eletroforético se mantém operando em polaridade normal. Após ajustes na injeção e demais parâmetros instrumentais foi verificada a seletividade da creatinina em relação aos demais analitos. Apesar da detecção direta ser mais seletiva do que a indireta pois é exclusiva à compostos cromóforos, por se tratar de uma matriz como a urina outros interferentes podem estar presentes. Sendo o método baseado na detecção direta dos analitos, foi possível a obtenção do espectro UV de cada pico registrado no eletroferograma, o que conseqüentemente permite a análise da pureza de pico.

A pureza de pico indica se há mais de um composto influenciando na intensidade de tal, logo, o valor foi calculado para o sinal de creatinina em um QC de urina (Tabela 11). Os parâmetros foram definidos da seguinte forma: Limiar de pureza, 995; Limiar de intensidade, 5 mAU; Ou seja, para qualquer sinal acima de 5 mAU de intensidade, só é considerado dentro da faixa de pureza aquele que apresentar um fator de pureza maior ou igual a 995. Toda a área do pico foi selecionada para cálculos de pureza. Todos os espectros de UV registrados para o pico de creatinina foram considerados para o cálculo dos espectros médios e subsequente cálculo do fator de pureza. Todos os cálculos foram

realizados usando ferramentas específicas do software Agilent ChemStation Rev. B. 04.03. Esses cálculos foram realizados independentemente em 6 eletroferogramas de QC de urina. Como resultado, não foram encontrados traços de impureza em nenhum eletroferograma verificado.

Tabela 11 - Pureza de Pico calculada para a creatinina urinária.

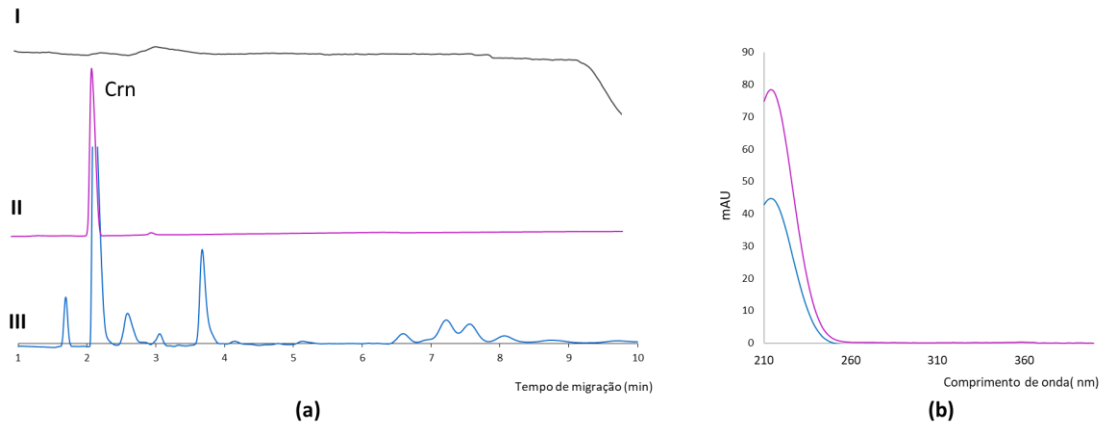
Fator de pureza	999.6 ± 0.12
Varição do tempo de migração	(2.01 ± 0.01) ; (2.30 ± 0.01)
Total de espectros UV registrados	31 ± 3
Intensidade de pico (mAU)	117.75 ± 17.4

Número de réplicas (n) = 6

Fonte: MOREIRA *et al.*, 2023.

Adicionalmente, foi realizado o estudo de matriz em busca de demais interferentes, submetendo uma amostra de urina sintética ao método analítico (Figura 22). Conforme esperado, apesar da matriz conter uma carga iônica significativa, cuja força iônica pode gerar as perturbações na linha base, compostos inorgânicos aniônicos e catiônicos que não absorvem na região do UV não serão registrados no eletroferograma. Assim como a ureia, composto neutro que não migrará devido ao baixo EOF Comparando com a amostra acrescida de padrão e com a amostra real, na região próxima ao tempo de migração de 2 minutos, nota-se que apenas o sinal da creatinina é registrado. Os espectros de UV do pico do padrão de creatinina está condizente com o da UCrn, confirmando que o analito está isolado das demais moléculas.

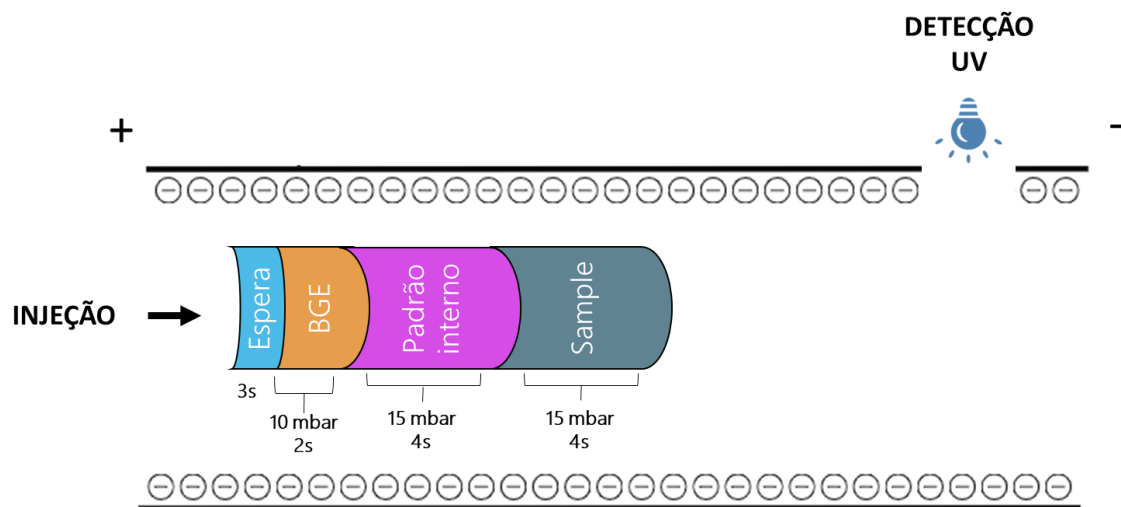
Figura 22 – Estudo do efeito de matriz para análise quantitativa de creatinina. (A) Eletroferogramas resultantes da análise de (I) matriz sintética; (II) matriz adicionada de padrão de creatinina; (III) QC de urina. (B) Espectros UV dos picos de creatinina em solução padrão (em lilás) e em urina (em azul).



Fonte: (MOREIRA et al., 2023a)

A fim de tornar o processo analítico mais automatizado, foi implementado neste trabalho um programa de injeção de amostras em sequências, chamado de injeção “*on-column*” na qual o padrão interno é adicionado à amostra automaticamente pelo sistema de eletroforese sem que haja a necessidade de misturar essas soluções antes de levá-las ao equipamento. São adicionados em sequência: amostra, padrão interno e um plug de BGE, seguido de um tempo de espera. (Figura 23).

FIGURA 23 - Representação ilustrativa do sistema de injeção com adição de padrão interno “*on-column*”.



Fonte: Elaborado pela autora (2024).

O BGE e o tempo de espera ao final têm a função de auxiliar na entrada das demais soluções no capilar evitando que os mesmos voltem para o *vial* por ação da gravidade, acomodando-os no capilar até que seja feita a troca automática para o *vial* de aplicação de voltagem. Nesse aspecto, a injeção de BGE com essa finalidade melhorou a reprodutibilidade do método em relação à injeção convencional (Tabela 12). Considerando que a viscosidade de cada urina pode ter variações, essa pode ser uma alternativa para melhorar a reprodutibilidade analítica de amostras com essas características, mesmo que não seja injetado nenhuma outra solução adicional. Durante os testes, foram programadas injeções de metanol e água entre corridas que utilizaram esse protocolo para verificar a contaminação entre corridas. Nenhum sinal de comprometimento de soluções foi identificado.

TABELA 12 - Dados comparativos de reprodutibilidade a partir de injeção convencional e “on-column” em termos das áreas e tempo de migração de Creatinina urinária (UCrn) e Imidazol (IS, 20 mmol L⁻¹).

	Média e desvio padrão		RSD (%)	
	uCrn	IS	uCrn	IS
Injeção convencional				
Área (mAU)	823 ± 187	688 ± 177	22.8	25.8
Tempo de migração (min)	2.09 ± 0.01	1.49 ± 0.01	0.61	0.90
“On-column”				
Área (mAU)	897 ± 26.9	801 ± 20.9	2.99	2.61
Tempo de migração (min)	2.07 ± 0.01	1.51 ± 0.01	0.80	0.53

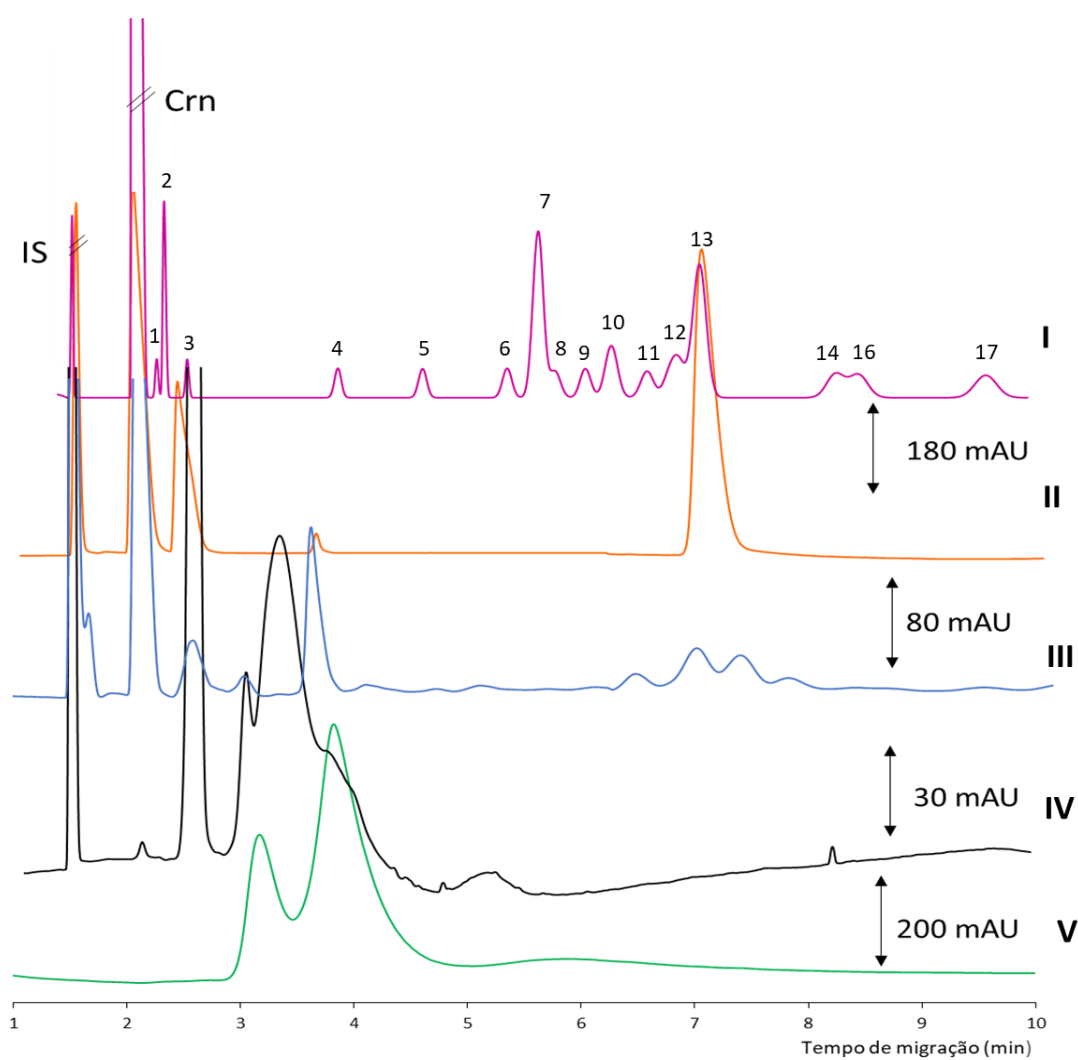
Número de réplicas (n) = 6

Fonte: MOREIRA *et al.*, 2023.

Uma vez que a creatinina também é um biomarcador em soro sanguíneo, a aplicabilidade do método foi avaliada nessas amostras (Figura 24). A partir dos eletroferogramas resultantes, temos que alguns aminoácidos são separados completamente dos demais, principalmente na região entre 2 e 6 minutos, de acordo com simulação. No entanto, visto que o foco em relação a esse grupo é a análise qualitativa, o

perfil obtido para as amostras de urina mesmo com sobreposição de sinal é suficiente para cumprir a proposta. A amostra de soro apresentou poucos sinais associados a um perfil de aminoácidos e mostrou um sinal intenso consequente de componentes da matriz, mesmo após diluição em água 1:100 (v/v)

FIGURA 24 - Eletroferogramas da aplicação do método em (I) simulação no software Peakmaster®; (II) padrões de creatinina e alguns aminoácidos; (III) QC de urina; (IV) QC de soro sanguíneo e (V) matriz artificial de soro sanguíneo.



IS, *internal standard* (Imidazol, 20 mmol L⁻¹); Crn, (Creatinina, 20 mmol L⁻¹); (1) Lisina; (2) Histidina; (3) Arginina; (4) Glicina; (5) Alanina; (6) Isoleucina; (7) Leucina + valina; (8) Serina; (9) Threonina + asparagina (10) Tryptophano; (11) Metionina; (12) Glutamina + Ácido Glutâmico; (13) Fenilalanina + Tirosina; (14) Prolina; (15) Cisteina; (16) Ácido aspártico. Eletroferograma da mistura de padrões: Histidina 5 mmol L⁻¹; Glicina 5 mmol L⁻¹ e fenilalanina 5 mmol L⁻¹;

Fonte: MOREIRA *et al.*, 2023.

O método de doseamento de creatinina foi aplicado nas 100 amostras de urina que totalizam o grupo teste e controle para COVID-19 considerando análise de três réplicas autênticas de cada amostra. As seqüências analíticas foram organizadas de forma que cada grupo de 100 amostras diferentes fossem analisadas de forma randômica intercaladas com: (a) curvas de calibração por adição de padrão interno em cinco níveis de concentração (1.0; 5.0; 10.0; 20.0; 30.0 mmol L⁻¹) e Imidazol fixo em 10 mmol L⁻¹; (b) amostra controle; (c) água como branco. Ao intervalo de cada batelada de 100 amostras, os eletrodos de platina e demais peças removíveis foram higienizadas com isopropanol em banho ultrassônico por 15 minutos, sendo novamente acopladas no equipamento e alinhadas. 10 amostras de soro sanguíneo também foram analisadas em triplicatas com curvas de calibração por padrão interno em cinco níveis de concentração (0.1; 0.5; 1.0; 2.0 e 5.0 mmol L⁻¹) e Imidazol fixo em 5 mmol L⁻¹. Os modelos de calibração e demais parâmetros estatísticos são descritos na Tabela 13.

TABELA 13 - Modelos de regressão linear por adição de padrão interno e parâmetros de performance analítica para doseamento de creatinina em urina e soro sanguíneo expressos em valores de média e desvio padrão. RSD em parênteses.

Parâmetros	UCrn	SCrn
Coefficiente linear	0.047 ± 0.049	0.023 ± 0.089
Coefficiente angular	0.228 ± 0.003	0.623 ± 0.034
Falta de ajuste	F _{cal} = 1.720 < F _{crit} = 2.991	F _{cal} = 0.036 < F _{crit} = 3.708
Significância da Regressão	F _{cal} = 6829 < F _{crit} = 4.195	F _{cal} = 323.7 < F _{crit} = 4.667
<i>r</i>	0.997	0.980
LOD (mg dL ⁻¹)	2.85 ± 1.78	0.04 ± 0.02
LOQ (mg dL ⁻¹)	8.24 ± 5.93	0.17 ± 0.06
Precisão (Crn, mg dL ⁻¹)	196 ± 4.78 (2.43)	0.62 ± 0.15 (14.0)
Exatidão (%)	82.4 ± 11.6	83.8 ± 12.2

Falta de ajuste: F_{crit} uCrn (v₁ = 3, v₂ = 25); F_{crit} sCrn (v₁ = 3, v₂ = 10);

Significância da Regressão: F_{crit} uCrn (v₁ = 1, v₂ = 28); F_{crit} sCrn (v₁ = 1, v₂ = 13);

Precisão: uCrn (n = 55 QCs); sCrn (n = 3 QCs); Exatidão (%): n = 3

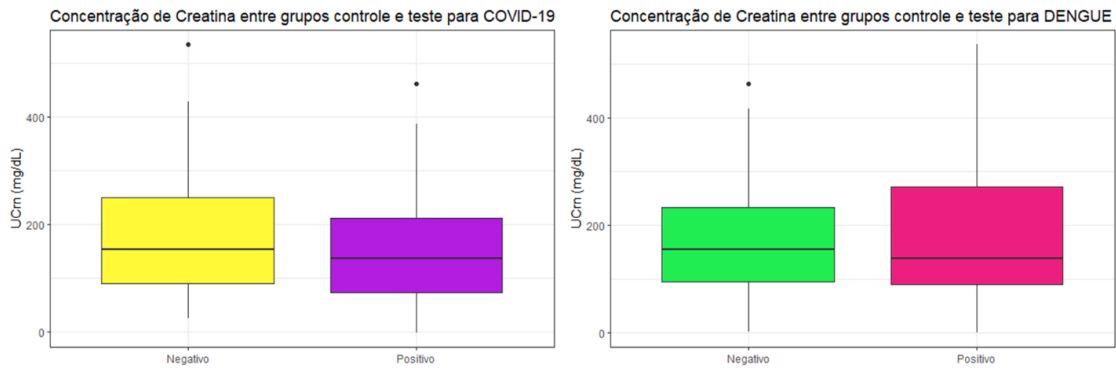
Fonte: MOREIRA *et al.*, 2023.

Os modelos construídos se mostraram aplicáveis visto que não apresentaram falta de ajuste dentro de um intervalo de confiança de 95% além de serem homocedásticos e normais. A precisão de 2.43 % para urina e 14.0% está dentro dos valores recomendados pelo ICH (*International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use*) e Anvisa para validação de métodos bioanalíticos na qual os erros devem se manter em até 20%. Para o soro sanguíneo, acredita-se que a baixa concentração do analito na matriz leva à valores de desvio mais elevados quando comparado com urina.

A Tabela 14, disposta no fim desta seção, contém os resultados de quantificação. Todos os valores encontrados estão condizentes com os valores de referência para urina e soro sanguíneo em mg dL^{-1} de fluido, também apresentando valores com precisão dentro da recomendação das agências reguladoras supracitadas. 10 amostras de urina selecionadas randomicamente foram submetidas ao método clássico por titulação com ácido pícrico em meio alcalino (*Jaffe's Reaction*) realizado pelo laboratório de análises clínicas parceiro. Os valores obtidos por CZE-UV geraram valores cerca de 30 % acima dos valores pelo método convencional, no entanto, o procedimento por titulação é influenciado por diversos fatores e envolve preparo de amostra extenso que justifica os valores subestimados.

Os valores de creatinina obtidos para o grupo controle de COVID-19 (RT-PCR negativo) mostraram-se ligeiramente elevados em relação ao grupo teste (RT-PCR positivo) considerando a distribuição em torno da mediana ($154.31 \text{ mg dL}^{-1}$ e $146.10 \text{ mg dL}^{-1}$, respectivamente), sendo que das amostras positivadas 15 % resultaram em valor de creatinina abaixo do limite de detecção, únicos valores discrepantes em todo o conjunto de dados. De posse do histórico de contaminação por arboviroses dos pacientes que doaram as amostras analisadas, 16 declararam a infecção por dengue. A mesma avaliação foi realizada, também apresentando medianas próximas com o grupo negativo contendo valores ligeiramente elevados ($153.95 \text{ mg dL}^{-1}$ e $137.77 \text{ mg dL}^{-1}$).

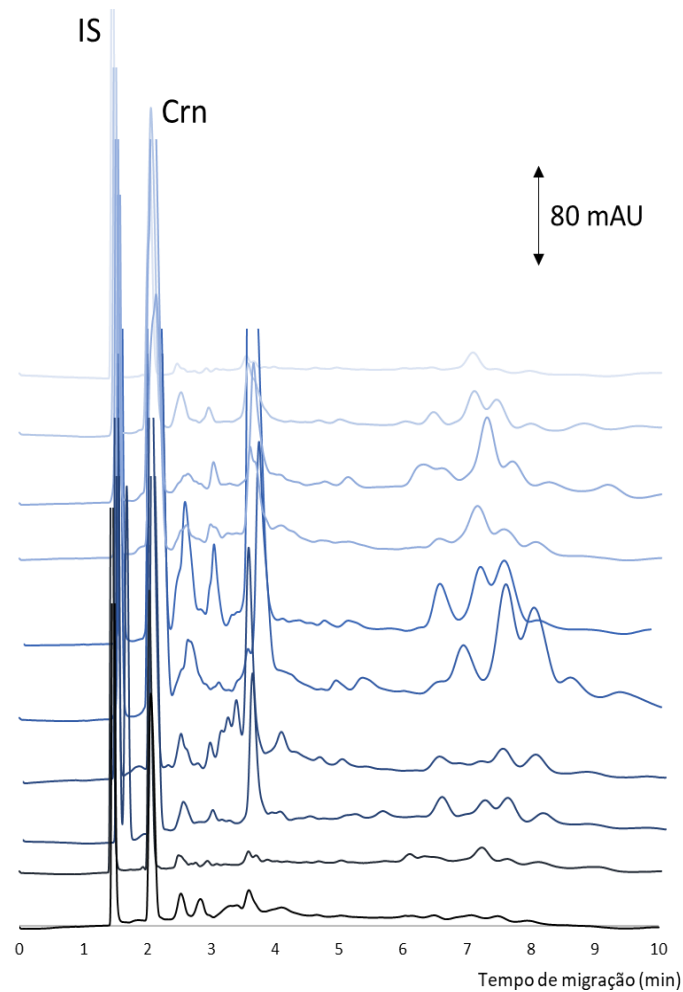
FIGURA 25 - *Box-Plot* da distribuição de concentração de creatinina em urina entre indivíduos saudáveis e com histórico de COVID-19 e Dengue.



Fonte: Elaborado pela autora (2024).

Em relação ao perfil de aminoácidos (Figura 26), semelhante às respostas obtidas pelo método para ácidos orgânicos, o conteúdo do conjunto de dados também tem potencial para construção de modelos de classificação, sendo essa uma etapa do trabalho ainda a ser concluída futuramente. Visto que o protocolo para o uso de eletroferogramas como dado de entrada para uso em aprendizado de máquina já foi otimizado anteriormente, pretende-se submeter esse novo conjunto de dados aos mesmos algoritmos comentados anteriormente na intenção de avaliar se o perfil de aminoácidos é mais propício para discriminação entre esses grupos.

FIGURA 26 - Eletroferogramas do perfil de aminoácidos de 10 amostras de urina aleatórias analisadas por CZE-UV.

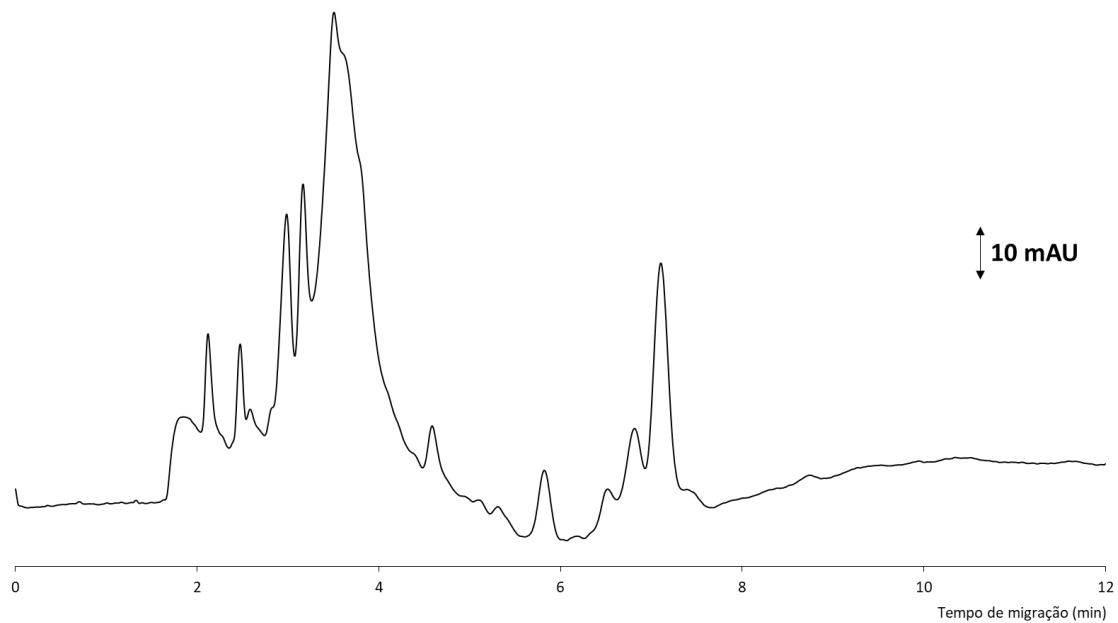


Fonte: MOREIRA *et al.*, 2023.

Esse método também foi utilizado para análise de amostras de saliva, sendo observada a necessidade de ajustes no método e no preparo de amostra a fim de diminuir prováveis efeitos de matriz (Figura 27). Alguns estudos já avaliaram a possibilidade de monitoramento de doenças renais crônicas pela creatinina salivar (LASISI; RAJI; SALAKO, 2016; TEMILOLA *et al.*, 2019), sendo essa uma aplicação do método, além da abordagem que envolve o perfil de aminoácidos, visto que mesmo após ajustes, não haverá definição de pico suficiente para análises quantitativas. Nenhum teste adicional em saliva foi realizado nesta parte do trabalho.

O desenvolvimento do método foi publicado contendo os resultados de quantificação de creatina em urina e soro sanguíneo em revista científica da área de Química Analítica com Qualis Capes A1 (DOI: 10.1016/j.talanta.2023.124465).

FIGURA 27 - Eletroferograma resultante da análise de aminoácidos em saliva por CZE-UV.



Fonte: Elaborado pela autora (2024)

TABELA 14 - Resultados da quantificação de creatinina em urina e soro sanguíneo a partir de método otimizado por CZE-UV expressos como média de cada triplicata e seus respectivos RSD.

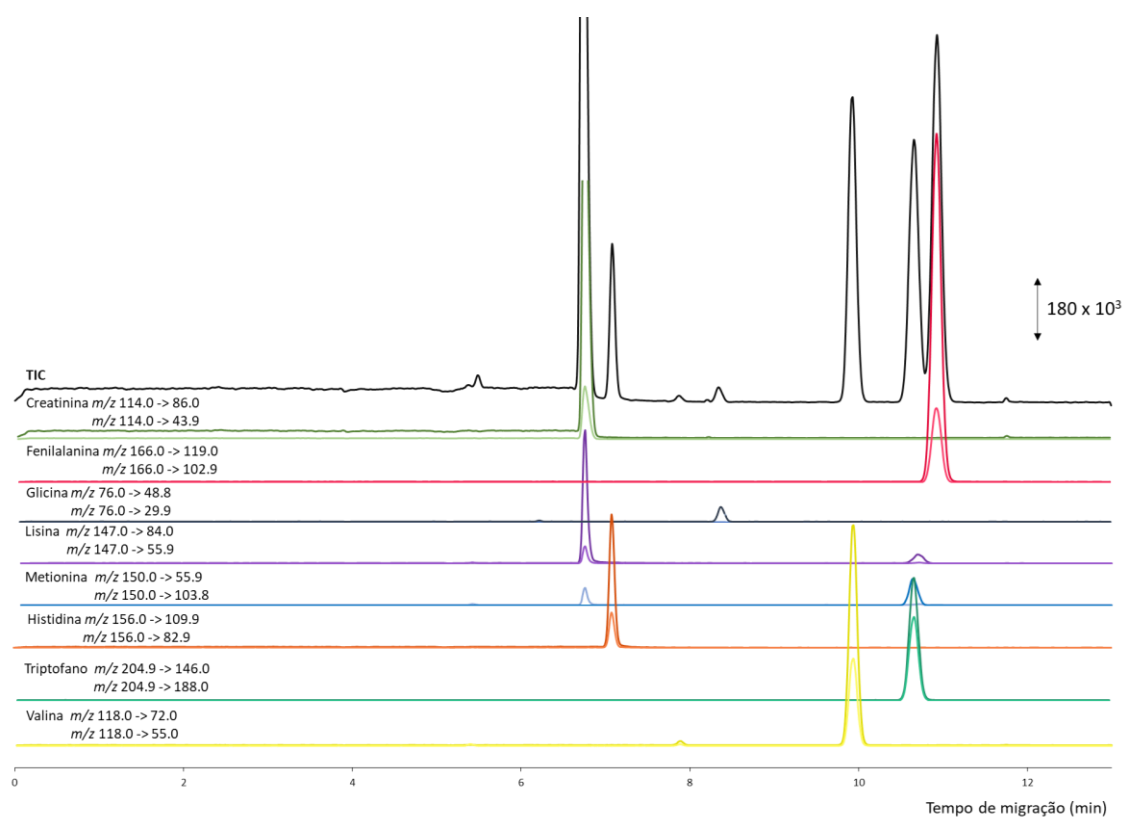
	Urina (mg dL ⁻¹)				Soro Sanguíneo (mg dL ⁻¹)					
	Homens até 40 anos	Homens acima de 40 anos	Mulheres até 40 anos	Mulheres acima de 40 anos	Homens	Mulheres				
	44.15 _(11.7)	252.4 _(14.6)	26.64 _(2.67)	201.5 _(4.21)	27.94 _(4.93)	232.0 _(6.70)	24.46 _(10.3)	189.3 _(14.5)	0.132 _(48.3)	0.105 _(35.9)
	60.85 _(5.47)	258.1 _(5.50)	41.74 _(8.71)	208.8 _(4.79)	38.53 _(8.14)	236.5 _(18.8)	42.12 _(2.99)	207.2 _(6.46)	0.308 _(13.5)	0.102 _(27.9)
	79.45 _(2.48)	286.3 _(0.30)	64.15 _(6.48)	222.2 _(11.04)	39.88 _(21.0)	255.5 _(4.49)	44.65 _(10.1)	211.9 _(4.57)	0.375 _(21.8)	0.307 _(7.28)
	91.63 _(4.22)	287.1 _(13.5)	93.57 _(4.61)	223.4 _(5.38)	56.64 _(1.79)	258.8 _(11.0)	45.44 _(12.5)	264.8 _(8.23)	0.622 _(18.3)	0.319 _(44.1)
	98.18 _(6.45)	288.9 _(7.11)	108.0 _(7.76)	238.2 _(6.70)	63.15 _(6.19)	415.6 _(5.64)	58.10 _(7.93)	-	0.695 _(16.9)	0.477 _(20.03)
	106.4 _(0.76)	325.3 _(7.12)	110.0 _(38.5)	258.2 _(21.6)	71.33 _(6.24)	-	67.67 _(2.62)	-	-	-
	110.6 _(2.84)	337.8 _(4.48)	114.9 _(6.28)	265.9 _(0.29)	100.6 _(5.22)	-	85.14 _(12.7)	-	-	-
	153.0 _(53.6)	340.0 _(2.77)	119.4 _(3.00)	310.9 _(1.32)	118.1 _(6.28)	-	93.69 _(11.7)	-	-	-
	153.9 _(5.53)	375.8 _(1.26)	143.8 _(11.4)	361.2 _(7.40)	140.4 _(4.26)	-	103.6 _(0.66)	-	-	-
	154.3 _(46.8)	384.7 _(7.29)	152.6 _(6.07)	386.9 _(6.21)	140.8 _(19.4)	-	113.8 _(11.7)	-	-	-
	193.8 _(5.19)	401.1 _(9.14)	186.2 _(14.7)	462.8 _(15.17)	146.1 _(10.9)	-	136.5 _(21.5)	-	-	-
	212.7 _(8.99)	-	189.1 _(11.2)	511.2 _(11.1)	153.7 _(4.41)	-	138.9 _(4.09)	-	-	-
	224.8 _(5.64)	-	192.3 _(12.1)	535.2 _(2.77)	164.0 _(18.5)	-	149.7 _(9.58)	-	-	-
	230.5 _(7.56)	-	195.2 _(4.24)	-	212.1 _(6.76)	-	168.7 _(4.19)	-	-	-
Média (mg dL ⁻¹)	209.3		222.6		144.7		120.6		0.245	0.427
Valores de Referência	24 - 392		22 - 238		16 - 327		15 - 287		0.7 - 1.3	0.5 - 1.0

Fonte:(MOREIRA et al., 2023a)

5.2.4 Análise de aminoácidos em urina por CZE-MS

Visando a obtenção de análises seletivas do perfil de aminoácidos de urina, com possibilidade de identificação e quantificação, um método por CE-MS/MS foi proposto. Utilizando um sistema de massas triplo quadrupolo, alguns aminoácidos foram identificados a partir do monitoramento em dois níveis de transição em modo positivo (Figura 28).

FIGURA 28 - Resultados MRM das análises de aminoácidos realizadas por CZE-MS/MS



Fonte: Elaborado pela autora (2024).

Em relação aos dados quantitativos gerados, os modelos de regressão falharam ao descrever relações adequadas para doseamentos destes analitos (Tabela 15). Em geral, a maior parte dos modelos não obedeceram aos pressupostos de homoscedasticidade e normalidade a 95 % de confiança. As regressões lineares, apesar de estarem ajustadas em

alguns casos, não representam dados bem correlacionados e todos falham no teste de significância da regressão

TABELA 15 - Modelos de regressão linear por adição de padrão interno para aminoácidos em urina por CZE-ESI-QQQ.

Analito	<i>r</i>	Falta de ajuste	Significância da regressão
Creatinina	0.3968	$F_{\text{cal}} = 1.381 < F_{\text{crit}} = 2.689$	$F_{\text{cal}} = 6.354 > F_{\text{crit}} = 4.13$
Fenilalanina	0.0643	$F_{\text{cal}} = 21.96 > F_{\text{crit}} = 2.689$	$F_{\text{cal}} = 0.523 < F_{\text{crit}} = 4.13$
Histidina	0.4532	$F_{\text{cal}} = 1.507 < F_{\text{crit}} = 2.689$	$F_{\text{cal}} = 7.509 > F_{\text{crit}} = 4.13$
Lisina	0.7258	$F_{\text{cal}} = 4.882 > F_{\text{crit}} = 2.689$	$F_{\text{cal}} = 7.324 > F_{\text{crit}} = 4.13$
Metionina	0.6372	$F_{\text{cal}} = 7.635 > F_{\text{crit}} = 2.689$	$F_{\text{cal}} = 9.918 > F_{\text{crit}} = 4.13$
Triptofano	0.6889	$F_{\text{cal}} = 4.645 > F_{\text{crit}} = 2.689$	$F_{\text{cal}} = 6.659 > F_{\text{crit}} = 4.13$
Valina	0.6838	$F_{\text{cal}} = 1.588 < F_{\text{crit}} = 2.689$	$F_{\text{cal}} = 4.841 > F_{\text{crit}} = 4.13$

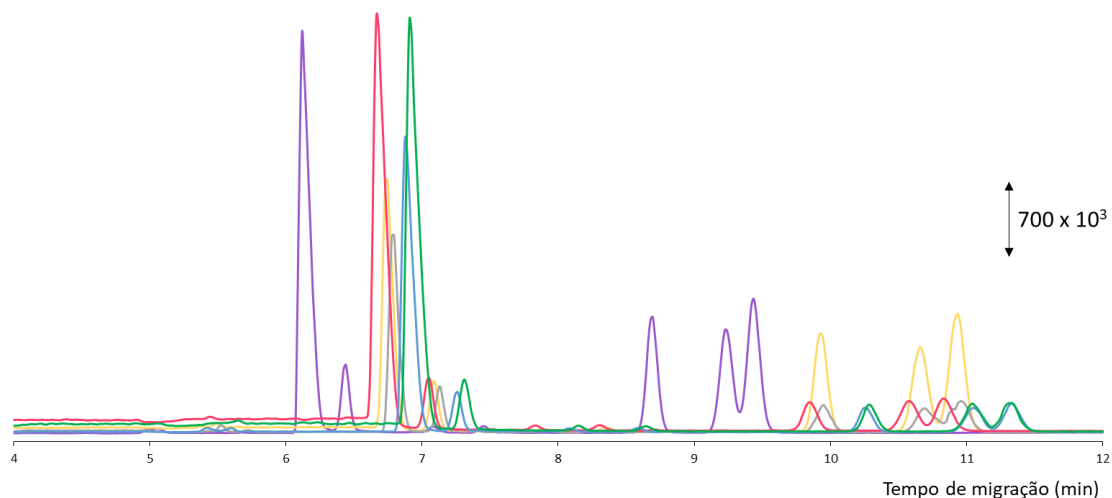
Falta de ajuste: $F_{\text{crit}} (v_1 = 4, v_2 = 30)$; Significância da Regressão: $F_{\text{crit}} (v_1 = 1, v_2 = 348)$;

Faixa linear: 1 – 40 $\mu\text{mol L}^{-1}$

Fonte: Elaborado pela autora (2024).

Os eletroferogramas apresentaram baixa reprodutibilidade (Figura 29), quanto à área e quanto ao tempo de migração dos analitos, mesmas observações foram feitas também nos resultados do tipo *scan*. mostrando que o método necessita de melhorias. Neste sentido, a partir desses resultados pretende-se adaptar o protocolo de condicionamento entre corridas otimizado para aliado à injeção de soluções em sequência otimizado CZE-UV, e ajustar demais condições que se mostrarem pertinentes para melhorar a precisão do método, incluindo parâmetros de ionização de fragmentação dos analitos, além de ajustar as diluições das amostras e prever outra faixa analítica de concentração.

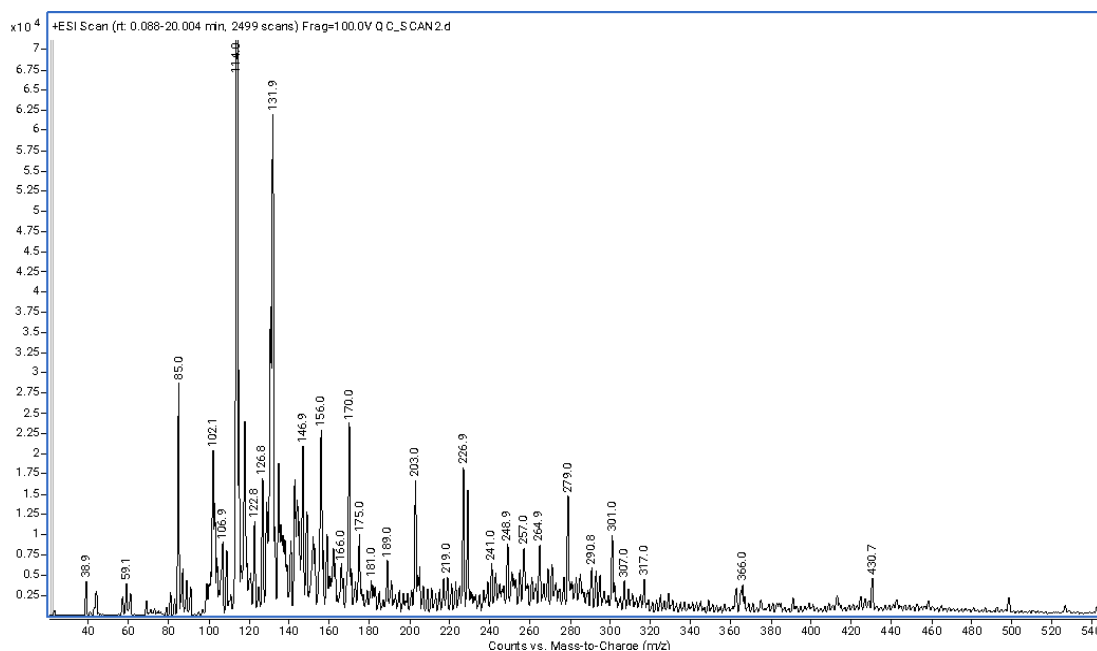
FIGURA 29 - Eletroferogramas totais da análise da mistura de padrões de aminoácidos por CZE-MS/MS na concentração de $10 \mu\text{mol L}^{-1}$.



Fonte: Elaborado pela autora (2024).

Os espectros de massa do modo *scan* foram também submetidos ao fluxograma de processamento inserido no software MassHunter Qualitative Analysis B10.0 para descoberta de compostos a partir da deconvolução dos eletroferogramas, sendo os espectros de massas obtidos no modo *profile* (Figura 30). Esse processamento inicial indicou uma média de identificação de cerca de 126 compostos para eletroferogramas de QC de urina. Ainda que esse resultado esteja em uma provável superestimação, pretende-se ainda fazer uma tentativa de identificação desses compostos, sabendo que esses resultados de m/z estão associados a um sistema de eletroforese capilar que implica em uma seletividade em relação a quais compostos têm possibilidades de estarem presentes.

FIGURA 30 - Espectro de massa do tipo *profile* de um QC de urina obtido a partir do método por CZE-MS.



Fonte: Elaborado pela autora (2024).

5.3 CONSIDERAÇÕES GERAIS

Os métodos por CZE-UV voltados para grupos de ácidos orgânicos mostraram-se aplicáveis à modelos de classificação a partir da associação com ferramentas de ciências de dados no caso de análises de urina. Sendo uma matriz complexa com grande quantidade de metabólitos, estudos quantitativos e de identificação de ácidos orgânicos por CZE-UV é uma limitação. Nesse caso, a adaptação destes métodos para medidas por CZE-MS ainda é uma perspectiva para este trabalho. A partir dos resultados obtidos, na qual foi descrito um modelo de classificação entre grupos teste e controle para a COVID-19, acredita-se que há potencial de identificação de biomarcadores dentro desse grupo de moléculas que pode futuramente trazer novas informações sobre a doença, sendo o desenvolvimento de métodos por CZE-MS uma estratégia para esse estudo investigativo/quantitativo.

A maior seletividade do método por detecção direta utilizado permitiu com sucesso o doseamento de marcados importante, como a creatinina. A partir dessa avaliação, acredita-se que o desenvolvimento de métodos por CZE-UV para

quantificação de biomarcadores com características físico-químicas que permitem gerar resultados seletivos seja uma alternativa interessante para rotinas laboratoriais, ainda que na maior parte dos casos o uso da espectrometria de massas não seja dispensado. Em relação aos aminoácidos, imagina-se que seja possível a discriminação entre grupos utilizando esse grupo de marcadores a partir de modelos preditivos, sendo essa uma etapa a ser realizada futuramente. E a exemplo do que foi discutido para o método para ácidos orgânicos, essa triagem pode indicar que estudos elaborados dentro desse grupo de compostos seja relevante.

Em relação à saliva, métodos utilizando instrumentação mais simples como a CZE-UV é uma possibilidade para estudos de metabólitos específicos. A partir da análise de amostras nas quais informações dos pacientes são disponibilizadas de modo que grupos teste e controle sejam definidos, a utilização desses métodos pode ser interessante tanto para construção de modelos classificatórios utilizando o perfil do eletroferogramas como para estudos segundo a metabolômica alvo, sobretudo o método para SCFA, cujos metabólitos são mais estudados, já sendo estabelecida a importância destes biomarcadores.

Quanto ao uso da eletroforese capilar, destaca-se que uma vez que a técnica permite a análise de matrizes complexas sem que seja necessários preparos de amostra extensos e por não fazer uso de colunas cromatográficas recheadas não há prejuízos em caso de comprometimento, tais métodos também podem se estender a outras aplicações sem restrições quanto ao tipo de amostra. Ainda na área clínica, sabe-se que outros fluidos biológicos também podem ser relevantes para esse tipo de análise, sendo todos os métodos passíveis de adaptação às matrizes biológicas.

Além disso, da mesma forma que um dos métodos descritos nesse trabalho foi adaptado de aplicações em fitoquímicos, muitos dos analitos citados nesse capítulo estão presentes em diversos cosméticos, fármacos, alimentos e bebidas, como por exemplo os ácidos glicólico, láctico, tartárico, butírico, acético e alguns aminoácidos como o triptofano, fenilalanina, metionina. Dessa forma, todos os métodos podem ser adequados à análise de controle de qualidade desses produtos. Por fim, a aplicação da CZE na obtenção de dados de entradas para modelos de predição, introduzidos nesse trabalho, pode também se expandir para diversas áreas.

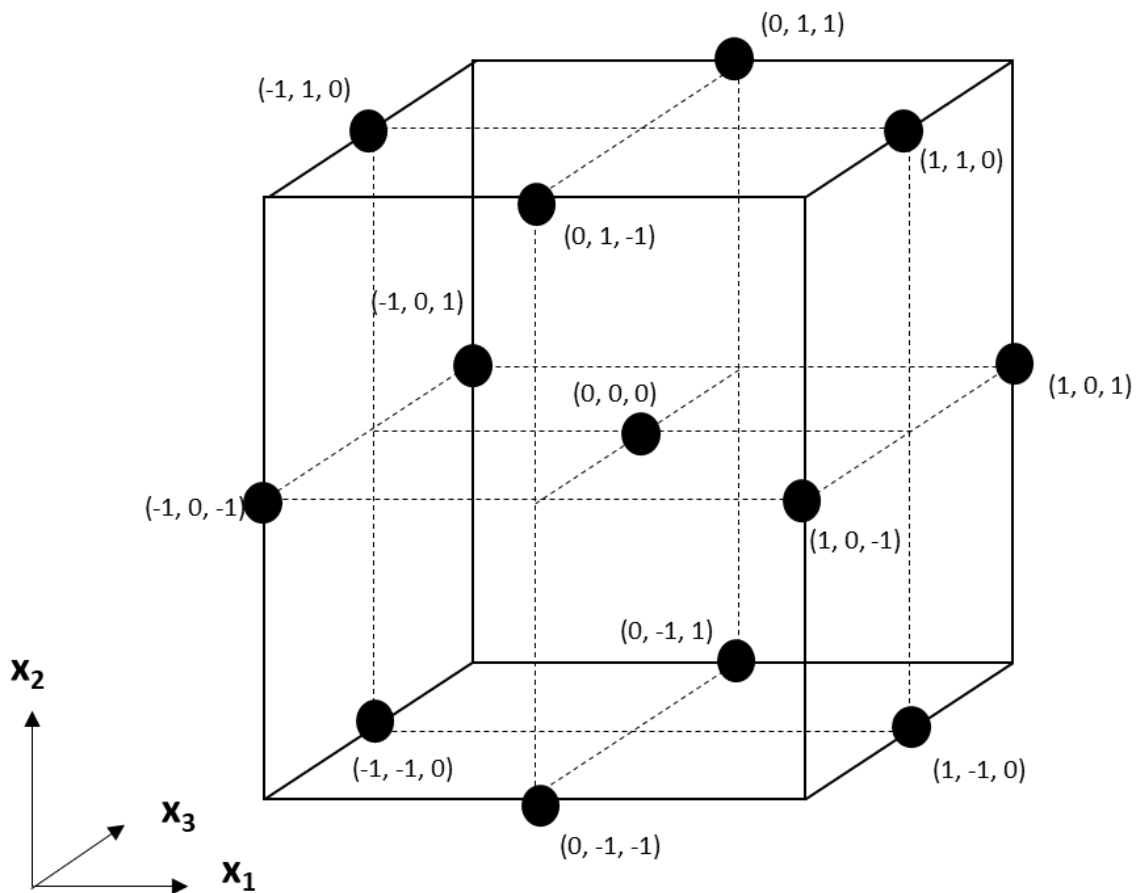
6 METABOLÔMICA GLOBAL POR ESPECTROMETRIA DE MASSAS DE ALTA RESOLUÇÃO

A metabolômica clássica considera a análise abrangente de amostras biológicas na busca por biomarcadores decisivos na diferenciação entre os grupos teste e controle selecionados. Ainda que análises segmentadas estão sendo consideradas neste trabalho, o fluxograma clássico que envolve equipamentos com alto poder de identificação de moléculas em matrizes complexas também foi explorado. Inicialmente, alguns métodos analíticos voltados para obtenção de dados por LC-MS de alta resolução foram testados, nesse sentido um planejamento de experimentos foi proposto para estudar as eventuais alterações nos resultados devido às modificações nos parâmetros de ionização e, ao mesmo tempo, verificar quais parâmetros são mais relevantes, a fim de gerar dados de alta qualidade para a etapa posterior de processamento. Uma vez otimizado, o método também foi aplicado nas 100 amostras pré-selecionadas associadas ao projeto COVID-19.

6.1 DESENVOLVIMENTO E OTIMIZAÇÃO DE MÉTODO ANALÍTICO POR HPLC-ESI-QTOF

Utilizando como ponto de partida, o método proposto por (ALBERICE et al., 2013) e outras referências semelhantes, três parâmetros da ESI foram selecionados para variação em três níveis considerando um planejamento fatorial parcial 3^3 (Box-Behnken). O planejamento de experimentos (*Design of Experiments*, DOE) é uma ferramenta que permite a otimização de processos com experimentos controlados na qual uma superfície de resposta (*Response Surface Methodology*, RSM) é construída a partir de um modelo quadrático polinomial para avaliação da interação das variáveis que se mostrarem estaticamente significativas. O Box-Behnken é um planejamento parcial que resulta em 15 experimentos incluindo uma triplicata no ponto central (Figura 31), dessa forma os dados gerados são suficientes para construção de um modelo quadrático com dados para verificação da qualidade de ajuste do modelo.

FIGURA 31 - Representação do planejamento fatorial 3^3 tipo Box-Behnken com fatores codificados.



Fonte: (MOREIRA et al., 2023b)

Uma vez que os experimentos foram realizados em sistema *Dual Jet Stream* na qual permite a aplicação de voltagem já na fonte de ionização (*Nozzle Voltage*, V), esse parâmetro foi selecionado como um fator de variação. A *Dual Jet Stream* é baseada em uma segunda camada de gás de secagem que envolve a saída do fluxo de eluição do HPLC e auxilia no direcionamento das partículas do ESI na entrada do analisador. Os valores associados a essas duas trajetórias de gás de secagem (temperatura e vazão) não foram selecionadas para evitar que as corridas fossem realizadas sem estabilidade. Desse modo, a pressão de nebulização foi o segundo fator selecionado. Por fim, a voltagem de fragmentação aplicada já no fim do capilar foi o terceiro parâmetro inserido no planejamento, sabendo que essa voltagem é decisiva em relação a integridade da molécula que chegará no detector. O intervalo de valores considerados para avaliação desse

parâmetro foi escolhido de modo a se obter uma informação abrangente também com base em referências relevantes (Tabela 16).

TABELA 16 - Níveis associados a cada experimento descrito pelo planejamento Box-Behnken.

Experimento	X ₁ Nozzle Voltage (V)	X ₂ Nebulizador (psig)	X ₃ Fragmentação (V)
1	0	30	175
2	500	30	175
3	0	50	175
4	500	50	175
5	0	40	100
6	500	40	100
7	0	40	325
8	500	40	325
9	250	30	100
10	250	50	100
11	250	30	325
12	250	50	325
0_R1	250	40	175
0_R2	250	40	175
0_R3	250	40	175

Fonte: (MOREIRA et al., 2023b)

Duas respostas foram avaliadas a partir dos resultados obtidos: (a) a qualidade do perfil do cromatograma de íons totais (*Total Ion Chromatogram*, TIC) considerando intensidade de sinal e resolução de picos, uma vez que cromatogramas estão sendo utilizados também como imagem de entrada para aprendizado de máquina e ciências de dados associadas; (b) a quantidade de compostos encontrados com potencial de serem selecionados como biomarcadores.

6.1.1 Reagentes e soluções

Acetonitrila, isopropanol, água (grau HPLC/MS) e ácido fórmico (grau HPLC) foram adquiridos da Sigma Aldrich (Saint Louis, EUA). As amostras utilizadas para essa foram preparadas conforme descrito na seção 4.3.1. Soluções utilizadas como fase móvel foram preparadas no momento da análise e submetidas à desgaseificação com auxílio de banho ultrassônico por 15 minutos e mantidas em temperatura ambiente. A fase móvel aquosa foi mantida em frasco âmbar durante toda fase analítica.

6.1.2 Instrumentação e métodos

Os experimentos foram realizados em um sistema de cromatografia a líquido de alta eficiência (HPLC) modelo 1260 Infinity II equipado com uma bomba quaternária e acoplado ao espectrômetro de massa de alta resolução modelo 6530 Accurate-Mass QTOF, este conectado a uma fonte de ionização por Eletrospray *Dual Jet Stream Technology* (Dual AJS-ESI), todos os módulos da marca Agilent Technologies (Palo Alto, CA, EUA). O sistema foi controlado pelo software de aquisição de dados MassHunter LC/MS B.08.00.

Foi utilizada uma coluna de fase reversa C18 Infinity Lab Poroshell 120 EC-C18 (4,6 x 100 mm x 2,7 μm), mantida a 30 °C. 5 μL de cada amostra foram injetadas enquanto o amostrador foi mantido a 10 °C. Como fase móvel foram utilizados os solventes: A (água e ácido fórmico 99,9: 0,1, v/v) e B (acetonitrila e ácido fórmico 99,9: 0,1, v/v) a 0,5 mL min⁻¹ em um gradiente de eluição definido da seguinte forma: 0% de solvente B de 0 a 0,5 min, aumentando para 9% até 2 min, 20% em 5 min, 45% em 8 min e atingindo 100% em 9,5 min, onde foi mantida até 11 min. A etapa de reequilíbrio foi mantida até 20 min para retorno à condição inicial (ALBERICE *et al.*, 2013).

Os dados foram inicialmente coletados no modo positivo operado em varredura com faixa de massa de 50 a 1500 a uma taxa de 1,02 espectros/s. O gás de secagem foi mantido a 325 °C com vazão de 11 L min⁻¹ e o gás de revestimento a 350 °C com vazão de 11 L min⁻¹. A voltagem capilar foi ajustada a 4000 V. As massas de referência habilitadas foram 121,0509 (Purina, C₅H₄N₄) e 922,0098 (Hexakis(1H, 1H, 3H-tetrafluoropropoxi)fosfazina, C₁₈H₁₈O₆N₃P₃F₂₄) para ESI (+). 119,0363 (Purina, C₅H₃N₄) e 301,9981 (Tris (2,4,6-trifluorometil)-1,3,5-triazina, C₆N₃F₉OH) considerando ESI (-). A aquisição de dados no modo negativo foi realizada seguindo a mesma condição otimizada definida para o modo positivo. Demais parâmetros foram otimizados a partir de planejamento de experimentos e será discutido na seção de resultados.

6.1.3 Experimental

Os 15 experimentos foram realizados de forma aleatória a partir da análise de um QC de 10 amostras de urina selecionadas randomicamente. Após equilíbrio da coluna cromatográfica, antes de cada experimento o sistema foi submetido apenas à análise de

fase móvel para limpeza e estabilização dos fatores selecionados como variáveis. Toda parte externa da fonte de ionização foi limpa com isopropanol (grau HPLC/MS) antes da análise.

6.1.4 Processamento de dados

Os cromatogramas de íons totais (TICs) resultantes foram trabalhados a partir da extração por *molecular features* seguido pela extração dos Cromatogramas de Pico Base (*Base Peak Chromatogram*, BPC). Informações sobre a extração do pico de creatinina foram obtidas com integração automática do pico (*Extracted Ion Chromatogram*, EIC) considerando o m/z de $[M + H]^+$ 144,0669 em um tempo de retenção (Rt) de 1,8 min previamente determinado pela análise de solução aquosa padrão de creatinina a 10 mg L⁻¹. Foram calculadas áreas totais de TIC em unidades arbitrárias (UAs) por meio de integração automática usando o algoritmo *Agile 2*. Todos os cálculos acima mencionados foram realizados usando o software *Mass Hunter Qualitative Analysis B.10.00* (Agilent Technologies). Não foram aplicados filtros ou outros protocolos de pré-processamento.

Os dados das características moleculares extraídas foram então analisados por RSM considerando uma função polinomial de segunda ordem, de acordo com equação:

$$\hat{y}_i = \beta_0 + \sum_{i=1}^k \beta_i x_i + \sum_{i=1}^k \beta_{ii} x_i^2 + \sum_{i=1}^k \sum_{j=2, i < j < k}^k \beta_{ij} x_i x_j + \varepsilon_i$$

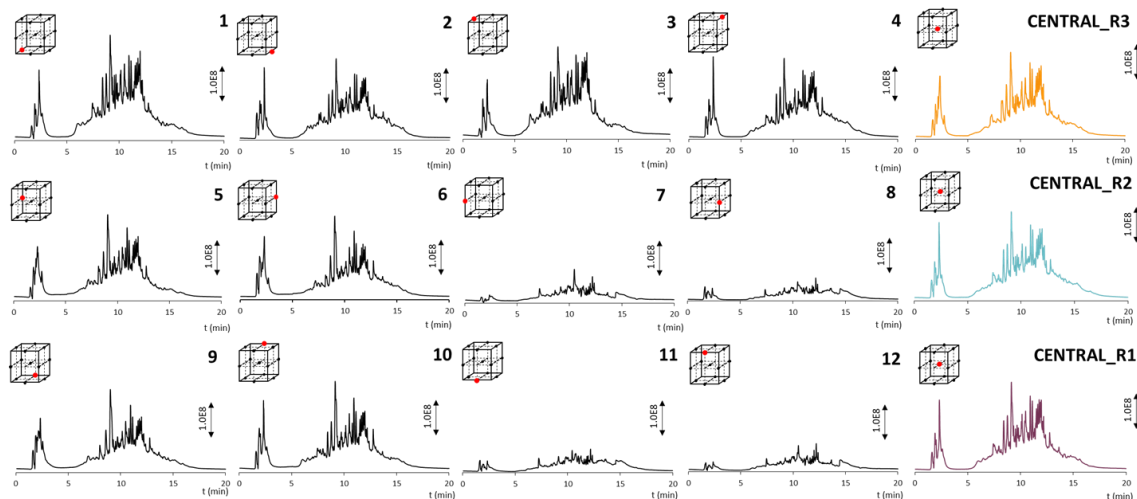
Em que n é o número de variáveis, x_i e x_j são os fatores (variáveis independentes), β representa o coeficiente de cada termo, sendo estes, β_0 o termo constante, β_i termo de primeira ordem, β_{ij} o quadrático ($i = j$) e β_{ij} o termo de interação ($i \neq j$). O resíduo associado ao experimento é expresso por ε . Foi utilizado o Microsoft Excel® 2019 para construção do modelo polinomial, falta de teste de ajuste e para o gráfico da superfície de resposta.

6.1.5 Resultados e discussões

A princípio, uma resposta qualitativa do TIC resultante de cada experimento (Figura 32) foi avaliada. Esta abordagem foi considerada desde que TIC, tanto de respostas de cromatografia líquida quanto gasosa, estão sendo usados recentemente como base de dados de entrada promissora para análise de reconhecimento de padrões como redes neurais artificiais (RNAs) e outras possibilidades de aprendizado profundo (Deep

Learning, DLNs) e aprendizado de máquina (MLs), a exemplo do que foi feito neste trabalho com os resultados por eletroforese capilar.

FIGURA 32 Cromatogramas resultantes da aplicação do planejamento de experimentos



Fonte: (MOREIRA et al., 2023b)

Os experimentos 7, 8, 11 e 12 foram imediatamente descartados como candidatos finais ao método devido à sua baixa intensidade e baixa resolução de pico, consequência de uma alta tensão do fragmentador. Escolhemos uma ampla gama deste parâmetro ESI específico visando um panorama abrangente de sua influência no resultado final, uma vez que deve ser definido em um valor que fragmente a molécula o suficiente para ser analisada, mas não tanto a ponto de destruí-la, como provavelmente ocorreu com a molécula de creatinina na condição de maior voltagem do fragmentador, por exemplo (Tabela 17). Neste contexto, o nível oposto deste fator aplicado nos experimentos 5, 6, 9 e 10 resulta em um cromatograma aprimorado em relação à intensidade, porém não melhor que o fator médio 175 V descrito nos experimentos 1 - 4 e o ponto central.

Entre os resultados promissores pré-selecionados, os experimentos na qual a *nozzle voltage* é aplicada apresentam pequenas variações em comparação com aquele em que esse parâmetro é zerado, considerando a resolução do pico, a intensidade das contagens e a quantidade de informações registradas na região de aumento de gradiente eluição da fase orgânica, entre minutos 5 e 10. O mesmo vale para as variações do fator de pressão do nebulizador. Observamos a quantidade de *molecular features* extraídas dentro da janela selecionada de 5 minutos onde há variações visuais no perfil do TIC, apresentam melhores resultados com fragmentador 175 V e sem *nozzle voltage*.

TABELA 17 - Informações extraídas dos cromatogramas resultantes da aplicação do planejamento factorial

Exp.	X ₁	X ₂	X ₃	Total MF	MF (Rt 5-10 min)	Área Total TIC (AU)	Creatinina (Crn)		
							Área (AU)	Altura de Pico	Rt (min)
1	0	30	175	3917	1388	1.27×10^{10}	0.56×10^7	1.64×10^6	1.866
2	500	30	175	4002	1266	0.49×10^{10}	1.30×10^7	3.70×10^6	1.881
3	0	50	175	3905	1481	1.23×10^{10}	0.58×10^7	1.55×10^6	1.885
4	500	50	175	3346	1008	0.48×10^{10}	1.15×10^7	3.67×10^6	1.886
5	0	40	100	3488	1122	0.81×10^{10}	0.41×10^7	1.22×10^6	1.850
6	500	40	100	2974	1029	0.37×10^{10}	1.29×10^7	3.64×10^6	1.848
7	0	40	325	823	269	0.11×10^{10}	-	-	-
8	500	40	325	640	190	0.13×10^{10}	-	-	-
9	250	30	100	3013	1022	0.44×10^{10}	0.83×10^7	2.59×10^6	1.847
10	250	50	100	3693	1065	0.70×10^{10}	1.05×10^7	2.59×10^6	1.863
11	250	30	325	801	240	0.10×10^{10}	-	-	-
12	250	50	325	751	228	0.19×10^{10}	-	-	-
0_R1	250	40	175	3482	1187	0.79×10^{10}	0.84×10^7	2.98×10^6	1.866
0_R2	250	40	175	3773	1067	0.73×10^{10}	0.81×10^7	2.68×10^6	1.860
0_R3	250	40	175	3973	1284	0.93×10^{10}	0.94×10^7	2.99×10^6	1.875

X₁: Nozzle Voltage (V); X₂: Pressão de nebulização (psig); X₃: Fragmentação (V); Rt = Tempo de retenção; MF = *molecular features*; AU: Unidades arbitrárias.

Fonte:(MOREIRA et al., 2023b)

A metodologia de superfície de resposta foi aplicada considerando o número de *molecular features* extraídas de cada experimento (Tabela 17), uma vez que expressam estritamente o número de compostos distintos encontrados, quanto maior a quantidade de informações, maior a contribuição para o aprimoramento de futuras aquisições, considerando o processamento estatístico e de identificação bioquímica envolvida no fluxograma de trabalho convencional da metabolômica global. Como informação complementar, informações sobre a molécula de creatinina (composto majoritário em urina) foram extraídas, junto o valor de área total dos cromatogramas como uma medida quantitativa de intensidade de sinais. À primeira vista, o número de *molecular features* encontradas pelo software para cada experimento foi de encontro às observações qualitativas supracitadas.

TABELA 18 - Coeficientes resultantes do modelo quadrático

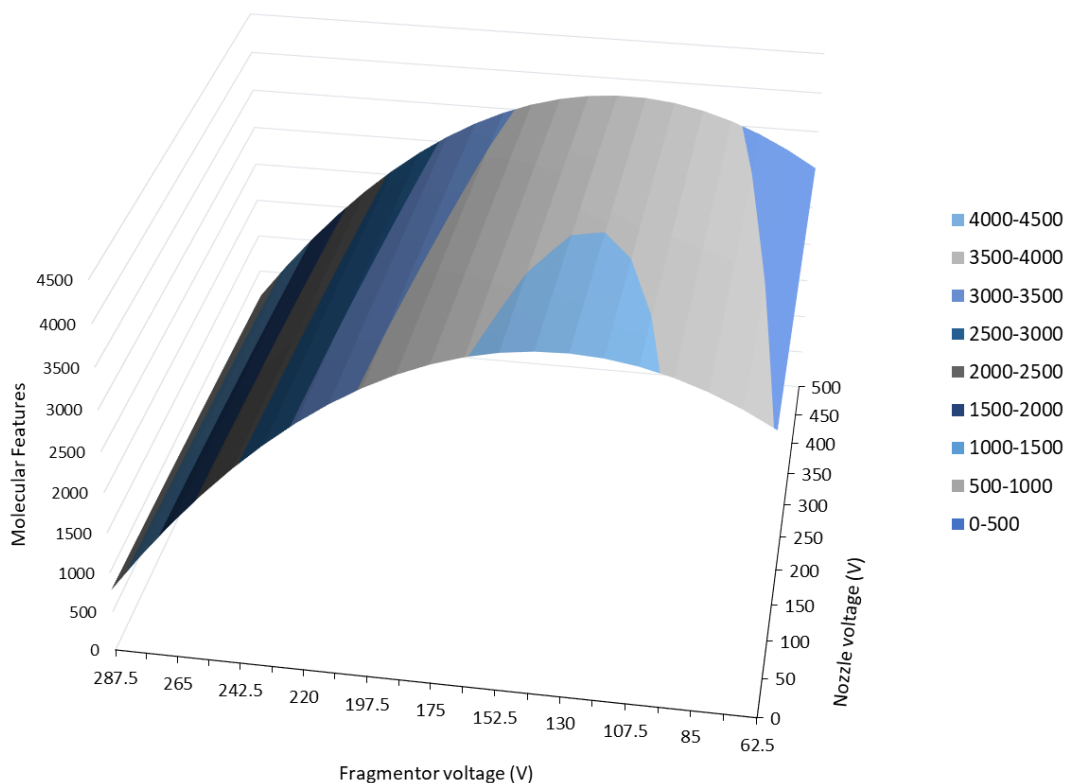
Coeficientes		Erro padrão	Est. <i>t</i>	p-valor
β_0	3742.7	142.5	26.2	0.001448
β_1	-146.4	87.3	-1.68	0.235572
β_2	-4.75	87.3	-0.05	0.961552
β_{3*}	-1269.1	87.3	-14.5	0.004698
β_{11}	-16.7	128.5	-0.13	0.908438
β_{22}	66.5	128.5	0.52	0.656142
β_{33*}	-1744.7	128.5	-13.6	0.00538
β_{12}	-161.0	123.4	-1.30	0.322075
β_{13}	82.7	123.4	0.67	0.571696
β_{23}	-182.5	123.4	-1.48	0.277399

Fonte: (MOREIRA et al., 2023b)

Com base nos valores dos coeficientes resultantes (Tabela 18), a voltagem de fragmentação foi o fator que mais influenciou no resultado final (valor $p < 0,05$), seguido pelo primeiro fator, *nozzle voltage*. De acordo com os coeficientes resultantes, a pressão do nebulizador não está influenciando tanto o sistema quanto os demais dentro dos níveis considerados pelo DOE. Portanto, a pressão do nebulizador não foi considerada relevante em avaliações posteriores. Ainda que os fatores combinatórios que consideram duas variáveis distintas não sejam matematicamente significativos, dado que a aplicação de voltagens é um fator crucial para a ionização molecular, o modelo RSM foi construído para a análise gráfica das pequenas contribuições do uso da *nozzle voltage* para análise de urina e para definição da voltagem de fragmentação (Figura 33). De acordo com os dados descritos na Tabela 18, há uma semelhança de interação nos valores de Área TIC e Aera/Intensidade de pico de creatinina de acordo com a combinação de voltagem do bocal e do fragmentador. Além disso, essa abordagem matemática pode ser interessante para otimizar a ionização de amostras que apresentam maior quantidade de substâncias químicas apolares em sua constituição.

O modelo ajustado não apresentou falta de ajuste ($p\text{-valor} = 0,48 > 0,05$) e significância da regressão ($p\text{-valor} = 7,82 \times 10^{-6} < 0,05$) com 99,9% de variação explicada, todos considerando um intervalo de confiança de 95%. (Tabela 19). A precisão calculada do DOE resultou em um RSD de 6,14% considerando o número de *molecular features* resultantes da triplicata no ponto central.

FIGURA 33 - Superfície de resposta da interação entre a voltagem do fragmentador (X_3) e Nozzle voltage (X_1) considerando o número de moléculas geradas pela aplicação de cada combinação de fatores



Fonte: (MOREIRA et al., 2023b)

TABELA 19 -Análise de variância associadas aos ajustes do modelo quadrático

Parâmetros	Soma quadrática	Graus de Liberdade	Média quadrática
Regressão	120790957.07	9	13421217.45
Resíduos	341740.42	5	68348.08
Falta de ajuste	219819.75	3	73273.25
Erro puro	121920.67	2	60960.33
Total	121132697.5	14	-

Fonte: (MOREIRA et al., 2023b)

A partir da superfície de resposta (Figura 33), nota-se que experimentos que consideram os níveis (-1) e (0) de fragmentação estão dentro da região ótima das condições de ESI. Os experimentos 1, 3, 5, 9, 10 e o ponto central aparecem entre os ótimos. Porém, juntamente com a avaliação qualitativa provida anteriormente, os experimentos 1, 3 e central apresentam-se como melhor ajuste considerando o perfil TIC.

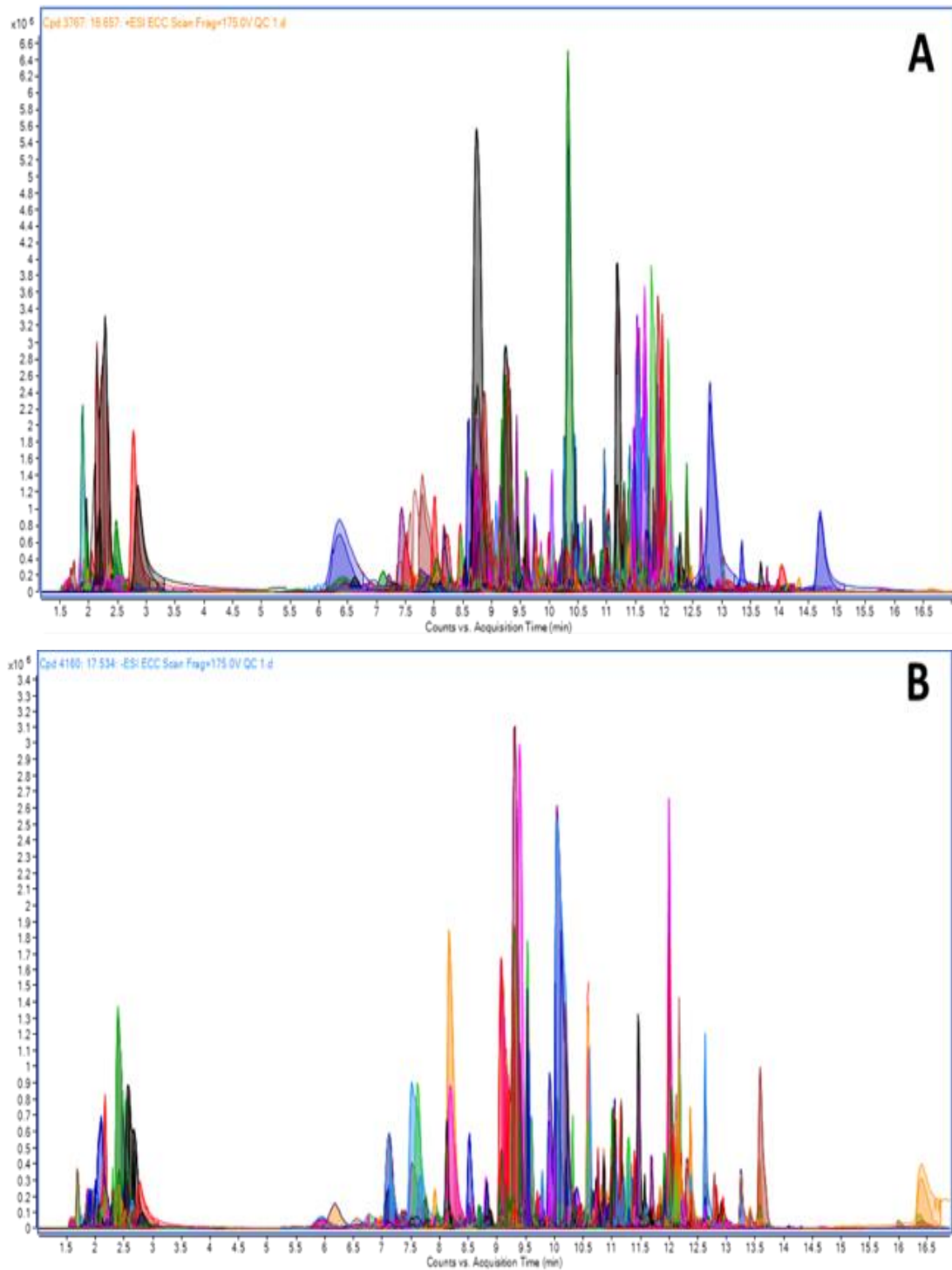
Ao todo, dentre essas opções, o experimento 3 foi selecionado como método otimizado. Todos os valores extraídos são estatisticamente iguais ao experimento 1, considerando o RSD associado, mas dadas as avaliações visuais prévias de melhor definição e resolução do pico, principalmente no início do gradiente de eluição, além de bons valores de área e intensidade de EIC/TIC, foi finalmente definido como o melhor ajuste. É importante ressaltar que não utilizar *nozzle voltage* é uma vantagem interessante, visto que qualquer equipamento pode ser utilizado para reproduzir esta condição, pois nem todos os sistemas MS incluem este acessório.

6.2 APLICAÇÃO DO MÉTODO

O método otimizado foi posteriormente aplicado como protocolo analítico para análise global de 100 amostras de urina em triplicatas autênticas. Semelhante às sequências organizadas nos experimentos por eletroforese capilar, foram organizados 3 grupos de análise contendo 100 amostras diferentes analisadas de forma randômica intercaladas com: (a) curvas de calibração externa de mistura de padrões contendo espécies de grupos distintos (aminoácidos e moléculas de baixo peso molecular, fármacos, ácidos orgânicos e estruturas apolares); (b) amostra controle; (c) água como branco. No intervalo de cada batelada de 100 amostras, a parte da externa da fonte de ionização foram lavadas com isopropanol (grau HPLC/MS), acessórios da fonte e demais peças removíveis foram levadas ao ultrassom imersas em isopropanol (grau HPLC/MS) por 15 minutos. Todo esse processamento foi feito para ambos modos de ionização (ESI + e ESI -).

Extraindo as massas do QC (Figura 34), foi encontrado um valor médio de 3.809 e 3.868, com RSD de 2,27% e 7,29% para modo positivo e negativo, respectivamente. A variação no modo negativo é esperada, uma vez que a ionização neste modo pode ser problemática, no entanto, os valores são aceitáveis para métodos bioanalíticos por cromatografias. Foram gerados 600 TICs correspondentes às análises, sendo contabilizadas as 100 amostras em réplicas autênticas e os dois modos de aquisição. O processamento de dados dessas amostras voltado para seleção de biomarcadores diferenciais entre os grupos teste e controle da COVID-19 ainda é uma das perspectivas para esse trabalho.

FIGURA 34 - Cromatogramas de pico base extraídos para uma amostra controle de urina a partir do método otimizado no modo ESI + (A) e ESI - (B).

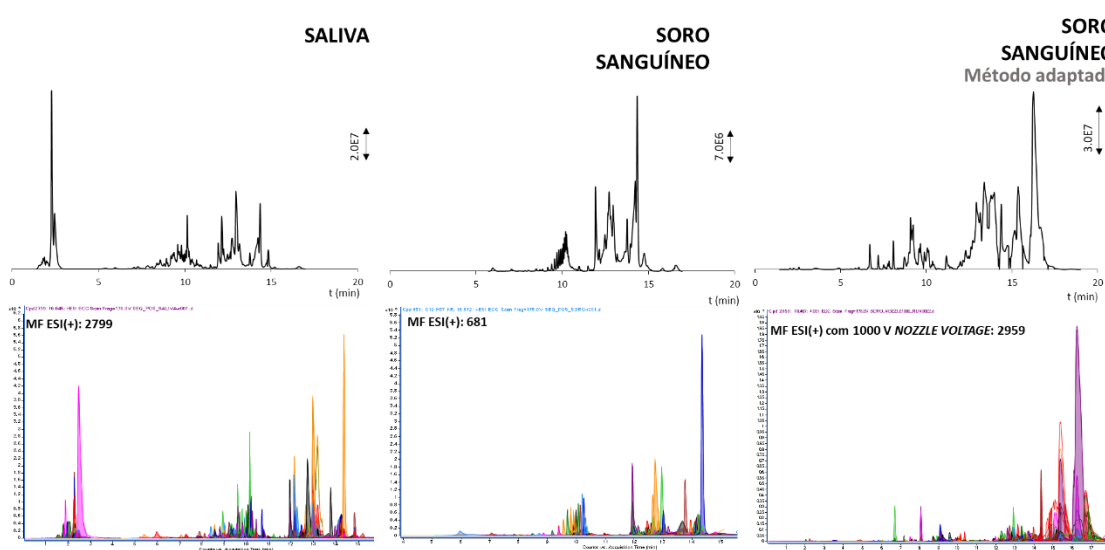


Fonte: (MOREIRA et al., 2023b)

O método proposto também foi aplicado em amostras de soro sanguíneo e saliva humana (Figura 35) com resultados promissores, porém, nesses casos a nova aplicação do DOE pode ser necessária. A saliva e o soro sanguíneo são amostras com maior

viscosidade, portanto a otimização da diluição ou outros tipos de preparo devem ser estudadas, além disso, essas amostras possuem uma quantidade significativa de componentes apolares como hormônios, esteroides, ácidos graxos, vitaminas, entre outros. Então, neste caso poderemos descobrir que a utilização da *nozzle voltage* e associados a outros níveis de tensão do fragmentador juntamente com ajustes de temperatura, pressão e outros parâmetros é mais adequada, uma vez que essas moléculas são mais difíceis de serem ionizadas.

Figura 35 - TIC e BPC resultante da aplicação do método em outras matrizes biológicas.



Fonte: (MOREIRA et al., 2023b)

6.3 CONSIDERAÇÕES GERAIS

O fluxograma de trabalho convencional da metabolômica global a partir do uso da espectrometria de massas de alta resolução envolve um processamento de dados elaborado na qual são extraídas as informações moleculares encontradas pelo analisador de massa de forma a ser construída uma planilha única contendo os dados das amostras para ser então inserida em softwares de reconhecimento de padrão. Os algoritmos irão então processar esses dados e prepará-los para que sejam inseridos em banco de dados que contenham informações químicas passíveis de identificação de biomarcadores. Esse processo exige cautela dos analistas para evitar o comprometimento dos resultados, sendo estes o mais fiel à realidade possível. Por ser dados volumosos e extensos, o processo

costuma ser demorado. Ao longo do desenvolvimento desse trabalho, rotinas semelhantes foram empregadas a outros tipos de conjunto de dados, porém exigem processamentos diferentes quando informações de massas em alta resolução estão envolvidas. Dessa forma, os estudos sobre a metabolômica e as ferramentas de bioinformáticas associadas ainda estão em andamento, visando que sejam obtidos resultados confiáveis a partir dos dados coletados por LC-MS.

A proposta do uso de planejamento de experimentos aplicado na otimização de parâmetros de ionização se mostra interessante para aplicações em diversos casos. Nota-se que ajustes simples, como o desse trabalho, fazem diferença na hora da obtenção dos resultados. O objetivo dessa otimização foi justamente mostrar a potencialidade dessa ferramenta quando associada com espectrometria de massas. Neste contexto, outros parâmetros podem ser envolvidos no DOE, incluindo possibilidades que envolvam colisão e outros sistemas, tudo isso voltado para aplicações diversas. Esse trabalho foi publicado em revista científica da área de Química Analítica com Qualis Capes A1 (DOI: 10.1039/d3ay00094j).

7 APRENDIZADO DE MÁQUINA ASSOCIADO À ESPECTROSCOPIA VIBRACIONAL POR INFRAVERMELHO.

A associação da espectroscopia vibracional com o aprendizado de máquina aplicado às pesquisas relacionadas com saúde é uma prática que tem sido amplamente explorada nas últimas duas décadas. Como Vigo e colaboradores (2023) elaboram em seu recente artigo de revisão, esta abordagem foi considerada primeiro para a avaliação de informações de células cancerígenas na urina, saliva e líquido cefalorraquidiano, posteriormente avançado para outros segmentos de saúde por meio da análise de outras amostras de fluidos corporais (VIGO et al., 2023).

O uso de técnicas espectroscópicas está associado ao volume de informações coletadas e à facilidade de coleta de dados. Para opções de infravermelho (*infrared*, IR), a possibilidade de instrumentação portátil contribui para os recursos vantajosos do uso de espectroscopia como aquisição de dados para modelagem computacional orientada para diagnóstico clínico e pesquisas relacionadas à saúde. As regiões espectrais do infravermelho médio (*Mid-Infrared Spectroscopy*, MIR) variam de 400 cm^{-1} a 4000 cm^{-1} e funcionam como a interação da radiação eletromagnética na região do espectro que compreende o infravermelho com as moléculas contidas na amostra, fornecendo informações químicas qualitativas sobre a composição molecular correspondentes a impressões digitais da amostra, úteis para fins de modelagem computacional.

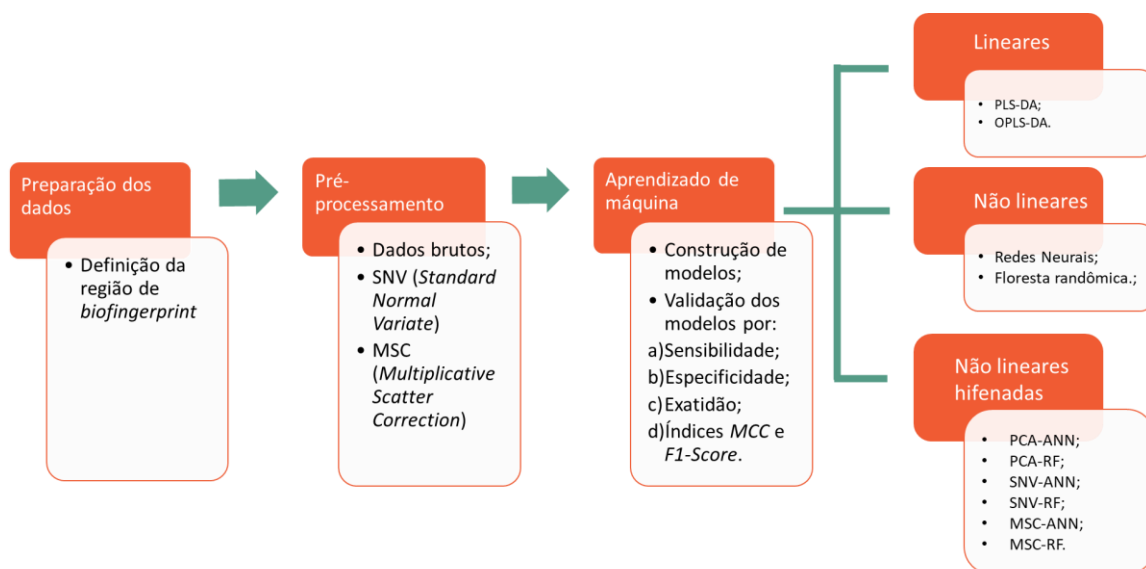
A exemplo da rotina analítica desenvolvida para resultados por CZE-UV discutidos na seção 5, os espectros brutos foram utilizados como base de dados para a construção de um novo modelo binário de classificação capaz de discriminar os grupos teste e controle das amostras de urina associados à COVID-19. De posse do histórico clínico relacionado à incidência de arboviroses das amostras coletadas sob CEP de projeto COVID-19, aquelas em que os voluntários declaram já ter sido acometido por quaisquer uma das arboviroses foram também analisadas por infravermelho a fim de construir um modelo de predição multiclasse entre as infecções de interesse deste trabalho.

7.1 AMOSTRAS COVID-19

As 100 amostras pré-selecionadas e divididas em grupos teste e controle com base nos resultados de RT-PCR para COVID-19 foram analisadas por MIR em três réplicas

autênticas cada. Como resultado, foram gerados 300 espectros, na qual a região com informação bioquímica relevante (*biofingerprint*) foi delimitada, e organizados em uma planilha única como conjunto de dados de entrada para processamento por aprendizado de máquina. Os dados foram randomicamente divididos em dois grupos (treinamento e teste) seguido da construção de modelos lineares e não lineares com ou sem pré-processamento, além da associação de modelos como forma de pré-seleção de variáveis para modelos não-lineares de aprendizado profundo. Essa etapa do trabalho que envolve rotinas quimiométricas foi também realizada em colaboração com outros estudantes do GQAQ. O fluxograma completo do processo de construção de modelos de predição está representado na Figura 36.

FIGURA 36 - Esquema representativo das etapas referentes ao processamento de dados obtidos por CZE-UV a partir de ferramentas de aprendizado de máquina



Fonte: Elaborado pela autora (2024)

7.1.1 Instrumentação e aquisição de dados

Os espectros de Espectroscopia no infravermelho por transformada de Fourier (*Fourier-transform infrared spectroscopy*, FTIR) foram adquiridos usando um espectrômetro Alpha FTIR acoplado a um dispositivo de amostragem universal de Reflexão Total Atenuada (*Attenuated Total Reflectance*, ATR) contendo um cristal de diamante/ZnSe (Bruker, Billerica, MA) e software de aquisição OPUS (Bruker, Billerica,

MA) disponível no departamento de química do Instituto de Ciências Exatas, Universidade Federal de Juiz de Fora.

20 μL de cada amostra preparada de acordo com o protocolo descrito na seção 4 foram colocados diretamente na superfície do cristal do instrumento ATR-FTIR. As amostras foram organizadas em uma sequência aleatória em que foram analisadas cerca de 40 amostras por dia. Foram analisadas as 100 amostras de urina pré-selecionadas em três réplicas autênticas cada. No início de cada lote de análise, os espectros de fundo foram registrados. A superfície do cristal foi limpa com etanol antes de cada medição. Todos os dados foram registrados em temperatura ambiente em sistema aberto usando modo transmitância com número de onda variando de 4000 a 400 cm^{-1} com resolução de 4 cm^{-1} e registrados como uma média de 32 varreduras.

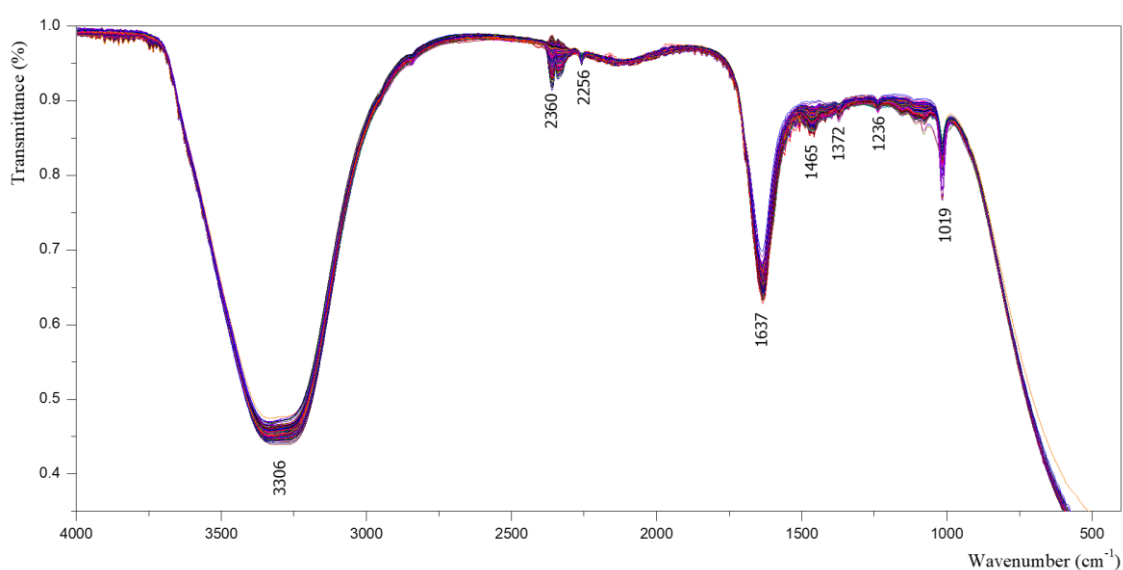
7.1.2 Processamento de dados

Inicialmente, o conjunto de dados de treinamento foi empregado para desenvolver um modelo PLS-DA construído utilizando um total de 10 componentes. O mesmo procedimento foi replicado na criação da OPLS-DA, na qual foram consideradas uma única componente linear e quatro componentes ortogonais. Ambas as metodologias foram processadas pelo pacote “*ropls*” na versão 1.32.0 da linguagem R. Modelos de ANN foram construídos através da utilização do pacote “*neuralnet*”, versão 1.44.236 com configuração de quatro camadas ocultas compostas por 300, 150, 150 e 150 neurônios respectivamente. O modelo RF foi desenvolvido utilizando o pacote “*RandomForest*” na versão 4.7-1.1. A construção do modelo envolveu a montagem de 5.000 árvores, onde foram empregadas 100 funcionalidades para cálculos em cada divisão. Na abordagem SNV o processamento é feito através das amostras em vez das variáveis utilizando função “*scale*” do pacote base o RStudio. MSC foi feito com o pacote “*prospectr*” versão 0.2.6 disponível no CRAN. A PCA foi realizada através da função PCA do pacote ‘*FactoMineR*’ na versão 2.8. A partir de sua associação com ANN e RF, as configurações de entrada foram atualizadas para 3 camadas ocultas com 15 neurônios cada 5.000 árvores com 3 amostras por divisão, respectivamente.

7.1.3 Resultados e discussões

O conjunto de dados brutos foi formado por 296 espectros (Figura 37) considerados como amostras independentes das quais 105 estão contidos no grupo teste (RT-PCR positivo) contra 191 do grupo controle (RT-PCR negativo), todos registrados nas mesmas condições experimentais. Ainda que o MIR em fase aquosa registre o sinal vibracional da molécula de água, algumas regiões do espectro são características da presença de metabólitos orgânicos e inorgânicos (OLIVER et al., 2016; PAN et al., 2022; PEREZ-GUAITA et al., 2020; SARIGUL et al., 2021; STEENBEKE et al., 2020).

FIGURA 37 - Espectros vibracionais por infravermelho médio de amostras de urina.



Fonte: Elaborado pela autora (2024).

Ureia, creatinina, ácido úrico e água estão entre os principais componentes da urina, contribuindo assim para as bandas de maior intensidade. Bandas associadas à deformação axial do N-H tanto da ureia quanto da creatinina, junto a outras aminas e uma pequena quantidade da porção lipídica possivelmente presente na amostra estão encobertas pela banda de maior intensidade em 3306 cm^{-1} , consequência da grande influência das vibrações de ligações O-H axial provenientes da grande quantidade de água. Por outro lado, esses principais componentes orgânicos também contribuem para a intensidade das bandas nas regiões mais baixas. As bandas em 2360 e 2256 cm^{-1} são o registro do CO_2 do ambiente, visto que todos os espectros são obtidos em um sistema aberto, em contato com o ar.

A creatinina e a ureia são as principais responsáveis pela intensidade acentuada em 1637 cm^{-1} e regiões adjacentes, juntamente com uma variedade de amidas devido às vibrações de estiramento C=O e bandas de N-H. Algumas proteínas residuais podem estar contribuindo para a banda em 1637 cm^{-1} dado o estiramento C-N, porém, como as amostras foram submetidas à desproteínização, essas macromoléculas não estão totalmente registradas nos espectros. A região entre 1465 e 1236 cm^{-1} é consequência das vibrações de deformação do NH_2 e principalmente do estiramento C-H presente em uma ampla variedade de metabólitos. Nesta região também são registrados sais inorgânicos, como fosfatos, sulfatos, citratos e outros. Os carboidratos, responsáveis por vibrações CO estão geralmente presentes em 1019 cm^{-1} , junto com o estiramento de O-H nas regiões próximas. Para posterior processamento de dados foram considerados apenas os sinais compreendidos entre 1800 e 900 cm^{-1} , região que representa a impressão digital das amostras de urina visto que contém a informação molecular mais rica não encoberta pela banda oriunda da grande quantidade de água, ou seja, um registro exclusivo dos metabólitos urinários humanos.

O conjunto de dados foi dividido aleatoriamente em treinamento e teste, considerando a proporção de 80:20 respectivamente, mantendo a proporção de amostras positivas e negativas do conjunto de dados original (considerando os resultados do teste RT-PCR para Sars-CoV-2), resultando em um conjunto de treinamento de 237 amostras, 84 GT e 153 GC. Enquanto o conjunto de dados de teste compreende 59 amostras no total, 21 GT e 38 GC. Todas as considerações feitas na seção 5.1.5 sobre a codificação das amostras dentro da programação R são também válidas para esse tópico do trabalho. Os parâmetros de desempenho dos modelos construídos estão listados na Tabela 20 e a matriz de confusão combinada dos modelos na Tabela 21.

TABELA 20 - Avaliação comparativa de desempenho de todos os modelos classificatórios por FTIR.

Modelo	Exatidão	Sensibilidade	Especificidade	MCC	F1-Score
PLS-DA	0.73	0.72	0.78	0.37	0.82
O-PLS-DA	0.64	0.65	0.50	0.08	0.77
ANN	0.78	0.84	0.68	0.52	0.83
RF	0.81	0.77	1.00	0.62	0.87
SNV-ANN	0.77	0.76	0.71	0.42	0.82

SNV-RF	0.80	0.76	1.00	0.55	0.86
MSC-ANN	0.73	0.79	0.61	0.41	0.79
MSC-RF	0.78	0.78	0.79	0.50	0.84
PCA-ANN	0.78	0.82	0.70	0.51	0.83
PCA-RF	0.80	0.78	0.84	0.54	0.86

Fonte: Elaborado pela autora (2024).

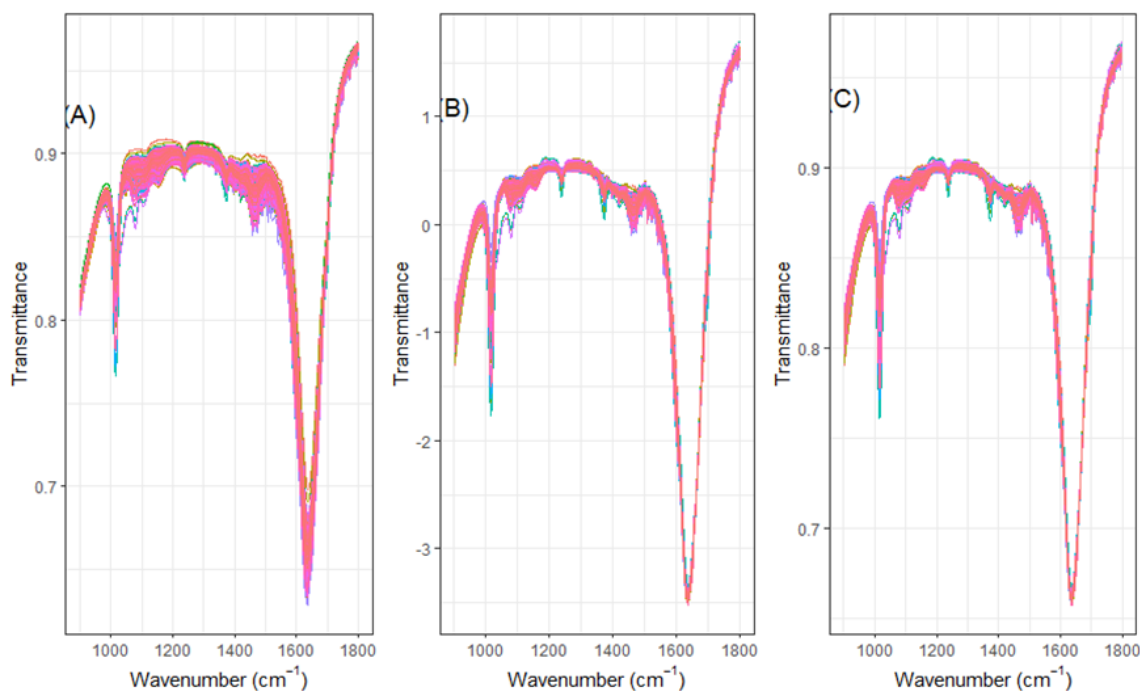
TABELA 21 - Matriz de confusão combinada dos modelos de aprendizado de máquina não lineares associados os dados de FTIR

Modelo	Classe real	Classe predita	
		Negative	Positive
PLS-DA	Negative	36	2
	Positive	14	7
OPLS-DA	Negative	36	2
	Positive	19	2
ANN	Negative	31	7
	Positive	6	15
RF	Negative	38	0
	Positive	11	10
SNV-ANN	Negative	34	4
	Positive	11	10
SNV-RF	Negative	38	0
	Positive	12	9
MSC- ANN	Negative	33	5
	Positive	8	13
MSC-RF	Negative	35	3
	Positive	10	11
PCA-ANN	Negative	32	6
	Positive	7	14
PCA-RF	Negative	32	2
	Positive	10	11

Fonte: Elaborado pela autora (2024).

Os modelos lineares, PLS-DA e OPLS-DA, demonstraram capacidade de predição limitada quando aplicadas à tarefa de classificar divergências metabólicas como consequência da infecção por Sars-CoV-2, sendo insuficientes para distinguir efetivamente entre positivo e negativo para infecção por Sars-CoV-2. Apesar da PLS-DA ter apresentado valores acima de 70 % para os parâmetros de exatidão, sensibilidade e especificidade, a modelo falha nos parâmetros que considera os resultados por completo, como o índice MCC. Além disso, ao comparar com os modelos não-lineares, nota-se que os resultados são inferiores. Este resultado sugere que a complexidade e a não linearidade inerentes aos dados MIR presentes nas amostras de urina podem exigir estratégias de modelagem mais avançadas, semelhante aos resultados gerados para os dados de CZE-UV. Modelos de aprendizado profundo sem pré-processamento de dados já apresentam uma melhora na descrição de dados, com o RF atingindo 100 % de especificidade. Com objetivo de mitigar possíveis variações instrumentais no sinal registrado, alguns algoritmos de pré-processamento foram utilizados (Figura 38), além da PCA como seleção de variáveis para ANN e RF

FIGURA 38 - Resultado comparativo da aplicação de pré-processamento aos dados de FTIR. (A) Sem tratamento; (B) SNV; (C) MSC.



Fonte: Elaborado pela autora (2024).

A variação normal padrão (*Standard Normal Variation, SNV*) envolve um processo de normalização que padroniza cada espectro a partir da subtração do valor de intensidade em cada comprimento de onda pela intensidade média do espectro total, dessa forma cada espectro tem valor médio zero e uma variância unitária, consequentemente reduzindo variações sistemáticas, como mudanças na linha de base e de intensidade, que podem distorcer a informação espectral. (DHANOA et al., 1994; PIGOZZO, 2011) Na avaliação do desempenho do modelo, o SNV-RF supera o SNV-ANN. Novamente o modelo RF alcançou 100 % de especificidade, o que sugere que a capacidade do RF de classificar corretamente os casos negativos é consistentemente confiável em diversos cenários clínicos. No entanto, comparando com os dados não processados, não houve uma melhora expressiva para RF, bem como, o ANN sem processamento apresenta desempenho superior, exceto no parâmetro de especificidade.

A correção multiplicativa de sinal (*Multiplicative Scatter Correction, MSC*) também é uma ferramenta de normatização amplamente utilizada que corrige o espectro de modo que ele se mantenha o mais próximo possível de um espectro de referência, na qual o algoritmo atua fazendo uma regressão linear das variáveis espectrais. A abordagem MSC corrige os fatores como tamanho de partícula e espessura da amostra, que podem obscurecer as verdadeiras diferenças espectrais (DHANOA et al., 1994; WINDIG; SHAVER; BRO, 2008). Comparando com os resultados sem pré-processamento, a MSC também falha em aprimorar os resultados obtidos, apesar de uma melhora no *F1-Score*, demais parâmetros não superaram os valores obtidos anteriormente, inclusive sendo inferiores à associação com o SNV. Por fim, a PCA foi usada para reduzir a dimensionalidade dos dados, selecionando os componentes principais como forma de seleção para o processamento seguinte. A exemplo do que foi observado como resultados dos processamentos anteriores, a PCA também não gera resultados superiores em relação aos dados sem processamento.

Este estudo teve como objetivo fornecer uma alternativa de teste de triagem utilizando amostragem não invasiva para análise em instrumentações de baixa manutenção, fáceis de operar e portáteis. Os dados adquiridos pelo MIR forneceram um registro espectral cujo processamento posterior nos permitiu inferir que há de fato uma diferença bioquímica/metabólica na urina de pessoas expostas ao Sars-Cov-2 daquelas que não o foram. Ao utilizar as tecnologias emergentes associadas à ciência de dados, a

foi possível criar um modelo de classificação que poderia ser posteriormente utilizado para discriminação qualitativa entre positivo e negativo para a presença de Sars-Cov-2. Por serem análises simples, também é uma estratégia interessante para triagem de estudos elaborados, ou seja, a partir dos resultados obtidos por FTIR para qualquer conjunto de dados é possível fazer uma previsão se vale investir em análises subsequentes. Os resultados referentes a esse estudo estão atualmente submetidos para apreciação em revista científica da área de instrumentação e ciência de dados indexada com conceito Qualis Capes A1.

8 CONCLUSÃO

As propostas analíticas desenvolvidas ao longo do trabalho mostraram a vasta gama de possibilidades de métodos aplicáveis ao estudo do metaboloma humano. Ainda que muitas etapas ainda estejam sendo concluídas e outras novas estratégias possam ser desenvolvidas dentro do grupo de pesquisa em que foram iniciadas, os resultados geraram discussões interessantes dentro do contexto analítico.

Em relação aos métodos, a associação da eletroforese capilar com detecção UV à ciência de dados envolvendo aprendizado profundo não tem expressivos registros na literatura, sendo uma novidade explorada dentro desse trabalho. Os resultados sugerem não apenas que a abordagem é possível como pode ser adaptada a realidades outras se não as bioanálises, além de servirem estrategicamente como forma de pesquisas exploratórias previamente ao uso de instrumentações mais elaboradas. Mesmo em matrizes complexas, mostrou-se ainda ser uma possibilidade para doseamento de alguns metabólitos específicos, também se apresentando como uma alternativa de doseamento de analitos de alta demanda. Acredita-se ainda que o uso de técnicas de eletromigração associadas a detectores universais como os espectrômetros de massa que possam segmentar o estudo do metaboloma humano em grupos de moléculas estratégicos seja uma alternativa interessante para obter resultados confiáveis de forma eficiente.

Quanto a cromatografia em meio líquido em associação com a espectrometria de massas, por mais que seja uma técnica consolidada e hoje talvez a mais utilizada para estudos do metaboloma humano, ainda há espaço para novas propostas. O uso de planejamento de experimentos para refinar as escolhas relacionadas ao uso da fonte de ionização também soma poucos registros na literatura. Apesar de simples, o estudo elaborado dentro deste projeto indica que tal estratégia pode ser útil para obtenção de

dados de qualidade, pré-requisito importante para análises dentro do conceito da metabolômica.

Cumprindo o objetivo do trabalho, foi possível determinar que há metabólitos em amostras de coleta não invasivas como a urina suficientes para que seja identificada a infecção pelo vírus causados da COVID-19. Nesse aspecto, tanto técnicas de eletromigração quanto a espectroscopia por infravermelho mostraram ser aplicáveis à proposta, com destaque para o infravermelho que gerou melhores resultados a partir de instrumentação simplificada, sendo possivelmente uma abordagem de primeira escolha a ser adaptada como modelo de triagem em espaços de atendimento em saúde, todas essas sendo associadas a ferramentas de processamento de dados atuais e vem se tornando populares nos últimos anos.

Ao longo deste trabalho propostas envolvendo Ressonância Magnética Nuclear, Cromatografia a Gás e outras propostas das instrumentações utilizadas foram também desenhadas, sendo ainda uma possibilidade de realização no futuro. Por fim, acredita-se que foram obtidos dados de qualidade e, principalmente, foram desenvolvidos métodos e rotinas analíticas com implicações clínicas, instrumentais e metodológicas que podem futuramente ser extrapoladas ao mercado e ao atendimento à população e ao impulsionamento de novas pesquisas.

REFERÊNCIAS

AHMED, Z. Practicing precision medicine with intelligently integrative clinical and multi-omics data analysis. **Human Genomics**, v. 14, n. 1, p. 1–5, 2020.

ALBERICE, J. V. et al. Searching for urine biomarkers of bladder cancer recurrence using a liquid chromatography-mass spectrometry and capillary electrophoresis-mass spectrometry metabolomics approach. **Journal of Chromatography A**, v. 1318, p. 163–170, 2013.

ANVISA. **RESOLUÇÃO DA DIRETORIA COLEGIADA - RDC Nº 786, DE 5 DE MAIO DE 2023**

BALAKRISHNAN, V. S. WHO launches global initiative for arboviral diseases. **The Lancet Microbe**, v. 3, n. 6, p. e407, 2022.

BALOGH, E. P.; MILLER, B. T.; BALL, J. R. **Improving diagnosis in health care**. Washinton DC: The National Academies Press, 2016.

BEISEL, W. R. METABOLIC RESPONSE TO INFECTION. **Annual Review of Medicine**, p. 9–20, 1975.

BELARD, A. et al. Precision diagnosis: a view of the clinical decision support systems (CDSS) landscape through the lens of critical care. **Journal of Clinical Monitoring and Computing**, v. 31, n. 2, p. 261–271, 2017.

BÉLTRÁN-SILVA, S. L. et al. Clinical and differential diagnosis: Dengue, chikungunya and Zika. **Revista Médica del Hospital General de México**, v. 81, n. 3, p. 146–153, 2018.

BEZERRA, J. M. T. et al. Entry of dengue virus serotypes and their geographic distribution in Brazilian federative units: A systematic review. **Revista Brasileira de Epidemiologia**, v. 24, 2021.

BÖGER, B. et al. Systematic review with meta-analysis of the accuracy of diagnostic tests for COVID-19. **American Journal of Infection Control**, v. 49, n. 1, p. 21–29, 2021.

BRASIL. MINISTÉRIO DA SAÚDE. **Dengue: Aspectos Epidemiológicos, Ministério da Saúde Diagnóstico e Tratamento**. Série A. N ed. Brasília, DF.: Editora MS, 2002.

BRASIL. MINISTÉRIO DA SAÚDE. **Transmissão**. Disponível em: <https://www.gov.br/saude/pt-br/assuntos/coronavirus/transmissao>. Acesso em: 24 jan. 2024.

BRASIL. MINISTÉRIO DA SAÚDE. **Sintomas**. Disponível em: <https://www.gov.br/saude/pt-br/assuntos/coronavirus/sintomas>. Acesso em: 24 jan. 2024.

BRASIL. MINISTÉRIO DA SAÚDE. **Zika Vírus**. Disponível em: [https://www.gov.br/saude/pt-br/assuntos/saude-de-a-a-z/z/zika-virus#:~:text=O diagnóstico do Zika Vírus,Único de Saúde \(SUS\)..](https://www.gov.br/saude/pt-br/assuntos/saude-de-a-a-z/z/zika-virus#:~:text=O diagnóstico do Zika Vírus,Único de Saúde (SUS)..) Acesso em: 23 jan. 2024a.

BRASIL. MINISTÉRIO DA SAÚDE. **Chegam ao Brasil as primeiras doses da vacina contra a dengue**. Disponível em: <https://www.gov.br/saude/pt-br/assuntos/noticias/2024/janeiro/chegam-ao-brasil-as-primeiras-doses-da-vacina-contr-a-dengue>. Acesso em: 24 jan. 2024b.

BRASIL. **Arboviroses**. Disponível em: <https://www.gov.br/saude/pt-br/assuntos/saude-de-a-a-z/a/arboviroses>. Acesso em: 9 jan. 2024.

BRASIL, M. DA S. **Chikungunya: Manejo Clínico**. [s.l.: s.n.]. v. 2ª edição

BYERS, N. M. et al. Metabolomic Insights into Human Arboviral Infections: Dengue, Chikungunya, and Zika Viruses. **viruses**, v. 11, n. 3, p. 1–30, 2019.

CALVO, E. P. et al. Easy and inexpensive molecular detection of dengue, chikungunya and zika viruses in febrile patients. **Acta Tropica**, v. 163, p. 32–37, 2016.

CANUTO, G. A. B. et al. Metabolômica: Definições, Estado-Da-Arte E Aplicações Representativas. **Quimica Nova**, v. 41, n. 1, p. 75–91, 2018.

CASERTA, T. M. V.; BAIENSE, A. S. R.; ANDRADE, L. G. DE. O IMPACTO DA RESOLUÇÃO–RDC Nº786, DE 05 DE MAIO DE 2023 NA ATENÇÃO FARMACÊUTICA EM FARMÁCIA COMUNITÁRIA. **Revista Ibero- Americana de Humanidades, Ciências e Educação-**, v. 9, n. 9, 2023.

CDC. **COVID-19 Testing: What You Need to Know**. Disponível em: <https://www.cdc.gov/coronavirus/2019-ncov/symptoms-testing/testing.html>. Acesso em: 25 fev. 2024a.

CDC. **Self-Testing At Home or Anywhere**. Disponível em:

<https://www.cdc.gov/coronavirus/2019-ncov/testing/self-testing.html>. Acesso em: 25 fev. 2024b.

CDC. **Symptoms of COVID-19**. Disponível em: <https://www.cdc.gov/coronavirus/2019-ncov/symptoms-testing/symptoms.html>. Acesso em: 25 jan. 2024.

CHAN, K. S. et al. Effect of serotypes on clinical manifestations of dengue fever in adults. **Journal of Microbiology, Immunology and Infection**, v. 42, n. 6, p. 471–478, 2009.

CORDEIRO, M. T. Laboratory diagnosis for dengue. **Revista do Instituto de Medicina Tropical de Sao Paulo**, v. 54, n. SUPPL.18, p. 10–12, 2012.

DA POIAN, A. T.; EL-BACHA, T.; LUZ, M. R. M. P. Nutrient Utilization in Humans: Metabolism Pathways. **Nature Education**, v. 3, n. 9, p. 11, 2010.

DALLMANN, R. et al. The human circadian metabolome. **Proceedings of the National Academy of Sciences of the United States of America**, v. 109, n. 7, p. 2625–2629, 2012.

DAME, Z. T. et al. The human saliva metabolome. **Metabolomics**, v. 11, n. 6, p. 1864–1883, 2015.

DE PAULA, S. O.; FONSECA, B. A. L. DA. Dengue: a review of the laboratory tests a clinician must know to achieve a correct diagnosis. **The Brazilian journal of infectious diseases: an official publication of the Brazilian Society of Infectious Diseases**, v. 8, n. 6, p. 390–398, 2004.

DHANOVA, M. S. et al. The Link between Multiplicative Scatter Correction (MSC) and Standard Normal Variate (SNV) Transformations of NIR Spectra. **Journal of Near Infrared Spectroscopy**, v. 2, n. 1, p. 43–47, 1994.

DO NASCIMENTO, M. P. et al. Mixture design of an electrolyte system for the simultaneous separation of Cl^- , SO_4^{2-} , NO_3^- , NO_2^- , and HCO_3^- in shrimp-farming water by CZE-UV. **Analytical Methods**, n. 3, p. 461–469, 2022.

DUTTA, A. COVID-19 waves: variant dynamics and control. **Scientific Reports**, v. 12, n. 1, p. 1–9, 2022.

DUTTA, D. et al. COVID-19 Diagnosis: A Comprehensive Review of the RT-qPCR Method for Detection of SARS-CoV-2. **Diagnostics**, v. 12, n. 6, p. 1–18, 2022.

EWE, K. **Dengue Fever Is Soaring Worldwide. Here's What to Know—and How to Stay Protected.** Disponível em: <https://time.com/6429963/dengue-fever-cases-symptoms-prevention-explainer/>. Acesso em: 22 jan. 2024.

FILCHAKOVA, O. et al. Review of COVID-19 testing and diagnostic methods. **Talanta**, v. 244, n. January, p. 123409, 2022.

FIOCRUZ. **Como o vírus zika é diagnosticado?** Disponível em: <https://portal.fiocruz.br/pergunta/como-o-virus-zika-e-diagnosticado>. Acesso em: 23 jan. 2024.

FIOCRUZ. **Chikungunya: sintomas, transmissão e prevenção.** Disponível em: <https://www.bio.fiocruz.br/index.php/br/chikungunya-sintomas-transmissao-e-prevencao>. Acesso em: 22 jan. 2024.

FIOCRUZ. **Zika.** Disponível em: <https://portal.fiocruz.br/zika>. Acesso em: 22 jan. 2024.

GARDNER, A.; CARPENTER, G.; SO, P. W. Salivary metabolomics: From diagnostic biomarker discovery to investigating biological function. **Metabolites**, v. 10, n. 2, 2020.

GÉRON, A. **Mãos à Obra: Aprendizado de máquina com Scikit-Learn e TensorFlow. Conceitos, ferramentas e técnicas para construção de sistemas inteligentes.** Rio de Janeiro: Alta Books, 2019.

GODOY, A. T.; EBERLIN, M. N.; SIMIONATO, A. V. C. Targeted metabolomics: Liquid chromatography coupled to mass spectrometry method development and validation for the identification and quantitation of modified nucleosides as putative cancer biomarkers. **Talanta**, v. 210, n. December 2019, p. 120640, 2020.

GOULD, E. et al. Emerging arboviruses: Why today? **One Health**, v. 4, n. June, p. 1–13, 2017.

GU, H. et al. Metabolomics Method to Comprehensively Analyze Amino Acids in Different Domains. **Analyst**, v. 140, n. 8, p. 2726–2734, 2016.

GUPTA, A. et al. Prevalence of Dengue Serotypes and Its Correlation with the Laboratory Profile at a Tertiary Care Hospital in Northwestern India. **Cureus**, v. 13, n. 5, p. 1–8, 2021.

HMDB. **Showing metabocard for Fumaric acid (HMDB0000134).**

HMDB. **Showing metabocard for Tartaric acid (HMDB0000956)**. Disponível em: <https://hmdb.ca/metabolites/HMDB0000956>. Acesso em: 12 fev. 2024b.

HMDB. **Showing metabocard for Oxalic acid (HMDB0002329)**. Disponível em: <https://hmdb.ca/metabolites/HMDB0002329>. Acesso em: 13 fev. 2024c.

INSTITUTO BUTANTÃ. **Vacina da Dengue**. Disponível em: <https://butantan.gov.br/dengue>. Acesso em: 24 jan. 2024.

JIN, S.; HOPPEL, L.; TSERNGS, K. Incomplete Fatty Acid Oxidation. v. 267, n. 1, p. 119–125, 1992.

KERKHOF, K. et al. Reliable Serological Diagnostic Tests for Arboviruses: Feasible or Utopia? **Trends in Microbiology**, v. 28, n. 4, p. 276–292, 2020.

KHOURY, M. J.; IADEMARCO, M. F.; RILEY, W. T. Precision Public Health for the Era of Precision Medicine. **American Journal of Preventive Medicine**, v. 50, n. 3, p. 398–401, 2016.

KHOVIDHUNKIT, W. et al. Effects of infection and inflammation on lipid and lipoprotein metabolism: Mechanisms and consequences to the host. **Journal of Lipid Research**, v. 45, n. 7, p. 1169–1196, 2004.

KINDHAUSER, M. K. et al. Zika: Origine et propagation d'un virus transmis par des moustiques. **Bulletin of the World Health Organization**, v. 94, n. 9, p. 675- 686C, 2016.

KÖNIG, I. R. et al. What is precision medicine? **European Respiratory Journal**, v. 50, n. 4, p. 1–12, 2017.

KOWALSKA, K. et al. The Influence of SARS-CoV-2 Infection on Lipid Metabolism—The Potential Use of Lipid-Lowering Agents in COVID-19 Management. **Biomedicines**, v. 10, n. 9, 2022.

KTSOYAN, Z. A. et al. Profiles of microbial fatty acids in the human metabolome are disease-specific. **Frontiers in Microbiology**, v. 1, n. JAN, p. 1–9, 2011.

LAMERS, M. M.; HAAGMANS, B. L. SARS-CoV-2 pathogenesis. **Nature Reviews Microbiology**, v. 20, n. 5, p. 270–284, 2022.

LASISI, T. J.; RAJI, Y. R.; SALAKO, B. L. Salivary creatinine and urea analysis in

patients with chronic kidney disease: A case control study. **BMC Nephrology**, v. 17, n. 1, p. 1–6, 2016.

LING, Z. N. et al. Amino acid metabolism in health and disease. **Signal Transduction and Targeted Therapy**, v. 8, n. 1, 2023.

MÁRCIA MIGUEL CASTRO FERREIRA. **Quimiometria: Conceitos, métodos e aplicações**. Campinas: Editora da Unicamp, 2015.

MARINI, F. Orthogonal PLS (O-PLS) and related algorithms. **Journal of Chemometrics**, v. 34, n. 1, p. 10–12, 2020.

MARQUES, C. D. L. et al. Recommendations of the Brazilian Society of Rheumatology for the diagnosis and treatment of chikungunya fever. Part 2 – Treatment. **Revista Brasileira de Reumatologia (English Edition)**, v. 57, n. S 2, p. 438–451, 2017.

MARTINS, L. O. O segmento da medicina diagnóstica no Brasil. **Revista da Faculdade de Ciências Médicas de Sorocaba**, v. 16, n. 3, p. 139–145, 2014.

MCCOLL, E. R. et al. The Age of Omics-Driven Precision Medicine. **Clinical Pharmacology and Therapeutics**, v. 106, n. 3, p. 477–481, 2019.

MONTAGU, A.; PRINCENTON, N. J. Adaptive Value of Labor Pains. **JAMA - Journal of the American Medical Association**, p. 2015, 2015.

MOREIRA, O. et al. Capillary Electrophoresis Applied to Human Urine Analysis for Clinical Diagnosis: New Trends and Perspectives. **Brazilian Journal of Analytical Chemistry**, n. April, p. 1–35, 2022.

MOREIRA, O. B. DE O. et al. Determination of creatinine in urine and blood serum human samples by CZE-UV using on-column internal standard injection. **Talanta**, v. 258, 2023a.

MOREIRA, O. B. DE O. et al. Factorial design applied to LC-ESI-QTOF mass spectrometer parameters for untargeted metabolomics. **Analytical Methods**, v. 15, n. 20, p. 2512–2521, 2023b.

MOULIN, E. et al. Simultaneous outbreaks of dengue, chikungunya and Zika virus infections : diagnosis challenge in a returning traveller with nonspecific febrile illness. **New Microbes and New Infections**, v. 11, p. 6–7, 2016.

NAVA, A. S. L. DE; RAJA, A. **Physiology, Metabolism**. Treasure Island: StatPearls Publishing, 2022.

NETO, B. B.; SCARMINIO, I. S.; BRUNS, R. E. **Como Fazer Experimentos: Pesquisa e Desenvolvimento na Ciência e na Indústria**. 4th. ed. Porto Alegre: Bookman, 2010.

NETO, S. R. D. S. et al. **Machine learning and deep learning techniques to support clinical diagnosis of arboviral diseases: A systematic review**. [s.l.: s.n.]. v. 16

O'RIORDAN, K. J. et al. Short chain fatty acids: Microbial metabolites for gut-brain axis signalling. **Molecular and Cellular Endocrinology**, v. 546, n. January, 2022.

OLIVEIRA, M. A. L. et al. Testes diagnósticos para o SARS-COV-2: uma reflexão crítica. **Química Nova**, v. 45, n. 6, p. 760–766, 2022.

OLIVER, K. V. et al. Infrared vibrational spectroscopy: A rapid and novel diagnostic and monitoring tool for cystinuria. **Scientific Reports**, v. 6, n. October, p. 1–7, 2016.

OLIVIER, M. et al. The need for multi-omics biomarker signatures in precision medicine. **International Journal of Molecular Sciences**, v. 20, n. 19, 2019.

ORGANIZAÇÃO PANAMERICANA DE SAÚDE. **Zika - Reporte de Casos Acumulados**. Disponível em: <<https://www3.paho.org/data/index.php/es/temas/zika-weekly-es/>>. Acesso em: 9 jan. 2024a.

ORGANIZAÇÃO PANAMERICANA DE SAÚDE. **Casos Reportados de Dengue en las Américas**. Disponível em: <https://www3.paho.org/data/index.php/es/temas/indicadores-dengue/dengue-nacional/9-dengue-pais-ano.html>. Acesso em: 9 jan. 2024b.

ORGANIZAÇÃO PANAMERICANA DE SAÚDE. **Chikungunya cases**. Disponível em: <https://www3.paho.org/data/index.php/en/mnu-topics/chikv-en/550-chikv-weekly-en.html>. Acesso em: 9 jan. 2024c.

PAIXÃO, E. S.; TEIXEIRA, M. G.; RODRIGUES, L. C. Zika , chikungunya and dengue : the causes and threats of new and re- emerging arboviral diseases. 2017.

PAN, S. W. et al. Using an ATR-FTIR Technique to Detect Pathogens in Patients with Urinary Tract Infections: A Pilot Study. **Sensors**, v. 22, n. 10, 2022.

PARK, M. et al. A systematic review of covid-19 epidemiology based on current evidence. **Journal of Clinical Medicine**, v. 9, n. 4, 2020.

PATTI, G. J.; YANES, O.; SIUZDAK, G. Metabolomics: the apogee of the omics trilogy. **Nature Reviews Molecular Cell Biology**, v. 13, n. 4, p. 263–269, 2012.

PENNA, P. M. M. et al. Biossegurança: Uma Revisão. **Arquivos do Instituto Biológico**, v. 77, n. 3, p. 555–565, 2010.

PEREZ-GUAITA, D. et al. Quantification and Identification of Microproteinuria Using Ultrafiltration and ATR-FTIR Spectroscopy. 2020.

PIGOZZO, R. J. B. **Espectroscopia de Infravermelho próximo em madeiras neotropicais: identificação e predição de propriedades físicas**. [s.l.] Universidade de São Paulo, 2011.

PLAZA, N. C.; GARCÍA-GALBIS, M. R.; MARTINEZ-ESPINOSA, R. M. Impact of the “Omics Sciences” in Medicine: New Era for Integrative Medicine. **Journal of Clinical Microbiology and Biochemical Technology**, v. 3, p. 9–13, 2017.

PUCCIONI-SOHLER, M. et al. Dengue infection in the nervous system : lessons learned for Zika and Chikungunya. **Arquivos de Neuro-Psiquiatria**, v. 75, n. 2, p. 123–126, 2016.

RABAAN, A. A. et al. Viral dynamics and real-time rt-pcr ct values correlation with disease severity in covid-19. **Diagnostics**, v. 11, n. 6, 2021.

SAFADY, N. G. **Ciências Ômicas e seu papel na medicina de precisão**. Disponível em: <https://blog.varsomics.com/o-que-sao-as-ciencias-omicas/>. Acesso em: 9 jan. 2024.

SARIGUL, N. et al. Urine analysis using FTIR spectroscopy: A study on healthy adults and children. **Journal of Biophotonics**, v. 14, n. 7, p. 1–14, 2021.

SCHMIDT, C.; SCHNIERLE, B. S. Chikungunya Vaccine Candidates: Current Landscape and Future Prospects. **Drug Design, Development and Therapy**, v. 16, p. 3663–3673, 2022.

SRISAWAT, N. et al. World Dengue Day: A call for action. **PLoS Neglected Tropical Diseases**, v. 16, n. 8, p. 2–10, 2022.

STEENBEKE, M. et al. Exploring the possibilities of infrared spectroscopy for urine sediment examination and detection of pathogenic bacteria in urinary tract infections. **Clinical Chemistry and Laboratory Medicine**, v. 58, n. 10, p. 1759–1767, 2020.

SUN, J. et al. COVID-19: Epidemiology, Evolution, and Cross-Disciplinary Perspectives. **Trends in Molecular Medicine**, v. 26, n. 5, p. 483–495, 2020.

TAHAMTAN, A.; ARDEBILI, A. Real-time RT-PCR in COVID-19 detection: issues affecting the results. **Expert Review of Molecular Diagnostics**, v. 20, n. 5, p. 453–454, 2020.

TANG; MIRANDA, C. **Biossegurança Biomédicos e de Microbiologia**. [s.l.: s.n.].

TANNER, J. E.; ALFIERI, C. The fatty acid lipid metabolism nexus in COVID-19. **Viruses**, v. 13, n. 1, p. 1–18, 2021.

TAVARES, A. **Agência reguladora dos Estados Unidos aprova vacina contra chikungunya do Butantan e da Valneva**. Disponível em: <https://butantan.gov.br/noticias/agencia-reguladora-dos-estados-unidos-aprova-vacina-contrachikungunya-do-butantan-e-da-valneva>. Acesso em: 24 jan. 2024.

TAVARES, A. **Butantan trabalha no desenvolvimento de vacina contra Zika para prevenir microcefalia em recém-nascidos**. Disponível em: <https://butantan.gov.br/noticias/butantan-trabalha-no-desenvolvimento-de-vacina-contrazika-para-prevenir-microcefalia-em-recem-nascidos>. Acesso em: 24 jan. 2024.

TAVARES, M. F. M. Separation mechanisms in capillary electrophoresis. **Quimica Nova**, v. 20, n. 5, p. 493–511, 1997.

TEBANI, A. et al. Omics-based strategies in precision medicine: Toward a paradigm shift in inborn errors of metabolism investigations. **International Journal of Molecular Sciences**, v. 17, n. 9, 2016.

TEIXEIRA, M. D. G.; BARRETO, M. L.; GUERRA, Z. Epidemiologia e Medidas de Prevenção do Dengue. **Informe epidemiológico do SUS**, v. 8, n. 4, p. 5–33, 1999.

TEMILOLA, D. O. et al. Salivary creatinine as a diagnostic tool for evaluating patients with chronic kidney disease. **BMC Nephrology**, v. 20, n. 1, p. 1–6, 2019.

THERMOFISHER SCIENTIFIC. **TaqPath COVID-19 CE-IVD RT-PCR Kit**.

TSERNG, K. Y.; JIN, S. J. Metabolic Origin of Urinary 3-Hydroxy Dicarboxylic Acids. **Biochemistry**, v. 30, n. 9, p. 2508–2514, 1991.

UNITED NATIONS. **Spike in dengue cases due to global warming, warns WHO**. Disponível em: <<https://news.un.org/en/story/2023/07/1138962>>. Acesso em: 22 jan. 2024.

UZUNIAN, A. Coronavírus SARS-CoV-2 e Covid-19. **J Bras Patol Med Lab**, v. 56, p. 1–4, 2020.

VAIDYANATHAN, K.; NARAYANAN, M. P.; VASUDEVAN, D. M. Organic acidurias: An updated review. **Indian Journal of Clinical Biochemistry**, v. 26, n. 4, p. 319–325, 2011.

VARGHESE, J.; DE SILVA, I.; MILLAR, D. S. Latest Advances in Arbovirus Diagnostics. **Microorganisms**, v. 11, n. 5, 2023.

VAZ, F. A. S. et al. Optimisation of a capillary zone electrophoresis methodology for simultaneous analysis of organic aliphatic acids in extracts of *Brachiaria brizantha*. **Phytochemical Analysis**, v. 23, n. 6, p. 569–575, 2012.

VIGO, F. et al. Vibrational Spectroscopy in Urine Samples as a Medical Tool: Review and Overview on the Current State-of-the-Art. **Diagnostics**, v. 13, n. 1, p. 1–14, 2023.

VINCENT, K. M. et al. The solubility of calcium oxalates explains some aspects of their underrepresentation in the oral cavity. **Archives of Oral Biology**, v. 121, n. August 2020, p. 104965, 2021.

WAHL, R.; KALLEE, E. Oxalic Acid in Saliva, Teeth and Tooth Tartar. **European Journal of Clinical Chemistry and Clinical Biochemistry**, v. 32, p. 821–825, 1994.

WALLER, J. V. et al. Diagnostic Tools for Coronavirus Disease (COVID-19): Comparing CT and RT-PCR Viral Nucleic Acid Testing. **American Journal of Roentgenology**, v. 215, n. 4, p. 834–838, 2020.

WANG, Y. et al. Current Advances in Zika Vaccine Development. **Vaccines**, v. 10, n. 11, p. 1–18, 2022.

WEBSTER, D. P.; FARRAR, J.; ROWLAND-JONES, S. Progress towards a dengue vaccine. **The Lancet Infectious Diseases**, v. 9, n. 11, p. 678–687, 2009.

WHO. **Coronavirus disease (COVID-19): How is it transmitted?** Disponível em: <https://www.who.int/news-room/questions-and-answers/item/coronavirus-disease-covid-19-how-is-it-transmitted>. Acesso em: 24 jan. 2024.

WHO. **Global Arbovirus Initiative.** Disponível em: <https://www.who.int/news-room/events/detail/2022/03/31/default-calendar/global-arbovirus-initiative>. Acesso em: 23 jan. 2024.

WHO. **Health topics: Infectious diseases.** Disponível em: <https://www.emro.who.int/health-topics/infectious-diseases/index.html>.

WHO. **Dengue - Global situation.** Disponível em: <https://www.who.int/emergencies/disease-outbreak-news/item/2023-DON498>. Acesso em: 10 jan. 2024b.

WHO. **Dengue and severe dengue.** Disponível em: <https://www.who.int/news-room/fact-sheets/detail/dengue-and-severe-dengue>.

WHO. **WHO Coronavirus (COVID-19) Dashboard.** Disponível em: <https://covid19.who.int/>.

WHO. **Chikungunya.** Disponível em: https://www.who.int/health-topics/chikungunya#tab=tab_1. Acesso em: 22 jan. 2024a.

WHO. **WHO COVID-19 dashboard.** Disponível em: <https://data.who.int/dashboards/covid19/cases?n=c>. Acesso em: 25 jan. 2024b.

WIDERA, M. et al. Evaluation of stability and inactivation methods of SARS-CoV-2 in context of laboratory settings. **Medical Microbiology and Immunology**, v. 210, n. 4, p. 235–244, 2021.

WINDIG, W.; SHAVER, J.; BRO, R. Loopy MSC: A simple way to improve multiplicative scatter correction. **Applied Spectroscopy**, v. 62, n. 10, p. 1153–1159, 2008.

XAVIER, A. R. et al. Clinical and laboratory diagnosis of Zika fever: An update. **Jornal Brasileiro de Patologia e Medicina Laboratorial**, v. 53, n. 4, p. 252–257, 2017.

YOON, J. G. et al. Clinical significance of a high SARS-CoV-2 viral load in the Saliva. **Journal of Korean Medical Science**, v. 35, n. 20, p. 1–6, 2020.

YÜCE, M.; FILIZTEKIN, E.; ÖZKAYA, K. G. COVID-19 diagnosis —A review of current methods. **Biosensors and Bioelectronics**, v. 172, n. October 2020, 2021.

ZHAO, T. et al. Altered Lipid Profile in COVID-19 Patients and Metabolic Reprogramming. **Frontiers in Microbiology**, v. 13, n. May, p. 1–11, 2022.

APÊNDICE I

PUBLICAÇÕES RELACIONADAS AO TEMA DO PROJETO

MOREIRA, OLÍVIA BRITO DE OLIVEIRA; QUEIROZ DE SOUZA, JÉSSICA CORDEIRO; BERALDO CANDIDO, JOÃO MARCOS; DO NASCIMENTO, MARIA PATRÍCIA; CHELLINI, PAULA ROCHA; DE LEMOS, LÚCIO MARCO; DE OLIVEIRA, MARCONE AUGUSTO LEAL. Determination of creatinine in urine and blood serum human samples by CZE-UV using on-column internal standard injection. *TALANTA*, v. 258, p. 124465, 2023. DOI:10.1016/j.talanta.2023.124465

MOREIRA, OLÍVIA BRITO DE OLIVEIRA; QUEIROZ DE SOUZA, JÉSSICA CORDEIRO; BERALDO CANDIDO, JOÃO MARCOS; DO NASCIMENTO, MARIA PATRÍCIA; CHELLINI, PAULA ROCHA; MARCO DE LEMOS, LÚCIO; LEAL DE OLIVEIRA, MARCONE AUGUSTO. Factorial design applied to LC-ESI-QTOF mass spectrometer parameters for untargeted metabolomics. *ANALYTICAL METHODS*, v. 15, p. 2512-2521, 2023. DOI: 10.1039/d3ay00094j.

MOREIRA, OLÍVIA; BERLANDE, BRUNA; DE SOUZA, JÉSSICA; CANDIDO, JOÃO MARCOS; JAEGER, LAUREN; ADRIANO, LUIZ HENRIQUE; E BRITO, MARCO ANTÔNIO; DO NASCIMENTO, MARIA PATRÍCIA; GERALDO, PATRÍCIA; CHELLINI, PAULA; DE OLIVEIRA, MARCONE AUGUSTO. Capillary Electromigration Techniques Applied to Clinical Chemistry Investigations and Alternative Diagnostic Assays: Possibilities and Perspectives. *JOURNAL OF THE BRAZILIAN CHEMICAL SOCIETY*, v. 00, p. 1-23, 2023. DOI: 10.21577/0103-5053.20230096.

LIMA, NERILSON M.; FERNANDES, BRUNO L.M.; ALVES, GUILHERME F.; DE SOUZA, JÉSSICA C.Q.; SIQUEIRA, MARCELO M.; PATRÍCIA DO NASCIMENTO, MARIA; **MOREIRA, OLÍVIA B. O.**; SUSSULINI, ALESSANDRA; DE OLIVEIRA, MARCONE A.L. Mass spectrometry applied to diagnosis, prognosis, and therapeutic targets identification for the novel coronavirus SARS-CoV-2: A review. *ANALYTICA CHIMICA ACTA*, v. 1195, p. 339385, 2022. DOI: 10.1016/j.aca.2021.339385.

MOREIRA, OLÍVIA B. O.; SOUZA, J. C. Q.; CANDIDO, J. M. B.; NASCIMENTO, M. P.; PENNA, E. A.; CHELLINI, P. R.; OLIVEIRA, M. A. L. Capillary Electrophoresis Applied to Human Urine Analysis for Clinical Diagnosis: New Trends and Perspectives. Brazilian Journal of Analytical Chemistry - BrJAC (Online), p. 17-51, 2022. DOI: 10.21577/0103-5053.20230096.

OLIVEIRA, MARCONE; WATANABE, ARIPUANÃ; CESAR, DIONÉIA; CANDIDO, JOÃO; LIMA, NERILSON; **MOREIRA, OLÍVIA;** CHELLINI, PAULA. Testes diagnósticos para o sars-cov-2: uma reflexão crítica. QUÍMICA NOVA (ONLINE), p. 1-7, 2022. DOI: 10.21577/0100-4042.20170895

APÊNDICE II

SÚMULA CURRICULAR

ARTIGOS PUBLICADOS EM PERIÓDICOS

ADRIANO, LUIZ HENRIQUE; CANDIDO, JOÃO MARCOS; **MOREIRA, OLÍVIA**; SCARMINIO, IEDA; BRUNS, ROY; DE OLIVEIRA, MARCONE AUGUSTO. Tutorial para aplicação didática de planejamentos de misturas utilizando Libreoffice e Rstudio. QUIMICA NOVA, v. X, p. 1, 2024. DOI: 10.21577/0100-4042.20240016.

PORTO, BRENDA; FARIA, ISAURA; DE SOUZA, JÉSSICA; **MOREIRA, OLÍVIA**; DA GAMA, MARCO; DE OLIVEIRA, MARCONE. Rapid method for simultaneous determination of cis-9, trans-11 and trans-10, cis-12 conjugated linoleic acid isomers in milk by gc-fid. QUIMICA NOVA, v. XY, p. 1, 2023. DOI: 10.21577/0100-4042.20230026.

DE SOUZA, DENISE SALVADOR; LOMEU, ALICE; DE OLIVEIRA **MOREIRA, OLÍVIA BRITO DE OLIVEIRA**; MARCONE AUGUSTO LEAL; MENDONÇA, HENRIQUE VIEIRA DE. New methods to increase microalgae biomass in anaerobic cattle wastewater and the effects on lipids production. BIOMASS & BIOENERGY, v. 176, p. 106915, 2023. DOI: 10.1016/j.biombioe.2023.106915.

GERALDO, PATRÍCIA ABRANCHES; VELASCO, ARTHUR WOYAMES DE CASTRO; BERLANDE, BRUNA MARCHIORI; SOUZA, JÉSSICA CORDEIRO QUEIROZ DE; CANDIDO, JOÃO MARCOS BERALDO; ADRIANO, LUIZ HENRIQUE CANTARINO; BRITO, MARCO ANTÔNIO DOMINGUES ASSAD E; DO NASCIMENTO, MARIA PATRÍCIA; **MOREIRA, OLÍVIA BRITO DE OLIVEIRA**; DE OLIVEIRA, MARCONE AUGUSTO LEAL. Fatty acids analysis by capillary electrophoresis: Fundamentals, advantages and applications. ELECTROPHORESIS, v. 00, p. 1, 2023. DOI: 10.1002/elps.202300144.

LOMEU, ALICE AZEVEDO; **DE OLIVEIRA MOREIRA, OLÍVIA BRITO**; DE OLIVEIRA, MARCONE AUGUSTO LEAL; DE MENDONÇA, HENRIQUE VIEIRA.

Applying Ozone in Cattle Wastewater to Maximize Lipid Production in Microalgae Biomass. *BioEnergy Research*, v. XY, p. 1, 2023. DOI: 10.1007/s12155-023-10564-z.

DE OLIVEIRA, PATRÍCIA LOPES; PORTO, BRENDA LEE SIMAS; **DE OLIVEIRA MOREIRA, OLÍVIA BRITO**; CANUTO, LÍVIA TORQUATO VENTURA; DE OLIVEIRA, MARCONE AUGUSTO LEAL. Trans-Fatty Acid Analysis Applied to Authenticity Monitoring of Bottled Butter by CZE-UV. *Food Analytical Methods*, v. XY, p. 1, 2022. DOI: 10.1007/s12161-022-02416-x.

BRITO DE OLIVEIRA MOREIRA, OLÍVIA; VINÍCIUS DE FARIA, LUCAS; MATOS, RENATO CAMARGO; ENES, KARINE BRAGA; COSTA COURI, MARA RÚBIA; DE OLIVEIRA, MARCONE AUGUSTO LEAL. Determination of hydroquinone and benzoquinone in pharmaceutical formulations: critical considerations on quantitative analysis of easily oxidized compounds. *Analytical Methods*, v. 14, p. 4784-4794, 2022. DOI: 10.1039/D2AY01631A.

MIGUEZ, FLÁVIO B.; **MOREIRA, OLÍVIA B. O.**; DE OLIVEIRA, MARCONE A. L.; DENADAI, ÂNGELO M. L.; DE OLIVEIRA, LUIZ F. C.; DE SOUSA, FREDERICO B. Reversible electrospun fibers containing spiropyran for acid and base vapor sensing. *JOURNAL OF MATERIALS RESEARCH*, v. XY, p. 1-10, 2022. DOI: 10.1557/s43578-022-00842-5

MOREIRA, OLÍVIA; CASTRO, LARISSA; OLIVEIRA, MARCONE. Cálculo e preparo de soluções tampão: guia completo usando o software peakmaster®. *QUIMICA NOVA*, v. 44, p. 783, 2021. DOI: 10.21577/0100-4042.20170702

CAPÍTULOS DE LIVRO

OLIVEIRA, M. A. L.; CAMPOS, G. P.; SOUZA, J. C. Q.; NASCIMENTO, M. P.; LIMA, NERILSON M.; **MOREIRA, O. B. O.**; CHELLINI, P. R.; AMORIM, T. L. Recent trends in the analysis of lipids, carbohydrates and proteins in food by capillary electrophoresis. In: María Castro-Puyana, Miguel Herrero, María Luisa Marina. (Org.). *Capillary Electrophoresis in Food Analysis*. 1ed. Singapore: Bentham Books, 2022, v., p.1-. DOI: 10.2174/97898150361521220201

TRABALHOS PUBLICADOS EM ANAIS DE CONGRESSOS

OLIVEIRA, M. A. L.; SOUZA, J. C. Q.; CANDIDO, J. M. B.; NASCIMENTO, M. P.; **MOREIRA, OLÍVIA B.O.**; CHELLINI, P. R. Capillary Electrophoresis with UV detection applied to analysis of urinary organic acids for targeted metabolomics. In: 20° Encontro Nacional de Química Analítica, 2022, Bento Gonçalves, 2022.

Macedo, A. N.; SILVA, M. C. M.; **MOREIRA, O. B. O.**; OLIVEIRA, M. A. L.; RECH, A, R. Chemical profile of Aroeira Honey from the Jequitinhonha Valley. In: 20° Encontro Nacional de Química Analítica, 2022, Bento Gonçalves, 2022.

RESUMOS PUBLICADOS EM ANAIS DE CONGRESSOS

LEAL, K. N. S.; **MOREIRA, O. B. O.**; SILVA, A. B. S.; LEMOS, L. M.; OLIVEIRA, M. A. L.; ARRUDA, M. A. Z. Development of an analytical method for determining trace elements in urine samples by ICP-MS and ICP OES. In: 46ª Reunião Anual da Sociedade Brasileira de Química, 2023, Águas de Lindóia, 2023.

MOREIRA, O. B. O.; ADRIANO, LUIZ HENRIQUE; LEMOS, L. M.; OLIVEIRA, M. A. L. Targeted metabolomics by CZE-UV associated with deep learning applied to the investigation of organic acid profile in human urine of COVID-19 patients. In: 7ª Escola de Separações, 2023, Unicamp., 2023.

MOREIRA, OLÍVIA; DE OLIVEIRA, MARCONE AUGUSTO LEAL. Response surface methodology applied to LC-QToF mass spectrometer parameters optimization for untargeted metabolomics investigations. In: III IBEROAMERICAN Conference on Mass Spectrometry, 2022, Rio de Janeiro, 2022.

SOUZA, J. C. Q. **MOREIRA, OLÍVIA B.O.**; CHELLINI, P. R.; OLIVEIRA, M. A. L. Potencialidades da eletroforese capilar de zona em análises de fármacos. In: 15° Congresso de Farmácia e Bioquímica de Minas Gerais, 2020, Belo Horizonte, 2020.

APÊNDICE III

Projeto “Investigação de estratégias para o desenvolvimento de metodologias analíticas aplicáveis ao diagnóstico de Dengue, Chikungunya e Zika em fluidos biológicos”.

INTRODUÇÃO

Segundo definição da Organização Mundial da Saúde (OMS) (*World Health Organization*, WHO), as arboviroses se enquadram no grupo de doenças infecciosas, ou seja, causadas por micro-organismos patogênicos, como bactérias, vírus, parasitas ou fungos e podem ser transmitidas direta ou indiretamente de uma pessoa a outra. A palavra arbovirose deriva do conceito de “vírus transmitidos por artrópodes”, artrópodes sendo o grupo de animais invertebrados no qual os mosquitos estão inseridos. Considerando as doenças infecciosas transmitidas por vetores, um dos mais notórios vetores é o *Aedes Aegypti*, responsável pela transmissão dos vírus da Dengue (DENV), da Chikungunya (CHIKV) e da Zika (ZIKV) (BRASIL, 2024; WHO, 2023a).

O ZIKV pertence à família *Flaviviridae* e foi identificado em primatas pela primeira vez em 1947, recebendo este nome por ter sido encontrado na floresta Zika em Uganda, África. Entre as décadas de 1960 e 1980, infecções em humanos começaram a ser reportadas e a partir do início dos anos 2000 a doença começou a expandir geograficamente, com o primeiro caso no Brasil sendo notificado em 2015 acompanhado de um surto de casos no ano seguinte, período no qual a OMS a insere na lista de doenças consideradas como emergência de saúde pública (*Public Health Emergency of International Concern*, PHEIC). O meio de transmissão principal é a mordida do mosquito *Aedes Aegypti*, no entanto, no caso do ZIKV, alguns estudos também alertam sobre a possibilidade de ser sexualmente transmissível (FIOCRUZ, 2024; KINDHAUSER et al., 2016).

A Chikungunya é causada pelo vírus CHIKV da família *Togaviridae*, reportado pela primeira vez em 1952 no sul da Tanzânia, África. O nome Chikungunya deriva de uma palavra de origem local da Tanzânia que significa “aqueles que se dobram” em referência aos primeiros pacientes acometidos, que apresentavam aparência curvada em devido aos problemas nas articulações causados pela doença. O vírus também pode ser transmitido pelo mosquito *Aedes Albopictus*, além do *Aedes aegypti*. A doença se

espalhou pela África e Ásia inicialmente e, de acordo com a OMS, desde 2005 até hoje mais de 2 milhões de casos foram reportados, incluindo a Europa e as Américas (FIOCRUZ, 2022; WHO, 2024a).

O DENV pertence ao grupo dos primeiros micro-organismos que foram denominados com o nome “vírus” no início do século XX, sendo a doença notificada desde 1845 aproximadamente. Desde que foi implementada essa nomenclatura, já havia a hipótese que o *Aedes Aegypt* seria o possível vetor da transmissão do DENV. No entanto, apenas 40 anos depois foi possível o acesso a tecnologias que permitiram obter informações relevantes sobre o vírus, incluindo o isolamento de cepas com características antigênicas distintas, consideradas então sorotipos de um mesmo vírus. Semelhante à Zika, também é um vírus da família *Flaviviridae* com quatro sorotipos (DENV-1, DENV-2, DENV-3 e DENV-4), sendo todos estes já identificados no Brasil (BEZERRA et al., 2021; PAIXÃO; TEIXEIRA; RODRIGUES, 2017; WHO, 2023b). Os sorotipos são diferentes justamente por ter diferentes interações com os anticorpos no sangue humano, podendo haver ainda alguma variação genética dentro do próprio sorotipo. De acordo com pesquisas clínicas, algumas diferenças no quadro sintomático e no prognóstico dos pacientes podem ser diferentes de acordo com cada sorotipo (CHAN et al., 2009; GUPTA et al., 2021)

Em geral, essas arboviroses especificamente têm epidemiologias e ciclos de transmissão similares, em que os indivíduos contaminados podem ser assintomáticos ou apresentar sintomas leves, moderados e graves. Quando presentes, os sintomas clínicos são muito semelhantes, sobretudo nas fases iniciais, e incluem febre alta, dor de cabeça e dores nas articulações. Destaca-se que para todas essas doenças, não há um tratamento específico, apenas a administração de medicamentos que aliviam os sintomas como analgésicos e antitérmicos, além de ingestão de líquidos. No entanto, elas têm diferenças fundamentais e complicações destoantes em fases agudas, fase na qual o tratamento desse ser bem definido (BÉLTRÁN-SILVA et al., 2018; KERKHOF et al., 2020; PAIXÃO; TEIXEIRA; RODRIGUES, 2017).

No caso da dengue deve ser feito um monitoramento dos hematócitos, em decorrência da possibilidade de dengue hemorrágica. Recomenda-se também evitar a administração de medicamentos que possam prejudicar funções hepáticas. Se o diagnóstico indicar Chikungunya, a prevalência de efeitos nas articulações em nível crônico deve ser considerada. Em relação à Zika, é necessário avaliar a potencial

transmissão materno-fetal, se pertinente, devido ao risco de microcefalia congênita, além do risco de complicações neurológicas do próprio paciente e do desencadeamento da síndrome de Guillain-Barré, que também compromete o sistema nervoso central (KERKHOF et al., 2020; MOULIN et al., 2016). Portanto, o diagnóstico assertivo nos estágios iniciais da doença é imprescindível para que a subsequente tomada de decisão seja feita de forma adequada.

Tais arboviroses são predominantes em países tropicais e de grande incidência nas Américas, principalmente na América Latina com destaque para o Brasil, atualmente o país com mais casos registrados no mundo. A dengue é listada como uma das doenças tropicais negligenciadas, grupo de doenças consideradas endêmicas em população de baixa renda (WHO, 2023c). Estima-se que, mundialmente, considerando apenas o vírus da dengue, 390 milhões de pessoas são infectadas por ano (CALVO et al., 2016; KERKHOF et al., 2020; WHO, 2023b). Nos últimos 5 anos, observou-se um crescimento no número de casos dessas doenças. Com destaque para os últimos dois anos, visto que nos anos anteriores uma possível subnotificação no número de casos em decorrência da pandemia de COVID-19 tenha ocorrido (ORGANIZAÇÃO PANAMERICANA DE SAÚDE, 2024a, 2024b, 2024c).

De acordo com especialistas, acredita-se que progressivo aumento de casos, com um importante “ressurgimento” mundial a partir do ano de 2016, está associado à rápida dispersão geográfica em consequência da adaptação dos artrópodes à vida urbana (GOULD et al., 2017; PAIXÃO; TEIXEIRA; RODRIGUES, 2017). No ano de 2023, a OMS alertou que o aumento significativo no número de casos também pode ser uma consequência das mudanças climáticas, com a dengue apresentando potencial de se tornar uma emergência global (EWE, 2023; UNITED NATIONS, 2023). Destaca-se que para desacelerar o avanço dessas doenças, a melhor estratégia de prevenção é o combate ao mosquito vetor de transmissão ao evitar o acúmulo de água parada que prolifera a criação de larvas do mosquito (TEIXEIRA; BARRETO; GUERRA, 1999).

Neste sentido, em 2022 foi lançada pela OMS junto com os departamentos de controle de doenças tropicais negligenciadas, de imunização e de biológicos, uma iniciativa global voltada ao combate às arboviroses (*Global Arbovirus Initiative*), que envolve uma estratégia integrada de atividades na qual estão previstos a colaboração multissegmentada entre órgãos públicos e setores privados, o preparo para uma possível emergência de dengue, o investimento em pesquisa e divulgação de estratégias de

prevenção e combate, o aumento da capacidade de cuidados em saúde com maior acesso da população, a participação da comunidade e a constante vigilância contra a proliferação dos vetores (BALAKRISHNAN, 2022; SRISAWAT et al., 2022; WHO, 2022).

Conforme comentado, o diagnóstico preciso e precoce é crucial para o manejo de cada uma dessas doenças. Para todas as arboviroses citadas, o exame físico aliado ao histórico do paciente é ainda a abordagem de primeira escolha. Devido à similaridade entre as manifestações sintomáticas supracitadas considerando arboviroses e outras condições com a Influenza e a própria COVID-19, distinguir a predominância de um vírus em específico baseado apenas em exame clínico, é desafiador (BÉLTRÁN-SILVA et al., 2018; CALVO et al., 2016; PAIXÃO; TEIXEIRA; RODRIGUES, 2017; PUCCIONI-SOHLER et al., 2016).

No caso da dengue, de acordo com guia disponibilizado pelo Ministério da Saúde, o quadro clínico é muito variável sendo a febre alta de início abrupto a primeira manifestação mais comum, no entanto, pode evoluir muito rápido para manifestações hemorrágicas em casos típicos de dengue hemorrágica. Em caso de sintomas leves, o diagnóstico diferencial pode incluir gripe, rubéola, sarampo e outras doenças virais e bacterianas. Já em casos moderados, a sintomatologia pode indicar além destas, meningococemia, febre amarela, malária, leptospirose, hepatite e outras doenças transmitidas por mosquitos. Quando a sintomatologia é grave e inespecífica, a recomendação é fazer a internação hospitalar com hidratação intravenosa acompanhado do monitoramento hemodinâmico, visto que não devem ser administrados medicamentos sem o diagnóstico correto (BRASIL, 2002).

Alguns exames laboratoriais são disponibilizados, porém o resultado confiável ainda é problema e, principalmente, não são facilmente encontrados e podem ter custo não acessível à população mais afetada (AHMED, 2020; KERKHOF et al., 2020; MOULIN et al., 2016; PUCCIONI-SOHLER et al., 2016). Os testes laboratoriais são baseados em métodos de isolamento, molecular ou sorológicos. O método de isolamento é o mais específico uma vez que é feita a identificação do vírus em sangue no período inicial da doença (3° ao 6° dia após aparecimento dos sintomas) (BRASIL, 2002; CORDEIRO, 2012). O teste molecular, realizado por exame de transcrição reversa seguida de reação em cadeia da polimerase (RT-PCR, *Reverse Transcription Polymerase Chain Reaction*), técnica que envolve detecção de RNA viral em amostras de sangue utilizando *kits* de extração de material genético e instrumentação automatizada, gerando

uma resposta qualitativa e quantitativa em relação à carga viral. De acordo com de Paula e Fonseca (2004), existem diversos protocolos para detecção de sorotipos de dengue por RT-PCR, e que apesar de ser uma técnica de alta sensibilidade analítica, para dengue falta uma padronização analítica desses exames (BRASIL. MINISTÉRIO DA SAÚDE., 2002; CORDEIRO, 2012; DE PAULA; FONSECA, 2004).

Mais facilmente encontrados em laboratórios de análises clínicas, os testes rápidos e sorológicos também são disponibilizados como teste diagnóstico de Dengue. Em geral, na fase inicial da doença recomenda-se o teste de antígeno NS1 por ensaios imunocromatográficos ou imunoenzimáticos, como o ELISA (*Enzyme Linked Immunosorbent Assay*) com sensibilidade analítica variando de 71 a 95 % e cerca 99 % de especificidade. Na segunda fase da doença, depois do 6º dia de sintomas, é possível fazer o teste sorológico de anticorpos IgM/IgG também por técnicas imunoenzimáticas ou imunocromatográficas, neste caso também apresentando mais de 70 % de sensibilidade e mais de 80 % de especificidade para ambos os casos (DE PAULA; FONSECA, 2004). Em geral, testes sorológicos são menos custosos, porém dependem de uma resposta do organismo à infecção, estando assim suscetíveis a falsos negativos se realizados antes do tempo, além de irem contra a ideia de diagnóstico precoce. No Brasil, esses exames são oferecidos pelo Sistema Único de Saúde (SUS), porém os resultados de imunoenaios podem demorar cerca de 15 dias para serem liberados enquanto que o RT-PCR são liberados em até 30 dias.

Em pacientes contaminados com o vírus da Zika, a primeira manifestação sintomática mais comum também inclui febre, podendo estar seguida de aparecimento de manchas na pele e dores nas articulações, também sintomas comuns às demais arboviroses (XAVIER et al., 2017). A primeira abordagem também é o diagnóstico clínico junto à avaliação das circunstâncias epidemiológicas locais (FIOCRUZ, 2016). Semelhante a dengue, os testes moleculares, de isolamento viral, sorológicos e testes rápidos também são adaptados ao vírus da Zika a partir da análise de amostras de sangue. Se o paciente for gestante, é recomendado que faça a testagem acompanhado de exames de imagem como ultrassom ou tomografias devido ao risco de alterações encefálicas no feto (BRASIL, 2024; FIOCRUZ, 2016; XAVIER et al., 2017)

O processo diagnóstico diferencial da Chikungunya também considera a dengue na fase aguda, febre reumática e artrite séptica devido ao forte comprometimento das articulações, além de outras doenças virais transmitidas por mosquitos. Semelhante às

demais, a primeira manifestação sintomática mais comum é a febre aguda e repentina e o protocolo médico é fazer a avaliação clínica e o monitoramento da progressão dos sintomas (BRASIL, 2017; MARQUES et al., 2017). As opções de testes laboratoriais descritas para as demais arboviroses supracitadas também são adaptadas para identificação do vírus da Chikungunya, porém após pesquisas nos portfólios de exames disponíveis em alguns conceituados laboratórios de análises clínicas no país, nota-se que a oferta desses exames para Zika e Chikungunya ainda é limitada.

Considerando as arboviroses, somado às questões mencionadas o maior desafio na testagem é falta de especificidade em consequência da similaridade entre as sequências de proteínas destes vírus, sobretudo dentro de vírus da mesma família como é o caso do DENV e ZIKV. Em algumas regiões geográficas inclusive, há outros vírus com características semelhantes como o Mayaro e o *Ross River*, por exemplo. Outra consequência da extensa circulação geográfica é que em alguns países como Brasil há incidência de várias dessas doenças concomitantemente. Logo, o ideal é que os dispositivos de testagem façam uma varredura plural que considera todas essas infecções, ou seja, mais do que oferecer uma resposta “sim-ou-não” para uma doença, o dispositivo deveria oferecer uma resposta assertiva em relação a qual doença o paciente pode estar infectado (KERKHOF et al., 2020).

Recentemente, algumas novas estratégias de diagnóstico e imunização dessas doenças foram disponibilizadas à população. Em relação aos novos meios de diagnósticos, a norma técnica correspondente à RDC 786/2023 publicada pela Agência Nacional de Vigilância Sanitária (ANVISA) dá as providências para o funcionamento de pontos de testagem em farmácias/drogarias justamente com o objetivo de amplificar os ambientes de testagem e facilitar o acesso da população a esses exames. Dentro da lista de testes disponibilizados, Dengue, Zika e Chikungunya estão incluídos. A partir da publicação deste regulamento técnico, são feitos testes cujos resultados são gerados em até 30 minutos como uma etapa de triagem na qual o farmacêutico irá direcionar o paciente para o especialista se houver necessidade (ANVISA, 2023; CASERTA; BAIENSE; ANDRADE, 2023).

Os testes de fluxo lateral são feitos a partir da coleta de uma gota de sangue do paciente, que ao ser misturada com uma solução tampão é inserida no dispositivo descartável. Se contido no sangue, o material genético do vírus irá reagir com um antígeno específico ao migrar pelo dispositivo. Havendo uma marcação controle e outra teste,

significa que houve reação com o antígeno, ou seja, um resultado positivo. Para a dengue são disponibilizados testes de antígeno NS1 e anticorpos, enquanto que para os demais, apenas o teste de antígeno está contido na lista de opções. Dentre o portfólio de testes também estão incluídas outras doenças que podem se assemelhar nas manifestações sintomáticas dessas arboviroses, como a própria febre amarela e outras como rubéola e influenza (CASERTA; BAIENSE; ANDRADE, 2023).

Em relação à imunização, em janeiro deste ano começaram a ser distribuídas pelo SUS as primeiras doses da vacina contra a dengue, destinadas a municípios com alto índice de transmissão (BRASIL, 2024b). Em artigo de revisão publicado por Webster e colaboradores (2009) são descritos os progressos científicos em torno da produção de uma vacina contra a dengue até a época. Na última década, estudos que compreenderam vacinas tetravalentes de vírus atenuado avançaram para fases clínicas por serem economicamente viáveis e induzir imunidade eficiente. Nesse tipo de imunizante o vírus está ativo, porém sem a capacidade de infectar o indivíduo, induzindo uma resposta imunológica do organismo e são estas que estão sendo atualmente disponibilizadas à população. Até o momento, a vacina QDenga® produzida pela Takeda Vaccines está aprovada pela ANVISA e sua distribuição está permitida no Brasil. A vacina desenvolvida pelo Instituto Butantã, também um imunizante tetravalente de vírus atenuado, se encontra em fase final de estudo clínico com previsão para finalização ainda em 2024 (BRASIL, 2024b; INSTITUTO BUTANTÃ, 2024; WEBSTER; FARRAR; ROWLAND-JONES, 2009)

A vacina de Zika, feita com vírus inativado, também está sendo produzida pelo Instituto Butantã na intenção de prevenir o avanço da microcefalia em recém-nascidos, com fase de estudos pré-clínicos previstos para serem iniciados em 2024. De acordo com artigo de revisão publicado por Wang e colaboradores em 2022, há muitos aspectos ainda desconhecidos sobre o ZIKV, o que desacelerou os avanços nas pesquisas de uma vacina (TAVARES, 2024; WANG et al., 2022). Além disso, por ser doenças com fases agudas endêmicas, existe uma certa dificuldade de avançar continuamente com estudos clínicos. Por outro lado, as pesquisas relacionadas ao desenvolvimento de vacinas para Chikungunya encontram-se em estágios mais avançados. A vacina de vírus atenuado desenvolvida pelo Instituto Butantã já foi aprovada pela agência reguladora americana (*Food and Drug Administration*, FDA) e seu registro está atualmente sendo avaliado pela ANVISA (SCHMIDT; SCHNIERLE, 2022; TAVARES, 2023).

Por fim, em resposta à iniciativa global ao combate às arboviroses, embora um grande avanço tenha sido alcançado durante a última década, o aumento no número de casos também vem sendo expressivo. Assim, muitas pesquisas voltadas para este tema ainda estão sendo o foco da comunidade científica levando em consideração todos os desafios e objetivos citados anteriormente. No que diz respeito ao diagnóstico, alguns artigos de revisão já foram publicados considerando os últimos avanços. Métodos baseados na ampliação isotérmica de material genético aparecem como uma alternativa mais acessível ao RT-PCR, em termos de custo para o paciente. Recentemente houve um avanço nos chamados testes *Point-of-Care* (POC), sem tradução para o português, que são dispositivos destinados à realização de testes fora do ambiente de laboratório. Neste contexto, são encontradas pesquisas voltadas para o desenvolvimento de Lab-on-a-chip (LOC), microfluídica em papel e outras estratégias de melhorias nos testes em fluxo laminar supracitados (VARGHESE; DE SILVA; MILLAR, 2023). Com o recente avanço das ciências voltadas ao estudo do metaboloma humano aliado às ferramentas de investigação química abrangentes como os espectrômetros de massas, ciências de dados e inteligência artificial, tecnologias promissoras vêm sendo elaboradas a fim de suprir as necessidades específicas da testagem para arboviroses (BYERS et al., 2019; NETO et al., 2022).

CEP DENGUE, ZIKA, CHIKUNGUNYA

Em relação às amostras de urina e saliva da pesquisa que considera Dengue, Zika e Chikungunya, a coleta foi autorizada pelo Comitê de Ética da Universidade Federal de Juiz de Fora (CEP UFJF) sob o parecer de número 5.267.924, na qual prevê o recolhimento de 500 amostras de urina e 500 de saliva, também a partir de parceria com Lemos Laboratório de Análises Clínicas. Descrito no TCLE correspondente, foi dada a opção ao voluntário de escolher o fluido a ser doado ou a doação dos dois pelo mesmo indivíduo. Da mesma forma, foram incluídos indivíduos maiores de 18 anos que concordem formalmente em participar da pesquisa e que fossem submetidos ao teste laboratorial para confirmação do diagnóstico da doença suspeita.

Devido ao grande foco na COVID-19 durante os anos iniciais dessa pesquisa, a demanda por exames de arboviroses foi reduzida em grande escala, o que dificultou a coleta do material. Desse modo, na intenção de expandir a possibilidade de aquisição dessas

amostras, ações foram promovidas para inserção de uma emenda no projeto aprovado junto ao CEP-UFJF que viabilizasse a participação das Unidades Básicas de Saúde (UBS) da cidade de Juiz de Fora como ponto de coleta.

Em reuniões com a Secretaria de Saúde da Prefeitura de Juiz de Fora e com a Faculdade de Enfermagem da UFJF, foi planejado que enfermeiros inscritos no Programa de Residência Multidisciplinar em Saúde Familiar da UFJF que já trabalhem nas UBSs da cidade fossem responsáveis pela abordagem de potenciais voluntários e coleta do material. Sendo os voluntários, indivíduos que tenham procurado atendimento médico em UBS distribuídas na cidade de Juiz de Fora em vista do reconhecimento de sintomas relacionados a algumas destas doenças. Seriam coletadas amostras de saliva e/ou urina de voluntários que também coletaram sangue para exame confirmatório a ser realizado pelo Laboratório Central da Prefeitura de Juiz de Fora (LACEN), cujos resultados também nos seriam disponibilizados via plataforma virtual. No entanto, devido a alguns percalços burocráticos, o adendo não foi protocolado junto ao CEP correspondente, assim não sendo possível realizar a coleta das amostras previstas inicialmente para o projeto contendo as arboviroses.

FTIR E MACHINE LEARNING

As amostras selecionadas para essa etapa do trabalho descritas no item 4.4 já foram analisadas por FTIR em mesmo espectrômetro Alpha FTIR (Bruker) seguindo os mesmos parâmetros instrumentais e condições experimentais descritas na seção e.7.1.1. Foram gerados 205 espectros contendo amostras de pessoas com histórico de contaminação por Dengue, Zika, Chikungunya, COVID-19 e indivíduos saudáveis. As amostras foram selecionadas de forma a montar um conjunto de dados com aproximadamente 50 % de amostras entre positivas. De modo complementar, aquelas amostras cujo paciente declarou ser infectado por mais de uma destas doenças também foram analisadas, totalizando 14 amostras.

O objetivo para esse conjunto de dados é construir um modelo preditivo para as cinco classes de amostras. Os demais estudos discutidos ao longo dos capítulos dependeram apenas de classificações binárias, portanto, novos estudos envolvendo a utilização desses algoritmos estão em andamento. Pretende-se ainda utilizar o fluxograma empregado nos dados para a COVID-19 para classificação dois a dois utilizando esse

conjunto de dados, considerando as amostras de dengue em relação às sadias e dengue em relação à COVID-19, visto que são os dados em maior volume. Além disso, os dados de dupla infecção serão testados nos modelos durante validação para avaliar o comportamento dos mesmo frente a essa situação. Para essas amostras foi escolhido fazer as primeiras análises por infravermelho para avaliar se exibirão um perfil passível de predição. De posse desses primeiros resultados, se pertinente, os outros métodos otimizados com capacidade de identificação de biomarcadores também serão aplicados às amostras futuramente.