UNIVERSIDADE DE SÃO PAULO
INSTITUTO DE QUÍMICA
DEPARTAMENTO DE QUÍMICA ANALÍTICA

ALGORITMO GENÉTICO E ESPECTROSCOPIA NO INFRAVERMELHO
– ALGUMAS APLICAÇÕES NA INDÚSTRIA COSMÉTICA

Marcos Coelho Amendola

Tese de Mestrado

Orientador: Prof. Dr. Ivano Gebhardt Rolf Gutz

São Paulo/SP

Data de depósito do trabalho na SPG: 06/11/2006

São Paulo, 02 de Janeiro de 2007

**FICHA CATALOGRÁFICA ELABORADA PELA
BIBLIOTECA DO INSTITUTO DE QUÍMICA DA USP**

Dissertação de mestrado submetida ao Instituto de Química da Universidade de São Paulo como parte dos requisitos necessários para obtenção do grau de mestre na área de Química Analítica. Aprovada pela comissão julgadora em 29 de Janeiro de 2007.

BANCA EXAMINADORA

Prof. Dr. Ivano G. R. Gutz

IQ – USP

(Orientador)

João Pedro Simon Farah

IQ-USP

Ronei Jesus Poppi

Unicamp - Campinas

Aos meus pais, Mariza e Valter,
e a minha querida esposa, Selma.

AGRADECIMENTOS

Ao prof. Dr. Ivano G. R. Gutz por orientar, incentivar e valorizar a realização deste trabalho.

Aos meus pais Valter e Mariza, e a minha avó, Delphina, pelos ensinamentos, apoio e confiança.

À minha esposa Selma, por compartilhar e apoiar esta jornada.

À minha irmã, Cristiane, e ao meu irmão, Fábio, pela convivência e companherismo.

À Colgate Palmolive e a Luiza Okubo, pelas oportunidades oferecidas para a execução deste trabalho.

A todos os amigos e companheiros da Colgate Palmolive, pela valiosa troca de idéias e momentos de descontração.

CURRÍCULO RESUMIDO

NOME COMPLETO: Marcos Coelho Amendola

LOCAL DE NASCIMENTO: São Paulo – SP

DATA DE NASCIMENTO: 15/02/77

EDUCAÇÃO

- Técnico em Química pela Escola Técnica Walter Belian (1992 – 1995).
- Bacharel em Química pela Universidade de São Paulo, com especialização em Química Industrial e Biotecnologia (1996 – 1999).
- Mestrado em Química pela Universidade de São Paulo. Título: Algoritmo genético e espectroscopia no infravermelho – Algumas aplicações na indústria cosmética.

EXPERIÊNCIA PROFISSIONAL

- Analista Químico na Colgate Palmolive (mar/2000 – nov/2005).
- Químico Instrumental na indústria Colgate Palmolive (dez/2006 – abril/2006).
- Coordenador de Produtos e Processos na indústria Colgate Palmolive (maio/2006 – atual).

RESUMO

ALGORITMO GENÉTICO E ESPECTROSCOPIA NO INFRAVERMELHO – ALGUMAS APLICAÇÕES NA INDÚSTRIA COSMÉTICA

Este trabalho discute o desenvolvimento de um algoritmo genético escrito em linguagem VBA para Excel e suas aplicações. O algoritmo elaborado foi utilizado em combinação com a técnica de FTIR-ATR para o desenvolvimento de metodologias aplicáveis na indústria cosmética e de saneantes, tais como a quantificação de surfactantes e bactericidas. Algumas modificações introduzidas no algoritmo foram estudadas através das aplicações selecionadas, destacando-se a introdução de técnicas de paralelismo que possibilitam a quantificação de mais de um analito em uma mesma execução do algoritmo. Este tipo de técnica foi aplicado na quantificação o-benzil p-cloro fenol, o-fenil fenol e etanol, em uma mistura dos três componentes, utilizada como matéria prima (bactericida) na indústria de saneantes. Também com o uso do algoritmo, FTIR-ATR e calibração linear múltipla, foram desenvolvidos métodos para determinação do ingrediente ativo total da mistura lauril sulfato de amônio/lauril éter sulfato de amônio (matéria prima para shampoos), e ainda da concentração de lauril éter sulfato de sódio em amostras de shampoo. Comparados aos métodos usuais de titulação, cromatografia líquida ou gasosa, os novos métodos se distinguem por não requererem preparo algum da mostra, nem consumirem solventes orgânicos, sendo mais rápidos e diretos, oferecendo exatidão comparável.

ABSTRACT

GENETIC ALGORITHM AND INFRARED SPECTROSCOPY – SOME APPLICATIONS IN THE COSMETIC INDUSTRY

A genetic algorithm was implemented in VBA for Excel and applied in combination with FTIR-ATR to develop analytical methods for some raw materials and finished products of the home care and personal care industries, like the quantification of surfactants and bactericides. Some modifications were introduced in the traditional genetic algorithm and evaluated for the mentioned applications, with emphasis on parallelism techniques that extend the algorithm to the definition of methods for more than one analyte simultaneously. The parallelistic approach was used for the quantification of o-benzyl p-chloro phenol, o-phenyl phenol and ethanol in a mixture of these three components (used as bactericides in home care products). Using the genetic algorithm, FTIR-ATR and multiple linear calibration it was also possible to develop methods to quantify the active ingredient of the ammonium lauryl sulfate/ammonium lauryl ether sulfate mixture (raw material for shampoos), and the concentration of sodium lauryl ether sulfate in commercial samples of shampoos. In comparison with usual methods of analysis, like titrations, gas chromatography or liquid chromatography, the new developed methods are faster and more direct, with comparable accuracy. Furthermore, they do not require any kind of sample pretreatment nor the use of organic solvents.

ÍNDICE

INTRODUÇÃO GERAL	1
1.1. Espectro Infravermelho e vibrações moleculares	1
1.2. O modelo do oscilador harmônico	2
1.3. O modelo do oscilador anarmônico	5
1.4. Espectroscopia no infravermelho baseada na transformada de Fourier (FTIR)	7
1.5. A reflectância total atenuada	11
OBJETIVOS DO TRABALHO	14
PARTE A: DESENVOLVIMENTO DO ALGORITMO GENÉTICO	15
A.1. INTRODUÇÃO	15
A.1.1. Algoritmo Genético	15
A.1.2. Programas de otimização baseados no algoritmo genético	23
A.2. PARTE EXPERIMENTAL	24
A.3. RESULTADOS E DISCUSSÃO	26
A.3.1. Definição dos parâmetros e elitismo	27
A.3.2. Geração da população inicial e codificação	28
A.3.3. Avaliação e classificação	30
A.3.4. Cruzamento e mutação	31
A.3.5. Saída de resultados e critério de parada	32
A.3.6. Migração	35

PARTE B: QUANTIFICAÇÃO DE LAURIL ÉTER SULFATO DE AMÔNIO 28%	39
B.1. INTRODUÇÃO	39
B.1.1. Alquil Sulfatos e Alquil Éter Sulfatos	39
B.1.2. Determinação de surfactantes – titulação volumétrica e potenciométrica	41
B.2. PARTE EXPERIMENTAL	44
B.3. RESULTADOS E DISCUSSÃO	48
 PARTE C: QUANTIFICAÇÃO DE O-BENZIL P-CLORO FENOL, O-FENIL FENOL E ETANOL (OPP/OBCP/ETANOL) NA MISTURA DOS TRÊS COMPONENTES	 55
C.1. INTRODUÇÃO	55
C.1.1. Os bactericidas o-fenil fenol e o-benzil p-cloro fenol	55
C.2. PARTE EXPERIMENTAL	56
C.3. RESULTADOS E DISCUSSÃO	59
 PARTE D: QUANTIFICAÇÃO DE LAURIL ÉTER SULFATO DE SÓDIO EM AMOSTRAS DE SHAMPOO	 73
D.1. INTRODUÇÃO	73
D.1.1. Formulações de shampoo	73
D.2. PARTE EXPERIMENTAL	76
D.3. RESULTADOS E DISCUSSÃO	79
 PARTE E: ESTUDO DA INFLUÊNCIA DOS PARÂMETROS DE CONFIGURAÇÃO NO DESEMPENHO DO ALGORITMO GENÉTICO	 88
E.1. INTRODUÇÃO	88
E.2. PARTE EXPERIMENTAL	90
E.3. RESULTADOS E DISCUSSÃO	93

COMENTÁRIOS GERAIS E PERSPECTIVAS	102
CONCLUSÕES	104
REFERÊNCIAS BIBLIOGRÁFICAS	107

LISTA DE FIGURAS

Figura 1: Exemplo de oscilador harmônico _____	3
Figura 2: Curva de energia para o oscilador harmônico simples ^[2] _____	4
Figura 3: Curva de energia para o oscilador anarmônico ^[2] _____	6
Figura 4: Esquema simplificado de um equipamento de FTIR _____	7
Figura 5: Diagrama esquemático do FTIR Spectrum BX (X), utilizado no presente trabalho ^[6] . _	8
Figura 6: Interferograma ilustrativo obtido com o equipamento Perkin Elmer Spectrum BX. ____	9
Figura 7: Exemplo de reflectância total atenuada _____	12
Figura 8: Acessório de reflectância total atenuada utilizado neste trabalho _____	13
Figura 9: Fluxograma das etapas de um algoritmo genético _____	17
Figura 10: Exemplo de seleção roleta. Imagine-se que, a cada seleção, a seta pare num ponto aleatório da barra de probabilidades representada; depreende-se que os cromossomos com respostas mais altas tem maior probabilidade de serem escolhidos frente aos demais _____	19
Figura 11: Representação do cruzamento simples _____	20
Figura 12: Exemplo dos primeiros dados de saída (resultados) do algoritmo _____	33
Figura 13: Representação esquemática da avaliação da resposta por diferentes populações ____	36
Figura 14: Representação esquemática da migração _____	36
Figura 15: Estruturas do Azul de Dissulfino, Brometo de Dimínio e Hiamina 1622 _____	43
Figura 16: Curva de titulação potenciométrica típica para LESA 28% _____	46
Figura 17: Espectro típico de Lauril Éter Sulfato de Amônio 28% _____	48
Figura 18: Espectros de LESA 28% e LESS 27% _____	50
Figura 19: Desenvolvimento de cada população quanto às respostas (R^2) , para LESA 28%. ____	52
Figura 20: Comprimentos de onda selecionados em uma das possíveis resoluções para quantificação de LESA 28% _____	54
Figura 21: Fórmulas estruturais para o-Fenil Fenol e o-Benzil p-Cloro Fenol _____	55
Figura 22: Espectros de OPP 50% em etanol, OBCP 20% em etanol, etanol puro e da mistura OPP/OBCP 70% em etanol _____	60

Figura 23: Curva de regressão e números de onda (cromossomo, em cm^{-1}) para o analito OPP. A figura também indica a correlação do cromossomo para cada analito _____	62
Figura 24: Resíduo (diferença entre o valor teórico e o previsto) para cada amostra em relação ao teor de OPP esperado. As amostras controle são indicadas no gráfico _____	63
Figura 25: Curva de regressão e números de onda (cromossomo, em cm^{-1}) para o analito OBCP. A figura também indica a correlação do cromossomo para cada analito _____	63
Figura 26: Resíduo (diferença entre o valor teórico e o previsto) para cada amostra em relação ao teor de OBCP esperado. As amostras controle são indicadas no gráfico _____	64
Figura 27: Curva de regressão e números de onda (cromossomo, em cm^{-1}) para o analito EtOH. A figura também indica a correlação do cromossomo para cada analito _____	64
Figura 28: Resíduo (diferença entre o valor teórico e o previsto) para cada amostra em relação ao teor de EtOH. As amostras controle são indicadas no gráfico _____	65
Figura 29: Estrutura química de uma dietanolamida _____	74
Figura 30: Estrutura química de um alquilpoliglicosídeo _____	75
Figura 31: Estrutura química da cocoamidopropilbetaína _____	76
Figura 32: Planejamento experimental para bases de shampoos _____	77
Figura 33: Espectro FTIR do lauril éter sulfato de sódio 27% (LESS 27%) _____	81
Figura 34: Comparação entre os espectros FTIR de LESS 27%, de LESS diluído à 10% em água e de um shampoo contendo 10% de LESS _____	81
Figura 35: Comprimentos de onda escolhidos pelo algoritmo, nas regiões pré-selecionadas ____	82
Figura 36: Curva de regressão linear para as amostras de calibração _____	83
Figura 37: Resultados de titulação potenciométrica e FTIR para as amostras-teste _____	83
Figura 38: Avaliação da Linearidade do Método de FTIR (amostras de validação) _____	86
Figura 39: Avaliação da precisão e exatidão do método de FTIR (amostras de validação, níveis 70%, 100% e 130%) _____	87
Figura 40: Estrutura típica dos dimetil polisiloxanos ^[55] _____	89
Figura 41: Espectro infravermelho do silicone 60000 _____	93
Figura 42: Espectro infravermelho típico de um sabonete comum _____	94
Figura 43: Espectros FTIR calculados, simulando 20 concentrações diferentes de silicone no sabonete, com indicação das regiões onde a linearidade foi preservada _____	95
Figura 44: Detalhe da região próxima a $775 - 791 \text{ cm}^{-1}$, para os 20 espectros simulados _____	95

Figura 45: Detalhe da região próxima a $1250 - 1270 \text{ cm}^{-1}$, para os 20 espectros simulados____ 96

Figura 46: Sumário dos resultados do planejamento experimental da tabela 15. As cores fazem referência à média da pontuação recebida nos testes em duplicata (também indicada na figura): pontuação baixa (amarelo), média (laranja) e alta (vermelho)_____ 100

LISTA DE TABELAS

Tabela 1: Estrutura do algoritmo genético desenvolvido _____	26
Tabela 2: Exemplos de codificação com os sequenciais 1-5 _____	29
Tabela 3: Tempo gasto pelo algoritmo de acordo com o número de gerações. _____	34
Tabela 4: Tempo gasto pelo algoritmo de acordo com o número de gerações. _____	35
Tabela 5: Concentrações de LESA 28 % utilizadas na calibração _____	47
Tabela 6: Concentração das amostras preparadas de OPP, OBCP e EtOH em relação a cada analito _____	57
Tabela 7: Comparação entre a concentração real e os resultados obtidos por FTIR-ATR para o analito OPP. _____	65
Tabela 8: Comparação entre a concentração real e os resultados obtidos por FTIR-ATR para o analito OBCP. _____	67
Tabela 9 : Comparação entre a concentração real e os resultados obtidos por FTIR-ATR para o analito EtOH. _____	69
Tabela 10: Média dos módulos de cada erro absoluto e módulo do erro máximo observado para cada analito, para cada conjunto de amostras. _____	71
Tabela 11: Regiões de limitação de números de onda para cada gene _____	79
Tabela 12: Resultados obtidos (FTIR) e esperados (Titulação Potenciométrica) para as amostras de shampoo _____	84
Tabela 13: Resultados Obtidos para as amostras de validação _____	86
Tabela 14: Concentrações teóricas de silicone nos espectros simulados _____	91
Tabela 15: Parâmetros e resultados do planejamento experimental para estudo de desempenho do algoritmo genético _____	98

INTRODUÇÃO GERAL

1.1. Espectro Infravermelho e vibrações moleculares

Por não ser a espectroscopia de absorção no infravermelho uma técnica corriqueira nos laboratórios analíticos em geral, será apresentada uma breve introdução conceitual e instrumental, sendo que uma visão mais aprofundada poderá ser encontrada na bibliografia especializada^[1-4].

A região do espectro eletromagnético correspondente ao infravermelho se estende de comprimentos de onda de aproximadamente 0,75 μm até quase 1 mm. A região do infravermelho médio está situada entre 2,5 – 25 μm (4000 – 400 cm^{-1}). A região de mais baixa frequência (600 – 200 cm^{-1}) é conhecida como infravermelho longínquo e a região de mais alta frequência (4000 cm^{-1} até a região do visível) como infravermelho próximo.

O espectro infravermelho é originário das vibrações dos átomos de uma molécula. As frequências das vibrações observadas no espectro podem ser relacionadas com os grupos funcionais presentes na molécula (principalmente na região do infravermelho médio, onde as atribuições são mais fáceis) tornando a técnica bastante útil na identificação de substâncias. Além disso, em grande parte dos casos a absorção de radiação obedece à lei de Lambert-Beer, tornando a técnica também útil em análise quantitativa.

A condição para que ocorra absorção da radiação infravermelha é que haja variação do momento de dipolo elétrico da molécula como consequência do seu movimento vibracional ou rotacional (o momento de dipolo é determinado pela magnitude da diferença de carga e a distância entre dois centros de carga). Somente nestas circunstâncias, o campo elétrico alternante da radiação incidente interage com a molécula, originando os espectros de absorção. Expresso de

outra maneira, isto significa que o espectro de absorção no infravermelho tem origem quando a radiação eletromagnética incidente tem uma componente com frequência correspondente a uma transição entre dois níveis vibracionais.

A vibração molecular mais simples é chamada estiramento e envolve a aproximação ou distanciamento dos átomos envolvidos no plano. Uma molécula contendo n átomos tem um total de $3n$ graus de liberdade, correspondendo à coordenadas cartesianas de cada átomo na molécula. Em uma molécula não linear, 3 dos graus de liberdade são rotacionais e 3 são translacionais, sendo que os demais são vibracionais. Em uma molécula linear temos 2 graus de liberdade rotacionais e 3 translacionais. Desta forma, o número total de vibrações previstas é $3n - 6$ para molecular não lineares e $3n - 5$ para moléculas lineares.

Em relação à vibrações moleculares, é importante lembrar que:

- Todos os átomos da molécula se movem periodicamente e em fase.
- O centro de massa da molécula permanece inalterado.
- A molécula não muda de orientação ou posição com a vibração.

1.2. O modelo do oscilador harmônico

O modelo mais simples para descrever a frequência da vibração é a do oscilador harmônico ideal, baseado na lei de Hook. Nesta aproximação, dois átomos conectados por uma ligação são tratados como um oscilador harmônico simples composto de 2 esferas iguais (os átomos) conectados por uma mola (a ligação), conforme mostra a figura 1.

Figura 1: Exemplo de oscilador harmônico

De acordo com a lei de Hook, a frequência da vibração (ν) é relacionada com a massa das esferas (m) e com a constante de força da mola (k), pela equação 1:

Equação 1
$$\nu = \frac{1}{2\pi} \sqrt{\frac{k}{m}}$$

No oscilador harmônico clássico, a energia é descrita pela equação 2:

Equação 2
$$E = \frac{1}{2} kx^2 = h\nu$$

Onde x é o estiramento da mola. Desta forma, a energia e a frequência são dependentes da distância da compressão ou estiramento entre as esferas. Se este modelo fosse 100% condizente com a realidade, uma molécula poderia absorver energia em qualquer comprimento de onda. Todavia, a vibração molecular é quantizada, de forma que as transições só podem ocorrer entre determinados níveis de energia, de acordo com a equação 3.

Equação 3
$$E = (n + \frac{1}{2}) h\nu$$

Onde ν é a frequência da vibração e n é o número quântico (0, 1, 2, ...). A curva da energia para um oscilador harmônico simples é ilustrada na figura 2.

Figura 2: Curva de energia para o oscilador harmônico simples^[2]

A menor energia é $E_0 = 1/2 hv$, ($n = 0$) a próxima é $E_1 = 3/2 hv$ ($n = 1$) e assim por diante. De acordo com este modelo, somente são permitidas as transições entre níveis vibracionais adjacentes, sendo que a diferença entre os níveis vibracionais é sempre a mesma (ou seja, todas as transições possuem frequências iguais a da transição fundamental).

Para molécula diatômicas, a equação 4 pode ser derivada da lei de Hook :

Equação 4

$$\bar{\nu} = \frac{1}{2\pi c} \sqrt{\frac{f(m_1 + m_2)}{m_1 m_2}}$$

Onde $\bar{\nu}$ é o número de onda da vibração (cm^{-1})

m_1 e m_2 são as massas dos átomos 1 e 2 respectivamente (g).

c é a velocidade da luz (cm/s)

f é a constante de força da ligação (dina/cm)

Da equação 4, conclui-se que quanto maior a constante de força da ligação, maior a frequência da vibração (assim, uma ligação C=C terá frequência vibracional maior que uma ligação C-C) e que quanto maior a massa dos átomos envolvidos, menor será a frequência.

Embora a equação 4 seja uma boa aproximação, deve-se levar em consideração que o movimento de dois átomos em uma molécula grande não pode ser isolado do movimento do restante dos átomos.

1.3. O modelo do oscilador anarmônico

Em se tratando de moléculas reais, o comportamento das moléculas é melhor representado pelo modelo do oscilador anarmônico. A expressão de energia potencial para este modelo inclui termos de ordem superior:

Equação 5
$$V = \frac{1}{2} kx^2 + k'x^3 + \dots$$
 sendo que $k' < k$

A função de Morse (equação 6) é a equação empírica mais utilizada para descrever a energia potencial da molécula.

Equação 6
$$V = D_e (1 - e^{-\beta x})^2$$

Onde D_e é a energia de dissociação da ligação, β é uma constante e x é o deslocamento dos átomos envolvidos.

Aplicando a teoria quântica à equação de Morse, a energia vibracional pode ser descrita pela equação 7:

Equação 7
$$E = h\nu \left(v + \frac{1}{2} \right) - x_m h\nu \left(v + \frac{1}{2} \right)^2$$

A curva de energia para este modelo é mostrada na figura 3, representando um oscilador anarmônico.

Figura 3: Curva de energia para o oscilador anarmônico^[2].

Como mostrado na figura, conforme a distância interatômica aumenta, a energia chega a um máximo, quando então a ligação se rompe.

O modelo do oscilador anarmônico também prevê a presença de sobretons e bandas de combinação nos espectros. Sobretons (*overtones*) são bandas com valores de frequência

correspondentes a múltiplos inteiros das vibrações fundamentais. Bandas de combinação são combinações lineares das bandas fundamentais ou múltiplos inteiros das mesmas.

1.4. Espectroscopia no infravermelho baseada na transformada de Fourier (FTIR)

Os equipamentos atuais de FTIR podem ser basicamente esquematizados conforme mostra a figura 4:

Figura 4: Esquema simplificado de um equipamento de FTIR.

A fonte é formada basicamente por uma cerâmica condutora ou por um metal revestido por cerâmica. Quando a cerâmica é aquecida, a emissão de radiação ocorre em todos os comprimentos de onda e a mesma aumenta com a temperatura (por isso o controle rigoroso da temperatura é essencial para a obtenção de espectros reproduzíveis). A temperatura típica de aquecimento varia entre 1000 - 1800 °C^[5].

O interferômetro é o coração do instrumento. O interferômetro clássico de Michelson envolve um divisor de feixe (vide item 16, figura 5), que é um componente que reflete metade da radiação incidente e transmite o resto. A parte da radiação que é transmitida encontra um espelho móvel (item 17, figura 5) e é novamente refletida de forma a ser recombinada com a radiação anterior (inicialmente refletida pelo divisor de feixe).

1) Laser HeNe	10) Janela
2) Janela	11) Compartimento da Amostra
3) Envoltório de Proteção	12) Foco da radiação Infravermelha
4) Blindagem	13) Espelho côncavo ajustável
5) Fonte	14) Detector para o laser
6) Espelho ajustável para o Laser	15) Espelho plano do interferômetro
7) Detector	16) Divisor de feixe
8) Espelhos Côncavos	17) Espelho móvel do interferômetro
9) Janela	

Figura 5: Diagrama esquemático do FTIR Spectrum BX (X), utilizado no presente trabalho^[6].

Conforme o espelho móvel se desloca, o feixe de radiação refletido no mesmo interage com a radiação que foi refletida pelo divisor de feixe, gerando interferências construtivas e destrutivas, descritas pela função cosseno. Se a radiação for composta por diversos comprimentos de onda,

teremos como resultado a soma de várias funções cosseno, típicas de todos os comprimentos de onda incidentes. O resultado final é chamado de interferograma (figura 6):

Figura 6: Interferograma ilustrativo obtido com o equipamento Perkin Elmer Spectrum BX.

A conversão da função cosseno versus diferença de caminho óptico em espectro (transmitância versus número de onda) é efetuada por um algoritmo denominado transformada de Fourier. Matematicamente, a transformada de Fourier requer uma integração de mais infinito à menos infinito. Como o deslocamento do espelho móvel é finito, é necessário um tratamento matemático adicional denominado apodização^[3,6]. Espectros de gases (bandas estreitas) são mais sensíveis à apodização.

Existem diversos tipos de detectores para FTIR, sendo que os mais comuns são foto-resistores, isto é, os mesmos possuem uma alta resistência no escuro, que decai

proporcionalmente com a incidência de radiação. Entre os detectores mais comuns estão o sulfato de triglicina deuterado (DTGS)^[7-8] e o telureto de mercúrio e cádmio (criogenicamente utilizado).

O sinal de saída do detector é amplificado e convertido em um sinal de tensão que varia com o tempo e depois convertido de analógico para digital. Um laser de He/Ne de 316,4 nm é normalmente utilizado para controlar a frequência de conversão do sinal, medir a diferença de caminho ótico entre o divisor de feixe e o espelho móvel e ainda ajudar a corrigir variações de temperatura na fonte.

Uma das características dos equipamentos FTIR é que a medida é extremamente rápida, já que ocorre em todos os comprimentos de onda simultaneamente e o cálculo da transformada é realizado por modernos e velozes processadores. Esta agilidade permite a obtenção de diversos espectros da mesma amostra (*scans*) em questão de segundos e a média dos *scans* possibilita a melhora da relação sinal-ruído (a melhora na relação sinal-ruído é proporcional a raiz quadrada do número de *scans*).

Antes de se iniciar a obtenção do espectro da amostra é necessário fazer um “espectro de fundo” ou “*background*”, ou seja, adquirir um espectro sem nenhuma amostra no compartimento. Este *background* terá absorções características de qualquer umidade ou gás carbônico que estejam no caminho ótico, bem como qualquer outro tipo de perda em virtude das respostas características do equipamento (detector/fonte/divisor de feixe). Quando se adquire um espectro, o *background* é subtraído de forma que o resultado final é somente o espectro da amostra, sem as interferências citadas anteriormente.

Outro parâmetro importante é a resolução utilizada para adquirir os espectros. O padrão mais usual da indústria está relacionado com a densidade de pontos da transformada de Fourier. Desta forma, considera-se a metade da resolução como a densidade de pontos (ou seja se a

resolução configurada for 2 cm^{-1} , teremos um ponto espectral para 4000, 3999 401, 400 cm^{-1} , etc). Para amostras líquidas, a resolução de 2 cm^{-1} é normalmente suficiente, já que as bandas são usualmente largas. Resoluções maiores requerem que o espelho móvel percorra distâncias maiores (o tempo de aquisição do espectro aumenta).

1.5. A reflectância total atenuada

A espectroscopia de reflectância total atenuada^[3-5] é uma técnica que permite a obtenção direta de espectros de pastas, adesivos e sólidos, sem a necessidade de dissolução ou preparo de pastilhas. A técnica baseia-se no fato de que quando um feixe de radiação passa de um meio mais denso (cristal da janela do dispositivo de ATR) para um meio menos denso (amostra), ocorre refração e/ou reflexão. A fração do feixe de luz incidente que é refletida aumenta conforme o ângulo de incidência, e quando excede um determinado ângulo crítico a reflexão é completa. No ponto de reflexão, o feixe atua como se penetrasse a uma pequena distância dentro da amostra, de forma que esta componente da radiação –designada de onda evanescente – pode ser absorvida pela amostra em contato com a superfície da janela (neste caso, se refere-se à radiação como “atenuada”). Dependendo da configuração do acessório utilizado, é possível fazer com que a radiação sofra diversas reflexões internas e conseqüentes atenuações (vide figura 7), aumentando a sensibilidade da técnica que, de qualquer forma, não é grande, dado o curto “caminho ótico” proporcionado pelas ondas evanescentes.

Figura 7: Exemplo de reflectância total atenuada

A profundidade da penetração (d_p) é dada pela equação 8:

Equação 8
$$d_p = \lambda / \{2\pi n_1 [\sin^2(\theta) - (n_2/n_1)^2]^{1/2}\}$$

Onde:

n_1 = índice de refração do cristal ATR

n_2 = índice de refração da amostra

θ = ângulo de incidência da radiação

λ = comprimento de onda da radiação

O uso do diamante como material de janela para ATR permitiu o desenvolvimento de um grande número de aplicações, devido às suas importantes propriedades em relação à análise de infravermelho. O alto índice de refração do diamante é adequado para obtenção de espectros de excelente qualidade para a maior parte das substâncias, com grau de penetração da radiação na amostra em torno de 1-2 μm (45° , 1000 cm^{-1}). Todavia, a sensibilidade pode ser insuficiente para análise de traços, vez que o caminho ótico para ATR com 9 reflexões se situa ao redor de 10-20 μm e o solvente pode encobrir regiões apreciáveis do espectro.

A dureza e estabilidade química do diamante permite a obtenção direta de espectros de material em meio aquoso, substâncias abrasivas, quimicamente hostis, e também permite a aplicação de pressão sobre as amostras (importante para análise de material sólido). Além disso, o diamante transmite luz do UV-Vis ao infravermelho, permitindo a obtenção de espectros em uma vasta gama de comprimentos de onda, dependendo da configuração do acessório.

A figura 8 mostra o acessório de reflectância atenuada (9 reflexões) com cristal de diamante utilizado neste trabalho.

Figura 8: Acessório de Reflectância total atenuada utilizado neste trabalho

OBJETIVOS DO TRABALHO

- Desenvolver um algoritmo genético de fácil utilização e que não dependa de *software* comercial, além do Excel (disponível na maioria dos computadores) com aplicabilidade voltada para a determinação de comprimentos de onda em calibrações multivariadas, visando otimizar o tempo envolvido no desenvolvimento de metodologias analíticas baseadas na espectroscopia.
- Estudar as etapas e parâmetros envolvidos na configuração do algoritmo genético clássico, buscando introduzir melhorias de forma a melhorar a performance do mesmo. Testar a implementação pseudo-paralela do algoritmo.
- Utilizar e testar o algoritmo genético desenvolvido em aplicações práticas envolvendo espectroscopia e calibração linear múltipla.
- Explorar as potencialidades da técnica FTIR-ATR no desenvolvimento de aplicações de interesse na indústria cosmética e de saneantes, visando a obtenção de métodos rápidos e robustos que possam ser aplicados no controle de qualidade de matérias primas e produtos acabados.

PARTE A: DESENVOLVIMENTO DO ALGORITMO GENÉTICO

A.1. INTRODUÇÃO

A.1.1. Algoritmo Genético

O algoritmo genético (AG) faz parte dos chamados métodos de inteligência artificial^[9] e vem sendo muito utilizado na seleção de comprimentos de ondas em aplicações espectroscópicas quantitativas^[10-11]. O algoritmo genético funciona simulando matematicamente a teoria de Darwin^[9-21], sendo aplicado por analogia ao que ocorre no mundo biológico: as condições experimentais (cromossomos), que levam as melhores respostas (indivíduos mais adaptados ao ambiente) tem maior chance de serem selecionadas (sobreviver), sendo transmitidas às novas gerações através da reprodução. Assim, a otimização das respostas (evolução da espécie) é alcançada por meio de recombinação das variáveis (cruzamento de cromossomos) e de algumas modificações aleatórias (mutações de genes).

Esse algoritmo foi idealizado para os casos onde os métodos tradicionais de otimização (simplex, planejamento experimental e operação evolucionária) não podem ser empregados. Isso ocorre quando as superfícies de resposta que descrevem sistemas reais, por serem complexas, apresentam vários máximos e mínimos locais e descontinuidades.

Em relação à eficiência, o GA seria uma classe intermediária entre os métodos altamente eficientes mas específicos em relação à determinados problemas e os métodos pouco eficientes para a totalidade dos problemas existentes. Neste raciocínio, o GA poderia ser classificado como um método razoavelmente eficiente para uma ampla variedade de problemas.

Entre as principais vantagens do GA, pode-se destacar:

- Busca direcionada pelo espaço amostral (economia de tempo).
- Realizam buscas simultâneas em várias regiões do espaço amostral, pois trabalham com uma população e não com um único ponto.
- Aplicável a uma ampla gama de otimizações (versatilidade).
- Não requer informações sobre a superfície de resposta.
- Robusto em relação à máximos e mínimos locais na superfície de resposta.
- Pode apresentar mais de um resultado satisfatório.
- Fornecem uma lista de parâmetros ótimos e não uma simples solução.
- Adaptam-se bem em computadores em rede, com processamento distribuído.

Um AG trabalha com populações de cromossomos que representa uma possível solução do problema a ser otimizado. Em geral, um cromossomo representa um conjunto de parâmetros da função objetivo cuja resposta será minimizada ou maximizada.

O algoritmo genético clássico pode ser dividido nas seguintes etapas sequenciais (figura 9):

Figura 9: Fluxograma das etapas de um algoritmo genético

Inicia-se o programa com a geração da população inicial e o mesmo prossegue em *loop* pelas etapas consecutivas, até que o critério de parada seja atingido. Existem diversas formas de realizar cada etapa, além de inúmeras configurações. A seguir descreve-se um breve resumo de cada etapa^[15, 18-21].

Tomando-se como exemplo o algoritmo aplicado à seleção de comprimentos de onda, a população inicial é constituída de n cromossomos, em que cada gene representa um comprimento de onda a ser incluído na avaliação. Normalmente utiliza-se codificação binária para representar o cromossomo, sendo a codificação real menos comum (na codificação real, o valor selecionado é armazenado no gene sem nenhum tipo de transformação). Embora a codificação binária seja mais simples de utilizar e manipular, a codificação real gera cromossomos menores e é mais naturalmente compreendida.

A avaliação da resposta é uma das etapas mais importantes do algoritmo. A busca orientada de soluções depende fortemente da função utilizada na etapa de avaliação e muito cuidado é

requerido na determinação desta função. Determinadas funções de avaliação são mais sensíveis à correlações randômicas, podendo ocasionar resultados insatisfatórios quando as soluções encontradas são aplicadas em amostras de validação (*overfitting*). Desta forma, a qualidade do algoritmo de avaliação vai influir diretamente na robustez da calibração. Normalmente esta etapa também é a parte mais lenta do algoritmo. Neste ponto, vale a pena fazer uma distinção entre função de avaliação e função de aptidão^[22]. A função de avaliação é utilizada para se avaliar a performance do cromossomo em relação ao problema específico que se deseja resolver. Um exemplo deste tipo de função é a calibração linear múltipla na seleção de comprimentos de onda em aplicações quantitativas. A função de aptidão, transforma o resultado da função de avaliação em critérios para seleção de cromossomos na continuidade do algoritmo (a seleção roleta, explicada mais adiante, é um exemplo de função de aptidão).

Na etapa de seleção, os melhores cromossomos são selecionados e irão compor uma população intermediária, onde serão aplicados os operadores de cruzamento e mutação. Esta população intermediária irá em uma etapa posterior substituir parcial ou totalmente a população atual, dando seguimento ao algoritmo.

Os métodos usuais de seleção são seleção roleta e seleção por torneio. Na seleção roleta, a probabilidade de um cromossomo ser escolhido para compor a próxima população aumenta de acordo com o resultado da avaliação do mesmo (a função de aptidão, utilizada na etapa de seleção, transforma o resultado da avaliação em probabilidade do cromossomo ser escolhido para compor as gerações futuras). A figura 10 exemplifica a seleção roleta. Neste exemplo, a correlação linear múltipla foi utilizada como avaliação da resposta utilizando-se 4 comprimentos de onda (genes). A figura indica que os cromossomos com maior valor de resposta possuem uma probabilidade proporcionalmente maior de serem escolhidos.

Figura 10: Exemplo de seleção roleta. Imagine-se que, a cada seleção, a seta pare num ponto aleatório da barra de probabilidades representada; depende-se que os cromossomos com respostas mais altas tem maior probabilidade de serem escolhidos frente aos demais.

Na seleção por torneio, são escolhidos aleatoriamente n cromossomos da população, e o cromossomo com maior aptidão é selecionado para a população intermediária. O processo repete-se até se preencher totalmente a população intermediária. Usualmente, utiliza-se $n = 3$ (31).

Na etapa de cruzamento, existe a troca de genes entre os cromossomos. Na prática, isso equivale a uma busca no espaço amostral representado pela população inicial, de forma a testar as diferentes combinações possíveis de comprimentos de onda. Algumas formas conhecidas de se realizar esta operação são: cruzamento simples, múltiplo e uniforme. O cruzamento simples envolve a quebra de dois cromossomos em uma posição aleatória e a troca de genes, conforme ilustra a figura 11.

Figura 11: Representação do cruzamento simples

O cruzamento múltiplo ocorre de forma similar, porém com dois ou mais pontos de quebra no cromossomo. No cruzamento uniforme, a troca de genes entre dois cromossomos ocorre de acordo com um padrão determinado aleatoriamente. Considerando dois cromossomos envolvidos no cruzamento, o padrão indicará, para cada posição no cromossomo resultante (“filho”), se o gene será recebido do cromossomo “pai” 1 ou do cromossomo “pai” 2. Em outras palavras, no cruzamento uniforme, para cada par de cromossomos é gerada aleatoriamente uma “máscara” que determina quais genes deverão ser trocados entre os mesmos.

No algoritmo tradicional, o cruzamento irá ocorrer de acordo com uma certa probabilidade (normalmente configurada entre 60 – 90%), desta forma, é possível que alguns cromossomos selecionados permaneçam sem alteração, preservando a resposta.

A etapa de mutação introduz variações aleatórias nos cromossomos (ou seja, um gene aleatório é “mutado”, modificando aleatoriamente o comprimento de onda representado por este gene). A mutação é uma forma eficiente de expandir o espaço amostral para valores não representados na população inicial, por outro lado destrói parcialmente a informação que o

cromossomo contém. A mutação pode ser realizada uma ou mais vezes em um mesmo cromossomo. Tipicamente, utiliza-se uma taxa de mutação em torno de 1%^[14].

Após aplicados os operadores de cruzamento e mutação na população intermediária, os cromossomos deverão ser reinseridos na população original (resultando na nova geração de cromossomos). Existem dois tipos básicos de reinserção: geracional e gradual (“*steady state*”)^[21].

Na reinserção geracional, a população intermediária substitui totalmente a população original. Já na reinserção gradual, os cromossomos da população intermediária são avaliados juntamente com os cromossomos da população original, e somente os melhores cromossomos entre as duas populações são selecionados para compor a próxima população. Neste tipo de reinserção, os cromossomos “pais” estarão presentes na população juntamente com os cromossomos “filhos”.

O uso da técnica de elitismo permite um meio termo entre a reinserção geracional e gradual. Esta técnica é utilizada a partir da reinserção geracional, porém os k melhores cromossomos da população original nunca são substituídos. Desta forma, a maior parte da nova população será formada pelos cromossomos da população intermediária (“filhos”), porém as melhores respostas da população original (“pais”) não serão perdidas. Usualmente $k=1$ e quanto maior o valor de k maior o risco de convergência prematura (máximos/mínimos locais).

As condições de parada podem ser baseadas na etapa de avaliação (ou seja, quando algum cromossomo atinge o valor ideal o algoritmo é interrompido), através de um número fixo de gerações, ou pela convergência (o algoritmo é interrompido quando não ocorrer melhora significativa dos melhores cromossomos para um dado número de gerações).

Uma implementação recente em algoritmos genéticos é o uso de paralelismo recorrendo a vários processadores. Desta forma, cada etapa é feita por um processador diferente, visando

otimizar o tempo gasto na resolução de problemas (particularmente útil na resolução de problemas complexos). Este tipo de implementação paralela é chamado de modelo de paralelismo global.

Alternativamente, é possível fazer com que cada processador desenvolva todas as etapas para populações diferentes de cromossomos, abrindo espaço para técnicas de migração entre as populações, acrescentando maximização do espaço amostral, melhorando a resposta do algoritmo e diminuindo o tempo de processamento^[18-20]. Este tipo de implementação é chamado modelo das ilhas ou de granulometria grossa.

A possibilidade de combinação de técnicas na programação do algoritmo genético é muito grande e a literatura revela vários estudos tanto de aplicabilidade prática do algoritmo como visando melhorar o rendimento do mesmo utilizando diferentes configurações, além de tentar solucionar alguns problemas comuns apresentados pelo mesmo^[20, 23-26]. Como exemplo de problemas a serem superados podemos citar:

- A configuração dos parâmetros depende muito da aplicação selecionada e não há regra aplicável a todos os casos (talvez sejam necessários vários testes até se chegar a uma condição ótima para a aplicação de interesse).
- Ainda é difícil encontrar um programa grátis flexível o suficiente para uso do algoritmo.
- O algoritmo consegue otimizar bem certos métodos, porém não há certeza que a resposta seja a melhor (independentemente de quão boa a mesma seja). Muitas vezes, respostas diferentes são encontradas cada vez que se inicializa o algoritmo. Isto é previsto pela própria dinâmica do algoritmo clássico, que não percorre completamente todo o espaço amostral e é fortemente dependente da população aleatória inicial.

A.1.2. Programas de otimização baseados no algoritmo genético

Existem hoje no mercado diversos programas de otimização baseados em algoritmos genéticos. Uma parte destes algoritmos são facilmente aplicáveis a problemas gerais de otimização numérica, sendo que nesta categoria se encontram a maior parte dos algoritmos desenvolvidos em Excel, como por exemplo o Generator^[27], Premium Solver^[28], OptWorks Excel^[29] e Evolver^[30]. Os programas deste tipo são bastante flexíveis já que trabalham com valores das células no Excel, permitindo que as mesmas contenham fórmulas que podem ser personalizadas pelo usuário. Por outro lado, problemas complexos e com muitas variáveis requerem um trabalho adicional de personalização de planilhas. Serve de exemplo o programa OptWorks, do fabricante Pi Blue, do qual é possível obter cópia via internet (*download*) da versão de demonstração, que funciona gratuitamente por 30 dias, com algumas limitações de parâmetros. O preço da licença sem limitação, para um usuário, é US\$29,99. O mesmo programa, gravado em CD e embalado custa US\$130,00.

Alguns algoritmos disponíveis requerem um trabalho maior de configuração para sua utilização, como é o caso do pacote “*Genetic Algorithm and Genetic search toolbox*” do programa Matlab^[31] ou dos diversos controles *ActiveX* existentes (OptGA^[32], GPC^[33]). Nestes casos, é necessário conhecimento da linguagem de programação envolvida para utilizar o algoritmo. A vantagem deste tipo de programa é a possibilidade de elaborar aplicações específicas utilizando algoritmo genético, porém sem o trabalho de desenvolver todas as etapas do mesmo. Para o programa Matlab há versão de demonstração disponível na internet (funcional por quinze dias). Ao preço da licença do algoritmo genético, US\$ 59,00, há que somar a licença do programa Matlab, de US\$ 99,00 para a versão educacional.

Adicionalmente, temos os programas já prontos para resolução de problemas específicos, como por exemplo o TSP Solver^[34] (problema do caixeiro viajante^[35-36]), MDL QSAR^[37] (envolvendo relações entre estrutura química e atividade^[38]) e GeneProTools^[39] (envolvendo modelagem de funções). Nestes casos, embora a aplicabilidade seja normalmente limitada ao problema de interesse, não é necessário nenhum trabalho extra de desenvolvimento. Um bom exemplo é o software GeneProTools 4.0, produzido pela Geptsoft, que tem interface bastante amigável e inúmeros recursos, e cujo preço varia desde £245,00 (versão *standard*) até £33.000,00 (versão *enterprise*). A versão de demonstração pode ser copiada via internet, porém não permite visualizar o código para os modelos gerados.

O algoritmo genético desenvolvido neste trabalho pertence a esta última categoria, com aplicabilidade específica na seleção de comprimentos de onda. A implementação do algoritmo foi feita integralmente em *Visual Basic for Applications* acoplado à planilha Excel da Microsoft. Desta forma, assegura-se a portabilidade para a maior parte dos computadores pessoais existentes na atualidade. Em nenhum dos trabalhos citados até aqui ou em buscas na literatura e na internet pôde ser encontrado um programa Excel+VBA pronto para esta finalidade.

A.2. PARTE EXPERIMENTAL

Inicialmente, foi realizado um levantamento sobre o funcionamento e muitas aplicações dos chamados algoritmos genéticos e, a partir dos dados iniciais, foi desenvolvido um algoritmo, em VBA-Excel, para seleção de comprimentos em aplicações quantitativas de FTIR. O estudo foi desenvolvido procurando-se otimizar e melhorar os resultados obtidos com o algoritmo genético, através da introdução de algumas modificações no mesmo.

Considerando a filosofia por trás do algoritmo, o programa foi estruturado a partir de um módulo de classe que representa o cromossomo. Este módulo de classe (cromossomo) possui rotinas internas para gerenciar coleções de objetos (genes). Desta forma o cromossomo possui rotinas embutidas para definir e alterar o valor do gene, adicionar ou excluir genes, além de variáveis para guardar informações sobre o gene e o cromossomo (como por exemplo, a geração a que o cromossomo pertence ou se o mesmo sofreu mutação).

Seguindo a lógica do algoritmo, foi definida uma matriz bidimensional, com cada coluna representado uma população diferente e os cromossomos organizados em linhas. A matriz foi chamada de “Bios”, onde cada elemento da matriz representa um cromossomo. Desta forma, o comando:

Bios(c,p).addgene 101, 0

Significa: Selecione o cromossomo “c”, pertencente à população “p” e adicione um gene com valor “101” na posição 0 (posição inicial).

A utilização de populações distintas gerenciadas pelo mesmo algoritmo possibilitará o desenvolvimento de aplicações envolvendo múltiplos analitos, como se verá nas seções posteriores.

O programa trabalha com vários tipos de “*loops*” baseados na função “*For next...*”, sendo que três “*loops*” são os mais importantes e recorrentes. O “*loop*” principal é referente ao número de gerações, o segundo trabalha com todas as populações (percorrendo cada cromossomo) e terceiro trabalha com todos os cromossomos (percorrendo cada gene).

As etapas do algoritmo foram estruturadas como mostra a tabela 1, e serão discutidas nas próximas seções. As etapas 1-4 ocorrem somente no início do algoritmo, que prossegue em *loop* pelas etapas 5-10 até que o critério de parada seja atingido.

Tabela 1: Estrutura do algoritmo genético desenvolvido

Etapa	Descrição
1	Definição dos parâmetros (atribuição das variáveis)
2	Geração da População inicial
3	Avaliação
4	Classificação
	<i>“Loop” (Critério de parada)</i>
5	Cruzamento
6	Mutação
7	Avaliação
8	Classificação (Inserção de marcas)
9	Reinserção da nova geração (Elitismo)
10	Migração
11	Saída de Resultados
	<i>“Loop”</i>

A.3. RESULTADOS E DISCUSSÃO

Dentro do escopo deste trabalho, o algoritmo genético desenvolvido foi testado e utilizado na seleção de comprimentos de onda em FTIR, para o desenvolvimento da metodologia para três aplicações distintas: Quantificação de lauril éter sulfato de amônio 28%, quantificação de o-

benzil p-Cloro Fenol, o-fenil fenol e etanol na mistura dos três componentes e quantificação de lauril éter sulfato de sódio em *shampoos*. Os resultados específicos obtidos para cada um dos métodos desenvolvidos serão discutidos nas partes B, C e D deste trabalho. Nesta sessão, serão discutidos alguns aspectos relevantes de cada etapa do algoritmo proposto, bem como as modificações efetuadas visando melhorar a performance do mesmo.

A.3.1. Definição dos parâmetros e elitismo

Antes de se iniciar o algoritmo propriamente dito, é feita a atribuição das variáveis com base nos parâmetros informados pelo usuário. Os parâmetros que devem ser informados são listados abaixo.

Tamanho da população

Número de populações

Número de gerações

Tamanho do cromossomo

Frequência de saída dos resultados

Probabilidade de mutação

% Elitistas

Intervalo de migração

Número de imigrantes

Classificação Mínima (utilizado na migração)

Número de amostras

Intervalos de valores permitidos para cada gene

No algoritmo desenvolvido, optou-se por utilizar populações e cromossomos de tamanho fixo e fazer uso da técnica de elitismo.

Em um algoritmo clássico, as novas gerações são inseridas ou substituem as gerações anteriores. Com o uso da técnica de elitismo, isto também ocorre porém os melhores cromossomos são obrigatoriamente preservados e permanecem inalterados. No algoritmo desenvolvido, os novos cromossomos substituem os cromossomos antigos, exceto os que são denominados “cromossomos elitistas” (os que foram preservados por terem melhor resposta na geração anterior).

Na etapa de definição dos parâmetros também é definida a matriz bidimensional $A \times N$, onde A representa a absorbância para cada número de onda informado e N o número de amostras. Desta forma, se a base de dados contiver espectros obtidos no intervalo $700 - 4000 \text{ cm}^{-1}$, com resolução de 1 cm^{-1} , para 35 amostras, teremos definida uma matriz 3300×35 .

Outra matriz também definida nesta etapa é a matriz $C \times N$, onde C representa a concentração para cada analito que se deseja quantificar e N é o número de amostras.

A.3.2. Geração da população inicial e codificação

A geração da população inicial é feita sequencialmente atribuindo-se valores para cada gene de cada cromossomo dentro de cada população, utilizando a função geradora de números aleatórios do Excel. A função gera um número aleatório (inteiro) entre os valores permitidos para cada gene. Estes valores inteiros estão relacionados com a codificação utilizada para representar as variáveis.

Um dos primeiros passos no desenvolvimento foi decidir a forma de codificação e optou-se pela codificação real ao invés da binária. Desta forma, ao tratarmos com números de onda inteiros, cada gene continha o valor do próprio número de onda, tornando a saída do programa mais “amigável” ao usuário. O programa inicial foi escrito desta forma, porém quando em determinado momento do trabalho passou-se a trabalhar com uma resolução espectral maior, alcançando Algarismos decimais dos números de onda, foi necessário adaptar a abordagem para não ter que reescrever o código do programa (em virtude da forma como o algoritmo lida com inteiros, decimais e números longos).

Assim, os números de onda passaram a ser representados e codificados por números inteiros sequenciais. A representação é feita de forma totalmente livre e flexível e não foram necessárias grandes alterações de código. A tabela 2 mostra três exemplos de codificações diferentes com os sequenciais de 1-5:

Tabela 2: Exemplos de codificação com os sequenciais 1-5

Codificação 1	Codificação 2	Codificação 3
1 = 11	1 = 5,32	1 = 1001,0
2 = 12	2 = 5,31	2 = 1001,5
3 = 13	3 = 5,29	3 = 1002,0
4 = 14	4 = 5,28	4 = 1002,5
5 = 15	5 = 5,27	5 = 1003,0

Com a representação utilizada e expandindo as aplicações, seria até mesmo possível trabalhar com letras e símbolos.

A codificação contribui para que todas as etapas do algoritmo sejam independentes da função de avaliação, ou seja, a operação de cada etapa pode ser realizada independentemente do tipo de função de avaliação ou do tipo de resposta obtida. Desta forma, todas as operações são

realizadas simplesmente com os números contidos em cada gene, porém é na função de avaliação que se vincula o número da codificação com o comprimento de onda e a respectiva absorbância, para cada amostra.

A.3.3. Avaliação e classificação

A função de avaliação utilizada baseou-se no valor da regressão linear múltipla^[40], obtido entre as absorbâncias das amostras nos comprimentos de onda codificados pelos genes e a concentração do analito que se pretendia quantificar. Para cada cromossomo, a função busca na matriz $A \times N$ as absorbâncias referentes aos comprimentos de onda codificados e as relaciona com a concentração de cada analito através da função LINEST. A função LINEST pode retornar diversos parâmetros, porém o valor utilizado como critério de avaliação foi o valor do coeficiente de regressão (R^2).

Em cada cromossomo é armazenado o valor do seu coeficiente de regressão; a seguir a função de classificação compara os coeficientes de toda a população e classifica os cromossomos na população em ordem decrescente de valor. Desta forma o cromossomo de maior valor de correlação passa a ser o primeiro cromossomo da população, enquanto o cromossomo de menor valor passa a ser o último.

Além do coeficiente de variação, o cromossomo também carrega consigo uma marca referente à sua posição na população (ou *rank*). No presente algoritmo, a posição do algoritmo na população é numericamente equivalente ao *rank* (ou classificação), porém o sistema de marcas utilizado permite que se modifique a posição do cromossomo sem perda da informação de

classificação (isto é importante quando se deseja variar os pares de cromossomos na etapa de cruzamento).

A.3.4. Cruzamento e mutação

A função de cruzamento gera um número aleatório que define o ponto de quebra para um par de cromossomos. Os genes a partir do ponto de quebra são trocados entre os dois cromossomos. O par de cromossomos é definido pela ordem em que os mesmos aparecem na população. Lembrando que os mesmos são classificados na etapa de avaliação, temos como consequência que as melhores respostas são combinadas entre si.

Os cromossomos resultantes do cruzamento são reinseridos sequencialmente na população, logo após os cromossomos marcados como “elitistas”. Como o número de cromossomos da população é fixo, os cromossomos excedentes são descartados. Note-se que o número de cromossomos excedentes é igual o número de cromossomos elitistas. Além disso, estes excedentes correspondem às combinações dos piores cromossomos da geração anterior.

O número de cromossomos elitistas utilizado no algoritmo deve ser cuidadosamente configurado de forma a facilitar a combinação dos melhores comprimentos de onda, porém sem inibir ou diminuir significativamente a busca por novas combinações. Se o número de elitistas for muito alto, a maior parte dos melhores cromossomos conterá comprimentos de onda idênticos, diminuindo a probabilidade de combinação com comprimentos de onda diferentes. Neste cenário, a maioria das mudanças significativas ficariam a cargo da mutação e a população progrediria mais lentamente em relação à resposta.

A mutação consiste em atribuir um valor aleatório para determinado gene (sempre dentro do intervalo de valores permitido para o mesmo) e ocorre no algoritmo de acordo com uma certa probabilidade definida. Desta forma, para cada cromossomo gera-se um número aleatório entre 1 e 100, e se o número gerado for menor que a probabilidade de mutação definida pelo usuário, um novo valor será definido para um gene do cromossomo. Tanto a posição do gene a ser mutado como o valor de mutação são definidos aleatoriamente por um sistema similar. Os cromossomos que sofreram mutação podem ser marcados para rastreamento posterior.

A.3.5. Saída de resultados e critério de parada

O critério de parada utilizado no algoritmo foi o número de gerações. Assim, se define inicialmente em que geração a execução do código irá se encerrar. A princípio, pode-se definir qualquer número como critério de parada e não há um número limite máximo, já que o número de cromossomos e populações é definido e fixo, e não há mudança significativa na memória exigida pela execução do código.

A única ressalva é quanto à saída dos resultados do programa, pois a configuração da frequência de saída também é feita de acordo com o número de gerações e os resultados são enviados para uma planilha em Excel. Se por exemplo a frequência de saída for igual a 10, então a cada dez gerações o programa enviará uma tabela para o Excel (incluindo a geração inicial). Esta tabela contém uma lista de todos os cromossomos, respostas e marcas da geração em questão. A figura 12 mostra um exemplo dos resultados de saída do algoritmo.

Pop	Cromo	Generation	Mark	Resp1	Resp2	Resp3	Resp4	Gene1	Gene2	Gene3	Gene4
0	0	0	0	0,9594	0,9594	0,9684	0,9684	745	1251	3085	3382
0	1	0	0	0,9565	0,9565	0,9626	0,9626	1245	1697	2559	3459
0	2	0	0	0,9682	0,9682	0,9714	0,9714	1221	1768	3021	3574
0	3	0	0	0,9535	0,9535	0,9602	0,9602	1226	1462	2970	3927
0	4	0	0	0,9786	0,9786	0,9822	0,9822	1000	1735	2957	3766
0	5	0	0	0,9698	0,9698	0,9759	0,9759	802	1777	3181	3843
0	6	0	0	0,9870	0,9870	0,9854	0,9854	778	1298	2893	3959
0	7	0	0	0,9679	0,9679	0,9677	0,9677	929	1542	2842	3760
0	8	0	0	0,9640	0,9640	0,9667	0,9667	1108	1265	2829	3600
0	9	0	0	0,9810	0,9810	0,9854	0,9854	1056	1711	3053	3478

Figura 12: Exemplo dos primeiros dados de saída (resultados) do algoritmo.

Pelo critério acima, seria possível termos registros de todas as gerações de um algoritmo configurado com 30 cromossomos, 3 populações e 100 gerações como critério de parada, pois se o intervalo de saída for 1 geração, teremos uma tabela com $30 \times 3 \times 100 = 9000$ linhas (sem problemas mesmo para computadores medianos). Por outro lado, se o número de gerações fosse 1000, não seria possível ter $30 \times 3 \times 1000 = 90000$ linhas de saída, vez que as atuais versões do Excel aceitam, no máximo, 65.536 linhas. Uma alternativa seria diminuir a frequência de saída para 10 gerações, já que teríamos uma tabela com $90000/10 = 9000$ linhas.

A menos que se queira estudar o comportamento do algoritmo ou as soluções intermediárias encontradas pelo mesmo, bastaria dispor da saída da última geração para encontrar os melhores resultados. Assim poder-se-ia configurar o algoritmo para 10000 gerações e intervalo de saída 5000, e o único fator limitante seria o tempo despendido (neste caso, registro da geração inicial, da geração 5000 e da geração 10000).

Uma estimativa do tempo gasto pelo algoritmo é mostrado na tabela 3. Para este teste, foi utilizado um computador Pentium IV 3.1 Gz e 512 M RAM e o algoritmo foi configurado com

64 espectros referentes à mistura de o-benzil p-cloro fenol, o-fenil fenol e etanol (o mesmo conjunto de dados utilizado para o trabalho com bactericidas, ver parte C), 30 cromossomos, 4 genes no cromossomo e, como critério de avaliação, os valores de R^2 nos quatro comprimentos de onda, obtidos por regressão linear múltipla. O intervalo de saída utilizado foi igual ao número de gerações dividido por 2. O intervalo de migração utilizado foi igual ao critério de saída.

Tabela 3: Tempo gasto pelo algoritmo de acordo com o número de gerações

Número de gerações	1 população		3 populações
	Duração (hora/min/seg)	9 x Duração (hora/min/seg)	Duração (hora/min/seg)
100	0:00:07	0:01:03	0:00:19
1000	0:00:34	0:05:06	0:02:46
10000	0:05:06	0:45:54	0:27:13
100000	0:50:21	7:33:09	4:31:40

Considerando 3 populações envolvidas, onde cada população se desenvolve em relação a um analito e é avaliada para os outros dois, temos um total de 3 populações para cada um dos três analitos, equivalendo a nove execuções simples do algoritmo (ou seja, considerando que para o mesmo número de avaliações, seria necessário executar o algoritmo três vezes para cada um dos três analitos). A tabela 3 mostra a duração experimental para o algoritmo executado com somente uma população e com 3 populações. Para efeito de comparação, também é mostrada a duração da execução para uma população, multiplicada por nove, demonstrando a economia de tempo obtida na quantificação de múltiplos analitos utilizando implementação paralela. Para as aplicações deste trabalho, o número de gerações utilizado para se chegar a resultados satisfatórios não passou de 100.

O tempo requerido pelo algoritmo em repetições com parâmetros iguais e realizadas no mesmo computador mostrou-se constante, em segundos, quer a repetição fosse imediata ou em dia subsequente.

Como esperado, a etapa de saída é uma das mais demoradas da algoritmo, já que nesta etapa os cromossomos e as respectivas respostas são armazenados em um arquivo em Excel. A tabela 4 compara o tempo gasto utilizando intervalo de saída igual à cinquenta e igual à um, para 100 gerações (todos os outros parâmetros foram mantidos iguais ao do teste da tabela 3, descrito previamente).

Tabela 4: Tempo gasto pelo algoritmo de acordo com o número de gerações

Número de Gerações	3 Populações	
	Intervalo de saída (em gerações)	Duração (hora/min/seg)
100	50	0:00:19
100	1	0:02:21

A.3.6. Migração

Embora a implementação tenha sido feita em um único computador (processador), foram utilizados os conceitos de paralelismo, de forma a permitir a geração de populações independentes. As vantagens desta implementação são:

- Permitir a avaliação de respostas diferentes, ou seja, é possível otimizar métodos para diversos analitos simultaneamente (cada população se desenvolve em relação a um determinado analito).

- Todas as populações são testadas para todas as respostas, isso acarreta redução de tempo de processamento, uma vez que a população que já foi gerada para determinada resposta (analito) é aproveitada na avaliação de respostas para outros analitos. Apesar disso, o direcionamento em relação à resposta não é negativamente afetado, pois cada população só se desenvolve em relação à sua própria resposta.

- Os indivíduos de melhor resposta migram entre as populações. Além de otimizar o algoritmo, isto insere um novo elemento para maximização do espaço amostral dentro da população.

As figuras 13 e 14 ilustram o conceito.

Figura 13: Representação esquemática da avaliação da resposta por diferentes populações.

Figura 14: Representação esquemática da migração

O algoritmo permite determinar o intervalo de migração e o número de cromossomos envolvidos. Quando a migração é ativada no algoritmo, todos os cromossomos são avaliados para todas as respostas, e para que a migração ocorra, o cromossomo deve possuir uma resposta mínima em relação a população destino. Esta resposta mínima é configurada no algoritmo sob a denominação “classificação mínima”, na forma de um número que varia de 1 até o número máximo de cromossomos. O candidato à migração deve possuir uma resposta melhor que o cromossomo com número equivalente à classificação mínima da população-destino (lembrando que como existe a etapa de classificação dos cromossomos, os mesmos são praticamente numerados em ordem crescente em relação ao seu desempenho).

Para exemplificar e tornar mais claro, consideremos um caso com critério de classificação mínima igual a 5, frequência de migração igual a 50, número de imigrantes igual a 3, número máximo de gerações igual a 200 e três populações envolvidas (denominadas A, B e C). Neste caso, poderá haver migração nas gerações 50, 100, 150 e 200. Os cromossomos da população A também serão avaliados em relação às respostas das populações B e C. Para migrar para a população B ou C, o cromossomo da população A tem que possuir resposta melhor que o quinto cromossomo da população-destino. O processo ocorre até que três cromossomos tenham migrado ou que não reste nenhum cromossomo com resposta melhor que a do quinto cromossomo da população destino (se nenhum dos cromossomos da população A satisfizer o critério, então não ocorrerá migração). Esta avaliação ocorre para a população A em relação à B e C, para a população B em relação à A e C e para a população C em relação à A e B. Desta forma, todas as populações são potenciais doadoras ou receptoras de cromossomos. Quando a migração ocorre, os cromossomos migrantes simplesmente substituem os cromossomos de pior classificação da população destino. No caso do exemplo, se 3 cromossomos migrarem de A para B, então os três

últimos cromossomos da população B serão substituídos (lembrando que posição do cromossomo na população é igual a sua classificação em relação à resposta).

PARTE B: QUANTIFICAÇÃO DE LAURIL ÉTER SULFATO DE AMÔNIO 28%

B.1. INTRODUÇÃO

B.1.1. Alquil Sulfatos e Alquil Éter Sulfatos

Alquil Sulfatos (AS), Alquil Benzeno Sulfonatos e Alquil Éter Sulfatos (AES) são os três tipos mais comuns de surfactantes utilizados em formulações de *personal care e home care*^[41]. Normalmente, os mesmos são utilizados como surfactantes primários na formulação, sendo os principais responsáveis pela remoção de sujeira e gordura do cabelo.

Os alquil sulfatos são conhecidos por suas propriedades como bons espumantes (espuma densa e em grande quantidade) e eficientes na limpeza (boa detergência). Todavia, sua solubilidade e performance são afetadas pela dureza da água e força iônica do meio, além de serem levemente irritantes quando em contato direto com a pele.

Os alquil sulfatos podem ser representados pela fórmula:

Onde $n = 8$ a 14 e $\text{M} = \text{Na}^+, \text{NH}_4^+, \text{etc.}$

O processo de síntese do tensoativo consiste primeiro na obtenção do álcool graxo, seguida de sulfatação. A parte hidrófoba do tensoativo é formada pela cadeia carbônica. Os processos preferenciais para obtenção da cadeia carbônica são baseados na rota petroquímica (a partir do etileno) ou na rota oleoquímica (a partir de óleos vegetais, como óleo de coco ou de palma). No

caso da rota oleoquímica, os óleos vegetais são inicialmente convertidos à ésteres ou ácidos e então hidrogenados para se produzir o álcool graxo. A distribuição da cadeia carbônica influencia as propriedades do tensoativo, já que o aumento da cadeia acarreta aumento da hidrofobicidade e do poder surfactante, porém, diminui a solubilidade. Os tensoativos com predominância de cadeia C12-C14 possuem boa solubilidade.

A sulfatação é feita normalmente com ácido sulfúrico saturado com trióxido de enxofre. Em seguida ocorre neutralização com base (normalmente hidróxido de sódio ou amônio, porém outras bases podem ser utilizadas, dependendo da característica que se deseja para o surfactante). Nas etapas de sulfatação e neutralização, a quantidade resultante de material não sulfatado e a quantidade de sal formada são importantes pois alteram a viscosidade do surfactante.

Em relação aos alquil éter sulfatos, o processo de obtenção é bastante similar ao dos alquil sulfatos, porém com uma etapa de etoxilação (efetuada com óxido de etileno na presença de catalisadores) antes da sulfatação. Os alquil éter sulfatos podem ser representados pela fórmula abaixo:

Em relação à outros surfactantes aniônicos, os alquil éter sulfatos são únicos porque tanto a parte hidrófoba quanto a parte hidrófila podem ser modificadas, de forma que é possível obter uma grande variedade de propriedades físicas e químicas. O tamanho médio da cadeia de óxido de etileno (EO) pode variar bastante (0 a mais de 10), sendo que o valor típico está entre 1-3. Nos valores típicos de etoxilação o nível de álcool não etoxilado ainda é significativo (15 - 42%) e impacta diretamente nas propriedades do tensoativo. O aumento no grau de etoxilação aumenta a solubilidade do tensoativo.

Em geral, os alquil éter sulfatos de cadeia C12-C14 são bastante solúveis em água (embora a diluição possa ser difícil dependendo da concentração e do tipo de tensoativo, já que a partir de 22-29% o tensoativo adquire a consistência de gel). Tensoativos do tipo AES são bons estabilizadores de espuma (embora não forneçam espuma tão densa como os alquil sulfatos) e possuem a propriedade de adquirirem viscosidade facilmente com a adição de sal (bastante útil em determinadas formulações). Além disso, os AES são menos irritantes que os alquil sulfatos e mais compatíveis com outros tipos de surfactantes.

B.1.2. Determinação de surfactantes – titulação volumétrica e potenciométrica

As titulações volumétricas visuais e potenciométricas de surfactantes estão entre os métodos mais usuais para determinação de surfactantes. A titulação de surfactantes^[42] baseia-se no princípio que surfactantes com íons de cargas opostas irão se combinar estequiometricamente formando um sal hidrofóbico insolúvel em água (porém solúvel em determinados solventes orgânicos, tipicamente clorofórmio). A insolubilidade do sal formado é de vital importância para a precisão do método. Os titulantes utilizados são normalmente bastante diluídos (1 a 4 mmol L⁻¹) em virtude da limitada solubilidade dos mesmos.

A titulação volumétrica com indicador misto foi inicialmente sugerida por Holness e Stone^[43] em 1957, sendo o método padrão atual para determinação de tensoativos aniônicos, incorporado na ISO 2271-1989 e BS 3762. O procedimento consiste em adicionar uma alíquota (20 mL) de uma solução do surfactante aniônico cujo teor se deseja determinar à uma proveta provida de tampa. Em seguida adiciona-se 10 mL de água, 10 mL de clorofórmio e 10 mL de solução do indicador misto. O titulante utilizado é uma solução 4 mmol L⁻¹ do tensoativo

catiônico Hyamine 1622. Após cada adição é necessário tampar a proveta e agitar vigorosamente para que ocorra mistura dos reagentes. O ponto final é determinado quando a fase clorofórmica, inicialmente rosa, se torna cinza-violeta (em presença de excesso de titulante, a mesma se torna azul).

O indicador misto é formado pelo tensoativo aniônico *disulfine blue* VN (B) e pelo catiônico *dimidium bromide* (D). Quando esta mistura é adicionada ao surfactante aniônico (A^-), um complexo rosa, preferencialmente solúvel na fase clorofórmica, é formado com o *dimidium*:

O *dimidium bromide* permanece na solução aquosa tornando-a verde-azulada. Conforme o titulante (Hy^+) é adicionado, o mesmo reage com o tensoativo aniônico, deslocando o íon *dimidium* de volta para a fase aquosa. No ponto de equivalência, todo íon *dimidium* terá retornado para a fase aquosa, enquanto na fase orgânica restará o sal formado pelos dois tensoativos:

Na presença de excesso de Hyamine, a mesma reagirá com *disulfine blue* (B), produzindo um complexo azul solúvel em clorofórmio:

A figura 15 mostra a estrutura dos indicadores e da Hyamine 1622. Este método produzirá melhores resultados se a solução de Hyamine também for fatorada seguindo o mesmo procedimento e se o tensoativo que se deseja determinar possuir cadeia carbônica igual ou maior que C10, para o caso de alquil sulfatos.

Figura 15: Estruturas do Azul de Dissulfino, Brometo de Dimínio e Hiamina 1622

A titulação volumétrica requer experiência do analista em avaliar o ponto final e ainda possui as desvantagens do uso e manipulação do clorofórmio. O método alternativo, empregando titulação potenciométrica, baseia-se no mesmo princípio de reação entre surfactantes de cargas opostas, porém não possui as desvantagens mencionadas acima. Ainda assim, a técnica possui alguns fatores que precisam ser melhorados:

- A solubilidade do sal formado pelos tensoativos aniônico e catiônico vai depender da cadeia e do tipo de surfactantes envolvidos (analito e titulante), bem como de suas respectivas concentrações, assim o ponto estequiométrico real pode não corresponder ao ponto indicado pela inflexão da curva de titulação.

- O tipo de eletrodo para surfactante empregado também interfere no método, já que a resposta do mesmo em relação aos diversos tipos de surfactantes pode variar. A seletividade do eletrodo em relação ao tamanho da cadeia carbônica dos surfactantes é um fator importante. A resposta do eletrodo também pode variar em função do condicionamento do mesmo (ou seja, em função do meio e das condições em que se encontrava anteriormente à titulação). Em relação aos diferentes tipos de eletrodo disponíveis no mercado, Buschmann and Schultz^[44] obtiveram desvios padrão relativos da ordem de 0,5 – 2,36%, em relação ao valor esperado.

- A formação de micelas diminui a atividade dos surfactantes e, conseqüentemente, a resposta do eletrodo, diminuindo o salto de potencial no ponto estequiométrico. Um salto de potencial típico está em torno de 67 mV (diferença entre as leituras referentes a adição de 90% e 110% do volume de equivalência da titulação).

- A presença de outros surfactantes ou substâncias influencia na formação de micelas.

Por todos estes motivos, a utilização da titulação potenciométrica para determinação de surfactantes deve ser utilizada somente se já houver um estudo ou uma validação do sistema como um todo (incluindo o tipo de eletrodo, titulador e titulante) para o analito que se deseja quantificar.

Em vista das dificuldades e deficiências dos métodos acima descritos, bem como das determinações cromatográficas^[1], que não serão detalhadas aqui, fica evidenciada a importância da busca de alternativas rápidas e confiáveis para a análise tanto das matérias primas como dos produtos formulados. Nos experimentos apresentados e discutidos a seguir, o algoritmo genético previamente desenvolvido foi aplicado com sucesso à seleção de comprimentos de onda para quantificação de teor ativo da mistura de lauril éter sulfato de amônio/lauril sulfato de amônio 28%.

B.2. PARTE EXPERIMENTAL

Para o estudo, foram utilizadas 12 amostras da mistura comercial de lauril éter sulfato de amônio/lauril sulfato de amônio 28% (1:3), provenientes de dois fornecedores diferentes.

O material usualmente oferecido no mercado cosmético contém aproximadamente 21% de lauril sulfato de amônio, 7% de lauril éter sulfato de amônio (totalizando 28% de ingrediente ativo), 1% de sulfato de amônio, 1% de material insulfatado livre e 0,5% de cloreto de amônio, sendo que o restante (69%) é composto por água.

O ingrediente destas amostras foi determinado através de titulação potenciométrica, utilizando titulador automático Orion, eletrodo para surfactante Orion e eletrodo de referência da mesma marca. Os principais parâmetros utilizados na titulação foram:

Pré Dose: 1 mL

Volume Máximo: 11 mL

Critério de estabilização: 3 mV/min

Incremento de potencial para cálculo do volume dispensado: 5 mv

Massa Molar: 315

Cálculo do Volume de equivalência: Primeira Derivada

A figura 16 mostra uma curva de titulação típica, que permite a determinação bastante precisa do ponto estequiométrico, sendo o DPR da ordem de 1,2%.

Uma das amostras, escolhida aleatoriamente, foi diluída em água em quatro concentrações diferentes (expressa como% massa), de forma a aumentar o intervalo da curva de calibração. A concentração destas amostras foi calculada com base no resultado de titulação potenciométrica da amostra original, considerando a diluição utilizada.

Figura 16: Curva de titulação potenciométrica típica para LESA 28%

Para obtenção dos espectros, foi utilizado o FTIR Spectrum BX (Perkin Elmer) com detector DTGS, equipado com acessório de reflectância atenuada (diamante - 9 reflexões). Os parâmetros utilizados foram:

Número de Scans: 16

Resolução: 2 cm^{-1}

Intervalo: $700 - 4000\text{ cm}^{-1}$

O software utilizado para aquisição dos dados foi o Spectrum v.2.0 (Perkin Elmer).

Para cada amostra foram adquiridos espectros em duplicata (totalizando 32 espectros). A tabela 5 exibe as concentrações utilizadas em cada espectro.

Tabela 5: Concentrações de LESA 28 % utilizadas na calibração

Espectro	Concentração	
	Titulação Potenciométrica (%)	Certificado do fornecedor (%)
20A	19,91*	19,90*
20B	19,91*	19,90*
22A	22,14*	22,13*
22B	22,14*	22,13*
24A	23,93*	23,91*
24B	23,93*	23,91*
26A	25,92*	25,90*
26B	25,92*	25,90*
Latd28A	28,07	28,05
Latd28B	28,07	28,05
067/04 A	28,11	28,08
067/04 B	28,11	28,08
194/04 A	27,67	27,80
194/04 B	27,67	27,80
071/05 A	28,33	28,20
071/05 B	28,33	28,20
124/06 A	28,53	28,40
124/06 B	28,53	28,40
208/06 A	28,07	27,90
208/06 B	28,07	27,90
028/07 A	28,81	28,50
028/07 B	28,81	28,50
054/07 A	27,92	27,95
054/07 B	27,92	27,95
078/07 A	28,18	27,95
078/07 B	28,18	27,95
080/07 A	27,58	27,50
080/07 B	27,58	27,50
059/08 A	27,60	27,60
059/08 B	27,60	27,60
091/08 A	28,46	28,20
091/08 B	28,46	28,20

* Concentrações calculadas com base nos resultados da amostra Latd28

Para efeito de comparação, além das titulações potenciométricas, também foram utilizados os resultados do certificado de análise dos fornecedores, que utilizam titulação volumétrica manual com indicador visual acima descrita.

B.3. RESULTADOS E DISCUSSÃO

Na figura 17 é mostrado o espectro típico do lauril éter sulfato de amônio 28%, com as atribuições vibracionais para as bandas principais^[45].

Figura 17: Espectro típico de Lauril Éter Sulfato de Amônio 28%

O íon amônio possui simetria tetraédrica e nove vibrações fundamentais: estiramento simétrico (3040 cm⁻¹, inativo no IR), deformação duplamente degenerada (1710 cm⁻¹, inativa no IR), estiramento antissimétrico triplamente degenerado (3140 cm⁻¹, muito forte) e deformação triplamente degenerada (1400 cm⁻¹, forte). Uma banda forte em 3040 cm⁻¹ é na verdade uma

banda de combinação ($1710 + 1400 \text{ cm}^{-1}$), intensificada pela ressonância de Fermi^[46] (no caso da figura 17, aparece junto com νOH e νNH). A hidratação do íon amônio também causa absorções adicionais na região de 3300 cm^{-1} .

O grupo OSO_3^- e o grupo oxi-etileno imediatamente vizinho interagem fortemente, de forma que o espectro dos alquil etóxi sulfatos não são simplesmente superposições dos espectros das estruturas parciais. Com o aumento do grau de etoxilação, a banda $\nu_{\text{as}}(\text{C-O-S})$, mostrada na figura 17 em 974 cm^{-1} , é deslocada para números de onda menores, enquanto a vibração $\nu_{\text{s}}(\text{OSO}_3^-)$ permanece inalterada ao redor de 1080 cm^{-1} .

A presença ou não de cátions como Na^+ ou K^+ no surfactante tem grande influência na frequência da bandas. No caso do íon amônio, é difícil demonstrar inequivocamente a presença do mesmo, pois as vibrações características deste grupo normalmente aparecem sobrepostas à outras vibrações (no caso da figura 17, νNH junto com νOH e δNH junto com δCH_2 e δCH_3). A figura 18 mostra a comparação entre os espectros de Lauril Éter Sulfato de Amônio 28% (LESA 28%) e Lauril Éter Sulfato de Sódio 27% (LESS 27%).

Para tensoativos etoxilados, a banda $\nu_{\text{as}}(\text{C-O-C})$ aparece ao redor de 1110 cm^{-1} . Esta banda não é vista no espectro de LESA 28%, devido à diluição, além de o tensoativo etoxilado aparecer em menor concentração na mistura dos tensoativos. A banda pode ser observada no espectro de LESS 27%, ainda que mal definida. A banda correspondendo ao modo vibracional degenerado $\nu_{\text{as}}(\text{SO}_3^-)$ aparece larga na figura 18, e em espectros de surfactantes sólidos do mesmo tipo, a mesma aparece duplicada, provavelmente em virtude das interações intermoleculares com o cátion ligado, que elimina a degeneração da vibração.

Figura 18: Espectros de LESA 28% e LESS 27%.

Para o teste inicial do algoritmo, foi utilizado um modelo simples de quantificação para Lauril Éter Sulfato de Amônio 28%, utilizando o R^2 da regressão linear múltipla como função de avaliação da resposta.

Dois tipos similares de respostas foram avaliadas, referentes à concentração de ingrediente ativo. Em um tipo, utilizou-se os resultados de ativo obtidos através da análise do surfactante pelo método potenciométrico (titulador automático com eletrodo para surfactante). Para o outro tipo de resposta, utilizou-se resultados do laudo de análise dos fornecedores destas matérias primas (titulação por duas fases utilizando *dimidium bromide/disulfine blue* como indicador). Cada resposta foi avaliada duas vezes visando maximizar o espaço amostral. Todas as avaliações foram realizadas simultaneamente e a execução total da avaliação não durou mais que 25 s (utilizando um Pentium 4, 1.5 GHz, 254 RAM).

Para fim de avaliação de desempenho do algoritmo, determinou-se que o mesmo não poderia utilizar mais que 4 comprimentos de onda na regressão (tamanho do cromossomo = 4), sendo que cada gene podia variar dentro de uma determinada faixa de comprimentos de onda, a saber:

Gene 1: 700 – 1300 cm^{-1}

Gene 2: 1301 – 1800 cm^{-1}

Gene 3: 2400 – 3300 cm^{-1}

Gene 4: 3301 – 4000 cm^{-1}

Outros parâmetros utilizados neste teste foram:

- Número de cromossomos: 30
- Critério de parada: 100 gerações
- % de elitismo: 10%
- % de mutação: 30%
- % cruzamento: 100%
- Tipo de cruzamento: Simples
- Tipo de mutação: em um único ponto
- Número de populações: 4 (duas para cada resposta)
- Migração: a cada 30 gerações.
- Tipo de Migração: Os melhores migram para qualquer outra população. A avaliação é feita com base na resposta da população que vai receber o cromossomo.
- Número de imigrantes: 3
- Número de amostras: 14 em duplicata (28 espectros)

As definições de resposta, correspondendo a cada população são:

Resposta 1 = Correlação x Concentração (laudo de análise do fornecedor)

Resposta 2 = Correlação x Concentração (laudo de análise do fornecedor)

Resposta 3 = Correlação x Concentração (resultados da titulação potenciométrica)

Resposta 4 = Correlação x Concentração (resultados da titulação potenciométrica)

A figura 19 a seguir mostra o desenvolvimento de cada população (somente no range 0,985 – 0,997, para facilitar a visualização).

Figura 19: Desenvolvimento de cada população quanto às respostas (R^2), para LESA 28%.

Cada ponto nestes gráficos representa um cromossomo (no caso, um conjunto de quatro comprimentos de onda, utilizadas na regressão linear). O gráfico mostra a resposta (correlação) de cada cromossomo em relação à concentração de surfactante, de acordo com a geração. Pode-se observar que cada população desenvolve-se independentemente das demais, mesmo havendo concentrações iguais (resposta 1 = Resposta 2 e Resposta 3 = Resposta 4). De 30 em 30 gerações, o patamar superior (maior correlação) é igualado devido à migração dos melhores cromossomos.

Observa-se que todas as populações se desenvolveram satisfatoriamente, no sentido de melhorar a correlação frente aos valores de concentração utilizados, alcançando ótimos valores de correlação em todos os casos.

A quantificação de lauril éter sulfato de amônio 28%, embora relevante e útil para a análise de rotina do material em controle de qualidade, se mostrou um problema simples, facilmente tratado pelo algoritmo, visto que diversas respostas satisfatórias (correlação maior que 0.99) foram encontradas rapidamente.

Vale ressaltar que qualquer modelo proposto necessita ser validado antes da efetiva implementação. Adicionalmente, deve-se conferir se os comprimentos de onda selecionados são quimicamente relevantes para a aplicação proposta. A figura 20 mostra (em vermelho) os comprimentos de onda selecionados em uma das possíveis respostas.

Figura 20: Comprimentos de onda selecionados em uma das possíveis resoluções para quantificação de LESA 28%

PARTE C: QUANTIFICAÇÃO DE O-BENZIL P-CLORO FENOL, O-FENIL FENOL E ETANOL (OPP/OBCP/ETANOL) NA MISTURA DOS TRÊS COMPONENTES

C.1. INTRODUÇÃO

C.1.1. Os bactericidas o-fenil fenol e o-benzil p-cloro fenol

As substâncias OPP, OBCP (vide figura 21) e suas misturas em diferentes proporções, utilizando etanol como solvente, são normalmente utilizadas como agentes desinfetantes em formulações de *home care* ^[47-48].

o-Fenil Fenol (OPP)

o-Benzil p-Cloro Fenol (OBCP)

Figura 21: Fórmulas estruturais para o-Fenil Fenol e o-Benzil p-Cloro Fenol.

Estes fenólicos são facilmente solúveis em solventes orgânicos, porém somente ligeiramente solúveis em água. Em geral, os derivados de fenólicos são bastante efetivos contra uma ampla variedade de organismos, incluindo bactérias, fungos e leveduras ^[48], incluindo *Pseudomonas spp*, *Mycobacterium tuberculosis*, *Rhizopus nigricans*, etc. A possibilidade de

mistura de diferentes fenólicos na formulação aumenta ainda mais o intervalo de eficiência em relação aos diferentes microorganismos.

O OPP é bastante utilizado em desinfetantes, adesivos, couro e até mesmo em frutas, possuindo toxicidade moderada (LD_{50} oral: 3000 mg/Kg em ratos), sendo porém irritante e corrosivo para as mucosas e moderadamente irritante para a pele. O OPP não é considerado agente mutagênico nem teratogênico. No ambiente, o OPP é degradado rápida e totalmente (100 ppm de OPP são degradados em 2 dias por microorganismos usuais da água de esgoto).

Em termos de variedade de microorganismos, o OBCP não atua em espectro tão amplo quanto o OPP, porém é mais eficiente contra alguns microorganismos específicos (ex: *Escherichia coli*). O OBCP é bastante irritante para pele e mucosas (LD_{50} oral: 15 g/Kg em ratos). Não é considerado mutagênico nem teratogênico. No ambiente, 50 ppm de OBCP são quase totalmente eliminados (75-100% eliminação) em 7 dias por microorganismos usuais da água de esgoto.

C.2. PARTE EXPERIMENTAL

Para quantificação da mistura o-benzil p-cloro fenol, o-fenil fenol e etanol (OPP/OBCP/Etanol) foram utilizadas 36 amostras preparadas em laboratório (através de pesagem das substâncias fenólicas de procedência Merck e dissolução em etanol), totalizando 46 espectros (espectros adquiridos em triplicata para algumas amostras).

Para obtenção dos espectros, foi utilizado o FTIR Spectrum BX (Perkin Elmer) com detector DTGS, equipado com acessório de reflectância atenuada (diamante - 9 reflexões). Os parâmetros utilizados foram:

Número de *Scans*: 16

Resolução: 2 cm^{-1}

Intervalo: $700 - 4000\text{ cm}^{-1}$

O programa utilizado para aquisição dos dados foi o *Spectrum* v.2.0 (Perkin Elmer).

Para teste do modelo de regressão escolhido foram utilizadas 6 amostras preparadas em laboratório, sendo adquiridos 3 espectros para cada amostra (totalizando 18 espectros). Estas 6 amostras (aqui denominadas amostras controle) não participaram em nenhum momento da modelagem da curva de calibração (conjunto de espectros independentes). Os espectros adquiridos foram numerados e as concentrações de cada analito são apresentadas na tabela 6.

Tabela 6: Concentração das amostras preparadas de OPP, OBCP e EtOH em relação a cada analito

Nº Espectro	Tipo	OPP	OBCP	EtOH
1	Calibração	40,18	19,27	39,82
2	Calibração	42,41	19,42	37,43
3	Calibração	45,33	19,27	34,68
4	Calibração	47,19	19,53	32,55
5	Calibração	49,98	19,47	29,83
6	Calibração	39,86	21,84	37,49
7	Calibração	42,53	22,09	34,58
8	Calibração	44,49	21,78	32,94
9	Calibração	47,42	21,93	29,86
10	Calibração	49,70	21,93	27,59
11	Calibração	40,27	24,24	34,62
12	Calibração	42,34	24,36	32,44

Tabela 6 (cont.): Concentração das amostras preparadas de OPP, OBCP e EtOH em relação a cada analito

Nº Espectro	Tipo	OPP	OBCP	EtOH
13	Calibração	44,29	25,46	29,36
14	Calibração	47,79	24,20	27,16
15	Calibração	49,93	24,30	24,91
16	Calibração	39,94	26,74	32,38
17	Calibração	41,97	27,22	29,86
18	Calibração	43,87	26,47	28,74
19	Calibração	47,46	26,67	24,95
20	Calibração	49,77	26,94	22,37
21	Calibração	40,24	29,01	29,76
22	Calibração	42,44	29,05	27,51
23	Calibração	44,87	29,21	24,92
24	Calibração	46,69	29,32	23,01
25	Calibração	49,96	29,29	19,77
26	Calibração	32,50	22,75	43,91
27	Calibração	46,96	22,70	29,53
28	Calibração	60,85	22,71	15,65
29	Calibração	46,87	15,80	36,70
30	Calibração	46,64	22,72	29,82
31	Calibração	46,26	29,19	23,56
32	Calibração	46,42	19,63	33,22
33	Calibração	46,42	19,63	33,22
34	Calibração	46,42	19,63	33,22
35	Calibração	46,26	23,34	29,57
36	Calibração	46,26	23,34	29,57
37	Calibração	46,26	23,34	29,57
38	Calibração	46,17	27,70	25,18
39	Calibração	46,17	27,70	25,18
40	Calibração	46,17	27,70	25,18
41	Calibração	41,61	23,22	34,33
42	Calibração	41,61	23,22	34,33
43	Calibração	41,61	23,22	34,33
44	Calibração	50,71	23,28	25,19
45	Calibração	50,71	23,28	25,19

Tabela 6 (cont.): Concentração das amostras preparadas de OPP, OBCP e EtOH em relação a cada analito

Nº Espectro	Tipo	OPP	OBCP	EtOH
46	Calibração	50,71	23,28	25,19
47	Controle	45,52	20,70	33,02
48	Controle	45,52	20,70	33,02
49	Controle	45,52	20,70	33,02
50	Controle	45,63	23,31	30,23
51	Controle	45,63	23,31	30,23
52	Controle	45,63	23,31	30,23
53	Controle	46,00	25,68	27,43
54	Controle	46,00	25,68	27,43
55	Controle	46,00	25,68	27,43
56	Controle	44,10	23,21	31,85
57	Controle	44,10	23,21	31,85
58	Controle	44,10	23,21	31,85
59	Controle	47,90	22,95	28,33
60	Controle	47,90	22,95	28,33
61	Controle	47,90	22,95	28,33
62	Controle	46,25	22,84	30,08
63	Controle	46,25	22,84	30,08
64	Controle	46,25	22,84	30,08

C.3. RESULTADOS E DISCUSSÃO

A figura 10 mostra os espectros de OPP 50% em etanol, OBCP 20% em etanol, etanol puro e da mistura OPP/OBCP 70% em etanol. Nota-se que todos os espectros são muito parecidos, com a bandas referentes ao etanol sobressaindo-se bastante, e espera-se que o algoritmo genético seja capaz de selecionar quatro comprimentos de onda para quantificação de cada um dos analitos de forma simultânea. Ao mesmo tempo, espera-se que o conjunto de comprimentos de onda selecionado para cada analito não apresente boa correlação para os outros analitos.

Figura 22: Espectros de OPP 50% em etanol, OBCP 20% em etanol, etanol puro e da mistura OPP/OBCP 70% em etanol.

O algoritmo foi inicializado com os seguintes parâmetros:

- Número de cromossomos: 30
- Critério de parada: 100 gerações
- % de Elitismo: 10%
- % de mutação: 30%
- % Cruzamento: 100%
- Tamanho do cromossomo: 4 genes (4 números de onda serão utilizados)
- Número de populações: 3 (uma para cada analito)
- Migração: A cada 10 gerações.

- Condição para migração: Melhor que o 5º cromossomo da população destino.

A implementação simultânea de três populações e o uso da técnica de migração permite que as respostas desejadas sejam encontradas mais rapidamente, já que os cromossomos “promissores” são automaticamente identificados em intervalos regulares e migram para as outras populações. Além disso, a exploração do espaço amostral é melhorada, pois, ao todo 90 cromossomos (30 de cada população) são avaliados a cada geração, enquanto a direção da evolução permanece diferente (cada população se desenvolve independentemente em relação a um analito). A execução do algoritmo leva menos que 2 min.

As figuras 23, 24 e 25 mostram as curvas de regressão e os melhores cromossomos (conjunto de quatro comprimentos de onda) selecionados para cada analito, bem como a correlação deste cromossomo para os outros analitos. Nas figuras correspondentes o eixo das abcissas traz as concentrações do analito (% m/m) nos padrões preparados, e o eixo das ordenadas traz o valor calculado com os 4 números de onda (em cm^{-1}) também apresentados na figura (“cromossomo”). Adicionalmente, também são mostradas na figura as respostas do cromossomo (correlação) em relação à cada um dos analitos.

Em todos os casos, nota-se boa correlação para o analito de interesse, enquanto a correlação para os outros analitos se manteve baixa, o que é preferível, pois indica baixa covariância. Isto decorre da seleção manual entre os dez melhores cromossomos, realizada buscando o melhor compromisso entre a correlação alta para o analito de interesse e baixa para os demais analitos. Na tabela de resultados obtidos pelo algoritmo é possível visualizar simultaneamente a correlação de cada resposta em relação ao três analitos, o que facilita seleção manual. Esta seleção poderia ser automatizada para aplicações rotineiras.

Para teste das curvas de regressão foram utilizadas 18 amostras controle, sendo que a diferença de resultados em relação ao esperado (resíduo = resultado esperado – resultado obtido) pode ser visualizada nas figuras 26, 27 e 28, para cada analito. As concentrações previstas para as amostras de regressão e controle são mostradas nas tabelas 7, 8 e 9.

Figura 23: Curva de regressão e números de onda (cromossomo, em cm^{-1}) para o analito OPP. A figura também indica a correlação do cromossomo para cada analito

Figura 24: Resíduo (diferença entre o valor teórico e o previsto) para cada amostra em relação ao teor de OPP esperado. As amostras controle são indicadas no gráfico

Figura 25: Curva de regressão e números de onda (cromossomo, em cm^{-1}) para o analito OBCP. A figura também indica a correlação do cromossomo para cada analito

Figura 26: Resíduo (diferença entre o valor teórico e o previsto) para cada amostra em relação ao teor de OBCP esperado. As amostras controle são indicadas no gráfico

Figura 27: Curva de regressão e números de onda (cromossomo, em cm^{-1}) para o analito EtOH. A figura também indica a correlação do cromossomo para cada analito

Figura 28: Resíduo (diferença entre o valor teórico e o previsto) para cada amostra em relação ao teor de EtOH. As amostras controle são indicadas no gráfico

Tabela 7: Comparação entre a concentração real e os resultados obtidos por FTIR-ATR para o analito OPP

Nº Espectro	Tipo	% Real	% Prevista (FTIR)	% Real - % Prevista
1	Calibração	40,18	39,80	0,38
2	Calibração	42,41	41,86	0,55
3	Calibração	45,33	44,70	0,63
4	Calibração	47,19	46,68	0,51
5	Calibração	49,98	49,86	0,12
6	Calibração	39,86	39,55	0,32
7	Calibração	42,53	42,32	0,21
8	Calibração	44,49	44,10	0,39
9	Calibração	47,42	47,48	-0,06
10	Calibração	49,70	49,62	0,08

Tabela 7 (cont): Comparação entre a concentração real e os resultados obtidos por FTIR-ATR para o analito OPP

Nº Espectro	Tipo	% Real	% Prevista (FTIR)	% Real - % Prevista
11	Calibração	40,27	40,10	0,17
12	Calibração	42,34	42,32	0,01
13	Calibração	44,29	44,62	-0,33
14	Calibração	47,79	47,60	0,18
15	Calibração	49,93	49,45	0,48
16	Calibração	39,94	40,06	-0,12
17	Calibração	41,97	42,53	-0,56
18	Calibração	43,87	44,44	-0,58
19	Calibração	47,46	47,73	-0,27
20	Calibração	49,77	49,85	-0,09
21	Calibração	40,24	41,05	-0,82
22	Calibração	42,44	43,22	-0,78
23	Calibração	44,87	45,29	-0,41
24	Calibração	46,69	47,04	-0,36
25	Calibração	49,96	50,02	-0,07
26	Calibração	32,50	32,74	-0,25
27	Calibração	46,96	47,01	-0,05
28	Calibração	60,85	60,48	0,37
29	Calibração	46,87	45,74	1,14
30	Calibração	46,64	46,66	-0,02
31	Calibração	46,26	45,76	0,50
32	Calibração	46,42	46,03	0,38
33	Calibração	46,42	45,93	0,48
34	Calibração	46,42	46,26	0,15
35	Calibração	46,26	46,30	-0,04
36	Calibração	46,26	46,34	-0,08
37	Calibração	46,26	46,30	-0,04
38	Calibração	46,17	46,69	-0,52
39	Calibração	46,17	46,63	-0,46
40	Calibração	46,17	46,64	-0,47
41	Calibração	41,61	41,42	0,19
42	Calibração	41,61	41,33	0,28
43	Calibração	41,61	41,51	0,10
44	Calibração	50,71	51,13	-0,41
45	Calibração	50,71	51,24	-0,53
46	Calibração	50,71	51,03	-0,32

Tabela 7 (cont): Comparação entre a concentração real e os resultados obtidos por FTIR-ATR para o analito OPP

Nº Espectro	Tipo	% Real	% Prevista (FTIR)	% Real - % Prevista
47	Controle	45,52	46,04	-0,52
48	Controle	45,52	45,75	-0,23
49	Controle	45,52	45,77	-0,25
50	Controle	45,63	46,18	-0,55
51	Controle	45,63	46,09	-0,46
52	Controle	45,63	46,31	-0,68
53	Controle	46,00	46,41	-0,41
54	Controle	46,00	46,43	-0,43
55	Controle	46,00	46,49	-0,49
56	Controle	44,10	44,73	-0,63
57	Controle	44,10	44,50	-0,40
58	Controle	44,10	44,39	-0,29
59	Controle	47,90	48,28	-0,39
60	Controle	47,90	48,15	-0,25
61	Controle	47,90	48,08	-0,18
62	Controle	46,25	46,46	-0,21
63	Controle	46,25	46,27	-0,02
64	Controle	46,25	46,32	-0,07

Tabela 8: Comparação entre a concentração real e os resultados obtidos por FTIR-ATR para o analito OBCP

Nº Espectro	Tipo	% Real	% Prevista (FTIR)	% Real - % Prevista
1	Calibração	19,27	19,44	-0,17
2	Calibração	19,42	19,37	0,06
3	Calibração	19,27	19,42	-0,15
4	Calibração	19,53	19,43	0,10
5	Calibração	19,47	19,44	0,03
6	Calibração	21,84	21,77	0,07
7	Calibração	22,09	22,12	-0,03
8	Calibração	21,78	21,92	-0,14
9	Calibração	21,93	21,88	0,04

Tabela 8 (cont): Comparação entre a concentração real e os resultados obtidos por FTIR-ATR para o analito OBCP

Nº Espectro	Tipo	% Real	% Prevista (FTIR)	% Real - % Prevista
10	Calibração	21,93	21,83	0,10
11	Calibração	24,24	24,29	-0,05
12	Calibração	24,36	24,35	0,01
13	Calibração	25,46	25,31	0,14
14	Calibração	24,20	24,26	-0,06
15	Calibração	24,30	24,29	0,02
16	Calibração	26,74	26,50	0,24
17	Calibração	27,22	27,19	0,03
18	Calibração	26,47	26,48	-0,01
19	Calibração	26,67	26,80	-0,13
20	Calibração	26,94	27,07	-0,13
21	Calibração	29,01	29,06	-0,05
22	Calibração	29,05	29,02	0,04
23	Calibração	29,21	28,96	0,26
24	Calibração	29,32	29,30	0,02
25	Calibração	29,29	29,48	-0,19
26	Calibração	22,75	22,67	0,08
27	Calibração	22,70	22,61	0,09
28	Calibração	22,71	22,53	0,18
29	Calibração	15,80	16,20	-0,40
30	Calibração	22,72	22,29	0,43
31	Calibração	29,19	29,96	-0,77
32	Calibração	19,63	19,81	-0,18
33	Calibração	19,63	19,67	-0,03
34	Calibração	19,63	19,66	-0,03
35	Calibração	23,34	23,16	0,18
36	Calibração	23,34	23,27	0,07
37	Calibração	23,34	23,16	0,18
38	Calibração	27,70	27,74	-0,05
39	Calibração	27,70	27,65	0,04
40	Calibração	27,70	27,35	0,35
41	Calibração	23,22	23,36	-0,14
42	Calibração	23,22	23,21	0,01
43	Calibração	23,22	23,35	-0,13
44	Calibração	23,28	23,23	0,05
45	Calibração	23,28	23,36	-0,08
46	Calibração	23,28	23,15	0,13
47	Controle	20,70	21,17	-0,47

Tabela 8 (cont.): Comparação entre a concentração real e os resultados obtidos por FTIR-ATR para o analito OBCP

Nº Espectro	Tipo	% Real	% Prevista (FTIR)	% Real - % Prevista
48	Controle	20,70	21,12	-0,42
49	Controle	20,70	20,98	-0,28
50	Controle	23,31	23,36	-0,05
51	Controle	23,31	23,57	-0,26
52	Controle	23,31	23,52	-0,21
53	Controle	25,68	25,93	-0,25
54	Controle	25,68	25,80	-0,13
55	Controle	25,68	25,85	-0,17
56	Controle	23,21	23,70	-0,49
57	Controle	23,21	23,42	-0,20
58	Controle	23,21	23,33	-0,11
59	Controle	22,95	23,17	-0,22
60	Controle	22,95	23,26	-0,31
61	Controle	22,95	23,09	-0,14
62	Controle	22,84	22,96	-0,11
63	Controle	22,84	23,09	-0,25
64	Controle	22,84	22,82	0,02

Tabela 9 : Comparação entre a concentração real e os resultados obtidos por FTIR-ATR para o analito EtOH

Nº Espectro	Tipo	% Real	% Prevista (FTIR)	% Real - % Prevista
1	Calibração	39,82	39,65	0,17
2	Calibração	37,43	37,72	-0,29
3	Calibração	34,68	35,00	-0,32
4	Calibração	32,55	32,99	-0,44
5	Calibração	29,83	30,09	-0,27
6	Calibração	37,49	37,34	0,15
7	Calibração	34,58	34,58	0,00
8	Calibração	32,94	33,03	-0,09
9	Calibração	29,86	30,07	-0,21
10	Calibração	27,59	27,74	-0,15

Tabela 9 (cont.): Comparação entre a concentração real e os resultados obtidos por FTIR-ATR para o analito EtOH

Nº Espectro	Tipo	% Real	% Prevista (FTIR)	% Real - % Prevista
11	Calibração	34,62	34,20	0,42
12	Calibração	32,44	32,30	0,13
13	Calibração	29,36	29,19	0,18
14	Calibração	27,16	27,18	-0,02
15	Calibração	24,91	24,98	-0,07
16	Calibração	32,38	31,97	0,41
17	Calibração	29,86	29,49	0,38
18	Calibração	28,74	28,11	0,64
19	Calibração	24,95	24,74	0,21
20	Calibração	22,37	22,49	-0,12
21	Calibração	29,76	28,86	0,90
22	Calibração	27,51	26,69	0,82
23	Calibração	24,92	24,44	0,48
24	Calibração	23,01	22,85	0,16
25	Calibração	19,77	20,43	-0,66
26	Calibração	43,91	43,00	0,91
27	Calibração	29,53	29,67	-0,14
28	Calibração	15,65	16,25	-0,60
29	Calibração	36,70	37,43	-0,74
30	Calibração	29,82	29,79	0,03
31	Calibração	23,56	23,57	-0,01
32	Calibração	33,22	33,87	-0,65
33	Calibração	33,22	33,94	-0,72
34	Calibração	33,22	33,75	-0,54
35	Calibração	29,57	29,82	-0,25
36	Calibração	29,57	29,91	-0,34
37	Calibração	29,57	29,79	-0,22
38	Calibração	25,18	25,07	0,11
39	Calibração	25,18	25,04	0,15
40	Calibração	25,18	24,70	0,48
41	Calibração	34,33	34,26	0,07
42	Calibração	34,33	34,58	-0,25
43	Calibração	34,33	34,23	0,10
44	Calibração	25,19	25,02	0,17
45	Calibração	25,19	24,99	0,20
46	Calibração	25,19	25,34	-0,15
47	Controle	33,02	31,89	1,13
48	Controle	33,02	32,30	0,71

Tabela 9 (cont.): Comparação entre a concentração real e os resultados obtidos por FTIR-ATR para o analito EtOH

Nº Espectro	Tipo	% Real	% Prevista (FTIR)	% Real - % Prevista
49	Controle	33,02	32,52	0,50
50	Controle	30,23	29,69	0,53
51	Controle	30,23	29,62	0,61
52	Controle	30,23	29,47	0,76
53	Controle	27,43	26,89	0,53
54	Controle	27,43	27,06	0,37
55	Controle	27,43	26,97	0,45
56	Controle	31,85	30,49	1,37
57	Controle	31,85	31,30	0,55
58	Controle	31,85	31,40	0,45
59	Controle	28,33	27,66	0,67
60	Controle	28,33	27,94	0,39
61	Controle	28,33	28,18	0,15
62	Controle	30,08	29,84	0,24
63	Controle	30,08	29,87	0,22
64	Controle	30,08	30,04	0,04

A tabela 10 mostra a média do módulo dos resíduos absolutos e o módulo do resíduo máximo obtido em cada modelo de regressão, para as amostras de regressão e para as amostras controle.

Tabela 10: Média dos módulos de cada erro absoluto e módulo do erro máximo observado para cada analito, para cada conjunto de amostras

Analito	Resíduo Médio, em %			
	Absoluto		Relativo	
	Amostras Regressão	Amostras Controle	Amostras Regressão	Amostras Controle
OPP	0,33	0,36	0,66	0,72
OBCP	0,13	0,23	0,26	1,15
Etanol	0,32	0,54	0,64	1,80

Para fins de controle de qualidade, o modelo obtido foi considerado aceitável, com baixo erro principalmente em relação ao analitos OPP e OBCP (principais ingredientes e responsáveis pela eficiência da matéria prima). O método manteve-se robusto durante o período de avaliação, que foi de um ano.

PARTE D: QUANTIFICAÇÃO DE LAURIL ÉTER SULFATO DE SÓDIO EM
AMOSTRAS DE SHAMPOO

D.1. INTRODUÇÃO

D.1.1. Formulações de shampoo

Essencialmente, *shampoos* são detergentes líquidos destinados à limpeza do cabelo e couro cabeludo. Todavia, as características estéticas deste grupo, como aparência, viscosidade, espuma e fragrância também são de vital importância para o consumidor.

Os principais ingredientes que fazem parte de uma formulação de *shampoo* são^[49-50]:

- Surfactante primário para limpeza e espuma
- Surfactante secundário para melhoria da espuma, viscosidade e/ou condicionamento.
- Espessantes: gomas, sal, amidas.
- Opacificantes para efeitos visuais
- Acidulantes para ajuste de pH
- Fragrância
- Preservante
- Componentes especiais destinados à propriedades específicas, como ativos vegetais, agentes condicionadores, substância para tratamento da caspa, etc.

A ação dos *shampoos* envolve a limpeza de sujidades e remoção de gordura do cabelo e ligação de substâncias condicionantes ao cabelo. Usualmente estas interações ocorrem na camada

superficial do cabelo (cutícula). Em termos de concentração, os *shampoos* são formados basicamente pelos surfactantes primários, surfactantes secundários e água.

Os surfactantes primários mais utilizados em formulações de *shampoo* são lauril sulfato de amônio e lauril éter sulfato de sódio (2-3 mols de óxido de etileno), devido à boa performance na remoção de gordura, sujeira e boa formação de espuma (a introdução da parte B trata o assunto com maiores detalhes).

Entre os surfactantes secundários, os tipos mais comuns são:

- Dietanolamida do ácido graxo de coco (alcanolamidas de ácidos graxos): São agentes sobreengordurantes para o cabelo, ajudam na solubilização de essências e no espessamento, além de serem estabilizantes de espuma. A estrutura química desses compostos é mostrada na figura 29.

Figura 29: Estrutura química de uma dietanolamida

- Alquilpoliglucosídeos (APGs): São obtidos a partir da condensação da glucose do milho com o óleo de coco ou óleo de palmito. Possuem a estrutura química mostrada na figura 30.

parte hidrófoba (m entre 6 e 14)

parte hidrófila (n entre 1 e 2)

Figura 30: Estrutura química de um alquilpoliglicosídeo

São tensoativos suaves, diminuem o poder de irritação dos surfactantes primários, ajudam a dar viscosidade (alguns tipos de APGs) e melhoram a espuma (diminuem o tamanho das bolhas, resultando numa espuma mais cremosa e mais estável).

- Betaínas: São tensoativos anfóteros (se comportam como catiônicos em valores baixos de pH e como aniônicos em pH alto. Em pH intermediário, apresenta carga positiva e negativa na molécula). Entre as betaínas, a cocoamidopropilbetaína é extensamente utilizada nas formulações de *shampoos*. Sua estrutura química é mostrada na figura 31.

Figura 31: Estrutura química da cocoamidopropilbetaína

As betaínas apresentam baixo potencial de irritação dos olhos e são efetivas como estabilizadoras de espuma. Possuem afinidade pela queratina e propriedades condicionadoras e ajudam na obtenção da viscosidade com o uso de eletrólitos, além de serem compatíveis com todas as outras classes de surfactantes.

Os ingredientes restantes, embora bastante importantes, são utilizados em baixa concentração (usualmente menos que 5%). No espectro infravermelho, é bastante difícil detectar a presença destes componentes.

D.2. PARTE EXPERIMENTAL

Para o desenvolvimento da curva de calibração, foram preparadas em laboratório 44 bases de *shampoos*, contendo somente os principais ingredientes da formulação: Lauril Éter Sulfato de Sódio, Cocoamidopropil Betaína, Cocodietanolamida e Água. O planejamento é mostrado na figura 32.

Figura 32: Planejamento experimental para bases de shampoos (em vermelho: bases preparadas)

Para obtenção dos espectros, foi utilizado o FTIR Spectrum BX (Perkin Elmer) com detector DTGS, equipado com acessório de reflectância atenuada (diamante - 9 reflexões). Os parâmetros utilizados foram:

Número de Scans: 16

Resolução: 2 cm^{-1}

Intervalo: $700 - 4000 \text{ cm}^{-1}$

O programa utilizado para aquisição dos dados foi o Spectrum v.2.0 (Perkin Elmer).

Para teste da curva de calibração, foram utilizadas 30 amostras de *shampoo* (amostras controle), que tiveram seu teor de surfactante aniônico determinado em duplicata pelo método de referência (titulação potenciométrica com eletrodo seletivo). As 30 amostras são representativas de 5 grupos de fórmulas diferentes (onde há variação nas concentrações dos principais componentes, além de diferenças nos extratos, perfumes e corantes utilizados).

Adicionalmente, iniciou-se a validação do método utilizando-se 5 amostras de validação, cada uma com um nível diferente de surfactante (70%, 85%, 100%, 115% e 130% da quantidade original do analito). As amostras de validação foram preparadas em laboratório, com fórmula representativa de um dos grupos testados. Os espectros para cada amostra foram adquiridos em triplicata e avaliou-se a linearidade do método, bem como a precisão e exatidão para as amostras níveis 70%, 100% e 130%. O protocolo de validação utilizado está baseado no protocolo do FDA - *Food and Drugs Association Validation Guindance* ^[51-53].

No desenvolvimento da curva da calibração, utilizou-se o algoritmo genético previamente desenvolvido, sem pré-tratamento dos dados espectrais. Os parâmetros utilizados na configuração do algoritmo foram:

- Número de cromossomos: 30
- Critério de parada: 100 gerações
- % de Elitismo: 10%
- % de mutação: 30%
- % Cruzamento: 100%
- Tamanho do cromossomo: 7 genes (7 números de onda serão utilizados)
- Número de populações: 1

Somente as regiões do espectro consideradas representativas foram utilizadas pelo algoritmo, conforme mostra a tabela 11 (a seleção das regiões é feita pelo usuário, que pode utilizar sua experiência prévia para determinar regiões mais promissoras). Desta forma, cada gene somente poderia assumir valores restritos a uma região específica e delimitada (denominada aqui intervalo de trabalho). Este método acelera o desenvolvimento do algoritmo, já que o espaço

amostral para cada gene se torna menor. Adicionalmente, esta abordagem obriga que os números de onda sejam selecionados na região de interesse. Por outro lado, esta abordagem não oferece a possibilidade de seleção de mais de um número de onda na região delimitada (a não ser que o algoritmo seja especificamente configurado para tal).

Tabela 11: Regiões de limitação de números de onda para cada gene

Gene	Região (cm ⁻¹)	
	Início	Fim
1	1011	1028
2	1058	1074
3	1075	1120
4	1195	1229
5	1230	1265
6	2847	2859
7	2915	2931

D.3. RESULTADOS E DISCUSSÃO

O método de referência utilizado para a elaboração da curva da calibração foi a titulação potenciométrica com eletrodo para surfactante (equipamento e eletrodos da marca Orion). Os principais parâmetros utilizados na titulação foram:

Pré Dose: 1 mL

Volume Máximo: 11 mL

Critério de estabilização: 3 mV/min

Incremento de potencial para cálculo do volume dispensado: 5 mv

Massa Molar: 384

Cálculo do Volume de equivalência: Primeira Derivada

O coeficiente de variação deste método, em condições normais de trabalho, foi calculado como sendo 1,28% (o cálculo envolveu 32 titulações de uma mesma amostra, realizadas ao longo de 3 meses). O salto de potencial é da ordem de 30 mV.

A figura 33 mostra um espectro FTIR típico de lauril éter sulfato de sódio 27%, com as devidas atribuições (vide parte 1 para comparação com o espectro do lauril éter sulfato de amônio e comentários adicionais). A figura 34 mostra o mesmo espectro em comparação com o espectro típico para uma formulação de *shampoo* contendo 10% de LESS e em comparação com o espectro de LESS 27% (matéria prima utilizada). É possível observar a presença das mesmas bandas em todos os espectros. Ao comparar o espectro do *shampoo* com o de LESS 10% em água, nota-se que o *shampoo* apresenta absorbância ligeiramente maior, particularmente na região $1000 - 1300\text{ cm}^{-1}$, provavelmente devido à interferência de outros componentes da formulação. Ainda assim, não é possível distinguir claramente outras bandas além das pertencentes ao lauril éter sulfato de sódio.

Figura 33: Espectro FTIR do lauril éter sulfato de sódio 27% (LESS 27%)

Figura 34: Comparação entre os espectros FTIR de LESS 27%, de LESS diluído à 10% em água e de um *shampoo* contendo 10% de LESS

Em relação ao desenvolvimento do método de quantificação por FTIR, o algoritmo genético foi inicializado e, ao término do processamento, foram selecionadas as dez melhores respostas encontradas pelo algoritmo (cada uma com sete comprimentos de onda). Estas dez respostas foram avaliadas e selecionou-se a melhor resposta considerando o residual (resultado da titulação – resultado FTIR) encontrado nas amostras-teste. A figura 35 mostra os comprimentos de onda escolhidos. Vale notar que as regiões com forte absorção da água foram evitadas já na etapa de seleção do intervalo de trabalho para cada gene.

Figura 35: Comprimentos de onda escolhidos pelo algoritmo, nas regiões pré-selecionadas

A tabela 12 mostra os resultados obtidos para as amostras de calibração e controle em relação ao esperado. A figura 36 (curva de regressão obtida) e a figura 37 (gráfico de controle para as amostras-teste) ilustram os resultados obtidos.

Figura 36: Curva de regressão linear para as amostras de calibração.

Figura 37: Resultados de titulação potenciométrica e FTIR para as amostras-teste

Tabela 12: Resultados obtidos (FTIR) e esperados (Titulação Potenciométrica) para as amostras de shampoo

Nº Espectro	Tipo	% Esperada (Titulação Potenciométrica)	% Obtida (FTIR)	% Esperada - % Obtida
1	Calibração	5,29	5,39	-0,10
2	Calibração	7,92	7,98	-0,06
3	Calibração	10,55	10,54	0,01
4	Calibração	13,12	13,05	0,07
5	Calibração	5,30	5,31	-0,01
6	Calibração	7,98	8,01	-0,03
7	Calibração	10,62	10,62	0,00
8	Calibração	13,13	13,16	-0,03
9	Calibração	5,27	5,22	0,05
10	Calibração	7,84	7,85	-0,01
11	Calibração	10,49	10,46	0,03
12	Calibração	13,26	13,28	-0,02
13	Calibração	5,15	5,11	0,04
14	Calibração	7,99	7,90	0,09
15	Calibração	10,54	10,74	-0,20
16	Calibração	13,27	13,24	0,03
17	Calibração	7,94	7,85	0,09
18	Calibração	7,94	7,81	0,13
19	Calibração	5,23	5,25	-0,02
20	Calibração	7,99	7,96	0,03
21	Calibração	10,48	10,44	0,04
22	Calibração	13,07	13,04	0,03
23	Calibração	5,30	5,26	0,04
24	Calibração	7,89	7,88	0,01
25	Calibração	10,52	10,52	0,00
26	Calibração	13,07	13,12	-0,05
27	Calibração	5,21	5,25	-0,04
28	Calibração	7,92	8,04	-0,12
29	Calibração	10,56	10,55	0,01
30	Calibração	13,13	13,18	-0,05
31	Calibração	5,29	5,25	0,04
32	Calibração	7,93	7,95	-0,02
33	Calibração	10,59	10,60	-0,01
34	Calibração	13,23	13,22	0,01
35	Calibração	5,29	5,29	0,00
36	Calibração	7,95	7,86	0,09
37	Calibração	10,59	10,70	-0,11

Tabela 12 (cont.): Resultados obtidos (FTIR) e esperados (Titulação Potenciométrica) para as amostras de shampoo

Nº Espectro	Tipo	% Esperada (Titulação Potenciométrica)	% Obtida (FTIR)	% Esperada - % Obtida
38	Calibração	13,18	13,04	0,14
39	Calibração	7,84	7,97	-0,13
40	Calibração	10,68	10,62	0,06
41	Calibração	5,41	5,39	0,02
42	Calibração	8,20	8,16	0,04
43	Calibração	8,09	8,18	-0,09
44	Controle	5,90	5,83	0,07
45	Controle	5,92	5,76	0,16
46	Controle	5,89	5,97	-0,08
47	Controle	5,77	6,02	-0,25
48	Controle	5,64	6,01	-0,37
49	Controle	6,00	6,00	0,00
50	Controle	5,88	5,84	0,04
51	Controle	5,95	5,91	0,04
52	Controle	5,67	5,90	-0,23
53	Controle	5,60	5,69	-0,09
54	Controle	5,56	5,91	-0,35
55	Controle	5,93	5,90	0,03
56	Controle	10,11	9,85	0,26
57	Controle	10,06	9,65	0,41
58	Controle	9,80	9,57	0,23
59	Controle	10,24	10,21	0,03
60	Controle	10,41	10,00	0,41
61	Controle	10,01	10,20	-0,19
62	Controle	10,46	10,55	-0,09
63	Controle	10,77	10,90	-0,13
64	Controle	10,61	10,72	-0,11
65	Controle	10,39	10,53	-0,14
66	Controle	10,11	10,63	-0,52
67	Controle	10,11	10,34	-0,23
68	Controle	10,10	10,44	-0,34
69	Controle	10,90	10,86	0,04
70	Controle	10,48	10,91	-0,43
71	Controle	11,10	11,30	-0,20
72	Controle	9,95	9,92	0,03
73	Controle	10,00	10,21	-0,21

Para validação do método, foram utilizadas as amostras preparadas especificamente para tal e a tabela 13 mostra os resultados obtidos. A figura 38 mostra a linearidade do método (critério de aceitação: $R^2 \geq 0,98$). A figura 39 mostra os resultados de precisão e exatidão obtidos (critérios de aceitação: coeficiente de variação menor que 2,5% e intervalo de confiança entre 95 e 105% para os três níveis).

Tabela 13: Resultados Obtidos para as amostras de validação

Nível (Calculado)	% Ativo Real	Análise 1	Análise 2	Análise 3
70,02	7,02	7,16	7,11	7,14
85,40	8,54	8,45	8,53	8,46
100,10	10,01	9,84	9,95	10,00
115,50	11,55	11,48	11,41	11,31
130,10	13,01	12,89	12,73	12,84

Nível	x		y			
	Adicionado	Média	Calculado	95 % ci (min)	95 % ci (max)	
1	7,02	7,137	7,089	6,988	7,188	Regressão linear: $y = a + bx$ $a = 0,3984$ $b = 0,9530$
2	8,54	8,480	8,537	8,468	8,608	
3	10,01	9,930	9,938	9,880	9,996	
4	11,55	11,400	11,406	11,334	11,477	Coef. De Correlação: 0,9995 Critério: $\geq 0,98$
5	13,01	12,820	12,797	12,697	12,898

Figura 38: Avaliação da linearidade do método de FTIR (amostras de validação)

Nível	x		y		Recuperação	
	Adicionado	Média	Desvio Padrão	Coefficiente de Variação	Média	95 % Intervalo de Confiança
1 (70 %)	7,02	7,1367	0,0252	0,35	101,7	100,8 a 102,6 %
2 (100%)	10,01	9,93	0,0819	0,82	99,2	97,2 a 101,2 %
3 (130 %)	13,01	12,82	0,0819	0,64	98,5	97,0 a 100,1 %

Crêterios: Coeficiente de variaçã0 menor que 2,5 % e intervalo de confiançã0 entre 95 e 105 %

Figura 39: Avaliação da precisão e exatidão do método de FTIR (amostras de validação, níveis 70%, 100% e 130%)

PARTE E: ESTUDO DA INFLUÊNCIA DOS PARÂMETROS DE CONFIGURAÇÃO NO DESEMPENHO DO ALGORITMO GENÉTICO

E.1. INTRODUÇÃO

Em virtude das próprias características do algoritmo genético, não é tarefa simples generalizar a definição dos parâmetros a serem utilizados para obter o melhor rendimento do mesmo. Os parâmetros utilizados para abordagem de um determinado problema podem não servir para outro tipo. O desempenho do algoritmo depende fortemente da função de avaliação utilizada, da superfície de resposta do problema estudado e mesmo das características particulares envolvidas na programação do algoritmo. Nesta etapa, pretende-se avaliar e discutir a influência de alguns parâmetros de configuração do algoritmo desenvolvido no problema específico de seleção de comprimentos de onda em calibração multivariada, utilizando FTIR-ATR.

Para se avaliar devidamente o desempenho do algoritmo, faz-se necessário conhecer previamente e com exatidão as combinações de números de onda que resultam nas melhores respostas. Para satisfazer tal requerimento, optou-se por simular matematicamente os espectros utilizados no estudo, a partir de espectros reais de uma possível aplicação da técnica de FTIR-ATR. A aplicação escolhida neste caso foi a quantificação de silicone em amostras de sabonete em barra.

Os sabonetes são obtidos a partir da saponificação de gorduras. O processo de fabricação de um sabonete envolve várias etapas. Resumidamente, o processo inicia com o branqueamento e desodorização da gordura bruta, que segue para a reação de saponificação com NaOH, dando origem ao sabão (normalmente o sabão resultante contém desde cadeias C12 até C18). O sabão é

então seco e cortado, sendo transformado no que se denomina “chips”. Os chips são posteriormente misturados com diversos aditivos (corantes, pigmentos, preservantes, substâncias emolientes, princípios ativos, etc), moldados na forma de barra e prensados, dando origem ao sabonete de uso pessoal.

O termo silicone é utilizado para descrever siloxanos organo modificados de fórmula geral R_2SiO ^[54], onde R pode ser um grupamento alquila, fenila, poliéter, etc. (R_2SiO é comumente chamado de grupamento siloxila). Os diferentes tipos de polímeros organo siloxanos possuem propriedades únicas e são bastante utilizados em formulações cosméticas.

Entre os diversos tipos de silicone, encontram-se os denominados dimetil polisiloxanos, cuja estrutura típica é mostrada na figura 40.

Figura 40: Estrutura típica dos dimetil polisiloxanos ^[55]

A viscosidade deste tipo de silicone pode variar de 0,65 até 100.000 centistokes (cS). O silicone 60000, fornecido comercialmente, é um exemplo deste tipo de silicone. O mesmo possui massa molar aproximada de 116.500 u.m.a, viscosidade em torno de 60.000 cS e ao redor de 1.570 unidades monoméricas.

O silicones normalmente são utilizados em produtos de *Personal Care* visando melhorar a percepção deixada pelo produto em relação a maciez, emoliência e espalhabilidade, entre outras características. Para que o consumidor percebe mudança sensorial, normalmente são necessários 0,1 – 5 % de silicone na formulação do produto.

E.2. PARTE EXPERIMENTAL

Para simulação matemática dos espectros utilizados no estudo, foram utilizados como ponto de partida espectros reais de uma amostra de sabonete comercial (sabonete Palmolive Suave, que não contém silicone em sua composição) e da matéria prima Silicone 60000 (fornecedor G.E.).

Para obtenção dos espectros, foi utilizado o FTIR Spectrum BX (Perkin Elmer) com detector DTGS, equipado com acessório de reflectância atenuada (diamante - 1 reflexão). Os parâmetros utilizados foram:

Número de *Scans*: 16

Resolução: 2 cm^{-1}

Intervalo: $700 - 4000\text{ cm}^{-1}$

O programa utilizado para aquisição dos dados foi o *Spectrum* v.2.0 (Perkin Elmer).

A partir dos espectros reais, foram obtidos matematicamente 20 espectros através da simples adição do espectro do silicone (multiplicado por um determinado fator) ao espectro do sabonete, ou seja, nesta primeira etapa considerou-se que não há desvios da lei de Lambert-Beer. As absorbâncias em cada número de onda foram somadas proporcionalmente de forma a simular espectros com as concentrações de silicone apresentadas na tabela 14.

Tabela 14: Concentrações teóricas de silicone nos espectros simulados

Número do espectro	Concentração de Silicone (%)
1	0,02
2	0,04
3	0,06
4	0,08
5	0,10
6	0,12
7	0,14
8	0,16
9	0,18
10	0,20
11	0,22
12	0,24
13	0,26
14	0,28
15	0,30
16	0,32
17	0,34
18	0,36
19	0,38
20	0,40

Para simplificar o tratamento matemático, optou-se por realizar o estudo com somente dois números de onda configurados no algoritmo, de forma que foi necessária a definição de duas regiões do espectro onde os números de onda, se selecionados ao mesmo tempo, apresentassem a maior correlação possível, em relação a qualquer outra combinação de números de onda. As duas regiões escolhidas, com absorções características do silicone, foram $775 - 791 \text{ cm}^{-1}$ e $1250 - 1270 \text{ cm}^{-1}$. Procurando distorcer minimamente o formato dos espectros nestas regiões, as absorbâncias foram manipuladas matematicamente de forma que uma região complementasse a outra frente ao coeficiente de correlação, em relação à concentração (na prática, a absorbância foi

considerada constante para os espectros 14-20, na região $775 - 791 \text{ cm}^{-1}$, e para os espectros 1-9, na região $1250 - 1270 \text{ cm}^{-1}$). A alta correlação nos números de onda restantes do espectro foi eliminada através da inserção de ruído aleatório em cada um dos espectros. A magnitude do ruído gerado foi no máximo 0,002 (valor absoluto). Em relação às absorbâncias, isto significa um ruído de aproximadamente 13 % em relação à mínima absorbância e 0,4 % em relação à máxima absorbância.

O passo seguinte à obtenção dos espectros simulados foi a definição das variáveis a serem testadas, e para isso foi utilizado um planejamento fatorial completo 2^4 [56], com as seguintes variáveis e níveis:

Número de gerações: 100 e 1000 gerações.

Tamanho da população: 30 e 100 cromossomos.

Probabilidade de mutação: 1 e 10 %.

Número de cromossomos elitistas: 1 e 6.

O planejamento fatorial completo, com os testes efetuados em ordem aleatória e em duplicata, é mostrado na tabela 15 (vide discussão). Para os testes efetuados, o algoritmo foi executado em um computador Pentium 4, 3.1 GHz, com 512 MB de memória RAM.

E.3. RESULTADOS E DISCUSSÃO

Os espectros do silicone e do sabonete, utilizados para simulação dos espectros utilizados no presente estudo são mostrados nas figuras 41 e 42.

Figura 41: Espectro infravermelho do silicone 60000

Figura 42: Espectro infravermelho típico de um sabonete comum

Em relação ao espectro do silicone^[45, 57], as principais bandas de absorção se encontram em 787 cm^{-1} (ν_s Si-CH₃), 1011 cm^{-1} (ν_{as} Si-O-Si) e 1259 cm^{-1} (δ_s CH₃). No caso do espectro do sabonete^[45], as bandas principais são em 1422 cm^{-1} (δ_s CH₃), 1442 cm^{-1} (ν_s CO₂⁻), 1557 cm^{-1} (ν_{as} CO₂⁻), 2849 cm^{-1} (ν_s CH₂) e 2918 cm^{-1} (ν_{as} CH₂).

A figura 43 mostra os 20 espectros simulados com a indicação das regiões onde a correlação em relação à concentração foi preservada (nas regiões restantes do espectro a correlação foi eliminada através da inserção de ruído aleatório). As figuras 44 e 45 mostram em detalhe as regiões indicadas. É fácil visualizar nestas figuras que a variação de absorbância nas regiões assinaladas é pequena em comparação com o ruído simulado e com as altas absorções de outras bandas do espectro, de forma que na escala da visualização completa do espectro não seria possível distinguir visualmente a linearidade implícita.

Figura 43: Espectros FTIR calculados, simulando 20 concentrações diferentes de silicone no sabonete, com indicação das regiões onde a linearidade foi preservada.

Figura 44: Detalhe da região próxima a 775 – 791 cm⁻¹, para os 20 espectros simulados.

Figura 45: Detalhe da região próxima a $1250 - 1270 \text{ cm}^{-1}$, para os 20 espectros simulados.

O algoritmo genético foi configurado com somente dois genes em cada cromossomo, de forma a procurar pelos dois números de onda que apresentassem melhor correlação linear múltipla em relação à concentração. Da forma como os espectros foram simulados, já se sabe que as melhores soluções possíveis são combinações dos números de onda das regiões $775 - 791 \text{ cm}^{-1}$ e $1250 - 1270 \text{ cm}^{-1}$. Para evitar que os dois números de onda fossem escolhidos na mesma região, limitou-se o espaço de busca de cada gene, de forma que o gene 1 somente poderia assumir valores no intervalo $700 - 1240 \text{ cm}^{-1}$ e o gene 2 somente poderia assumir valores no intervalo $1241 - 4000 \text{ cm}^{-1}$. Qualquer combinação entre números de onda das regiões $775 - 791 \text{ cm}^{-1}$ ou $1250 - 1270 \text{ cm}^{-1}$ resulta na solução ideal e possui o mesmo valor de correlação linear (numericamente igual a 0,992607). Este é o melhor coeficiente que se poderia obter nos testes realizados.

Para comparar o desempenho do algoritmo nos testes realizados, foi realizada análise posterior dos resultados gerados, de acordo com o seguinte sistema de pontos:

- Regra 1: Para a melhor solução obtida, cada número de onda identificado nas regiões de interesse ($775\text{-}791\text{ cm}^{-1}$ ou $1250\text{-}1270\text{ cm}^{-1}$) vale 10 pontos.

- Regra 2: Cada solução diferente contendo os dois números de onda nas regiões de interesse vale 1 ponto adicional.

Para exemplificar, consideremos a seguintes situações:

- Se nenhum número de onda for identificado nas regiões de interesse, o teste receberá pontuação igual a zero.

- Se somente um número de onda for identificado na região de interesse (considerando o melhor cromossomo), o teste receberá pontuação igual a 10 (regra 1).

- Se somente uma solução for identificada contendo os dois valores nas regiões de interesse, o teste receberá pontuação igual a 21 (regra 1 + regra 2).

Se tivermos 3 soluções diferentes onde os dois números de onda foram identificados nas regiões de interesse, o teste receberá pontuação igual a 23, e assim por diante.

A tabela 15 mostra os parâmetros utilizados nos testes, bem como os números de onda encontrados na melhor solução, o número de soluções diferentes contendo os dois números de onda na região de interesse e a pontuação final. Em destaque são mostrados os números de onda pertencentes às regiões ideais.

Tabela 15: Parâmetros e resultados do planejamento experimental para estudo de desempenho do algoritmo genético

#	Tamanho da população	Número de Gerações	Probab. de Mutação	Número de elitistas	Valor do Gene 1	Valor do Gene 2	Número de soluções ideais diferentes	Pontuação
1a	30	100	1	1	784	3681	0	10
2a	100	100	1	1	779	1264	2	22
3a	30	1000	1	1	787	3014	0	10
4a	100	1000	1	1	787	1264	1	21
5a	30	100	10	1	778	1720	0	10
6a	100	100	10	1	778	1260	2	22
7a	30	1000	10	1	781	1263	1	21
8a	100	1000	10	1	789	1251	2	22
9a	30	100	1	6	779	3977	0	10
10a	100	100	1	6	791	3936	0	10
11a	30	1000	1	6	776	2448	0	10
12a	100	1000	1	6	780	1250	8	28
13a	30	100	10	6	776	3324	0	10
14a	100	100	10	6	789	1264	5	25
15a	30	1000	10	6	785	1261	1	21
16a	100	1000	10	6	785	1261	9	29
1b	30	100	1	1	789	2504	0	10
2b	100	100	1	1	780	3577	0	10
3b	30	1000	1	1	786	3156	0	10
4b	100	1000	1	1	789	1251	1	21
5b	30	100	10	1	776	3275	0	10
6b	100	100	10	1	790	2918	0	10
7b	30	1000	10	1	779	2990	0	10
8b	100	1000	10	1	780	1265	2	22
9b	30	100	1	6	783	2865	0	10
10b	100	100	1	6	790	1263	3	23
11b	30	1000	1	6	778	1263	1	21
12b	100	1000	1	6	788	1261	8	28
13b	30	100	10	6	786	2948	0	10
14b	100	100	10	6	787	1263	7	27
15b	30	1000	10	6	790	1258	1	21
16b	100	1000	10	6	786	1258	8	28

Pela análise da tabela 15 já é possível observar que em todos os testes o algoritmo conseguiu identificar na região ideal o número de onda codificado pelo gene 1, e em muitos casos foi identificada a melhor solução possível, com os dois genes codificando valores nas regiões ideais. Também é importante notar que em alguns casos, mais que uma solução possível foi apresentada.

Conforme citado anteriormente, cada gene podia assumir valores dentro de um intervalo previamente definido. O gene 1 podia assumir 540 valores diferentes ($700\text{-}1240\text{ cm}^{-1}$), sendo que 16 eram valores ideais ($775\text{-}791\text{ cm}^{-1}$), enquanto o gene 2 podia assumir 2759 valores diferentes ($1241\text{-}4000\text{ cm}^{-1}$), sendo que 20 eram valores ideais ($1250\text{-}1270\text{ cm}^{-1}$). Com isto em vista, torna-se compreensível porque a incidência de soluções ideais no gene 1 é maior que no gene 2, já que o intervalo de valores possíveis é menor para o primeiro, considerando um número parecido de soluções ideais nos dois casos

A figura 46 sumariza a média das pontuações obtidas para as replicatas de cada teste, de acordo com os parâmetros de configuração que foram testados. Nesta figura, são representadas em amarelo as pontuações consideradas baixas, em laranja as pontuações consideradas médias e em vermelho as altas pontuações.

Geração	1000	10	16	16	21	21	28	22	29
	100	10	10	10	10	16	17	16	26
Elitistas >		1	6	1	6	1	6	1	6
Mutação >		1		10		1		10	
População >		30				100			

Figura 46: Sumário dos resultados do planejamento experimental da tabela 15. As cores fazem referência à média da pontuação recebida nos testes em duplicata (também indicada na figura): pontuação baixa (amarelo), média (laranja) e alta (vermelho).

Pela análise da figura 46, percebe-se que o tamanho da população afeta bastante o desempenho do algoritmo, com os melhores resultados alcançados utilizando-se populações maiores. O número de gerações também se mostra bastante significativo, e o uso de um número maior de gerações aumentou bastante a pontuação obtida. Tanto o aumento no número de cromossomos quanto o aumento no número de gerações acarreta um número maior de soluções sendo avaliadas, levando a melhores resultados, mas também aumenta o tempo despendido, de forma que vale ressaltar que é necessário procurar manter um compromisso entre o desempenho e o tempo necessário. No presente estudo o tempo não foi um problema, já que a função de avaliação é relativamente simples e o número de espectros utilizados foi de apenas vinte. O tempo total de processamento em cada teste não ultrapassou 73 segundos.

Em relação à probabilidade de mutação, a influência não foi tão acentuada quanto os parâmetros anteriores, sendo que é possível notar algum sinergismo somente em determinadas configurações, como por exemplo, no caso específico de níveis altos de elitismo e população,

aliados ao baixo número de gerações. Conclui-se que, dentro dos intervalos testados, este parâmetro não trouxe influência negativa, sendo que pode ajudar em casos específicos. É importante salientar que a mutação introduz valores que não estavam presentes na população original, o que, por um lado, aumenta a exploração do espaço amostral, e, por outro, expõe ao risco de eliminação de soluções potenciais antes que as mesmas sejam identificadas em combinações promissoras.

No caso do elitismo, nota-se que o número maior de cromossomos elitistas favoreceu, em alguns casos, a identificação do valor ideal no gene 2, mas além disso, aumentou consideravelmente o número de soluções ideais diferentes, quando utilizada com altos níveis de população e número de gerações. Observa-se na tabela 15 a ocorrência de até oito combinações ideais diferentes sugeridas pelo algoritmo utilizando-se níveis altos de elitismo. Em problemas reais, a identificação de diferentes soluções com correlações parecidas é bastante útil, principalmente para comparação, escolha e validação do método a ser implementado. A influência do elitismo pode ser explicada por uma característica particular deste algoritmo, que automaticamente causa mutação dos cromossomos idênticos. As altas taxas de elitismo acarretariam uma ocorrência maior de cromossomos redundantes, porém com a mutação forçada, o efeito geral é que existe um direcionamento adicional no espaço amostral, envolvendo os cromossomos de melhor resposta. O presente teste demonstra que este direcionamento é efetivo e leva a melhores resultados.

COMENTÁRIOS GERAIS E PERSPECTIVAS

A técnica de FTIR-ATR possui a grande vantagem de permitir a obtenção direta de espectros de substâncias com alta concentração de água ou solvente. Esta vantagem aumenta bastante o campo de aplicação da técnica na indústria e inclui a possibilidade de acompanhamento e estudo de reações químicas efetuadas diretamente sobre o acessório de diamante. Esta mesma vantagem, aliada aos recentes avanços quimiométricos, colaborou para que as técnicas baseadas no infravermelho próximo (NIRS) tenham ganhado rápido destaque na indústria. Em relação ao NIRS, o infravermelho médio possui a vantagem de prover informação qualitativa mais facilmente, com menor dependência de recursos quimiométricos avançados (embora os mesmos possam ser utilizados). Os estudos envolvendo estas duas técnicas e a comparação entre as mesmas poderão servir para um melhor direcionamento na escolha da técnica para determinada aplicação.

A análise de substâncias aquosas, tanto pela técnica de FTIR quanto NIRS, possui o problema referente à alta absorção de água em amplas regiões do espectro, o que impede ou dificulta a quantificação de analitos muito diluídos. A modificação química do cristal de ATR é um dos caminhos para aumentar a sensibilidade e seletividade da técnica de FTIR, buscando contornar o problema da alta absorção de água. Estudos mais profundos e propostas de novas opções para este problema devem contribuir para a maior popularização da técnica, bem como facilitar a quantificação de múltiplos analitos em uma mesma formulação (independente da concentração).

Em relação aos métodos quimiométricos, os chamados métodos de inteligência artificial, têm ganhado bastante destaque na literatura e têm alto potencial de aplicação na indústria, já que

a procura de métodos eficientes, com alto grau de automação e facilmente implementáveis é necessária, em virtude da alta competitividade, tanto em termos produtivos quanto em termos de qualidade. Neste sentido, o algoritmo genético possui a vantagem de ser facilmente utilizável, contribuindo bastante para o desenvolvimento automático supervisionado de soluções (onde o computador realiza as etapas mais demoradas e repetitivas trabalho, enquanto o cientista direciona e avalia as soluções apresentadas). O algoritmo genético ainda é pouco utilizado na indústria, porém a popularização poderá se acelerar com o surgimento de *softwares* prontos para uso e com o aumento do número de aplicações demonstradas na literatura, sob iniciativa dos centros de pesquisa acadêmica.

Em virtude da própria natureza do algoritmo genético, fortemente dependente da função de avaliação de resposta e do problema em si, ainda é difícil utilizar parâmetros gerais de configuração pré-determinados para problemas específicos, de forma que muitas vezes se torna necessária a realização de diversos testes antes se obter a configuração mais próxima possível da ideal. Um estudo sistemático das etapas, parâmetros de configuração e performance do algoritmo genético em relação à diversos tipos de problemas poderá propiciar não só um melhor entendimento e desenvolvimento das técnicas utilizadas em algoritmos genéticos, bem como acelerar a aplicação direta destes algoritmos em atividades de pesquisa e na indústria. O atual crescimento do número de publicações envolvendo utilização de algoritmos genéticos em todos os campos da ciência é um indício forte do potencial da técnica.

CONCLUSÕES

O bem sucedido desenvolvimento de uma variante do algoritmo genético, implementada em Excel, utilizando programação em Visual Basic (VBA) é demonstrado nesta dissertação. As etapas básicas do algoritmo genético clássico foram mantidas, porém algumas modificações foram feitas de forma a melhorar o desempenho em relação às aplicações testadas. As modificações relevantes incluem o uso da codificação real (em substituição à codificação binária), a implementação pseudo-paralela, que permite a quantificação de mais de um analito simultaneamente, e o uso de altas taxas de elitismo que, aliadas aos critérios de classificação e seleção utilizados, procura maximizar a diversidade de soluções e, ao mesmo tempo, mantém a busca orientada de soluções. O algoritmo desenvolvido também é bastante flexível, permitindo a fácil configuração de vários parâmetros, como tamanho da população inicial, número de genes no cromossomo, número de gerações, porcentagem de elitismo, probabilidade de mutação, número de populações, taxa de migração, etc.

A variante do algoritmo genético foi capaz de selecionar comprimentos de onda apropriados para a definição de métodos analíticos baseados em FTIR-ATR para as três aplicações de interesse na indústria cosmética investigadas: i) quantificação do teor ativo de lauril éter sulfato de amônio 28% (LESA), ii) quantificação da mistura de o-benzil p-cloro fenol (OBCP), o-fenil fenol (OPP) e etanol (matéria prima para a indústria de saneantes) e iii) quantificação de lauril éter sulfato de sódio (LESS) em *shampoos*.

As aplicações desenvolvidas são relevantes do ponto de vista da indústria, pois, uma vez divulgadas, poderão vir a substituir os métodos de análise clássicos (titulação potenciométrica ou em duas fases para os tensoativos e cromatografia gasosa para a mistura OPP/OBCP/etanol),

possuindo vantagens em termos de facilidade (não requer preparo prévio da amostra nem depende da experiência do analista) e tempo (espectros adquiridos em menos de 1 minuto), além de não acarretar impacto no meio ambiente (não gera resíduo e não requer manuseio ou dispêndio de reagentes). Para efeito de comparação, uma titulação potenciométrica de tensoativo, fazendo uso de eletrodo para surfactante, demora em torno de 20 min para ser realizada, já a titulação volumétrica em duas fases, embora mais rápida (5-10 min), tem a desvantagem de utilizar clorofórmio e de depender bastante da experiência do analista na visualização do ponto final. O método usual para quantificação da mistura OPP/OBCP/etanol, por cromatografia gasosa, é ainda mais demorado (3 h por amostra) pois cada corrida cromatográfica demora em torno de 30 min (3 injeções para a amostra e 3 para a curva de calibração).

As aplicações também serviram para estudo da performance do algoritmo genético durante o desenvolvimento do trabalho. No caso da quantificação de OPP/OBCP/etanol, a implementação paralela do algoritmo foi realizada com sucesso, considerando a qualidade do método desenvolvido e a facilidade com que a calibração foi realizada, já que as três curvas de regressão linear foram elaboradas simultaneamente pelo algoritmo. Os erros médios observados neste método para as amostras controle foram 0,72%, 1,15% e 1,80% para OPP, OBCP e etanol respectivamente. Também o método desenvolvido para *shampoo* apresentou bom desempenho, sendo que a exatidão e precisão do método foram demonstradas através da validação com amostras preparadas em laboratório (intervalo de confiança entre 95 – 105% e coeficiente de variação menor que 1%). O algoritmo também se mostrou bastante eficiente no estudo de desempenho efetuado com os espectros simulados de sabonete, já que conseguiu identificar com sucesso e em pouco tempo os números de onda ideais para a calibração.

Em suma, esta dissertação não só evidencia o valor do algoritmo genético – na variante proposta e com implementação em Excel+VBA – como ferramenta para acelerar a definição de novos métodos analíticos ou otimizá-los, como também demonstra que seu uso em combinação com a técnica de FTIR-ATR com acessório de diamante conduz a excelentes alternativas para desenvolvimento de métodos rápidos e precisos na indústria cosmética e de saneantes.

REFERÊNCIAS BIBLIOGRÁFICAS

- [1] Skoog, A. Douglas; Leary, James J.; Principles of instrumental analysis; Saunders College Publishing; 4º Ed.; 1992.
- [2] CU Boulder Department of Chemistry and Biochemistry; Handbook for Organic Chemistry Lab; University of Colorado, 8º Ed.
- [3] The Internet Journal of Vibrational Spectroscopy - <http://www.ijvs.com> (visitado em 18/09/06).
- [4] Handbook of instrumental techniques for analytical chemistry; Settle, Frank A.; Upper Saddle River, NJ : Prentice Hall PTR, ©1997.
- [5] Fahrenfort, J.; Spectrochimica Acta; 1961; Vol. 17; 698-709.
- [6] Perkin Elmer; Spectrum BX FTIR User's Guide.
- [7] Lin, Rong; L. White, Robert; Anal. Chem., 1994, 66, 2976-2980.
- [8] King, P.L., Ramsey, M.S., McMillan, P.F., & Swayze, G.A., 2004. Laboratory Fourier Transform Infrared Spectroscopy Methods for Geologic Samples. In: King, P.L., Ramsey, M. & Swayze, G. (eds.), 2004. Infrared Spectroscopy in Geochemistry, Exploration Geochemistry and Remote Sensing. Mineralogical Association of Canada, Short Course 33, 57-91.
- [9] Cartwright, H. M.; Applications of Artificial Intelligence in Chemistry, Oxford Science Publications, New York (1995).
- [10] Leardi, R.; J. Chemom. (1994), 8, 65.
- [11] Jouan - Rimbaud, D.; Massart, D. L.; Leardi, R.; Noord, O. E.; Anal. Chem. (1995), 67, 4295.
- [12] Sharaf, M. A.; Illman, D. L.; Kowalski, B. R.; Chemometrics, John Wiley & Sons, New York (1986).
- [13] Lucasius, C. B. and Kateman, G.; Chemom. Intell. Lab. Syst. (1993), 19, 1.
- [14] Costa Filho, P. A. and Poppi, R. J.; Química Nova (1999), 22, 3.
- [15] Saldanha, T. C. B.; Araújo, M. C. U.; Neto, B. B.; Química Nova (1999), 22, 6.
- [16] Leardi, R.; Boggia, R.; Terrile, M.; J. Chemom. (1992), 6, 267.
- [17] Arcos, M. J.; Ortiz, M. C.; Villahoz, B.; Sarabia, L. A.; Anal. Chim. Acta (1997), 339, 63.
- [18] Evonet Online Tutorial - <http://evonet.lri.fr>, (visitado em 18/09/06).

- [19] Almeida, Pedro Henrique de; Furtado, João Carlos; Carvalho, Claudia Wollmann; Ferrão, Marco Flores; Molz, Rolf Fredi; Bassani, Irionson Antonio; Huning, Silvio Luiz; Pesquisa Operacional, v. 23, n.1, 189-207 - 2003.
- [20] Shyh-Chang Lin, W. F. Punch III, E. D. Goodman, Parallel & Distributed Processing, Dallas TX, Oct 1994.
- [21] Lacerda, E.G.M e Carvalho, A.C.P.L. (1999) "Introdução aos algoritmos genéticos", In: Sistemas inteligentes: aplicações a recursos hídricos e ciências ambientais. Editado por Galvão, C.O., Valença, M.J.S. Ed. Universidade/UFRGS: Associação Brasileira de Recursos Hídricos. p. 99-150. (Coleção ABRH de Recursos Hídricos; 7); Porto Alegre.
- [22] Whitley, Darrel; A genetic algorithm tutorial; Technical Report CS-93-103; Department of Computer Science; Colorado State University; August 1993; 10.
- [23] Haaland, D. M.; Thoms, E. V.; Anal. Chem. 1988, 60, 1193
- [24] U. Depczynski, V.J. Frost, K. Molt, Anal. Chim. Acta, 420 (2000) 217-227.
- [25] Riccardo leardi, Amparo Lupiães González, Chemom. Intell. Lab. System, 41 (1998), 195-207.
- [26] B. K. Lavine, C. E. Davidson, A. J. Moores, Chemom. Intell. Lab. System, 60 (2002), 161-171
- [27] http://www.nli-ltd.com/products/genetic_algorithms/demos.htm (visitado em 18/09/06).
- [28] <http://www.solver.com> (visitado em 18/09/06).
- [29] http://www.piblu.com/products/optworks_ex.html (visitado em 18/09/06).
- [30] <http://www.palisade-europe.com/evolver/> (visitado em 18/09/06).
- [31] <http://www.mathworks.com> (visitado em 18/09/06).
- [32] <http://www.optiwater.com/> (visitado em 19/09/06).
- [33] <http://webservice.hhsoft.com> (visitado em 19/09/06).
- [34] <http://www-cse.uta.edu/%7Ecook/ai1/lectures/applets/gatasp/TSP.html> (visitado em 19/09/06).
- [35] <http://aaai.org/AITopics/html/tsp.html> (visitado em 19/09/06).
- [36] <http://www.tsp.gatech.edu/> (visitado em 19/09/06).
- [37] <http://www.mdli.com/products/predictive/qsar/index.jsp> (visitado em 19/09/06).
- [38] Ferreira, Márcia Miguel Castro, Montanari, Carlos Alberto and Gaudio, Anderson Coser Seleção de variáveis em QSAR. Quím. Nova, Maio 2002, vol.25, no.3, p.439-448.

- [39] <http://www.gepsoft.com/gxpt.htm> (visitado em 19/09/06).
- [40] Martens, H.; Naes, T.; "Multivariate Calibration"; John Wiley & Sons, New York, 1989
- [41] Lange, K. Robert; Detergents and Cleaners - A Handbook for Formulators; Carl Hanser Verlag, Munich, Vienna, New York; 1994.
- [42] Cross, John; Anionic Surfactants - Analytical Chemistry, Surfactant Science Series, Vol. 73; 2º Ed.; Marcel Dekker, Inc.; New York, 1998.
- [43] H. Holness and W. R. Stone; Analyst 82:166 (1957).
- [44] N. Buschmann and R. Schultz, Jorn. Com. Esp. Deterg. 23:323 (1992).
- [45] O. Hummel, Dietrich; Analysis of Surfactants - Atlas of FTIR Spectra with interpretations
- [46] Stuart, Barbara H.; Infrared Spectroscopy: Fundamentals and Applications; Wiley Interscience; 2004.
- [47] Dowicide® 1 Antimicrobial; Product Information; Dow.
- [48] Paulus, Wilfried; Microbiocides for the protection of materials: A handbook; 1º Edition; London; 1993.
- [49] Yann Lai, Kuo; Liquid Detergents, Surfactant Science Series, Vol. 67; 1º Ed.; Marcel Dekker, Inc.; New York; 1997.
- [50] Sagarin, Edward; Cosmetics - Science and Technology, Interscience Publishers, Inc; New York; 1957.
- [51] Guidance for Industry - Analytical Procedures and methods validation, Agosto 2002, do site <http://www.fda.gov/cder/Guidance> (visitado em 18/10/06).
- [52] Guideline for Industry - Text on validation of analytical procedures, ICH Q2A, Março 1995, do site <http://www.fda.gov/cder/Guidance> (visitado em 18/10/06).
- [53] Guideline for Industry - Validation of analytical procedures: Methodology, ICH Q2B, Novembro 1996, do site <http://www.fda.gov/cder/Guidance> (visitado em 18/10/06).
- [54] Schueller, Randy; Romanowski, Perry; Conditioning Agents for Hair and Skin, Cosmetic Science and Technology Series – Volume 21; Marcel Kekker, Inc; New York; 1999.
- [55] Silicone Science On line, http://www.silicones-science.com/chemistry_pdms.html, visitado em 29/10/2006
- [56] Barros Neto, Benício de; Spacino Scarminio, Ieda; Bruns, Roy Eduard; Como fazer experimentos – Pesquisa e desenvolvimento na ciência e na indústria; 2º Edição; Editora Unicamp, Campinas; 2003.

- [57] B. Colthup, Norman; H. Daly, Lawrence; E. Wiberley, Stephen; Introduction to infrared and raman spectroscopy; Third Edition; Academic Press, New York.