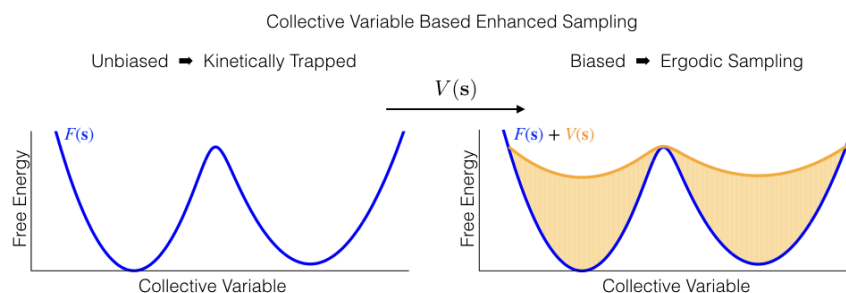


Modelowanie zachowania się złożonych układów dynamicznych w długiej skali czasowej jest podstawowym zadaniem w naukach chemicznych. Symulacje dynamiki molekularnej (MD) pozwalają na badanie szczegółów czasoprzestrzennych procesów molekularnych. Jednakże, tzw. problem próbkowania Boltzmanniana poważnie ogranicza ich użyteczność w praktyce. Problem próbkowania Boltzmanniana wynika z faktu, że typowy krajobraz energii swobodnej składa się z wielu stanów metastabilnych oddzielonych barierami energii swobodnej znacznie przewyższającymi energię termiczną. W skali czasu, którą można symulować, przekroczenie bariery jest rzadkim zdarzeniem, a układ pozostaje kinetycznie uwięziony w stanie metastabilnym. Chociaż powolna ścieżka reakcji może scharakteryzować rzadkie przejście pomiędzy dwoma stanami metastabilnymi, wyprowadzenie układu z równowagi wzdłuż tej ścieżki jest niewykonalne przy użyciu standardowych metod MD.

Jednym ze sposobów na częściowe rozwiązanie problemu próbkowania Boltzmanniana jest zastosowanie symulacji z próbkowaniem wzmocnionym, w których fluktuacje kilku krytycznych stopni swobody, zwanych zmiennymi kolektywnymi (CV), są wzmocnione przez zewnętrzny potencjał. Wydajność takich metod zależy od jakości CV. Efektywne CV powinny rozróżniać odpowiednie stany metastabilne i zawierać istotne stopnie swobody. Zazwyczaj CV są wybierane na podstawie intuicji. Jednak znalezienie CV, które kwantyfikują kluczowe cechy rzadkiego zdarzenia, może nie być trywialne. Metody próbujące skonstruować CV bezpośrednio z danych MD są ograniczone do uczenia się tylko ze standardowych symulacji MD. Zatem, jeśli dane MD nie są w stanie opisać zdarzeń o długiej skali czasowej, CV będą pozbawione ważnych informacji. Takie metody są testowane głównie na istniejących wcześniej zestawach danych lub prostych modelach, dla których znamy już wolne CV.



Obecnie żadna metoda nie jest w stanie uczyć się z danych wygenerowanych przez wzmocnione MD, ograniczając zastosowanie istniejących metod tylko do krótkich skal czasowych. Dlatego, dla bardziej złożonych układów, w których występują długie skale czasowe, musimy generować dane w oparciu o ulepszone techniki próbkowania. W tym projekcie rozważamy kluczowe problemy związane z metodologią estymacji CV: Jak konstruować CV bez uciekania się do wiedzy eksperckiej specyficznej dla danego systemu? Czy możliwe jest skonstruowanie powolnych CV bezpośrednio na podstawie danych z symulacji próbkowania rozszerzonego?

Dzięki udoskonaleniu naszej ostatniej metody, *Multiscale Reweighted Stochastic Embedding* (MRSE) [J. Phys. Chem. A **2021**, 125], będziemy w stanie dostarczyć statystyczne oszacowania powolnych CV wywnioskowanych z danych z symulacji wzmocnionego próbkowania. Będziemy w stanie użyć naszej metody uczenia iteracyjnie w różnych czasach symulacji, aby oszacować powolne CV, a następnie wzmocnić próbkowanie. Ogólnie rzecz biorąc, planujemy opracować metodę, która będzie w stanie nauczyć się powolnych CV z danych pochodzących z dynamiki wzmocnionej w sposób niemalże ślepy, czyniąc ją dostępną dla wielu użytkowników bez szczegółowej wiedzy na temat teorii próbkowania wzmocnionego. Spodziewamy się, że metoda ta będzie miała znaczący wpływ na obecny stan środowiska MD i będzie mogła być zastosowana do procesów długoskalowych w chemii, fizyce i biologii i biologii.