



*Univerza v Ljubljani*  
*Fakulteta za matematiko in fiziko*  
*Oddelek za fiziko*

Seminar za 4. letnik

# Uvod v zvijanje proteinov za fizike

Anže LOŠDORFER BOŽIČ

*Mentor:* Prof. Dr. Rudolf PODGORNIK

21. marec 2008

## **Povzetek**

V seminarju predstavimo osnove teorije zvijanja proteinov z vidika statistične fizike. Spočetka obravnavamo biološko zgradbo proteinov, nato pa definiramo omejeni problem zvijanja proteinov z vsemi privzetki. Zatem se posvetimo osnovnim enačbam, ki jih dobimo s statistično-mehansko obravnavo problema, in izpostavimo pomen energijske funkcije za razumevanje zvijanja proteinov. Na koncu si ogledamo še model spinskih stekel, s katerim je moč kvalitativno opisati lastnosti energijskih funkcij proteinov.

# Kazalo

<b>1</b>	<b>Uvod</b>	<b>3</b>
<b>2</b>	<b>Zgradba proteinov</b>	<b>4</b>
2.1	Aminokislina in primarna struktura . . . . .	4
2.2	Sekundarna struktura . . . . .	5
2.3	Terciarna in kvartarna struktura . . . . .	7
<b>3</b>	<b>Problem zvijanja proteinov</b>	<b>8</b>
3.1	Mikroskopski opis mehanizmov zvijanja . . . . .	9
3.2	Energijske funkcije in Levinthalov paradoks . . . . .	12
<b>4</b>	<b>Model spinskih stekel</b>	<b>13</b>
4.1	Analogija s spinskimi stekli . . . . .	14
4.2	Princip minimalne frustracije . . . . .	15
<b>5</b>	<b>Zaključek</b>	<b>17</b>

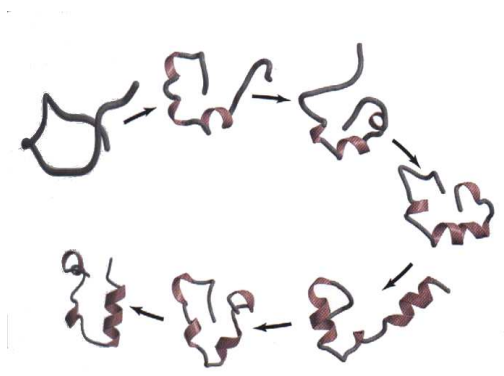
# 1 Uvod

Proteini so najbolj razširjene biološke makromolekule – moč jih je najti v vseh živih bitjih na Zemlji, kjer se pojavljajo v številnih oblikah. Raznolike biološke funkcije, ki jih opravljajo, imajo pomembno vlogo pri skoraj vseh opravilih, ki so bistvena za preživetje celice in celega organizma. Med drugim sodelujejo pri transportu snovi, imajo sposobnost vezanja in razcepljanja točno izbranih molekul ter delujejo kot prenašalci signalov [1, 2]. Manko proteina ali njega napačno delovanje sta tako vzrok številnim boleznim, zaradi česar je proteinska znanost vedno večjega pomena pri razvoju moderne medicine.

Sama funkcija proteina je odvisna od njegove strukture. Vsi proteini so sestavljeni iz istega nabora dvajsetih različnih aminokislin, povezanih v različne kombinacije in zaporedja, katerih posledica so raznolike lastnosti in namembnosti proteinov. Izolirani proteini pa običajno obstajajo v zgolj eni stabilni trodimenzionalni strukturi, ki ji rečemo *nativna konformacija*. Večina proteinov opravlja svojo vlogo le v tej nativni obliki, zato je poznavanje prostorske oblike proteinov pomembno za napovedovanje in razumevanje življenjskih mehanizmov.

Eksperimenti Christiana Anfinsena so pokazali, da je celotna informacija, potrebna, da protein doseže nativno obliko, zajeta že v zaporedju aminokislin (in torej ni posledica kakšnih drugih celičnih mehanizmov) [3]. Ta trditev nam omogoča, da izoliramo sistem, ki ga preučujemo, in natančno definira *problem zvijanja proteinov* – napovedovanje trodimenzionalne nativne strukture proteinov iz poznavanja njihovih zaporedij aminokislin ter fizikalnih zakonov.

Problem zvijanja proteinov je eden izmed najpomembnejših še ne rešenih problemov bioloških znanosti. Dandanes se napovedovanja prostorske strukture proteinov iz zaporedja aminokislin lotevamo na dva načina [1]. Prvi način (t.i. *knowledge-based*) je hiter in pragmatičen, uporablja vse empirične informacije, ki so že na voljo, in preko številnih poskusov izboljšuje postopke za napovedovanje brez poznavanja podrobnosti fizikalnih procesov, ki so v ozadju. Drugi je počasen, premišljen pristop, kjer na osnovi začetnih predpostavk iščemo pot do nativne konformacije preko fizikalnih zakonov (*ab initio protein folding*).



Slika 1: Računalniška simulacija postopnega zvijanja segmenta proteina *villina* s 36 aminokislinskimi ostanki [4]

## 2 Zgradba proteinov

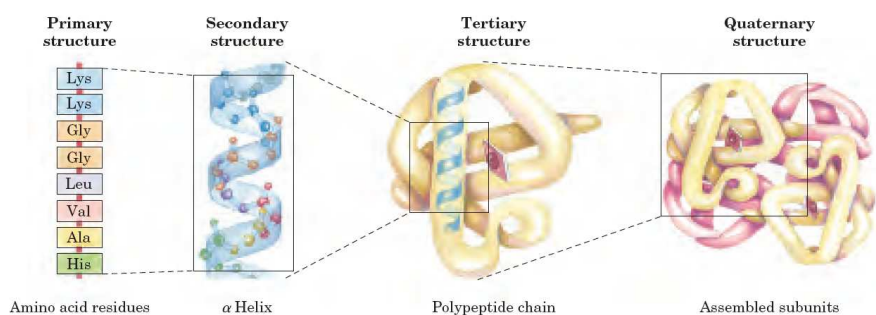
Navzlic raznovrstnim biološkimi funkcijami so proteini z vidika kemijske sestave razmeroma homogen razred molekul. So namreč linearni heteropolimeri – nerazvejane verige različnih monomernih enot, aminokislin, ki lahko same obstajajo kot stabilne molekule. Obliko proteina opisujemo na štirih različnih nivojih (Slika 2):

**primarna** struktura podaja zaporedje aminokislin kot gradnikov verige,

**sekundarna** struktura opisuje ponavljajočo se lokalno zgradbo delov polipeptida,

**terciarna** struktura je prostorska oblika celotnega proteina,

**kvartarna** struktura pa lahko nastane kot tvorba nekaterih proteinskih molekul, ki se med seboj povezujejo in tvorijo proteinski kompleks.



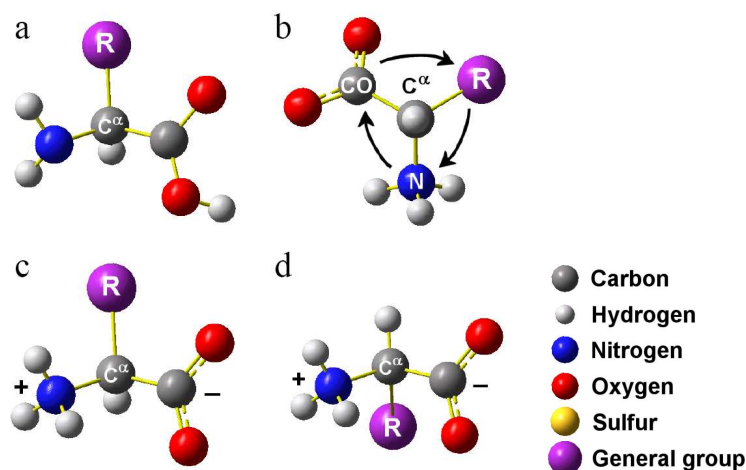
Slika 2: Različni nivoji zgradbe proteina – primarna, sekundarna, terciarna in kvartarna struktura [4]

### 2.1 Aminokislina in primarna struktura

Aminokislina so spojine, ki sestojijo iz centralnega ogljikovega atoma ( $\alpha$ -ogljik), na katerega se vežejo štiri skupine: amino ( $\text{NH}_2$ ) in karboksilna ( $\text{COOH}$ ) skupina, vodikov atom in četrta, stranska skupina (R) (Slika 3). V vodni raztopini in pri fizioloških pogojih sta tako amino kot karboksilna skupina električno nabiti<sup>1</sup>; prva sprejme en proton in je pozitivno nabita, dočim druga odda proton in je negativno nabita [1, 5]. Ta oblika ne spodbuja polimerizacije – zanjo je potrebna še pomoč encima (ribosoma).

Kadar stranska veriga R ni enaka eni izmed preostalih treh skupin, vezanih na  $\alpha$ -ogljik, je aminokislina *kiralna* – obstaja v dveh oblikah, ki sta zrcalni sliki druga druge in ju torej ne moremo prekriti zgolj z rotacijo v prostoru. Pravimo, da  $\alpha$ -ogljik predstavlja kiralni center, aminokislina pa obstaja v dveh različnih *enantiomerih*, imenovanih L- in D- oblika (Sliki 3c in 3d). Iz razlogov, ki so še vedno neznani, so skoraj vse aminokislina v proteinih L-enantiomeri [1, 4].

<sup>1</sup>t.i. zwitterionska oblika aminokislin



Slika 3: Aminokislina. (a) Nevtralni L-enantiomer. (b) CORN mnemotehnično pravilo za določanje L-oblike. (c) Nabiti L-enantiomer, prevladujoča oblika aminokislin v živih bitjih. (d) Nabiti D-enantiomer. [1]

Čeravno ena cela nedoločena stranska skupina ponuja navidez neomejeno število različnih možnosti, nastopa v proteinih živih bitij le dvajset *standardnih aminokislin*, ki so določene z genskim zapisom. Razvrstimo jih lahko v pet glavnih skupin glede na lastnosti njihovih stranskih skupin, predvsem polarnosti in električnega naboja le-teh (Slika 4).

Aminokislina se v polimerno verigo vežejo kovalentno s *peptidno vezjo*, ki nastane z odcepitvijo molekule vode. En vodikov atom prispeva amino skupina prve aminokislina, drugi vodikov in en kisikov atom pa karboksilna skupina druge (Slika 5). Zaporedje aminokislin tako nastalega proteina imenujemo primarna struktura. Glavno verigo tvori ponavljajoče se zaporedje atomov  $C^\alpha$ , C in N, nanjo pa so pripete stranske skupine R. Posamezne aminokislina v polipeptidu imenujemo *ostanki*<sup>2</sup>, da jih ločimo od njihovih samostojnih oblik. Specifičnost vsakega proteina je posledica različnih lastnosti dvajsetih stranskih skupin in njihovih položajev v zaporedju.

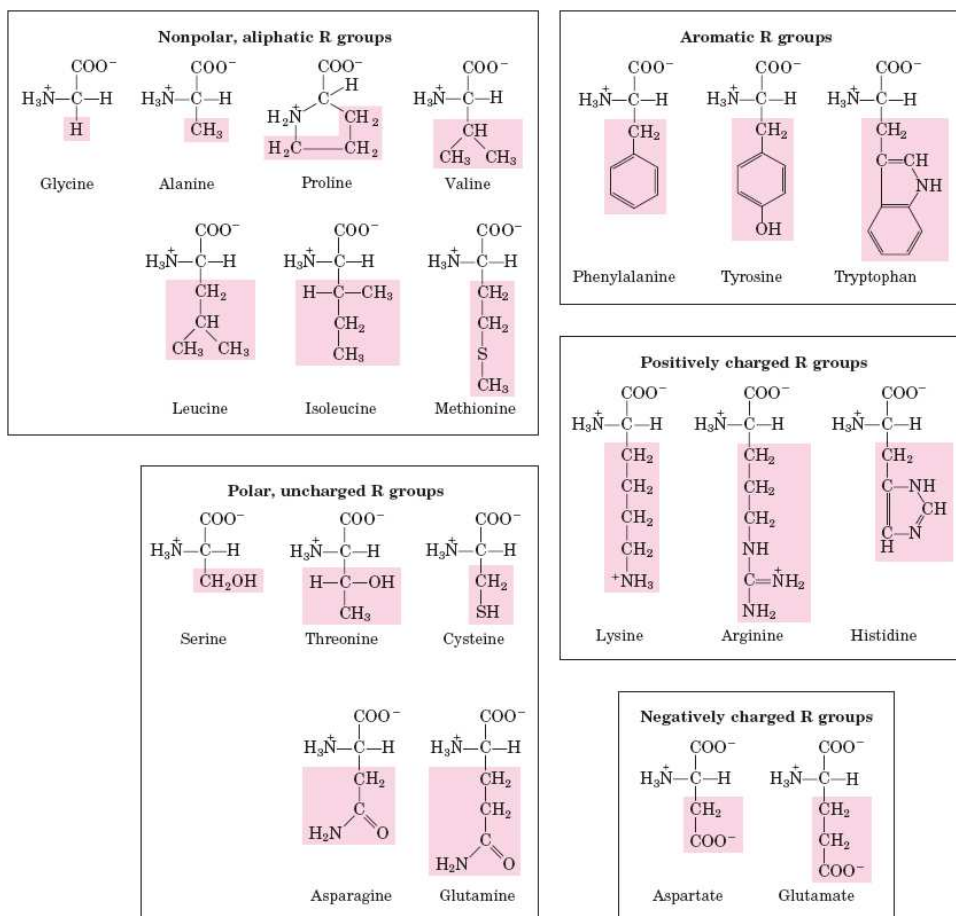
## 2.2 Sekundarna struktura

Izkaže se, da med atomi peptidne vezi dvojna vez ne postoji zgolj med karboksilnim ogljikom in kisikovim atomom, marveč pride do delokalizacije dveh elektronskih parov med vezema C–O in C–N. Zavrlo tega resonančnega efekta se (delna) dvojna vez razteza od dušikovega atoma preko karboksilnega ogljika do kisikovega atoma [4, 5].

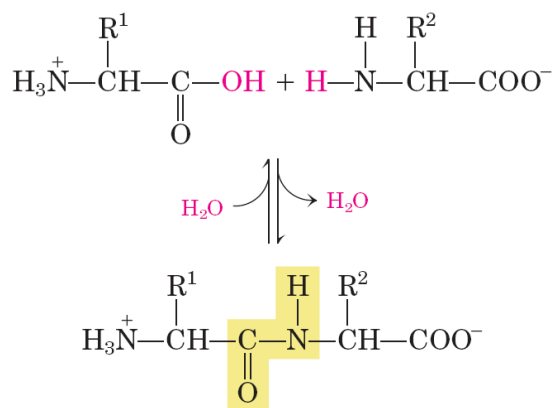
Peptidna vez ima torej naravo delne dvojne vezi, zaradi česar ti trije atomi (N, C in O) težijo h koplanarnosti. V ravnini, ki jo le-ti definirajo<sup>3</sup>, prav tako ležita  $\alpha$ -ogljika in vodikov atom (Slika 6). Ta koplanarnost dovoljuje zgolj dve konformaciji: v prvi, imenovani *trans*,  $\alpha$ -ogljika ležita na različnih straneh linije, ki vsebuje vez C–N, v drugi,

<sup>2</sup>amino acid residues

<sup>3</sup>imenujemo jo tudi *peptidna ravnina*



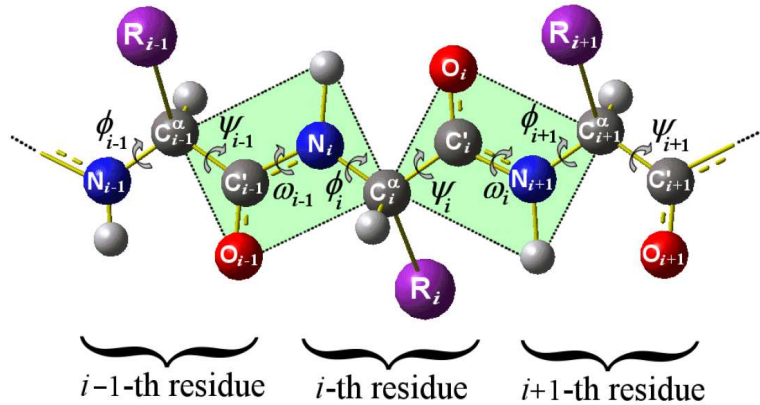
Slika 4: Dvajset standardnih aminokislin. Strukturne formule prikazujejo stanje ionizacije, ki prevladuje pri pH 7.0, rdeča območja pa označujejo stranske skupine. [4]



Slika 5: Nastanek peptidne vezi z odcepitvijo molekule vode [4]

imenovani *cis*, pa ležita na isti strani te linije. Velika večina peptidnih vezi v proteinih je v *trans* konformaciji [1, 4].

Edini rotaciji, ki sta tako še dovoljeni, sta rotaciji okrog vezi  $N-C^\alpha$  in  $C^\alpha-C$ . Ti prostostni stopnji ponavadi označujemo s  $\varphi$  in  $\psi$ , imenujemo pa ju *Ramachandranova kota*. Glavno verigo polipeptida si lahko sedaj zamišljamo kot niz togih ravnin, kjer si zaporedne ravnine delijo skupno točko rotacije v  $C^\alpha$  (Slika 6).



Slika 6: Rotacijske prostostne stopnje v glavni verigi polipeptida. Peptidne ravnine so označene z zelenimi pravokotniki. [1]

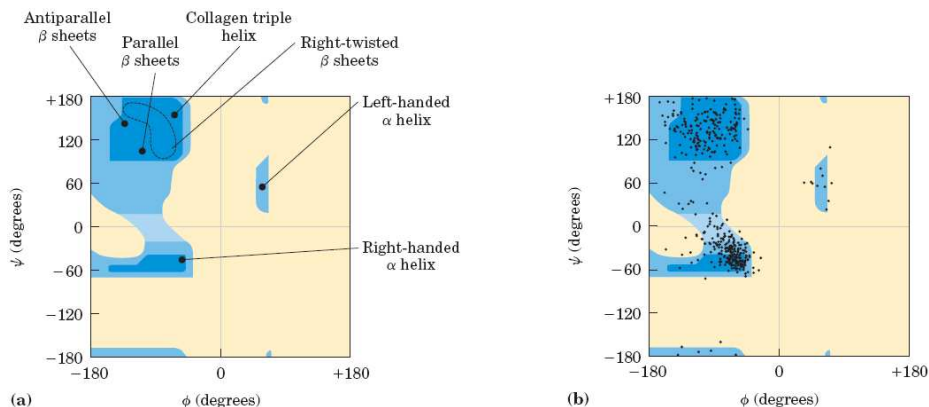
Dovoljene vrednosti Ramachandrovih kotov lahko grafično prikažemo na faznem diagramu ( $\varphi, \psi$ ). Primer je prikazan na Sliki 7. Diagram razpade v več območij, ki ustrezajo posebnim (lokalnim) konformacijam – elementom sekundarne strukture. Kljub temu, da lahko vsak tip sekundarne strukture povsem opišemo s koti zvijanja pri posameznem ostanku, je bolje označiti skupine z ozirom na vzorec vodikovih vezi, ki jih tvorijo, saj je to lastnost, zaradi katere so te oblike prevladujoče [4, 5].

Prvi tak element je  $\alpha$ -vijačnica, spiralna struktura s  $\sim 3.6$  ostanka na zavoju, kjer karbonilna skupina ( $C=O$ ) vsakega *i*-tega ostanka tvori vodikovo vez z amino skupino ( $N-H$ ) ostanka  $i + 4$ . To ni edina takšna struktura; omenimo lahko še  $\pi$ -vijačnico in  $3_{10}$ -vijačnico, a se pojavljata bistveno redkeje.

Poleg desnosučne  $\alpha$ -vijačnice se v proteinih najpogosteje pojavlja še  $\beta$ -list, kjer se vodikove vezi tvorijo med sosednjimi verigami, ki potekajo (anti-)vzporedno in tako tvorijo ravninsko strukturo. Na mestih, kjer polipeptidna veriga spremeni smer, pa se pogosto nahaja še  $\beta$ -obrat.

### 2.3 Terciarna in kvartarna struktura

Naslednji nivo zgradbe proteinov, ki je posledica sestave elementov sekundarne strukture in segmentov verige brez pravilnosti v dobro določeno prostorsko strukturo, imenujemo terciarna struktura. Glede na slednjo delimo proteine na dve vrsti [4]: *globularne* proteine, sestavljene iz različnih sekundarnih struktur, in *fibrilarne* proteine, ki jih tvori



Slika 7: Ramachandranov diagram za različne strukture. **(a)** Vrednosti kotov  $\varphi$  in  $\psi$  za številne dovoljene sekundarne strukture. **(b)** Vrednosti kotov  $\varphi$  in  $\psi$  za vse aminokislinske ostanke z izjemo Gly v encimu *piruvat-kinaze*. [4]

večinoma sekundarna struktura ene vrste. Obe se med seboj razlikujeta tudi po funkcijah, ki jih opravljajo proteini iz posamezne skupine. S kvartarno strukturo, ki opisuje združevanje posameznih proteinskih monomerov v bolj kompleksne sisteme, se na tem mestu ne bomo ukvarjali.

Problem zvijanja proteinov lahko glede na to opredelimo kot poskus napovedovanja sekundarne in terciarne strukture proteinov iz znane primarne strukture.

### 3 Problem zvijanja proteinov

Bioško funkcionalna nativna konformacija proteina je zelo kompleksna. Najbolj opazni značilnosti zgradbe proteina sta njena iregularnost in pomanjkanje simetrije, zaradi česar je bilo sprva mišljeno, da je za zapleten proces zvijanja odgovoren nek poseben celični mehanizem [1].

Velik korak k razumevanju zvijanja proteinov predstavljajo, kot je bilo že omenjeno, eksperimenti Anfinsena, ki je opazoval ponovno zvijanje proteina ribonukleaze [3]. Le-to tvori enojna veriga s 124 aminokislinskimi ostanki, stabilizirajo pa jo štirje disulfidni mostički. V prvem koraku je proteinu dodal reducirajočo snov, s katero je razcepil mostičke, nato pa še ureo, snov, ki spodbuja denaturacijo ("razvitje"). Od tod je nadaljeval na dva načina. Pri prvem je najprej odstranil ureo, nato pa dodal oksidirajočo snov za ponovno vzpostavitev disulfidnih vezi, pri drugem pa je dodal oksidirajočo snov pred odstranitvijo uree. S slednjim načinom je nastala neaktivna mešanica produktov, ki je imela približno 1% aktivnosti nativnega encima, dočim je prvi način proizvedel homogen vzorec, nerazločljiv od začetnega nativnega proteina, ki je obdržal vso biološko aktivnost [1, 5].

Pomemben zaključek teh eksperimentov je, da je vsa informacija, potrebna za zvitje proteinov, zajeta v zaporedju aminokislin. Resda obstajajo tudi t.i. molekularni šaperoni, ki pomagajo pri zvijanju proteinov v celičnem okolju, a po najbolj splošno sprejetem pogledu ne dodajo nikakršne strukturne informacije k procesu. V nadaljevanju bomo to-



vrstne mehanizme zanemarili, tako kot razne izjeme, denimo nekatere proteine, ki se zvijajo že med sintezo v ribosomu, proteine, ki vsebujejo *cis* konformacije aminokislin, . . . , in predpostavili, da bodo ugotovitve o mehanizmih zvijanja *in vitro* majhnih proteinov, pri katerih zvijanje poteka hitro, prav tako veljale za njihovo zvijanje *in vivo*, v večji meri pa tudi za zvijanje posameznih domen večjih proteinov.

Tako lahko definiramo *omejeni problem zvijanja proteinov*<sup>4</sup> kot [1] celostni opis fizikalnega obnašanja povsem sintetiziranih proteinov, v vodni raztopini in pri fizioloških pogojih, sestavljenih zgolj iz dvajsetih standardnih aminokislin, brez dodatnih pripojenih molekul, ki posledično zajema tudi napovedovanje njihove nativne konformacije.

### 3.1 Mikroskopski opis mehanizmov zvijanja

Zmogljiv teoretični okvir za razumevanje in napovedovanje konformacijskih lastnosti proteinov predstavlja statistična mehanika. V tem razdelku si bomo ogledali pomembnejše relacije, ki veljajo za tovrstne sisteme, pri čemer bomo sledili idejam iz [1] in [6].

Omejimo se na primer, kjer se makroskopski parametri, kot sta denimo temperatura  $T$  in število molekul vode  $N_w$ , ne spreminjajo. Fizikalni sistem je torej *en protein, obkrožen z  $N_w$  molekulami vode*, in ustreza prej definiranemu omejenemu problemu zvijanja. Seveda so vsi nadaljnji argumenti in izpeljava povsem enaki za razredčeno vodno raztopino makroskopskega števila proteinov, med katerimi ni interakcij.

Predpostavimo še, da se sistem ravna po zakonih klasične mehanike, tako da lahko vsako mikroskopsko stanje v celoti opišemo z generaliziranimi koordinatami in momenti atomov, pripadajočih proteinu (označimo jih z  $x^\mu$  in  $\pi_\mu$ ,  $\mu = 1, \dots, N$ ), in atomov, ki pripadajo molekulam vode (označimo jih z  $X^m$  in  $\Pi_m$ ,  $m = N + 1, \dots, N + N_w$ ). Fazni prostor vseh mikroskopskih stanj označimo z  $\Gamma \times \Gamma_w$ , s čimer smo poudarili, da je direktni produkt faznega prostora proteina  $\Gamma$  in faznega prostora molekul vode  $\Gamma_w$ .

Časovni razvoj sistema določa Hamiltonova funkcija [7]:

$$H(x^\mu, X^m, \pi_\mu, \Pi_m) = \sum_{\mu} \frac{\pi_\mu^2}{2M_\mu} + \sum_m \frac{\Pi_m^2}{2M_m} + V(x^\mu, X^m) \quad , \quad (1)$$

kjer so  $M_\mu$  in  $M_m$  atomske mase,  $V(x^\mu, X^m)$  pa potencialna energija sistema, odvisna zgolj od koordinat proteina in molekul vode. Po vzpostavitvi ravnovesnega stanja pri temperaturi  $T$  lahko zapišemo fazno vsoto za kanonični ansambel kot [1, 6]

$$Z(T, V) = \frac{1}{h^{N+N_w} N_w!} \int_{\Gamma \times \Gamma_w} \exp(-\beta H(x^\mu, X^m, \pi_\mu, \Pi_m)) dx^\mu dX^m d\pi_\mu d\Pi_m \quad , \quad (2)$$

kjer je  $\beta = 1/k_B T$ , faktor  $N_w!$  pa je posledica nerazločljivosti molekul vode; integral teče po celotnem faznem prostoru  $\Gamma \times \Gamma_w$ . V našem primeru je multiplikativni faktor pred integralom zgolj konstanta in kot tak predstavlja le premik ničle v prosti energiji, zato ga lahko v nadaljevanju brez izgube splošnosti opuščamo.

Ker nas zanima predvsem obnašanje konformacij polipeptidne verige, v naslednjem koraku izpovprečimo koordinate in momente molekul vode. To moremo storiti, saj je v

---

<sup>4</sup>*restricted protein folding problem*

večini primerov ravnovesno stanje raztopine doseženo zelo hitro v primerjavi z gibanjem makromolekule [6]. Integracija po momentih  $\Pi_m$  doprinese zgolj člen, sorazmeren s temperaturo  $T$ , ki ga lahko izpustimo iz nadaljnje obravnave iz istih razlogov, ki so navedeni v prejšnjem odstavku. Manj trivialna je integracija koordinat  $X_m$ , katero lahko, z izjemo zelo preprostih potencialov, opravimo samo formalno. V ta namen definiramo *efektivno potencialno energijo*<sup>5</sup>

$$W(x^\mu; T) \equiv -k_B T \ln \left( \int \exp(-\beta V(x^\mu, X^m)) dX^m \right) , \quad (3)$$

ki je funkcija zgolj temperature in koordinat atomov proteina, fazno vsoto  $Z$  (2) pa prepisemo v

$$Z = \int_{\Gamma} \exp(-\beta H_{\text{eff}}(x^\mu, \pi_\mu; T)) dx^\mu d\pi_\mu , \quad (4)$$

kjer smo uvedli *efektivno Hamiltonovo funkcijo*

$$H_{\text{eff}}(x^\mu, \pi_\mu; T) = \sum_{\mu} \frac{\pi_{\mu}^2}{2M_{\mu}} + W(x^\mu; T) . \quad (5)$$

Dobljeni izraz (4) lahko še poenostavimo, če izpovprečimo tudi proteinske momente  $\pi_{\mu}$ . Pri tem moramo biti pozorni na dejstvo, da zaradi tega koraka verjetnostna gostota v  $x^\mu$ -prostoru ni več invariantna na kanonične transformacije [1, 8]. S tem v mislih dobimo z integracijo momentov  $\pi_{\mu}$  znova člen, ki je sorazmeren s temperaturo  $T$ , zaradi česar ga v nadaljevanju lahko opustimo. Tako dobimo za fazno vsoto

$$Z = \int_{\Omega} \exp(-\beta W(x^\mu; T)) dx^\mu , \quad (6)$$

kjer smo z  $\Omega$  označili prostorski del faznega prostora proteina  $\Gamma$ .

Vredno je pripomniti, da lahko, vkolikor je moč prvotno potencialno energijo  $V(x^\mu, X^m)$  ločiti na vsoto interakcij med proteini, interakcij med vodnimi molekulami in interakcij protein-raztopina, pišemo efektivno potencialno energijo  $W(x^\mu; T)$  kot vsoto dveh delov: proteinske energije v vakumu in efektivne energije raztopine [6, 9]. Ta poenostavitev ni nujna za nadaljnje sklepe, zato se je ne bomo poslužili.

Sedaj lahko zapišemo verjetnostno gostoto v faznem prostoru  $\Omega$ :

$$p(x^\mu) = \frac{\exp(-\beta W(x^\mu))}{Z} , \quad (7)$$

kjer smo izpustili eksplicitno odvisnost energije od temperature  $T$ , saj je v obravnavanem primeru le-ta konstantna [1]. Od tod vidimo, da so konformacijske lastnosti polipeptidne verige v termodinamskem ravnovesju povsem določene z energijsko funkcijo  $W(x^\mu)$ . Makroskopski opis stanja proteina brez mikroskopskih podrobnosti pa dobimo iz izraza za *Helmholtzovo prosto energijo* [7]:

$$F = -k_B T \ln Z . \quad (8)$$

---

<sup>5</sup>*potential of mean force*

Pogosto se uporablja tudi opis konformacij, kjer ne gledamo posameznih stanj v faznem prostoru  $\Omega$  ali celotne množice, marveč definiramo *stanja*, končne podmnožice  $\Omega_i \subset \Omega$ , vsebujoča številne različne konformacije, ki pa so si v nekem oziru sorodne. Ta stanja morajo seveda biti dobro definirana ter zadoščati določenim pogojem – najpomembnejši izmed teh je, da se nobena konformacija ne sme nahajati v dveh različnih stanjih hkrati,  $\Omega_i \cap \Omega_j = \emptyset \forall i \neq j$  [1].

Tako definiramo še fazno vsoto stanja  $\Omega_i$  kot

$$Z_i \equiv \int_{\Omega_i} \exp(-\beta W(x^\mu; T)) dx^\mu \quad , \quad (9)$$

verjetnost, da se sistem nahaja v stanju  $\Omega_i$ , pa tedaj podaja

$$P_i = \frac{Z_i}{Z} \quad . \quad (10)$$

Prosta energija takšnega stanja se izraža kot

$$F_i = -k_B T \ln Z_i \quad , \quad (11)$$

za razliko katerih velja zveza:

$$\Delta F_{ij} = F_j - F_i = -k_B T \ln \frac{Z_j}{Z_i} = -k_B T \ln \frac{P_j}{P_i} = -k_B T \ln \frac{[j]}{[i]} = -k_B T \ln K_{ij} \quad , \quad (12)$$

kjer  $[i]$  označuje koncentracijo stanja  $i$ ,  $K_{ij}$  pa je reakcijska konstanta ravnovesja  $i \leftrightarrow j$ .

Če upoštevamo še dejstvo, da sta pri fizioloških pogojih razlika Gibbsovih prostih energij  $\Delta G$  in razlika Helmholtzovih prostih energij  $\Delta F$  približno enaki, saj je člen  $p\Delta V$  zanemarljiv [6], predstavlja relacija (12) povezavo med koncentracijama nativnega in denaturiranega stanja  $[N]$  in  $[U]$ , ki ju ponavadi merimo v laboratoriju, ter prosto energijo zvijanja pri konstantni temperaturi in konstantnem tlaku  $\Delta G_{\text{zvitje}} \equiv G_N - G_U$ , ki opisuje stabilnost proteina in jo je moč tudi teoretično oceniti [1].

Naposled definiramo še verjetnostno gostoto v stanju  $\Omega_i$  kot verjetnostno gostoto iz enačbe (7), pogojeno z dejstvom, da konformacija  $x^\mu$  leži v  $\Omega_i$ :

$$p_i(x^\mu) \equiv p(x^\mu | x^\mu \in \Omega_i) = \frac{p(x^\mu)}{P_i} = \frac{\exp(-\beta W(x^\mu))}{Z_i} \quad , \quad (13)$$

s katero lahko izračunamo tudi notranjo energijo  $U_i$  in entropijo  $S_i$  stanja  $\Omega_i$  [1, 7]:

$$U_i = \langle W \rangle_i = \int_{\Omega_i} W(x^\mu) p_i(x^\mu) dx^\mu \quad , \quad (14)$$

$$S_i = -k_B \int_{\Omega_i} p_i(x^\mu) \ln p_i(x^\mu) dx^\mu \quad . \quad (15)$$

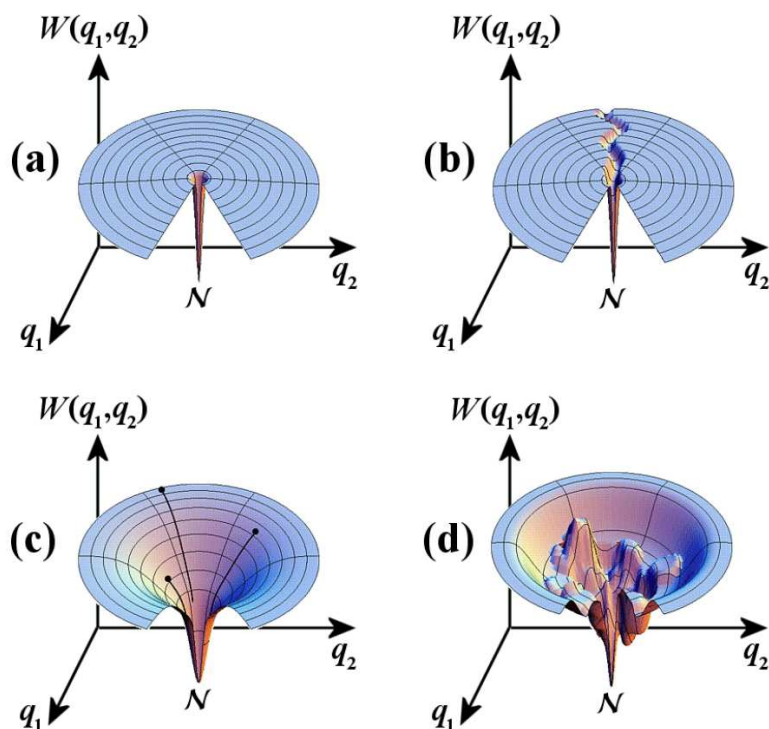
Pokazati je moč tudi, da velja običajna termodinamska zveza med potenciali

$$\Delta F_{ij} = \Delta U_{ij} - T \Delta S_{ij} \simeq \Delta G_{ij} = \Delta H_{ij} - T \Delta S_{ij} \quad , \quad (16)$$

kar izraža intuitivno idejo, da je stabilnost proteina posledica ravnovesja med efektivno energijo, ki favorizira nativno stanje, ter konfiguracijsko entropijo, ki daje prednost denaturiranemu stanju.

### 3.2 Energijske funkcije in Levinthalov paradoks

Efektivna potencialna energija  $W(x^\mu)$  očitno povsem določa konformacijske lastnosti polipeptidne verige. Pri tem opisu smo izpovprečili prostostne stopnje raztopine, zaradi česar so sile, ki so posledica tega potenciala, odvisne tudi od temperature. Dinamično izpovprečevanje teh prostostnih stopenj za opis časovnega razvoja proteina je vsekakor zelo zahtevna naloga, zato bomo v nadaljevanju privzeli, da je relaksacija raztopine bistveno hitrejša od gibanja polipeptidne verige. Ta privzetek je sicer težko preveriti v eksperimentalnih situacijah, vendar pa simulacije molekularne dinamike nakazujejo, da je približno izpolnjen. Posledično je energijska funkcija  $W(x^\mu)$  tudi efektivni “dinamični” potencial, ki določa mikroskopski časovni razvoj proteinskih prostostnih stopenj [1, 10, 11].



Slika 8: Možne energijske funkcije  $W(q_1, q_2)$  proteina v dvodimenzionalnem konformacijskem prostoru  $\Omega = \{q_1, q_2\}$  z izpovprečenimi prostostnimi stopnjami raztopine.  $\mathcal{N}$  označuje nativno stanje, ki je tu privzeto za globalni minimum. (a) *Igrišče za golf*: energijska funkcija v primeru Levinthalovega paradoksa. (b) *Mravlja pot*: prvotno zamišljena rešitev Levinthalovega paradoksa. (c) *Gladek lijak*: novejši pogled na energijsko funkcijo proteina. (d) *Realnejši, delno naguban lijak*. [1]

Poznavanje oblike energijske funkcije je tako osrednjega pomena za razumevanje zvijanja proteinov. Pri tem hitro opazimo, da je število prostostnih stopenj  $N$  v povprečnem polipeptidu dovolj veliko, da je velikost konformacijskega prostora, ki raste eksponentno

z  $N$ , astronomska. To dejstvo je predmet *Levinthalovega paradoksa*, ki ga je prvi podal Cyrus Levinthal leta 1969, in ki pravi, da protein ne bo nikoli dosegel nativnega stanja, vkolikor mora tekom zvijanja obhoditi vse možne konformacije, konformacije posameznega ostanka pa so neodvisne od konformacij preostalih delov [1].

Za primer vzemimo protein, sestavljen iz 100 aminokislin, kjer ima vsaka od teh le dve možni konformaciji. Število vseh možnih konformacij za celoten protein je tako  $2^{100} \sim 10^{30}$ . Če ocenimo potreben čas, da protein zavzame vsako izmed njih,  $s \sim 1$  ps, bi protein potreboval približno  $10^{18}$  s (več kot  $10^{10}$  let) za obhod vseh konformacij [10].

Seveda je takšen argument očitno napačen, saj se proteini nenazadnje zvijajo, od koder lahko sklepamo, da zvijanje proteinov ne more biti zgolj naključen sprehod v konformacijskem prostoru. To paradoksalno situacijo ponazarja energijska funkcija oblike igrišča za golf (Slika 8a), kjer točka, ki opisuje konformacijo verige, tava po ravnini denaturiranih stanj, dokler po naključju ne najde nativnega stanja.

Že Levinthal je kot rešitev problema predlagal, da proces zvijanja poteka po dobro definiranih poteh, kjer protein preko vmesnih stanj preide – kakor urejena vrsta mravelj – iz denaturiranega stanja v nativno stanje (Slika 8b). Ta, starejši pogled na zvijanje predvideva le eno takšno pot z minimalno energijo, ki prevladuje nad ostalimi. Vendar pa se je kasneje izkazalo, da trajektorije sistema redko opišejo enako pot, tako da lahko protein doseže nativno konformacijo na več različnih načinov, zato je takšen primer mravlje poti malo verjeten [1, 10].

Na podlagi tega ter ugotovitev iz teorije spinskih stekel se je razvil opis mehanizmov zvijanja, ki predvideva, da ima energijska funkcija obliko lijaka (Sliki 8c in 8d) [1, 10, 12]. Lijak je usmerjen k nativnem stanju, zato ima vsaka mikroskopska trajektorija v vsaki točki faznega prostora večjo verjetnost za razvoj k nativni konformaciji kot za razvoj v nasprotni smeri. Deterministično zvijanje po eni sami poti torej nadomesti statistična obravnava, kjer je zvijanje heterogena reakcija, ki zajema širok ansambel struktur, vmesna stanja, ki so včasih opažena, pa predstavljajo bolj ali manj izrazite grbine v stenah lijaka [1]. Takšna oblika energijske funkcije ni značilnost vseh polimerov s katerikoli zaporedjem aminokislin, marveč je posledica evolucije. Čeprav novejši pogled ne prepoveduje starejšega, saj je slednji le njegov poseben primer, je med obema pomembna biološka razlika – mutacije lahko bistveno bolj dramatično vplivajo na dinamiko, ki poteka po enolični poti, kot na tisto, kjer je poti več [10, 13].

## 4 Model spinskih stekel

Dasiravno energijska funkcija oblike lijaka predstavlja edino konsistentno sliko, s katero je moč pojasniti vsa eksperimentalna dejstva o zvijanju proteinov, še vedno umanjka pojasnilo, zakaj je energijska funkcija ravno takšna. Na prvi pogled se zde zaporedja aminokislin v proteinih, ki se pojavljajo v naravi, domala naključna. Povprašamo se lahko torej, če se naključni heteropolimeri obnašajo enako kot proteini v živih organizmih. Odgovor na to vprašanje sta s kvantitativnim opisom zvijanja podala Bryngelson in Wolynes [12], tako da sta na teh naključnih sekvencah uporabila ideje iz statistične fizike razurejenih sistemov – natančneje, spinskih stekel.

## 4.1 Analogija s spinskimi stekli

Spinska stekla so magnetni sistemi, kjer so interakcije med spini enako pogosto – in tudi naključno – feromagnetne (spini težijo k poravnavi v isto smer) kot antiferomagnetne (spini žele kazati v nasprotni smeri). Nobena razporeditev spinov ne more zadostiti vsem lokalnim interakcijam hkrati, kar vodi do pojava, imenovanega *frustracija*. Enak pojav opazimo v primeru polipeptida z naključnim zaporedjem aminokislin, ki, zavoljo številnih prostostnih stopenj, vezi, povezanih s sestavljenostjo verige, ter interakcij monomernih enot s sosedi in z okoljem, ne more hkrati optimizirati vseh interakcij. Takšna frustracija vodi do energijske funkcije oblike plitvega, grobo nagubanega lijaka z mnogo lokalnimi minimi, visokimi pregradami in pastmi, kar je površina, značilna za spinska stekla [1, 10].

Bryngelson in Wolynes sta kot prva obravnavala posledice grobih oblik energijskih funkcij heteropolimerov za razumevanje zvijanja proteinov. Njun model energijske funkcije proteina je bil [12]

$$H = - \sum_i E_i(\alpha_i) - \sum_i J_{i,i+1}(\alpha_i, \alpha_{i+1}) - \sum_{i,j} K_{i,j}(\alpha_i, \alpha_j, r_i, r_j) \quad , \quad (17)$$

kjer  $\alpha_i$  označuje stanje vsake aminokislina v zaporedju,  $-E_i(\alpha_i)$  je energija vsakega takega stanja, z  $-J_{i,i+1}(\alpha_i, \alpha_{i+1})$  popišemo interakcije med najbližjimi sosedi, interakcije med ostanki, ki so razmeroma daleč v verigi, pa opisuje člen  $-K_{i,j}(\alpha_i, \alpha_j, r_i, r_j)$ , kjer je  $r_i$  položaj  $i$ -tega ostanka. Vsota teče od 0 do  $\nu$ , kjer je z 0 vedno označeno nativno stanje,  $\nu$  pa je število možnih stanj vsakega ostanka v denaturiranem proteinu in je reda velikosti  $\nu \sim 10$ .

V nadaljevanju postuliramo, da lahko energije stanj naključnega heteropolimera približno opišemo z naborom naključnih, neodvisnih energij [12]. Za vsako izmed zgoraj navedenih interakcij ( $E_i$ ,  $J_{i,i+1}$  in  $K_{i,j}$ ) privzamemo, da je porazdeljena po Gaußovi porazdelitvi, zato je takšna tudi porazdelitev energij, dobljenih iz Hamiltonove funkcije (17):

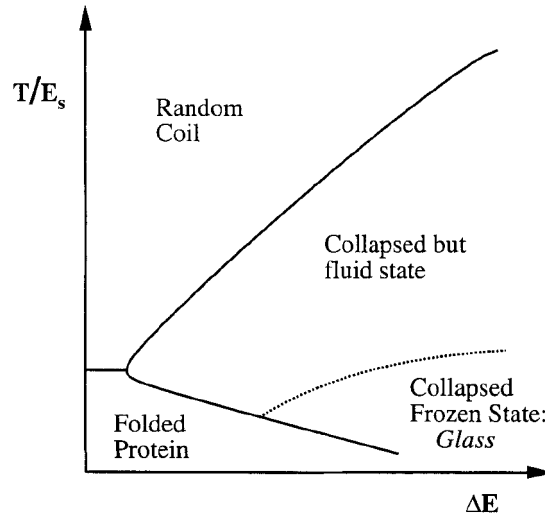
$$P(E) = \frac{1}{\sqrt{2\pi\Delta E^2}} \exp\left(-\frac{(E - \bar{E})^2}{2\Delta E^2}\right) \quad , \quad (18)$$

kjer je  $\bar{E}$  srednja vrednost,  $\Delta E$  pa standardna deviacija. Ta model je v teoriji spinskih stekel znan pod imenom *približek naključnih energij* [10, 14].

Odtod lahko nadaljujemo s statistično obravnavo ter dospemo do izrazov za entropijo, prosto energijo, . . . , s čimer se v tem seminarju ne bomo ukvarjali; vedoželnejši bralec si izpeljavo lahko ogleda v [12] in [10]. Zanimivejši sta namreč dve posledici modela, obstoj kritične temperature steklastega prehoda  $T_g$  ter različnih faz zvitega proteina. Na Sliki 9 vidimo, da model napove obstoj štirih različnih faz proteina v odvisnosti od parametra  $\Delta E$ , ki predstavlja “grobost” površine lijaka energijske funkcije, in parametra  $T/E_s$ . Tu je  $E_s$  energijska razlika med naborom stanj, ki so močno podobna nativnemu stanju, in med naborom stanj, za katere je ta podobnost majhna (denaturirana stanja) [10]. Oba parametra sta seveda odvisna še od temperature, ki bi predstavljala tretjo dimenzijo v diagramu.

Prvo izmed območij, na katera razpade fazni diagram, je urejena faza, kjer proteinske molekule težijo k ureditvi v nativno konformacijo. Naslednja je razurejena faza, v kateri

je protein moč najti v katerikoli konformaciji. V preostalih fazah se proteinske molekule nahajajo v številnih nenativnih stanjih; pod kritično temperaturo prehoda  $T_g$ , ki jo določa nagubanost površine lijaka, pa sistem zamrzne v enega od teh mnogih lokalnih minimov energijske funkcije, zato ta faza spominja na fazo spinskih stekel. Eksperimentalni podatki in fazni diagrami preprostih proteinskih modelov tudi dejansko nakazujejo na obstoj različnih faz [11].



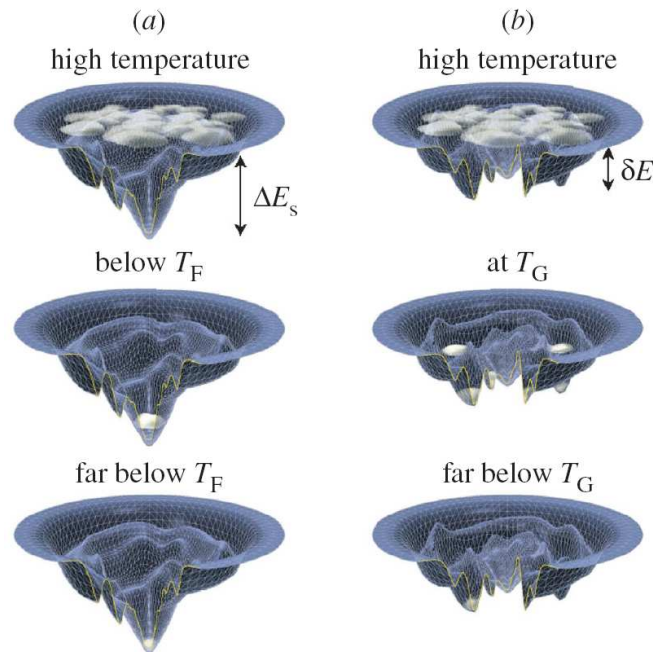
Slika 9: Fazni diagram zvijanja proteina. Diagram prikazuje končna stanja zvijanja proteina, ki bi ga dobili iz eksperimenta. Polne črte nakazujejo fazni prehod prvega reda, črtkana fazni prehod drugega reda. Podrobnosti so navedene v besedilu. [10]

## 4.2 Princip minimalne frustracije

Raznovrstne interakcije v naključnem heteropolimeru so frustrirane – kakorkoli že, protein ni heteropolimer z naključnim zaporedjem. Tekom evolucije so bila izbrana zaporedja, ki se zvijejo v biološko smiselnem času, tako da so konflikti med interakcijami minimalni, zaradi česar se proteini zanesljivo zvijejo v nativno strukturo [13]. Bryngelson in Wolynes sta takšno situacijo, kjer je manj konfliktnih interakcij, kot bi jih pričakovali pri naključnem heteropolimeru, poimenovala *princip minimalne frustracije* [12]. Protein lahko torej definiramo kot polipeptidno verigo, katere zaporedje je bilo evolucijsko izbrano tako, da zadosti principu minimalne frustracije [1].

Princip minimalne frustracije ni zgolj nek abstrakten pojem, saj zanj najdemo dobro merilo v deležu frustracije v sistemu. Ta delež dobimo s primerjavo temperature zvitja proteina  $T_f$  in temperature steklastega prehoda  $T_g$  [13]. Pri tem temperaturo zvitja  $T_f$  običajno definiramo kot temperaturo, pri kateri je delež proteinov v nativnem stanju enak deležu proteinov v vseh ostalih konfiguracijah, odvisna pa je od energijske razlike  $E_s$  [10]. Moč je pokazati, da je razmerje  $T_f/T_g$  monotona funkcija  $E_s/\Delta E$  [15]. Če naj ima energijska funkcija proteina obliko lijaka, mora vrednost  $T_f$  presežati vrednost  $T_g$

– v tem primeru je zaporedje minimalno frustrirano. Tovrstna energijska funkcija je na Sliki 10 primerjana z energijsko funkcijo naključnega heteropolimera.



Slika 10: **(a)** Lijak energijske funkcije evolviranega proteina. Entropija narašča v radialni smeri, energija in podobnost nativni strukturi naraščata z globino. Pri visokih temperaturah se konformacije proteinov (oblački) nahajajo pri vrhu lijaka, pod temperaturo zvijanja  $T_f$  pa nenadoma padejo na njegovo dno. **(b)** Plitkejši in bolj naguban lijak naključnega heteropolimera. Z nižanjem temperature se entropija manjša, a pri temperaturah, dovolj nizkih za ustalitev v nekem stanju, ujetost v pasti močno upočasni zvijanje. To se primeri pri temperaturi steklastega prehoda  $T_g$ . [15]

Vsekakor velja še poudariti, da princip minimalne frustracije ne pomeni, da so energijske funkcije proteinov povsem brez frustracije. Oblika lijaka je namreč posledica občutljivega ravnovesja med entropičnim in entalpičnim prispevkom k prosti energiji (16):

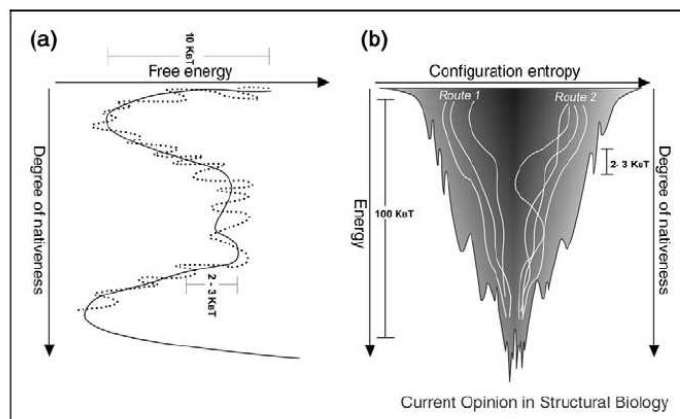
$$\Delta G_{\text{razvitje}} = \Delta H_{\text{razvitje}} - T\Delta S_{\text{razvitje}} \quad . \quad (19)$$

Prosta energija razvitja proteina v raztopini  $\Delta G_{\text{razvitje}}$  redko preseže  $k_B T$ , ta majhna vrednost pa je razlika količin ( $\Delta H_{\text{razvitje}}$  in  $T\Delta S_{\text{razvitje}}$ ), ki sta tipično nekaj redov velikosti večji ( $\sim 100k_B T$ ) [1, 13].

Obe količini sta hkrati še močno odvisni od podrobnosti energijske funkcije  $W(x^t)$ , katere energijska skala je mestoma zelo raznolika (Slika 11). Majhne spremembe v konformaciji, ki predstavljajo pregrade pri zvijanju, so ponavadi reda velikosti nekaj  $k_B T$  in predstavljajo le majhen delež celotne energije zvijanja, ki je reda velikosti  $100k_B T$ . Večje



mutacije dosežejo tudi do  $10k_B T$ , premikanje med temi globljimi, različnimi lokalnimi minimi pa je lahko zelo počasno [10, 13].



Slika 11: Površina energijske funkcije evolviranega proteina. Prepovedane mutacije lahko dosežejo vrednosti  $10k_B T$ , pasti pri zvijanju pa so reda velikosti nekaj  $k_B T$ , kar je le majhen delež celotne energije zvijanja (približno  $100k_B T$ ). [13]

Zavoljo tega je, če uporabimo slabe modele energijskih funkcij, v teoretičnih simulacijah zvijanja velika verjetnost, da model zgreši nativno stanje oziroma celo ne proizvede energijskega lijaka. Dandanašnji modeli energijskih funkcij v splošnem niso zmožni zvijanja proteinov “od začetka”, mnogo pa je tudi dokazov, da so se dejanska zaporedja v proteinih razvila tako dobro, da so manj frustrirana od najboljših obstoječih modelov [1, 13]. In četudi model spinskih stekel ne poskuša posnemati fizikalnih procesov zvijanja proteinov, nudijo nekatere njegove posledice, kot sta denimo princip minimalne frustracije in obstoj faznih diagramov zvijanja, uvid v možne izboljšave modelov energijskih funkcij, z daljnosežnejšim ciljem zanesljivega zvijanja *ab initio*.

## 5 Zaključek

Evolucija je tekom let rešila problem zvijanja proteinov – razumevanje samega procesa pa je vprašanje, ki ostaja pereče še danes. Zvijanje proteinov je namreč nadvse kompleksen proces, kar kaže tudi obravnava problema z orodji statistične fizike, ki pripomorejo k boljšemu poznavanju le-tega. Z njimi je moč izpostaviti pomembnost energijske funkcije proteinov, ki ima osrednjo vlogo pri zvijanju. Če naj se proteini zanesljivo zvijejo v nativno stanje, mora energijska funkcija imeti obliko lijaka, pri čemer je bil velik korak k razumevanju tega storjen z uporabo modela spinskih stekel, ki pa se je naposled izkazal za nezmožnega pojasniti dejanski proces zvijanja. Zvijanje proteinov z vidika osnovnih fizikalnih principov tako še vedno ostaja uganka, vendar lahko v prihajajočih desetletjih pričakujemo, hkrati z napredkom eksperimentalnih in simulacijskih metod, številna nova spoznanja.

## Literatura

- [1] P. Echenique, "Introduction to protein folding for physicists." [arXiv:0705.1845v1](https://arxiv.org/abs/0705.1845v1) [physics.bio-ph], 2007.
- [2] I. Urbančič, "Proteini." Fakulteta za matematiko in fiziko, Seminar za 3. letnik, 2006.
- [3] C. B. Anfinsen, "Studies on the principles that govern the folding of protein chains." Nobel Lecture, 1972.
- [4] D. L. Nelson and M. M. Cox, *Lehninger Principles of Biochemistry*. Freeman, 4th ed., 2005.
- [5] R. Cotterill, *Biophysics: An Introduction*. Wiley, 2002.
- [6] T. Lazaridis and M. Karplus, "Thermodynamics of protein folding: a microscopic view," *Biophys. Chem.* **100** (2003) 367–395.
- [7] K. Huang, *Lectures on Statistical Physics and Protein Folding*. World Scientific, 2005.
- [8] L. D. Landau and E. M. Lifšic, *Mechanics*, vol. 1 of *Course of theoretical physics*. Butterworth-Heinemann, 3rd ed., 2006.
- [9] T. Lazaridis and M. Karplus, "Effective energy function for proteins in solution," *Proteins: Struct. Fun. Gen.* **35** (1999) 133–152.
- [10] J. D. Bryngelson, J. N. Onuchic, N. D. Socci, and P. G. Wolynes, "Funnel, pathways and the energy landscape of protein folding: a synthesis," *Proteins: Struct. Fun. Gen.* **21** (1995) 167–195.
- [11] C. M. Dobson, A. Šali, and M. Karplus, "Protein folding: a perspective from theory and experiment," *Angew. Chem. Int. Ed.* **37** (1998) 868–893.
- [12] J. D. Bryngelson and P. G. Wolynes, "Spin glasses and the statistical mechanics of protein folding," *Proc. Natl. Acad. Sci. USA* **84** (1987) 7524–7528.
- [13] J. N. Onuchic and P. G. Wolynes, "Theory of protein folding," *Curr. Opin. Struct. Biol.* **14** (2004) 70–75.
- [14] J. Slanovec, "Dinamični procesi v magnetno frustriranih sistemih." Fakulteta za matematiko in fiziko, Podiplomski seminar, 2007.
- [15] P. G. Wolynes, "Energy landscapes and solved protein-folding problems," *Phil. Trans. R. Soc. A* **363** (2005) 453–467.