

QUÍMICA COMPUTACIONAL 2010/2011

Bloco de Químio-informática

A. Organização e avaliação

Neste bloco (Químio-informática, 1ª metade do semestre) utilizamos a metodologia *Team-Based Learning* (<http://tblc.camp9.org/>). Este é um método de aprendizagem activa em que são minimizados os tempos de “aula teórica” tradicional (em que os alunos ouvem passivamente a matéria antes de a ter estudado). Ao contrário, neste método, antes de cada aula os alunos terão que estudar a matéria que vai ser tratada na aula, a partir de material disponibilizado previamente. O tempo de aula é utilizado para realizar actividades que não podem ser feitas por cada aluno sozinho, como sejam esclarecer dúvidas com o professor ou com os colegas, realizar tutoriais acompanhados, ou participar em actividades de aplicação que requerem colaboração em equipa.

Grande parte das actividades na aula são realizadas em equipa. Os alunos são divididos em equipas pelo docente de modo a que as equipas sejam equilibradas e contenham diversidade máxima.

A primeira actividade a realizar em cada aula é um Teste para Garantir a Preparação individual (TGPI): é um teste de escolha múltipla, individual, sem consulta, que conta para a nota final e que pretende confirmar que o aluno domina os conteúdos principais do capítulo para poder realizar actividades de aplicação. Depois do TGPI ter sido respondido e entregue ao Professor, o mesmo teste é realizado em equipa (TGPe) e a chave é conhecida imediatamente. As equipas que não concordarem com a solução de alguma questão devem elaborar um recurso no momento. No final, o Professor resolve o teste para todos, discute dúvidas e faz uma “mini aula teórica” onde reforça os pontos mais difíceis do capítulo.

Na 2ª parte de cada aula, os alunos têm algum tempo para resolver tutoriais (problemas com resolução guiada) no computador, individualmente ou em grupo de 2 pessoas consoante a disponibilidade de computadores.

Depois dos tutoriais, na mesma aula ou na aula seguinte, é dado um problema de aplicação exigente para ser resolvido em equipa no tempo de aula.

Durante as aulas não é assegurado apoio na instalação de *software* em *laptops* privados, nem na utilização de sistemas operativos / programas diferentes dos instalados nos computadores da sala de aula (ver lista no documento “Software”).

A avaliação tem em conta as actividades das aulas (avaliação contínua, 60%) e a avaliação final (40%).

As actividades das aulas que contam para avaliação são

- Testes TGP individuais e de equipa
- Actividades de resolução de problemas em equipa
- Mini-seminário a apresentar na última aula do bloco
- Avaliação pelos colegas da equipa (no final do bloco, dentro de cada equipa cada membro distribui 100 pontos pelos restantes membros de acordo com o contributo que considerou que cada um deu para a obtenção de bons resultados).

O mini-seminário é realizado em equipa, a sua apresentação na última aula do bloco tem a duração de 20 min, seguida de perguntas. Os mini-seminários são apresentados obrigatoriamente através de vídeo-conferência, sendo fornecidos os meios necessários. Os trabalhos para os mini-seminários são atribuídos pelo docente a cada equipa, podendo cada equipa escolher entre três possibilidades: a) apresentação e discussão de um artigo científico envolvendo tópicos da disciplina; b) resolução de um problema; c) demonstração de um *software* de Químio-informática.

A nota que cada aluno recebe das actividades das aulas é a média dos TGP individuais (50%) e dos resultados da equipa (com pesos 25% para TGPs de equipa, 60% para problemas de aplicação e 15% para mini-seminário). A nota da equipa que conta para cada membro é ajustada com a avaliação que os colegas fazem (nota da equipa x soma de pontos dados pelos colegas /100).

Com base nas notas da resolução de TGPs de equipa, de problemas de aplicação e do mini-seminário é feita uma ordenação (*ranking*) das equipas. No final do bloco, a equipa vencedora tem um bónus de 1 valor na sua nota e a equipa perdedora tem um desconto de 0,5 valor na nota.

A avaliação final é feita por exame escrito na época de exames.

B. Resumo do Programa

1. Introdução à Químio-informática

2. Representação da estrutura molecular e reacções. Necessidade e estratégias para a representação de compostos químicos. A notação linear SMILES. Os formatos MDL Molfile e Sdfile. *Software* para a interconversão de ficheiros e para a standardização de estruturas. *Hashed fingerprints e hash codes*. Definições de similaridade entre moléculas. Representação de estruturas 3D. Representação de reacções nos formatos SMILES e MDL Rdf file.

3. Descritores moleculares. Descritores constitucionais. Descritores de fragmentos. Descritores topológicos. O índice de Wiener. Vectores de autocorrelação 2D. Descritores 3D. Funções de distribuição radial. *Software* para o cálculo de descritores moleculares.

4. Introdução às relações quantitativas estrutura-propriedade (QSPR) e estrutura-actividade (QSAR). Selecção de conjuntos de treino, validação e previsão. Selecção de descritores. Treino e avaliação de modelos.

5. Métodos para análise de dados. Regressões multilíneas. Árvores de decisão. Redes neuronais de *back-propagation*. Redes neuronais de Kohonen. Redes neuronais de *counterpropagation*.

C. Bibliografia

- *Chemoinformatics - a Textbook*, Gasteiger, J. Engel, T., Eds.; Wiley-VCH: Weinheim, 2003.
- Leach, A. R.; Gillet, V. J. *An Introduction to Chemoinformatics*, 2^a ed.; Springer: Dordrecht, 2007.
- Handbook of Chemoinformatics, Johann Gasteiger, Wiley-VCH, 2003.
- **Moodle:** <http://moodle.fct.unl.pt/course/view.php?id=2753>

D. Definição de Químio-informática

A Químio-informática é uma área científica que utiliza metodologias informáticas para resolver problemas de Química frequentemente associados à utilização de informação sobre estruturas moleculares. Apesar de ter um nome recente, a Químio-informática tem uma história de mais de 40 anos, criou revistas científicas próprias bem

estabelecidas, produziu livros de referência, foi incorporada em variados programas curriculares e teve um enorme impacto industrial nomeadamente na indústria farmacêutica ao nível dos processos de descoberta de novos fármacos. Também hoje o acesso de todos os químicos a vastíssimas fontes de informação envolvendo estruturas moleculares é suportado por infra-estruturas químio-informáticas frequentemente imperceptíveis.

Apesar de para alguns autores a Químio-informática englobar a aplicação de cálculos teóricos de mecânica quântica, na prática esta última tem permanecido sob a designação de “Química Computacional”, “Química Teórica” ou mesmo “Modelação Molecular”. Exemplos de aplicações típicas da Químio-informática são a gestão de bases de dados químicas, análise de grandes conjuntos de dados experimentais associados a estruturas moleculares, utilização de métodos estatísticos e de aprendizagem automática para a previsão de actividades biológicas ou outras propriedades observáveis (QSAR/QSPR), aplicação de métodos de inteligência artificial para elucidação estrutural, ou a visualização de estruturas e informação química.

E. Objectivos Globais do Bloco

A frequência cadeira deverá contribuir para que o aluno:

- Compreenda a necessidade de utilizar métodos informáticos para processar grandes volumes de informação química.
- Compreenda a vantagem de transformar uma estrutura química numa linguagem para a representação e processamento por um computador.
- Aprenda as principais estratégias para a representação computacional de estruturas moleculares e reacções químicas.
- Aprenda a representar aspectos específicos da estrutura molecular por descritores moleculares.
- Adquira competências para a utilização de técnicas de aprendizagem automática.
- Adquira competências para a aplicação da metodologia QSAR/QSPR – relações quantitativas estrutura-actividade e estrutura-propriedade.
- Reconheça a versatilidade das metodologias computacionais estudadas e aprenda a reduzir problemas concretos a formatos tratáveis pelas mesmas.
- Reconheça as limitações dos métodos aprendidos e a necessidade de novos desenvolvimentos.
- Consolide conceitos fundamentais de Química (como estereoquímica ou influência de aspectos estruturais nas propriedades físicas dum composto).
- Reflicta sobre as possibilidades e limitações da construção de modelos preditivos em Química.
- Desenvolva competências para aceder à literatura de Químio-informática e para aprender autonomamente novos tópicos nesta área.
- Desenvolva capacidades de multidisciplinaridade (por exemplo em estudos de QSAR).
- Desenvolva competências transferíveis no domínio das tecnologias de informação.

F. Objectivos Específicos do Bloco

Representação da estrutura molecular

- Aprender as regras fundamentais da notação linear SMILES
- Ser capaz de codificar uma estrutura molecular simples (<10 átomos, hidrogénios excluídos) em SMILES
- Ser capaz de desenhar a fórmula de estrutura de uma molécula simples (<10 átomos, hidrogénios

excluídos) a partir da representação SMILES

- Reconhecer a necessidade de representações canónicas de estruturas moleculares
- Interpretar estruturas de Markush
- Aprender a construir as matrizes de adjacências, ligações e distâncias dada uma estrutura molecular
- Aprender a interpretar tabelas de conectividade
- Aprender a interpretar os formatos MDL Molfile e SDfile
- Aprender a utilizar *software* para a interconversão de ficheiros e para a standardização de estruturas
- Aprender a representar estruturas moleculares por *hashed fingerprints*
- Aprender a utilizar *software* para a geração rápida de modelos moleculares 3D e seu arquivo em formatos moleculares MDL Molfile e PDB

Análise de diversidade molecular

- Reconhecer a utilidade de análises de diversidade molecular no planeamento de experiências laboratoriais de *screening*
- Aprender a calcular distâncias entre moléculas através do coeficiente euclidiano e coeficiente de Tanimoto com base em *hashed fingerprints*

Representação de reacções químicas

- Aprender a interpretar e utilizar representações SMILES de reacções químicas
- Aprender a interpretar ficheiros no formato MDL RDfile, incluindo a especificação do centro reaccional e mapeamento átomo-a-átomo

Descritores moleculares

- Compreender o que são descritores moleculares
- Distinguir entre diferentes tipos de descritores moleculares, nomeadamente descritores constitucionais, descritores de fragmentos, descritores topológicos e descritores 3D
- Aprender a definição de alguns descritores representativos como o índice de Wiener, vectores de autocorrelação 2D e funções de distribuição radial 3D
- Aprender a usar *software* para o cálculo de descritores moleculares

Introdução às relações quantitativas estrutura-propriedade e estrutura-actividade (QSPR e QSAR)

- Aprender o que são relações quantitativas estrutura-propriedade e estrutura-actividade (QSPR e QSAR)
- Compreender os vários passos para o estabelecimento de QSPR e QSAR
- Aprender a seleccionar descritores para um modelo QSPR/QSAR

Métodos para análise de dados

- Aprender a utilizar regressões multilíneas para QSPR/QSAR
- Familiarizar-se com os aspectos fundamentais do algoritmo e com a estrutura de árvores de decisão
- Compreender o funcionamento de redes neuronais de *back-propagation*, Kohonen e de *counterpropagation*
- Ser capaz de utilizar *software* para treinar e aplicar árvores de decisão e redes neuronais.
- Distinguir entre aprendizagem supervisionada e não-supervisionada

- Aprender a aplicar técnicas de aprendizagem automática a problemas de Química