

# Molekularna dinamika RNA molekula

Kemijski seminar I

Stipe Mustać

# 1 Uvod

RNA molekule smatraju se jednim od najvažnijih i najraznovrsnijih kemijskih vrsta u molekularnoj biologiji [1]. Prema centralnoj dogmi molekularne biologije, mRNA služi kao posrednik pri prijenosu genetske informacije, dok tRNA služi kao molekularni adapter za prenošenje te informacije u proteinsku sekvencu [2, 3]. Uz to, RNA ima i brojne druge uloge ključne za održavanje, regulaciju i procesiranje genske informacije [1, 4]. Prema tome RNA može biti klasificirana u protein-kodirajuću i nekodirajuću RNA (ncRNA). Velika većina RNA transkriptata, njih barem 80 000, pripadaju nekodirajućoj RNA i imaju fundamentalnu ulogu u velikom broju regulatornih procesa [1]. Kao posljedica toga, istraživanja RNA molekula izašla su iz okvira proučavanja njihove fiziološke funkcije vezane uz translaciju gena. Primjerice, mutacije u ncRNA povezane su s rakom i nekim drugim bolestima, a riboprekidači (eng. *riboswitch*) i ostale ncRNA postaju mete djelovanja novih antibiotika [1, 5–8].

Kako se širi znanje o RNA molekulama i otkrivanju novi oblici i funkcije tih molekula, ograničenja eksperimentalnih alata postala su limitirajući čimbenik pri proučavanju ovisnosti funkcije RNA molekula o njihovoj strukturi i dinamici. U svladavanju ovih ograničenja pomažu nam brojni računalni alati i to na način da nam olakšavaju interpretaciju postojećih eksperimentalnih podataka i proučavanje povezanosti između strukture, dinamike i funkcije RNA molekula, kao i postavljanje daljnjih eksperimentalnih hipoteza [1].

## 2 Principi strukture i dinamike RNA

RNA je linearni polimer sastavljen od četiriju nukleotidnih monomera: A, C, G i U. Svaki se nukleotid sastoji od planarnih aromatskih baza vezanih za ribozu, pentozni monosaharid oblika furanoznog prstena sa 5'-fosfatnom skupinom. RNA lanci građeni su formiranjem fosfodieterskih veza između uzastopnih šećernih ostataka, pri čemu fosfatne skupine vežu 3'-ugljik svakog šećernog ostatka s 5'-ugljikom onog idućeg. Pri tome ostaju slobodni 5'-ugljik na jednom kraju lanca (5'-kraj) i 3'-ugljik na drugom kraju (3'-kraj). Ovo čini RNA lance asimetričnima. 5'-kraj se smatra početkom lanca jer je to mjesto gdje sinteza lanca započinje u živim organizmima [1].

## 2.1 Primarne, sekundarne i tercijarne RNA strukture

RNA molekule mogu se opisati analiziranjem njihove primarne, sekundarne (2D) i tercijarne (3D) strukture. Primarna struktura definira RNA sekvencu u smislu nukleotida od kojih je RNA sastavljena, koji naknadno prolaze post-transkripcijsku kemijsku modifikaciju. 2D struktura definirana je obrascem Watson-Crick kanonskih parova baza (A=U i G=C). Kanonski parovi baza su posebno stabilni te stabiliziraju duplekse RNA za  $1-3 \text{ kcal mol}^{-1}$  po baznom paru [9]. Iako su RNA molekule jednolančane, većina RNA molekula formira niz antiparalelnih dvostrukih heliksa na način da se presavijaju i sami sa sobom stvaraju nizove Watson-Crick komplementarnih baza. Za razliku od DNA, gdje prevladava B-heliks, RNA molekule češće formiraju A-heliks. Uz A=U i G=C parove baza, dvostruki heliksi A-oblika često sadrže i značajan udio G/U "kolebljivih" parova baza. Ostali kolebljivi parovi baza uključuju A=C<sup>+</sup>, U=U i G=A parove, ali i brojne druge kombinacije [10].

Kanonski dvostruki heliksi izmjenjuju se s regijama nukleotida koji ne formiraju kanonske parove baza te su stoga nominalno nesporeni. 2D strukture su prema tome prikaz susjednih kanonskih parova baza uz preostale nukleotide prikazane kao "petlje" u tim 2D grafovima. Međutim, mnogi od tih nesporenih nukleotida formiraju nekanonske bazne parove. U većini RNA molekula, kanonski heliksi teže tomu da budu što kraći te općenito ne sadrže preko 12 uzastopnih kanonskih baznih parova [1]. Dugi nizovi kanonskih baznih parova smatraju se premonotonima i pretabilnima da bi RNA bila dovoljno konformacijski dinamična što joj omogućuje izvršavanje njezinih bioloških uloga. Također, na taj se način razlikuju i vlastite od patogenih RNA u ljudskom organizmu. [1, 11].

Bazirajući se na 2D strukturi, 3D struktura RNA određena je brojnim kanonskim i nekanonskim interakcijama baznih parova koji često uključuju stvaranje vodikovih veza posredstvom 2'-OH skupine. Za razliku od DNA, RNA molekule mogu formirati puno više tipova parova baza uz već poznate Watson-Crick bazne parove. Razlog je tomu to što RNA molekule imaju tri različita brida dostupna za stvaranje vodikovih veza. Uz spomenuti Watson-Crick brid, preko kojega se prikazuju interakcije vodikovih veza u većini klasičnih prikaza interakcija dušičnih baza (Watson-Crick sparivanje), kod RNA molekula uobičajeno je i vezanje preko Hoogsteenova brida (Hoogsteenovo sparivanje), vezanje preko 2'-OH skupine na ribozu, kao i sve moguće kombinacije veza između navedenih područja (tj. rubova) koji sudjeluju u stvaranju vodikovih veza. Tako se prema

načinju vezanja, RNA bazni parovi klasificirani su u 12 geometrijskih različitih tipova parova baza [1, 12–14]. Ova je klasifikacija kasnije proširena na interakcije baza-fosfat specifične za pojedine sekvence, što omogućuje da se sparivanja RNA baza objasni kao kao mreža internukleotidnih interakcija specifičnih za određenu sekvencu [1, 15]. Upravo je ova klasifikacija baznih parova u smislu njihove izosteričnosti ključna za razumijevanje razvoja evolucije RNA molekule. Princip izosteričnosti nalaže da je očuvanje točne geometrije interakcije baznih parova ključni uvjet koji upravlja evolucijom i kovarijacijskim obrascima RNA sekvenci. To znači i da sekundarna i tercijarna struktura određuju primarnu, a ne obrnuto [1, 10, 12, 16]. Zamjena baze, ili baznog para drugačije geometrije promijenila bi geometriju čitave RNA molekule, što bi izravno utjecalo i na njenu biološku funkciju. To se događa jer promjena geometrije baznog para utječe i na cijelu strukturnu osnovicu, čime bi se konformacijski poremećaj prenio duž čitave RNA molekule. Iz tog su razloga i slučajevi u kojima dolazi do prekršaja principa izosteričnosti rijetki. Analiza izosteričnosti je u biti jedan od tipičnih primjena MD simulacija, budući da tehnika prilično pouzdano opisuje oblike nukleobaza i njihove reakcije, što omogućava proučavanja supstitucije baza.

Unutar zamotane RNA molekule, svaka prisutna vodikova veza ima dakle važnu ulogu u očuvanju strukture te su RNA strukture puno strože određene u odnosu na tipične proteinske strukture. Specifično strukturirani RNA blokovi, nazvani još i motivima, opisani su karakterističnim vodikovim vezama i interakcijama slaganja koje su ključne za njihovu formaciju i diktiraju nukleotidne nizove koje ih grade. Stroga određenost RNA struktura mora biti uzeta u obzir i pri izvršenju molekularnodinamskih istraživanja budući da u slučaju da se princip izosteričnosti ne uzme u obzir, može doći do pogrešne interpretacije rezultata simulacije. Gubitak interakcija određenih baznih parova u trodimenzionalnoj strukturi tijekom MD simulacije može primjerice biti posljedica strukturne nestabilnosti koja nastaje zbog uvedenih aproksimacija kao posljedica primjene polja sila.

## 2.2 Dinamika RNA molekula

Kao što je i ranije spomenuto, 3D struktura RNA molekula uvelike je određena sekvencom nukleotida koja gradi tu određenu RNA molekulu. Međutim, važno je svojstvo RNA molekula da njihove strukture nisu statične. Jedna molekula RNA sa točno definiranom sekvencom često ima više dostupnih 2D i 3D struktura koje leže u uskom rasponu slo-

bodne energije strukturiranja, a zbog termalne fluktuacije i interakcije sa okolinom takve su strukture i uobičajene. Vremenski raspon u kojem te alternativne strukture nastaju varira između mikrosekunda za jednostavne promjene parova baza do sekunda (ili više) za kompleksna ponovna sklapanja 2D struktura [17, 18].

Usprkos nevjerojatnoj važnosti opisivanja strukture RNA, eksperimentalna je karakterizacija dinamike puno zahtjevnija od istraživanja njene statičke strukture. Iz toga se razloga u analizi dinamike molekula RNA koriste napredne molekularnodinamske simulacije koje služe upotpunjavanju dostupnih eksperimentalnih tehnika. To se postiže na način da simulacijama eksplicitno i u realnom vremenu simuliramo strukturnu dinamiku RNA molekula što nam omogućuje bolju interpretaciju i lakše planiranje eksperimentalnih mjerenja. Nadalje, računalnim simulacijama dolazimo do pouzdanih pretpostavki vezanih uz strukturalnu dinamiku koje zatim eksperimentalno provjeravamo. Konačno, simulacije nam daju uvide u određene dinamičke aspekte koje ne možemo dobiti koristeći trenutne eksperimentalne metode, a pomažu nam objasniti određene funkcionalne uloge RNA molekula. U nekim slučajevima aproksimacije koje moramo koristiti u MD simulacijama ne reproduciraju eksperimentalno dobivene podatke, no čak i tada dobiveni rezultati simulacija, u kombinaciji sa kemijskom i fizikalnom intuicijom, mogu biti iznimno korisni u dizajniranju novih pokusa [1]. U sklopu ovog rada objasnit ću metode korištenja nekih od tih simulacija, točnije one koje omogućuju praćenje strukturne dinamike RNA kroz nešto dulje vremenske intervale.

### 3 Molekularna dinamika

Metode molekularne dinamike služe rješavanju jednadžbi gibanja kojima opisujemo sustav te nam omogućuju praćenje vremenske evolucije sustava u realnom vremenu. Vremenski rasponi u kojima se događaju konformacijske tranzicije RNA molekula mogu se uvelike razlikovati. Interna dinamika malih "ukosnica" RNA može trajati u rasponima milisekunda do mikrosekunda, dok sporiji procesi kao što su ligandima inducirana sklapanja riboprekidača imaju trajanje reda veličine sekundi, minuta, ili čak i dulje [19, 20]. Uz to, budući da se simulacije provode na jedinstvenim kopijama molekula, potrebno je primijetiti više individualnih rijetkih događaja kako bismo dobili statistički dokazana svojstva, a time i realistična termodinamička i kinetička svojstva procesa kojeg istražujemo. Takve su opsežne simulacije često izvan dosega današnjih računala i zbog toga se u svrhu simulacije većih i kompliciranijih sustava često koriste računalni sustavi koji imaju veliku računalnu snagu (superračunala ili računalni klasteri) u kombinaciji s određenim tehnikama simulacije.

#### 3.1 Načela molekularne dinamike

Molekularnodinamske simulacije su računalne simulacijske metode kojima možemo analizirati fizičke pomake atoma i molekula u nekom sustavu. Atomi i molekule u molekularnodinamskoj simulaciji stupaju u interakciju tijekom nekog vremena čime dobivamo prikaz dinamičke evolucije sustava. Budući da se tim metodama mogu simulirati prilično veliki sustavi (stotine tisuća ili milijuni molekula) te da su reakcije u biološkim sustavima često spore (simulacije traju i po nekoliko stotina nanosekundi ili čak i nekoliko mikrosekundi), simulacije molekularne dinamike uključuju milijune računa u svrhu dobivanja kvalitetno simulirane putanje molekule.

U molekularnoj dinamici, uzastopne se konfiguracije sustava dobivaju integriranjem Newtonovih jednadžbi gibanja. Putanja čestice dobiva se rješavanjem diferencijalne jednadžbe:

$$\frac{d^2 r_i}{dt^2} = \frac{F_{x_i}}{m_i} \quad (1)$$

Ova jednadžba prikazuje gibanje čestice mase  $m_i$  duž jedne koordinate ( $x_i$ ), pri čemu je  $F_{x_i}$  sila na česticu koja se giba.

U ovom je slučaju bitno razlikovati tri tipa problema na koje se Newtonovi zakoni

gibanja mogu primijeniti. U najjednostavnijem slučaju, prije sudara dviju čestica na njih ne djeluje niti jedna sila. Nakon jednog sudara i prije drugoga, pozicija čestice promijeni se za iznos  $v_i \delta t$ , pri čemu je  $v_i$  konstantna brzina, a  $\delta t$  vrijeme između sudara. U drugom slučaju, na čestice tijekom gibanja može djelovati konstantna sila (npr. gibanje nabijenih čestica u homogenom električnom polju). Treći je slučaj u kojemu sila na čestice ovisi o njenoj relativnoj poziciji u odnosu na druge čestice. Budući da su gibanja čestica spregnuta, u posljednjem je slučaju vrlo često nemoguće analitički opisati sustav.

Prema tome, u realnim modelima dolazi do promjene sile koja djeluje na svaku česticu kad god čestica promijeni poziciju ili kad bilo koja od čestica s kojom je ta čestica u interakciji promijeni poziciju. Gibanja svih čestica su spregnuta i dolazi do problema više tijela, za koji ne postoji analitičko rješenje. Iz tog razloga za integraciju jednadžbi gibanja koriste se *metode konačnih diferencija*.

Tehnike konačnih diferencija koriste se za generiranje putanja molekularne dinamike molekula sa kontinuiranom funkcijom potencijala. Ideja tih tehnika je da se integracija jednadžbi gibanja razdjeli na malene segmente, razmaknute za fiksirano vrijeme  $\delta t$ . Ukupna sila na svaku česticu u vremenu  $t$  je izračunana kao suma vektora interakcije te čestice sa drugim česticama. Iz sile možemo nadalje odrediti akceleraciju čestice, koje onda kombiniramo sa pozicijom i brzinama u vremenu  $t$  kako bi odredili poziciju i brzine u vremenu  $t + \delta t$ . Uzima se aproksimacija da je sila tijekom tog vremenskog koraka konstantna. Sile na česticu zatim se određuju za novi dobiveni položaj čestice, iz čega se onda ponovno određuju položaj i brzina za  $t + 2\delta t$ , itd.

Postoji velik broj algoritama za integriranje jednadžbi gibanja korištenjem metoda konačnih diferencija. Svi oni pretpostavljaju da svojstva sustava koje određujemo jednadžbama gibanja, položaj, brzinu, akceleraciju, itd. možemo aproksimirati ekspanzijom Taylorova niza:

$$r(t + \delta t) = r(t) + \delta t v(t) + \frac{1}{2} \delta t^2 a(t) + \frac{1}{6} \delta t^3 b(t) + \frac{1}{24} \delta t^4 c(t) + \dots \quad (2)$$

$$v(t + \delta t) = v(t) + \delta t a(t) + \frac{1}{2} \delta t^2 b(t) + \frac{1}{6} \delta t^3 c(t) + \dots \quad (3)$$

$$a(t + \delta t) = a(t) + \delta t b(t) + \frac{1}{2} \delta t^2 c(t) + \dots \quad (4)$$

$$b(t + \delta t) = b(t) + \delta t c(t) + \dots, \quad (5)$$

pri čemu su  $r(t)$  položaj čestice u vremenu  $t$ ,  $v(t)$  brzina,  $a(t)$  akceleracija, a jedan je

od najčešćih algoritama korišten za integriranje jednadžbi gibanja *Verlet algoritam*. Za izračunavanje položaja u vremenu  $t + \delta t$ ,  $r(t + \delta t)$ , Verlet algoritam koristi položaj i akceleraciju u vremenu  $t$  te položaj u prethodnom koraku,  $r(t - \delta t)$ .

$$r(t + \delta t) = r(t) + \delta t v(t) + \frac{1}{2} \delta t^2 a(t) + \dots \quad (6)$$

$$r(t - \delta t) = r(t) - \delta t v(t) + \frac{1}{2} \delta t^2 a(t) - \dots \quad (7)$$

Zbrajanjem ove dvije jednadžbe dolazimo do

$$r(t + \delta t) = 2r(t) - r(t - \delta t) + \delta t^2 a(t) \quad (8)$$

Dalje se iz tog algoritma može izračunati i brzina u vremenu  $t$ ,  $v(t)$ .

$$v(t) = [r(t + \delta t) - r(t - \delta t)] / 2\delta t \quad (9)$$

Osim Verlet algoritma, za integraciju jednadžbi gibanja često se koriste i *"leap frog"* algoritam, *Beemanov algoritam*, *Verlet po brzinama* (eng. *Velocity Verlet*) i drugi.

Ukoliko je sila koja djeluje na atome u sustavu konzervativna (ovisi samo o položaju atoma), kao što je slučaj u sustavima molekularne dinamike, sila i potencijalna energija povezane su relacijom:

$$F_{x_i} = - \frac{dV}{dr_i} \quad (10)$$

Što znači da jednadžbu (8) možemo zapisati u obliku:

$$r(t + \delta t) = 2r(t) - r(t - \delta t) - \frac{\delta t^2}{m} \frac{dV(t)}{dr} \quad (11)$$

Prema tome, kako bismo iz integrirali jednadžbu gibanja koristeći Verlet algoritam te došli do položaja u idućoj točki,  $r(t + \delta t)$ , potrebne su informacije o položaju čestice u vremenu  $t$ , položaju u vremenu  $t - \delta t$  te gradijentu potencijalne energije  $-dV(t)/dr$  u vremenu  $t$ . Potrebno je i zadati inicijalne uvjete. Inicijalne položaje atoma uzimamo iz eksperimentalno dobivenih struktura ili teoretskih modela, dok brzine atoma dobivamo pomoću Maxwell-Boltzmannove distribucije na nekoj temperaturi.

Navedeni gradijent potencijalne energije moguće je izračunati na više načina:

- korištenjem polja sila, pri čemu se onda radi o klasičnoj molekularnoj dinamici



- rješavanjem Schrödingerove jednačbe  $\hat{H}\psi = E\psi$
- kombinacijom oba navedena načina, QM/MM (eng. *Quantum Mechanics/Molecular Mechanics*) molekularna dinamika

Osim odabira adekvatnog načina računanja gradijenta potencijalne energije, za provođenje kvalitetne simulacije važan je i adekvatan odabir trajanja vremenskog koraka,  $\delta t$ . Ukoliko je  $\delta t$  prekratko, putanja simulacije pokrit će premali komad faznog prostora molekule, a ukoliko je preveliko, javit će se pogreške u integracijskim algoritmima zbog velikih energetskih preklapanja između atoma.

U idealnim slučajevima, našom bi simulacijom mogli posjetiti sve točke faznog prostora i particijska bi se funkcija:

$$Q_{NVT} = \frac{1}{N!} 1h^{3N} \iint dp^N dr^N \exp\left[-\frac{E(p^N, r^N)}{k_B T}\right] \quad (12)$$

mogla dobiti sumiranjem funkcije  $\exp(-E/k_B T)$ . Tada bi se putanja faznog prostora nazivala *ergodičnom* te bi rezultati simulacije bili neovisni o inicijalnoj konfiguraciji. Zbog veličine faznog prostora većih sistema, ergodičnu putanju u pravilu nije moguće postići.

Termodinamička svojstva mogu ovisiti o derivaciji particijske funkcije ili o samoj particijskoj funkciji. Ona koje ovise o derivaciji particijske funkcije nazivaju se mehanička svojstva i mogu se rutinski dobiti Monte Carlo ili molekularnodinamskom simulacijom. To su primjerice unutrašnja energija, toplinski kapacitet ili tlak. Ona koja ovise o izravno o samoj particijskoj funkciji nazivamo entropijskim ili termalnim svojstvima. To su npr. slobodna energija, kemijski potencijal, ili sama entropija.

Entropijska se svojstva zbog njihove relacije s particijskom funkcijom mogu dobiti puno teže u odnosu na mehanička svojstva i kako bismo ih adekvatno procijenili potrebno je koristiti specijalne tehnike uzorkovanja. Razlog je tomu što spomenute veličine jako ovise o konfiguracijama visoke energije, a spomenute metode, Monte Carlo i molekularna dinamika, naginju uzorkovanju područja niže energije faznog prostora. U simulacijama se zbog toga područja više energije slabije uzorkuju, što dovodi do netočnosti i nepreciznosti u izračunima slobodne energije i ostalih entropijskih svojstava.

Potrebno je napomenuti i da se ovi problemi ne javljaju u slučaju malih, izoliranih molekula koje imaju mali broj dobro okarakteriziranih energetskih minimuma. U takvim slučajevima može se dobiti particijska funkcija korištenjem standardnih metoda statističke

termodinamike koje uključuju sumiranje preko stanja minimalne energije, uzimajući u obzir pri tome i doprinose unutarnjih vibracijskih gibanja [21].

### 3.2 Izračun slobodne energije

Slobodna energija jedna je od najvažnijih termodinamičkih veličina kojima opisujemo promatrani sustav. Ta veličina određuje spontanost određenog procesa vezanog za taj sustav, kao i vjerojatnost da će sustav zauzeti određeno stanje. Slobodnu energiju izražavamo kao Helmholtzovu slobodnu energiju,  $A$ , ili Gibbsovu slobodnu energiju,  $G$ . Helmholtzova slobodna energija prikladna je za sustave u kojima imamo konstantan broj čestica, temperaturu i volumen (NVT ili kanonski ansambl), dok je Gibbsova slobodna energija prikladna za sustave sa konstantnim brojem čestica, temperaturom i tlakom. Većina eksperimenata provodi se u uvjetima konstantnih temperature i tlaka, za koje je Gibbsova slobodna energija prikladna veličina.

Kao što je u prethodno objašnjeno, slobodnu energiju vrlo teško možemo dobiti za sustave koji imaju veliki broj energetskih minimuma. Prema jednadžbi za Helmholtzovu slobodnu energiju [21]:

$$A = k_B T \ln \left( \iint dp^N dr^N \exp \left( \frac{E(p^N, r^N)}{k_B T} \right) \rho(p^N, r^N) \right) \quad (13)$$

Važna značajka ove jednadžbe je to da eksponencijalni član,  $\exp(E(p^N, r^N)/k_B T)$ , značajno doprinosi vrijednosti integrala. To znači i da konfiguracije sustava s visokom energijom isto tako značajno doprinose vrijednostima integrala. Budući da često korištene metode, Monte Carlo i molekularna dinamika, imaju tendenciju uzorkovanja dijelova faznog prostora niže energije, posljedica neadekvatnog uzorkovanja dijelova prostora više energije bit će loša konvergencija i netočnost dobivenih vrijednosti slobodne energije (i ostalih entropijskih svojstava).

### 3.3 Ploha slobodne energije

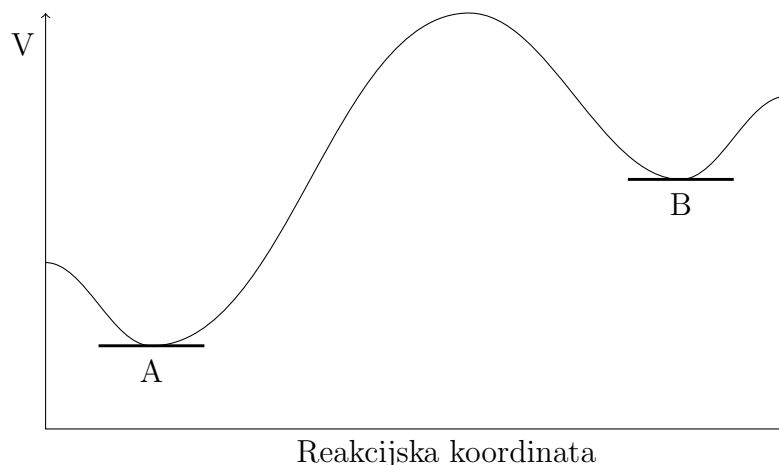
Često je slučaj da tijekom nekog procesa u sustavu možemo pratiti tako da promatramo promjenu samo jednog ili manjeg niza parametara koji su relevantni za taj proces (razmaci između određenih atoma, duljine određenih veza, torzijski kut neke veze u molekuli, itd.). Taj se parametar ili skupina parametara tada naziva reakcijska koordinata. U računalnoj

kemiji ti se parametri često nazivaju i kolektivnim varijablama (CV, eng. *collective variable*). Promatramo li kako se mijenja slobodna energija sustava u ovisnosti o promjeni reakcijske koordinate, možemo konstruirati plohu slobodne energije. U literaturi je vrlo čest i engleski naziv *potential of mean force* (PMF). Kada simuliramo sustav koji je u otapalu, plohom slobodne energije obuhvaćeni su i učinci otapala, kao i intrinzične interakcije tog sustava koji promatramo. Kretanje plohom slobodne energije odgovara istovremeno i fizički realnim promjenama. Prema tome, točka najviše energije na plohi slobodne energije odgovara prijelaznom stanju, iz kojeg je moguće izvesti kinetičke veličine kao što je primjerice konstanta brzine.

Zbog spomenutih nedostataka pri uzorkovanju dijela faznog prostora više energije, u cilju zaobilazanja tih poteškoća razvijena je tehnika "*umbrella sampling*", US.

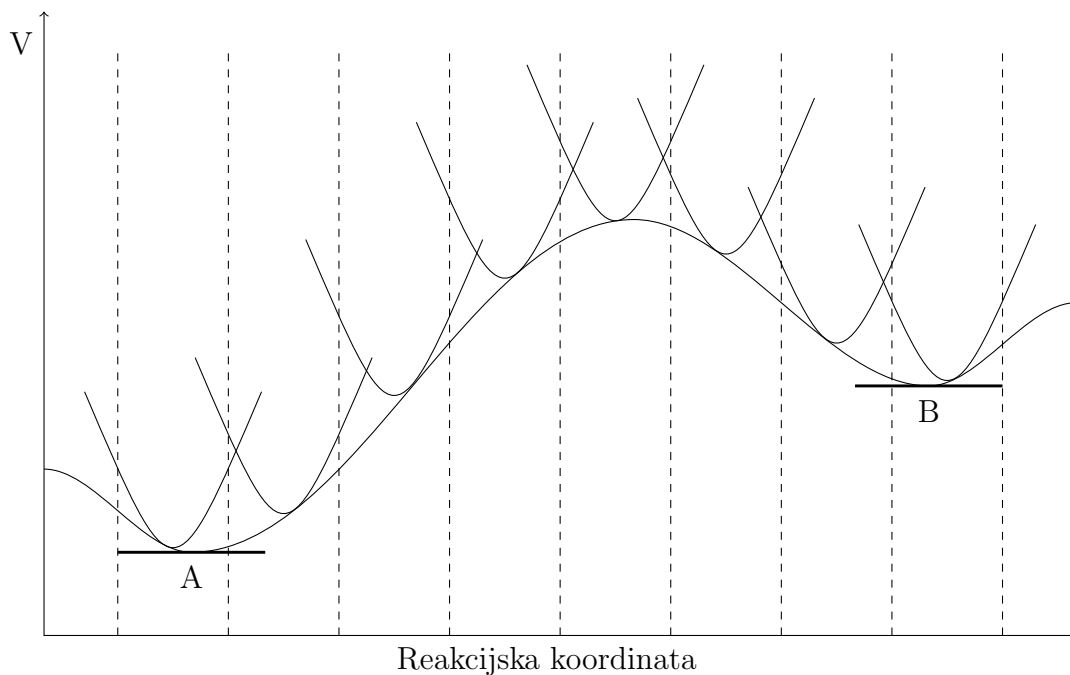
### 3.4 Umbrella sampling

Cilj umbrella sampling metode je prema tome savladati probleme uzorkovanja modificiranjem funkcije potencijala tako da se i energetske nepovoljne stanja dostatno uzorkuju. Zamislamo li da simuliramo prijelaz  $A \rightarrow B$  s velikom reakcijskom barijerom kako je prikazano na slici ispod.



Simulacija ovog procesa konvencionalnom molekularnom dinamikom trajala bi izrazito dugo jer je vjerojatnost prijelaza preko ovoliko visoke barijere vrlo mala. Umbrella sampling je metoda kojom možemo ubrzati dinamiku molekule i popuniti populaciju B, kao i istražiti konformacijski prostor između molekula A i B i na taj način dobiti podatke o mehanizmu kojim se reakcija odvija. Iako to nije nužan uvjet da bismo započeli

simulaciju, pri primjeni Umbrella sampling tehnike pretpostavka je da nam je poznata reakcijska koordinata (ili kolektivne varijable) kojom se reakcija odvija. Kolektivnim varijablama možemo opisati i razlikovati različita metastabilna stanja u našem sustavu. Prije provođenje same simulacije, potrebno je primijeniti kratku simulaciju povlačenja duž reakcijske koordinate. Na taj način odredimo niz konfiguracija koje koristimo kao početne konfiguracije za simulaciju u svakom "prozoru". Unutar svakog prozora zatim primijenimo dodatan potencijal, prikazan parabolama na slici ispod, povrh stvarnih potencijala.



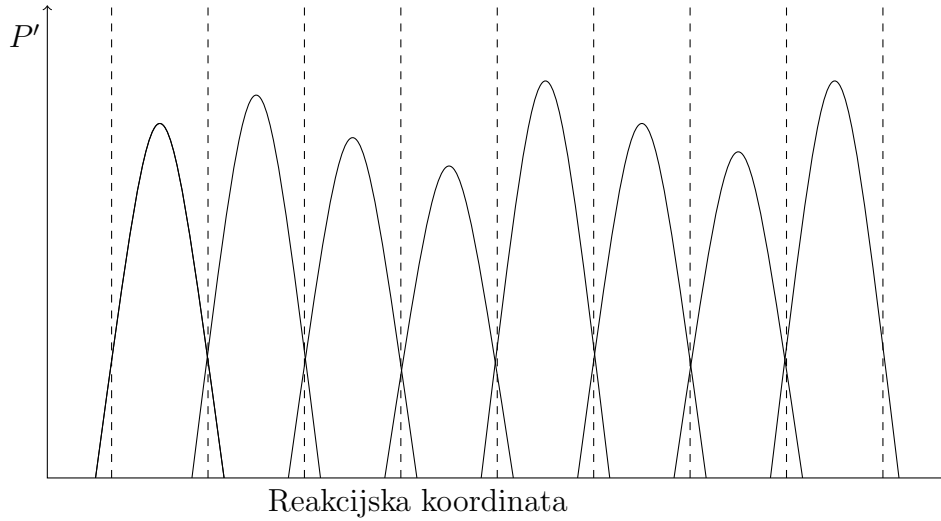
Ukupni potencijal tada iznosi:

$$V'(r^N) = V(r^N) + W(r^N), \quad (14)$$

pri čemu je  $W(r^N)$  oblika:

$$W(r^N) = k_W(r^N - r_0^N)^2, \quad (15)$$

a  $r_0^N$  je ravnotežno stanje. Ovaj se prilagođeni potencijal (eng *bias potential*)  $V'$  koristi pri integriranju jednadžbi gibanja Verlet algoritmom. Nadalje, iz simulacija dobijemo za svaki prozor gustoću vjerojatnosti  $P'$ . Ta je gustoća vjerojatnosti  $P'$  prilagođena gustoća vjerojatnosti i nije jednaka pravoj gustoći vjerojatnosti.



Iz te je prilagođene gustoće vjerojatnosti moguće primjenom matematičkih tehnika doći do prave gustoće vjerojatnosti.

$$G = f(P') \quad (16)$$

Bitno je pri tome da iznos konstante sile  $k_W$  iz jednadžbe (15) nije niti premalen, što bi dovelo do tog da ne dođe do prijelaza barijere, niti prevelik, što bi dovelo do uzorkovanja preuskih dijelova reakcijske koordinate te posljedično i do preuskih distribucija vjerojatnosti koje se ne bi presijecale. To bi dalje onemogućilo precizno određivanje plohe slobodne energije.

Metoda koja se najčešće koristi za dobivanje plohe slobodne energije iz prilagođene gustoće vjerojatnosti je metoda ponderiranih histograma (eng. *weighted histogram method*, WHAM) [22].

Gustoća vjerojatnosti duž neke reakcijske koordinate dobije se eliminacijom svih drugih stupnjeva slobode, osim  $\xi$ ,

$$P(\xi) = \frac{\int \delta[\xi(r) - \xi] \exp(-V(r^N)/k_B T) dr^N}{\int \exp(-V(r^N)/k_B T) dr^N}, \quad (17)$$

gdje  $\xi$  predstavlja reakcijsku koordinatu, a  $\delta$  je Diracova delta funkcija. Iz toga se može izračunati slobodna energija za reakcijsku koordinatu:

$$G(\xi) = -\frac{1}{\beta} \ln P(\xi), \quad (18)$$

pri čemu je  $\beta = 1/k_B T$ . Budući da nam je iz umbrella samplinga poznata  $P'(\xi)$ , potrebno

je iz  $P'(\xi)$  dobiti  $P(\xi)$  pa iz toga slobodnu energiju. Analogno onome kako je prikazano u jednadžbi (17), moguće je dobiti i prilagođenu gustoću vjerojatnosti  $P'$ . Odnos između prilagođene i stvarne gustoće vjerojatnosti prikazan je jednadžbom:

$$\frac{P(\xi)}{P'(\xi)} = \exp(W(\xi(r^N))/k_B T) \frac{\int \exp\left(-\frac{V(r^N) + W(\xi(r^N))}{k_B T}\right) dr^N}{\int \exp\left(-\frac{V(r^N)}{k_B T}\right) dr^N} \quad (19)$$

Iz ovog se izraza može dobiti očekivana vrijednost  $\exp\left(-\frac{W(\xi(r^N))}{k_B T}\right)$ :

$$\left\langle \exp\left(-\frac{W(\xi(r^N))}{k_B T}\right) \right\rangle = \frac{\int \exp\left(-\frac{W(\xi(r^N))}{k_B T}\right) \exp\left(-\frac{V(r^N)}{k_B T}\right)}{\int \exp\left(-\frac{V(r^N)}{k_B T}\right)} \quad (20)$$

Dalje iz toga vrijedi:

$$P(\xi) = P'(\xi) \exp\left(\frac{W(\xi(r^N))}{k_B T}\right) \left\langle \exp\left(-\frac{W(\xi(r^N))}{k_B T}\right) \right\rangle \quad (21)$$

Kada uvrstimo jedn. (21) u jedn. (18), dobijemo:

$$G(\xi) = -\frac{1}{\beta} \ln P(\xi) = -\frac{1}{\beta} \ln \left[ P'(\xi) e^{\beta W(\xi(r^N))} \left\langle e^{-\beta W(\xi(r^N))} \right\rangle \right] \quad (22)$$

Uzmemo li da je  $-\frac{1}{\beta} \ln \left\langle e^{-\beta W(\xi(r^N))} \right\rangle = F$ , dobijemo:

$$G(\xi) = \frac{1}{\beta} \ln P'(\xi) - W(\xi(r^N)) + F \quad (23)$$

Primijenimo li jedn. (23) za svaki prozor na koji smo podijelili reakcijsku koordinatu, dobije se:

$$G_i(\xi) = \frac{1}{\beta} \ln P'_i(\xi) - W_i(\xi(r^N)) + F_i \quad (24)$$

Da bi dobili svaki pojedini  $G_i(\xi)$  i  $G(\xi)$  potrebno je dobiti sve pojedine  $F_i$ . Spomenuta WHAM najkorištenija je metoda kojom se dolazi do  $F_i$  i  $G(\xi)$ .

Prema toj metodi, ako je  $F_i = -\frac{1}{\beta} \ln \left\langle e^{-\beta W(\xi(r^N))} \right\rangle$ , tada prema definiciji očekivane

vrijednosti ansambla ponovno vrijedi:

$$e^{-\beta F_i} = \left\langle e^{\beta W(\xi(r^N))} \right\rangle = \int P(\xi) e^{-\beta W(\xi(r^N))} d\xi \quad (25)$$

Gustoća vjerojatnosti  $P(\xi)$  izračuna se iz ponderirane srednje vrijednosti gustoća vjerojatnosti u individualnim prozorima simulacija:

$$P(\xi) = \sum_i^N p_i(\xi) P_i(\xi) \quad (26)$$

pri čemu je  $N$  broj prozora, a svaki  $P_i(\xi)$  pomnožen sa svojim ponderom  $p_i(\xi)$ . Ponder gustoće vjerojatnosti za svaki pojedini prozor dobije se minimaliziranjem statističke pogreške  $P_i(\xi)$ :

$$\frac{\partial \sigma^2(P(\xi))}{\partial p_i(\xi)} = 0 \quad (27)$$

Suma svih pondera mora biti jednaka 1:

$$\sum_i^N p_i(\xi) = 1 \quad (28)$$

Uvjeti (27) i (28) dovode nas do sljedećih jednadžbi [23]:

$$p_i(\xi) = \frac{a_i(\xi)}{\sum_j^N a_j} \quad (29)$$

Prema jednadžbi (29),  $p_i(\xi)$  je omjer parametra  $a_i(\xi)$  koji je karakterističan za svaki prozor, podijeljenog sumom svih parametara. Parametri  $a_i(\xi)$  dobiju se jednadžbom:

$$a_i(\xi) = N_i e^{-\beta W_i(\xi(r^N)) + \beta F_i} \quad (30)$$

gdje je  $N_i$  broj uzorkovanih koraka u prozoru  $i$ . Ukratko, WHAM algoritam funkcionira na način da se prvo izračuna

$$e^{-\beta F_i} = \int P(\xi) e^{-\beta W(\xi(r^N))} d\xi \quad (25)$$

$P(\xi)$  potreban za tu jednadžbu dobije se iz:

$$P(\xi) = \sum_i^N p_i(\xi) P_i(\xi), \quad (26)$$

a  $p_i(\xi)$  iz jednadžbi:

$$p_i(\xi) = \frac{a_i(\xi)}{\sum_N^j a_j} \quad (29)$$

$$i \ a_i(\xi) = N_i e^{-\beta W_i(\xi(r^N)) + \beta F_i} \quad (30)$$

Iz ovog niza vidimo da ukoliko želimo izračunati konstantu  $F_i$ , potrebne su nam gustoće vjerojatnosti iz jedn. (26). Za rješavanje te jednadžbe potrebni su nam opet ponderi iz jedn. (29), koji ovise o parametrima  $a_i(\xi)$ . Ti parametri  $a_i(\xi)$  prema jedn. (30) ponovno ovise o  $F_i$ . To nam govori da ovaj niz jednadžbi mora biti iterativno riješen dok rezultati ne konvergiraju.

Ukratko, umbrella sampling simulaciju provodimo da podijelimo reakcijsku koordinatu u određeni broj prozora i pokrenemo u svakom prozoru simulaciju molekularne dinamike uz primjenu dodatnog potencijala kako sustav ne bi izlazio iz prozora u kojem ga želimo zadržati. Iz toga dobijemo prilagođenu gustoću vjerojatnosti, iz koje matematičkim tehnikama (najčešće WHAM) možemo konstruirati plohu slobodne energije duž naše reakcijske koordinate, što je i cilj provođenja metode umbrella samplinga.



## Literatura

- [1] Jiří Šponer i dr. „RNA Structural Dynamics As Captured by Molecular Simulations: A Comprehensive Overview”. *Chemical Reviews* 118.8 (travanj 2018.), str. 4177–4338.
- [2] F. Crick. „Central Dogma of Molecular Biology”. *Nature* 227.5258 (kolovoz 1970.), str. 561–563.
- [3] S. J. Sharp i dr. „Structure and Transcription of Eukaryotic tRNA Genes”. *CRC critical reviews in biochemistry* 19.2 (1985.), str. 107–144.
- [4] Ross C. Wilson i Jennifer A. Doudna. „Molecular Mechanisms of RNA Interference”. *Annual Review of Biophysics* 42 (2013.), str. 217–239.
- [5] Thomas A. Cooper, Lili Wan i Gideon Dreyfuss. „RNA and Disease”. *Cell* 136.4 (veljača 2009.), str. 777–793.
- [6] Christina E. Lünse, Anna Schüller i Günter Mayer. „The Promise of Riboswitches as Potential Antibacterial Drug Targets”. *International journal of medical microbiology: IJMM* 304.1 (siječanj 2014.), str. 79–92.
- [7] Laura P. W. Ranum i John W. Day. „Myotonic Dystrophy: RNA Pathogenesis Comes into Focus”. *American Journal of Human Genetics* 74.5 (svibanj 2004.), str. 793–804.
- [8] Savannah Colameco i Marie A. Elliot. „Non-Coding RNAs as Antibiotic Targets”. *Biochemical Pharmacology* 133 (lipanj 2017.), str. 29–42.
- [9] D. H. Mathews i dr. „Expanded Sequence Dependence of Thermodynamic Parameters Improves Prediction of RNA Secondary Structure”. *Journal of Molecular Biology* 288.5 (svibanj 1999.), str. 911–940.
- [10] Ali Mokdad i dr. „Structural and Evolutionary Classification of G/U Wobble Base-pairs in the Ribosome”. *Nucleic Acids Research* 34.5 (2006.), str. 1326–1341.
- [11] Chelsea M. Hull i Philip C. Bevilacqua. „Discriminating Self and Non-Self by RNA: Roles for RNA Structure, Misfolding, and Modification in Regulating the Innate Immune Sensor PKR”. *Accounts of Chemical Research* 49.6 (lipanj 2016.), str. 1242–1249.

- [12] Blake A. Sweeney, Poorna Roy i Neocles B. Leontis. „An Introduction to Recurrent Nucleotide Interactions in RNA”. *Wiley interdisciplinary reviews. RNA* 6.1 (2015.), str. 17–45.
- [13] N. B. Leontis i E. Westhof. „Geometric Nomenclature and Classification of RNA Base Pairs”. *RNA (New York, N.Y.)* 7.4 (travanj 2001.), str. 499–512.
- [14] Neocles B. Leontis, Jesse Stombaugh i Eric Westhof. „The Non-Watson-Crick Base Pairs and Their Associated Isostericity Matrices”. *Nucleic Acids Research* 30.16 (kolovoz 2002.), str. 3497–3531.
- [15] Craig L. Zirbel i dr. „Classification and Energetics of the Base-Phosphate Interactions in RNA”. *Nucleic Acids Research* 37.15 (kolovoz 2009.), str. 4898–4918.
- [16] Jesse Stombaugh i dr. „Frequency and Isostericity of RNA Base Pairs”. *Nucleic Acids Research* 37.7 (travanj 2009.), str. 2294–2312.
- [17] Jörg Rinnenthal i dr. „Mapping the Landscape of RNA Dynamics with NMR Spectroscopy”. *Accounts of Chemical Research* 44.12 (prosinac 2011.), str. 1292–1301.
- [18] Hashim M. Al-Hashimi i Nils G. Walter. „RNA Dynamics: It Is about Time”. *Current opinion in structural biology* 18.3 (lipanj 2008.), str. 321–329.
- [19] Elizabeth A. Dethoff i dr. „Visualizing Transient Low-Populated Structures of RNA”. *Nature* 491.7426 (studeni 2012.), str. 724–728.
- [20] Andrea Haller, Marie F. Soulière i Ronald Micura. „The Dynamic Nature of RNA as Key to Understanding Riboswitch Mechanisms”. *Accounts of Chemical Research* 44.12 (prosinac 2011.), str. 1339–1348.
- [21] Andrew R. Leach. *Molecular Modelling: Principles and Applications*. 2nd ed. Harlow, England ; New York: Prentice Hall, 2001.
- [22] Johannes Kästner. „Umbrella Sampling”. *WIREs Computational Molecular Science* 1.6 (studeni 2011.), str. 932–942.
- [23] Shankar Kumar i dr. „THE Weighted Histogram Analysis Method for Free-Energy Calculations on Biomolecules. I. The Method”. *Journal of Computational Chemistry* 13.8 (1992.), str. 1011–1021.