

NMRDB para atividades interativas investigativas com a informática química (cheminformatic)

Alba D. Q. Ferreira¹, Bruna Souza², Gustavo H. Carvalho², Nicolas C. Scorpelini²

¹Instituto de Ensino Superior São Paulo– Faculdade de Paulínia (FACP)
Paulínia – SP – Brasil

²Instituto de Ensino Superior Campo Limpo Paulista– Faculdade Campo Limpo Paulista (FACCAMP)

alba.facp@gmail.com

Abstract. *NMRDB is a tool of chemical informatics for predictions of spectra of Nuclear Magnetic Resonance (NMR), an essential technique for the determination of molecular structures. The development of nmrdb illustrates the ANNs (Artificial Neural Networks) solving chemical problems. The characteristics of this educational tool facilitate interactivity and investigative thinking, as demonstrated in the activity with acetylcholine, histamine and oxytocine molecules.*

Resumo. *NMRDB é uma ferramenta da informática química para previsões de espectros de Ressonância Magnética Nuclear (RMN), uma técnica essencial para a determinação de estruturas moleculares. O desenvolvimento do nmrdb ilustra as RNAs (Redes Neurais Artificiais) na solução de problemas químicos. As características desse recurso educativo facilitam a interatividade e o pensamento investigativo, conforme demonstrado na atividade com as moléculas da acetilcolina, histamina e oxitocina.*

1. Introdução

A quantidade crescente de informações sobre as substâncias químicas, desde pequenas moléculas até macromoléculas, associada à necessidade de mecanismos de busca de dados usando a linguagem química, contribuiu para o desenvolvimento da informática química (na língua inglesa pode ser *cheminformatic* ou *chemoinformatic*). Neste caso, o interesse maior está no planejamento, criação, organização, gestão, recuperação, análise, disseminação, visualização e o uso dos dados para resolver problemas químicos, incluindo as previsões de propriedades. Em termos de conceitos básicos, a química computacional utiliza os da mecânica clássica e da química quântica [Skaf 2013], ao passo que a informática química necessita de um novo conceito, o espaço químico, onde as moléculas equivalem às estrelas desse universo [Varnek & Baskin 2010].

A informática química, além de ilustrar as aplicações da matemática para a compreensão de fenômenos químicos, forneceu nos últimos anos recursos para o desenvolvimento de atividades investigativas que possibilitam a visualização de estruturas químicas e a previsão de propriedades de moléculas. Os algoritmos que fazem previsões utilizam,

entre outros recursos, a abordagem das redes neurais artificiais (RNAs), como exemplificado na ferramenta usada neste trabalho, o *www.nmrdb.org* [Banfi & Patiny 2008]. Esta ferramenta *online* é uma interface alternativa do projeto SPINUS (*Structure-based Predictions In Nuclear magnetic resonance Spectroscopy*) [Aries-de-Sousa & Binev 2011].

A Ressonância Magnética Nuclear (RMN) é uma das mais eficientes para a elucidação de estruturas de moléculas orgânicas, incluindo as biomoléculas. Para esta técnica espectroscópica não há necessidade de luz, pois o sistema químico é perturbado por um campo magnético (B_0) com energia na faixa de rádio frequência (MHz). Os núcleos atômicos que geram sinal de RMN são de isótopos que possuem spin e momento magnético. Por exemplo, para o hidrogênio (por ter carga positiva é chamado de próton, ^1H) há dois estados de spin: $+1/2$ (de energia baixa, alinhado com B_0) e $-1/2$ (de energia alta, em sentido oposto de B_0). Na presença desse campo, o spin nuclear sofre transições entre dois níveis energéticos permitidos pela química quântica, sendo que este sinal poderá sofrer desdobramentos em função do acoplamento entre os núcleos ativos vizinhos [Pavia 2010]. A instrumentação para a realização desses experimentos é complexa e cara, mas os recursos tecnológicos de livre acesso para previsão de dados experimentais provam ser suficientes para o ensino - aprendizagem dos fatos, conceitos e princípios essenciais sobre RMN na graduação.

A seguir são apresentados exemplos de aplicações das RNAs nas soluções de problemas para análises de RMN e os resultados das atividades de simulações para a previsão de espectros de moléculas com atividade biológica, a saber, acetilcolina, histamina e oxitocina. Dentro desse contexto, a expectativa é divulgar um software educativo (SE) /objeto de aprendizagem (OA) [Machado 2016] de qualidade, acessível e que promove a interdisciplinaridade, o uso de habilidades investigativas e discussões sobre a importância da informática química para o processo de ensino-aprendizagem do pensamento químico.

2. As RNAs e estratégias para o desenvolvimento do NMRDB

Evidências desde o início da década de 90 apontam para o interesse de químicos na aplicação de RNAs para o desenvolvimento de ferramentas de informática química para a investigação de problemas analíticos como os de espectroscopia de RMN [Burns & Whitesides 1993], [Gasteiger & Zupan 1993], [Spining *et al.* 1994]. As características inerentes do fenômeno que resulta nos sinais de RMN facilitaram as investigações para previsão de parâmetros espectrais, correlação das estruturas com os espectros, determinação de propriedades das substâncias até análises de misturas biológicas complexas [Meusinger & Himmelreich 2003].

A construção de modelos para as previsões de valores de deslocamento químico (δ) requer a coleta de dados experimentais e da literatura envolvendo estruturas com vários tipos de prótons em diferentes ambientes químicos. Por outro lado, tais previsões podem ser melhoradas testando as RNAs com arquiteturas e algoritmos de aprendizado distintos. Por exemplo, para a previsão de valores de δ de hidrocarbonetos a rede neural de contrapropagação ou CPNN (*ConterPropagation Neural Network*) foi treinada com

744 exemplos de prótons tendo descritores físico-químicos, topológicos e geométricos, sendo que os algoritmos genéticos foram aplicados na seleção dos descritores [Aires-de-Sousa *et al.* 2002]. Com o mesmo número de exemplos de prótons, resultados mais exatos foram obtidos com um conjunto de redes diretas ou FFNN (*Feed Forward Neural Network*) [Binev & Aires-de-Sousa 2004]. Na versão atual do *nmrdb* há 75 FFNNs treinadas para cada classe de prótons, sendo 6.700 prótons ligados às estruturas alifáticas rígidas, 5.320 prótons ligados às estruturas alifáticas não rígidas, 5.230 prótons ligados aos sistemas aromáticos e 872 prótons π de sistemas não aromáticos, resultando em 18.122 valores de δ [Aires-de-Sousa & Binev 2011].

Porém, além de determinar os deslocamentos químicos, outro desafio computacional estava em eliminar os acoplamentos fracos e reduzir o tempo do cálculo das simulações de moléculas com mais de 13 núcleos ativos. Esta limitação foi contornada por meio de um algoritmo de Hamiltoniano aproximado, resultado da soma de um grande número de Hamiltonianos menores, que remove as interações fracas de acoplamento, faz o desdobramento dos sistemas de spins de modo eficiente e depois corrige o efeito dos acoplamentos que foram descartados [Castillo *et al.* 2011]. O procedimento de associações de redes neurais (ASNN) com as FFNNs foi aplicado para melhorar e corrigir as previsões dos deslocamentos químicos e também determinar, com uma ASNN secundária contendo dados experimentais, os valores das constantes de acoplamento (J, em Hz) [Binev *et al.* 2007].

3. Exemplos de simulações de espectros de RMN-¹H com o *NMRDB* e os aspectos investigativos da atividade

O tema “moléculas da vida” foi adotado como contexto da aprendizagem das aplicações da espectroscopia de RMN na disciplina de Análise Instrumental. As substâncias selecionadas (acetilcolina, histamina e oxitocina) possuem em comum a neurotransmissão, mas a histamina e, em especial a oxitocina, apresentam outras funções relevantes [Hastings *et al.* 2013]. A investigação estrutural de tais moléculas com o SE/OA proposto possibilita a visualização de fenômenos do universo quântico, o que pode contribuir para o desenvolvimento do pensamento químico [Wartha *et al.* 2010].

A seguir estão os passos essenciais para a interação com o *NMRDB* e os princípios básicos da origem do sinal de RMN que auxiliam na discussão dos espectros simulados.

Os estímulos para as RNAs do *NMRDB* são as estruturas químicas. Há duas opções para inserir esses dados: a) usando o código SMILES da molécula [Weininger 1988] ou b) desenhando a estrutura com o editor molecular JSME [Bienfait & Ertl 2013] que está acessível no *nmrdb* como indicado na Figura 1.

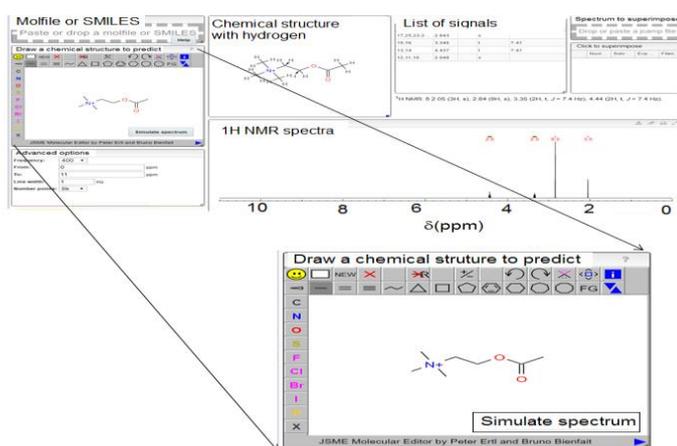


Figura 1. Estrutura da acetilcolina editada com o JSME no *NMRDB*.

Na prática, a frequência de ressonância exata de cada núcleo depende dos elétrons que estão no seu entorno causando o efeito de blindagem, o que resulta no campo magnético efetivo do núcleo. Como as diferenças de frequências são muito pequenas, então a medida é realizada em função de uma referência, para prótons é o TMS (tetrametilsilano). O deslocamento químico (δ) é a unidade de medida em partes por milhão (ppm) que indica a diferença de deslocamento da frequência em relação ao TMS (em Hz) em função da frequência do espectrômetro (em MHz). A posição do sinal de ressonância no espectro é influenciada pelo ambiente químico ao redor do núcleo. Quando os prótons estão próximos de grupos eletronegativos, os sinais são observados em valores de δ maiores, e neste caso os prótons estão mais desblindados. Para os prótons blindados, os sinais estão em valores de δ menores [Pavia 2010]. Para prótons em diferentes ambientes químicos é possível encontrar os valores de δ tabelados dentro de uma faixa de 0-12 ppm, o que facilita a verificação dos resultados do *NMRDB*. Na sequência será exemplificada a identificação de um conjunto de prótons no espectro simulado. É importante destacar que este SE/OA não faz previsões para prótons que estão diretamente ligados ao oxigênio e ao nitrogênio.

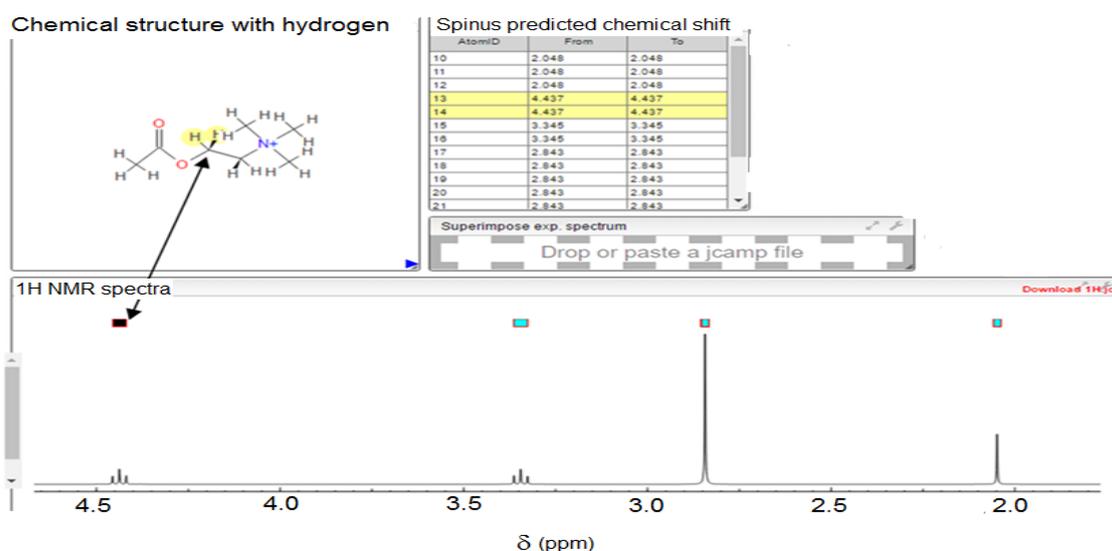


Figura 2. Espectro de RMN-¹H para acetilcolina com a versão anterior do programa.

Ao passar o cursor sobre algum núcleo destaca-se, com mudança de cor, o pico correspondente e vice-versa. Nesse exemplo os núcleos ativos em destaque correspondem ao conjunto de picos à esquerda, pois esses prótons estão mais próximos do grupo éster, o mais eletronegativo da estrutura, ou seja, os picos à esquerda são da parte menos blindada da molécula. Além do valor de δ , outro dado importante para elucidação estrutural de moléculas é o desdobramento dos sinais em função do número de núcleos ativos vizinhos ao núcleo que está sendo medido. Para o ^1H vale a regra $n + 1$, sendo n o número de núcleos ativos vizinhos [Pavia 2010]. Por exemplo, o sinal assinalado na Figura 2 confirma a presença de 2 prótons vizinhos, pois se desdobrou em um tripleto.

É possível observar que para a próxima molécula este princípio é ilustrado de modo semelhante. O grupo mais eletronegativo da histamina é o anel imidazólico e os 2 prótons a ele ligados (que estão em δ 2,71) têm como vizinhos 2 prótons um pouco mais blindados (δ 2,81), fornecendo os dois tripletos indicado na Figura 3. Esta pequena variação de δ é visualizada com a simulação em 400 MHz. A magnitude da interação entre os prótons vizinhos depende da estrutura molecular e pode ser avaliada com a constante de acoplamento (J , em Hz). A versão atual do *NMRDB* prevê esses valores e a comparação com os dados experimentais tabelados é uma atividade sugerida para avaliação crítica dos resultados calculados.

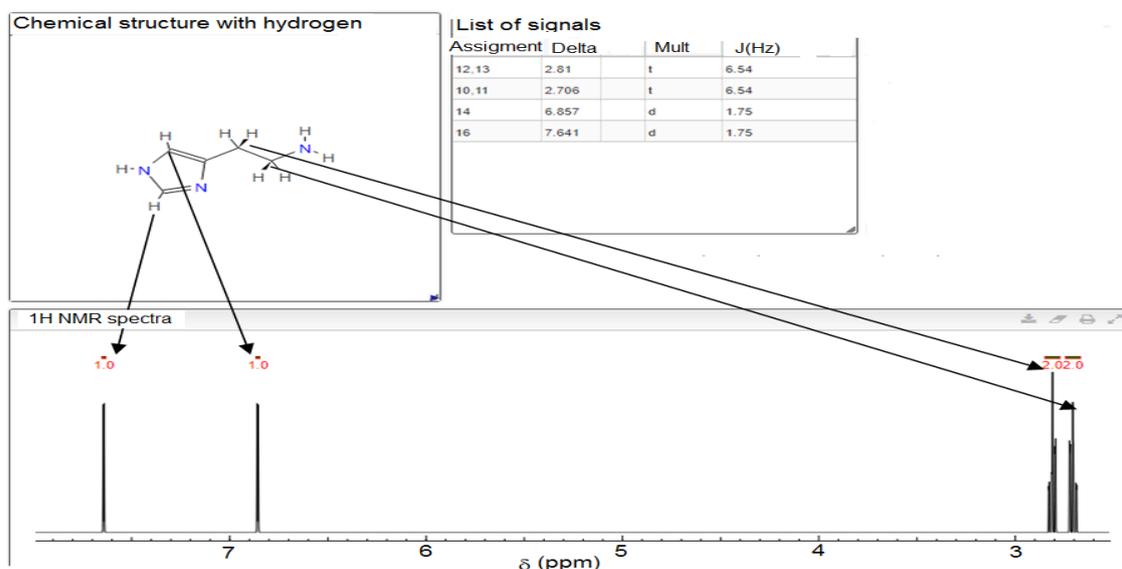


Figura 3. Espectro de RMN- ^1H para histamina (400 MHz) com a versão atual do programa.

Em termos estruturais, o grau de complexidade cresce da acetilcolina, para a histamina, sendo a da oxitocina a mais problemática. A estrutura 3 D da oxitocina na Figura 4 foi gerada com o *Molview* [Bergwef 2015]. Esta molécula é a que a forneceu o maior desafio para a simulação, pois possui 66 prótons e uma estrutura cíclica não rígida. A limitação está na complexidade do espectro simulado e não no tempo para o cálculo, pois para as três moléculas os resultados são apresentados em segundos. Exceto para poucas semelhanças que estão em destaque na Figura 5, os espectros não correspondem ao dado experimental [Kellenbach *et al.* 2008]. Desse modo, foi verificado que a simulação para a oxitocina abre possibilidades para novos questionamentos.

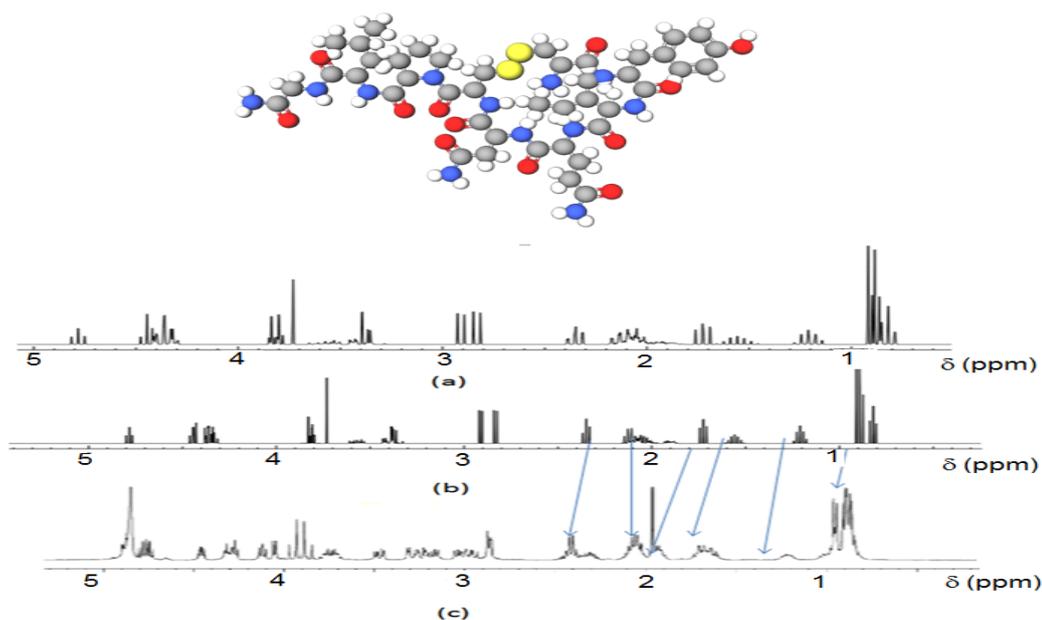


Figura 4. Espectros de RMN-¹H da oxitocina: a) simulado em 200 MHz, b) simulado em 400 MHz, c) experimental em 400 MHz. (As esferas mais claras na estrutura 3D representam os prótons da oxitocina).

De fato, quando as moléculas são maiores é comum partir para técnicas de RMN bidimensional e RMN-¹³C. É importante destacar que a versão atual do *NMRDB* já faz previsões de espectros 2D e de ¹³C, mas simular e analisar esses espectros está além dos objetivos deste trabalho.

Com respeito às atividades investigativas, foi observado que as características pertinentes desse recurso digital criam oportunidades para aplicar a taxonomia de Bloom [Anderson & Krathwohl 2001] integrada com os modelos de questões para o desenvolvimento do pensamento crítico [Barton 1997]. Isto deve facilitar a elaboração de perguntas de nível cognitivo elevado e que só podem ser respondidas com a análise dos resultados do *NMRDB*, por exemplo: a) “Que fatos você selecionaria para mostrar que a estrutura está relacionada com o espectro calculado”? b) “Como você resolveria o problema de interpretação espectral para esta molécula usando o que tem aprendido sobre a origem do sinal de RMN-¹H”? c) “Como você provaria que o espectro simulado corresponde à estrutura da molécula”? Conforme foi exemplificado para a acetilcolina e histamina, as duas primeiras questões podem ser respondidas com os princípios básicos de interpretação de RMN-¹H e com a simulação, mas a última requer o uso de dados experimentais, abrindo a possibilidade para novos questionamentos, comparando os dados disponíveis no *SDBS data base* [SDBS 2017] com os obtidos no *NMRDB*.

4. Considerações finais

O *NMRDB* ilustra como o desenvolvimento das RNAs contribui para melhorar as aplicações da análise de RMN. Este SE/OA exemplifica o uso integrado de ferramentas da informática química como o SMLIES, JSME e estimula a consulta de banco de dados espectrais, como o SDBS para verificação dos resultados da previsão. Tem o potencial para promover o desenvolvimento do pensamento químico, pois possibilita a visualização de fenômenos do universo quântico da origem do sinal de RMN e de fatos

experimentais essenciais, como o efeito da eletronegatividade nos deslocamentos químicos e a relação da estrutura com o desdobramento do sinal no espectro. Facilita fazer a conferência imediata das previsões feitas manualmente, estimulando as correções e questionamentos adequados para atividades investigativas, como foi exemplificado para as moléculas dos neurotransmissores, acetilcolina e histamina (com estruturas mais simples) e oxitocina (mais complexa e que favoreceu observar uma limitação do simulador). Com o *NMRDB* o estudo sobre RMN pode se tornar mais envolvente e interativo. Isto é tão próximo como ter a disposição um instrumento de RMN real.

Referências

- Aires-de-Sousa, J., Hemmer, M. and Gasteiger, J. (2002) “Prediction of $^1\text{H-NMR}$ Chemical Shifts Using Neural Networks”. In: *Analytical Chemistry*, v. 74, n.1, p. 80-90.
- Aires-de-Sousa, J. and Binev, Y. (2011) “SPINUS-WEB”, <http://www2.chemie.uni-erlangen.de/services/spinus/>. Acesso: julho 2017.
- Anderson, L.W. & Krathwohl, D. R. (Eds.) (2001). *A Taxonomy for Learning, Teaching and Assessing: A Revision of Bloom’s Taxonomy of Educational Objectives*, Addison Wesley/Longman.
- Banfi, D. and Patiny, L. (2008) “www.nmrdb.org: Resurrecting and processing NMR spectra on-line”. In: *Chimia*, v. 62, n. 4, p. 280-281.
- Barton, L. G.(1994). *Quick Flip Questions for Critical Thinking*, Edupress.
- Bergwef, H. (2015). Molview <https://confchem.cce.divched.org/2015FallCCCENLP9>, acesso: julho 2017.
- Bienfait, B. and Ertl, P. (2013). “JSME: a free molecule editor in JavaScript”. In: *Journal of Cheminformatics*. v.5, n.24, p. 2-6
- Binev, Y. and Aires-de-Sousa, J. (2004) “Structure-based predictions of $^1\text{H-NMR}$ chemical shifts using feed-forward neural networks”. In: *Journal of Chemical Information and Computer Sciences*. v. 44, n.3, p. 940–945.
- Binev, Y., Corvo M. and Aires-de-Sousa, J. (2004) “The Impact of Available Experimental Data on the Prediction of $^1\text{H-NMR}$ Chemical Shifts by Neural Networks”. In: *Journal of Chemical Information and Computer sciences*. v.44, n.3, p. 946–949.
- Binev, Y., Marques, M.M.B., and Aires-de-Sousa, J. (2007) “Prediction of $^1\text{H-NMR}$ Coupling Constants with Associative Neural Networks Trained for Chemical Shifts”. In: *Journal of Chemical Information and Modeling*. v.47, n.6, p. 2089–2097.
- Burns, J.A. and Whitesides, G.M. (1993) “Feed-Forward Neural Networks in Chemistry: Mathematical System for Classification and Pattern Recognition”. In *Chemical Review*, v. 98, n.8, p. 2583-2601.

- Castillo, A.M., Patiny L. and Wist, J. (2011) “Fast and Accurate Algorithm for the Simulation of NMR spectra of Large Spin Systems”. *Journal of Magnetic Resonance*, v. 209, n. 3, p. 123–130.
- Hastings, J., de Matos, P., Dekker, A., Ennis, M., Harsha, B., Kale, N., Muthukrishnan, V., Owen, G., Turner, S., Williams, M., and Steinbeck, C. (2013) “The ChEBI reference database and ontology for biologically relevant chemistry: enhancements for 2013”. In: *Nucleic Acids*, v. 41, D1, p. D456-D463.
- Gasteiger, J. and Zupan, J. (1993) “Neural Networks in Chemistry”. In: *Angewandte Chemie International Edition*, p. 503-527.
- Kellenbach, E., Sanders, K. and Overbeek, P.L.A. (2008) “The Use of Proton NMR as an Alternative for Amino Acid Identity Test for Peptides”. In: *NMR Spectroscopy in Pharmaceutical Analysis*, Edited by Ulrike Holzgrabe, Bernd Diehl and Iwona Wawer, Elsevier, Oxford.
- Machado, A.S. (2016) “Uso de Softwares Educacionais, Objetos de Aprendizagem e Simulações no Ensino de Química”. In: *Química Nova na Escola*, v. 38, n. 2, p.104-111.
- Meusinger, R. and Himmelreich, U. (2003). “Neural Networks and Genetic Algorithms Applications in Nuclear Magnetic Resonance Spectroscopy”. In: *Nature inspired Methods in Chemometrics: Genetic Algorithms and Neural Networks*, Edited by Riccardo Liardi, Elsevier, Oxford.
- Pavia, D. L., Lampman, G. M., Kriz, G. S. and Vyvyan, J. R (2010). “Espectroscopia de Ressonância Magnética Nuclear, Parte 1: Componentes Básicos”. In: *Introdução à Espectroscopia*. 4ª ed. São Paulo: Cengage Learning, p. 101-310.
- Skaf, M.S. (2013) “O Prêmio Nobel de Química 2013”. In: *Química Nova na Escola*, v.35, n. 4, p. 243-246.
- Spining, M.T., Darsey, J.A., Sumpter, B.G. and Noid, D.W. (1994) “Opening Up the Black Box of Artificial Neural Networks”. In: *Journal of Chemical Education*, p. 406-411.
- SDBSWeb (2017): “Spectral Data Base”, <http://sdbs.db.aist.go.jp>, acesso: julho 2017.
- Varnek, A. and Baskin, I.I. (2010) “Chemoinformatics as a Theoretical Chemistry Discipline”. In: *Molecular Informatics*, v. 30, p. 20-32.
- Wartha, E. J., Alves, L. C., Sá, L; P., Sanjuan, M. E. C. and Santos, C. V. (2010). “Uma proposta didática para a elaboração do pensamento químico sobre elemento químico, átomos, moléculas e substâncias”. In: *Experiências em Ensino de Ciências*. v. 5, n.1 p. 7-20.
- Weininger, D. (1988), “SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules”. In: *Journal of Chemical Information and Computer Sciences*, v. 28, n.1, p. 31-36.