

ANÁLISE E PROCESSAMENTO DE DADOS ESTRUTURAIS E TERMODINÂMICOS EM SOLUÇÃO

Rogério Ribeiro Macêdo¹ (IC), Juliana Fedoce Lopes (PQ)¹

¹Universidade Federal de Itajubá.

Palavras-chave: química computacional, processamento de dados, python.

Introdução

Estudos em Química Computacional (simulação computacional), como aqueles envolvendo o estudo de propriedades físico-químicas em solução geram uma grande quantidade de dados. Por esse motivo, o desenvolvimento de aplicações para extrair, tratar e analisar dados é uma fase do trabalho neste campo de estudo, pois, todo o processo de simulação inclui a escolha de modelos, sua implementação e a visualização e o estudo dos dados gerados (WINSBERG, 2018).

Metodologia

Um dos primeiros passos ao se desenvolver uma aplicação computacional para algum problema é a escolha da linguagem a ser utilizada na escrita do mesmo. Atualmente há uma grande diversidade de linguagens de programação que são usadas para o desenvolvimento de aplicações web, desktop ou até mesmo para dispositivos como celulares e eletrodomésticos. Neste caso em específico, por se tratar de programação científica, escolheu-se a linguagem de programação Python. Seu uso, quando comparada com outras linguagens, é justificado pela simplicidade, facilidade de aprender e, como é bastante utilizada na área de programação científica, tem ao seu dispor uma grande quantidade de bibliotecas científicas e fóruns públicos de compartilhamento de informações (OLIVEIRA, 2016).

Resultados e discussão

A realização deste trabalho resultou no desenvolvimento de três scripts em Python: `laqc_mod.py`, `laqc_spectrum.py` e `laqc_pop.py`. Todos eles são executados via linha de comando, Windows ou Linux (ambientes de trabalho presentes no Laboratório de Química Computacional da Unifei), como exemplificado na figura 1:

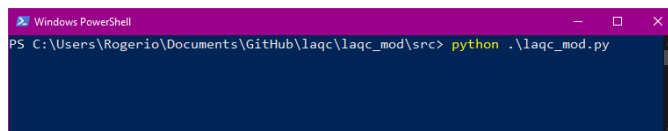


Figura 1 - Executar script python (Windows)

O script `laqc_mod.py`, realiza modificações no arquivo de input do Gaussian e pode ser usado para modificar apenas um arquivo ou vários (em lote). Seu uso envolve utilizar um arquivo de configuração de entrada (input) que deve ser construído pelo usuário informando o que será alterado ou, caso não se deseje usar tal arquivo, o próprio script irá questionar o usuário sobre as mudanças a serem realizadas. Abaixo, um exemplo de um arquivo de input do Gaussian antes e depois da execução do script.

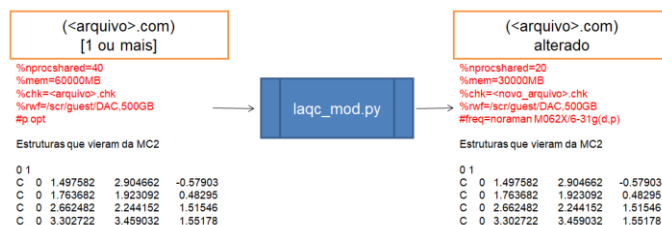


Figura 2 - Uso do `laqc_mod.py`

O exemplo representado na Figura 2 é resultado da execução do script `laqc_mod.py` de acordo com as respostas do usuário. Essas alterações em lote são extremamente importantes pois são processados muitos arquivos com coordenadas cartesianas diferentes que descrevem os sistemas químicos em solução, e, portanto, para cálculos subsequentes precisariam de edições manuais e individuais. O script então, além de evitar eventuais erros do usuário nesse processo de criação dos arquivos, otimiza o tempo de trabalho e de pesquisa.

Na Figura 3, é apresentada os arquivos de saída após a execução do script.

```

Windows PowerShell
PS C:\Users\Rogério\Documents\GitHub\laqc\laqc_mod\src> python .\laqc_mod.py

-----
UNIFEI - Universidade Federal de Itajubá
LaQC - Laboratório de Química Computacional
-----

obs 1: digite [sair] para encerrar o programa
obs 2: não informar valor considera que o parâmetro não será modificado.

-----
Modificação em lote dos arquivo de input do Gaussian
-----

Usar arquivo de configuração? [N].....: n
Local de origem dos arquivos.....: C:\Users\Rogério\Documents\exemplos_input_com
Local de destino dos arquivos.....: C:\Users\Rogério\Documents\exemplos_input_destino
Parâmetros de cálculo.....: #Freq=noraman M062X/6-31g(d,p)
Número de processadores.....: 20
Quantidade de memória.....: 30000MB
RWF.....:
Uso de camadas para cálculo ONIOM? [N].....: N
Congelar átomos ('freeze')? [N].....: N
Criar script de execução? [S].....: s

Processando arquivos...
- Modificando arquivos: mc20000080.com
- Modificando arquivos: mc20000160.com
- Modificando arquivos: mc20000240.com
- Modificando arquivos: mc20000320.com
    
```

Figura 3 - Execução do laqc_mod.py

A geração de espectros de UV/Vis é uma parte do trabalho de Química Computacional em alguns projetos e seu resultado advém do cálculo de estados excitados das estruturas moleculares. No entanto, pode ser necessária a visualização e interpretação de vários espectros simultaneamente e também o cálculo de um espectro médio. Nesse sentido, o script laqc_spectrum.py foi escrito para suprir tal necessidade. A primeira fase é a geração de um arquivo de entrada (input.dat) contendo os valores dos estados excitados das estruturas de referências, como mostrado na Figura 4:

```

Windows PowerShell
PS C:\Users\Rogério\Documents\GitHub\laqc\laqc_spectrum\src> python .\laqc_spectrum.py

-----
UNIFEI - Universidade Federal de Itajubá
LaQC - Laboratório de Química Computacional
-----

>>> digite [sair] para encerrar o programa <<<

-----
Gráfico UV-VIS de cálculo dos Estados Excitados
-----

O input.dat não existe. Deseja gerá-lo? (S or N) [N].....: s
Origem do arquivo de saída (Gaussian ou Orca) [Gaussian]:: Gaussian
Local dos arquivos.....: C:\Users\Rogério\Documents\dados_debora2
- Extraindo estado excitado do arquivo: poptd_mc_20000160.log
- Extraindo estado excitado do arquivo: poptd_mc_20000240.log
- Extraindo estado excitado do arquivo: poptd_mc_20000480.log
- Extraindo estado excitado do arquivo: poptd_mc_20000560.log
- Extraindo estado excitado do arquivo: poptd_mc_20000720.log
- Extraindo estado excitado do arquivo: poptd_mc_20000960.log
    
```

Figura 4 - Execução do laqc_spectrum.py (parte 1)

A segunda fase é a deconvolução dos dados de input.dat em um arquivo contendo os valores para geração do gráfico de UV/Vis, como mostrado na Figura 5:

```

Windows PowerShell
PS C:\Users\Rogério\Documents\GitHub\laqc\laqc_spectrum\src> python .\laqc_spectrum.py

-----
UNIFEI - Universidade Federal de Itajubá
LaQC - Laboratório de Química Computacional
-----

>>> digite [sair] para encerrar o programa <<<

-----
Gráfico UV-VIS de cálculo dos Estados Excitados
-----

Tipo do ajuste (gaussian or lorentzian) [gaussian].....: gaussian
Número de onda (100-800) [100-800].....: 100-800
Intervalo dos números de onda [10,0].....: 0.5
Tipo de média (aritmética) [aritmética].....: aritmética

Arquivo spectrum_gaussian.dat gerado!
Arquivo average_spectrum.dat gerado!
    
```

Figura 5 - Execução do laqc_spectrum.py (parte 2)

Por último os dados gerados pelo laqc_spectrum.py podem ser visualizados no programa gnuplot como apresentado na Figura 6:

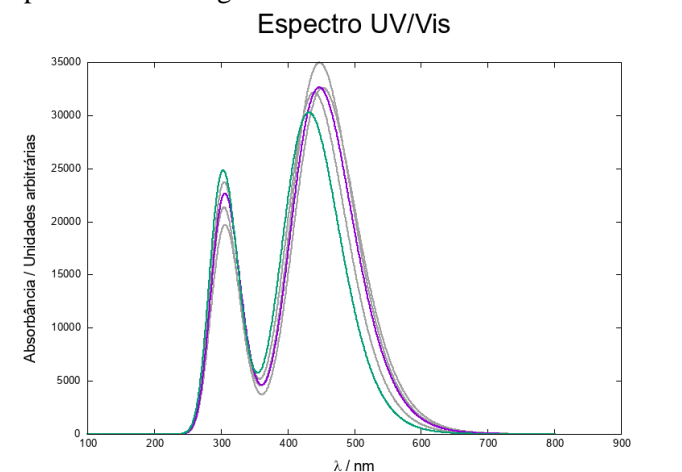


Figura 6 - Espetro gerado pelo laqc_spectrum.py

Por último, temos o laqc_pop, que tem por objetivo extrair valores de energia livre de Gibbs (“Sum of electronic and thermal Free Energies”) dos arquivos de log do Gaussian e assim gerar uma análise populacional das estruturas.

```

Windows PowerShell
PS C:\Users\Rogério\Documents\GitHub\laqc\laqc_pop\src> python .\laqc_pop.py

-----
UNIFEI - Universidade Federal de Itajubá
LaQC - Laboratório de Química Computacional
-----

Análise populacional de configurações

-----

Extrair dados de... C:\Users\Rogério\Documents\dados_debora_configuracoes2
- Extraindo energias do arquivo: DAC_CS2_optfreq_mc20000400.log
- Extraindo energias do arquivo: optfreq_mc20000160.log
- Extraindo energias do arquivo: optfreq_mc20000240.log
- Extraindo energias do arquivo: optfreq_mc20000320.log
- Extraindo energias do arquivo: optfreq_mc20000560.log
- Extraindo energias do arquivo: optfreq_mc20000720.log
- Extraindo energias do arquivo: optfreq_mc20000960.log
- Extraindo energias do arquivo: optfreq_mc20001040.log
- Extraindo energias do arquivo: optfreq_mc20001280.log
- Extraindo energias do arquivo: optfreq_mc20001360.log
- Extraindo energias do arquivo: optfreq_mc20001440.log
- Extraindo energias do arquivo: optfreq_mc20001520.log
- Extraindo energias do arquivo: optfreq_mc20001600.log
- Extraindo energias do arquivo: optfreq_mc20001680.log
- Extraindo energias do arquivo: optfreq_mc20001760.log
- Extraindo energias do arquivo: optfreq_mc20001840.log
- Extraindo energias do arquivo: optfreq_mc20001920.log
- Extraindo energias do arquivo: optfreq_mc20002080.log
- Extraindo energias do arquivo: optfreq_mc20002160.log
- Extraindo energias do arquivo: optfreq_mc20002240.log
- Extraindo energias do arquivo: optfreq_mc20003040.log
Arquivo C:\Users\Rogério\Documents\dados_debora_configuracoes2\dados_energias.dat salvo!
Arquivo C:\Users\Rogério\Documents\dados_debora_configuracoes2\analise_populacao.dat salvo!
    
```

Figura 6 - Execução do laqc_pop.py

Após a execução do script o usuário poderá utilizar o analise_populacao.dat para avaliar a contribuição de cada estrutura, como apresentado na Figura 7:

Arquivo	Sum of electronic and thermal Free Energies	Contribuição
1 optfreq_mc20002080.log	-3263,541248	64,592039
2 optfreq_mc20001680.log	-3263,540665	34,833278
3 optfreq_mc20001600.log	-3263,536788	0,573513
4 optfreq_mc20000720.log	-3263,530788	0,000997
5 optfreq_mc20000480.log	-3263,528669	0,000106
6 optfreq_mc20001440.log	-3263,528255	0,000068
7 DAC_CS2_optfreq_mc20000400.log	-3263,513042	0

Figura 7 - Contribuição das estruturas

Conclusões

Os trabalhos que envolvem a Química Computacional demandam uma grande dedicação à análise de muitos dados, portanto, a geração de aplicações que possam de alguma forma, facilitar o trabalho de pesquisa é de suma importância. Nesse sentido, por mais simples que um processo possa ser, como, por exemplo, a edição de arquivos de input, o desenvolvimento de uma ferramenta que permita otimizar este procedimento permite ao pesquisador, pesquisadora, o direcionamento de seu tempo para outras tarefas. Nesse sentido as ferramentas aqui desenvolvidas tem este propósito. Evidentemente que elas não são um fim, já que melhorias sejam na usabilidade ou a adição de novas funcionalidades devem e podem ser feitas à medida que tais ferramentas são utilizadas por diferentes níveis de usuário.

Agradecimentos

Ao Dr. Asdrubal Lozada pela ajuda à questão matemática envolvida no trabalho. Em especial, agradeço ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), à Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG) e ao Programa Institucional de Bolsas de Iniciação Científica (PIBIC) pela bolsa de estudo que me foi concedida para a realização deste trabalho.

Referências

OLIVEIRA FILHO, A. G. S. d.; BATISTA, A. P. d. L. ESPECTRO ELETRÔNICO DA MOLÉCULA I2: UMA BREVE INTRODUÇÃO À PROGRAMAÇÃO CIENTÍFICA. *Química Nova*, 39, n. 1, p. 118-122, 2016.

VAN ROSSUM, G., & DRAKE, F. L., Ed. *Python 3 Reference Manual*. Scotts Valley, CA: CreateSpace. 2009.

WINSBERG, E. Computer simulations in science. In: ZALTA, E. N. (Ed.). *The Stanford Encyclopedia of Philosophy*. Summer 2018. [S.l.]: Metaphysics Research Lab, Stanford University, 2018.