
Calculating the effect of a point mutation with free energy perturbation

A tutorial making use of VMD and NAMD

Jeffrey Comer

Department of Anatomy and Physiology
Kansas State University

November 7, 2023

1 Introducción a perturbación de energía libre

Energía libre es un concepto importante en la termodinámica, la fisicoquímica y el diseño de los fármacos. Transforma la segunda ley de termodinámica, que dice que la entropía del universo (incluyendo el sistema y su entorno) es maximizado en equilibrio, en una minimización de las propiedades del sistema solamente, su energía y su entropía: $F_{\text{sys}} = E_{\text{sys}} - TS_{\text{sys}}$, donde F_{sys} es la energía libre de Helmholtz, E_{sys} es la energía interna del sistema, S_{sys} es la entropía del sistema y T es la temperatura del sistema y su entorno. Un aspecto importante de la energía libre es que es una función de estado que significa que el cambio de energía libre de un estado a otro no depende en el camino. El método de perturbación de energía libre (FEP), particularmente perturbación de energía libre alquímica, aprovecha de este hecho, usando caminos no físicos sino convenientes y eficientes.

FEP es uno de los métodos más antiguos para calcular la energía libre de simulaciones moleculares y uno de los más usados en la academia y la industria farmacéutica [1]. FEP es basado en la ecuación de Zwanzig [2]. Para un sistema atómico, definimos dos Hamiltonianos ($H_{\mathbf{A}}$ y $H_{\mathbf{B}}$) que dan la energía del sistema desde las coordenadas y los momentos. Típicamente, no nos interesan los momentos y consideramos solo las energías potenciales, $U_{\mathbf{A}}(\mathbf{X})$ y $U_{\mathbf{B}}(\mathbf{X})$, que son funciones de las coordenadas de todos los átomos, $\mathbf{X} \equiv x_1 \dots x_{3N}$. En este contexto, la ecuación de Zwanzig rinde la diferencia de la energía libre de Helmholtz desde un promedio exponencial sobre la diferencia de las energías $U_{\mathbf{A}}$ y $U_{\mathbf{B}}$:

$$\Delta F_{\mathbf{A} \rightarrow \mathbf{B}} = -k_{\text{B}}T \ln \left\langle \exp \left(\frac{U_{\mathbf{B}}(\mathbf{X}) - U_{\mathbf{A}}(\mathbf{X})}{k_{\text{B}}T} \right) \right\rangle_{\mathbf{A}}. \quad (1)$$

Un aspecto importante y poderoso del método es que este promedio es hecho sobre un ensamble de coordenadas generado con la función de energía $U_{\mathbf{A}}$ solamente. Entonces, de una simulación con la energía potencial $U_{\mathbf{A}}$, determinamos la energía libre con una otra potencial $U_{\mathbf{B}}$. Un uso importante de este método es el cálculo eficiente de la energía libre de variaciones en la estructura química de un fármaco o una proteína con respecto a una estructura de referencia. Este método se llama “alchemical free energy perturbation calculations” porque se transforma un grupo químico al otro en una manera que se parece a los objetivos de la alquimia.

Normalmente, no es posible aplicar ecuación 1 directamente porque $U_{\mathbf{B}}$ y $U_{\mathbf{A}}$ para los estados de interés difiere por mucho más de $k_{\text{B}}T$, dejándonos con mala estadística (solo pocos juegos de coordenadas contribuyen al promedio). Entonces, lo más común es hacer el cálculo en muchas etapas (llamadas “windows”):

$$\Delta F_{\mathbf{A} \rightarrow \mathbf{B}} = -k_{\text{B}}T \sum_{i=1}^n \ln \left\langle \exp \left(\frac{U_{i+1}(\mathbf{X}) - U_i(\mathbf{X})}{k_{\text{B}}T} \right) \right\rangle_i. \quad (2)$$

Entre etapas consecutivas, asumimos que la energía no cambia demasiado ($|\Delta U| \not\gg k_{\text{B}}T$). El cambio es representado por el variable λ que va desde el valor 0 (en estado \mathbf{A}) al valor 1 (en estado \mathbf{B}). Para una mutación, por ejemplo de una fenilalanina a un triptófano, $\lambda = 0$ representa un estado en que la fenilalanina es completamente “encendido” y el triptófano es completamente apagado. Con $\lambda = 0.5$, la fenilalanina y el triptófano están medio encendidos (sus interacciones con las otras moléculas son reducidas). Finalmente, a $\lambda = 1$, la fenilalanina ha desaparecido completamente (no interactúa con su entorno) y el triptófano es completamente presente. Es por eso que decimos que el camino de la transición es no físico porque hay dos cadenas laterales al mismo tiempo y algunos átomos parcialmente existentes.

Porque los grupos químicos típicamente tienen diferentes números de grados de libertad y el campo de fuerza no representa los enlaces covalentes en una manera verdadera, el ΔF solo para

una mutación normalmente no es interpretable. Mejor, hacemos la misma mutación para el ligando en su sitio de unión y para el ligando libre. En esta manera, calculamos la energía libre de unión relativa:

$$\Delta\Delta F_{\mathbf{A}\rightarrow\mathbf{B}}^{\text{unión}} = \Delta F_{\mathbf{A}\rightarrow\mathbf{B}}^{\text{complejo}} - \Delta F_{\mathbf{A}\rightarrow\mathbf{B}}^{\text{libre}} \quad (3)$$

Por ejemplo, si diseñamos un péptido que forma una unión con un receptor inmunológico, podemos calcular con FEP alquímica si una mutación de residuo 3 desde fenilalanina a triptófano hace más o menos fuerte la interacción con el receptor.

2 Construcción del sistema

Para permitir un cálculo adentro de el horario de este taller, vamos a usar un sistema muy pequeño. Queremos imitar es un cálculo de cambios de energía libre para mutaciones de aminoácidos de un péptido en una unión con una proteína. Nuestro péptido consiste de solo tres aminoácidos. En lugar de una proteína, usaremos una capa de grafeno que permitirá que el cálculo se converja más rápido (¿por qué?). Nuestros sistemas no incluirán agua para que las simulaciones terminen rápidamente.

1. **Crea un péptido con la secuencia PPF (prolina-prolina-fenilalanina) con VMD.** Elige Extensions → Modeling → Molefacture. En la ventana Molefacture, elige Build → Protein Builder. Después de entrar la secuencia, pulsa el botón Build. Guarda el resultado por File → Write psf and pdb files. Usa el nombre mf_PPF.
2. **Construye un archivo de topología (psf) para este péptido.** Coloca el código abajo en un archivo script que se llama “build.tcl”. Abre el VMD Tk Console (Extensions → Tk Console) Ejecuta el script con: `source build.tcl`. (Si usas Windows y un editor como NotePad, el nombre completo puede ser “build.tcl.txt”, aunque Windows esconde el “.txt”.)

```
set pdb mf_PPF.pdb
set outName pep_PPF

package require psfgen
# La topología CHARMM para las proteínas
topology top_all36_prot.rtf

segment A {
    # Proline tiene un "patch" especial en el terminal N
    first PROP
    last CTER
    pdb $pdb
    auto angles dihedrals
}
coordpdb $pdb
guesscoord

writepsf $outName.psf
writepdb $outName.pdb
```

3. **Combina el péptido y el grafeno.** La estructura de grafeno está incluida con los archivos de este tutorial (graphTiny.psf y graphTiny.pdb).

```

set psf0 pep_PPF.psf
set pdb0 pep_PPF.pdb
set psf1 graphTiny.psf
set pdb1 graphTiny.pdb
set outName graph_PPF

```

```

package require psfgen
resetpsf
readpsf $psf0
coordpdb $pdb0
readpsf $psf1
coordpdb $pdb1

```

```

writepsf $outName.psf
writepdb $outName.pdb

```

4. Mueve el péptido para que no choque con el grafeno.

```

mol new graph_PPF.psf
mol addfile graph_PPF.pdb

set s [atomselect top "segname A"]
$s moveby {0 0 3}

set all [atomselect top all]
$all writepdb graph_PPF_pos.pdb

```

References

1. Schindler CE, Baumann H, Blum A, Böse D, Buchstaller HP, Burgdorf L, Cappel D, Chekler E, Czodrowski P, Dorsch D, Eguida MKI, Follows B, Fuchß T, Grädler U, Gunera J, Johnson T, Lebrun CJ, Karra S, Klein M, Knehans T, Koetzner L, Krier M, Leiendecker M, Leuthner B, Li L, Mochalkin I, Musil D, Neagu C, Rippmann F, Schiemann K, Schulz R, Steinbrecher T, Tanzer EM, Lopez AU, Follis AV, Wegener A, , Kuhn D (2020) Large-scale assessment of binding free energy calculations in active drug discovery projects. *J Chem Inf Model* 60(11):5457–5474
2. Zwanzig RW (1954) High-temperature equation of state by a perturbation method. I. Nonpolar gases. *J Chem Phys* 22(8):1420–1426