

Modelowanie białek
ab initio / de novo



Słowniczek

- ❖ *de novo* - od początku, na nowo
- ❖ *ab initio* - od początku

Słowniczek

- ❖ *de novo* - kategoria metod przewidywania struktury, w których nie używa się wzorców homologicznych struktur z PDB
- ❖ *ab initio* - przewidywanie struktury wyłącznie na podstawie praw fizyki

Modelowanie ab initio

Ab initio opiera się na hipotezie
„termodynamicznego” zwijania białek
czyli
natywna struktura sekwencji białkowej
odpowiada stanowi globalnego minimum
energetycznego

Dwa podejścia

PODEJŚCIE EWOLUCYJNE — SZKOŁA DARWINOWSKA

Rekonstrukcja procesu powstania sekwencji i struktury białka na drodze ewolucji (przyrodzie zabiera to miliony lat)

PODEJŚCIE FICZYCZNE — SZKOŁA BOLTZMANNOWSKA

Modelowanie zwijania białka (procesu poszukiwania przez łańcuch konformacji o najniższej energii swobodnej, który w komórkach trwa zaledwie ułamki sekundy) korzystając z praw fizyki statystycznej

Hipoteza Anfinsena

„Sekwencja aminokwasowa białka ściśle determinuje jego strukturę przestrzenną, która w danych warunkach fizjologicznych odpowiada globalnemu minimum energii swobodnej.”

Hipoteza Anfinsena

Zatem znajomość sekwencji aminokwasowej białka powinna wystarczyć do obliczenia konformacji o najniższej energii.

Struktura natywna białka

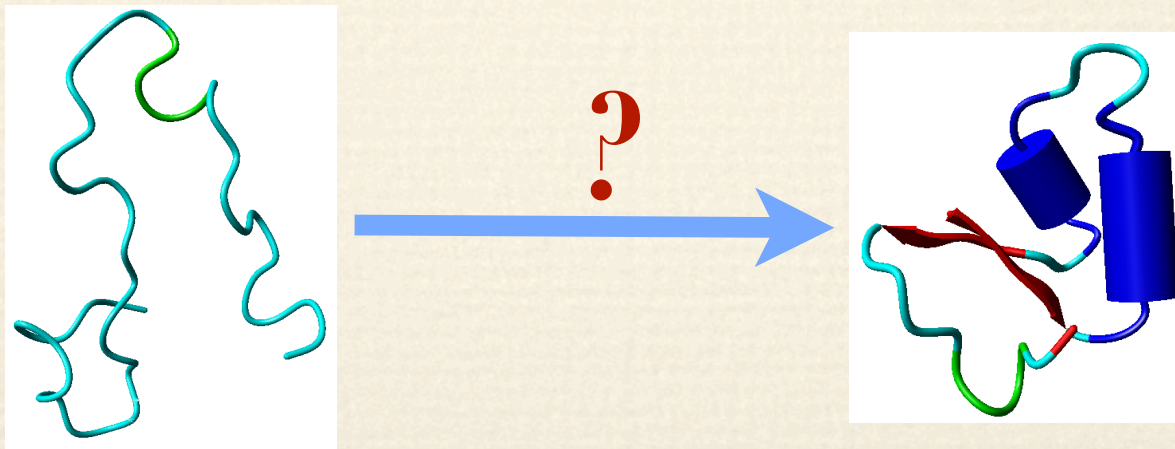
Struktura natywna zależy od wielu czynników:

- temperatura
- siła jonowa roztworu
- pH roztworu
- obecność jonów metali, cząsteczek organicznych, grup prostetycznych (np. hemu)

Modelowanie ab initio

problemy:

-paradoks Liventhala: liczba konformacji swobodnego łańcucha dąży do nieskończoności.



Modelowanie ab initio

Paradoks Liventhala

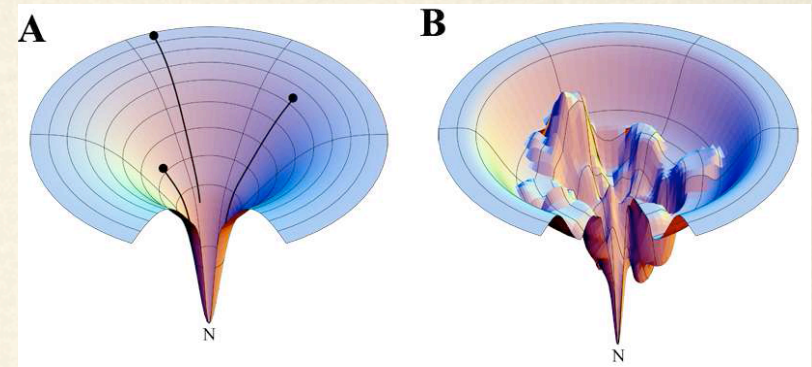
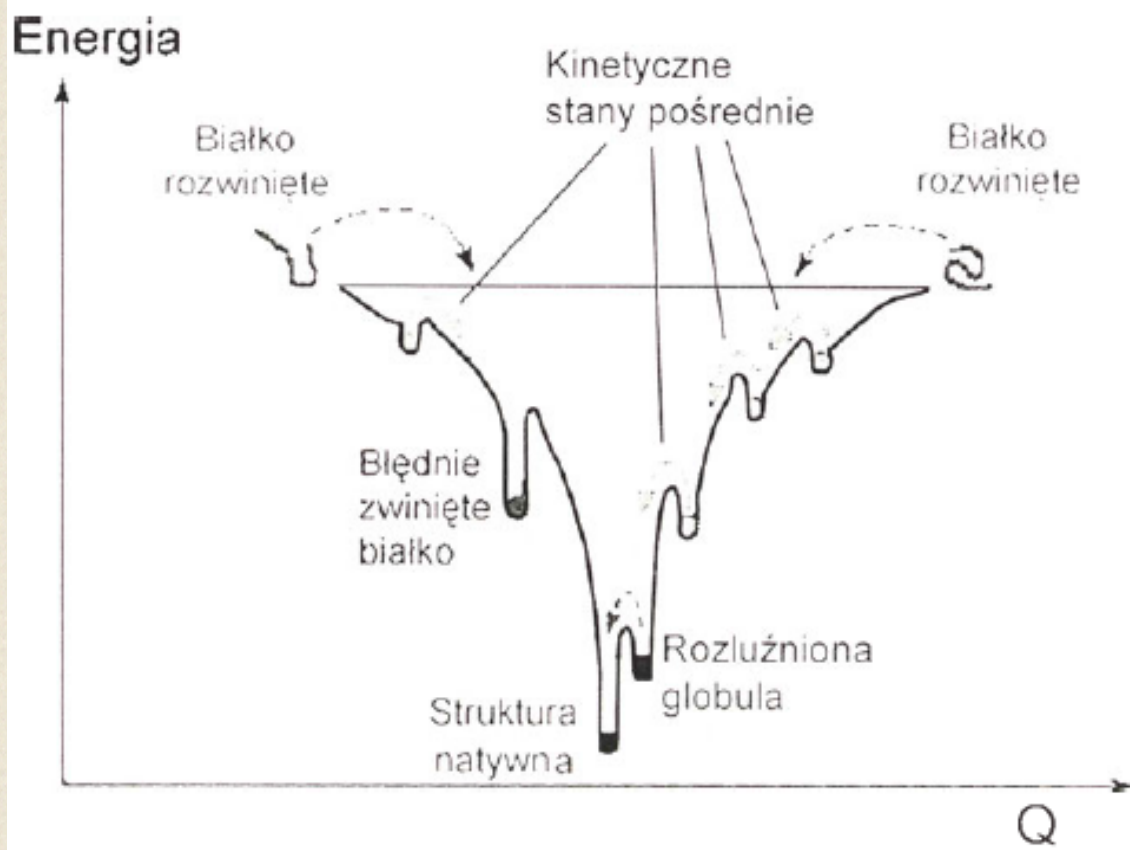
-jeśli liczba możliwych stopni swobody jednego aminokwasu wynosiłaby 2, to dla łańcucha N=150 aminokwasów liczba możliwych konformacji wynosi 2^{150}

-jeśli czas sprawdzania 1 konformacji wynosi 10^{-13} s, to dla takiego białka sprawdzenie wszystkich stanów wynosi:

$$2^{150} \times 10^{-13} \text{s} = 4.6 \times 10^{24} \text{ lat}$$

(czas „życia” Ziemi = 4.5×10^9 lat)

Hiperpowierzchnia energetyczna



A - wyidealizowany kształt lejka o gładkiej powierzchni, która mogłaby poprowadzić szybko do minimum

B - przykładowy kształt lejka z całym krajobrazem wzgórz, pułapek i barier energetycznych.

Zależność energii swobodnej (E) od współrzędnej Q opisującej proces zwijania białka za pomocą tzw. hipotezy lejka.

Hiperpowierzchnia energetyczna

Badania problemy:

- jeśli białko jest niekompletne może nie osiągnąć stabilnej struktury podobnej do struktury natywnej (rybonukleaza A pozbawiona 4 reszt z C-końca)
- w białkach wielodomenowych domeny mogą zwijać się niezależnie od siebie z różnymi prędkościami, a domeny są połączone giętkimi łącznikami

Podejście fizyczne

- ❖ Nawet krótki łańcuch białkowy teoretycznie może zwinąć się na tyle różnych sposobów, że obliczenie ich energii najszybszym współczesnym komputerem zajęłoby więcej czasu, niż upłynęło od początku wszechświata.
- ❖ Metody oceny energii oddziaływań wewnątrz cząsteczki białka oraz między białkiem a otoczeniem są niedoskonałe - dlatego możliwa jest sytuacja że komputer generuje właściwą (natywną) konformację, jednak algorytm oblicza jej energię tak nieprecyzyjnie, że okazałaby się wyższa od energii oszacowanej dla innych (nienatywnych) konformacji i dobry model zostaje odrzucony.
- ❖ Dlatego nie udało się jeszcze opracować metody „ab initio”, tzn. opierającej się wyłącznie na prawach fizyki, która poprawnie i dokładnie przewidywałaby natywną konformację całego białka. Udaje się w najlepszym razie przewidzieć strukturę peptydów długości kilkudziesięciu aminokwasów, a i tak rzadko można mieć pewność, że jest ona zgodna z rzeczywistością.
- ❖ Symulacja nawet niewielkich zmian konformacyjnych w cząsteczce białka, które w naturze zachodzą w ułamku mikrosekundy, wymaga długotrwałych obliczeń podczas gdy przejście od rozwiniętej konformacji przypadkowej do globularnej formy, ściśle określonej przez sekwencję aminokwasów, zajmuje rzeczywistemu białku od kilku milisekund do kilku minut.

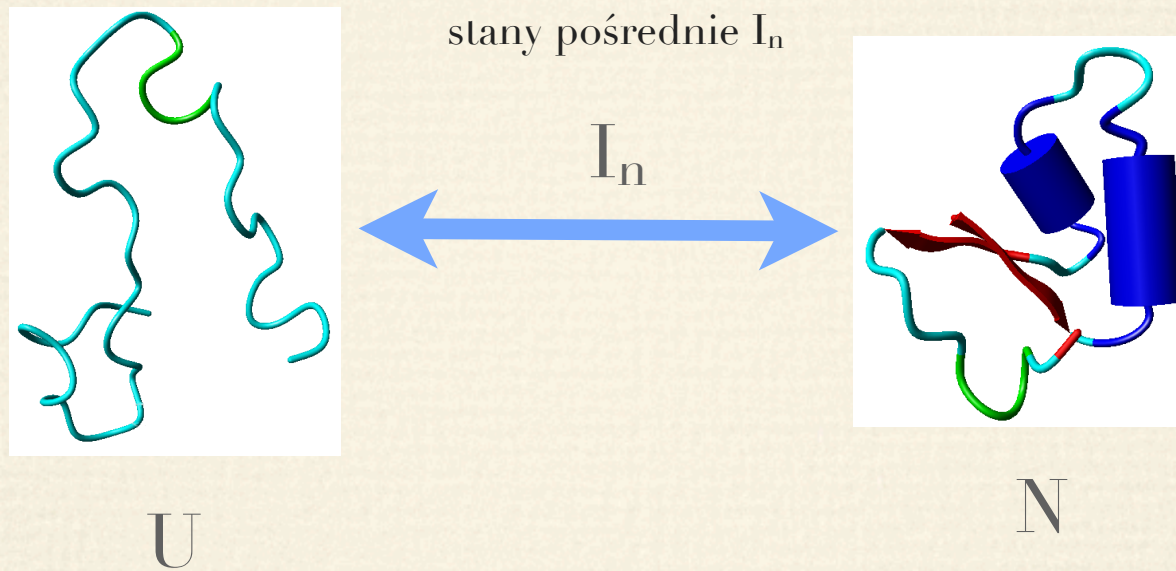
Podójście fizyczne

Rozwiązanie problemu:

Aby przyśpieszyć obliczenia i umożliwić symulację całego procesu zwijania, już od lat siedemdziesiątych podejmowano próby zastosowania uproszczonych modeli białek, w których unifikuje się grupy atomów (np. łańcuchy boczne), traktując je jako pojedyncze pseudoatomy.

1. zmniejsza to liczbę oddziaływań, które trzeba obliczyć
2. wygładza tzw. krajobraz energetyczny, który opisuje minima i maksima w przestrzeni konformacyjnej

Zwijanie białka



stanu U (unfolding) niezwinięty

stan N (native) zwinięty-natywny

Zwijanie białka

model dyfuzyjno-kolizyjny
(ang. *diffusion-collision
model*)

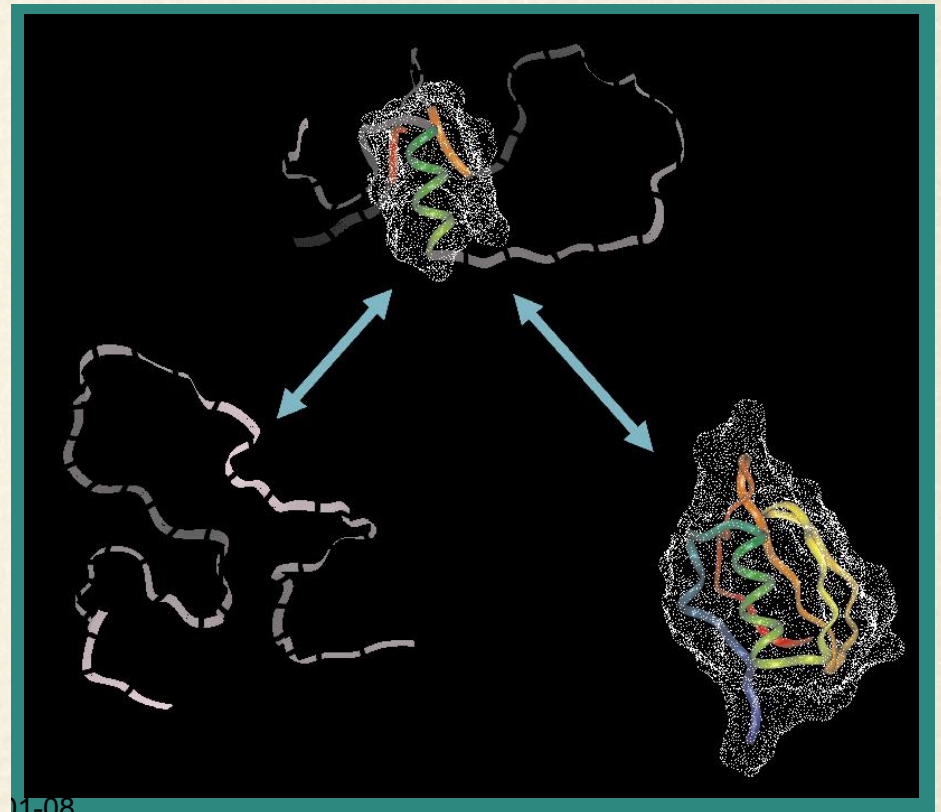
model nukleacyjny (ang.
nucleation model)

model hydrofobowego
kolapsu (ang. *hydrophobic
collapse model*)

Zwijanie białka

model dyfuzyjno-kolizyjny
(ang. *diffusion-collision
model*)

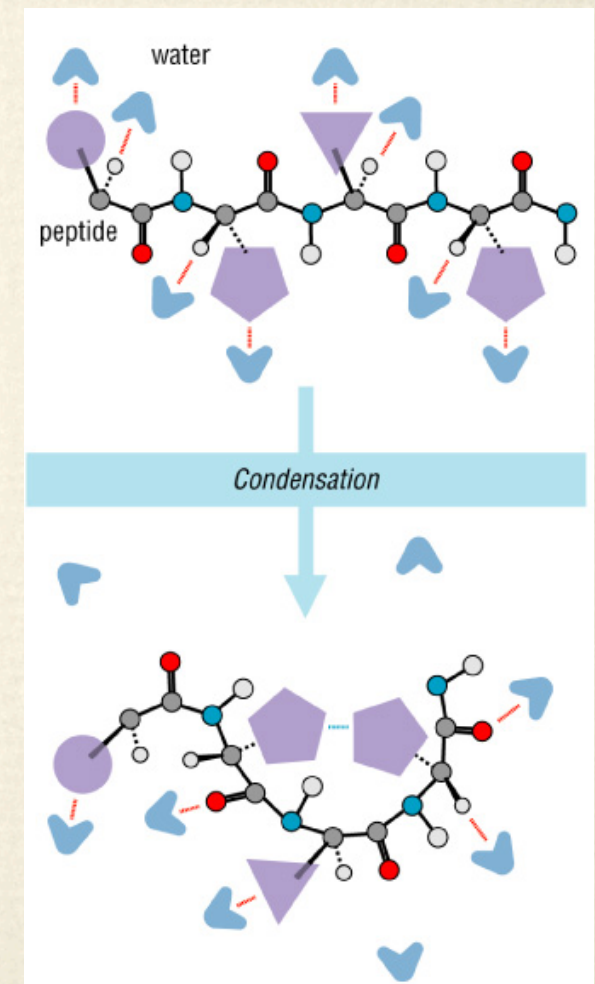
model nukleacyjny (ang.
nucleation model)



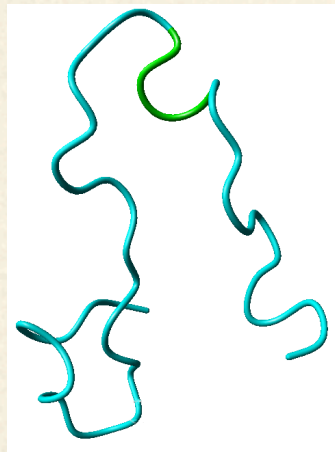
Zwijanie białka

model hydrofobowego kolapsu
(ang. *hydrophobic collapse model*)

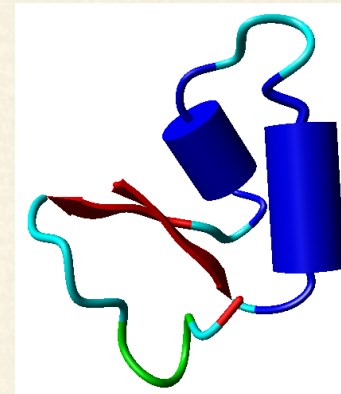
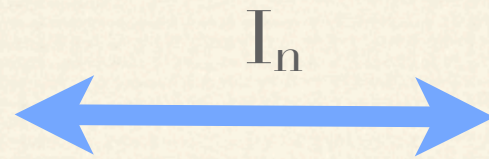
- minimalizacja powierzchni hydrofobowej dostępnej dla rozpuszczalnika
- zbliżenie polaryzowalnych grup hydrofobowych umożliwia powstanie między nimi oddziaływań van der Waalsa
- tym samym „wciągnięcie” polarnych grup C=O i N-H łańcucha głównego staje się siłą sprawczą powstania struktur drugorzędowych



Zwijanie białka

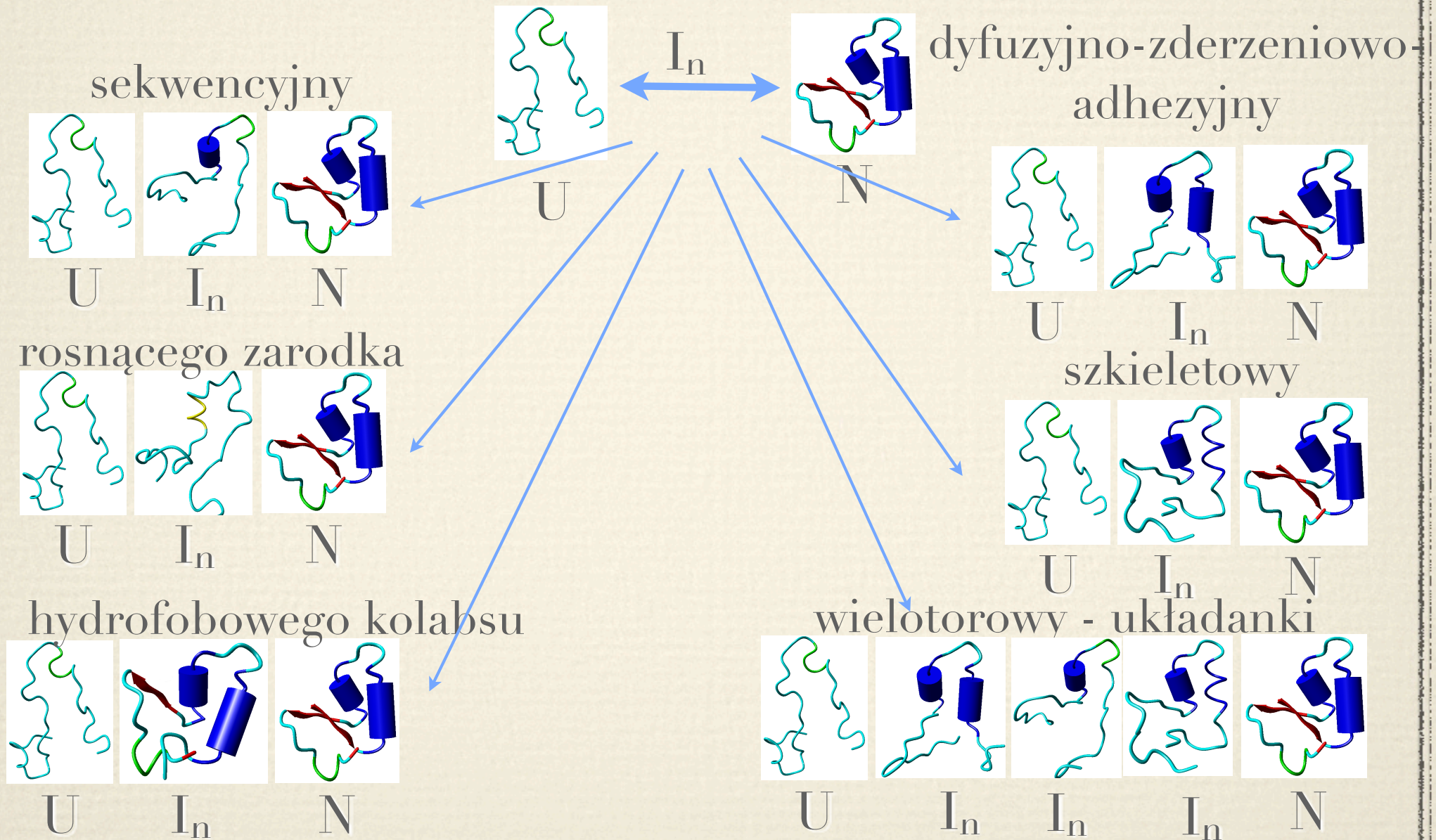


U



N

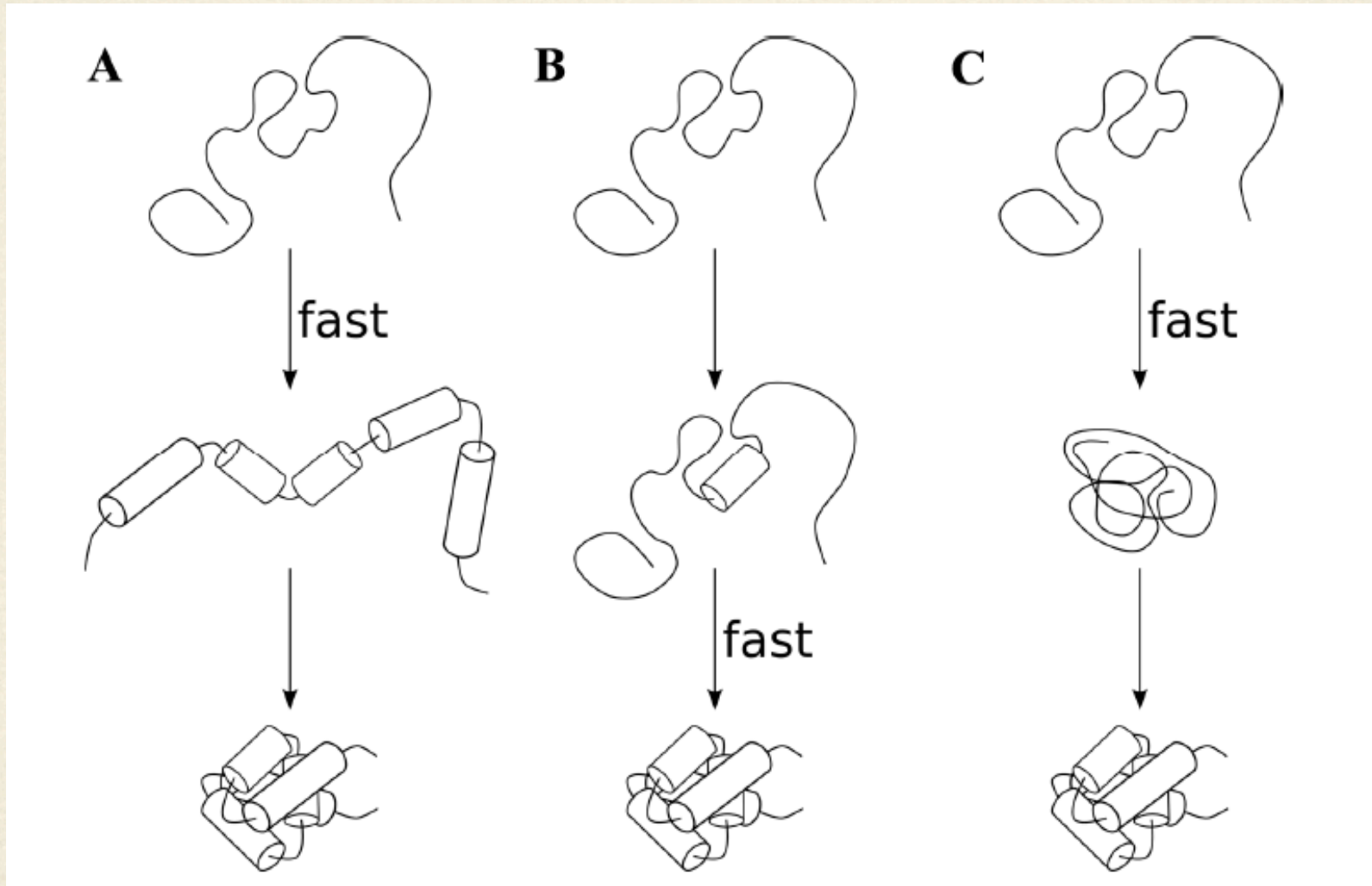
Zwijanie białka



Zwijanie białka

- sekwencyjne zwijanie do struktury natywnej
- szybkie zwinięcie do rozluźnionej globuli
- zwinięcie przy pomocy białek opiekuńczych (czaperonów)

Zwijanie białka

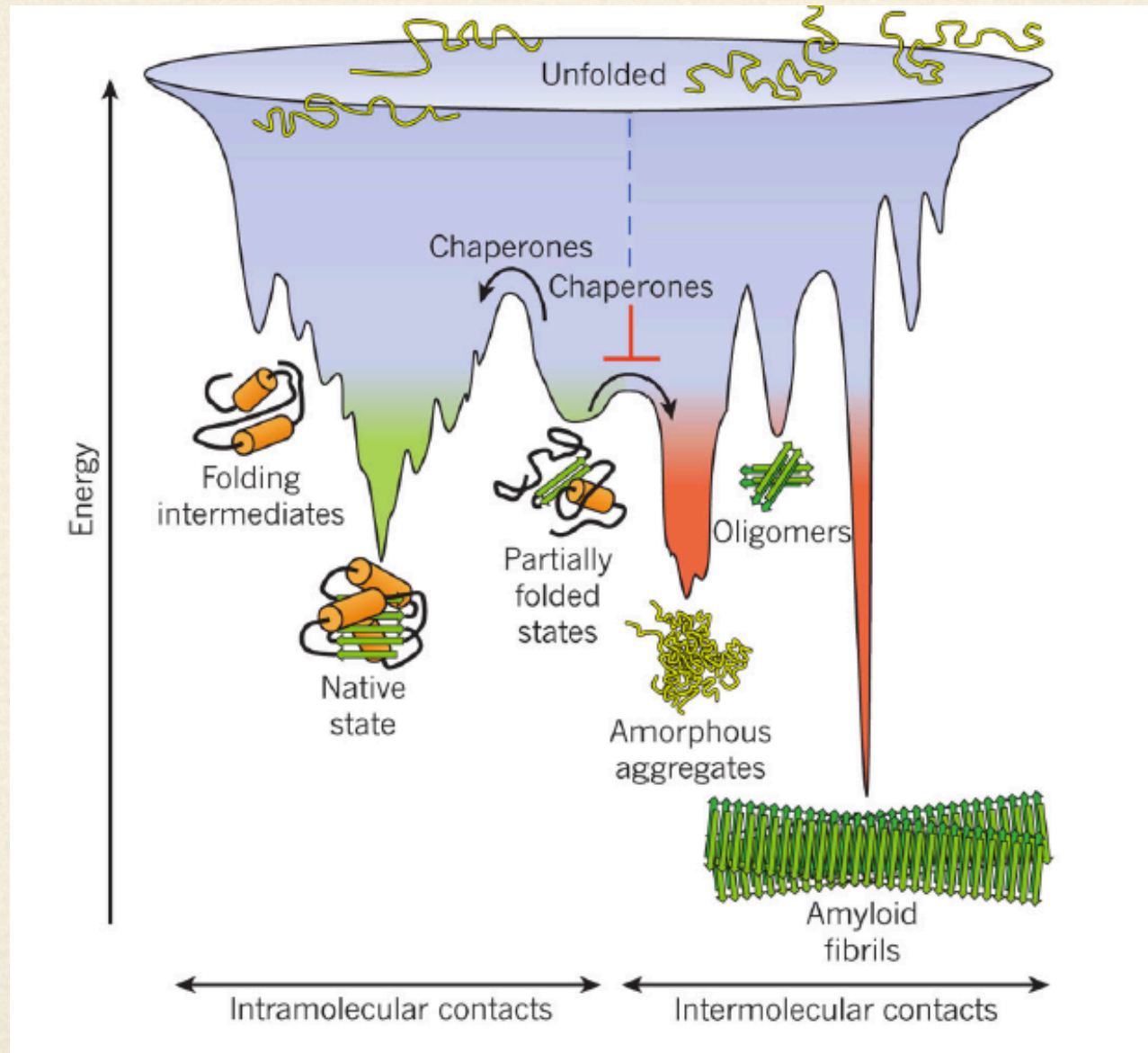


model dyfuzyjno-kolizyjny

model nukleacyjny

model hydrofobowego kolapsu

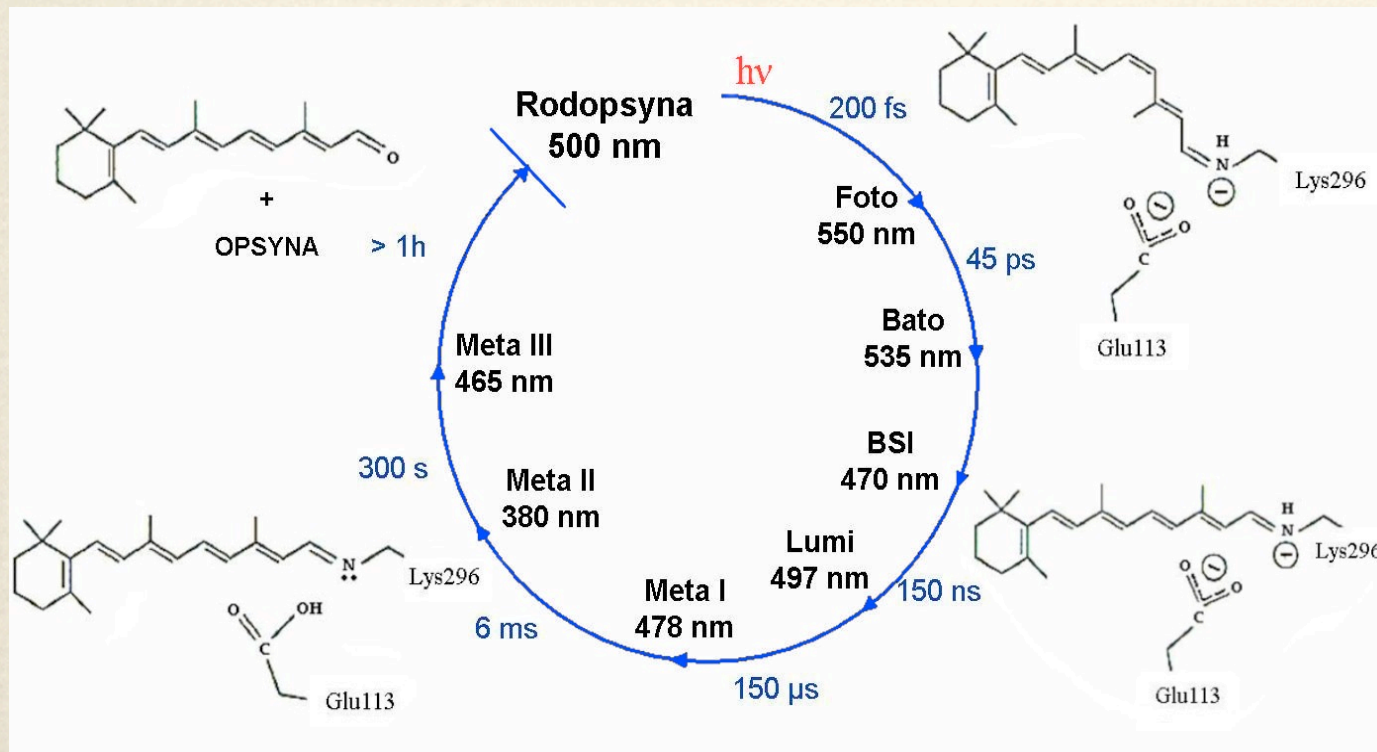
Hiperpowierzchnia energetyczna



Zwijanie białka

Są dwa zasadnicze powody, dla których pełna znajomość praw rządzących fałdowaniem jest tak istotna.

1. Po pierwsze zdobywanie i gromadzenie danych o sekwencji jest stosunkowo szybkie, a informacje na ten temat są coraz szerzej rozpowszechniane. Z kolei zdobywanie informacji na temat trójwymiarowej struktury białek jest nadal czasochłonne i ograniczone tylko do molekuł tworzących regularne kryształy (do badań metodami dyfrakcji promieni rentgenowskich) lub wystarczająco małe (do ok. 30 kDa), aby możliwe było określenie ich struktury metodami NMR.



Przemiany rodopsyny po zadziałaniu bodźca świetlnego. Pod nazwami izoform są umieszczone maksima absorpcji odpowiednie dla każdej z nich, a na niebiesko został przedstawiony czas jaki potrzebny jest do przejścia z jednej izoformy w drugą.

Zwijanie białka

Są dwa zasadnicze powody, dla których pełna znajomość praw rządzących fałdowaniem jest tak istotna

1. Po pierwsze zdobywanie i gromadzenie danych o sekwencji jest stosunkowo szybkie, a informacje na ten temat są coraz szerzej rozpowszechniane. Z kolei zdobywanie informacji na temat trójwymiarowej struktury białek jest nadal czasochłonne i ograniczone tylko do molekuł tworzących regularne kryształy (do badań metodami dyfrakcji promieni rentgenowskich) lub wystarczająco małe (do ok. 30 kDa), aby możliwe było określenie ich struktury metodami NMR.
2. Po drugie jesteśmy obecnie w stanie syntetyzować białka o dowolnej sekwencji metodami inżynierii genetycznej, więc synteza np. enzymów o zadanych własnościach katalitycznym jest wyzwaniem, które czeka na zrealizowanie. Aby jednak zadanie to mogło być wykonane musi być spełnionych kilka warunków:
 1. możliwość przewidzenia najbardziej stabilnego sfałdowania dla danej sekwencji;
 2. możliwość zaprojektowania pożądanej struktury;
 3. możliwość przewidzenia, czy osiągnięcie pożądanej struktury jest kinetycznie osiągalne;
 4. możliwość zaprojektowania precyzyjnych cech wiązań wewnątrz struktury;
 5. możliwość zaprojektowania precyzyjnych orientacji grup w białku dla otrzymania jak największej efektywności katalitycznej.

Zwijanie białka

Dlaczego?

- Muszą w stanie rozwiniętym przekraczać błony komórkowe
- Muszą być elastyczne, aby sprawnie pełnić swoje funkcje (np. katalityczne, czy podczas wiązania ligandów)
- Mogą potrzebować przyjmować różne kształty w zależności od warunków w ich otoczeniu
- Mogą mieć "zaprogramowany" czas "życia" poprzez podatność na proteolizę

Zwijanie białka

Stan zdenaturowany posiada znaczną swobodę konformacyjną. Nie jest zorganizowaną i ustaloną strukturą, a poszczególne segmenty łańcucha mogą się względem siebie poruszać, a poszczególne grupy chemiczne mogą rotować wokół pojedynczych wiązań. Mówi się o wysokiej entropii konformacyjnej, którą określić można jako:

$$S = k \ln W$$

gdzie: k - stała Boltzmann, W - liczba dostępnych stanów konformacyjnych

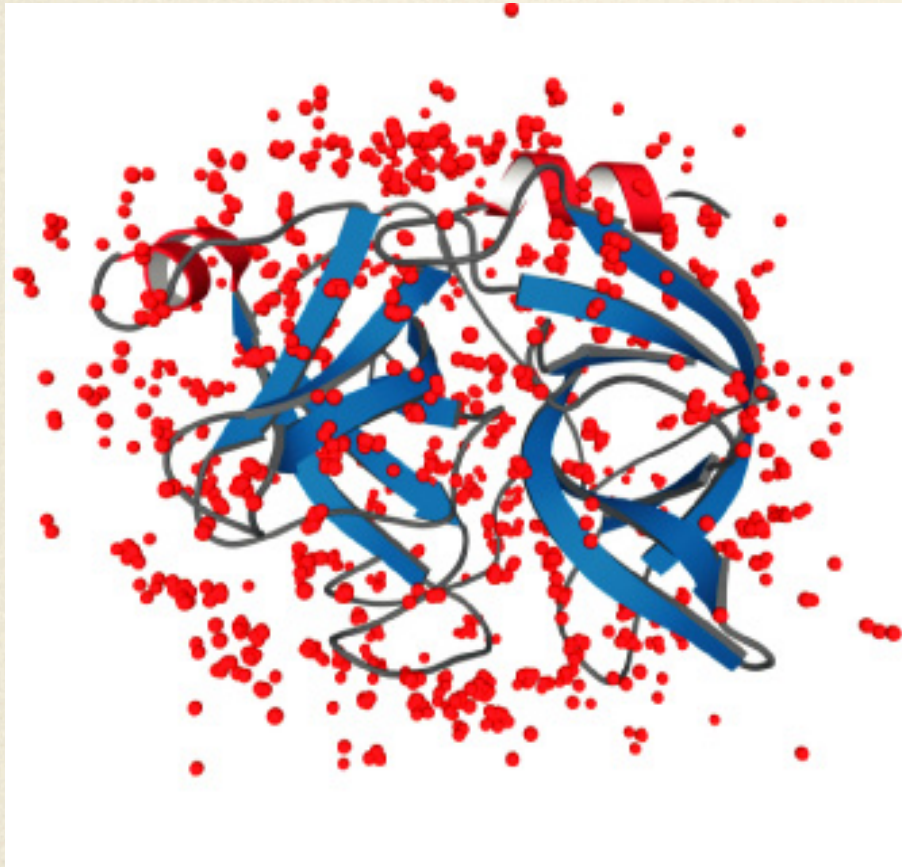
Stan natywny jest konformacyjnie ograniczony, a jego entropia konformacyjna jest bardzo niska. Jak to się dzieje, iż stan o niższej entropii jest bardziej preferowany niż stan o entropii wyższej?

Otóż "utrata" entropii musi być zrekompensowana przez efekty energetyczne, aby spełniony był podstawowy warunek samorzutności procesu:

$$\Delta G_{\text{fałdowania}} = \Delta H_{\text{fałdowania}} - T \Delta S_{\text{fałdowania}} < 0$$

Dla prawidłowego opisu zjawiska należy brać pod uwagę nie tylko zmiany funkcji termodynamicznych samego białka, ale także rozpuszczalnika.

Woda i białko



Pierwsza powłoka hydratacyjna elastazy

- cząsteczki wody ściśle związane z białkiem stanowią część jego struktury
- atomy we wnętrzu są upakowane prawie jak w ciele stałym
- kanały i szczeliny pozwalają jednak na ruch atomów i zapewniają elastyczność
- jeśli w rdzeniu znajdują się większe przestrzenie, to wypełnione są cząsteczkami wody, które mogą oddziaływać z okolicznymi grupami polarnymi

Zwijanie białka

Wkład rozpuszczalnika na proces zwijania białek:

1. Entropia uwolnienia cząsteczek wody podczas fałdowania: w stanie zdenaturowanym występuje wyeksponowanie hydrofobowych reszt do rozpuszczalnika.
2. Ciepło właściwe rozwijania: rozpuszczalnik ma także inny wpływ na proces zwijania/rozwijania łańcucha polipeptydowego, a mianowicie poprzez swoje ciepło właściwe C_p (definiowane jako energia potrzebna do podniesienia temperatury jednego mola substancji o 1K).

Wysoka wartość C_p ma kilka poważnych konsekwencji:

- Entalpia rozwijania zmienia się znacząco wraz ze zmianą temperatury
- Entropia rozwijania rośnie wraz z temperaturą

Powoduje to wraz ze wzrostem temperatury zmianę znaku przy wartości G

Modelowanie ab initio

problemy:

- fizyczne podstawy stabilności strukturalnej białek ciągle nie są w pełni poznane
- sekwencja I-rzędowa może nie determinować struktury III-rzędowej czasami łańcuchowi białka towarzyszą czaperony (ang. chaperones) białka, które mają zdolność do indukowania określonej ścieżki zwijania innych białek
- struktura natywna uzależniona jest od warunków środowiska i nie zawsze biologicznie czynne białko ma strukturę preferowaną termodynamicznie
- symulacja zwijania się białka w strukturę natywną przy pomocy metod dynamiki molekularnej nie jest powszechnie stosowane zarówno ze względów teoretycznych jak i praktycznych.

Modelowanie ab initio

Modelowanie struktur białkowych *de novo /ab initio* polega na zbudowaniu trójwymiarowego modelu białka „od zera”.

- Zredukowane modele siatkowe
- Rosetta

Zredukowana reprezentacja struktury białka

- uwzględnia się tylko kilka pseudoatomów na resztę aminokwasową
- np. $C\alpha$ $C\beta$
- ograniczona liczba możliwych położeń pseudoatomów
- specjalne potencjały dla pseudoatomów
- problem przejścia od reprezentacji zredukowanej do pełnej (atomowej)
- metody często łączone z elementami rozpoznawania kształtów

grupa Andrzeja Kolińskiego biocomp.chem.uw.edu.pl

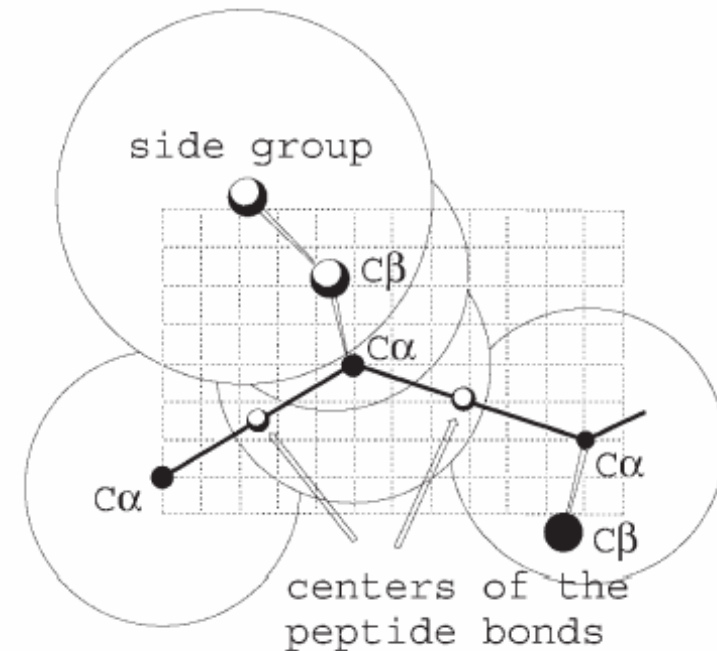
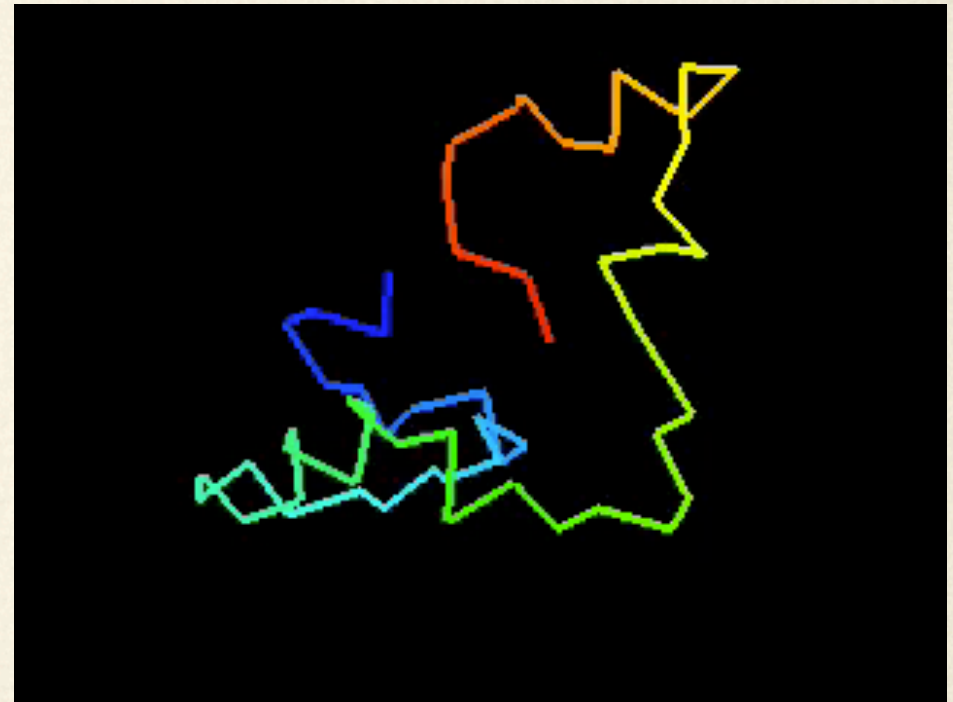
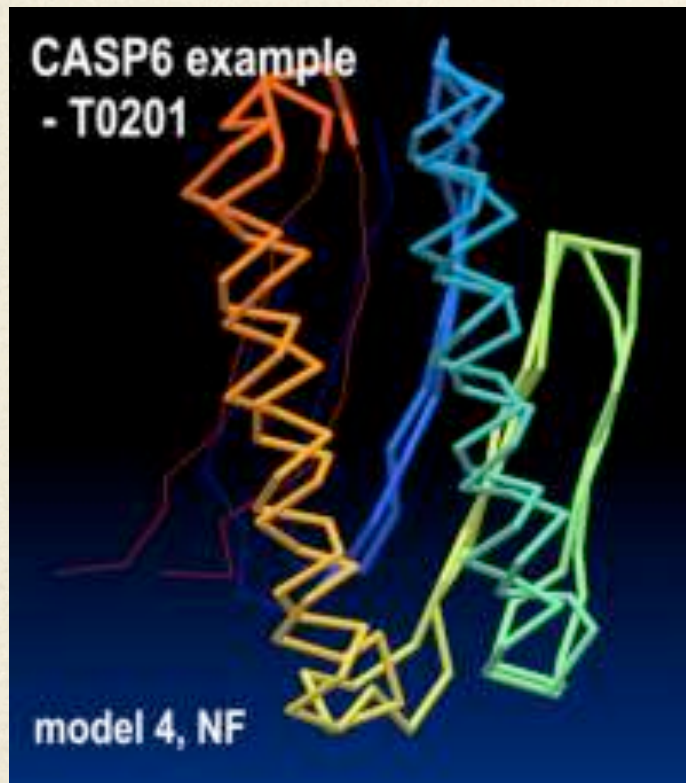


Figure 2. Schematic drawing of the reduced representation of a fragment of a protein chain.

2000-01-09

Zredukowana reprezentacja struktury białka



Dynamika Monte Carlo

dynamika Monte Carlo (MC) - polega na wprowadzaniu losowych zmian we fragmentach struktury i obliczaniu energii nowej konformacji.

- kroki prowadzące do konformacji o niższej energii są zawsze akceptowane
- kroki prowadzące do wyższych temperatur są akceptowane lub odrzucane w zależności od prawdopodobieństwa

Modelowanie ab initio

problemy:

-dokładna i wiarygodna funkcja oceniająca energię swobodną układu

Modelowanie ab initio

problemy:

-dokładna i wiarygodna funkcja oceniająca energię swobodną układu

-zamiast energii „fizycznej” oblicza się „pseudoenergię”, czyli potencjał statystyczny wyprowadzony z analizy częstości występowania oddziaływań danego typu w uprzednio poznanych strukturach

Modelowanie ab initio

metody wykorzystujące oba podejścia
(fizyczne i ewolucyjne) do modelowania

1. alternatywne przyrównania pomiędzy sekwencją celu i białkami o znanych strukturach

Modelowanie ab initio

metody wykorzystujące oba podejścia
(fizyczne i ewolucyjne) do modelowania

1. alternatywne przyrównania pomiędzy sekwencją celu i białkami o znanych strukturach
2. budowa alternatywnych modeli i ocena ich jakości

Modelowanie ab initio

metody wykorzystujące oba podejścia
(fizyczne i ewolucyjne) do modelowania

1. alternatywne przyrównania pomiędzy sekwencją celu i białkami o znanych strukturach
2. budowa alternatywnych modeli i ocena ich jakości
3. poprawa modelu

Modelowanie ab initio

metody wykorzystujące oba podejścia
(fizyczne i ewolucyjne) do modelowania

1. alternatywne przyrównania pomiędzy sekwencją celu i białkami o znanych strukturach
2. budowa alternatywnych modeli i ocena ich jakości
3. poprawa modelu
4. niekompletny model można uzupełnić przez lokalne przeszukiwanie przestrzeni konformacyjnej

Rosetta

-Modele budowane są z krótkich, 9- i 3-aminokwasowych fragmentów znanych struktur, tworzących „bibliotekę” możliwych konformacji.

-Danemu regionowi sekwencji nie jest przyporządkowywana na stałe jedna konformacja.

-ROSETTA przeprowadza symulację, w trakcie której 9- lub 3-aminokwasowe odcinki sekwencji celu przyjmują różne konformacje w oparciu o model probabilistyczny opisujący związki między konformacją i sekwencją fragmentów.

-Lista konformacji dopuszczalnych dla wszystkich odcinków sekwencji ustalana jest na początku symulacji na podstawie lokalnego podobieństwa odcinków sekwencji celu i przewidywanej struktury drugorzędowej do sekwencji i obserwowanej konformacji fragmentów tworzących „bibliotekę”.

-Fragmenty do budowy modelu pobierane są z niespokrewnionych struktur, które mogą wykazywać globalnie odmienną architekturę.

-Ostateczny model generowany jest przez ROSETTE w oparciu o ocenę energii oraz/lub przez identyfikację globalnych konformacji, które najczęściej powtarzały się w całej symulacji.

Rosetta

Możliwości programu ROSETTA zostały wykorzystane w meta-serwerze ROBETTA, który umożliwia konstruowanie modelu białka częściowo w oparciu o szablon, a częściowo „*de novo*”.

ROBETTA automatycznie dzieli sekwencję celu na regiony, które można wymodelować w oparciu o szablon i na takie, które nie wykazują globalnego podobieństwa do żadnej ze znanych struktur. Część białka jest modelowana poprzez „tradycyjną” homologię, natomiast pozostała część jest zwijana poprzez wstawianie 9- i 3- aminokwasowych fragmentów.

Modelowanie ab initio

Wiele struktur zbudowanych metodami „Darwinowskimi” jest wystarczająco poprawnych i dokładnych, aby stanowić dogodny punkt wyjścia do rozważań dotyczących funkcji badanego białka i może pomóc np. w identyfikacji aminokwasów odpowiedzialnych za stabilność lub oddziaływanie z innymi cząsteczkami w komórce.

Standardowa minimalizacja energii całego modelu „Darwinowskiego” praktycznie zawsze prowadzi do pogorszenia jego jakości, na skutek wprowadzenia go w lokalne, a nie globalne minimum energetyczne.

Dobre modele „Darwinowskie” mają konformację stosunkowo bliską konformacji natywnej i z tego powodu mogą służyć jako punkt wyjścia do lokalnego przeszukiwania przestrzeni konformacyjnej metodami „Boltzmanowskimi”.

Liczba możliwych struktur pozostaje zbyt wielka, by można było obliczyć energię dla wszystkich.

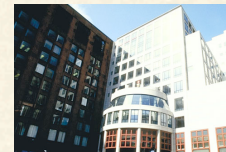
Modelowanie ab initio

metody przewidywania możliwych globalnych zmian konformacyjnych w oparciu o analizę plastyczności białka

Modelowanie ab initio



=



Wartość pracy jednego „twórcy” wirtualnych białek