

Mechanika i Dynamika Molekularna



i nie tylko ...

Równanie Schrödingera

E energia całkowita stanu molekuly opisana jest równaniem Schrödingera:

$$\hat{H}\Psi = E\Psi$$

postać hamiltonianu:

$$\hat{H} = \hat{T} + \hat{V} = -\frac{\hbar^2}{2m} \sum_{i=1}^N \nabla_i^2 - \frac{\hbar^2}{2} \sum_{k=1}^K \frac{1}{M_k} \nabla_k^2 + V(\mathbf{r}, \mathbf{R})$$

operator energii kinetycznej wszystkich e i jąder

operator energii potencjalnej $V(\mathbf{r}, \mathbf{R})$

K ilość jąder

N ilość elektronów

$Z_k e$ ładunek jądra

r_i –współrzędne elektronów

M_k masa jądra

m masa elektronu

$-e$ ładunek elektronu

R_k współrzędne jąder

(elementarny)

Przybliżenie Born-Oppenheimera

zaniedbuje się wszystkie możliwe oddziaływania między ruchem jąder i elektronów

separacja ruchu jąder od ruchu elektronów

HAMILTONIAN CAŁKOWITY = wkład od elektronów + wkład od energii kinetycznej jąder

$$\hat{H} = \hat{H}_0 + \hat{H}'$$

STAN MOLEKULARNY = elektronowa funkcja falowa + jądrowa funkcja falowa
(całkowita funkcja falowa)

$$\Psi_{n,i}(\mathbf{r}, \mathbf{R}) = \phi_n^{\text{el}}(\mathbf{r}) \times \chi_{n,i}(\mathbf{R})$$

ENERGIA CAŁKOWITA = energia kinetyczna jąder + energia elektronów
(uśredniona po ruchu jąder)

$$E_{n,i} = T_{\text{nuc}}(\mathbf{R}) + E_n^0(\mathbf{R}) = \text{const.} \quad \text{niezależna od } \mathbf{r} \text{ i } \mathbf{R} \text{ (parametr)}$$

Mechanika molekularna

Analityczna postać energii potencjalnej jest zazwyczaj wyrażona jako suma członów **wiązanych i niewiązanych**. Składnik oddziaływań **wiązanych** reprezentowany jest w postaci sumy łatwo interpretowalnych i weryfikowalnych eksperymentalnie członów: deformacji wiązań (energia rozciągania), deformacji kątów między wiązaniami (energia zginania), obrotów wokół wiązań (energia torsyjna). Człon **niewiązający** (daleko zasięgowy) złożony jest z członu energii oddziaływań elektrostatycznych oraz oddziaływań van der Waalsa.

Parametryzacja oparta jest na założeniu, że właściwości atomów i fragmentów molekularnych są podobne. Taką parametryzację nazywamy polem siłowym (ang. *force field*).

Mechanika molekularna

Pola siłowe stosowane w badaniach obiektów biologicznych:

- AMBER (Kollman i współpracownicy, UCSF)
- CHARMM (Karplus i współpracownicy, Harvard)
- Gromos (Berendsen i współpracownicy, Nijmegen)
- CVFF i CFF9X (Accelrys Inc.)
- Tripos (Tripos Inc.)

Energia rozciągania wiązania

Krzywa Morse'a:

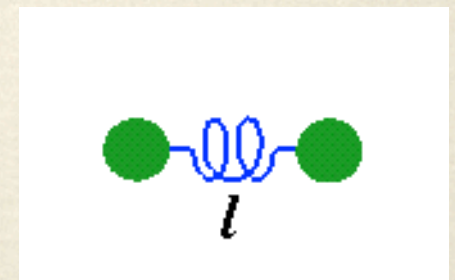
$$U(l) = D_e \{1 - \exp[-a(l - l_0)]\}^2, \text{ gdzie } a = \omega \sqrt{\mu / 2D_e} \text{ i } \omega = \sqrt{k / \mu}$$

gdzie:

μ - masa zredukowana, D_e - głębokość studni potencjału w minimum energii potencjalnej, k - stała siłowa dla danego wiązania, l - odległość pomiędzy atomami, l_0 - odległość pomiędzy atomami w stanie równowagi (w minimum energetycznym).

Potencjał harmoniczny reprezentowany prawem Hooke'a:

$$U(l) = \frac{1}{2}k(l - l_0)^2$$



Energia rozciągania wiązania

Potencjał harmoniczny reprezentowany prawem Hooke'a:

$$U(l) = \frac{1}{2}k(l-l_0)^2$$

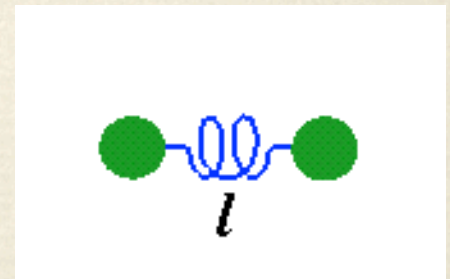
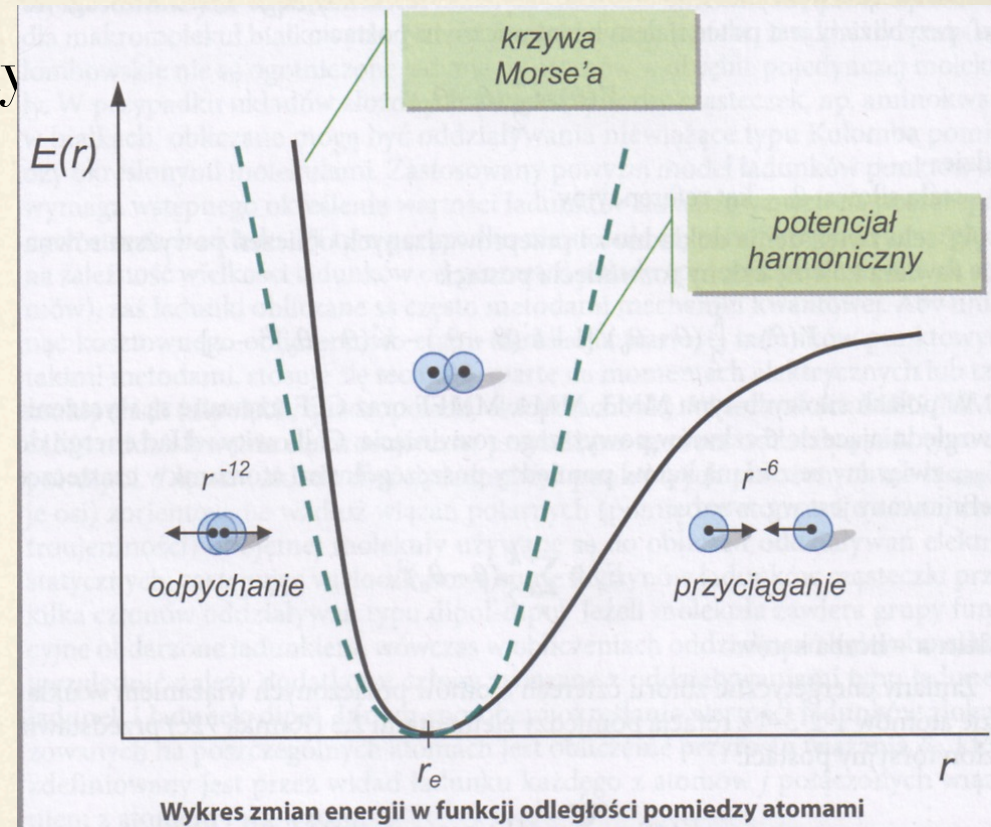
Dla poprawy opisu dysocjacji wiązania chemicznego zapis prawa Hooke'a został rozszerzony o człony anharmoniczne:

$$U(l) = \frac{1}{2}k(l-l_0)^2 [1 - k'(l-l_0) - k''(l-l_0)^2 - \dots]$$

Wkład energii rozciągania wszystkich wiązań chemicznych w cząsteczce to suma:

$$U_{\text{bond}} = \sum_b \frac{1}{2}k(l-l_0)^2$$

gdzie b - liczba wiązań



Energia zginania

Energia zginania związana z odchyleniem kąta pomiędzy trzema atomami od stanu referencyjnego Θ_0 (równowagowego) przybliżona jest potencjałem harmonicznym:

$$U(\Theta) = \frac{1}{2}k(\Theta - \Theta_0)^2$$

gdzie k - stała siłowa, Θ_0 - kąt referencyjny

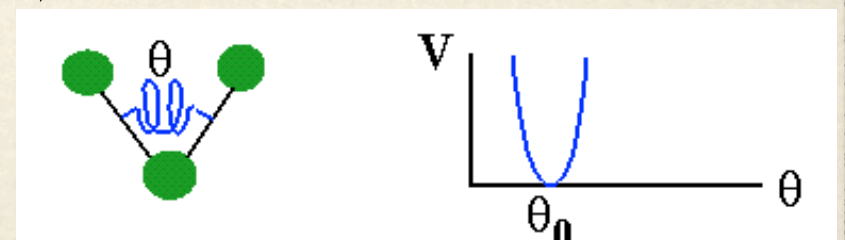
Dla poprawy opisu:

$$U(\Theta) = \frac{1}{2}k(\Theta - \Theta_0)^2 [1 - k'(\Theta - \Theta_0) - k''(\Theta - \Theta_0)^2 - \dots]$$

Wkład energii zginania kątów pomiędzy poszczególnymi atomami w cząsteczce to suma:

$$U_{\text{angle}} = \sum_a \frac{1}{2}k(\Theta - \Theta_0)^2$$

gdzie a - liczba kątów

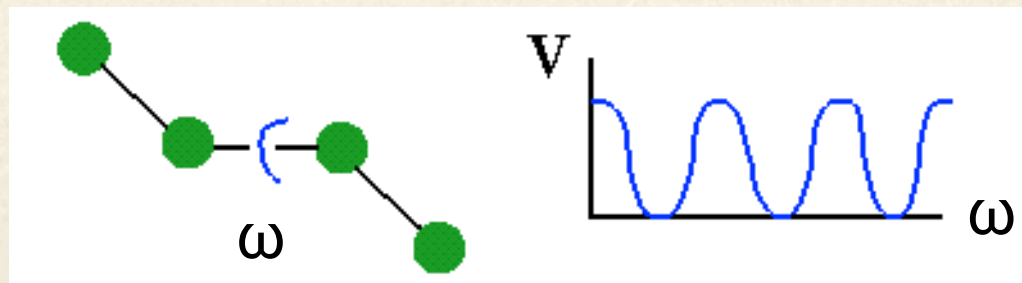


Energia torsyjna

Energia torsyjna:

$$U(\omega) = \sum_{n=1} 1/2 V_n [1 + \cos(n\omega - \gamma)]$$

gdzie n - liczba minimów funkcji w czasie rotacji wiązania ω do kąta pełnego ($\omega = 0^\circ \div 360^\circ$), γ - określa dokładne położenie minimum funkcji, V_n - bariera rotacji



Energia elektrostatyczna

$$U_{\text{Coulomb}} = U(\mathbf{q}) = \sum_{i < j}^N \frac{1}{4\pi\epsilon_0\epsilon} * \frac{q_i q_j}{r_{ij}}$$

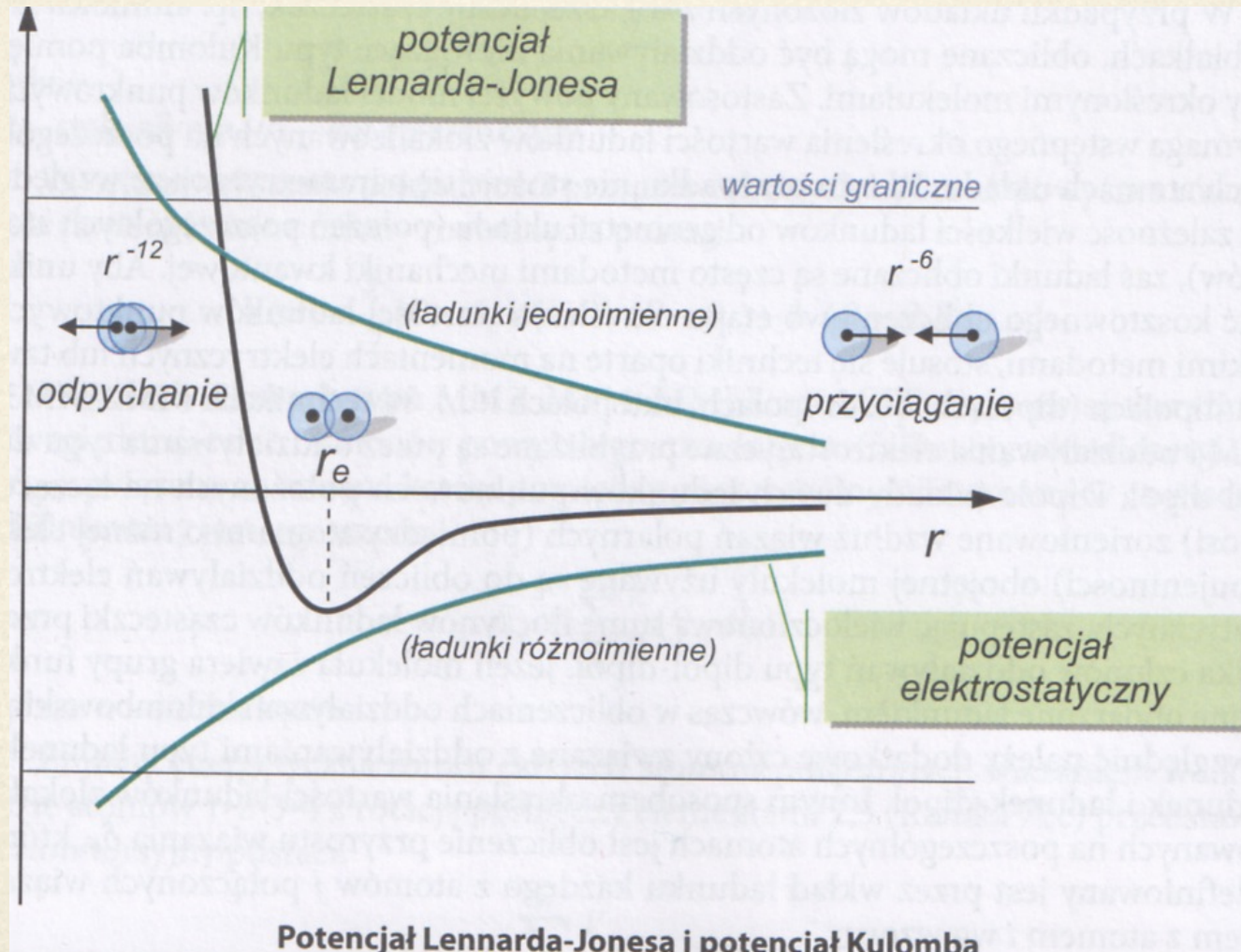
gdzie N - liczba ładunków punktowych molekuly, ϵ_0 - stała dielektryczna w próżni, ϵ - stała dielektryczna w danym ośrodku

Energia Van der Waalsa

$$U_{\text{vdW}} = U(r_{ij}) = \epsilon_{ij}[(R_{ij}^*/r_{ij})^{12} - 2(R_{ij}^*/r_{ij})^6]$$

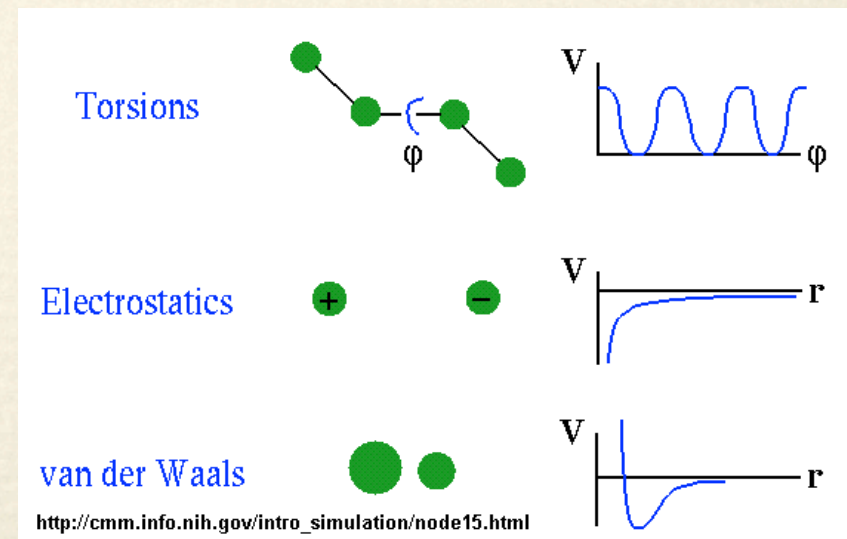
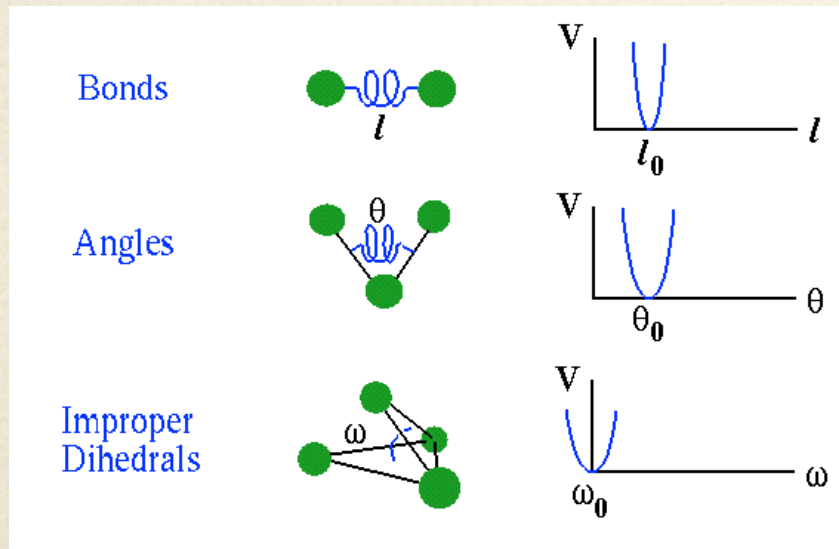
gdzie R_{ij}^* - odległość pomiędzy atomami i i j , gdzie funkcja $U(r_{ij})$ osiąga minimum (suma promieni Van der Waalsa oddziałujących atomów), ϵ_{ij} - głębokości studni potencjału w odległości równowagowej atomów i i j , r_{ij} - odległość pomiędzy atomami i i j

Potencjał L-J i elektrostatyczny



Energia potencjalna układu

$$U_{total} = U_{bond} + U_{angle} + U_{dihedral} + U_{VdW} + U_{Coulomb}$$



Mechanika molekularna

Mechanika molekularna jest oparta na minimalizacji energii układu. W oparciu o fizykę statystyczną można przyjąć, że w niskich temperaturach, właściwości układu są określone przez właściwości stanu podstawowego - czyli stanu o najniższej energii.

Mechanika molekularna

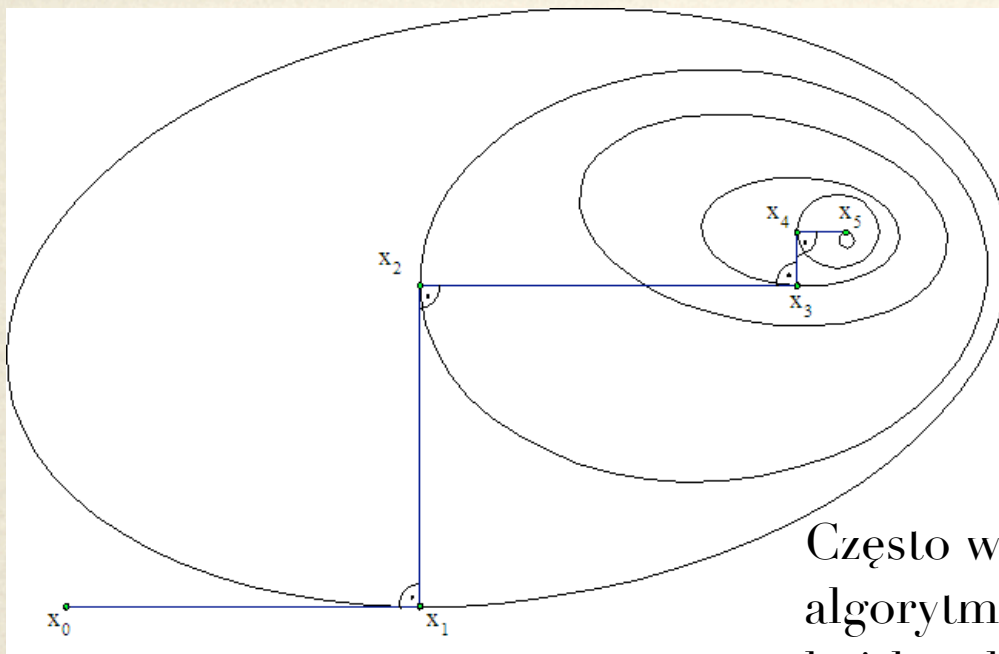
Najczęściej używane algorytmy:

- algorytm najszybszego spadku (ang. steepest descent)
- algorytm sprzężonych gradientów (ang. conjugate gradients)
- algorytm Newtona-Raphsona

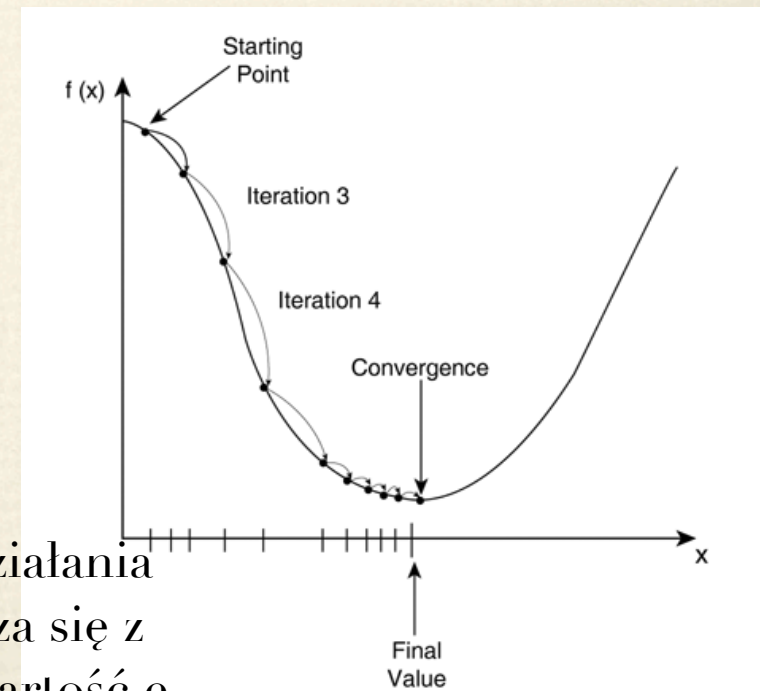
Mechanika molekularna- algorytm najszybszego spadku

Algorytm:

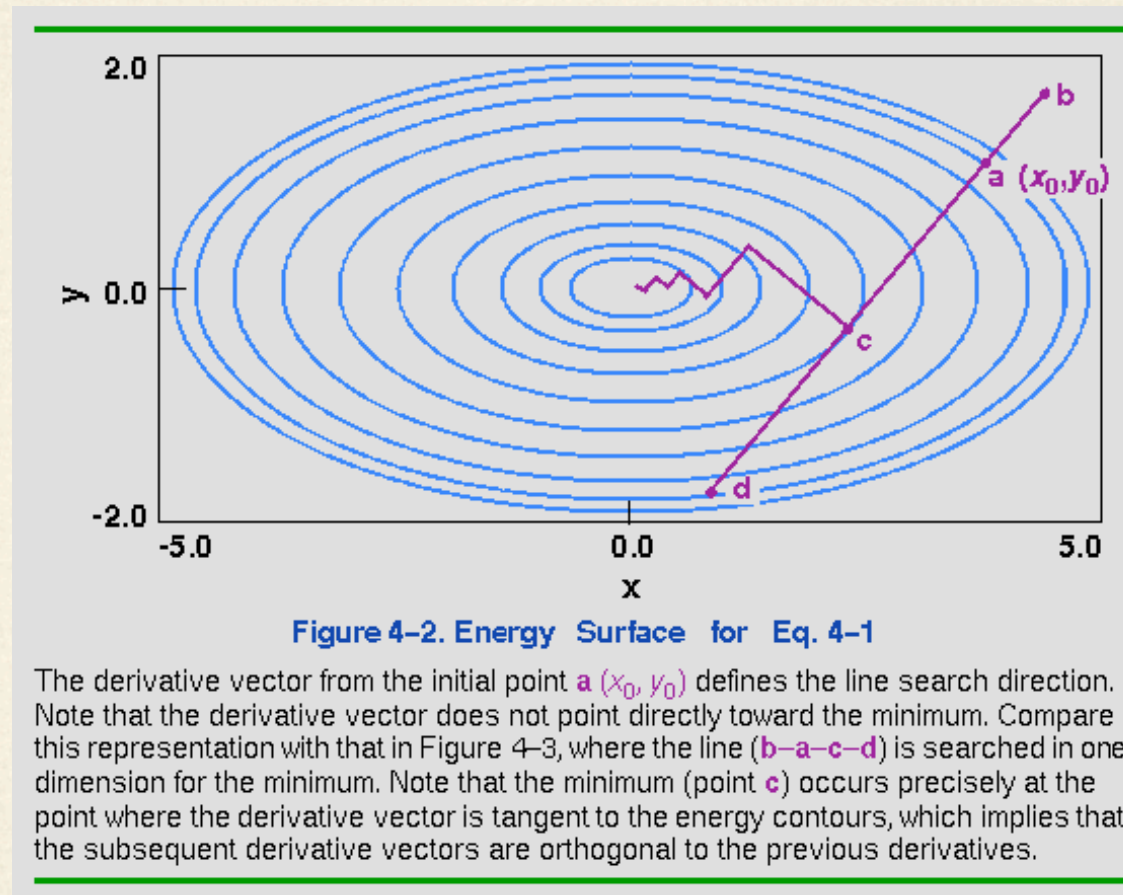
1. wybieramy punkt startowy x_0 , obliczamy wartość funkcji $F_0=f(x_0)$ oraz jej gradient $g_0=g(x_0)$
2. wyznaczamy kierunek poszukiwań jako $x=-g_0$
3. wzdłuż kierunku x wykonujemy krok o długości e oraz określamy współrzędne nowego punktu $x_{i+1}=x_i+ex$ (dla pierwszej iteracji $x_i = x_0$)
4. obliczamy w nowym punkcie wartość $F=f(x_{i+1})$ oraz gradientu $g=g(x_{i+1})$
5. jeżeli krok był pomyślny ($F<F_0$) to powtarzamy od drugiego punktu podstawiając g w miejsce g_0



Często w trakcie działania algorytmu zmniejsza się z każdym krokiem wartość e



Mechanika molekularna- algorytm najszybszego spadku

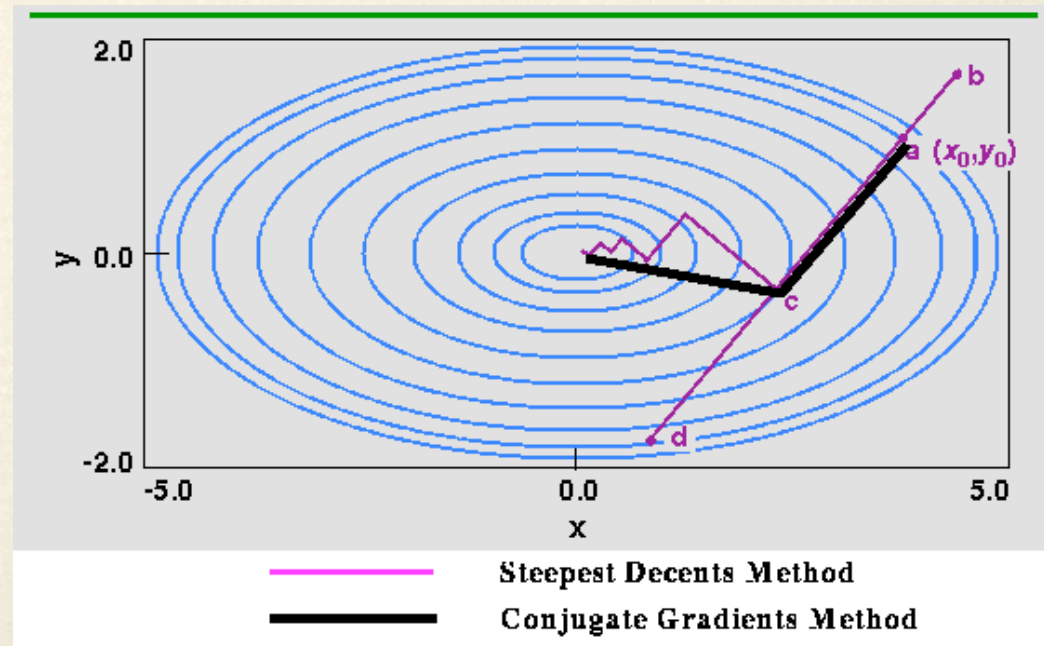


Algorytm ten jest prosty i niezawodny, z małymi wymaganiami ale niestety w okolicy minimum mało efektywny.

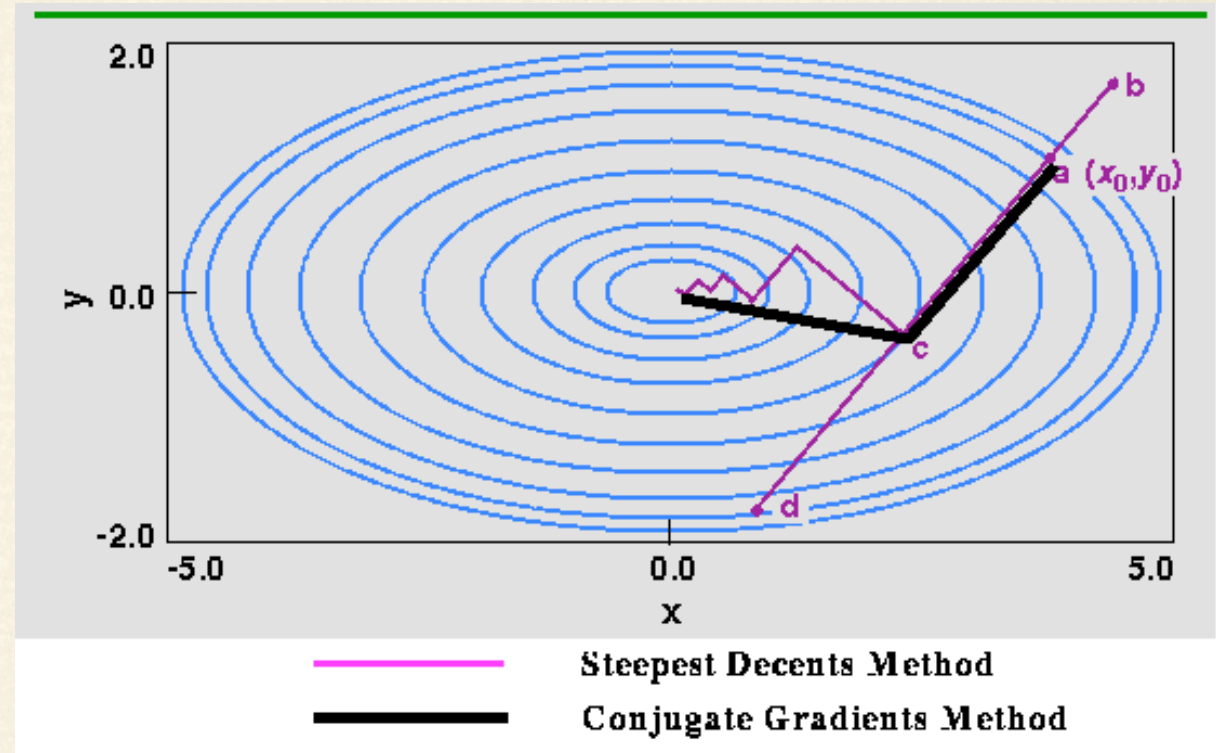
Mechanika molekularna- algorytm sprzężonych gradientów

Algorytm:

1. wybieramy punkt startowy x_0 , obliczamy wartość funkcji $F_0=f(x_0)$ oraz jej gradient $g_0=g(x_0)$
2. wyznaczamy kierunek poszukiwań jako $s_1=-g_0$
3. wzdłuż kierunku x wykonujemy krok o długości e oraz określamy współrzędne nowego punktu $x_{i+1}=x_i+e s_i$
4. obliczamy w nowym punkcie wartość $F=f(x_{i+1})$ oraz gradientu $g=g(x_{i+1})$
5. wyznaczamy kierunek poszukiwań jako $s_{i+1}=-g(x_{i+1})+ [|g(x_{i+1})|^2/|g(x_i)|^2]s_i$ kolejny kierunek poszukiwań sprzężony z poprzednim



Mechanika molekularna- algorytm sprzężonych gradientów



Szybszy od algorytmu najszybszego spadku, z małymi wymaganiami ale zawodny daleko od minimum.

Mechanika molekularna

Algorytm najszybszego spadku
i algorytm sprzężonych gradientów
są algorytmami używającymi jedynie informacji
o pierwszych pochodnych. Dzięki temu są niezawodne ale
i wolno zbieżne, zwłaszcza w pobliżu minimum.

Mechanika molekularna- algorytm Newtona - Raphsona

Algorytm Newtona Raphsona (i pochodne):

1. znajdź powierzchnię paraboloidalną najbliższą do powierzchni energii potencjalnej
2. zejść do minimum paraboloidy
3. oblicz gradient energii
4. jeśli gradient jest mniejszy niż założony próg idź do punktu 6
5. jeśli gradient nie jest mniejszy niż założony próg oblicz macierz drugich pochodnych i wróć do punktu 1.
6. to jest minimum

Mechanika molekularna- algorytm Newtona - Raphsona

Algorytm Newtona wykorzystuje informacje o drugich pochodnych, dlatego jest szybko zbieżny w pobliżu minimum. Ale kiedy molekula znajduje się daleko od minimum jest niestabilny.

Wymaga on również obliczania i przechowywania macierzy drugich pochodnych rozmiaru $N \times N$, co dla dużych molekuł jest niepraktyczne.

Mechanika molekularna

Używamy kombinacji algorytmów: niewielka ilość kroków (~ 100) algorytmem najszybszego spadku, dla usunięcia największych naprężeń

Dla małych układów (< 100 atomów): podobna ilość kroków przy użyciu algorytmu gradientów sprzężonych a na zakończenie obliczenia kilkadziesiąt iteracji algorytmu Newtona

Dla dużych układów (> 100 atomów): kilkaset (może być kilka tysięcy w zależności od wielkości układu) kroków przy użyciu algorytmu gradientów sprzężonych

Mechanika molekularna

Zastosowania algorytmów minimalizacji energii:

- Znajdowanie stabilnych konformacji molekuł i układów molekularnych. Minimalizacja energii prowadzi do znalezienia lokalnego minimum energii
- Badanie widma (badanie drgań harmoniczných)
- Przewidywanie struktur molekuł
- Dokowanie
- Przewidywanie energii wiązania

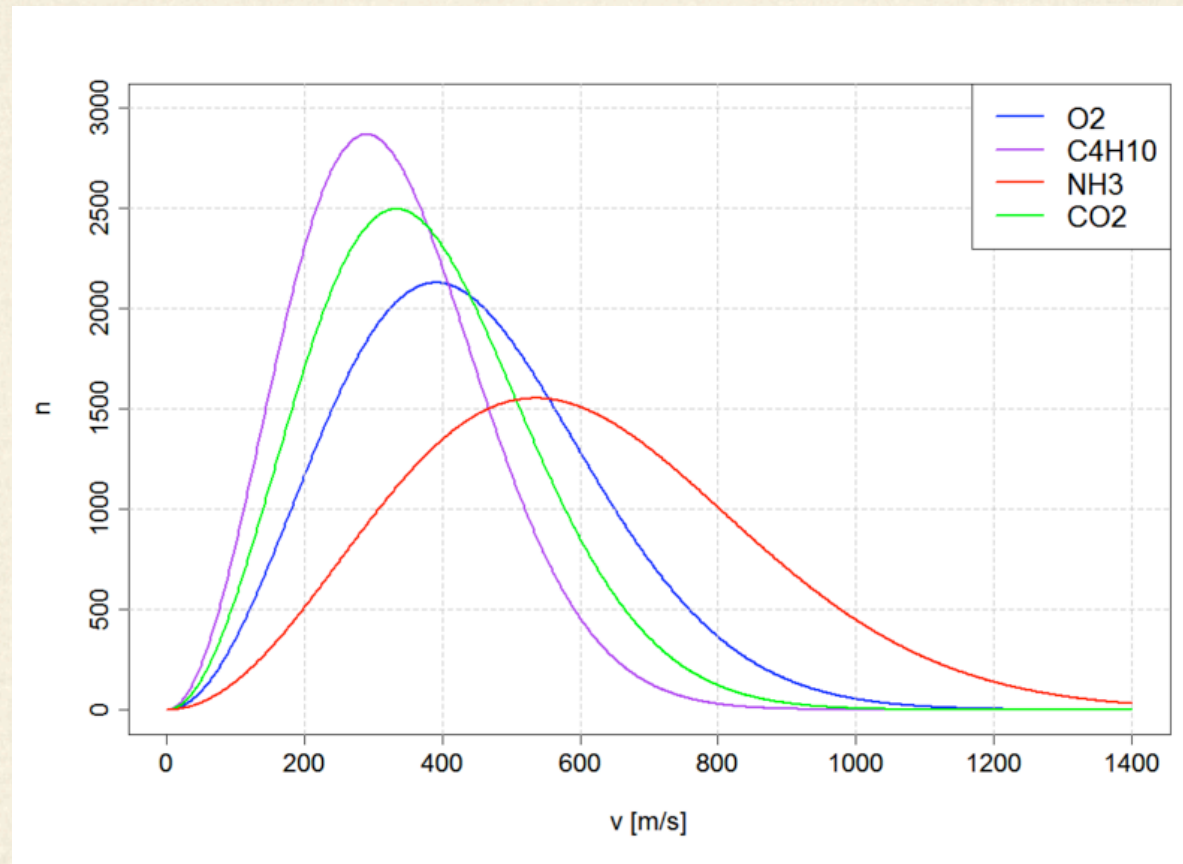
Dynamika molekularna

Dynamika molekularna (MD) służy do obliczania trajektorii w przestrzeni fazowej zbioru molekuł, z których każda oddzielnie podlega klasycznym równaniom ruchu.

(D. W. Heerman; Podstawy Symulacji komputerowych w Fizyce)

$$\vec{F}_i = m_i \frac{d^2 \vec{r}_i}{dt^2} \quad \vec{F}_i = -\vec{\nabla}_{\vec{r}_i} V(\vec{r}_1, \dots, \vec{r}_N)$$

Dynamika molekularna



Rozkład Maxwella dla **tlenu**, **butanu**, **amoniaku** i **dwutlenku węgla** w temperaturze pokojowej (20 °C). Na 1 milion cząsteczek, n będzie poruszać się z prędkością v

Algorytm Verleta

Równanie ruchu:

$$\ddot{\vec{r}}_i(t, \vec{r}_i(t)) = \frac{\mathbf{F}_i(t, \vec{r}_i(t))}{m_i}, i=1, 2, \dots, N$$

I po przekształceniach możemy wyznaczyć prędkości:

$$r(t + \Delta t) = 2r(t) - r(t - \Delta t) + \frac{\mathbf{F}(t)}{m} \Delta t^2$$

$$\mathbf{v}(t) = \frac{r(t + \Delta t) - r(t - \Delta t)}{2\Delta t}$$

gdzie: Δt - krok czasowy $\mathbf{v}(t)$ - prędkość w chwili t

$r(t)$ - położenie w chwili t $\mathbf{F}(t)$ - siła w chwili t

Algorytm Verleta

- ❖ aby wyznaczyć przyszłe położenia musimy znać położenia obecne i przeszłe
- ❖ algorytm ten jest szybki i nie wymaga dużej ilości pamięci
- ❖ jest stabilny, numerycznie efektywny i wiernie odwzorowuje trajektorie atomów
- ❖ łatwo go zaimplementować
- ❖ algorytm wymaga wygenerowania prędkości do wystartowania
- ❖ aby wyznaczyć prędkość w kroku n -tym potrzebna jest informacja o położeniu w kroku $n+1$

Algorytm Verleta

algorytm prędkościowy (velocity Verlet)

$$r(t + \Delta t) = r(t) + \mathbf{v}(t)\Delta t + \frac{\mathbf{F}(t)}{2m}\Delta t^2$$

$$\mathbf{v}(t + \Delta t) = \mathbf{v}(t) + \frac{\mathbf{F}(t + \Delta t) + \mathbf{F}(t)}{2m}\Delta t$$

algorytm skokowy (leap-frog method)

$$r(t + \Delta t) = r(t) + \Delta t\mathbf{v}(t + \Delta t/2)$$

$$\mathbf{v}(t + \Delta t/2) = \mathbf{v}(t - \Delta t/2) + \Delta t\frac{\mathbf{F}(t)}{m}$$

gdzie: Δt - krok czasowy $\mathbf{v}(t)$ - prędkość w chwili t

$r(t)$ - położenie w chwili t $\mathbf{F}(t)$ - siła w chwili t

Algorytm Verleta

Standardowy krok całkowania (Δt) dla układów biologicznych w reprezentacji pełnoatomowej to 1fs

Zespoły MD

Zespół mikrokanoniczny (NVE) ustalona, stała energia, objętość i stała liczba atomów/cząsteczek

$$H = \frac{1}{2} \sum_i \frac{1}{m} p_i^2 + \sum_{i < j} V(r_{ij}) = E$$

Zespół kanoniczny (NVT) stała jest temperatura, objętość i stała liczba atomów/cząsteczek

$$\frac{1}{2} \sum_i m v_i^2 = \frac{1}{2} \sum_i \frac{1}{m} p_i^2 = \text{const}$$

Zespół wielki kanoniczny (μ VT) stała jest temperatura i objętość i tzw. potencjał chemiczny, układ może wymieniać atomy/cząstki z otoczeniem

Zespół izotermiczno-izobaryczny (NPT) stała jest temperatura i ciśnienie, i i stała liczba atomów/cząsteczek

MD

Jest kilka warunków, których spełnienie jest wymagane podczas symulacji.

Warunek I

minimalizacja + termalizacja układu

minimalizacja układu dla usunięcia wstępnych naprężeń

termalizacja układu:

-wielokrotne losowanie prędkości początkowych dla stopniowo rosnącej temperatury

test stabilności

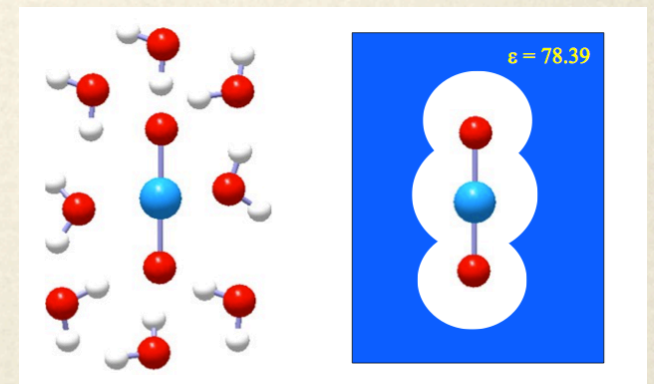
-symulacja układu dla stałej energii i monitorowanie temperatury

Warunek II

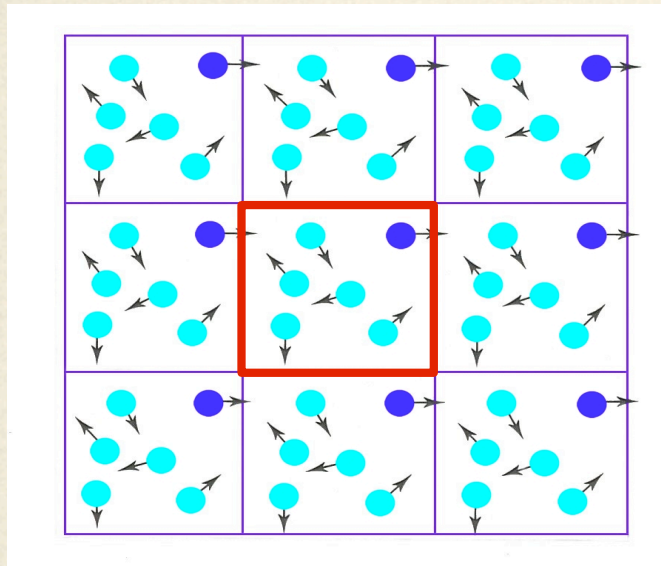
Białko nie jest w próżni !!!

Wprowadzamy środowisko do układu

- Przez reprezentację wody z rozpuszczonymi w niej jonami explicite.
- Przez reprezentację wody za pomocą modelu kroplowego.
- Przez reprezentację wody za pomocą modelu ciągłego (PCM - ang. *Polarizable Continuum Model*).
- Wprowadzenie uproszczonej błony komórkowej.



Warunek III



Włączenie rozpuszczalnika do obliczeń musi wprowadzić do symulacji periodyczne warunki brzegowe (ang. *Periodic Boundary Conditions*). Układ jest wtedy zbudowany z nieskończonej liczby identycznych elementów zwanych „pudelkami” (ang. *Periodic Box*), ułożonymi w taki sposób aby między ich granicami nie było wolnej przestrzeni.

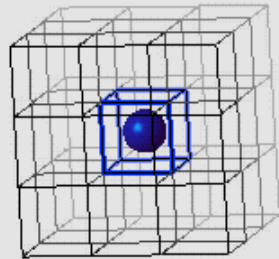
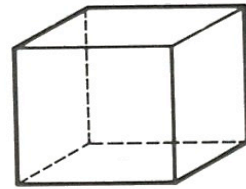


Figure 2-3. Solute Surrounded by Solvent

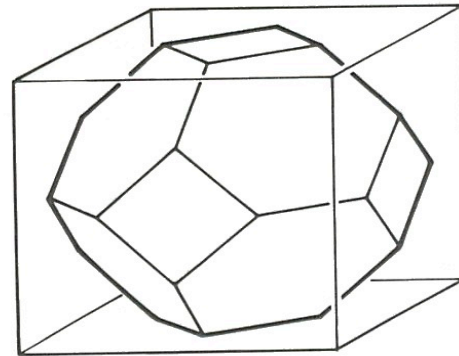
A solute surrounded by an isolated cube of solvent is replicated periodically in three dimensions in order to better represent a bulk or crystalline environment.

Warunek III

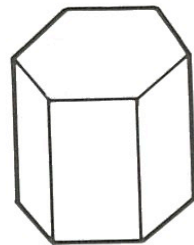
278 Computer simulation methods



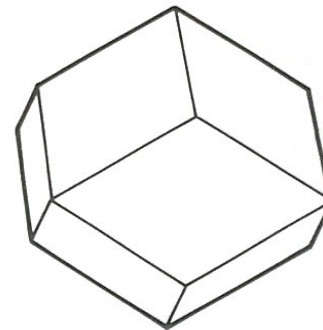
Cube



Truncated octahedron



Hexagonal prism



Rhombicuboctahedron

Warunek IV

Wprowadzamy uproszczony model

obliczania oddziaływań dalekozasięgowych

Istnieje kilka metod obliczania oddziaływań dalekozasięgowych, które korzystają z tego, że maleją one wraz ze wzrostem odległości:

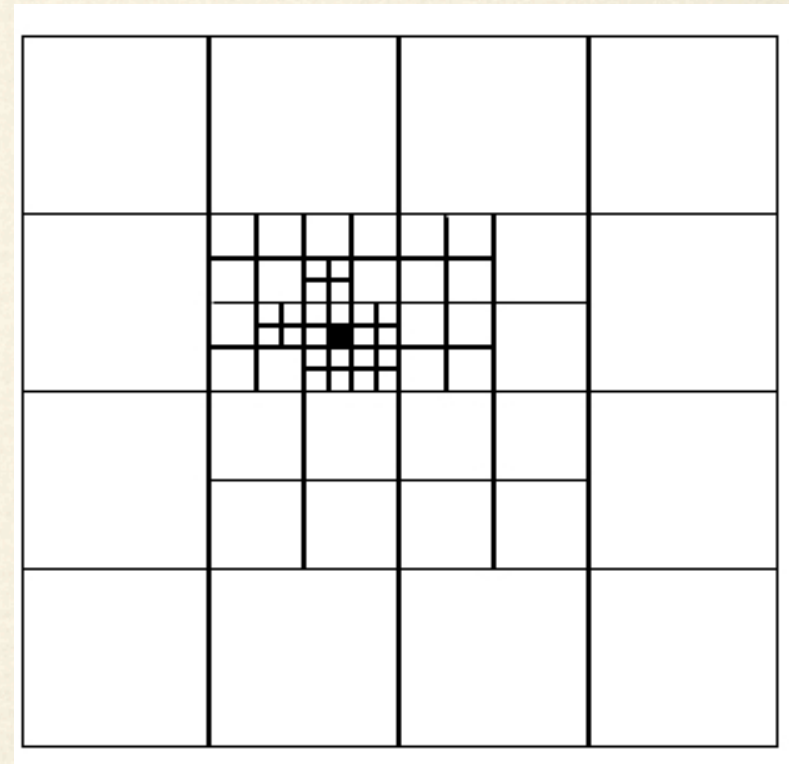
- Metoda odcinania oddziaływań (ang. *cutoff*)
- Metoda multipoli komórkowych (ang. *Cell Multipole Method*)
- Metoda pola reakcji
- Metoda Ewalda

Warunek IV

Metoda odcinania oddziaływań (ang. *cutoff*) zakłada, że dla dostatecznie dalekich obiektów możemy nie liczyć oddziaływań między nimi. Wprowadza się do obliczeń sferyczną granicę oddziaływania potencjału, do której liczymy oddziaływania.

Warunek IV

Metoda multipoli komórkowych (ang. *Cell Multipole Method*) dzieli układ na małe zawierające kilkanaście atomów kostki. Kostki są grupowane w kostki wyższego rzędu. Dla każdej kostki liczymy wypadkowy ładunek dipol i wyższe multipole. W kostkach elementarnych obliczamy te wielkości z wielkości atomowych. Dla komórek wyższego rzędu obliczamy je z uśredniania wartości dla komórek podstawowych.



Warunek IV

Metoda pola reakcji opiera się na fizycznym modelu oddziaływania układu z ciągłym rozpuszczalnikiem o określonej stałej dielektrycznej i zadany rozkładzie gęstości jonowej oraz z rozkładem dipoli na brzegu.

Warunek IV

Metoda Ewalda (PME - ang. *Particle - Mesh Ewald*), jest szybką numeryczną metodą wykorzystującą sumę Ewalda.

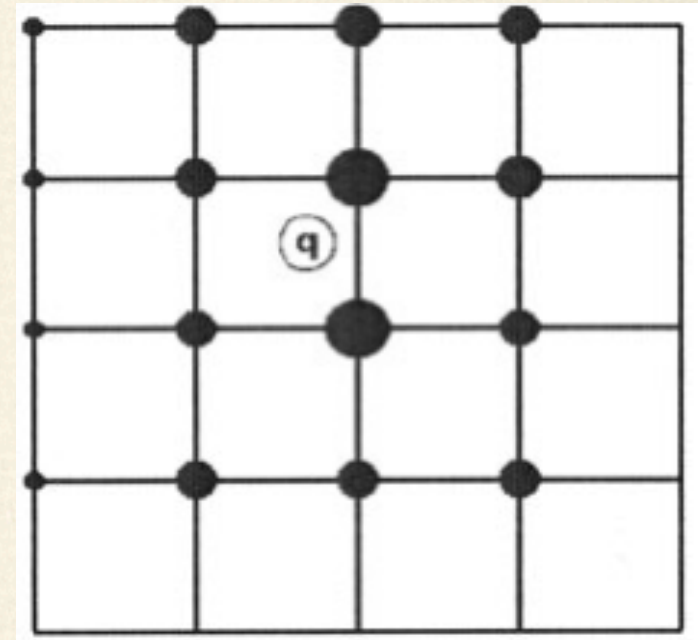
Suma Ewalda jest to nieskończona suma opisująca oddziaływania między ładunkami dla elektrycznie obojętnego układu, dla której wynik zależy od kolejności sumowania. Sumuje się w następującej kolejności: najpierw suma po pierwszym pudle, następnie sumy po kolejnych sferach pudeł o coraz większych promieniach.

Metodę Ewalda opisuje równanie składające się z czterech składników:

$$E_{Ewald} = E_{bezp} + E_{odw} + E_{enw^3} + E_{pow}$$

Warunek IV

W algorytmie PME parametr β dobierany jest tak, aby ciężar obliczeń spoczywał głównie na sumie odwrotnej, podczas gdy koszt obliczeń sumy bezpośredniej jest proporcjonalny do N . Wtedy koszt obliczeń PME jest proporcjonalny do $M \log N$. W PME ładunki są rozkładane na sieci punktów przy pomocy funkcji wagi odpowiednich do odległości punktów sieci od oryginalnego położenia ładunku. Rozłożenie wszystkich ładunków na siatce umożliwia zastosowanie szybkiej transformacji Fouriera (FFT ang. *Fast Fourier Transform*) do obliczenia sumy odwrotnej, co wydatnie skraca czas obliczeń



Warunek IV

Dla oddziaływań van der Waalsa stosuje się najczęściej metodę odcinania. Dla oddziaływań elektrostatycznych pozostałe metody. Metoda multipoli komórkowych, metoda Ewalda i metoda pola reakcji, zapewniają wysoką dokładność przy umiarkowanym koszcie obliczeniowym.

Przygotowanie układu topologia

```

RESI MET          0.00
GROUP
ATOM N    NH1    -0.47 ! |
ATOM HN   H      0.31 ! HN-N
ATOM CA   CT1    0.07 ! | HB1 HG1  HE1
ATOM HA   HB     0.09 ! | | | |
GROUP                    ! HA-CA--CB--CG--SD--CE--HE3
ATOM CB   CT2   -0.18 ! | | | |
ATOM HB1  HA    0.09 ! | HB2 HG2  HE2
ATOM HB2  HA    0.09 ! O=C
GROUP                    ! |
ATOM CG   CT2   -0.14
ATOM HG1  HA    0.09
ATOM HG2  HA    0.09
ATOM SD   S     -0.09
ATOM CE   CT3   -0.22
ATOM HE1  HA    0.09
ATOM HE2  HA    0.09
ATOM HE3  HA    0.09
GROUP
ATOM C    C      0.51
ATOM O    O     -0.51
BOND CB CA  CG CB SD CG CE SD
BOND N HN  N CA C CA C +N
BOND CA HA  CB HB1 CB HB2 CG HG1 CG HG2
BOND CE HE1 CE HE2 CE HE3
DOUBLE O C
IMPR N -C CA HN C CA +N O
DONOR HN N
ACCEPTOR O C
IC -C CA *N HN  1.3478 124.2100 180.0000 114.3900 0.9978
IC -C N  CA C   1.3478 124.2100 180.0000 106.3100 1.5195
IC N  CA C  +N  1.4510 106.3100 180.0000 117.7400 1.3471
IC +N CA *C O   1.3471 117.7400 180.0000 120.6400 1.2288
IC CA C  +N +CA 1.5195 117.7400 180.0000 124.5200 1.4471
IC N  C  *CA CB 1.4510 106.3100 121.6200 111.8800 1.5546
IC N  C  *CA HA 1.4510 106.3100 -116.9800 107.5700 1.0832
IC N  CA CB CG  1.4510 111.2500 180.0000 115.9200 1.5460
IC CG CA *CB HB1 1.5460 115.9200 120.5600 106.9000 1.1153
IC CG CA *CB HB2 1.5460 115.9200 -124.8000 109.3800 1.1129
IC CA CB CG SD  1.5546 115.9200 180.0000 110.2800 1.8219
IC SD CB *CG HG1 1.8219 110.2800 120.5000 110.3400 1.1106
IC SD CB *CG HG2 1.8219 110.2800 -121.1600 109.6400 1.1119
IC CB CG SD CE  1.5460 110.2800 180.0000 98.9400 1.8206
IC CG SD CE HE1 1.8219 98.9400 -179.4200 110.9100 1.1111
IC HE1 SD *CE HE2 1.1111 110.9100 119.9500 111.0300 1.1115
IC HE1 SD *CE HE3 1.1111 110.9100 -119.9500 111.0900 1.1112

```


Przygotowanie układu woda

Minimalizacja i ekwilibracja wody:

- Minimalizacja energii potencjalnej wody i jonów
- 300K Stałe ciśnienie (zespół izotermiczno-izobaryczny)
- Białko „zamrożone”
- Woda i jony ulegają ekwilibracji przez ok. 100-200 ps

Przygotowanie układu

Podgrzewanie układu

- „Odmrażamy” białko.
- Cały układ minimalizujemy (energię potencjalną)
- Podgrzewamy układ od 0K do 300K w trakcie 50-100 ps

Przygotowanie układu

Dynamika właściwa

- wykonujemy symulację w temperaturze 300K (zespół kanoniczny lub izotermiczno-izochoryczny lub wielki kanoniczny)
- w szczególnych przypadkach temperaturę można zmienić
- długość symulacji – im symulacja dłuższa tym lepiej.

Charakterystyka MD

- ❖ algorytmy wielokrokowe – wykorzystywanie danych w oparciu o poprzednie kroki
- ❖ generowania warunków początkowych
- ❖ duża ilość zmiennych
- ❖ wysoki koszt obliczeń
- ❖ symulacja procesów o dużej rozpiętości skal czasowych (od pikosukund do miliardów lat - symulacje)

Zadania MD

- ❖ wyznaczanie trajektorii złożonych układów cząsteczek czyli zależności położeń cząstek od czasu
- ❖ modelowanie dynamiki układów biologicznych (białka, błony lipidowe, kwasy nukleinowe), chemicznych (reakcje chemiczne), fizycznych (adsorpcja, dynamika kryształów)
- ❖ modelowanie zwijania białek
- ❖ wyznaczanie równowagowych konformacji układów
- ❖ badanie oddziaływania białko-białko lub białko-ligand
- ❖ wyznaczanie wielu średnich wielkości fizycznych w badanych układach (np. energii kinetycznej, potencjalnej, odległości atomów, ciśnienia, objętości, korelacji ruchów atomów, korelacji czasowych itp.)

SMD

Sterowana dynamika molekularna (SMD – ang. *Streered Molecular Dynamics*) umożliwia modelowanie takich biologicznych procesów w których niezbędne jest przejście z jednego stanu równowagi do drugiego

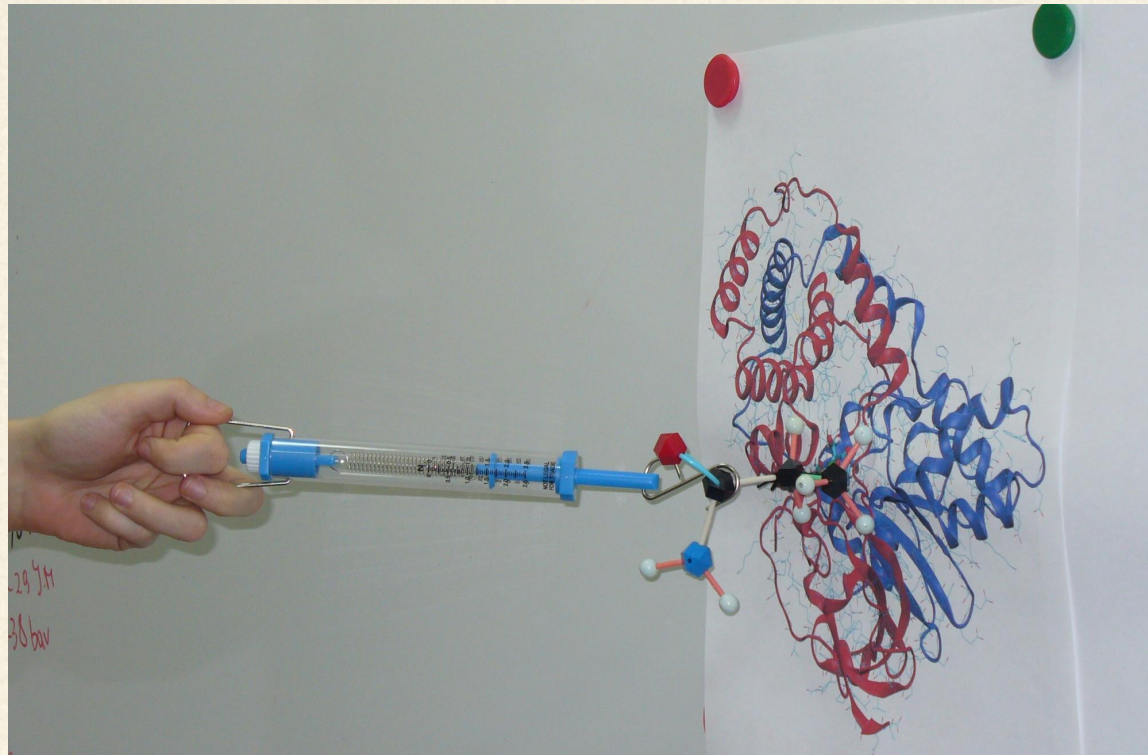
SMD

Metoda ta opiera się na wprowadzeniu sił zewnętrznych do układu, które będą działały na wybrany atom lub grupę atomów.

Siły wprowadzamy jako:

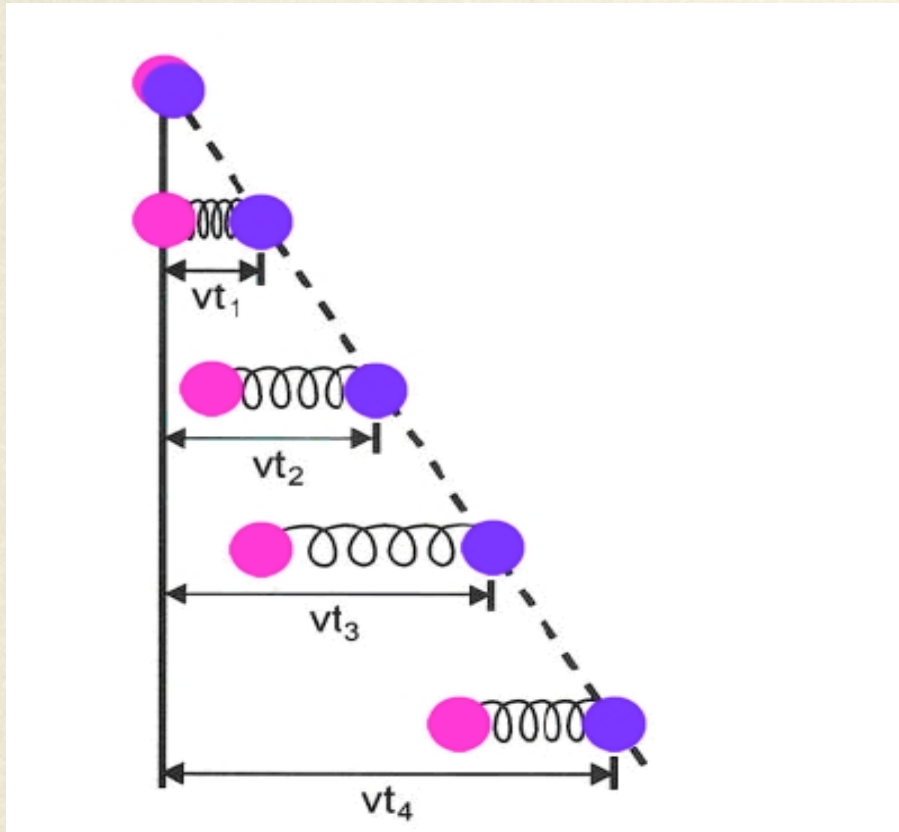
- stałą siłę przyłożoną do atomu (lub grupy atomów),
- więzy harmoniczne przyłożone do atomu (lub grupy atomów).

SMD



W pierwszym przypadku przykładany jest stały wektor siły do atomu (lub grupy atomów). Wektor ten jest dodawany do sił określonych w polu siłowym. Wymusza to ruch w określonym kierunku.

SMD

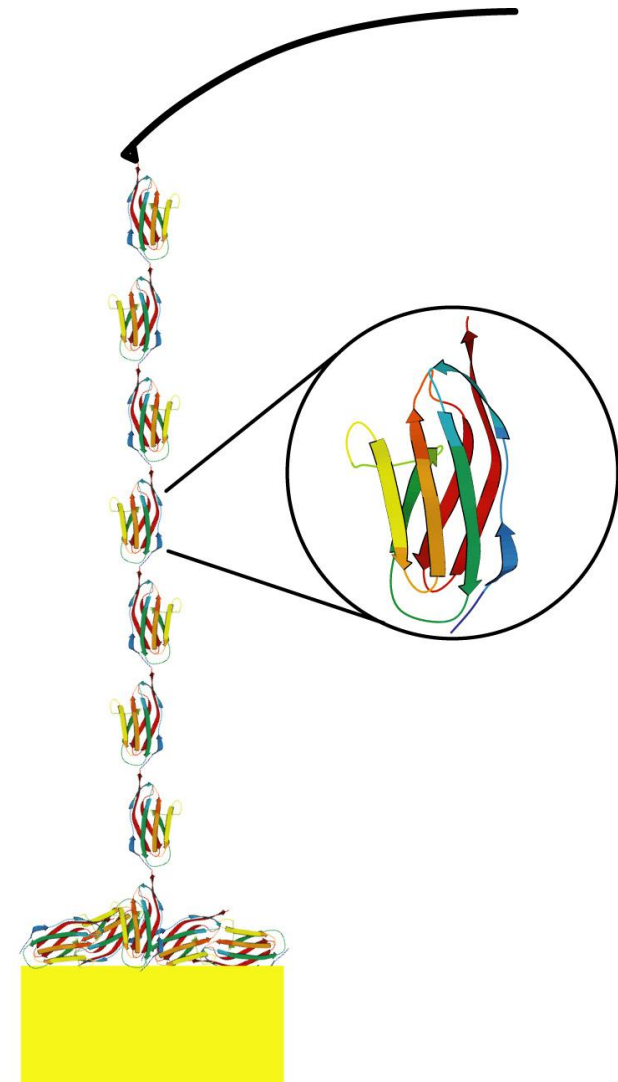
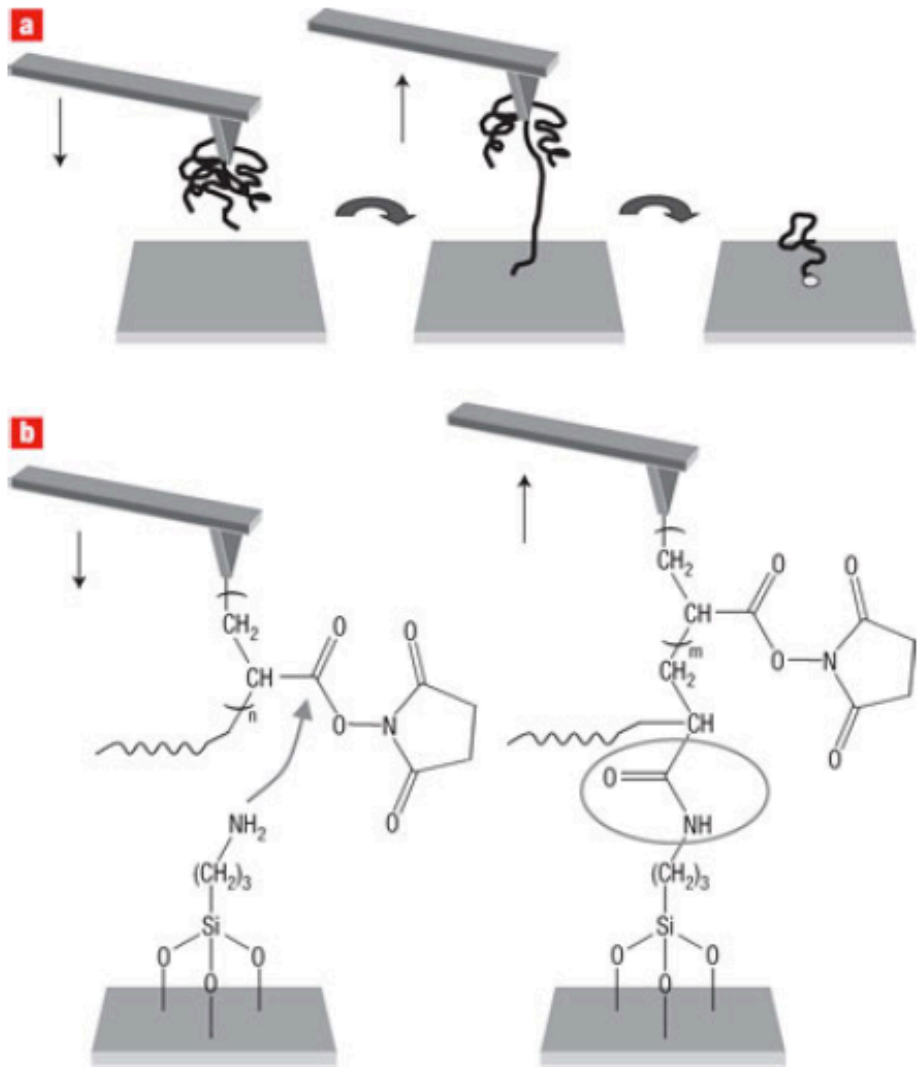


W drugim przypadku pojedynczy atom (lub środek masy grupy atomów) łączy się więzami harmonicznymi z wybranym punktem w przestrzeni, który następnie jest przesuwany z zadaną prędkością. Jego ruch wymusza ruch zaczepionego do niego atomu.

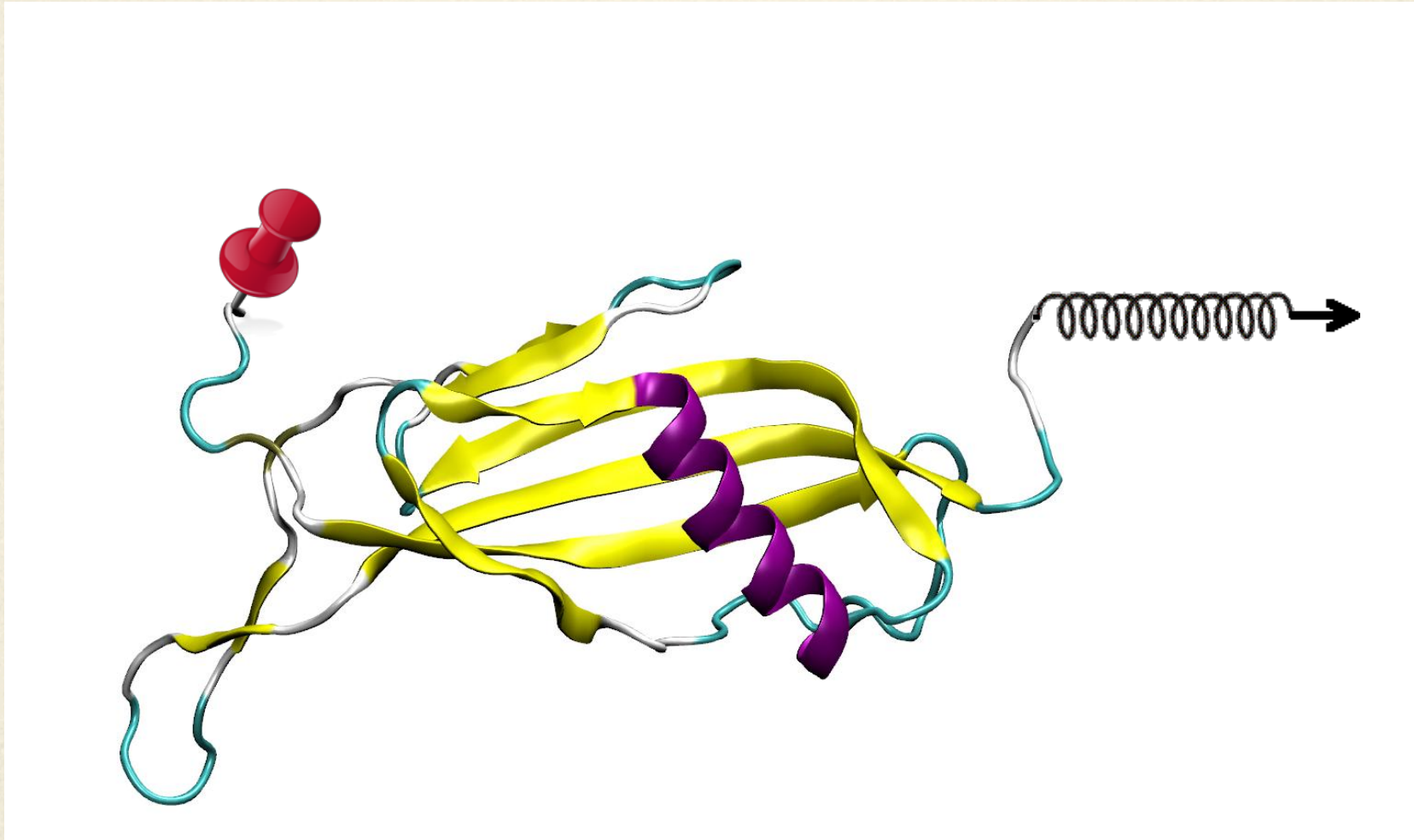
Proces ciągnięcia ze stałą prędkością w przypadku jednowymiarowym. Punkt ciągnięty oznaczony jest kolorem chabrowym, zaczepiony atom kolorem różowym

AFM

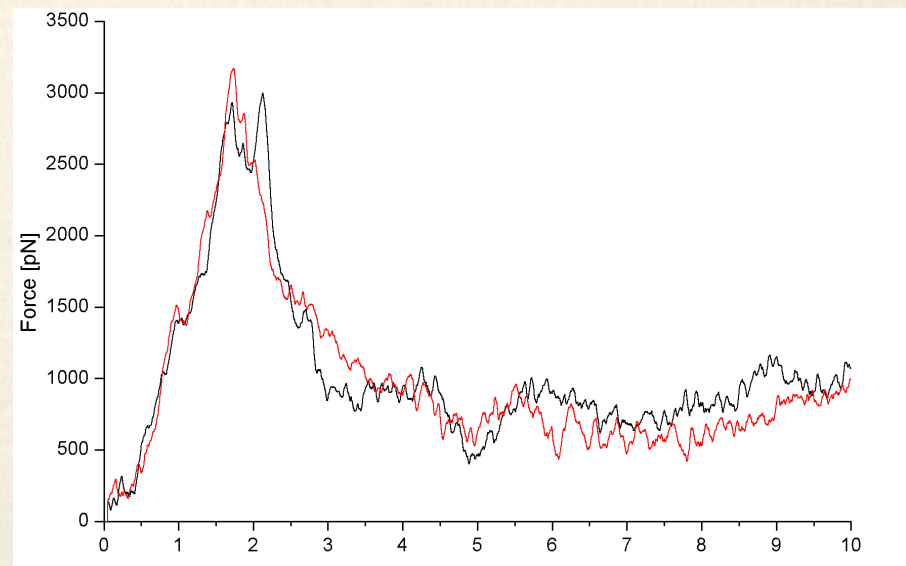
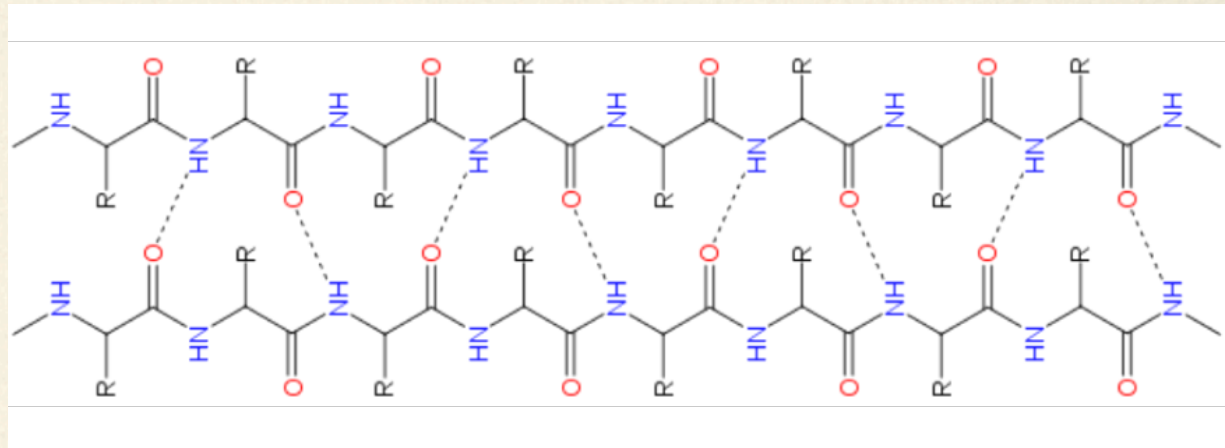
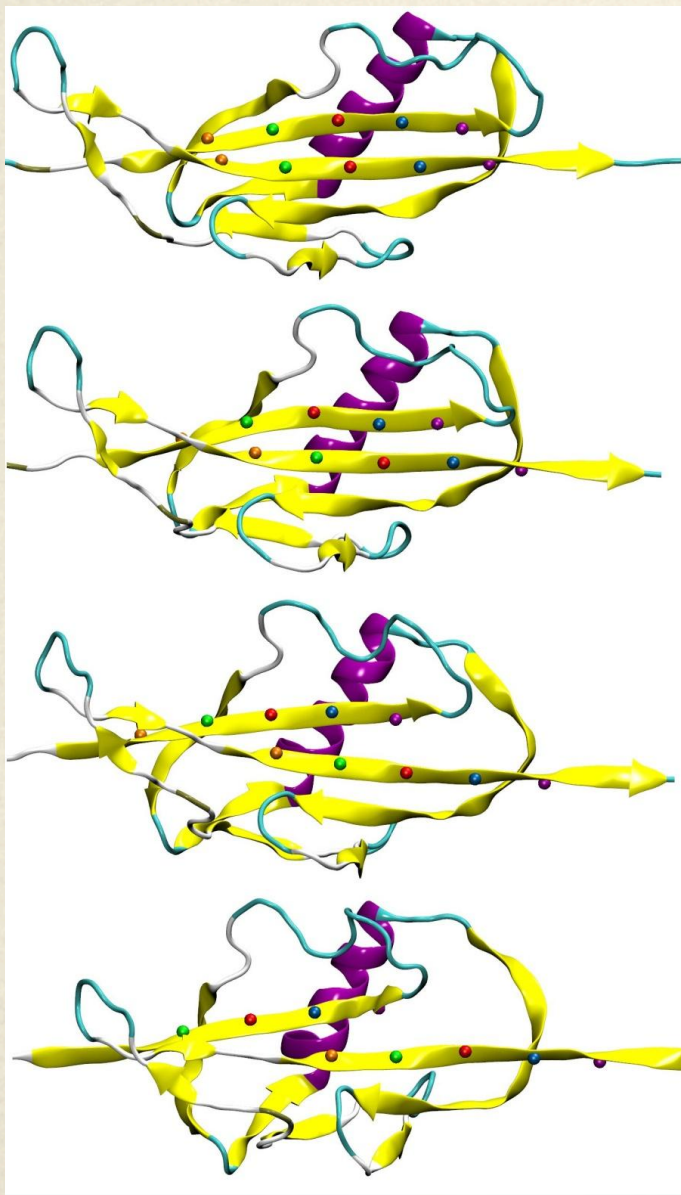
From
<http://www.physics.rice.edu/bio.cfm>



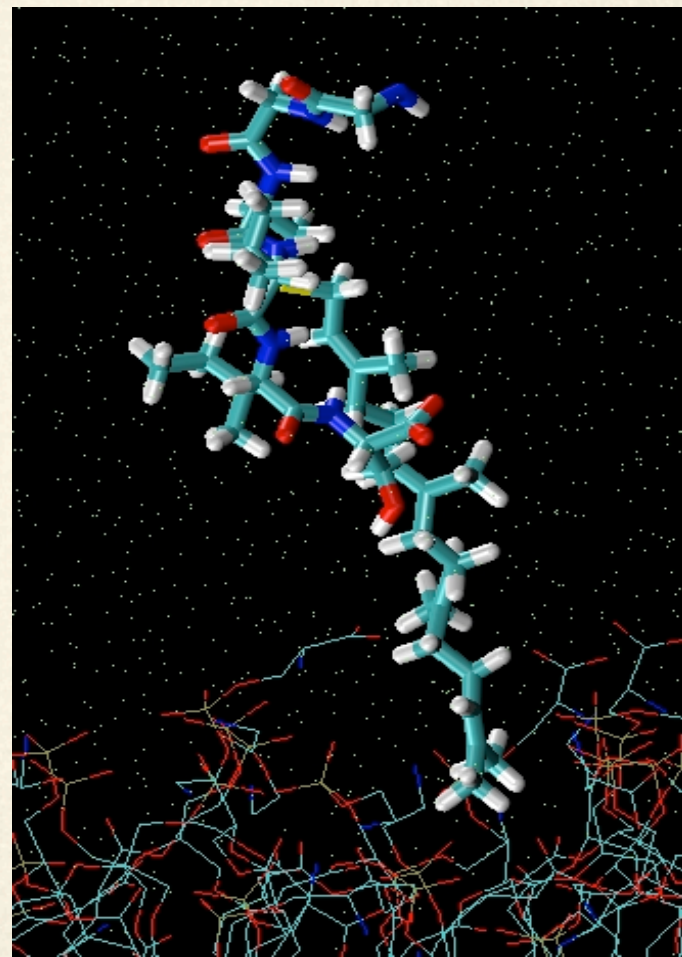
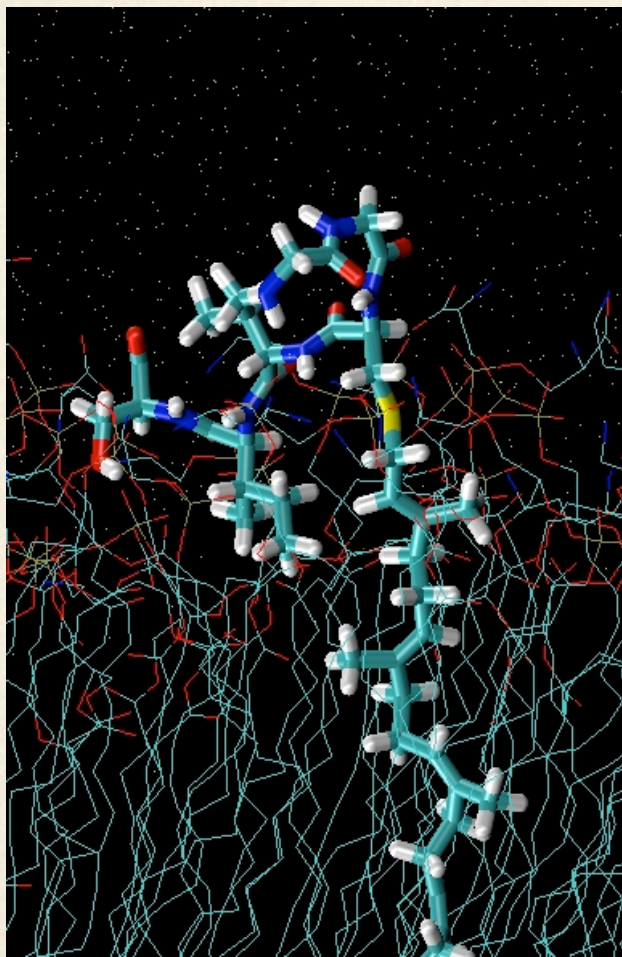
SMD



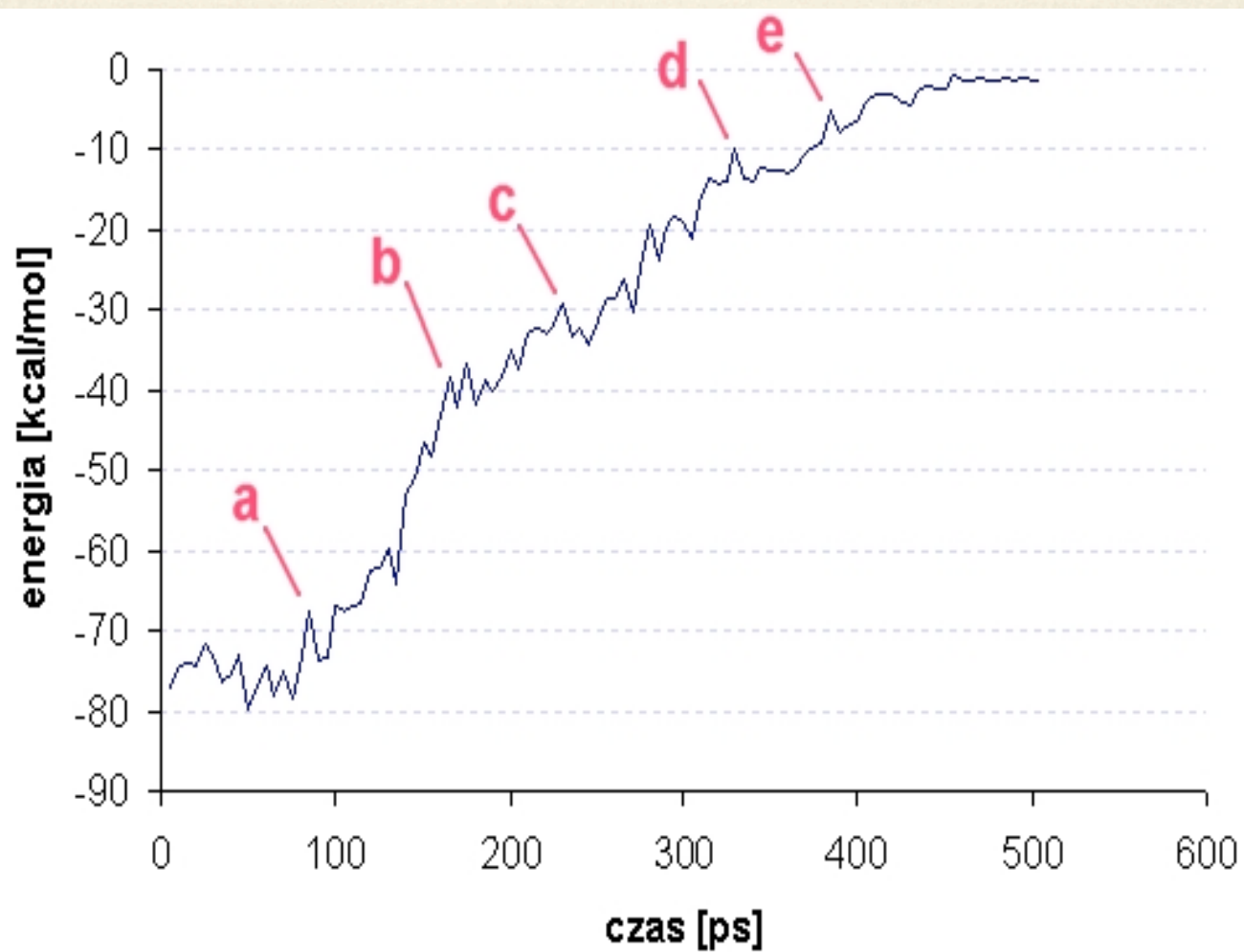
SMD



SMD



SMD



IMD

IMD – Interaktywna Dynamika Molekularna
Symulacja, w której „na żywo” przykładamy siły,
obserwując od razu wyniki na monitorze.



Podział symulacji

- ❖ symulacje deterministyczne (dynamika molekularna)
- ❖ symulacje stochastyczne (dynamika Monte Carlo)
- ❖ symulacje hybrydowe (dynamika Brownowska)

Donald E. Knuth

Metody stochastyczne

Metody stochastyczne są zbudowane na teorii prawdopodobieństwa i mechaniki statystycznej.

Proces stochastyczny to funkcja losowa, której wartości leżą w przestrzeni zdarzeń losowych, tzn. pewnej wielkości przypisane jest zdarzenie losowe - np. rzut monetą lub kostką.

Kluczowym elementem symulacji tego typu, jest stosowanie procesu/łańcucha Markowa.



Podstawowe pojęcia teorii prawdopodobieństwa

Zdarzenie elementarne jest to możliwy wynik doświadczenia losowego, do którego zwykle przypisane jest pewne prawdopodobieństwo jego wystąpienia.

Prawdopodobieństwo jest to funkcja $P(x)$, która przyporządkowuje każdemu elementowi zbioru zdarzeń losowych pewną nieujemną wartość rzeczywistą.

Podstawowe pojęcia teorii prawdopodobieństwa

Definicja Laplace'a z 1812 roku (tzw. klasyczna):

Prawdopodobieństwem zajścia zdarzenia A nazywamy iloraz liczby zdarzeń sprzyjających zdarzeniu A do liczby wszystkich możliwych przypadków, zakładając, że wszystkie przypadki wzajemnie się wykluczają i są jednakowo możliwe.

Podstawowe pojęcia teorii prawdopodobieństwa

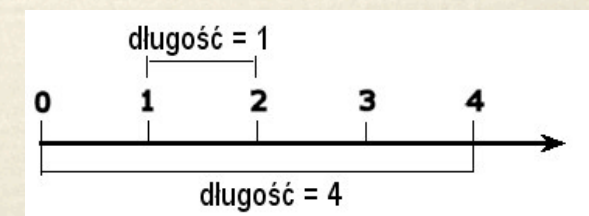
Prawdopodobieństwo zajścia zdarzenia A możemy zapisać w postaci:

$$P(A) = \frac{|A|}{|\Omega|}$$

gdzie $|A|$ oznacza liczbę elementów zbioru A , zaś $|\Omega|$ liczbę elementów zbioru Ω .

Podstawowe pojęcia teorii prawdopodobieństwa

Definicja klasyczna nie pozwala obliczać prawdopodobieństwa w przypadku, gdy zbiory A i Ω są nieskończone, jeśli jednak zbiory te mają interpretację geometryczną, zamiast liczebności zbiorów można użyć miary geometrycznej (długość, pole powierzchni, objętość) - tzw. prawdopodobieństwo geometryczne.



Proces Markowa

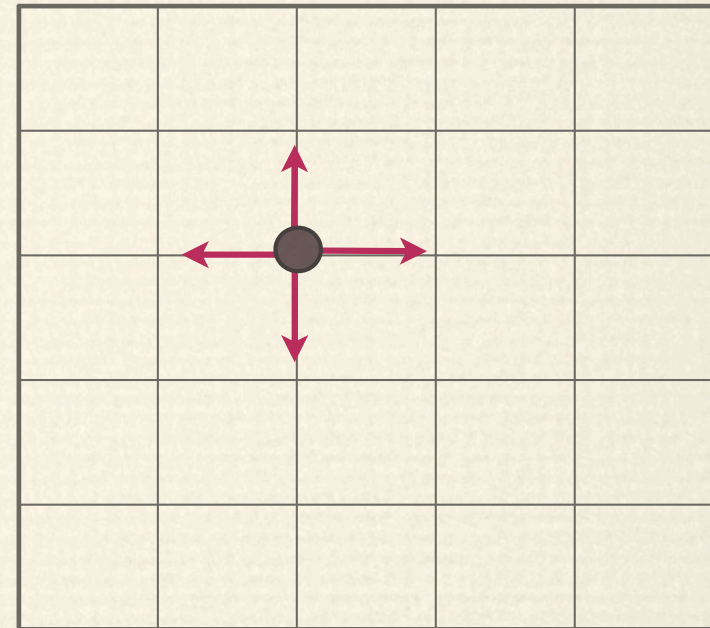
Proces Markowa jest procesem, w którym prawdopodobieństwo każdego zdarzenia przyszłego zależy tylko od wyniku poprzedzającego go.

Łańcuch Markowa to proces Markowa zdefiniowany na dyskretnej przestrzeni stanów.

Proces Markowa- przykład

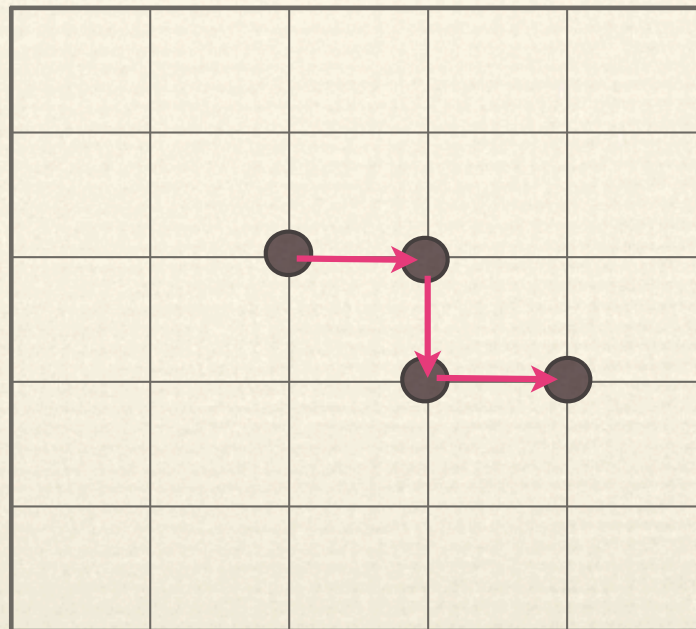
Przy każdym kroku czasowym cząstka przeskakuje do jednego z najbliższych, sąsiednich węzłów.

1. Cząstka nie pamięta skąd przyszła - czyli błądzenie przypadkowe
2. Cząstka pamięta skąd przyszła, i musi unikać przecinania drogi, którą już raz przyszła - błądzenie bez przecięć



Proces Markowa

Prawdopodobieństwo znalezienia układu w i -tym stanie zależy wyłącznie od stanu poprzedniego $i-1$.



Proces Markowa

ciągi stanów układu: $x_0, \dots, x_n, \dots,$

każdy stan x_i powstaje w wyniku próby losowej, czyli zmienna losowa X_i przyjmie wartość x_i i dokona tego z bezwzględnym prawdopodobieństwem a_i

prawdopodobieństwo warunkowe: $P(x_n | x_{n-1}, \dots, x_0)$

Proces Markowa

Ciąg x_0, \dots, x_n, \dots , jest nazwany łańcuchem Markowa, gdy dla każdego n zachodzi równość:

$$P(x_n | x_{n-1}, \dots, x_0) = P(x_n | x_{n-1})$$

Wynik każdej próby zależy od próby poprzedzającej i tylko od niej !!!!

Proces Markowa

Podstawową własnością łańcucha Markowa jest istnienie niezmiennego rozkładu stanów.

stan początkowy x_0 , gdzie $a_0 = 1$

Proces Markowa

Podstawową własnością łańcucha Markowa jest istnienie niezmiennego rozkładu stanów.

stan początkowy x_0 , gdzie $a_0 = 1$

Rozkład prawdopodobieństwa (u_k) musi być niezmienny lub stacjonarny dla danego łańcucha Markowa.

Łańcuch Markowa jest ergodyczny gdy zawiera tylko elementy ergodyczne.

Stan x_i jest ergodyczny, gdy jest aperiodyczny i trwały ze skończonym czasem rekurencji.

Metoda Monte Carlo



“Pomysł ten, nazwany później metodą Monte Carlo, wpadł mi do głowy, kiedy podczas choroby stawiałem pasjansa. Zauważyłem, że znacznie praktyczniejszym sposobem oceniania prawdopodobieństwa ułożenia pasjansa jest wykładanie kart, czyli eksperymentowanie z tym procesem i po prostu zapisywanie procentu wygranych, niż próba obliczenia wszystkich możliwości kombinatorycznych, których liczba rośnie wykładniczo”

Stanisław M. Ulam, Przygody matematyka.

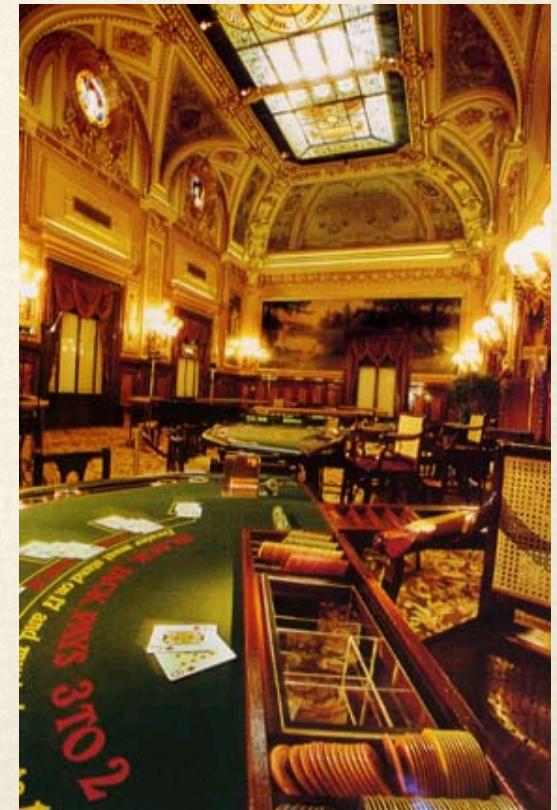


Metoda Monte Carlo



“Wydaje mi się że nazwa Monte Carlo bardzo przyczyniła się do jej popularyzacji. Metoda została tak nazwana z powodu roli przypadku: generowania liczb losowych, które decydują o przebiegu gry.”

Stanisław M. Ulam, Przygody matematyka.



Metoda Monte Carlo

Metoda Monte Carlo polega na przedstawieniu rozwiązania postawionego problemu w postaci parametru pewnej hipotetycznej populacji i używaniu losowej sekwencji liczb do tworzenia próbki tej populacji, na podstawie której można dokonać statystycznego oszacowania wartości badanego parametru.

Metoda Monte Carlo

Metodę Monte Carlo stosuje się w różnych działach matematyki numerycznej. Podstawą jej jest modelowanie statystycznego eksperymentu za pomocą środków techniki obliczeniowej i rejestracja charakterystyk liczbowych otrzymanych z tego eksperymentu. Metoda ta stosowana jest do modelowania matematycznego procesów zbyt złożonych (obliczanie całek, łańcuchy procesów statystycznych), aby można było przewidzieć ich wyniki za pomocą podejścia analitycznego. Istotną rolę w metodzie Monte-Carlo odgrywa losowanie (wybór przypadkowy) wielkości charakteryzujących proces, przy czym losowanie dotyczy rozkładów znanych skądinąd (np. z badania procesów prostszych lub niekiedy - z odpowiednio uzasadnionych lub oczywistych założeń). Rozwiązanie zadań rachunkowych za pomocą tej metody jest bliższe doświadczeniu fizycznemu niż klasycznym metodom rachunkowym.

Metoda Monte Carlo

Punktem wyjścia w metodzie Monte Carlo jest opisanie układu za pomocą hamiltonianu oraz wybór zespołu statystycznego odpowiedniego dla rozwiązywanego zagadnienia. Wszystkie obserwable są obliczane przy użyciu stowarzyszonej z układem funkcji rozkładu prawdopodobieństwa oraz funkcji rozdziału (sumy stanów).

Metoda Monte Carlo

Istotną rolę w metodzie Monte Carlo jest losowanie przypadkowych wielkości charakteryzujących proces, dotyczy to zarówno rozkładów procesów prostych lub złożonych. Składa się ona z następujących głównych części: sformułowanie modeli stochastycznych badanych procesów realnych, modelowania zmiennych losowych o danym rozkładzie prawdopodobieństwa, rozwiązywania problemu statystycznego z zakresu teorii estymacji.

Metoda Monte Carlo

Ogólny podział metod Monte Carlo:

Metoda von Neumanna

Metoda Metropolisa (łańcuchów Markowa)



Metoda Monte Carlo

Algorytm:

1. określ punkt początkowy x_0 w przestrzeni fazowej
2. wygeneruj nowy stan x'
3. oblicz prawdopodobieństwo przejścia $W(x, x')$
4. wygeneruj liczbę losową $R \in [0, 1]$ (rozkład równomierny)
5. jeżeli prawdopodobieństwo przejścia W jest mniejsze niż liczba losowa R , to potraktuj stary stan jako nowy i wróć do kroku 2
6. w przeciwnym razie, zaakceptuj nowy stan i wróć do kroku 2

Metoda Monte Carlo

Algorytm:

1. określ punkt początkowy x_0 w przestrzeni fazowej
2. wygeneruj nowy stan x'
3. oblicz prawdopodobieństwo przejścia $W(x, x')$
4. wygeneruj liczbę losową $R \in [0, 1]$ (rozkład równomierny)
5. jeżeli prawdopodobieństwo przejścia W jest mniejsze niż liczba losowa R , to potraktuj stary stan jako nowy i wróć do kroku 2
6. w przeciwnym razie, zaakceptuj nowy stan i wróć do kroku 2

Metoda Monte Carlo

Algorytm:

1. określ punkt początkowy x_0 w przestrzeni fazowej
2. wygeneruj nowy stan x'
3. oblicz prawdopodobieństwo przejścia $W(x, x')$
4. wygeneruj liczbę losową $R \in [0, 1]$ (rozkład równomierny)
5. jeżeli prawdopodobieństwo przejścia W jest mniejsze niż liczba losowa R , to potraktuj stary stan jako nowy i wróć do kroku 2
6. w przeciwnym razie, zaakceptuj nowy stan i wróć do kroku 2

Metoda Monte Carlo

Metoda Metropolisa polega na takim użyciu łańcucha Markowa, że rozpoczynając od stanu początkowego x_0 generuje się kolejne stany rozłożone dokładnie według $P(x)$. Kolejne stany generowane przez łańcuch Markowa mają tę właściwość, że stany potomne są położone w ścisłym sąsiedztwie poprzedzających. Czyli istnieje dobrze zdefiniowana korelacja między kolejnymi stanami.

gdzie $P(x)$ to prawdopodobieństwo wyboru stanów

Metoda Monte Carlo

Do oszacowania całki w przestrzeni wielowymiarowej, potrzebujemy wielkiej liczby stanów, z których większość będzie dawała pomijalny udział w wartości całki

Próbkowanie wazone polega na selekcji punktów fazowych z pewnym prawdopodobieństwem $P(x)$

Metoda Monte Carlo

Algorytm:

1. określ punkt początkowy x_0 w przestrzeni fazowej
2. wygeneruj nowy stan x'
3. oblicz prawdopodobieństwo przejścia $W(x, x')$
4. wygeneruj liczbę losową $R \in [0, 1]$ (rozkład równomierny)
5. jeżeli prawdopodobieństwo przejścia W jest mniejsze niż liczba losowa R , to potraktuj stary stan jako nowy i wróć do kroku 2
6. w przeciwnym razie, zaakceptuj nowy stan i wróć do kroku 2

Metoda Monte Carlo

ustalenie pokrewieństwa

Ograniczenia:

1. Dla wszystkich komplementarnych par (S, S') zbiorów punktów fazowych istnieją stany $x \in S$ oraz $x' \in S'$ takie, dla których $W(x, x') \neq 0$

narzuca warunek aby prawdopodobieństwa były dodatnie

2. Dla wszystkich par x, x' : $W(x, x') \geq 0$

3. Dla każdego x : $\sum_{x'} W(x, x') = 1$

orzeka, że całkowite prawdopodobieństwo tego, że w układzie nastąpi przejście do jakiegokolwiek stanu x' jest równe jedności

4. Dla każdego x : $\sum_{x'} W(x, x') P(x') = P(x)$

jest warunkiem, aby rozkład graniczny był rozkładem równowagowym, zgodnie z którym chcemy mieć rozłożone nasze stany

gdzie $P(x)$ to prawdopodobieństwo wyboru stanów

Metoda Monte Carlo

Ponieważ prawdopodobieństwo przejścia nie jest jednoznacznie określone, jest wiele możliwych wyborów przejść.

Metoda Monte Carlo

Jakiego rodzaju wartości średnie uzyskujemy z tych symulacji?

Otrzymujemy średnią po czasie, podobną do uzyskanej w MD, tylko tam trajektoria biegnie w przestrzeni położeniowo-pędowej, a w symulacji MC przebiega w przestrzeni konfiguracyjnej.

Metoda Monte Carlo

- ❖ dla zespołu mikrokanonicznego
- ❖ dla zespołu kanonicznego
- ❖ dla zespołu izotermiczno-izobarycznego

Metoda Monte Carlo

Łańcuch Markowa zaczyna się od stanu x_0 .

Jak wiele stanów początkowych musi zostać odrzuconych?

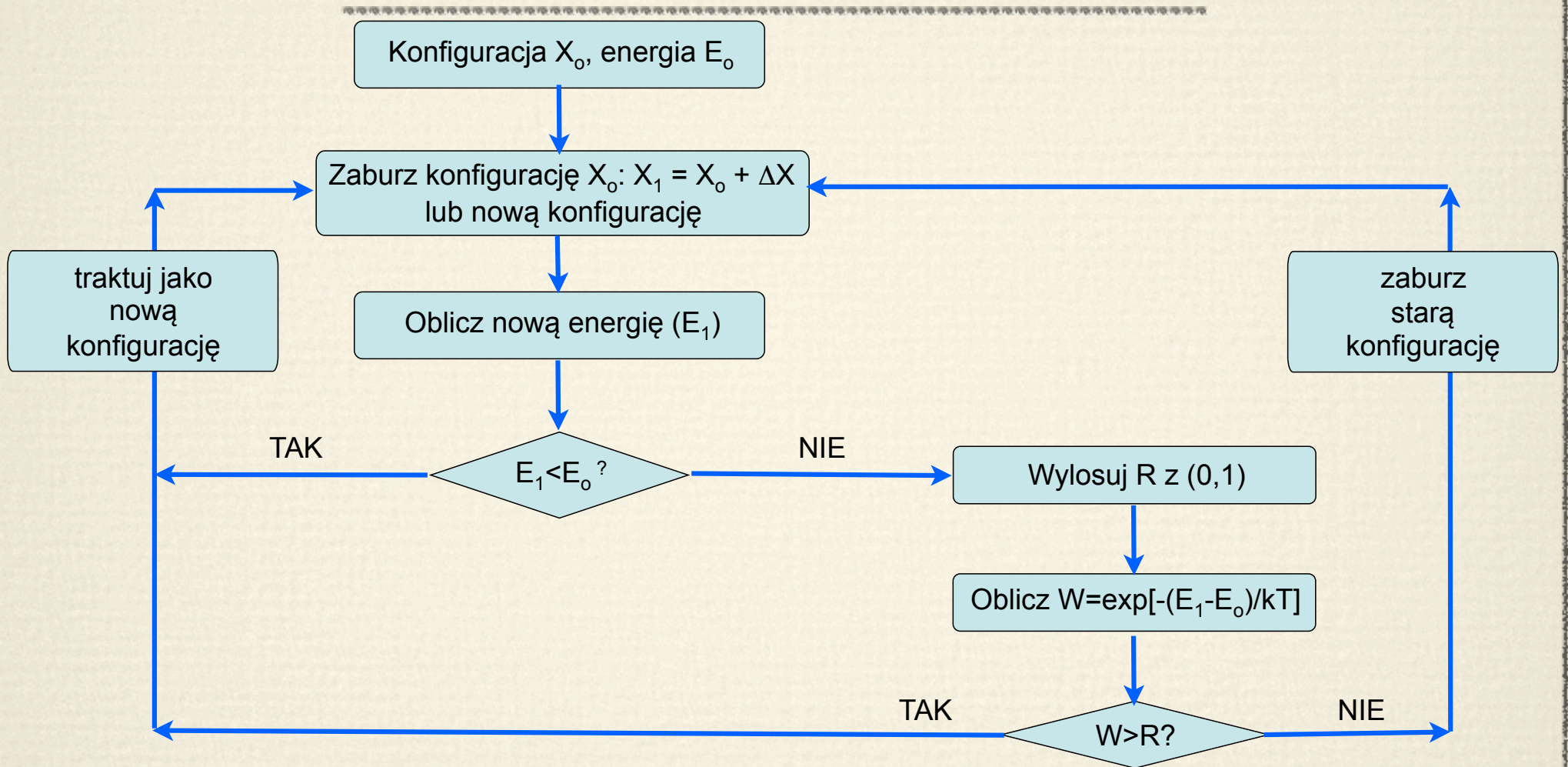
W ogólnym przypadku stan początkowy nie odpowiada stanowi równowagi. Układ musi mieć możliwość relaksacji, zanim będzie można wybrać jakiegokolwiek średnie. Relaksacja znacząco zależy od praw zachowania czyli od rodzaju zespołu, w którym wielkości są obliczane.

Należy określić stosowne skale czasowe modelu, aby można oszacować liczbę stanów początkowych, które zostaną odrzucone.

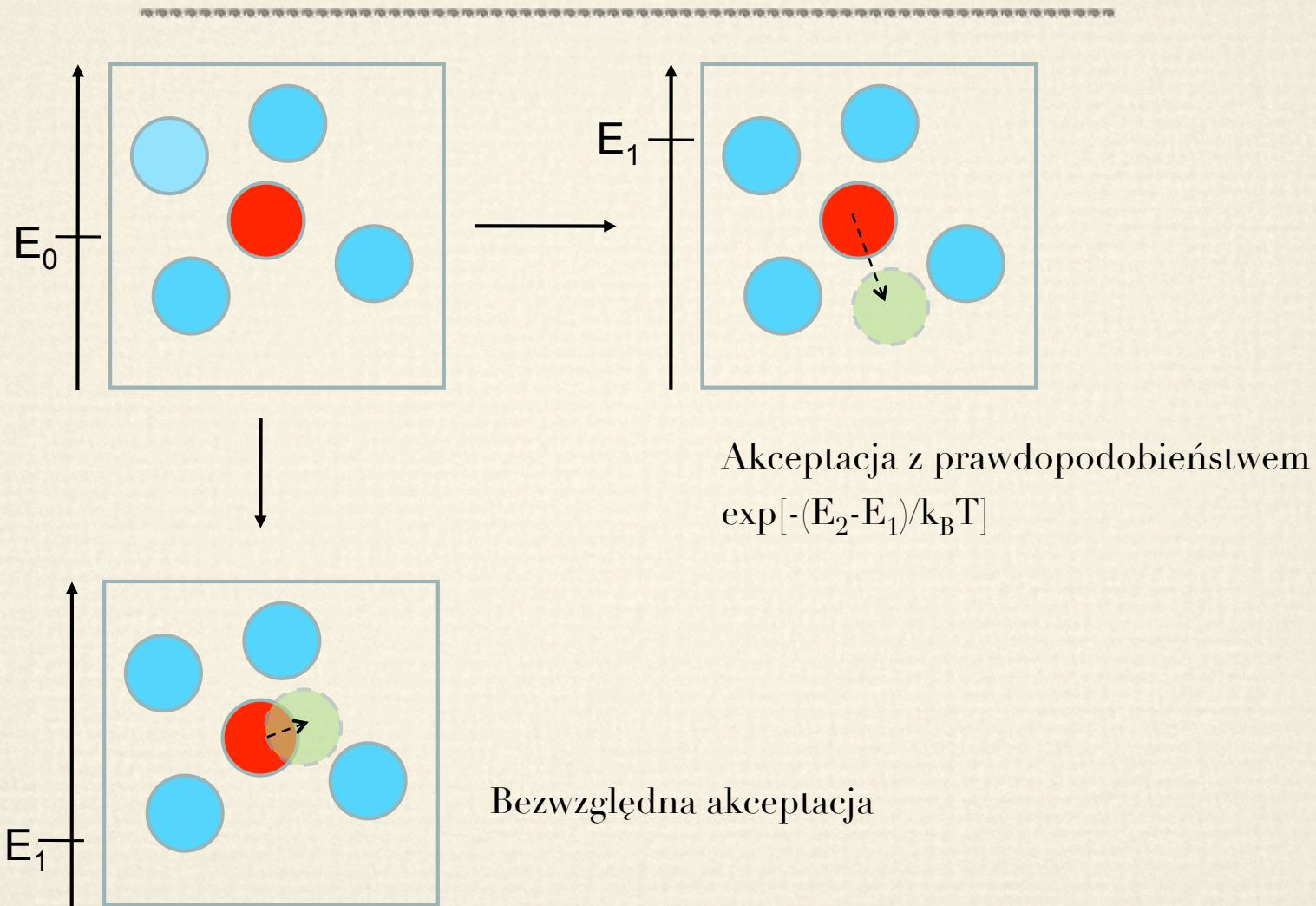
Metoda Monte Carlo

1. Bierzemy startową konfigurację układu daną współrzędnymi $(x_1^0, y_1^0, z_1^0, \dots, x_n^0, y_n^0, z_n^0)$; tej konfiguracji odpowiada energia E_0 .
2. Zaburzamy losowo wybraną współrzędną, np. x_i^0 lub Δx_i (mała wartość).
3. Obliczamy energię nowej konfiguracji i oznaczamy ją jako E_1 .
4. Jeżeli $E_1 < E_0$ to nową konfigurację akceptujemy traktując ją jako nową konfigurację startową i przechodzimy do punktu 2; w przeciwnym przypadku przechodzimy do punktu 5.
5. Wykonujemy test Metropolis'a:
 - a) Generujemy liczbę losową R z przedziału $(0,1)$.
 - b) Jeżeli $\exp[-(E_1 - E_0)/kT] > R$, (k jest stałą Boltzmann'a) akceptujemy nową konfigurację, w przeciwnym przypadku przechodzimy do punktu 2 ze starą konfiguracją.

Metoda Monte Carlo



Metoda Monte Carlo



Metoda Monte Carlo

Brak zmiennej czasowej. Zamiast całkować równania ruchu małym krokiem czasowym, powierzchnia energii potencjalnej próbkowana jest losowo.

Dyskretyzacja ruchów jest wymagana, ponieważ bez stosowania równań ruchu nieznane są siły poprzedniego kroku. Definiuje się więc zestawy ruchów lokalnych (perturbacji), naśladujące możliwe (fizyczne) ruchy białka, a wygenerowana konformacja akceptowana jest poprzez prawdopodobieństwo przejścia ze stanu poprzedniego do obecnego.

Metoda Monte Carlo

- nadaje się do badania własności fizycznych oraz struktury w stanie równowagowym
- pozwala dość dokładnie obliczać takie wielkości jak: energię potencjalną, ciśnienie, ciepło właściwe, radialną funkcję rozkładu
- jako metoda statystyczna raczej nie powinno się jej używać do badania wielkości zależnych od czasu, takich jak wszelkiego rodzaju czasowe funkcje korelacji.
- nie można za jej pomocą śledzić toru cząstki wynikającego z równań ruchu.

Metoda Monte Carlo

- wyznaczanie wielkości mechanicznych i termodynamicznych (gęstość, średnia energia, pojemność cieplna, przewodnictwo, współczynniki wirialne)
- symulacje przemian fazowych
- symulacje właściwości polimerów
- symulacje zwijania białek i innych biopolimerów
- symulacje wiązania ligandów z receptorami oraz szacowanie energii swobodnej tego procesu (projektowanie leków)
- symulacje reakcji chemicznych

Dynamika Brownowska

Metoda DB polega na obliczaniu trajektorii w przestrzeni fazowej zbioru molekuł, z których każda spełnia równania Langevina w polu sił.

$$m_i \frac{d^2 x_i}{dt^2} = -\frac{\partial V}{\partial x_i} - \gamma_i \frac{dx_i}{dt} + f_i^{rand}$$

Dynamika Langevina

$$\gamma_i = 6\pi (r_i + r_w) \eta_w$$

Prawo Stokesa

$$f_i^{rand} = \sqrt{\frac{2\gamma_i RT}{\Delta t}} N(0,1)$$

Proces Wienera

$$\gamma_i \frac{dx_i}{dt} = -\frac{\partial E}{\partial x_i} + f_i^{rand}$$

Dynamika Brownowska

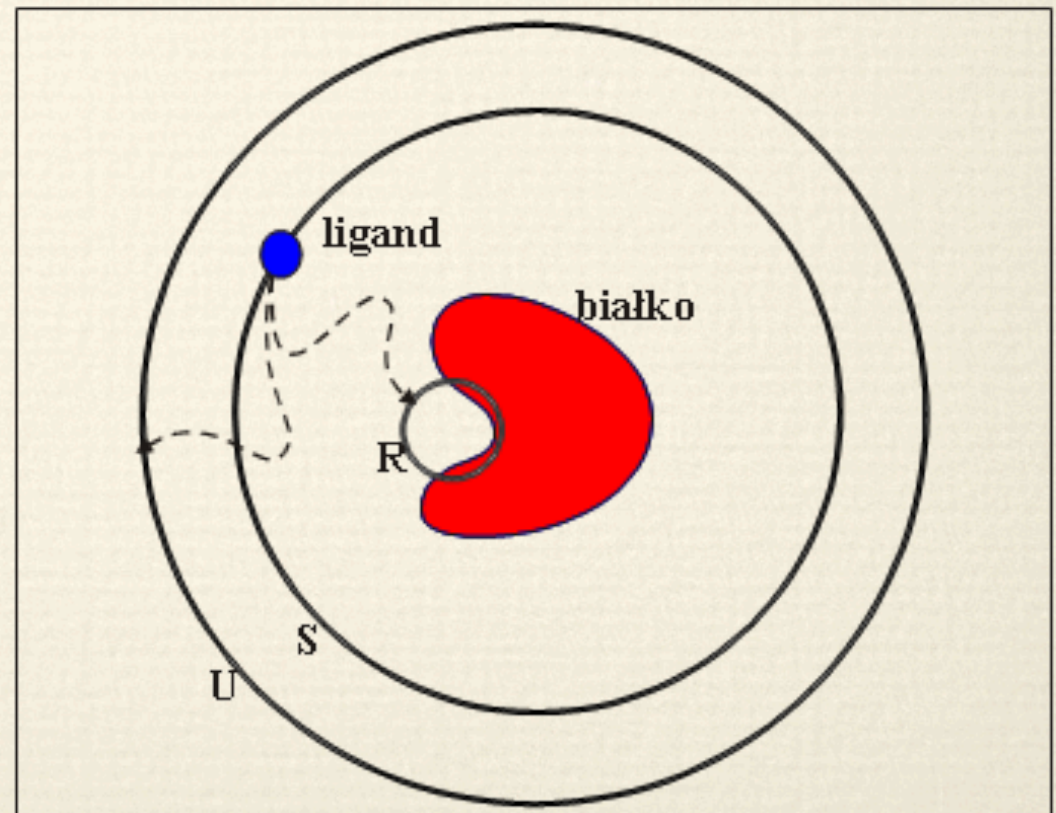
Dynamika Brownowska

Trajektoria ruchu cząsteczki jest wyznaczana w oparciu o algorytm Ermaka i McCammona:

$$\mathbf{r}(t + \delta t) = \mathbf{r}(t) + \frac{D\mathbf{F}\delta t}{kT} + \mathbf{R}$$

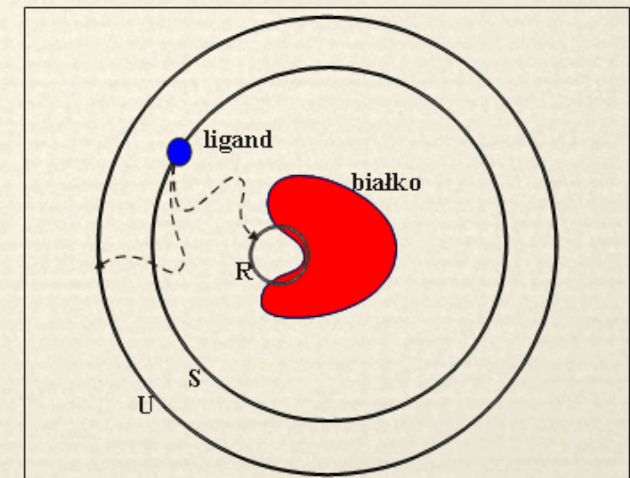
gdzie:

- D oznacza współczynnik dyfuzji cząsteczki,
- \mathbf{F} oznacza położenie i siłę elektrostatyczną działającą na ligand,
- T - temperaturę,
- k - stałą Boltzmann,
- δt - krok czasowy dynamiki,
- \mathbf{R} - przesunięcie stochastyczne, które jest dodawane, aby symulować zderzenia z cząsteczkami wody.



Dynamika Brownowska

Najpierw obliczany jest potencjał elektrostatyczny wewnątrz i na zewnątrz białka poprzez rozwiązanie równania Poissona-Boltzmanna. Siatka potencjału jest używana jako plik wejściowy do dalszych obliczeń. Następnie symulowany jest ruch ligandu w polu elektrostatycznym białka w oparciu o powyższe równanie; w każdym kroku czasowym obliczana jest siła elektrostatyczna działająca na ligand, dodawane jest do niej losowe przesunięcie i w ten sposób znajdowane jest nowe położenie ligandu. Aby uniknąć zderzeń białka z ligandem wszystkie kroki, które prowadzą do zachodzenia na siebie tych cząsteczek są odrzucane i dla takich przypadków jest losowane nowe R. S oznacza sferę, na której losowo wybierane jest położenie startowe ligandu. Ligand porusza się po trajektorii wyznaczonej przez siły oddziaływania z polem elektrostatycznym oraz przez siły stochastyczne. R i U oznaczają odpowiednio sfery "reakcji" i "ucieczki", na których kończą się trajektorie klasyfikowane jako prowadzące albo nie prowadzące do związania ligandu z białkiem. W modelu tym ligand jest przeważnie reprezentowany jako sfera o całkowitym ładunku lub jako sztywny zbiór atomów z ładunkami cząstkowymi. Symulacje, w których ligand startuje ze sfery S są przeprowadzane wielokrotnie tak, aby mieć wiarygodną informację statystyczną na temat częstości utworzenia kompleksu ligand-białko. Do obliczeń musi być także odpowiednio zdefiniowane tzw. kryterium reakcji, czyli obszar, w którym zakładamy, że ligand jest już związany z białkiem



Dynamika Brownowska

Ponieważ w dynamice brownowskiej nie interesują nas ruchy wewnętrzne molekuł, a jedynie ich wzajemne oddziaływanie stosuje się krok czasowy o wiele dłuższy niż w dynamice molekularnej ($> 1ps$).

Metoda dynamiki brownowskiej pozwala obliczyć stałe asocjacji molekuł, określić mechanizm asocjacji, a także zbadać czy siły elektrostatyczne mają istotny wpływ w procesie asocjacji molekuł.