



SOLVATE

MANUAL DO USUÁRIO: GUIA RÁPIDO

SOLVATE versão 2024.02

Manual do Usuário: Guia Rápido

VISÃO GERAL

Este é um manual do *Solvate Suite*. Os módulos do programa principal fazem interface com diversos programas amplamente utilizados na comunidade científica (como *ORCA*, *xTB*, *GROMACS*, *PackMol* e *Q-Force*), possibilitando o tratamento sequencial de todas as etapas da modelagem molecular envolvendo interação explícita das espécies de interesse com qualquer solvente utilizando procedimentos de simulação molecular e cálculos de estrutura eletrônica em uma abordagem híbrida (solvente explícito/implícito).

(1) O primeiro problema que o programa pretende resolver é a criação da caixa de simulação, considerando vários critérios especificados pelo usuário (como o número de moléculas de solvente, forma e densidade da caixa de simulação, concentração de soluto/solvente, ou proporção entre misturas de solventes), bem como a construção da topologia do sistema e dos parâmetros do campo de forças (no caso de métodos de simulação baseados em dinâmica clássica).

(2) O segundo problema é submeter a simulação com um conjunto mínimo de parâmetros ajustados (alguns dos quais podem ser alterados pelo usuário durante a execução; alternativamente, o usuário pode realizar esta etapa com seus próprios parâmetros de simulação).

(3) A terceira é a análise das propriedades simuladas (pressão, temperatura, densidade, interação de curto alcance, energia potencial total e função de distribuição radial), incluindo análise estatística (como valores médios estocásticos, erro padrão e desvio padrão), análise gráfica e extração de propriedades por diferentes critérios (como análise de blocos e filtro de pressão).

(4) O quarto envolve a seleção de um número determinado pelo usuário de estruturas ligadas a hidrogênio ou clusters de microsolvatação e moléculas de solvente por cluster identificadas a partir do arquivo de trajetória de simulação com base no critério de conectividade máxima (ligações de hidrogênio) ou no critério de energia mínima (ligações de hidrogênio e microsolvatação). aglomerados).

(5) A quinta envolve o tratamento (por métodos semi-empíricos de Química Quântica) de clusters de microsolvatação para o cálculo da energia livre, considerando três critérios: na geometria fixa extraída da simulação, em uma estrutura de cluster completamente otimizada, ou em uma cavidade de solvente relaxada e geometria de soluto completamente otimizada. De qualquer forma, o procedimento remove todas as frequências imaginárias (que têm peso significativo na composição da energia livre), e utiliza a correção quase-harmônica (para lidar com os modos intermoleculares de baixa frequência), seguida do uso de métodos de extrapolação para a energia livre. energia para o nível teórico especificado pelo usuário (o que também permite a extrapolação da restrição do conjunto de bases nos métodos de solvente implícitos ao considerar clusters de microsolvatação com pelo menos a primeira camada de solvatação).

(6) A sexta e última etapa envolve o gerenciamento do número significativo de arquivos gerados ao longo do processo de modelagem.

Etapa 1: Empacotamento

➤ `solvate <soluto>.ext <solvente>.ext [-nm I -it -bk] [I=inteiro]`

➔ **OBJETIVO:** Geração da caixa de empacotamento para a simulação.

- ✓ **Arquivos:** Os arquivos `<soluto>.com` e `<solvente>.com` contêm as geometrias do soluto e do solvente, respectivamente. O programa *Solvate* identifica as geometrias com base na ordem em que os arquivos são passados em linha de comando, sendo o arquivo com a estrutura do soluto obrigatoriamente o primeiro. É possível especificar mais de um solvente, informando-se ao programa o número ou a proporção de cada espécie no empacotamento. Por padrão, o programa gera uma caixa de simulação com uma única molécula de soluto no seu centro.¹⁻²
- ✓ **Nmol [-nm]:** Especifica o número de moléculas de solvente a serem incluídas no processo de empacotamento. Se este parâmetro for especificado o programa determinará automaticamente as dimensões da caixa de simulação que satisfaçam o critério de mínima distância entre as moléculas na caixa de simulação.
- ✓ **Npck [-np]:** Especifica o número de “camadas de empacotamento” (as quais são definidas de modo que haja ao menos *I* moléculas de solvente em torno do soluto em qualquer direção). Caso este parâmetro seja especificado o programa determinará automaticamente o número de moléculas de solvente (parâmetro *nmol*) e as dimensões da caixa de simulação que satisfaçam simultaneamente os critérios do número de camadas de empacotamento e de mínima distância entre as moléculas na caixa de simulação (o padrão atual do *Solvate* é de 2,0 Å, não sendo necessário informá-lo na linha de comando; este parâmetro pode ser ajustado com a opção de linha de comando *tolera*).³
- ✓ **Shape [-sp]:** Especifica formato da caixa de simulação. Os formatos disponíveis são *sphere* (esférico, padrão para o *ORCA*), *box* (caixa ortorrômbica) e *cube* (cúbico, padrão para o *GROMACS*).
- ✓ **Density [-ds]:** Especifica a densidade da caixa de simulação. A densidade solicitada é atingida a partir do ajuste automático do critério de mínima distância (parâmetro *tolera*), mantendo-se os demais critérios (como número de camadas de empacotamento ou número de moléculas de solvente) inalterados.

1 Caso as estruturas iniciais não tenham sido previamente minimizadas, utilize o módulo *runCREST* do suite de aplicativos para realizar a análise conformacional das estruturas de entrada para o empacotamento.

2 Utilize preferencialmente entradas no formato *.com* (*Gaussian*) ou *.inp* (*ORCA*) devido à informação da carga e da multiplicidade do sistema. Os demais formatos aceitos são *.xyz*, *.pdb*, *.log* (*Gaussian*) e *.out* (*ORCA*).

3 Os números reais devem ser passados ao programa com um ponto (não uma vírgula) como separador decimal. Neste manual optamos por utilizar a vírgula como separador devido ao idioma em que foi escrita.

Etapa 2: Simulação

➤ `solvate <solut>.com <solvente>.com -nmol=I -mdrun [-resub][-tsteps=N -nsteps=I]`

➔ **OBJETIVO:** Realização da simulação a partir da caixa gerada na etapa anterior.

- ✓ **MDSim:** Solicita a realização de uma simulação de *dinâmica molecular de Born-Openheimer* (BOMD) com o programa *Orca* no nível semi-empírico GFN2-xTB. Caso nenhum parâmetro adicional seja fornecido o programa realizará uma primeira submissão em um *ensemble* NVT na temperatura de 298 K, com 1000 passos de 1,0 fs.⁴⁻⁵
- ✓ **ensNPT:** Define o *ensemble* como NPT (caixa de simulação de tamanho variável).⁶
- ✓ **Resub:** Ressubmete a simulação, continuando do ponto onde a dinâmica anterior terminou.
- ✓ **Reini:** Ressubmete a simulação do início. Útil em casos de erro nos parâmetros iniciais.
- ✓ **mdPres:** Define a pressão a ser atingida na simulação.⁷
- ✓ **mdTemp:** Define a temperatura a ser atingida na simulação.⁸
- ✓ **tSteps:** Tamanho do passo da simulação (em fs; padrões: *Solvate* 1,0 fs; *Orca* de 0.5 fs).
- ✓ **nSteps:** Número de passos da simulação (padrões: *Solvate* 1000, *Orca* não existe).
- ✓ **Optim:** Realiza a otimização da caixa de simulação inicial.⁹

Notas: Em uma mesma pasta com múltiplos arquivos é possível realizar diferentes empacotamentos com número variável de moléculas de solvente, de modo que nesta etapa é necessário informar o número de moléculas de solvente com o parâmetro *nmol*, o que possibilita o programa identificar corretamente os arquivos gerados. Embora seja possível realizar a simulação diretamente com o *Orca*, sem a interface com o *Solvate*, a submissão pelo programa garante que os arquivos terão o formato adequado a ser lido nas demais etapas. Com a conclusão da simulação obtém-se os arquivos de trajetória (.trj.xyz, útil para visualizar o que está ocorrendo com a simulação; recomenda-se a utilização do programa *Avogadro*) e de dados (.dat, para visualizar a convergência dos parâmetros da simulação – tempo vs pressão, temperatura e energia potencial; recomenda-se a utilização do *QtiPlot*).

-
- 4 No *ensemble* NVT o volume é mantido impondo-se à caixa de simulação uma força definida por uma constante de mola que, por padrão, é de $10,0 \text{ kJ} \cdot \text{mol}^{-1} \text{Å}^{-2}$. Este critério pode ser ajustado no arquivo de entrada modificando-se o parâmetro *cell/spring*. Para mais informações e opções de ajuste consulte o manual do *Orca*.
 - 5 Os testes realizados até aqui sugerem que os critérios iniciais adotados são suficientes para se atingir a equilíbrio da temperatura e da energia. Trabalho ainda precisa ser feito para verificar o que ocorre no *ensemble* NPT.
 - 6 Os testes com micro-solvatação tem levado a destruição da caixa ao longo da simulação no *ensemble* NPT. Na dinâmica *ab initio* com o *Orca* não existe a opção de condições periódicas de contorno, o que possivelmente evitaria este problema. O ponto importante é que este fato ainda precisa ser estudado pois o ideal seria realizar a dinâmica neste *ensemble*. No entanto, como no momento a simulação tem sido destruída no final para se calcular a energia livre de *clusters* soluto-solvente este fato não deve levar a maiores implicações.
 - 7 Em um *ensemble* NVT este parâmetro não impacta a simulação. No entanto, a sua definição no arquivo de entrada possibilita que o *Orca* registre a pressão ao longo da simulação. No *ensemble* NPT o valor padrão (tanto do *Solvate* quanto do *Orca*) é de 1,0 bar.
 - 8 A temperatura é mantida na simulação por um *termostato*, que opera reescalando as velocidades ao longo da simulação. Atualmente o *Orca* só possui o modelo de termostato simples de Berendsen.
 - 9 Esta opção não tem sido testada no momento devido ao seu custo computacional. Possivelmente este recurso seja importante no *ensemble* NPT.

Etapa 3: Refinamento

➤ `solvate <solute>.com <solvente>.com -nmol=I -hbond[=N] [-nsvt=I -nsol=I] [-opt/conf]`

➔ OBJETIVO: Identificação das interações mais relevantes.

- ✓ **hBond:** Nesta etapa o *Solvate* analisa o arquivo de trajetória da simulação, selecionando a configuração de maior número de “conexões” soluto-solvente (este critério é baseado unicamente na distância entre sítios de recepção/doação de prótons no soluto e sítios de doação/recepção de prótons no solvente; caso sejam localizadas estruturas com igual número de conexões, considera-se a estrutura de menor energia; esta é a opção padrão deste módulo do programa) ou a estrutura de menor energia global da última trajetória (esta seleção é conduzida com a opção *minimal*). As ligações de hidrogênio são caracterizadas por dois critérios: (i) o primeiro é a distância soluto-solvente (apenas entre sítios de doação/recepção de prótons; este critério pode ser ajustado informando-se o valor da distância de corte com o parâmetro *N*; o padrão do *Solvate* é de 2,0 Å, não sendo necessário informá-lo na linha de comando; o limite superior é de 3,0 Å), e (ii) o segundo é a energia de estabilização (o programa caracteriza cada interação soluto-solvente individualmente; interações repulsivas possibilitam descartar moléculas de solvente; interações atrativas com energias de estabilização maiores que 1,2 kcal/mol – calculadas a nível GFN2-xTB – são preservadas).¹⁰ A análise de ligações de hidrogênio é conduzida com as geometrias relaxadas (determinadas a nível GFN2-xTB), de modo que o ângulo da ligação de hidrogênio não é levado em consideração.
- ✓ **Nsvt:** Especifica o número de moléculas de solvente nas vizinhanças do soluto a serem selecionadas (até o limite superior definido por *nmol*; o programa determina as *I* moléculas de solvente mais próximas do soluto a partir da menor distância determinada entre cada átomo do soluto a cada átomo de cada uma das *nmol* moléculas de solvente, e não em relação ao centro geométrico de cada molécula). A opção *nsol* especifica o número de “camadas de solvatação” (o número de moléculas de solvente é determinado de modo a haver ao menos *nsol* moléculas de solvente em cada direção envolvendo o soluto).
- ✓ **Opt:** Especifica a otimização da estrutura gerada após a seleção das moléculas de solvente em torno do soluto (realizada a nível GFN2-xTB).
- ✓ **Conf:** Especifica a realização da análise conformacional (por *simulated annealing*) da estrutura gerada após a seleção das moléculas de solvente em torno do soluto (realizada a nível GFN2-xTB).

Notas: Ao final do processo o *Solvate* seleciona as moléculas de solvente (formando ligações de hidrogênio com a opção *hbond*, ou certo número de moléculas de solvente mais próximas do soluto com a opção *nsvt*, ou ainda certo número de camadas de solvatação com a opção *nsol*) a partir a estrutura selecionada (de máximas conexões ou de menor energia ao longo da última trajetória gerada na etapa de simulação). Em todo caso, o programa gerará os *inputs* para a otimização de geometria e cálculo de estrutura eletrônica e de frequências vibracionais para o *Gaussian* (arquivo .com) e *Orca* (arquivo .inp), que devem ser ajustados com as opções de cálculo desejadas antes de proceder a próxima etapa. Além disso, o *Solvate* criará arquivos auxiliares com o mesmo nome do arquivo contendo o soluto, porém com as extensões .gjf (*Gaussian*) e .ojf (*Orca*), evitando-se a sobrescrição dos arquivos originais, de modo que os demais arquivos intermediários podem ser compactados, limpando a pasta de trabalho. Este procedimento de limpeza e organização de arquivos pode ser realizado na 4ª e última etapa de utilização do *Solvate*.

10 O critério de corte energético utilizado na caracterização das ligações de hidrogênio tem por base o artigo de revisão “*Experimental Binding Energies in Supramolecular Complexes*”, *Chem. Rev.*, 116(9) 5216 (2016).

Etapa 4: Gerenciamento de Arquivos

➤ `solvate <solute>.com <solvente>.com [-nmol=I [-nsvt=I]] -compact/expand`

➔ **OBJETIVO: Preservação e gerenciamento de arquivos intermediários.**

- ✓ Após a realização das etapas anteriores o número de arquivos gerados cresce bastante e rapidamente. Este módulo tem por finalidade facilitar o gerenciamento dos arquivos, compactando-os ou expandindo-os de acordo com o padrão de nomes de arquivos adotados pelo *Solvate*.
- ✓ Por exemplo, para ficar com um único arquivo de várias simulações de uma estrutura de referência, execute:

`solvate soluto.com solvente.com -compact`

→ **Resultado: compactação no arquivo soluto+solvente.tgz**

Para arquivar apenas as simulações com *nmol* moléculas de solvente, em uma pasta que contém diversas simulações com caixas de diferentes tamanhos:

`solvate soluto.com solvente.com -nmol=I -compact`

→ **Resultado: compactação no arquivo soluto+solvente_molI.tgz**

Para arquivar apenas as simulações com *nmol* moléculas de solvente e *nsvt* moléculas vizinhas:

`solvate soluto.com solvente.com -nmol=I -nsvt=J -compact`

→ **Resultado: compactação no arquivo soluto+solvente_molI_svtJ.tgz**

Para recuperar os arquivos o procedimento é o mesmo, porém substituindo o parâmetro *compact* pelo parâmetro *expand*.

Obs.: Execute o programa *Solvate* sem qualquer parâmetro para ver a lista completa das opções de execução.

Etapa 5: Submissão

➔ **OBJETIVO: Realização da otimização de geometria e cálculo da estrutura eletrônica.**

- ✓ Com a conclusão das etapas anteriores, tem-se um conjunto de arquivos de entrada gerados automaticamente pelo programa para a realização de cálculos de otimização de geometria, cálculo de estrutura e eletrônica e propriedades termodinâmicas. Esta etapa deve ser realizada em separado do programa *Solvate*, de modo que outras análises e ajustes sejam realizados de maneira independente. Por exemplo, pode-se utilizar o módulo *Conform*, que acompanha a suite de aplicativos do *Solvate*, para realizar a análise conformacional das estruturas geradas em diferentes estados de oxidação, o que é útil, por exemplo, no estudo de reações de oxirredução, no qual se tem o interesse no cálculo do potencial de redução.

Obs.: Em todo caso, os arquivos de entrada gerados automaticamente devem ser ajustados com as opções desejadas antes de se proceder a etapa de cálculo com o programa de interesse.