

# SAMTa - Uma nova proposta para análise computacional de dinâmicas moleculares

Bernardo de Assis Carvalho<sup>1</sup>, Teodorico Castro Ramalho<sup>2\*</sup>, Elaine Fontes Ferreira da Cunha<sup>2</sup> e Daniel Guimarães do Lago<sup>1</sup>

<sup>1</sup>Departamento de Ciência da Computação, Universidade Federal de Lavras, Cx. Postal 3037, 37200-000, Campus Universitário, Lavras, Minas Gerais, Brasil. <sup>2</sup>Departamento de Química, Universidade Federal de Lavras, Lavras, Minas Gerais, Brasil.  
\*Autor para correspondência. E-mail: teo@dqj.ufla.br

**RESUMO.** Este artigo trata de uma síntese sobre a aplicação SAMTa (Sistema de Análise Molecular de Trajetórias). Aborda sobre a simulação de compostos moleculares, utilizando recursos computacionais, o método de seleção por análise matemática e sua proposta de solução pelo algoritmo de clustering K-Means. Apresenta resultados acerca do desenvolvimento e utilização do software, ilustrando com um estudo de caso de desenvolvimento de inibidores da EPSP sintase.

**Palavras-chave:** SAMTa, análise de dinâmicas moleculares, química computacional, análise matemática, algoritmo K-Means.

**ABSTRACT. SAMTa - A new proposal for computational analysis of molecular dynamics.** This article provides an overview of the implementation of the System for Molecular Analysis of Trajectories (SAMTa, acronym in Portuguese). It addresses the simulation of molecular compounds using computer approaches, the method of selection by mathematical analysis using K-Means. The results on the development and use of the software are illustrated with a case study concerning EPSP synthase inhibitors.

**Key words:** SAMTa, molecular dynamics analysis, computational chemistry, mathematical analysis, K-Means algorithm.

## Introdução

Simulação computacional tem se constituído numa ferramenta poderosa para a compreensão dos fenômenos físicos e químicos em sistemas da matéria condensada. Dentre as técnicas mais utilizadas de simulação computacional, encontra-se a Dinâmica Molecular, que estuda os movimentos em um sistema de partículas em função do tempo. Assim, com o objetivo de auxiliar químicos, físicos, biólogos e farmacêuticos, o SAMTa (Sistema de Análise Molecular de Trajetórias) visa à análise qualitativa de dados oriundos de uma simulação de um composto por meio do software GROMACS<sup>1</sup>. O GROMACS é um dos softwares já existentes mais usados pela comunidade acadêmica da área de Química para simulações de dinâmicas moleculares, por isso ele se tornou a aplicação-base para desenvolvimento do SAMTa.

A proposta de uma análise qualitativa foi descrita por Haiech et al. (1995). As ideias presentes em Coquet et al. (1996) e Ramalho et al. (2004) completaram a base teórica de projeto para o SAMTa, que

estabeleciam metodologias matemáticas detalhadas por Gleich e Zhukov (2004) para mineração dos dados obtidos por meio do GROMACS.

Essa mineração de dados elimina resultados desnecessários ao estudo, e o tratamento desses dados que utiliza metodologias matemáticas em recursos computacionais aprofundaria o estudo e proporcionaria maior confiança nos resultados obtidos da simulação feita nos dois softwares.

Dessa forma, o objetivo do nosso artigo é relatar o desenvolvimento de uma nova proposta teórica para análise de simulações de dinâmicas moleculares.

## Material e métodos

### Simulação de dinâmica molecular

Na simulação de um sistema molecular no software GROMACS, vários tipos de energias são envolvidas, que dão origem às conformações na molécula, modificando sua disposição no espaço durante o tempo. Essa mudança na disposição da molécula no espaço durante o tempo se faz pela variação de todas as energias que envolvem a simulação, influenciando diretamente nos resultados.

A proposta do SAMTa consiste em selecionar qualitativamente os quadros (momentos, tempos) da

<sup>1</sup>Disponível em: <http://www.gromacs.org>

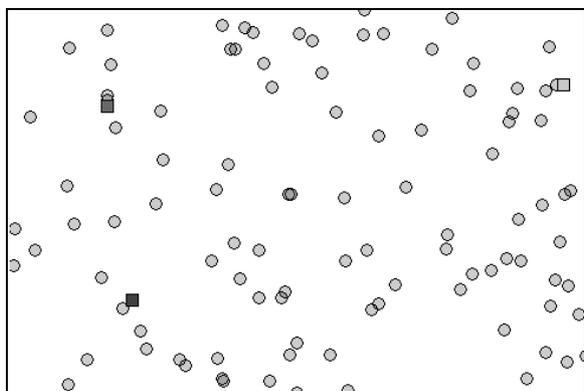
simulação, de duas formas: pelo deslocamento dos clusters no espaço da molécula, ou pelo valor da energia (força) desejada para a simulação, utilizando metodologias matemáticas.

Para efeito de seleção, existem dois métodos de comparação: (i) entre o primeiro quadro e o quadro atual da simulação ou (ii) entre o quadro anterior ao quadro atual e o quadro atual da simulação.

### O Algoritmo K-Means e sua aplicação no SAMTa

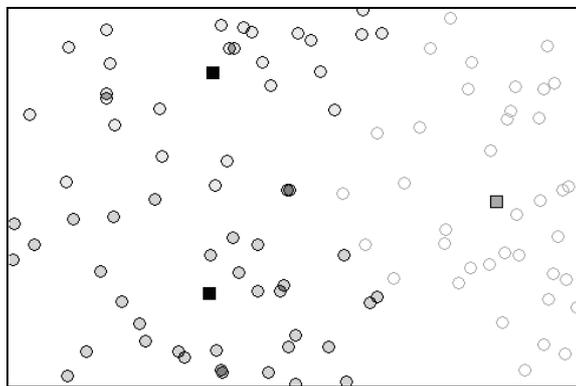
O Algoritmo K-Means é reconhecido na comunidade científica da Ciência da Computação como uma ferramenta para seleção de elementos em um espaço, utilizando metodologia matemática. A sua execução baseia-se na “clusterização” dos elementos presentes num determinado espaço. No caso dessa aplicação, no primeiro quadro da simulação os clusters são inicializados em pontos aleatórios no espaço da molécula (funcionando como um átomo fictício), sendo esses pontos posicionados posteriormente nos centros das maiores concentrações de átomos da molécula, de forma que todos os átomos sejam envolvidos pelos clusters presentes na molécula e que cada cluster possua pelo menos um átomo.

A Figura 1 ilustra a inicialização de três centros de clusters para análise de uma simulação com 100 átomos. Os centros dos clusters (átomos fictícios) são demonstrados por meio de quadrados, sendo um localizado na parte superior direita da figura, outro, na parte superior esquerda da figura e o último, na parte inferior esquerda da figura.



**Figura 1.** Inicialização de três clusters num espaço com 100 átomos para execução do algoritmo K-Means.

A Figura 2 ilustra o posicionamento final dos três clusters inicializados. Os centros dos clusters (identificados por quadrados) estão rodeados pelos átomos que pertencem ao seu cluster. De acordo com a coloração dos átomos disposta na figura, percebem-se quais átomos pertencem a que cluster.



**Figura 2.** Estado final da estrutura molecular mostrada na Figura 1 após execução do algoritmo K-Means.

Esse número de clusters é determinado pelo usuário da aplicação e limita-se, no mínimo, a três clusters e, no máximo, ao número de átomos da molécula menos 1.

Ao final do posicionamento dos clusters na molécula no primeiro quadro da simulação, as posições dos pontos dos clusters no espaço serão utilizadas no próximo quadro da simulação. De acordo com a movimentação dos átomos que pertencem ao cluster, o seu centro poderá ter sua posição modificada, e um átomo pertencente ao cluster pode trocar de cluster se o centro de algum outro cluster estiver mais próximo do átomo em questão. Essa diferença entre a posição do centro do cluster no primeiro quadro e no quadro seguinte será somada às outras diferenças de posição dos centros dos outros clusters presentes na molécula e serão a base dos resultados da análise para o método de seleção qualitativa por meio de conformações moleculares.

Para o método de comparação entre o primeiro quadro e o quadro atual da simulação, as posições dos centros de clusters presentes na molécula no primeiro quadro da simulação serão comparadas a todas as posições dos centros dos clusters presentes na molécula nos outros quadros da simulação. O fator que determina se uma conformação é significativa em relação à primeira, é o valor do somatório dos deslocamentos dos clusters entre o primeiro quadro e o quadro atual da simulação.

Para o método de comparação entre o quadro predecessor ao quadro atual e o quadro atual da simulação, as posições dos centros dos clusters presentes na molécula no quadro anterior ao quadro atual da simulação serão comparadas às do quadro atual da simulação, e o fator que determina se uma conformação é significativa em relação à anterior é o valor do somatório dos deslocamentos dos clusters entre o quadro anterior e o quadro atual da simulação.

Quanto maior o valor do deslocamento, maior a significância da conformação (Figura 3) na simulação.

O Algoritmo K-Means foi descrito detalhadamente por Cheung (2003) e Hartigan e Wong (1979). A aplicação proposta foi implementada com pequenas modificações em relação à proposta original para o algoritmo K-Means, com o objetivo de focalizar o software para a análise qualitativa de conformações moleculares. As mudanças implicam o tratamento dado aos resultados da execução do algoritmo, tais como o somatório dos deslocamentos dos clusters em cada quadro e a subtração pelo somatório dos deslocamentos dos clusters no quadro utilizado como referência (de acordo com o método de comparação escolhido).

O método de seleção por energias leva em consideração como fator determinante para a significância de uma conformação apenas o valor da energia escolhida para a simulação na aplicação. Quanto menor o valor da energia, maior a significância da conformação (Figura 3) na simulação. Para fins de informação e comparação de resultados, o algoritmo K-Means é executado, porém não participa como fator determinante para significância de qualquer conformação.

Os métodos de comparação para a seleção por energia funcionam comparando-se os valores das movimentações dos clusters entre os quadros, dependendo-se somente da escolha do método de comparação desejado. Portanto, fazem parte dos resultados aqueles que envolvem a execução do algoritmo K-Means, não influenciando no resultado da análise para este caso.

### Introdução da Lógica Fuzzy no SAMTa

Foi implementado um recurso auxiliar de detalhamento de resultados de análise executada no SAMTa, que utiliza a Lógica Fuzzy como base para determinar a porcentagem de significância da conformação molecular presente em cada quadro da simulação. A Lógica Fuzzy é descrita em Rao (1995), Gomide e Gudwin (1994) e Lee (1990). O usuário deverá fornecer os valores de limite mínimo e máximo em porcentagem para que uma conformação seja, ou não, significativa.

A aplicação utiliza os resultados da execução do algoritmo K-Means para quaisquer métodos de análise e de comparação e também os valores de todas as energias envolvidas na simulação para fazer o cálculo do grau de veracidade (possibilidade em porcentagem de o quadro ser uma conformação significativa) de cada quadro.

A partir do limite mínimo e máximo em porcentagem fornecida pelo usuário, a aplicação calculará os valores mínimos e máximos para movimentação dos clusters e energias e verificará quais quadros terão os graus de veracidade calculados. Primeiramente, será feita a análise sobre os valores de movimentação dos clusters. Se algum quadro tiver valor de movimentação dos clusters entre os valores mínimos e máximos, terá o grau de veracidade calculado. Se algum quadro tiver valor de movimentação dos clusters menor que o valor mínimo, o grau de veracidade será automaticamente 0. Se algum quadro tiver valor de movimentação dos clusters maior que o valor máximo, o grau de veracidade será automaticamente 1.

Os valores dos graus de veracidade para a movimentação dos clusters serão armazenados, para que sejam feitos os cálculos dos graus de veracidade dos quadros para as energias em cada quadro. Logo após será feita a média entre os valores dos graus de veracidade da movimentação dos clusters e das energias de cada quadro. O valor resultante dessa média será o grau de veracidade de cada quadro. O cálculo do grau de veracidade para as energias é semelhante ao da movimentação dos clusters.

Se algum quadro tiver valor de energia entre os valores mínimos e máximos, terá o grau de veracidade calculado. Se algum quadro tiver valor de energia menor que o valor mínimo, o grau de veracidade será automaticamente 0. Se algum quadro tiver valor de energia maior que o valor máximo, o grau de veracidade será automaticamente 1.

## Resultados e discussão

### Estatu do desenvolvimento

O SAMTa foi desenvolvido com a intenção de implementar métodos modernos de seleção e clusterização ainda utilizados para mineração de dados de dinâmica molecular (DM). Podem-se citar algumas vantagens dessa aplicação:

- interação com a aplicação GROMACS para conversão de arquivos relativos à simulação no GROMACS em arquivos para análise no SAMTa;
- interface gráfica amigável desenvolvida para ambiente JAVA, possibilitando a execução do SAMTa em qualquer sistema operacional que possua a máquina virtual JAVA instalada;
- leitura de arquivos no formato PDB (Relativo à simulação da molécula) e no formato XVG (Relativo às energias envolvidas na simulação da molécula), todos originados a partir do GROMACS;
- efetuação da análise da simulação, usando-se métodos matemáticos, comparando-se o primeiro

quadro com o quadro atual da simulação ou o quadro predecessor ao quadro atual com o quadro atual da simulação;

- lista do resultado da análise em ordem decrescente no painel da janela principal da aplicação;

- disponibilização da interpretação do resultado da análise por meio de um gráfico, com possibilidade de seleção do quadro desejado no gráfico e a localização automática desse quadro no painel da janela principal da aplicação;

- possibilidade de se salvar o quadro desejado da simulação pela seleção do quadro no painel da janela principal da aplicação, com o objetivo de visualizá-lo em um software dotado de dispositivos de visão em 3D (VMD, Pymol, entre outros);

- ajuda para a familiarização de um novo usuário com a interface gráfica e as funções da aplicação.

### Outros detalhes referentes à implementação

Quanto à manipulação dos dados: a aplicação interage com o usuário de forma que o mesmo efetue a entrada de dados por meio de leitura de arquivos. Para uma análise, o usuário fornece um arquivo relativo à estrutura da molécula no tempo de simulação e um arquivo relacionado com as energias envolvidas no tempo de simulação.

A aplicação, inicialmente, foi desenvolvida com a função de armazenar todos os dados em memória principal, para um acesso mais rápido e eficaz dos dados. Em contrapartida, constatou-se que, pela extensão dos dados armazenados relativos apenas à molécula de sistema, casos especiais de simulações não são executados em computadores com baixa capacidade de processamento e armazenamento de dados.

Quanto à interface gráfica: a aplicação foi desenvolvida com o objetivo de simplificar a interação com o usuário da maior forma possível.

### Estudo de caso aplicado no SAMTa

No intuito de se avaliar a performance do SAMTa em sistemas reais, empregaram-se as técnicas implementadas na análise de simulações moleculares da EPSP sintase e inibidores.

A enzima 5-enolpiruvil shikimato-3-fosfato sintase (EPSP sintase) participa do sexto caminho do shikimato e procede a um ponto intermediário importante, o clorismato, o qual é requerido para a síntese de metabólitos aromáticos incluindo os aminoácidos aromáticos essenciais fenilalanina, tirosina e triptofano (BERMAN et al., 2000). O glifosato é um herbicida pertencente ao grupo químico das glicinas substituídas e é o inibidor da EPSP sintase (GALLI; MONTEZUMA, 2005). É

indicado no controle de ervas daninhas anuais e perenes, monocotiledôneas ou dicotiledôneas, em culturas de arroz irrigado, cana-de-açúcar, café, citros, maçã, milho, pastagens, soja (plântio direto ou indireto), fumo, uva e soqueira em cana-de-açúcar. É indicado, ainda, para as culturas de ameixa, banana, cacau, nectarina, pera, pêssego, seringueira e plântio direto do algodão. Pode, ainda, ser aplicado na água para o controle das ervas aquáticas (AMARANTE JÚNIOR et al., 2002).

Dada a importância do desenvolvimento de herbicidas mais ativos, este trabalho também tem por objetivo estudar o modo de interação entre o glifosato e três análogos com a enzima EPSP sintase da *E. Coli*, nativa e mutante Gly96Ala, utilizando a técnica de dinâmica molecular.

Os arquivos fornecidos para análise fazem parte de simulações do glifosato na enzima EPSP sintase em sua forma nativa e mutante. Assim, nossa análise tem como objetivos identificar e investigar os quadros da simulação em que houve maior movimentação dos clusters para que estes sejam analisados em software de visualização gráfica 3D. Dessa forma, pode-se possibilitar a investigação de quadros da simulação que tornam o glifosato um inibidor de grande potencial de ação.

Neste artigo, apresentaremos os resultados de uma análise, utilizando o método de análise matemática (utilizando algoritmo K-Means) com seleção por conformações, método de comparação do primeiro quadro da simulação e a utilização de três clusters.

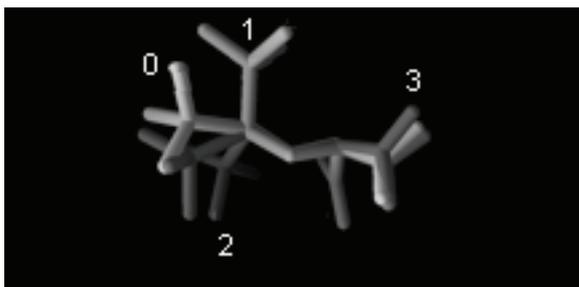
Os resultados referentes às análises podem ser visualizados na Tabela 1.

**Tabela 1.** Resultados da análise da simulação do glifosato em sua forma nativa e mutante pelo deslocamento dos clusters.

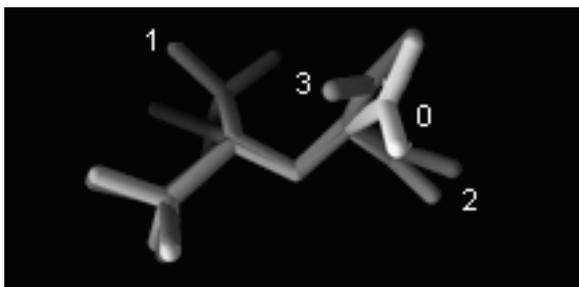
Análise de Simulação de DM por Deslocamento dos Clusters			
Molécula Nativa			
	Tempo	Deslocamento	Energia Total
Mais Alto	3340	5,516072058	-414234,188
Médio	1710	2,758173186	-413812,13
Mais Baixo	700	0,567216525	-414243,31
Molécula Mutante			
	Tempo	Deslocamento	Energia Total
Mais Alto	930	7,337448379	-413917,22
Médio	1790	2,218181218	-415642,38
Mais Baixo	2500	0,380710619	-414838,31

Comparando-se os quadros em que houve a movimentação mais alta dos clusters na forma nativa e mutante da EPSP sintase, verifica-se diferença positiva para a molécula mutante. Do ponto de vista molecular, essa diferença é pertinente, pois a enzima mutante tem menor volume para a cavidade do sítio ativo, o que resulta em menor movimentação da molécula pelo impedimento estérico e pela maior dificuldade na interação do grupo carboxilato com os átomos da enzima EPSP Sintase.

As Figuras 3 e 4 contêm as sobreposições das moléculas do glifosato na enzima EPSP sintase nativa e mutante, o quadro inicial (indicado pelo número 0) e os quadros de movimentação mais baixa (indicado pelo número 1), média (indicado pelo número 2) e mais alta (indicado pelo número 3). Nelas, pode-se visualizar a mudança de ângulo dos grupos carboxilato e fosfato, que realizam as interações com a enzima EPSP Sintase.



**Figura 3.** Sobreposição das moléculas do glifosato na forma nativa nos quadros da simulação selecionados pela análise por conformações efetuada no SAMTa.



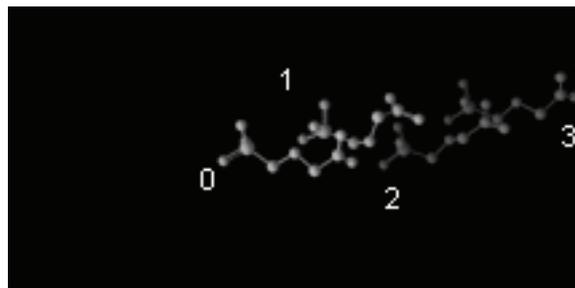
**Figura 4.** Sobreposição das moléculas do glifosato na forma mutante nos quadros da simulação selecionados pela análise por conformações efetuada no SAMTa.

Observando as Figuras 3 e 4 e comparando-se as moléculas com a molécula do quadro 2 (zero) inicial, percebe-se que as sobreposições consideram apenas as deformações das moléculas e não consideram as rotações e as translações que as moléculas fazem em tempo de simulação. Analisando-se as Figuras 5 e 6, a percepção sobre a movimentação das moléculas do glifosato é bem diferente. Elas mostram todos os movimentos que as moléculas fazem (rotação, translação e deformação) em comparação com o quadro inicial.

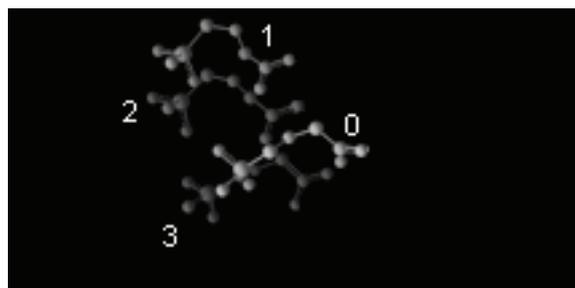
As diferenças nas posições das moléculas nas figuras têm por objetivo facilitar a visualização dos movimentos de rotação, translação e deformação das moléculas, principalmente dos grupos fosfato e carboxilato.

A Figura 5 deixa clara a eficácia da análise, já que no quadro de maior movimentação (molécula 3) ocorreu grande translação e uma rotação na região

do grupo fosfato, em relação ao quadro inicial (Figura 4) (molécula 0). O quadro de movimentação média (molécula 2) teve translação, rotação na região do grupo fosfato e leve rotação na região central da molécula, comparando-se ao quadro inicial (Figura 4). O quadro de menor movimentação (molécula 1) teve leve translação, com rotações nas regiões do grupo fosfato, central e carboxilato. As deformações apresentadas na Figura 3 indicam que o quadro de menor movimentação (molécula 1) teve forte deformação no grupo fosfato e carboxilato. Aparentemente, as deformações não somaram grandes distâncias, o que leva a crer que a deformação é um movimento de fraca influência nos valores de posicionamento dos átomos no espaço.



**Figura 5.** Estados reais das moléculas do glifosato na forma nativa no espaço, nos quadros selecionados pela análise por conformações efetuada no SAMTa.



**Figura 6.** Estados reais das moléculas do glifosato na forma mutante no espaço, nos quadros selecionados pela análise por conformações efetuada no SAMTa.

A Figura 6 já apresenta uma análise com compreensão dificultada. O quadro de maior movimentação (molécula 3) teve leve translação, rotações dos grupos fosfato e carboxilato e pequena deformação em relação ao quadro inicial (molécula 0). O quadro de movimentação média (molécula 2) teve translação considerável, deformação na área central da molécula e rotações do carboxilato e do fosfato em relação ao quadro inicial. O quadro de menor movimentação (molécula 1) teve forte translação, deformação na área central da molécula e rotações nos grupos carboxilato e fosfato, em relação ao quadro inicial.

O fator-chave determinante para que o quadro de maior movimentação (molécula 3) tenha realmente a maior movimentação em comparação aos outros quadros é a posição como a molécula se encontra no espaço. A visualização em foto permite visão muito limitada em três dimensões.

Na Figura 4, novamente se visualiza deformação das moléculas de movimentação média (molécula 2) e baixa (molécula 1), respectivamente. Como na Figura 3, aparentemente, as deformações não influenciaram nos valores de posicionamento dos átomos no espaço significativamente.

### Conclusão

Os sucessivos testes levaram a equipe de desenvolvimento a concluir que os resultados obtidos são pertinentes ao objetivo do software e ao objetivo do estudo, e os quadros selecionados nessa análise serão estudados de forma a identificar as propriedades do glifosato que o torna um inibidor de grande potencial na configuração encontrada.

A equipe de desenvolvimento pretende completar a aplicação com um novo método de análise implementado por meio das Redes Neurais Artificiais, um recurso computacional que dá poder à máquina de solucionar problemas, buscando o máximo de precisão, comparando com os neurônios humanos.

### Referências

- AMARANTE JÚNIOR, O. P.; SANTOS, T. C. R.; BRITO, N. M.; RIBEIRO, M. L. Glifosato: Propriedades, toxicidade, usos e legislação. **Química Nova**, v. 25, n. 4, p. 589-593, 2002.
- BERMAN, H. M.; WESTBROOK, J.; FENG, Z.; GILLILAND, G.; BHAT, T. N.; WEISSIG, H.; SHINDYALOV, I. N.; BOURNE, P. E. The protein data bank. **Nucleic Acids Research**, v. 28, n. 28, p. 235-242, 2000.
- CHEUNG, Y. K. Means: a new generalized K-means clustering algorithm. **Pattern Recognition Letters**, v. 24, n. 1, p. 2883-2893, 2003.
- COQUET, R.; TROXLER, L.; WIPFF, G. The STATIS method: characterization of conformational states of flexible molecules from molecular dynamics simulations in solution. **Journal of Molecular Graphics**, v. 14, n. 1, p. 206-212, 1996.
- GALLI, A. J. B.; MONTEZUMA, M. C. **Alguns aspectos da utilização do herbicida glifosato na agricultura**. São Paulo: Monsanto do Brasil, 2005.
- GLEICH, D.; ZHUKOV, L. **Soft clustering with projections: PCA, ICA, and Laplacian**. New York: Research Labs, 2004.
- GOMIDE, F. A. C.; GUDWIN, R. R. Modelagem, controle, sistemas e lógica Fuzzy. **SBA Controle e Automação**, v. 4, n. 3, p. 97-115, 1994.
- HAIECH, J.; KOSCIELNIAK, T.; GRASSY, G. Use of TSAR as a new tool to analyze the molecular dynamics trajectories of proteins. **Journal of Molecular Graphics**, v. 13, n. 1, p. 46-48, 1995.
- HARTIGAN, J. A.; WONG, M. A. Algorithm AS 136: a K-Means clustering algorithm. **Applied Statistics**, v. 28, n. 1, p. 100-108, 1979.
- LEE, C. C. Fuzzy logic in control systems: fuzzy logic controller, part I and II. **IEEE Transactions on Systems, Man and Cybernetics**, v. 20, n. 14, p. 404-435, 1990.
- RAMALHO, T. C.; CUNHA, E. F. F.; ALENCASTRO, R. B. Solvent effects on  $^{13}\text{C}$  and  $^{15}\text{N}$  shielding tensors of nitroimidazoles in the condensed phase: a sequential molecular dynamics/quantum mechanics study. **Journal of Physics: Condensed Matter**, v. 16, n. 16, p. 6159-6170, 2004.
- RAO, V. B. **C++ Neural networks and fuzzy logic**. Foster City: IDG Books Worldwide, 1995.

*Received on April 10, 2008.*

*Accepted on September 10, 2009.*

License information: This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.