

Químio-informática.

Conteúdos que urge ensinar.

JOÃO AIRES DE SOUSA*

Resumo

Na gestão de informação, computação, comunicação e edição, a Química usa hoje intensivamente e com uma especificidade muito própria, as principais funcionalidades dos computadores. O aproveitamento da cada vez mais abundante informação química requer o desenvolvimento constante de metodologias novas, que não pode ser feito sem uma forte formação em Química e sólidos

conhecimentos de Informática. A existência de uma gama extensa de software para Química exige dos potenciais utilizadores conhecimentos específicos sobretudo ao nível dos métodos envolvidos. Por tudo isto, sugere-se a incorporação de conteúdos de Químio-informática no sistema de ensino universitário e são propostos temas a tratar nesse contexto.

Introdução

A rápida disseminação de suportes electrónicos para arquivo de informação, o desenvolvimento da instrumentação laboratorial, nomeadamente no sentido da automatização e da miniaturização e o aparecimento de novas técnicas como a química combinatorial e HTS (*high-throughput screening*) provocaram nos últimos anos uma verdadeira explosão na produção de informação química. Uma empresa farmacêutica média que realize HTS pode produzir resultados de dezenas de milhar de experiências por dia¹. O Chemical Abstracts adiciona actualmente cerca de 900 000 novos compostos por ano à sua base de dados². Isto requer tecnologias eficientes para arquivo e acesso a dados; o problema mais sério que se põe é o de como, para uma dada aplicação, desprezar o "ruído" e usar a informação útil. Mais ainda, a utilização desta informação permite estabelecer correlações, aprender, extrair conhecimento, objectivos que só são alcançáveis usando metodologias computacionais e de inteligência artificial.

Por outro lado, a Internet trouxe possibilidades inovadoras de comunicação, de colaboração à distância e de acesso a vastíssimas fontes de informação, o que tem contribuído para a implementação de estratégias de representação e manuseamento de informação específicas da Química.

Simultaneamente tem-se observado um extraordinário incremento do poder computacional e uma notável sofisticação das técnicas de computação em Química. A modelação molecular tornou-se clássica, ganhou honras de prémio Nobel³ e a Química Teórica estabeleceu-se como um reconhecido ramo da Química.

Os frutos obtidos em todas estas frentes estão acessíveis ao químico experimental comum através de software de fácil utilização para os sistemas operativos que lhe são familiares e através de interfaces WWW. São ferramentas que invadiram gradualmente os laboratórios (e os universos mentais) e que se estão a impor mesmo aos químicos avessos à modernidade electrónica. Mas são utili-

zadas na maioria dos casos como caixas negras que ora fazem milagres, ora produzem resultados sem sentido!

Discute-se que nome dar a esta área nova, na fronteira entre a Química e a Informática, que aplica e desenvolve sistemas de informação, algoritmos, técnicas computacionais e métodos estatísticos para problemas de Química. "Cheminformatics", que aqui traduzo por Químio-informática, é a designação que vai sendo mais consensualmente aceite. "Química Computacional" é uma expressão ainda usada como sinónimo, embora seja normalmente conotada com a utilização de software (como o Gaussian⁴, MOPAC, DOCK⁵, Autodock⁶, ...) para modelação molecular.

Em todo o mundo industrializado, a indústria química e farmacêutica que faz investigação e desenvolvimento, utiliza fortemente recursos de Químio-informática, mas tem dificuldade em encontrar especialistas na área. Os informáticos raramente têm apetência para a Química e os químicos sabem normalmente muito pouco de Informática. Para suprir

*Investigador na área de Químio-informática, na FCT/UNL e na Universidade de Erlangen-Nürnberg (Alemanha). Morada: Departamento de Química, CQFB, campus FCT-UNL, Quinta da Torre, 2829 Monte de Caparica, Portugal. Email: jas@fct.unl.pt

estas necessidades, estão a implementar-se rapidamente cursos de mestrado em Químio-informática, sendo os pioneiros os da Universidade de Sheffield (Reino Unido)⁷, UMIST (Universidade de Manchester, Reino Unido)⁸ e IUPUI (Universidade de Indiana-Purdue, E.U.A.)⁹. A nível de pré-graduação, incorporaram-se já disciplinas de Químio-informática em numerosas Universidades¹⁰.

E enquanto tudo isto acontece, que adaptações sofrem os conteúdos das licenciaturas de Química nas Universidades Portuguesas? Que formação é proporcionada aos alunos portugueses nestas tecnologias?

Não são tecnologias futuristas, nem distantes, nem inacessíveis, nem caras... São exigidas por qualquer químico que queira hoje produzir um relatório, manusear um espectro, fazer pesquisa bibliográfica, analisar resultados obtidos, fazer previsões para os problemas que tem em mão...

O presente artigo pretende ser, por um lado, uma reflexão sobre as intervenções possíveis e necessárias no campo da Químio-informática nas instituições de ensino superior em Portugal e, por outro, um contributo para o planeamento dos cursos deste tipo que inevitavelmente vão ser criados.

A situação Portuguesa

A Químio-informática e a Bio-informática têm sido consistentemente nos últimos anos áreas científicas com excesso de procura de mão-de-obra face à oferta existente. A necessidade deste tipo de especialistas tem origem nas indústrias química e farmacêutica que fazem investigação de novos produtos e que

utilizam nesse processo métodos computacionais e técnicas intensivas baseadas em informação. É essa a razão por que o fenómeno não aconteceu ainda em Portugal, onde aquelas indústrias não têm tradicionalmente dimensão para desenvolver produtos novos. No entanto, a tendência da indústria multinacional para recorrer ao *outsourcing* também na investigação, e a abundante mão-de-obra científica formada em Portugal na última década, permitem perguntar se não haverá oportunidade para empresas de investigação portuguesas...

ção vai rapidamente permeando os métodos de decisão, as técnicas laboratoriais, os métodos de avaliação de resultados. Se numa primeira fase o fenómeno se centrou na indústria de inovação aplicada, ele vai estender-se ao ensino, à investigação fundamental, à optimização de processos e aos sistemas de controlo de qualidade, áreas que empregam a maioria dos químicos em Portugal.

Onde ir buscar profissionais com competências nesta área? O problema pôs-se há décadas, em pequena escala, nos grupos académicos pioneiros da Químio-informática. Nessa primeira geração recorreu-se normalmente a químicos

que adquiriram conhecimentos de informática de forma mais ou menos estruturada. O problema estendeu-se à indústria, em grande escala, quando esta adoptou metodologias computacionais, e tem sido resolvido por químicos com formação complementar em matemática ou informática.

No contexto actual exige-se uma resposta mais sistematizada por parte das instituições de ensino e, nos EUA e Reino Unido, estão a criar-se cursos de mestrado em Químio-informática e cadeiras nesta área a incluir nos cursos de graduação.

Em Portugal, a criação de disciplinas de Químio-informática nas licenciaturas de Química parece, por tudo o que acima se disse, urgente. Já a criação de mestrados será provavelmente desajustada no panorama presente. Para além do mais, o recrutamento para este campo, de licenciados em Química, será sempre uma estratégia com poucos frutos. Sempre serão raros os químicos que, depois de uma licenciatura em Química de 4 ou 5 anos, decidirem entrar numa

NCI Number	Formula	CAS	#Names	Sample Name
297343	C ₁₇ H ₂₁ N ₃ O ₂	(None)	1	(2-ethyl-5-methyl-1H-imidazol-3-yl)(3-pyridyl)carbamate
297344	C ₁₇ H ₂₁ N ₃ O ₂	(None)	1	(2-ethyl-5-methyl-1H-imidazol-3-yl)(4-pyridyl)carbamate

Tradicionalmente tem-se exigido aos químicos 'conhecimentos de informática na óptica do utilizador' e, mesmo esses, não lhes são normalmente fornecidos pelas universidades que frequentam. Conhecimentos e capacidades muito mais vastos nesta área vão ser necessários no futuro próximo à medida que os métodos computacionais em Química se desenvolvem e estabelecem, que o volume de informação química disponível começa a ser reconhecido e que a instrumentação gera e arquiva mais e mais dados. A computa-

área com tão forte componente de Informática. O mesmo problema acontece em Biologia, para encontrar cientistas na área da Bio-informática.

Por outro lado, seria muito mais fácil propor a finalistas do ensino secundário uma licenciatura em Bio- e Químio-informática cujos conteúdos principais seriam a Química, a Biologia e a Informática. A licenciatura daria formação básica em Biologia e em Química, com ênfase na instrumentação e nas aplicações informáticas; e daria formação sistemática em Informática. Produziria licenciados com formação e motivação próprias para trabalhar em Bio-informática e em Químio-informática; mas forneceria também profissionais para empregos mais frequentemente oferecidos nas áreas comerciais e de controlo de qualidade de empresas químicas e farmacêuticas, e mesmo para empregos na área da informática. Outra alternativa é a criação de um ramo de Bio- e Químio-informática em licenciaturas de Química ou Biologia. Mas será algum destes caminhos possível, em tempo de diminuição de número de alunos, num sistema universitário cansado de licenciaturas com nomes novos e problemas velhos ?...

Mais fácil será talvez introduzir cadeiras de Químio-informática nas licenciaturas de Química e criar cursos de formação profissional nesta área. Embora com nomes diversos, em várias universidades portuguesas têm sido incluídas matérias de informática e computação em algumas disciplinas de Química, ainda que de forma pouco abrangente (existindo ainda licenciaturas em Departamentos nacionais de prestígio sem uma disciplina de Química Computacional). As aproximações já feitas são resumidamente as seguintes:

- a) Disciplinas de Química Teórica que tratam da teoria subjacente aos métodos de modelação molecular (química quântica, soluções aproximadas da equação de Schrödinger, modelo de Hückel, método de Hartree-Fock,...). Incluem conteúdos de termodinâmica estatística, teoria de grupos e por vezes também de tratamento de erros.
- b) Disciplinas de Química Computacional ou de Modelação Molecular que têm por vezes conteúdos sobrepostos aos de cadeiras de Química Teórica. Incidem especialmente sobre 1) aplicações de cálculos quânticos e de campos de forças empíricos ao estudo de mecanismos reaccionais e de interacções ligando-receptor; 2) dinâmica molecular; 3) generalidades sobre sistemas operativos e programação.
- c) Disciplinas de Quimiometria que tratam de aplicações de estatística e redes neuronais à análise de dados químicos. Aplicações de referência são a classificação de amostras a partir de análises com múltiplas variáveis e o estabelecimento de relações entre estrutura molecular e função.
- d) Disciplinas de Documentação Científica que abordam, para além das fontes tradicionais de informação em papel, as fontes de informação química na Internet e em suportes digitais.
- e) Disciplinas de Química várias em cujos programas os docentes incluíram a utilização de software aplicado às respectivas matérias.

Propostas de Conteúdos

Sem entrar em pormenores quanto às reestruturações necessárias nas disciplinas existentes, proponho de seguida pistas de conteúdos que deveriam ser ensinados sistematicamente em disciplinas de Químio-informática.

Edição e visualização de informação química. Através de editores moleculares (por exemplo ChemDraw¹¹, ISIS Draw¹² ou applets Marvin¹³ em Java) criam-se e editam-se ficheiros electrónicos contendo estruturas químicas. Estruturas 2D ou 3D de moléculas simples ou de macromoléculas, assim como cromatogramas ou espectros, são visualizados através de programas próprios, de que existem muitos exemplos, gratuitos ou comerciais.

Para apresentação de informação na WWW usam-se principalmente docu-

mentos HTML, que podem ser editados com editores próprios, e que podem incorporar, de forma interactiva, estruturas químicas 2D ou 3D e espectros, por exemplo utilizando o *plug-in* Chime¹² para *browsers* de WWW. Tem sido desenvolvida uma linguagem própria para codificação de informação química (CML — *Chemical Mark-up Language*) tendo em vista sobretudo a Internet^{14,15}. Estruturas e propriedades moleculares podem ser visualizadas em ambiente de realidade virtual usando o formato VRML¹⁶. Informação química pode ser transmitida por email e reconhecida automaticamente pelo cliente de email de modo a utilizar editores ou visualizadores adequados. Tal é possível pela utilização de MIME químico¹⁷.

Arquivo de informação química. A existência de uma estrutura molecular está na identidade dos seus átomos, nas ligações que estabelecem entre si e na orientação espacial destas. Têm sido propostas várias maneiras de codificar esta informação e existem consequentemente vários formatos e tipos de ficheiros em que são arquivadas estruturas moleculares. Entre os mais comuns estão os ficheiros .mol¹⁸ e .pdb¹⁹. Também a informação espectroscópica é representada em formatos próprios que usam frequentemente estratégias para reduzir o tamanho dos ficheiros necessários. Há uma tendência para que o formato JCAMP²⁰ se torne um padrão.

Volumes grandes de informação arquivam-se de preferência em bases de dados. Existem vários sistemas para a construção de bases de dados com informação química (como o ChemFinder¹¹, ISIS Base¹², CACTVS²², ACD Spec-Manager²³) que permitem arquivar estruturas químicas, reacções, propriedades numéricas ou alfanuméricas, fórmulas, espectros. Vários tipos de pesquisas podem ser efectuadas, sendo específicas da Química as pesquisas de sub-estruturas, de sub-espectros, de reacções ou de semelhança estrutural. Através de bases de dados relacionais pode incorporar-se informação existente em várias bases de dados.

Computação química. Através de métodos estatísticos e de redes neuronais é

possível estabelecer correlações entre estrutura química e propriedade molecular (QSPR) ou entre estrutura química e actividade biológica (QSAR). Para tal, é frequente representar estruturas químicas por números, mais concretamente por um número fixo de parâmetros. Tem-se desenvolvido uma grande variedade de descritores moleculares baseados principalmente em grafos moleculares, em propriedades físico-químicas, ou em características geométricas.

A utilização de descritores moleculares, de impressões digitais moleculares e de outros tipos de códigos permite analisar o grau de diversidade estrutural existente num conjunto de compostos, o que é frequentemente útil no design de bibliotecas combinatoriais. Por outro lado permite estabelecer medidas de semelhança estrutural entre duas ou mais moléculas.

A aplicação mais clássica da computação química consiste no recurso à mecânica molecular ou à mecânica quântica para o cálculo de propriedades termodinâmicas de compostos e reacções. Vários tipos de programas estão disponíveis, com aproximações de vários níveis e têm tido muitas aplicações no design de fármacos e de reacções.

Algoritmos muito utilizados são os algoritmos evolucionários, principalmente os algoritmos genéticos. Simulam a evolução de uma população em que os indivíduos com características mais desejáveis sobrevivem e reproduzem-se, fazendo sobressair, ao fim de várias gerações, as características que tornam os indivíduos mais aptos. São utilizados para problemas de optimização e têm tido aplicação desde a análise conformacional até à selecção de descritores moleculares mais adequados ao estabelecimento de QSAR.

Bases de dados com informação bioquímica. Estes tipos de bases de dados têm a especificidade de conterem principalmente estruturas 3D de macromoléculas biológicas, sequências (de aminoácidos ou ácidos nucleicos) e vias metabólicas. É de especial relevo neste contexto o projecto do genoma humano. A análise para extrair conhecimentos re-

levantes destas bases de dados envolve algoritmos e software para alinhamento de sequências, pesquisa de semelhanças, estimativa de árvores filogenéticas, previsão estrutural e inferência funcional.

Fontes de informação química. O aparecimento e omnipresença da Internet veio revolucionar as fontes de informação a que os químicos têm acesso.^{24,25,26} Para além de inúmeras fontes novas, também as antigas (revistas, enciclopédias, catálogos, índices) migraram ou tendem a migrar (pelo menos em regime complementar) para a Internet. Estão disponíveis bases de dados espectroscópicas e estruturais, toxicológicas e de segurança, de reacções químicas, de propriedades moleculares e catálogos comerciais. Existem ferramentas comerciais de pesquisa bibliográfica (como o Chemical Abstracts ou o Beilstein) mas também *sites* gratuitos como o da Ingenta²⁷ (herdeiro do BIDS e do Uncover) ou o Medline²⁸, e serviços de alerta. Outra fonte formidável de informação disponível gratuitamente na Internet e frequentemente esquecida nas Universidades são as bases de dados de patentes²⁹. E as revistas científicas tradicionais estão presentes na Internet a par com outras revistas exclusivamente electrónicas.

Comunicação em Química. As comunidades virtuais de químicos na Internet têm tomado formas variadas. Nas *mailing lists* (ou *listserv*) a comunicação funciona por email sendo cada mensagem distribuída por todos os membros da lista³⁰. Um exemplo é a ORGLIST³¹ – *mailing list* internacional de Química Orgânica, que tem sede em Portugal. Nos *bulletin boards* os membros da comunidade podem entrar num *web site* para visualizar as mensagens publicadas e criar as suas próprias mensagens. Os *newsgroups* são os mais antigos, têm um protocolo próprio e as mensagens são públicas. Na ChemWeb³², os membros têm acesso a *fora* de discussão, a uma biblioteca, bases de dados, reportagens sobre congressos, informação sobre emprego, ou centro comercial. Idealmente a interacção com outros químicos numa comunidade virtual permi-

te a consulta rápida à "base de dados" estabelecida pela ligação em rede dos conhecimentos dos químicos envolvidos. Para além disso, os arquivos das mensagens trocadas nestas comunidades, quando existem e são pesquisáveis, constituem fontes de informação valiosas.

Mas os canais mais consistentes para a disseminação de informação química de qualidade continuam a ser as publicações com sistema de *refereeing*. Estas têm posto na Internet muitas das suas funcionalidades³³, incluindo a possibilidade de publicar material de forma interactiva, impossível no suporte de papel. A submissão de artigos tende a ser feita preferencialmente por email ou por transferência de ficheiros num *web site* e os artigos podem muitas vezes ser consultados na WWW pelos assinantes, havendo várias soluções consoante os editores. Normalmente o material suplementar dos artigos está disponível gratuitamente na WWW. É também frequente as revistas proporcionarem pesquisas nos *abstracts* ou nos títulos dos seus artigos em arquivo, sendo estas pesquisas alargadas, em alguns casos, a todas as revistas de um mesmo editor (como no caso da ACS³⁴).

Conclusão

A Químio-informática é uma área emergente que utiliza métodos computacionais para problemas de Química. Tem sido intensamente utilizada pela indústria química e farmacêutica na descoberta de novos produtos e tem trazido interpretações novas a questões científicas fundamentais. A presença destes conteúdos no ensino universitário da Química em Portugal é escassa e a sua incorporação ao nível da graduação e pós-graduação deve ter em consideração a) o panorama nacional e europeu de emprego para químicos e cientistas; b) que muitos destes métodos passaram já para o núcleo dos métodos básicos da Química. Neste artigo são lançadas pistas de reflexão sobre o assunto.

Bibliografia

- 1 D. J. Ausman, *Modern Drug Discovery*, **2001**, 4(5), 33-39.
- 2 <http://www.cas.org>
- 3 <http://www.nobel.se/chemistry/laureates/1998/index.html>
- 4 <http://www.gaussian.com>
- 5 <http://www.cmpchem.ucsf.edu/kuntz/dock.html>
- 6 <http://www.scripps.edu/pub/olsonweb/doc/autodock/>
- 7 <http://www.shef.ac.uk/~is/courses/pgrad/mscci/mscci.html>
- 8 <http://www.ch.umist.ac.uk/MScCheminf.htm>
- 9 <http://www.chem.iupui.edu/>
- 10 Por exemplo
- a) <http://www.ch.ic.ac.uk/>.
- b) <http://www.sfx.ca/people/chemist/chem375.html>.
- c) <http://www.vsch.cz/syllab/Subjects/ChemicalInformatics.html>
- 11 <http://www.cambridgesoft.com/>
- 12 <http://www.mdli.com>
- 13 <http://www.chemaxon.com/>
- 14 <http://www.xml-cml.org/>
- 15 P. Murray-Rust, H. S. Rzepa, *J. Chem. Inf. Comput. Sci.* **1999**, 39, 928.
- 16 <http://www2.chemie.uni-erlangen.de/services/vrmlcreator/index.html>
- 17 H. S. Rzepa, P. Murray-Rust, and B. J. Whitaker, *J. Chem. Inf. Comput. Sci.*, **1998**, 38 (6), 976-982.
- 18 <http://www.mdli.co.uk/cgi/dynamic/downloadsect.html?uid=&key=&id=1#file>
- 19 http://www.rcsb.org/pdb/info.html#File_Formats_and_Standard
- 20 <http://www.isasdortmund.de/projects/jcamp/jcamp.htm>
- 21 <http://www.chemfinder.com>
- 22 <http://www2.chemie.uni-erlangen.de/software/cactvs/index.html>
- 23 <http://www.acdlabs.com>
- 24 J. Aires de Sousa, *Química*, **1996**, 61, 16-20.
- 25 G. Wiggins, *J. Chem. Inf. Comput. Sci.*, **1998**, 38(6), 956-965.
- 26 <http://www.indiana.edu/~cheminfo/>
- 27 <http://www.ingenta.com>
- 28 <http://www.ncbi.nlm.nih.gov/entrez/query.fcgi>
- 29 <http://ep.espacenet.com/>
- 30 a) J. Aires-de-Sousa, *Internet Journal of Chemistry*, **1999**, 2.
- b) P. van der Valk, *Chemical Web Marketing and Technology*, **1997**, 1(3) (<http://www.chemconnect.com/cwmt/cwmt03.html#mailing>).
- c) W. Warr, *J. Chem. Inf. Comput. Sci.* **1998**, 38, 966-975.
- d) P. van der Valk, *Chemical Web Marketing and Technology*, **1998**, 2(8) (<http://www.chemconnect.com/cwmt/cwmt208.html#vc>).
- e) A. Pisanty, J. Labanowski, *Trends Anal. Chem.* **1996**, 15, 53.
- (<http://www.elsevier.nl:80/homepage/saa/trac/emltrac.htm>) f) S. Miller, *Trends Anal. Chem.* **1996**, 15, 164. (<http://www.elsevier.nl:80/homepage/saa/trac/lms.htm>)
- 31 <http://www.dq.fct.unl.pt/orglist>
- 32 <http://www.chemweb.com>
- 33 <http://www.chemconnect.com/library/journals/index.html>
- 34 <http://pubs.acs.org>



NÚMERO/Number **96/CEP.410**

A Associação Portuguesa de Certificação (APCER)
The Portuguese Association for Certification (APCER)
certifica que o sistema da qualidade da
certifies that the quality system of

SOQUÍMICA – SOCIEDADE DE REPRESENTAÇÕES DE QUÍMICA, LDA.
Rua Coronel Santos Pedroso, 15
1500 - 207 LISBOA
PORTUGAL

Implementado na comercialização, manutenção e calibração de
equipamentos de laboratório, cumpre os requisitos da
implemented in the supply, servicing and calibration of laboratory equipments, meets the requirements of

NP EN ISO 9002:1995

Sistemas da Qualidade. Modelo de garantia da qualidade na
produção, instalação e assistência após-venda.
Quality Systems. Model for quality assurance in production, installation and servicing

O presente certificado é emitido no âmbito do Sistema Português da Qualidade.
This certificate is issued within the Portuguese System for Quality

Data de emissão/Date of issue 1999-06-18 Válido até/Valid until 2002-06-17

Luís Fonseca
Director Geral/General Director



SOQUÍMICA
Sociedade de Representações de Química, Lda

R. Coronel Santos Pedroso 15 • 1500-207 Lisboa Tel 21 716 5160 • Fax 21 716 5169

R. 5 de Outubro 269 • 4100-175 Porto Tel 22 609 3069 • Fax 22 600 0834

E-mail: soquimica@mail.telepac.pt

www.soquimica.pt