



Avaliação de esquemas de otimização para o desenvolvimento de método composto em cálculos de entalpias de formação

Pedro Giraldi Faccin

Prof. Dr. Rogério Custodio - Unicamp

Palavras-chave: entalpia de formação, método composto, métodos de otimização

1 Introdução

1.1 Métodos compostos

O cálculo de propriedades eletrônicas, termoquímicas e espectroscópicas de átomos e moléculas constitui um grande desafio dentro da química quântica. O principal objetivo no desenvolvimento de novos métodos é que eles sejam capazes de alcançar valores compatíveis com dados experimentais acurados e, se possível, possuir um custo computacional reduzido. A utilização de cálculos ab initio, teoria do funcional de densidade e métodos compostos, são os meios mais amplamente empregados, já que moléculas das mais variadas dimensões e classes químicas, podem ser estudadas. Nos últimos anos, a elaboração e otimização de tais métodos, possibilitaram o estudo de diversos sistemas que antes eram muito difíceis de serem analisados, como, por exemplo, átomos e moléculas em estados excitados encontrados no espaço e compostos intermediários de reações químicas que possuem tempo de vida muito curto.

Com o crescente aumento no poder computacional, cálculos cada vez mais sofisticados estão sendo desenvolvidos. Contudo, o acesso a supercomputadores não é muito fácil para a maior parte dos pesquisadores e, portanto, surge a necessidade da elaboração de alternativas mais rápidas, mas que garantam confiabilidade nos resultados.

Os chamados métodos compostos representam uma via para se atingir tal necessidade, já que a obtenção desses resultados é compatível com valores obtidos a partir de cálculos de alto nível. Esse método possui uma característica particular, pois os valores calculados são resultados de uma série de combinações de cálculos ab initio de nível inferior, a fim de corrigirem os desvios de um valor de referência.

Os mais bem-sucedidos e amplamente utilizados métodos compostos são os denominados Gaussian-n ou simplesmente Gn desenvolvidos por Pople, Curtiss e col. [1–5]. A teoria Gaussian-3 ou G3 é a terceira na série de métodos Gn [4] e a energia final é definida por:

$$E_{G3} = E_{MP4/6-31G(d)} + \Delta E_{(+)} + \Delta E_{(2df,p)} + \Delta E_{QCI} + \Delta E_{G3\ large} + E_{SO} + E_{ZPE} + E_{HLC} \quad (1)$$

Em que a energia de referência $E_{MP4/6-31G(d)}$ é modificada pelas seguintes correções:

- $\Delta E_{(+)} = E_{MP4/6-31+G(d)} - E_{MP4/6-31G(d)}$, para funções difusas
- $\Delta E_{2df,p} = E_{MP4/6-31G(2df,p)} - E_{MP4/6-31G(d)}$, para funções de polarização em átomos com exceção do hidrogênio e funções-p em hidrogênios
- $\Delta E_{QCI} = E_{QCISD(T)/6-31G(d)} - E_{MP4/6-31G(d)}$ para efeitos de correlação eletrônica
- $\Delta E_{G3\ large} = E_{MP2(full)/G3\ large} - E_{MP2/6-31G(2df,p)} - E_{MP2/6-31+G(d)} + E_{MP2/6-31G(d)}$, para funções de bases extensas

- e) E_{SO} , para correção de spin-órbita extraída de experimentos atômicos ou cálculos teóricos
- f) E_{ZPE} , para correção da energia térmica
- g) E_{HLC} , para levar em conta outros efeitos de correlação de alto nível

O termo de correção HLC, é definido como $E_{HLC} = -An_{\beta} - B(n_{\alpha} - n_{\beta})$ no caso de moléculas e no caso de átomos $E_{HLC} = -Cn_{\beta} - D(n_{\alpha} - n_{\beta})$, em que n_{α} e n_{β} são o número de elétrons de valência com spins alfa e beta, respectivamente, e A, B, C e D são parâmetros otimizados para a produzir o menor erro absoluto médio possível em relação a valores experimentais [6]

Nos métodos Gn chama a atenção que o termo E_{HLC} tem por objetivo corrigir outras deficiências no cálculo dos termos de correlação e função de base. Isto sugere que talvez seja possível substituir o ajuste E_{HLC} por fatores de escalamento realizados diretamente nos termos de correção presentes, por exemplo, na energia G3. Em outras palavras, a expressão para a energia G3 final seria dada por:

$$E_{G3} = E_{MP4/6-31G(d)} + a\Delta E_{(+)} + b\Delta E_{(2df,p)} + c\Delta E_{QCI} + d\Delta E_{G3large} + eE_{SO} + fE_{ZPE} \quad (2)$$

Uma vantagem em utilizar esta formulação é que podemos ajustar os termos de correlação eletrônica e função de base diretamente e constatar através dos valores otimizados de cada parâmetro a importância do efeito de cada termo no cálculo da propriedade de interesse, como a entalpia de formação.

2 Metodologia

2.1 Cálculo de entalpia de formação

O procedimento padrão para o cálculo das entalpias de formação pode ser expresso através de uma sequência de equações bem conhecidas na literatura. O procedimento desenvolvido por Curtiss e col. inicia-se através da energia de atomização:

$$D_0 = \sum xE_{\text{átomo}} - E_{\text{molécula}} - E_{ZPE}$$

Sendo $\sum xE_{\text{átomo}}$, o somatório das energias de cada átomo presente na molécula, e o termo $E_{\text{molécula}}$ a energia da molécula e E_{zpe} a energia de ponto-zero. O termo $E_{\text{átomo}}$ será calculado usando a equação do método G3 padrão, enquanto o termo $E_{\text{molécula}}$ será obtido através da equação definida pelo método G3 modificado.

Com base na energia de atomização, pode-se calcular a entalpia de formação na temperatura de 0 K:

$$\Delta_f H^0(0 \text{ K, molécula}) = \sum x\Delta_f H^0(0 \text{ K, átomos}) - D_0$$

O primeiro termo à direita da igualdade, $\Delta_f H^0(0 \text{ K, molécula})$, simboliza o somatório das entalpias de formação a 0 K dos elementos químicos. Os valores para esses termo são obtidos através de dados experimentais. Finalmente, a entalpia de formação a 298 K é calculada através de correções térmicas:

$$\Delta_f H^0(298 \text{ K, molécula}) = \Delta_f H^0(0 \text{ K, molécula}) + [H^0(298 \text{ K}) - H^0(0 \text{ K})]_{\text{molécula}} - \sum x[H^0(298 \text{ K}) - H^0(0 \text{ K})]_{\text{átomo}}$$

O termo $[H^0(298 \text{ K}) - H^0(0 \text{ K})]_{\text{átomo}}$ é a correção térmica dos elementos, e $[H^0(298 \text{ K}) - H^0(0 \text{ K})]_{\text{molécula}}$ é obtido a partir da energia de ponto-zero e de H_{corr} , denominado correção térmica da entalpia, também fornecido pelos cálculos de frequência:

$$[H^0(298 \text{ K}) - H^0(0 \text{ K})]_{\text{molécula}} = H_{corr} - E_{ZPE}$$

A determinação de H_{corr} envolve contribuições das energias translacionais, vibracionais, rotacionais e eletrônica, além da constante de Boltzmann e da temperatura, sendo expressa como:

$$H_{corr} = E_{eletr} + E_{vib} + E_{rot} + E_{trans} + k_b T$$

2.2 Programação e cálculos quânticos

Para a realização de todos os cálculos e processamento de dados, foi necessário desenvolver uma série de programas e scripts auxiliares. A linguagem de programação escolhida foi a Python, devido a certas características, tais como fácil prototipagem, grande quantidade de tutoriais e documentações, além da existência de diversas funções presentes em bibliotecas e pacotes disponibilizados gratuitamente. Apesar da recorrente indicação que essa linguagem possui uma baixa performance, a quantidade de dados analisada não foi suficientemente alta para que isso fosse um problema, além do uso, sempre que foi possível, de vetorização de contas feitas dentro de laços utilizando funções e vetores da biblioteca *numpy*.

Para o armazenamento de dados experimentais e de resultados, utilizou-se arquivos no formato *.json*, já que esse formato permite uma fácil classificação e categorização e ser bastante simples de fazer a leitura e escrita utilizando as estruturas de dados padrão do *python*.

Todos os códigos foram feitos utilizando um ambiente cujo sistema operacional é *linux*, porém, devido a alta portabilidade da linguagem escolhida, os programas devem funcionar em qualquer outro sistema, desde que as dependências e bibliotecas estejam instaladas e as versões sejam compatíveis.

Para realizar os cálculos de energia baseando-se no método G3 original, utilizou-se o programa *Gaussian 09*. Os arquivos de entrada e saída foram gerenciados através de scripts escritos em *bash*, devido a melhor integração com o console de comandos do sistema. Todos os cálculos foram feitos no cluster Kahuna disponibilizado pelo CEPID CCES Unicamp, utilizando o software licenciado para fins acadêmicos.

3 Resultados e discussão

3.1 Entalpia de formação

Para o cálculo da entalpia de formação, foi necessário criar um protótipo de uma função que calcula essa propriedade para uma molécula qualquer, baseando-se apenas nos parâmetros mostrados na seção anterior. O primeiro modelo desenvolvido foi para calcular o $\Delta_f H^0(298K)$ utilizando a equação do método G3 original (H298) e a equação do método G3 modificado (H298mod). O código abaixo, mostra a função que realiza tal tarefa:

```
1 def enthalpy(molecule):
2
3     conv = 627.5095 #from Hartree to kcal/mol
4
5     H = molecules[molecule]["data"]["G3Enthalpy"]
6     Hmod = a*DEPlus + b*DE2DF + c*EQCISDT + d*EDeltaG3 + e*ETermal + RT
7
8     EX = EXmod = HX = HX298 = 0
9
10    for atom, num in molecules[molecule]["atoms"].items():
11
12        EX += num * atoms[atom]["EX"]
13        EXmod += num * (f*Ef + g*Eg + h*Eh + k*Ek + l*El)
14        HX += num * atoms[atom]["HX"] / conv
15        HX298 += num * atoms[atom]["HX298"] / conv
16
17    H298 = (HX - EX + H - HX298) * conv
18    H298mod = (HX - EXmod + Hmod - HX298) * conv
19
20    return H298, H298mod
```

Como é possível observar nas linhas 5, 6, 11 e 13, há o acesso de valores experimentais necessários para realizar os cálculos da contribuição entálpica dos átomos presentes na molécula. Todos os valores utilizados estão compilados em arquivos de troca de dados e foram retirados das referências presentes na literatura [7]. Os parâmetros de escalamento marcados de *a* até *e* são referentes a moléculas e os marcados de *f* até *l* são dos átomos.

A partir da diferença entre o dado experimental e o valor calculado, foi possível construir o histograma apresentado na figura 1. Esse resultado mostra que para as 148 moléculas do conjunto, 95 tiveram um desvio que foi menor ou igual a 1.0 kcal/mol. As maiores discrepâncias ocorreram nas moléculas orgânicas em que os hidrogênios foram substituídos por algum outro átomo mais eletronegativo, como por exemplo o C_2F_4 e C_2Cl_4 , que apresentaram um desvio de 4.9 kcal/mol e 3.4 kcal/mol respectivamente, ou em algumas moléculas inorgânicas, como o P_2 e N_2O , que apresentaram um desvio de -6.8 kcal/mol e 4.1 kcal/mol.

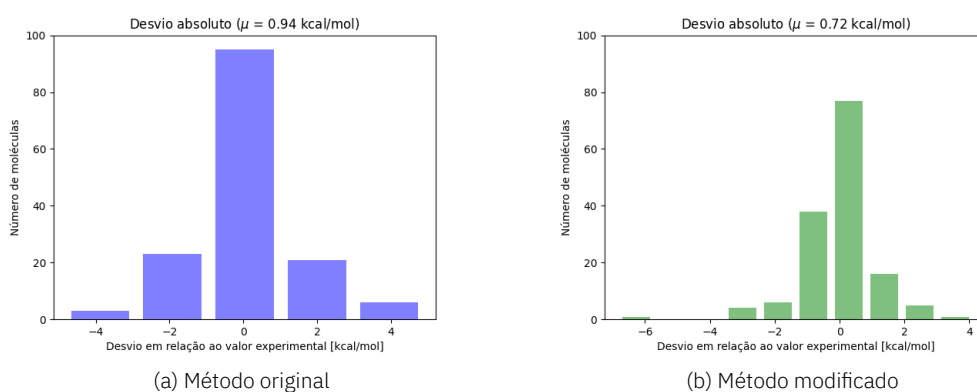


Figura 1: Histograma de frequências em relação ao desvio dos valores experimentais

Apesar de certas moléculas apresentarem uma discrepância elevada em relação aos dados experimentais, o erro absoluto médio ficou em 0.94 kcal/mol no caso do modelo original e 0.72 kcal/mol no modelo modificado.

3.2 Algoritmo de minimização

Para encontrar os valores dos coeficientes de escalamento da Equação 2, aplicou-se o método simplex modificado por Nelder-Mead. Tal função está implementada no pacote *scipy*. Então houve apenas a necessidade de fazer algumas alterações nas funções que calculam a entalpia. Com isso, a principal mudança é que em vez de calcular apenas o valor de $\Delta_f H^0(298\text{ K})$, ela retorna o valor $|\Delta_f H^0(298\text{ K})_{\text{expt.}} - \Delta_f H^0(298\text{ K})_{\text{calc.}}|$, para então ser minimizado de forma a convergir essa diferença a um mínimo.

Os valores obtidos a cada minimização foram escritos em um arquivo de dados, bem como os valores de $\Delta_f H^0(298\text{ K})$ experimentais utilizados. Os trechos de código abaixo representam tais considerações:

```

1 # função para determinar os parâmetros ótimos de cada moléculas
2 initial_param = np.array([1.0, 1.0, 1.0, 1.0, 1.0])
3 optimal_param = minimize(mod_enthalpy, initial_param, method = "Nelder-Mead")
4 # arquivo contendo os parâmetros ótimos para cada molécula
5 "SiF4": {
6     "parameters": {
7         "a": 1.0125258481056447,
8         "b": 1.0088369597141122,
9         "c": 1.0000089408431174,
10        "d": 1.0034558904386466,
11        "e": 1.0107403437183333
12    }
13 }

```

Como é possível observar, é necessário inicializar um vetor contendo valores iniciais para cada parâmetro. Embora não haja uma forma exata de saber quais valores utilizar, aparenta ser uma escolha razoável que eles sejam iguais a 1, já que isso implicaria uma contribuição igual de cada termo. De fato, ao obter os valores minimizados, nota-se que são todos muito próximos de um. Deve-se frisar também, que ao converter os valores em unidades atômicas de energia e entalpia para kcal/mol, mesmo que ínfimas, essas diferenças exercem uma grande influência nos resultados finais.

4 Conclusão

O método G3 apresenta uma forma bastante robusta no quesito de previsão de valores de propriedades termoquímicas de uma vasta gama de moléculas. Apesar da sua ampla difusão, há diversas questões por trás dos motivos de sua eficácia, como utilização do termo HLC, que é obtido de forma a minimizar o erro absoluto de uma série de propriedades.

Baseando-se nos termos principais de correção de energia e postos fatores de escalamento para ponderar cada um deles, notou-se que é possível obter resultados bastante satisfatórios quanto ao cálculo da entalpia de formação. O modelo original possui um erro médio absoluto de 0.94 kcal/mol, enquanto o modelo proposto apresentou uma redução do erro médio absoluto para 0.72 kcal/mol.

As principais vantagens dessa mudança foram em relação ao fato de que ao substituir um termo sem significado por fatores de escalamento, é possível saber quais termos são mais preponderantes na correção do valor de energia de referência, já que ao ter em mãos os valores dos parâmetros otimizados para cada classe de moléculas, basta selecionar em qual grupo uma nova molécula pertence e, muito rapidamente, calcular o valor da entalpia de formação da mesma. Ao saber quais termos exercem maior influência, é possível utilizar tais informações para o desenvolvimento de novos métodos, onde o uso de funções de bases que corrijam melhor tais desvios diminuam ainda mais o erro associado a tais divergências. De modo geral, os fatores de escalamento relacionados com a correção para funções difusas, $\Delta E_{(+)}$, foram os que tiveram uma maior influência, principalmente em moléculas contendo átomos do terceiro período.

5 Referências bibliográficas

1. Pople, J. A., Head-Gordon, M., Fox, D. J., Raghavachari, K. & Curtiss, L.A. Gaussian-1 theory: A general procedure for prediction of molecular energies. *J. Chem. Phys.* 90, 5622–5629 (1989).
2. Shao, Y. et al. Advances in methods and algorithms in a modern quantum chemistry program package. *Phys. Chem. Chem. Phys.* 8, 3172–3191 (2006).
3. Blaudeau, J.-P., McGrath, M. P., Curtiss, L. A. & Radom, L. Extension of Gaussian-2 (G2) theory to molecules containing third-row atoms K and Ca. *J. Chem. Phys.* 107, 5016–5021 (1997).
4. Curtiss, L. A., Raghavachari, K., Redfern, P. C., Rassolov, V. & Pople, J.A. Gaussian-3 (G3) theory for molecules containing first and second-row atoms. *J. Chem. Phys.* 109, 7764–7776 (1998).
5. Curtiss, L. A., Redfern, P. C. & Raghavachari, K. Gn theory. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* 1, 810–825 (2011).
6. Curtiss, L. A., Raghavachari, K. & Pople, J. A. Gaussian-2 theory using reduced Moller-Plesset orders. *J. Chem. Phys.* 98, 1293–1298 (1992).
7. Narayanan, B., Redfern, P. C., Assary, R. S. & Curtiss, L. A. Accurate quantum chemical energies for 133 000 organic molecules. *Chem. Sci.* 10, 7449–7455 (2019).