ARTIGO

SOBRE O USO DE MÉTODOS QUIMIOMÉTRICOS EM QUÍMICA COMBINATÓRIA

Andrei Leitão, Carlos A. Montanari * e Claudio L. Donnici
Departamento de Química-ICEx - Universidade Federal de Minas Gerais

Recebido em 15/1/99; aceito em 16/8/99

THE USE OF CHEMOMETRIC METHODS ON COMBINATORIAL CHEMISTRY. Combinatorial chemistry has emerged as a tool to circumvent a major problem of pharmaceutical industries to discover new lead compounds. A rapid and massive evaluation of a myriad of newly synthesised compounds can be carried out. Combinatorial synthesis leads to high throughput screening *en masse* towards another myriad of biological targets. The design of a set of compounds based upon combinatorial chemistry may be envisaged by using of QSPR-SIMCA and QSAR-SIMCA as tools for classification purposes. This work deals with the definition and establishment of a spanned substituent space (*SSS*) that reduces the analogue numbers with no exclusion of global content. The chemical diversity may be set properly within a specified pharmacological field. This allows a better use of its potentiality without loosing information.

Keywords: combinatorial chemistry; chemical diversity; QSPR-SIMCA.

INTRODUÇÃO

Um problema comum aos nossos dias refere-se ao acesso, manipulação e quantidade de informações disponíveis. Não obstante, mais importante do que o seu conhecimento está a necessidade de sua organização racional[1]. No que diz respeito à química medicinal, existem inúmeros bancos de dados listando os mais diferentes tipos de substâncias químicas e suas atividades farmacológicas.

No início deste século os métodos de descobrimento de novas drogas eram empíricos ou estavam quase subjugados ao acaso. O sucesso da equação de Hammett, entretanto, possibilitou a racionalização química de pequenas regiões subestruturais que permitiu o aparecimento da QSAR na década de 1960[2]. Desde então, a busca reducionista de informações capazes de descreverem biomacromoléculas tornou-se atividade comum em inúmeros centros de pesquisas em todo o mundo. Mais recentemente, contudo, um aparente "retorno ao passado" tem sido postulado. O advento da química combinatória trouxe um novo avanço, desta vez não somente na busca e identificação em massa de novas drogas como também na síntese combinatória[3]. O principal objetivo da síntese combinatória e do ensaio em batelada é identificar compostos biologicamente ativos através do uso de um grande número de substâncias sintetizadas e depois realizar o ensaio biológico "simultaneamente". Os dados obtidos passam a constituir grandes "coleções " de compostos (compound libraries) que contêm informações sobre a natureza biológica de seus integrantes.

A diversidade química constitui um excelente espectro do espaço multivariado de trabalho que pode, teoricamente, facilitar a descoberta de novas drogas. A transposição daquilo que se conhece como QSAR clássica para a química combinatória pode ser estabelecida dentro da dicotomia *diversidade na bioatividade = diversidade paramétrica.* A atividade do composto matriz pode ser identificada através de descritores físico-químicos. Mas, o estabelecimento de coleções de compostos (com até centenas de milhares) também precisa de descritores das coleções.

QUÍMICA COMBINATÓRIA

A química combinatória é uma das novas metodologias desenvolvidas por acadêmicos e pesquisadores de indústrias farmacêutica, agroquímica e biotecnológica para reduzir o tempo e o custo associados com a produção, mercado e introdução de novas drogas competitivas. De uma maneira simples, os cientistas usam a química combinatória para criar enormes populações (coleções) de moléculas - ou bibliotecas, que podem ser ensaiadas eficientemente *en masse.* Pela produção de compostos diversos em bibliotecas, há um aumento da probabilidade de encontrar novos compostos de valor terapêutico e comercial. O campo representa a convergência da química e biologia.

Durante a última década, uma nova fonte de compostos apareceu: aqueles obtidos a partir de geração química rápida (e algumas vezes também biológica) - formando as coleções de compostos[4-8]. Esse método de gerar novos compostos juntamente com a capacidade de ensaiá-los biologicamente, representa uma mudança importante no paradigma tradicional de geração e otimização de novas entidades químicas (NCEs)[9]. O estágio inicial de desenvolvimento para geração rápida deu-se através de um grande número de peptídios[10-12]; entretanto, muitos pesquisadores, atualmente, estão procurando desenvolver método de geração para compostos não-peptídicos.

Os métodos de geração de coleções de compostos diferem consideravelmente nos tipos e números de compostos preparados (dezenas até dezenas de milhares), e, se os compostos são obtidos como entidades estruturais simples ou como misturas. Associado a isso está o fato de como os compostos são preparados, em fase sólida ou solução, onde vantagens e desvantagens podem existir[13-29].

Os primeiros exemplos de geração diversa de moléculas de baixo peso molecular não-poliméricas são benzodiazepinas[30-32] e hidantoínas[32], em suporte sólido.

Uma "coleção universal" envolve o conceito de que qualquer macromolécula biológica (receptor, enzima, anticorpo, etc.) reconhece substratos através de interações físico-químicas precisas. Ao nível fundamental, essas interações podem ser divididas em diferentes parâmetros ou dimensões tais como tamanho, capacidade de formar ligações de hidrogênio, interações hidrofóbicas, etc..

O impacto da síntese combinatória está no descobrimento da substância matriz e, posteriormente, na otimização da substância matriz que propiciará a seleção da droga potencial[3].

A indústria farmacêutica multinacional possui, em geral, aquilo que se conhece por "coleção corporativa de compostos".

montana@dedalus.lcc.ufmg.br, http://www.qui.ufmg.br/~nequim

Essas coleções podem conter até 500.000 (ou mais) estruturas individuais prontas para serem ensaiadas. Portanto, elas não são de domínio público, mas, certamente, constituem fontes ricas em compostos matrizes. Essas coleções apresentam uma enorme diversidade química mas, nem sempre, é possível estabelecer se há falta de diversidade química ou excesso. Não obstante, um aspecto implícito ao conhecimento da diversidade química está associado ao modo de ação que os membros da coleção precisam demonstrar.

A Figura 1 mostra o impacto que a química combinatória tem demonstrado em química medicinal.

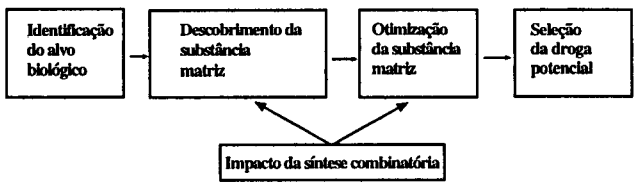


Figura 1. Impacto da síntese combinatória

Há, basicamente, dois níveis onde a química combinatória está delineada: o primeiro diz respeito à descoberta de novas substâncias químicas com atividade farmacológica e, o segundo, é estabelecido pela necessidade de otimização da potência.

O primeiro nível não pode ser aplicado em estudos de QSAR-SIMCA ou QSPR-SIMCA porque nesse nível a substância matriz está em fase de ser descoberta e não desenvolvida. O objetivo principal desse nível é o de estudar qualquer substância química existente frente a qualquer sistema biológico passível de ensaio. Trata-se, portanto, de uma verdadeira busca aleatória de novas substâncias químicas com atividade biológica em novos alvos ou alvos já conhecidos ou, substâncias químicas já disponíveis em novos e/ou conhecidos alvos.

No segundo, entretanto, o composto matriz já é conhecido; sua atividade farmacológica já foi estabelecida. O objetivo é otimizar sua potência, estudar o mecanismo de ação, guiar a rota sintética, etc. Nesse aspecto, as seguintes questões podem ser formuladas: qual é o grupo farmacofórico? Onde modificar? Quantas posições a alterar simultaneamente? Quantos substituintes devem ser usados?, etc.

A seleção do alvo biológico estabelece, previamente, o questionamento de quanta diversidade química é realmente necessária para definir-se apropriadamente uma coleção de compostos. Por isso, a quantificação da diversidade química é de fundamental importância e tem, certamente, um papel importantíssimo na academia que pode responder aos propósitos do limite espaço-paramétrico necessário para a perfeita compreensão do modo de ação. Portanto, as seguintes questões precisam ser respondidas durante a execução do projeto: (i) quanta diversidade química está perdida? (ii) quanta diversidade química é necessária?

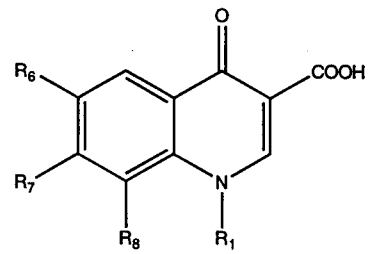
A quantificação pode ser estabelecida através do método de trabalho que vem sendo empregado em nosso grupo e também em outras instituições com projetos nessa área.

Em primeiro lugar, há que se calcular e medir as propriedades moleculares, tais como índices topológicos, presença-ausência de grupos funcionais, lipofilia, etc. Em segundo lugar, uma análise estatística pode ser realizada com o objetivo do entendimento da descrição estabelecida.

Com o objetivo de explicitar os fundamentos desse modelo, considere os extremos de uma coleção de hexapeptídios contendo as seguintes características: 21 aminoácidos essenciais são empregados na síntese combinatória de hexapeptídios que tenham como "cabeça" o grupo acetato e como cauda o grupo amino. Uma simples análise combinatória dessas poucas características resultará na síntese de 64 milhões de compostos! Admitamos que a biodisponibilidade seja um fator importante para a descrição da atividade farmacológica de interesse e que o coeficiente de partição, log P, seja utilizado como parâmetro para a sua estimativa. Esta assunção é bastante razoável haja vista que a droga precisa alcançar o sítio receptor e para isso "viaja" por caminhos lipofílicos-hidrofílicos" aleatórios.

Analisemos, portanto, dois extremos dessa enorme coleção de compostos: um dos possíveis hexapeptídios será Ac-Phe-Phe-Phe-Phe-Phe-Phe-NH_2 e seu coeficiente de partição calculado é CLOGP = 5,5. No outro extremo teríamos o seguinte hexapeptídio: Ac-Arg-Arg-Arg-Arg-Arg-Arg-NH_2, com CLOGP = -13. Admitindo-se a possibilidade de protonação quando em pH fisiológico, o CLOGD muda para -37! Então, parece natural assumir a questão não apenas do ponto de vista filosófico mas, muito certamente, deve-se ater ao seu aspecto prático. E, nesse caso, parece pouco provável que o espaço de trabalho lipofílico a ser estudado precisará conter tal variância.

Em linguagem cotidiana tentemos analisar o que isso representa quando quisermos definir o que tem sido a química do século XX, respondendo à seguinte questão: Quantos livros podemos ler? E, quais? Usando uma citação do "Chemical Abstract": em 1994, foram publicados 653.055 resumos, o que resulta em 1789/dia! Perca 10 dias e você estará atrás de 17.890 artigos ou patentes! O problema, então, não é somente obter informações, mas como organizá-las.



Quimicamente, a síntese de ácidos quinolinocarboxílicos como agentes antibacterianos contendo apenas quatro posições de substituição e 166 substituintes que descrevem apenas um pequeno espaço químico-diverso, geraria $166^4 = 7{,}6.10^8$ moléculas. Entretanto, se isto for fantasioso, poderíamos usar o corolário que estabelece 5 substituintes para cada parâmetro e, dessa forma ainda precisaríamos de 625 moléculas!

O que isso representa verdadeiramente pode ser visto através de um outro exemplo mais prático, que considera drogas que atuam no sistema nervoso central, CNS[34]. O gráfico da Figura 2 mostra que há uma pequena distribuição de um grande número de drogas com log P = 2-3. Nesse caso, portanto, falar-se em diversidade química fora do espaço de trabalho lipofílico delineado parece ser redundante!

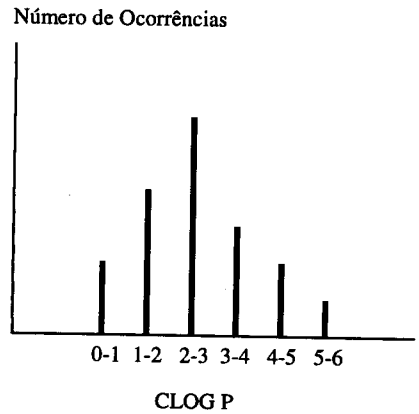
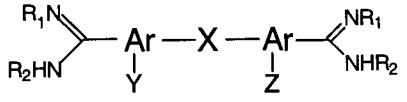


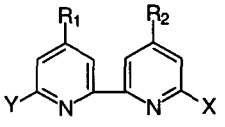
Figura 2. Drogas que atuam no Sistema Nervoso Central: prevalência de um grande número de drogas com lipofilia delimitada.

A questão primordial que pode ser colocada agora é: como resolver esse problema? A resposta pode estar na *representação*. Por representação entende-se o planejamento da série dentro do espaço de trabalho do substituinte, SSS[35].

A nossa capacidade de realizar síntese combinatória através de sintetizadores e depois realizar os ensaios biológicos em massa está longe de ser uma realidade. Além disso, a questão que permeia esta investigação não é somente esta. Qualquer químico sabe que propriedades estruturais e físico-químicas tornam-se isostéricas e o mesmo ocorre com as propriedades biológicas que se tornam bioisostéricas. Então, com o objetivo de tentar solucionar esse problema, pelo menos parcialmente, exploramos um espaço de trabalho multiparamétrico. O planejamento de coleções que orientem os grupos responsáveis pelas interações específicas descritas nesse espaço tem o objetivo de explorar tamanho, formas e volumes (topológicos), além de parâmetros físico-químicos através de modificações químicas simples. Uma sub-coleção representando a coleção universal planejada para explorar o espaço multiparamétrico e, consequentemente, identificar candidatos com atividades farmacológicas antileishmaniose, antitumoral, antibacteriana e antifúngica foi estabelecida para nortear as sínteses de compostos bipiridínicos[36] e bisamidínicos[37], Figura 3. Esses compostos constituem o fundamento desta investigação. Com base em suas estruturas químicas, atividades farmacológicas e sistemas biológicos de estudo, os substituintes foram selecionados. Para as bisamidinas, R_1, R_2 e Y representam substituintes escolhidos a partir da seleção do banco de 59 substituintes, objeto deste estudo; X representa o grupo bisamidínico e seus análogos. Para as bipiridinas, R_1 e R_2 não fazem parte, necessariamente do banco de dados, já que essas posições não estão sendo otimizadas; R_3-R_7, entretanto, o fazem.



R_1 = H, alq., etc....
R_2 = H, alq., etc...
Y = Z = H, Me, MeO, Br, NO_2, etc..
X = $-O(CH_2)_nO-$, $-S(CH_2)_nS-$, $-N=N-\overset{H}{N}-$,etc



$R_1 = R_2$ = H, $SO_3^-Na^+$, $CO_2^-Na^+$,NO_2, etc...
Y = X = $CSNR_3R_4$, $CSSR_5$, CH_2SR_6, CH_2CSSR_7 , etc...

Figura 3. *Estruturas matrizes estudadas. (Veja o texto para explicação).*

Para isso, métodos quimiométricos foram empregados. Os estudos das relações quantitativas entre estrutura química e propriedade físico-química, QSPR, constituem uma ferramenta poderosa no delineamento e estabelecimento de "sub-coleções". Esse método foi estabelecido por ser um método computacional rápido e eficiente.

MÉTODO

Inicialmente, realizamos a seleção do espaço de trabalho necessário para produzir uma diversidade suficiente para as diferentes substituições a realizar. Um banco de dados contendo 94 descritores químicos e 59 substituintes[38-41] foi usado para a seleção inicial da coleção de compostos.

O método constitui em, inicialmente, uma seleção por reconhecimento de padrões para a classificação dos substituintes em "famílias" ou grupos químicos similares, das quais "representantes" são escolhidos. Como esses representantes podem ser em menor número, uma sub-série ou sub-coleção pode ser criada.

A junção de métodos quimiométricos e modelagem molecular empregados neste trabalho, constitui, dentro de nossa realidade, talvez, uma alternativa mais eficiente para o estabelecimento de critérios mais práticos e exequíveis. Entretanto, essa simplicidade da química combinatória não pode ser entendida como o ponto final. Ainda mais pelo fato de que sua aplicação pode ter diferentes formas, cada uma necessitando de um sistema complexo de técnicas de síntese orgânica clássica, estratégias de planejamento racional de drogas, robotização e gerenciamento de informações científicas.

A "química combinatória virtual" pode conter os seguintes itens:

(i) identificação do alvo biológico. A aplicação do método precisa, necessariamente, do alvo para garantir o processo de otimização e o espaço farmacológico;
(ii) definição da estrutura molecular e geração do composto matriz: *De novo*, Busca 3D ou geração 3D são fontes comuns na obtenção das estruturas moleculares;
(iii) definição dos grupos químicos (substituintes) que serão usados para a modificação molecular da estrutura farmacofórica previamente definida. Delineamento do espaço físico-químico;
(iv) modelagem molecular;
(v) cálculo de parâmetros físico-químicos. Cálculo de parâmetros tridimensionais.
(vi) Análise quimiométrica: (a) análise de componentes principais, PCA; (b) KNN-QSPR; (c) SIMCA-QSPR, SIMCA-QSAR;
(vii) Síntese e ensaio biológico de análogos representativos;
(viii) reciclagem do processo.

O banco de dados constituído de 94 descritores e 59 substituintes foi estudado através dos programas contidos no pacote ARTHUR[42], operando em um microcomputador PentiumII, 166 MHz. Inicialmente, uma matriz 59x94 foi construída usando o programa ENTER; os descritores foram escalonados (média/desvio padrão), usando o programa SCAL. O escalonamento é necessário para evitar problemas decorrentes das diferentes unidades utilizadas na obtenção dos diferentes descritores físico-químicos; o banco de dados escalonado foi submetido a uma análise de componentes principais usando o programa PCA; os resultados, descritos em três componentes principais PC1, PC2 e PC3, foram estudados graficamente através das seguintes relações: PC1xPC2, PC1xPC3 e PC2xPC3. Uma análise prévia foi realizada para identificar as eventuais separações em classes ou grupamentos (famílias de compostos). Os grupamentos foram então numerados de 1 a 4 e submetidos ao programa SIMCA, que classificou-os de acordo com as propostas obtidas das componentes principais.

Principais vantagens

A rotina assim estabelecida proporciona várias vantagens: (i) racionaliza o número de análogos; (ii) estabelece critérios paramétricos de reconhecimento molecular; (iii) define similaridade química; (iv) estabelece relações bioisostéricas; (v) na reciclagem do método, o banco de dados já está pronto.

Seleção de variáveis

O banco de dados constituído de 94 descritores físico-químicos para 59 substituintes diferentes foi submetido, inicialmente, a uma análise de PCA. O gráfico da Figura 4 mostra os resultados da PC1 versus PC2 para os autovetores dos 94

descritores, enquanto que o gráfico da Figura 5 mostra os autovalores para a seleção dos 59 substituintes.

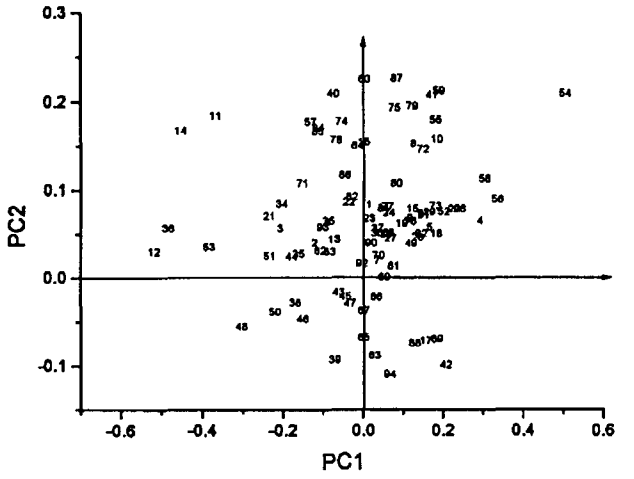


Figura 4. PC2 versus PC1 para os 94 descritores físico-químicos constituintes do banco de dados em estudo.

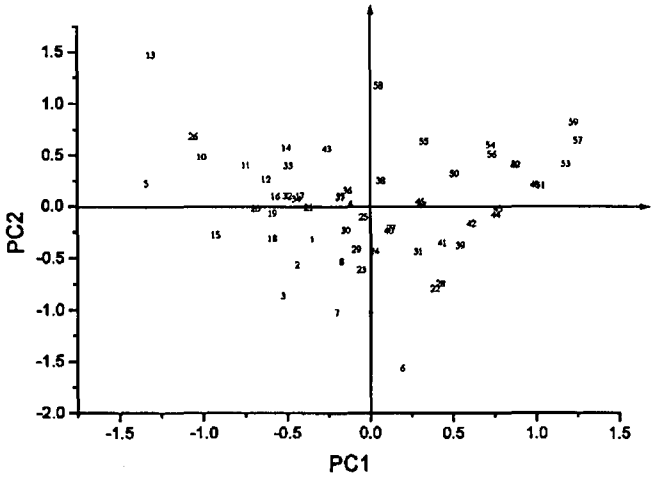


Figura 5. PC2 versus PC1 para os 59 substituintes constituintes do banco de dados em estudo.

A Figura 4 representa a distribuição paramétrico-espacial. Inúmeros descritores têm, essencialmente, o mesmo significado físico-químico. Por exemplo, aqueles descritores com valores de PC1 próximo a zero não são significativos nesse nível e, portanto, representam propriedades similares. Análise semelhante pode ser realizada para aqueles membros descritos em PC2. Há, portanto, clara evidência sobre a distinção inequívoca entre esses descritores e os demais que são adequadamente separados por força dessas componentes principais.

A importância dessa distinção encontra-se no fato de serem os descritores físico-químicos oriundos de diferentes características representadas através da natureza intrínseca de cada substituinte. Uma importante aplicação desse conceito está na utilização dos resultados em similaridade química: diferentes estruturas moleculares com propriedades físico-químicas semelhantes traduzem o conceito de isosterismo. Essa aparente dicotomia constitui a interface da distinção entre o que é estrutura química e propriedade físico-química. Por exemplo, a ligação de hidrogênio pode ser encontrada tanto em substituintes OH quanto NH, mas de suas estruturas intrínsecas apenas inferências paramétricas podem ser estabelecidas[9].

As três primeiras componentes principais descrevem 60% da variância total. Os autovetores podem ser encontrados nas equações abaixo e a Tabela 1 identifica os parâmetros significativos. Os termos incluídos nessas equações correspondem às magnitudes dos descritores, que representam a melhor variância do sistema. Os critérios utilizados para o estabelecimento do limite de exclusão não são óbvios. No presente caso, os descritores com valores menores que 0,2 foram excluídos.

PC1 ~ - 0.207(FARR) + 0.294(FALR) - 0.367(VW) - 0.519(VTSAR) - 0.453(MR) + 0.228(SPHL) - 0.345(SMSW) + 0.242(SPE) - 0.203(R) - 0.484(RNEW) + 0,208(E) + 0.303(LAMDAL) - 0.220(HD) - 0.233(HB) + 0.202(X0AR) - 0.382(X0VAR) + 0,508(X1AR) + 0,305(FASAL)

PC2 ~ 0.208(SSAR) + 0.207(SSAL) + 0.208(X1AR) + 0.211(W2) + 0.224(W3) + 0.225(S1)

PC3 ~ - 0.216(LAMDAL) + 0.208(HA) + 0.232(C) - 0.21785(E2) + 0.224(H1)

Equações para a identificação das principais variáveis descritoras do espaço inicial de 94 descritores físico-químicos, reduzidos para 29.

Tabela 1. Seleção de parâmetros físico-químicos usados no estudo.

PC1		Parâmetro
03 FAAR	Constante fragmental lipofílica para aromáticos	Lipofílico
04 FALR	Constante fragmental lipofílica para alifáticos	Lipofílico
11 VW	Volume de van der Waals	Estereoquímico
12 VTSAR	Volume fragmental de substituintes aromáticos	Estereoquímico
14 MR	Refratividade molar	Estereoquímico/ Polarizabilidade
20 SPHL	Constante de Hammett para substituintes na posição *para-*	Eletrônico
21 SMSW	Constante de Hammett para substituintes na posição *meta-*	Eletrônico
26 SPE	Constante de Hammett para substituintes na posição *para-*	Eletrônico
34 R	Parâmetro de ressonância	Eletrônico
36 RNEW	Parâmetro de ressonância corrigido	Eletrônico
42 E	Parâmetro eletrônico de substituinte baseado nas diferenças de OM ocupados	Eletrônico
48 LAMDAL	Constante lipofóbica para substituinte alifáticos	Eletrônico
50 HD	Grupo doador de hidrogênio (0 ou 1)	Eletrônico
51 HB	Parâmetro de ligação de hidrogênio = número de átomos capazes de formar ligação de hidrogênio	Eletrônico
52 X0AR	Conectividade molecular de order zero para grupos aromáticos	Topológico
53 X0VAR	Conectividade molecular de valência ordem zero para grupos aromáticos	Topológico
54 X1AR	Conectividade molecular de primeira ordem para grupos aromáticos	Topológico

Tabela 2. Matrizes de correlações das variáveis significativas de PC1.

Matriz de Correlação das Variáveis Significativas de PC1 (CF_3, CH_3, CCH, CH_2COOH, $CH_2CH_2C_6H_5$)

	VTSAR	SNSW	SPE	R	E	LAMDAL	HD	HB	XOAR	XOVAR
FALR	0,542	0,521								
VW		0,970				0,886			0,948	
VTSAR										0,624
MR			0,565		0,975	0,797				
SPHL			0,747	0,545				0,632		
SMSW						0,870			0,976	
SPE					0,713					
RNEW						0,598				0,864
E							0,763			
LAMDAL									0,936	

Matriz de Correlação das Variáveis Significativas de PC1 (F, SO_2CF_3, $NHCONH_2$, $NHCO_3$, $CH_2CH_2C_6H_5$)

	MR	SMSW	SPE	R	RNEW	E	LAMDAL	HD	XVAR	XOVAR
FALR		0,564							0,621	0,832
VW	0,602									
VTSAR				0,671						0,634
MR			0,838		0,563	0,978		0,752		
SMSW									0,972	
SPE					0,765	0,922		0,753		
R							0,587			
RNEW						0,681		0,718		0,782
E								0,766		

Pode-se observar, claramente, que apenas alguns poucos descritores físico-químicos foram suficientes para descrever toda a variância, sem prejuízo do espaço de trabalho previamente definido. Entretanto, não pode ser esquecido que alguns dos descritores ainda estão classificados em grupos paramétricos globais: lipofílico, eletrônico e topológico. Isso significa que a redução para 29 variáveis ainda não está finalizada. Dependendo dos substituintes selecionados, a partir dessa tabela reduzida, correlações elevadas podem ser encontradas, caso os substituintes não sejam selecionados adequadamente. Portanto, a seleção de substituintes precisa ser realizada com critério e dela uma nova análise estatística estabelecerá possíveis multicolinearidades entre os mesmos. A Tabela 2 mostra os descritores multicorrelacionados para os seguintes membros: (i) CF_3, CH_3, CCH, CH_2CO_2H, $CH_2CH_2C_6H_5$ e (ii) F, SO_2CF_3, $NHCONH_2$, $NHCO_2$, $CH_2CH_2C_6H_5$. Os espaços em branco demonstram falta de colinearidade. A Tabela 3 identifica os substituintes em estudo, classificados nas PCs. Esses substituintes foram selecionados através dos seguintes critérios, usando-se para isso os autovalores da Figura 5: (i) um representante de cada grupamento (quadrantes: +,+; +,-; -,-; -,+) mais um do grupamento central; (ii) viabilidade sintética.

Os resultados das PCAs foram analisados pelo método SIMCA[42, 43] com o objetivo de classificar os substituintes em famílias, os resultados chegaram a 97% para 4 categorias diferentes, o que demonstra muito boa classificação desses substituintes. Essas 4 categorias foram selecionadas com base nos autovalores da Figura 5, de acordo com os quadrantes já mencionados. Essa informação precisa ser, inicialmente, fornecida como dado de entrada para o método SIMCA.

Os dados resultantes (a série de treinamento) podem ser usados para derivar modelos baseados na estrutura que podem, então, ser usados para classificar novos compostos de uma classe desconhecida (da série de teste). Além disso eles também servem para classificar compostos similares e/ou dissimilares, naquilo que está sendo postulado como QSPR-SIMCA. Ou seja, compostos pertencentes a uma mesma família com propriedades físico-químicas multicorrelacionadas são similares enquanto os demais são dissimilares.

Tabela 3. Substituintes usados no estudo.

01	Br	21	$OCONH_2$	41	$N(CH_3)$
02	Cl	22	CH_3	42	C_3H_5
03	F	23	OCH_3	43	$COOC_2H_5$
04	I	24	CH_2OH	44	C_3H_7
05	NO_2	25	$NHCONH_2$	45	$CH(CH_3)$
06	H	26	SO_2CH_3	46	OC_3H_7
07	OH	27	SCH_3	47	$OCH_9CH_3)_2$
08	SH	28	$NHCH_3$	48	C_4H_9
09	NH_2	29	C_2H	49	$C(CH_3)_3$
10	SO_2NH_2	30	CH_2CN	50	OC_4H_9
11	CF_3	31	C_2H_3	51	NHC_4H_9
12	OCF_3	32	$COCH_3$	52	$N(C_2H_5)_2$
13	SO_2CF_3	33	$COOCH_3$	53	C_5H_{11}
14	SCF_3	34	$OCOCH_3$	54	C_6H_5
15	CN	35	CH_2COOH	55	OC_6H_5
16	SCN	36	OCH_3COOH	56	NHC_6H_5
17	NCS	37	$NHCOCH_3$	57	C_6H_{11}
18	CHO	38	$NHCOOCH_3$	58	COC_6H_5
19	COOH	39	C_2H_5	59	$CH_2CH_2C_6H_5$
20	$CONH_2$	40	OC_2H_5		

Esse nível de classificação, entretanto, é simplista do ponto de vista no qual assume-se que todas as características físico-químicas sejam correlacionadas ou não. Não obstante, ele é robusto no que concerne à classificação de famílias similares que pode conduzir à redução do número de candidatos em estudo, já que apenas alguns dos membros de cada família serão suficientes para representar toda a família, pelo menos dentro do princípio de que nas relações quantitativas uma determinada propriedade precisa ser caracterizada de forma apropriada.

Embora essa classificação seja apenas qualitativa, será possível ainda estabelecer uma nova família para compostos em teste não-similares e nem dissimilares e, portanto, não pertencentes a quaisquer dos conjuntos previamente estabelecidos.

CONCLUSÃO

Este trabalho estabelece uma rotina para a seleção de variáveis e de substituintes necessários para o delineamento de um determinado espaço "físico-químico" que representa uma "subcoleção" que pode estar constituída de poucos grupos químicos factíveis sinteticamente e que poderão fornecer informações importantes nos estudos de relações entre estrutura e propriedade e também atividade biológica. Os compostos estão em fase de síntese e os resultados biológicos serão motivo de publicação em outro tempo.

É necessário enfatizar, no entanto, que o crescimento do número de substituintes e mesmo o número de descritores poderá dificultar a presente análise, mas provavelmente não a invalidará. Nesse aspecto, portanto, faz-se mister que a diversidade química e biológica estejam perfeitamente delineadas e delimitadas. Nesse caso, certamente, o método mostrar-se-á eficiente e poderá constituir uma alternativa elegante ao tratamento experimental de imensos bancos de dados. De particular interesse, encontramos as substâncias oriundas de produtos naturais que poderão ser classificadas quanto às suas potencialidades dentro de uma determinada família terapêutica (tanto estrutural quanto biologicamente)[44]. Nosso grupo já está trabalhando nesse sentido e os resultados serão motivo de nova comunicação futura.

Os métodos quimiométricos usados no presente trabalham representam o "estado-da-arte" em matéria de reconhecimento molecular[41, 43]. Entretanto, outros métodos já estão disponíveis e podem ser usados com a mesma finalidade, como por exemplo o uso de algoritmo genético[45]. Métodos de seleção de variáveis como PLS[41,43], GOLPE[47] e Partição Recursiva (RP)[55] também são importantes nesse processo. Apesar disso, é bom ressaltar que não se trata de usar o método disponível, mas sim aquele que seja capaz de resolver um determinado problema. Além disso, esses métodos têm apresentado essencialmente os mesmos resultados, mas estes últimos talvez sejam mais eficientes quando um número muito elevado de variáveis estiver envolvido (10 até 10k) e algumas vezes são mais robustos na descrição adequada dessas variáveis. Mas, em outras ocasiões são apenas capazes de descrever essencialmente o mesmo resultado, como por exemplo em aplicações de redes neurais genéticas[48, 49]. É o que acontece com o SIMCA. Os resultados apresentados por esse método são, em geral, superiores àqueles apresentados pelo método KNN.

Para uma revisão exaustiva sobre o assunto, sugere-se ao(à) leitor(a) outras obras recentes[50-55].

AGRADECIMENTOS

Os autores agradecem ao CNPq, FAPEMIG e FINEP pelas bolsas e auxílios financeiros concedidos para a realização deste projeto.

REFERÊNCIAS

1. Hansch, C.; Hoekman, D. e Gao, H.; *Chem. Rev.* **1996**, *96*, 1045
2. Hansch, C. e Fujita, T.; *J. Am. Chem. Soc.* **1964**, *86*, 1616
3. Furlán, R. L. E.; Labadie, G. R.; Pellegrinet, S. C. e Ponzo, V. L.; *Quím. Nova* **1996**, *19*, 411
4. Moos, W. H., Green, G. D., Pavia, M. R.; *Ann. Rep. Med. Chem.* **1993**, *28*, 315.
5. Young, S. S.; Sheffield, C. F. e Farmen, M.; *J. Chem. Inf. Comp. Sci.* **1997**, *37*, 892
6. M. R., Sawyer, T. K.; Moos, W. H.; *Bioorganic Medicinal Chem. Lett.* **1993**, *3*, 387
7. Gallop, M. A.; Barrett, R. W.; Dower, W. J.; Fodor, S. P. A.; Gordon, E. M.; *J. Med. Chem.* **1994**, *37*, 12337.
8. Gordon, E. M.; Barrett, R. W.; Dower, W. J.; Fodor, S. P. A.; Gallop, M. A.; *J. Med. Chem.* **1994**, *37*, 1385
9. Montanari, C. A.; *Quím. Nova* **1995**, *18*, 56
10. Merrifield, R. B.; *J. Am. Chem. Soc.* **1963**, *85*, 2149.
11. Merifield, R. B. e Stewart, J. M.; *Nature* **1965**, *207*, 522.
12. Domling, A.; *Nucleos. & Nucleot.* **1998**, *17*, 1667
13. Geysen, H. M.; Meleon, R. H.; Barteling, S. J.; *Proc. Nat. Acad. Sci. U.S.A.* **1984**, *81*, 3998.
14. Ganesa, A.; *Angew. Chem.-Int. Ed.* **1998**, *37*, 2828
15. Furka, A.; Sebestyen, F.; Asgedom, M.; Dibo, G. 1988, Abstr. 14th Int. Congr. Biochem., Prague, Czechoslovakia, Vol 5, pg 47. Abstr. 10th Intl. Symp. Med. Chem., Budapest, Hungary, pg 288.
16. (a) Houghten, R. A.; *Proc. Natl. Acad. Sci. U.S.A.* **1985**, *82*, 5131.
17. Boger, D. L.; Chai, W. Y. e Jin, Q.; *J. Am. Chem. Soc.* **1998**, *120*, 7220
18. Houghten, R. A.; Pinilla, C.; Blondelle, S. E.; Appel, J. R.; Dooley, C. T.; Cuervo, J. H.; *Nature* **1991**, *354*, 84.
19. Kim, S. W.; Shin, Y. S. e Ro, S. G.; *Bioorg. Med. Chem. Lett.* **1998**, *8*, 1665
20. Lam, K. S.; Salmon, S. E.; Hersh, E. M.; Hruby, V. J.; Kazmierski, W. M.; Knapp, R. J.; *Nature* **1991**, *354*, 82
21. Brenner, S.; Lerner, R.A.; *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 5381
22. Nielsen, J.; Brenner, S.; Janda, K. D.; *J. Am. Chem. Soc.* **1993**, *115*, 9812
23. Needels, M. C.; Jones, D. G.; Tate, E. H.; Heinkel, G. L.; Kochersperger, L. M.; Dower, W. J.; Barrett, R. W.; Gallop, M. A.; *Proc. Natl. Acad. Sci. USA* **1993**, *90*, 10700.
24. Ohlmeyer, M. H.; Swanson, R. N.; Dillard, L. W.; Reader, J. C.; Asouline, G.; Kobayashi, R.; Wigler, M.; Still, W. C.; *Proc. Natl. Acad. Sci. USA* **1993**, *90*, 10922
25. Smith G. P.; *Science* **1985**, *228*, 1315
26. Cwirla, S.; Peters, E. A.; Barrett, R. W.; Dower, W. J.; *Proc. Natl. Acad. Sci. USA* **1990**, *87*, 6378
27. Scott, J. K.; Smith, G. P.; *Science* **1990**, *249*, 386
28. Devlin, J. J.; Panganiban, L. C.; Devlin, P. E.; *Science* **1990**, *249*, 404
29. Zuckermann, R. N.; Martin, E. J.; Spellmeyer, D. C.; et al, *J. Med. Chem.* **1994**, *37*, 2678
30. Bunin, B. A.; Ellman, J. A.; *J. Am. Chem. Soc.* **1992**, *114*, 10997
31. Liu, D. X.; Jiang, H. L.; Chem, R. X. e Ji, R. Y.; *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 233
32. DeWitt, S. H.; Kiely, J. S.; Stankovic, C. J.; Schroeder, M. C.; Reynolds Cody, D. M.; Pavia, M. R.; *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 6909
33. Xiang, X. D. e Schultz, P. G.; *Physica* **1997**, *282, 428*
34. Hansch, C. e Leo, A.; em *"Exploring QSAR: Fundamentals and Applications in Chemistry and Biology.* 1995, ACS, Washington, USA, p. 392-403
35. Volpe, P. L. O. e Montanari, C. A.; *Quim. Nova.* **1997**, *20*, 125
36. Antonini, I.; Claudi, F.; Cristalli, G.; Franchetti, P.; Grifantini, M. e Martelli, S.; *J. Med. Chem.* **1981**, *24*, 1181
37. Montanari, C. A.; Tute, M. S.; Beezer, A. E. e Mitchell, J. C.; *J. Comp.-Aided Mol. Des.* **1996**, *10*, 67
38. Van de Waterbeemd, H. e Testa, B.; *Adv. Drug Res.* **1987**, *16*, 85
39. Van de Waterbeemd, H.; El Tayar, N.; Carrupt, P. A. e Testa, B.; *J.Comput,-Aided Des.* **1989**, *3*, 111
40. Van de Waterbeemd, H.; Carrupt, P. A.; Testa, B. e Kier, L. B. em: *"Trends in QSAR and Molecular Modeling 92"*, Ed. C. G. Wermuth, Escom, The Netherlands 1993 pp. 69-75.
41. Van de Waterbeemd, H.; Costantino, G.; Clementi, S.; Cruciani, G. e Valigi, R. em: *"Chemometric Methods in Molecular Design"*, Ed. H. van de Waterbeemd, VCH, Weinheim 1995 pp. 113-164 e pp. 179-194.
42. Scarminio, I. S. e Bruns, R. E.; *Trends Anal. Chem.* **1989**, *8*, 326

43. Dunn III, W. J. e Wold, S.; em *"Methods and Principles in Medicinal Chemistry*, Mannhold, R,, Krogsggard-Larsen, e H. Timmerman, (eds.), Vol. 2, "Chemometric Methods in Molecular Design, Van de Waterbeem, H., (ed.) 1995, VCH, Weinheim, p. 179-194
44. Dutton, G.; *Genet. Eng. News* **1998**, *18*, 18
45. Brown, R. D. e Martin, Y. C.; *J. Med. Chem.* **1997**, *40*, 2304
46. Referência 41, Wold, S., pp. 195-218
47. Baroni, M.; Costantino, G.; Cruciani, G.; Riganelli, D.; Valigi, R. e Clementi, S.; *Quant. Struct.Act. Relat.* **1993**, *12*, 9
48. So, S.-S. e Karplus, M.; *J. Med. Chem.* **1997**, *40*, 4347
49. So, S.-S. e Karplus, M.; *J. Med. Chem.* **1997**, *40*, 4360
50. Zheng, Q. e Kyle, D. J.; *Bioorg. Med. Chem.* **1996**, *4*, 631
51. Endereço eletrônico na Internet: The ChemicalGeneration of Molecular Diversity. http://www.awod.com/netsci/Science/Combichem/
52. Waller, C. L.; Bradley, M. P.; *J. Chem. Inf. Comp. Sci.* **1999**, *39*, 345
53. Robinson, D. D.; Winn, P. J.; Lyne, P. D.; Richards, W. G.; *J. Med. Chem.* **1999**, *42*, 573
54. Linusson, A.; Wold, S.; Norden, B.; *Chemomet. Intel. Lab. Syst.* **1998**, *44*, 213
55. Young, S. S.; Hawkins, D. M.; *SAR-QSAR Environ. Res.* **1998**, *8*, 183