

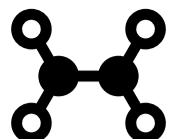
**UNA STANKOVIĆ**  
**BIOHEMIJA ZA**  
**INFORMATIČARE**  
**BELEŠKE SA PREDAVANJA**

<b>Predgovor .....</b>	<b>3</b>
<b>Uvod u aminokiseline.....</b>	<b>4</b>
Klasifikacija aminokiselina po polarnosti.....	4
Prostorna organizacija aminokiselina .....	8
Ekspresija gena I biosinteza proteina .....	9
<b>Peptidne veze.....</b>	<b>12</b>
Osnove termodinamike .....	12
Peptidna veza .....	15
<b>Nivoi strukture proteina.....</b>	<b>18</b>
Određivanje sekvence aminokiselina .....	20
Velike duplikacije .....	23
StrukturaE proteina .....	26
<b>Tercijarna struktura .....</b>	<b>32</b>
Kako se određuje tercijsna struktura?.....	33
Supersekundarne strukture Ili motivi uvijanja.....	36

# PREDGOVOR

Ova skripta predstavlja beleške sa predavanja Biohemija za informatičare na doktorskim studijama Informatike na Matematičkom fakultetu Univerziteta u Beogradu. Skripta nastaje na osnovu predavanja koja drži profesorka Natalija Polović sa Hemijskog fakulteta Univerziteta u Beogradu tokom školske 2019/2020 godine.

Skripta služi kao pomoćno sredstvo pri učenju, verovatno sadrži greške i ne predstavlja nikakav vid zvanične literature za kurs. U slučaju da radoznali čitalac primeti greške može ih prijaviti na mejl [unastankovic1310@gmail.com](mailto:unastankovic1310@gmail.com).



# UVOD U AMINOKISELINE

## KLASIFIKACIJA AMINOKISELINA PO POLARNOSTI

*Aminokiseline* su jedinjenja koja se sastoje iz amino grupe i karboksilne grupe, vodonika i R-ostatka. *R-ostatak* zavisi od konkretnе aminokiseline, varijabilan je, dok se ostatak strukture ne menja. Prikaz navedene strukture se nalazi na slici ispod. Postoji 20 proteinskih aminokiselina, koje mogu biti:

1. Nepolarne

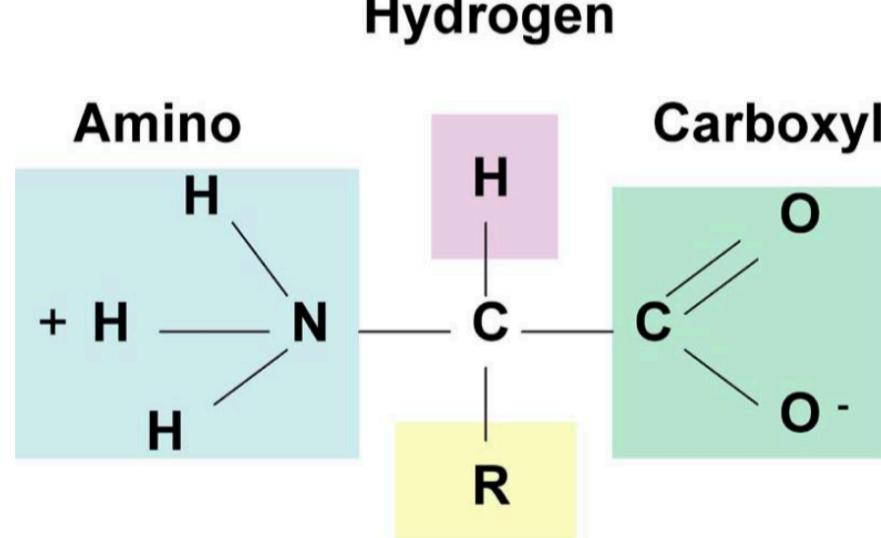
2. Polarne:

A. Nenaelektrisane

B. Naelektrisane:

a. Kisele

b. Bazne



Prikaz 20 aminokiselina

podeljenih prema njihovom

naelektrisanju mogu se videti u

tabeli ispod. Posmatranjem tabele lako se uočava koji deo je zajednički za sve aminokiseline (obeležen crnom bojom), a koji deo predstavlja R-ostatak (obeležen plavom bojom).

### Nepolarne aminokiseline

Narandžastom bojom u tabeli su označene nepolarne aminokiseline. Treba naglasiti da nisu sve aminokiseline podjednako nepolarne, naime, one aminokiseline koje imaju manji R-ostatak su manje nepolarne od onih kod kojih je taj ostatak mnogo veći. Primera radi, alanin je mnogo manje nepolaran od triptofana koji sadrži i aromatični (benzenov) prsten koji je ogroman. Kada posmatranjem tabele uočimo  $CH_3$ ,  $CH$ ,  $CH_2$ , itd. radi se o alkanima. To su

Alanine A	Valine V	Leucine L	Isoleucine I	Proline P
<chem>CC(C)N[C+](C)C(=O)[O-]</chem>	<chem>CC(C)C(C)N[C+](C)C(=O)[O-]</chem>	<chem>CC(C)CC(C)N[C+](C)C(=O)[O-]</chem>	<chem>CC(C)CC(C)C(C)N[C+](C)C(=O)[O-]</chem>	<chem>CC(C)CC(C)C(C)N[C+](C)C(=O)[O-]</chem>
Methionine M	Phenylalanine F	Tryptophan W	Glycine G	Serine S
<chem>CC(C)C(C)C(C)N[C+](C)C(=O)[O-]</chem>	<chem>CC(C)C(C)C(C)N[C+](C)C(=O)[O-]</chem>	<chem>CC(C)C(C)C(C)N[C+](C)C(=O)[O-]</chem>	<chem>CC(C)N[C+](C)C(=O)[O-]</chem>	<chem>CC(C)C(C)C(C)N[C+](C)C(=O)[O-]</chem>
Threonine T	Cysteine C	Asparagine N	Glutamine Q	Tyrosine Y
<chem>CC(C)C(C)C(O)N[C+](C)C(=O)[O-]</chem>	<chem>CC(C)C(C)C(S)N[C+](C)C(=O)[O-]</chem>	<chem>CC(C)C(C)C(=O)N[C+](C)C(=O)[O-]</chem>	<chem>CC(C)C(C)C(C)C(N)C(=O)N[C+](C)C(=O)[O-]</chem>	<chem>CC(C)C(C)C(C)C(O)N[C+](C)C(=O)[O-]</chem>
Aspartic Acid D	Glutamic Acid E	Lysine K	Arginine R	Histidine H
<chem>CC(C)C(C)C(=O)N[C+](C)C(=O)[O-]</chem>	<chem>CC(C)C(C)C(=O)N[C+](C)C(=O)[O-]</chem>	<chem>CC(C)C(C)C(C)C(N)C(=O)N[C+](C)C(=O)[O-]</chem>	<chem>CC(C)C(C)C(C)C(N)C(=O)N[C+](C)C(=O)[O-]</chem>	<chem>CC(C)C(C)C(C)C(C)C(N)C(=O)N[C+](C)C(=O)[O-]</chem>

Nepolarne

Polarne  
nenaelektrisane

Polarne  
naelektrisane:

kisele  
bazne

npr. benzin, ili parafin (od koga se prave sveće), čija je jedna od glavnih karakteristika da se ne mogu pomešati sa vodom, zbog čega se smatraju nepolarnim (voda je polarna). Kažemo i da su **hidrofobni** - "beže" od vode (ali ne istom brzinom). Brzina "bežanja" od vode zavisi od voluminoznosti (veličine) aminokiseline. Uzmimo za primer alanin, koji je jako mali, voda sa njime, zbog njegove veličine, ima manji problem, tako da ga sporije istiskuje. Dok, sa druge strane, leucin je ogroman i voda sa njime ima veći problem. Isto važi i za fenilalanin, jer je svaki ugao u benzenovom prstenu *CH*.

Iz tablice važno je posmatrati dva parametra:

1. Koliki je "repić" (R-ostatak) - posmatranjem golim okom
2. Koliko se "repić" i voda slažu međusobno - podela je izvršena po bojama

### Polarne nenaelektrisane aminokiseline

U drugoj kategoriji se nalaze polarne nenaelektrisane aminokiseline. One nemaju ni plus ni minus, ali imaju *OH*, zbog čega ih voda posmatra kao slične i ne smetaju joj. Kod tirozina, treba obratiti pažnju na postojanje prstena.

## Polarne naelektrisane aminokiseline

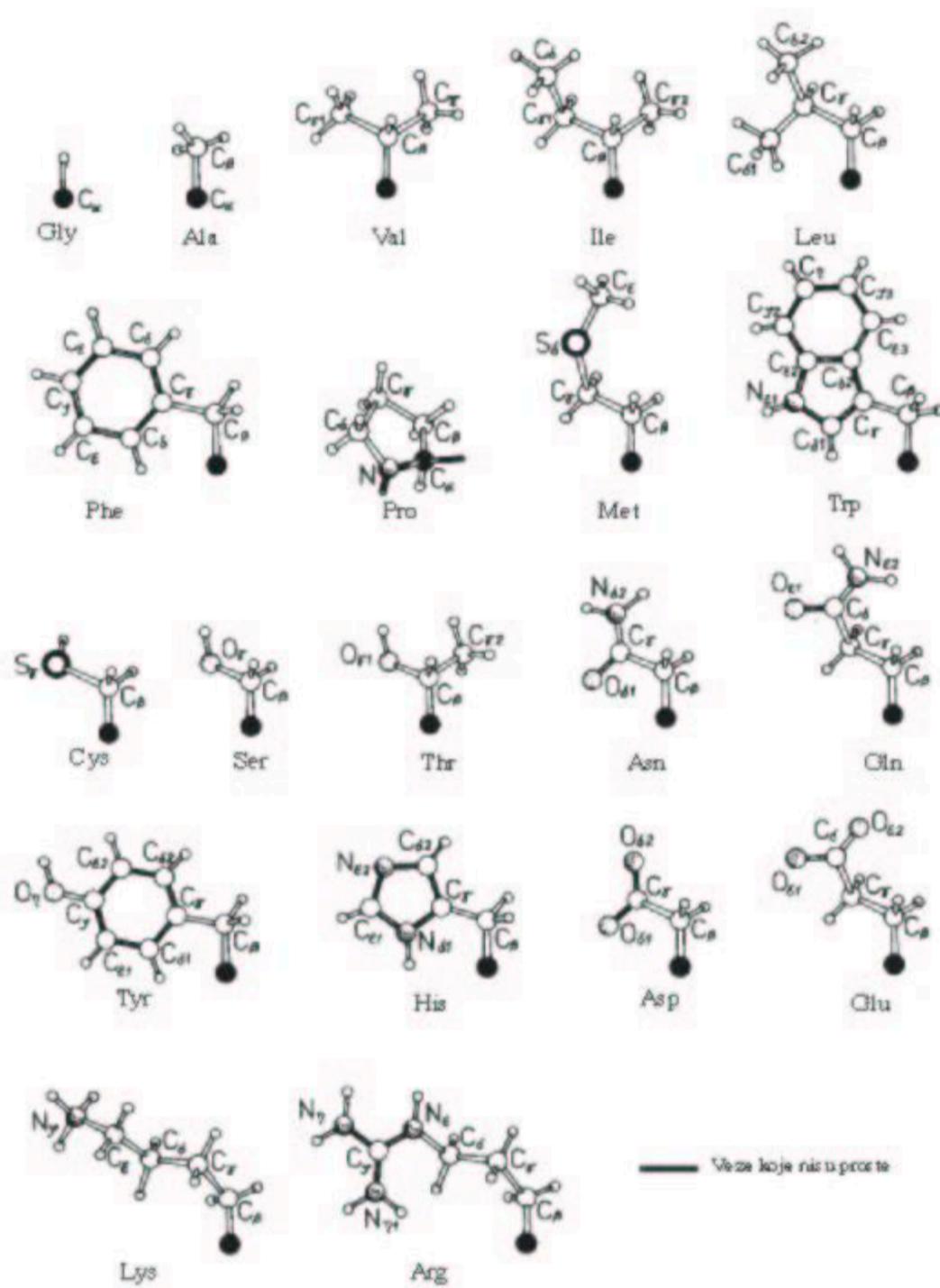
U poslednjoj grupi se nalaze polarne naelektrisane aminokiseline. One u vodi disosuju na  $H^+$  ion koji ode i ostaje  $COO^-$ , ponašajući se kao soli. Natrijum hlorid kad se rastvori u vodi disosuje na  $Na^+$  i hloridni ion  $Cl^-$ . Veoma je važno povezati osobine (hidrofobnost) i nazive aminokiselina.

U narednoj tabeli mogu se videti puni nazivi, troslojne skraćenice i jednoslovne skraćenice naziva aminokiselina. Smatralo se da je tri slova dovoljno za skraćenice, međutim, razvojem bioinformatike uvidelo se da to ipak nije idealno tako da su uvedene jednoslovne skraćenice.

Tabela 2.2: Skraćenice za 20 standardnih aminokiselina. Simboli Glx i Asx označavaju Gln ili Glu odn. Asn ili Asp. Ovi simboli potiču iz laboratorijske prakse, jer pri određivanju aminokiselinskog sastava i sekvene često nije moguće sa sigurnošću utvrditi da li se radi o amidima (Asn, Gln) ili o odgovarajućim kiselinama (Asp, Glu).

Aminokiselina ili ostatak	Troslojni simbol	Jednoslovni simbol	Mnemonik za jednoslovni simbol
Alanin	Ala	A	Alanin
Glutamat	Glu	E	od engl. "gluEtamic acid"
Glutamin	Gln	Q	od engl. "Q-tamine"
Aspartat	Asp	D	od engl. "asparDic acid"
Asparagin	Asn	N	asparagiN
Leucin	Leu	L	Leucin
Glicin	Gly	G	Glicin
Lizin	Lys	K	K je u abuci ispred L
Serin	Ser	S	Serin
Valin	Val	V	Valin
Arginin	Arg	R	aRginin
Treonin	Thr	T	Treonin
Prolin	Pro	P	Prolin
Izoleucin	Ile	I	Izoleucin
Metionin	Met	M	Metioni
Fenilalanin	Phe	F	Fenilalanin
Tirozin	Tyr	Y	engl. tYrosine
Cistein	Cys	C	Cistein
Triptofan	Trp	W	engl. "tWo rings"
Histidin	His	H	Histidin

Kada se posmatra bočni ostatak treba imati na umu **polarnost, veličinu bočnog ostatka i disocijaciju**. Disosovane i nedisosovane imaju ili nemaju plus i minus. Na narednoj slici mogu se videti strukture bočnih R-ostataka.

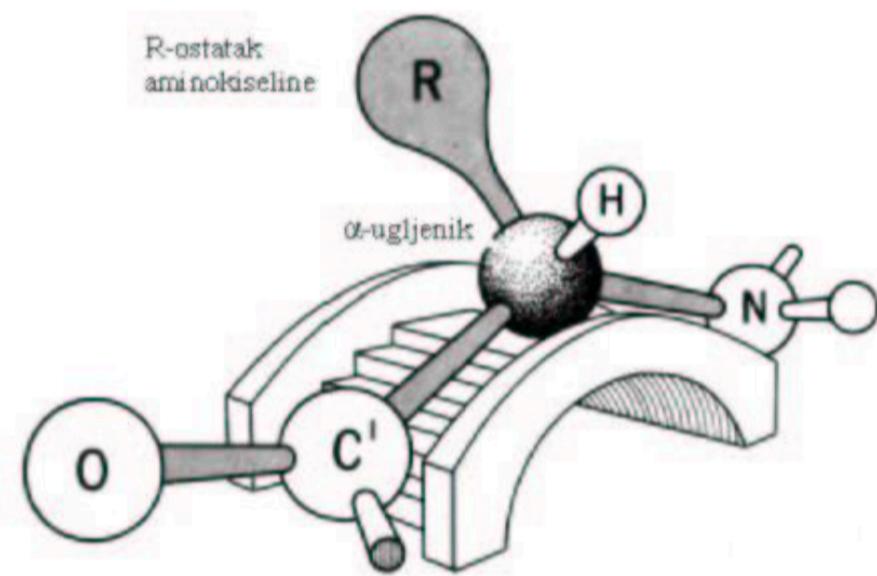


Slika 2.2: Strukture bočnih (R) ostataka standardnih aminokiselina; bočni ostaci su grupisani po sličnosti.

Kada su kiseline u vodenoj sredini menja im se struktura. Kiselina u vodenoj sredini oslobađa  $H^+$  jone i daje anjon i katjon. Baze, sa druge strane, primaju  $H^+$  jone. Oni koji disosuju na fiziološkom pH (7.4) neke aminokiseline mogu postati pozitivno nanelektrisane, a neke negativno.

## PROSTORNA ORGANIZACIJA AMINOKISELINA

Prostorna organizacija aminokiselina je u vidu tetraedra. R-ostatak bi trebalo da je potpuno svejedno sa koje je strane, međutim, evolutivno, u izgradnju proteina su ušle samo one aminokiseline kod kojih je R-ostatak sa leve strane. Ovakve aminokiseline nazivamo alfa-aminokiseline (L - aminokiseline). U prirodi nastaju i desne i leve aminokiseline i one se odnose kao lik u ogledalu. Prostorno to nije isti molekul, ali ponašanje mu je identično - samo je geometrijska organizacija drugačija.

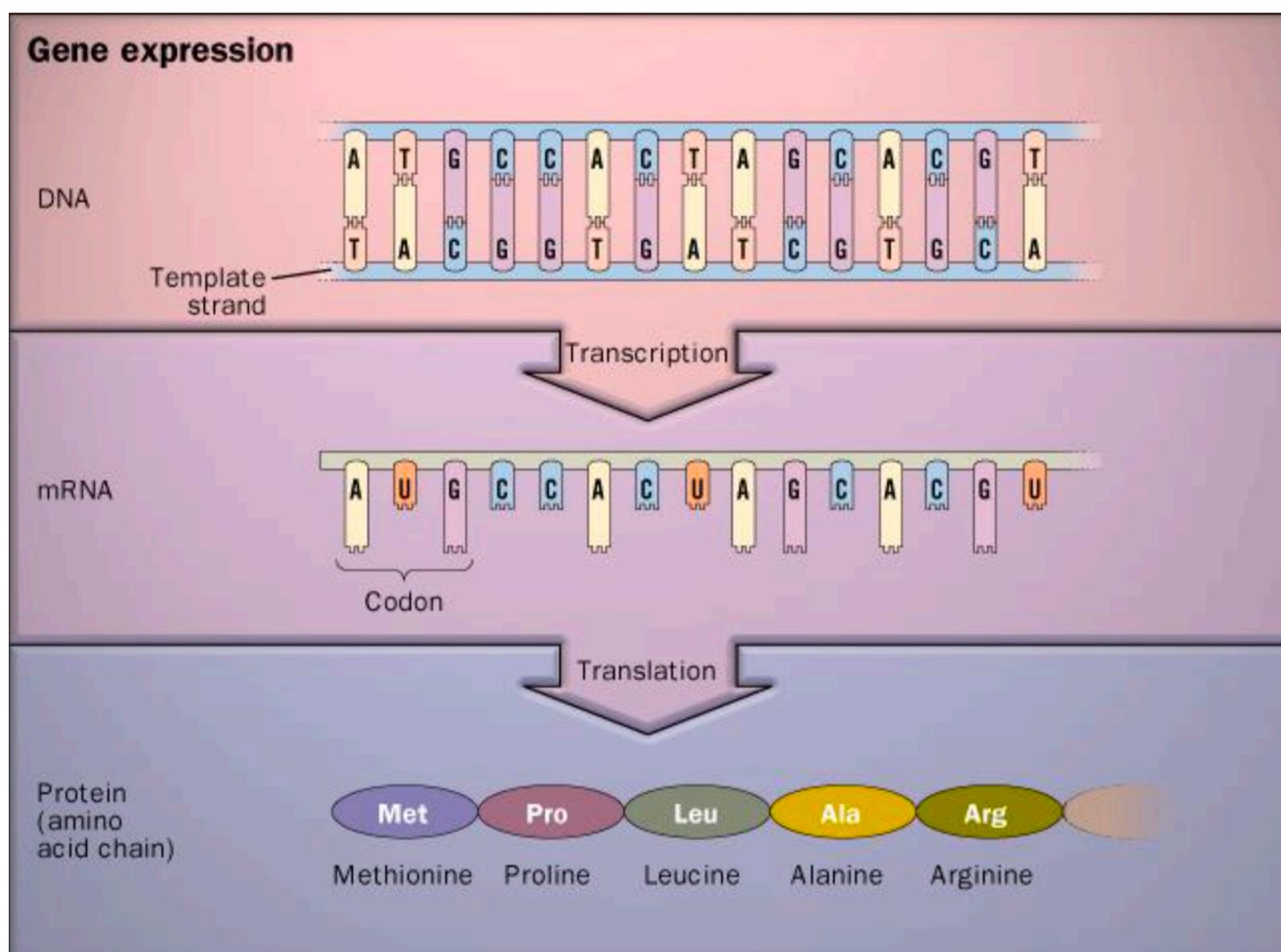


Slika 2.6: Mnemonik za prepoznavanje L-aminokiselina: Ako namestimo aminokiselinu tako da se čita "ON" (na engleskom  $\approx$  napred) R ostatak će biti sa leve strane mosta.

# EKSPRESIJA GENA I BIOSINTEZA PROTEINA

## Zašto baš 20 aminokiselina?

U proteinima se nalazi baš 20 aminokiselina, pitanje je zašto baš ovih dvadeset, a ne nekih drugih 20?



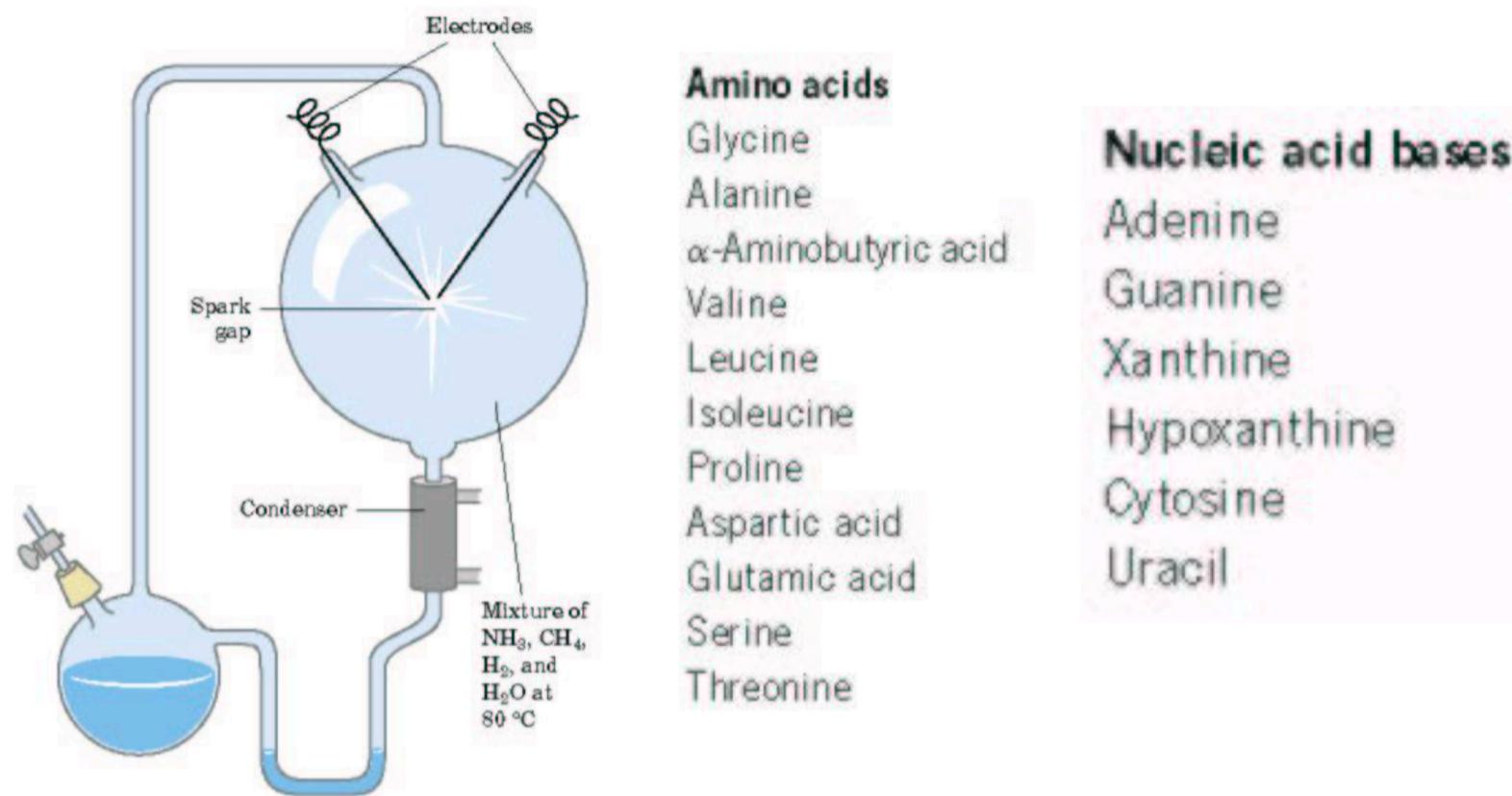
DNK je nosilac genetske informacije i sastoji se iz dva lanca. Svaki od lanaca sačinjen je iz nukleotida, odnosno 4 slova: A, C, G i T. Kako se informacije sa jednog lanca - templatnog (engl. Template - uzorak), prepisuju na informacionu RNK (engl. Messenger), po sistemu komplementarnosti T - A (U - A, kod RNK), C - G, ovim postupkom nastaje jednolančana RNK koja ide na ribozome. Ona služi da bi u procesu translacije svaki triplet nukleotida - kodon - bio prepoznat antikodonom sa transportne RNK i on odgovara jednoj aminokiselini.

Tablica 2.8: Genetski kod aminokiselina: kodoni predstavljaju sekvence nukleotidnih triplata u i-DNK. Prvi nukleotid u tripletu naveden je sa leve strane, drugi gore, a treći sa desne strane tablice.

	U	C	A	G	
U	Phe Phe Leu Leu	Ser Ser Ser Ser	Tyr Tyr End End	Cys Cys End Trp	U C A G
C	Leu Leu Leu Leu	Pro Pro Pro Pro	His His Gln Gln	Arg Arg Arg Arg	U C A G
A	Ile Ile Ile Met	Thr Thr Thr Thr	Asn Asn Lys Lys	Ser Ser Arg Arg	U C A G
G	Val Val Val Val	Ala Ala Ala Ala	Asp Asp Glu Glu	Gly Gly Gly Gly	U C A G

U tablici iznad, može se videti koje troslovne skraćenice odgovaraju kojim kombinacijama slova. Čitanje se vrši levo - gore - desno. Na primer, U - U - U je fenilalanin, U - U - A je leucin, itd. Ovaj genetski kod je identičan za sve što je živo na Zemlji što znači da je genetski kod **univerzalan**. Univerzalnost koda govori da su svi potekli od istog pretka, od iste praćelije od koje je sve poteklo i menjalo se evolucijom. Gde god se najde na genetsku sekvencu ona se može prevesti u polipeptidnu korišćenjem ove tablice.

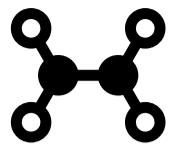
Druga osobina koda je da je kod **degenerisan**. Mogućih kombinacija triplata je  $4^3 = 64$ , a imamo 20 aminokiselina - ima više triplata za jednu aminokiselinu što se može videti i u tablici. Aminokiseline sa komplikovanim formulama su evolutivno ušle kasnije, pred kraj evolucije i imaju jedan triplet. Često se dešava da jedna aminokiselina ima po četiri kombinacije i to takve da su prve dve pozicije identične, a da se na trećoj može nalaziti bilo šta - varijabilna je. Na primer, za kombinaciju A - C - X, gde je X bilo koji od G, C, A, U dobija se treonin. Ta osobina da postoji više od jedne mogućnosti za neku aminokiselinu se naziva **degeneracija koda**, ona omogućava da kada se dešavaju mutacije na DNK (izmena slova) promene ne utiču na to koja aminokiselina se pojavljuje u kodu. Odnosno, da se ta mutacija *atenuira - utišava*. Tada postoji mutacija na nivou DNK, ali ne i na nivou proteina.



Još je ostalo da se odgovori na pitanje zašto baš tih 20 aminokiselina, a ne nekih drugih. Naime, kada je nastajala Zemlja postojalo je more i anaerobna atmosfera, munje su udarale u to more, bilo je mnogo toplije nego sad, itd. Isti uslovi su simulirani eksperimentalno korišćenjem elemenata za koje se smatra da su tada bili dostupni na Zemlji, to su uglavnom neki gasovi amonijak, metan, vodonik, itd. Zagrevanjem pare vode i korišćenjem elektroda koje simuliraju udare munja i ostavljanjem neko vreme, sintetišu se baš ove aminokiseline koje su ušle u izgradnju proteina, a ne ostale koje bi mogle teorijski da postoje. Tada su pronađeni baš oni nukleotidi koji su ušli u izgradnju nukleinskih kiselina.

### Zašto su L, a ne D-aminokiseline?

Zašto ne mogu da budu malo L, malo D-aminokiseline? To bi bilo kao da se kružne stepenice prave malo od levih stepenika, malo od desnih. Prepostavlja se da je L odabранo slučajno, a da je kasnije samo nastavljeno. Moguće je napraviti protein samo od D-aminokiselina i on izgleda kao lik u ogledalu u odnosu na prirodni protein, imaće i slične osobine.



## PEPTIDNE VEZE

### OSNOVE TERMODINAMIKE

Termodinamika ima nekoliko važnih formula Gibbs-Helmholtzova jednačina:

$$\Delta G = \Delta H - T * \Delta S$$

gde je  $\Delta G$  promena slobodne energije,  $\Delta H$  promena entalpije,  $T$  je temperatura u Kelvinima i  $\Delta S$  je promena entropije.

Ako neki proces može da se desi sam od sebe, onda se smatra da je taj proces **spontan**, bez uzimanja u obzir vremenske skale kao parametra. Samo se posmatraju početno i krajnje stanje.

Ako je krajnje stanje u nekom procesu stabilnije od početnog stanja, onda se taj proces smatra spontanim i posmatra se kao proces koji će se desiti sam od sebe, nezavisno koliko je vremena potrebno da se dođe od početnog do krajnjeg stanja. Zbog toga i govorimo u terminima  $\Delta G$  - promenama slobodne energije između početnog i krajnjeg stanja.  $G$ , kao takva, ako i postoji nama je nedostižna da razumemo, ali kada posmatramo neki proces možemo bez problema da razumemo  $\Delta G$ , pa čak i izmerimo. Neki proces je spontan ako je promena slobodne energije za taj proces manja od 0, odnosno, ako se sistem stabilizuje.

### Na koji sve način neki proces može biti spontan?

Može biti spontan zbog entalpije ili zbog entropije. Neophodno je da  $\Delta G$  bude manje od 0 kako bi proces bio spontan. Proces može biti spontan ako je  $\Delta H$  manje od 0 ili ako je  $T * \Delta S$  veće od 0 (jer ima minus kao predznak).

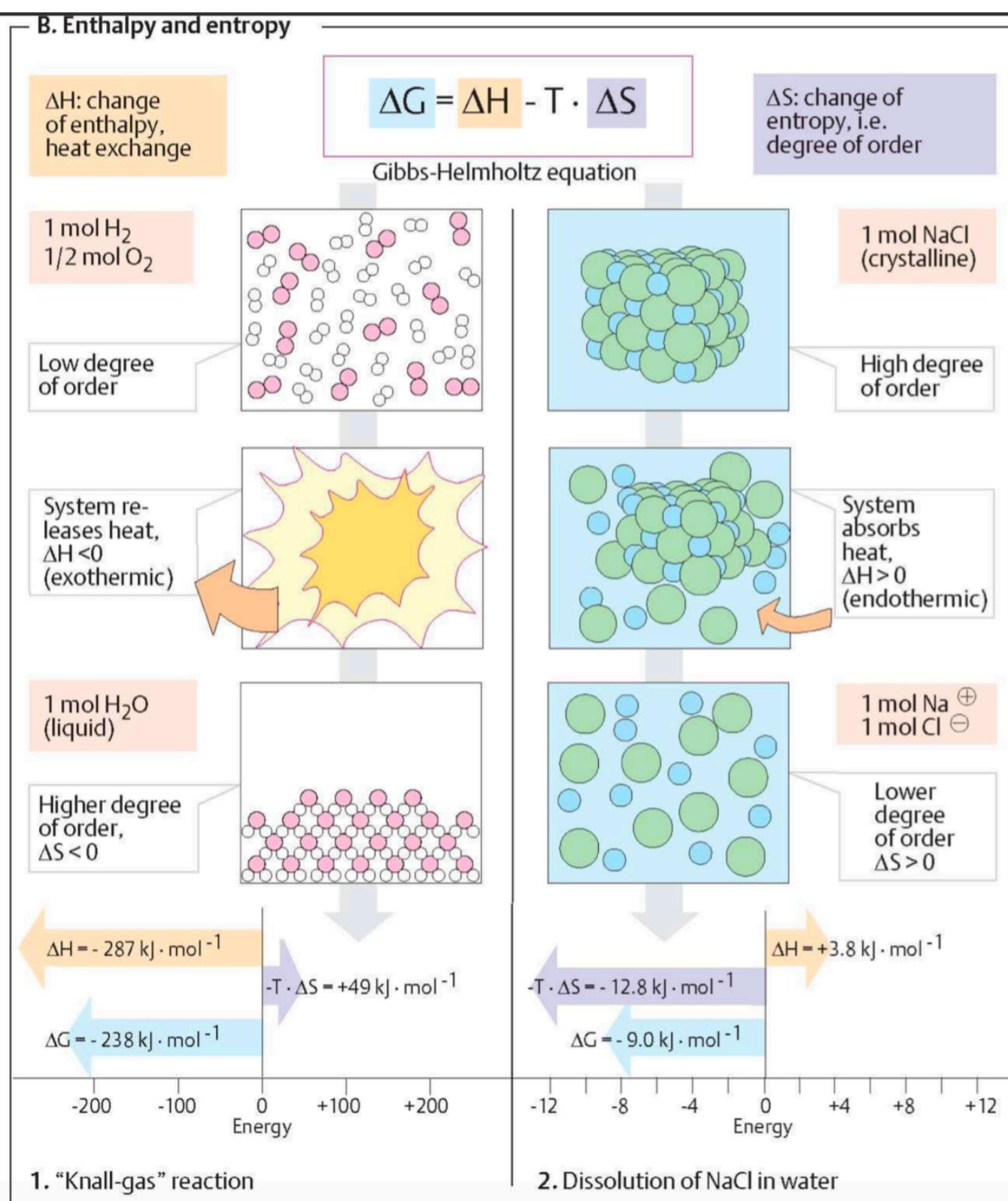
$\Delta H$  je promena entalpije i govori o jačini interakcija između neke dve stvari u procesu. Posmatraju se interakcije na početku i na kraju, na primer između A i B. Na početku je

$A + B$  i tu nema interakcije, jer su odvojeni, a na kraju je  $A * B$ , što znači da interaguju ( $\Delta H$  je manje od 0).

Drugi zakon termodinamike veoma pojednostavljen glasi: **“Entropija u svemiru raste (Entropija uvek raste)”**. Entropija je količina haosa, neuređenost sistema i zavisi od temperature. Svaki sistem, u ma kojoj dimenziji svemira, uvek teži ka povećanju neuređenosti. Na primer, ako bismo uzeli obojeni gas i pustili ga u nekom delu prostorije, on neće ostati zarobljen u tom delu gde je pušten, već će težiti da se raširi po celoj prostoriji.

Neuređenost zavisi od temperature - množi se temperaturom. Što je nešto bliže absolutnoj nuli Kelvina kretnje se smanjuju, pa se zbog toga entropija ne može povećavati. Što je temperatura veća to entropija više raste.

Ako se posmatraju vodonik i kiseonik, oba su gasovi koji reaguju dajući vodu. Vratimo se na primer sa eksperimentom (na slici ispod leva polovina). Tokom praistorije oni su kao gasovi



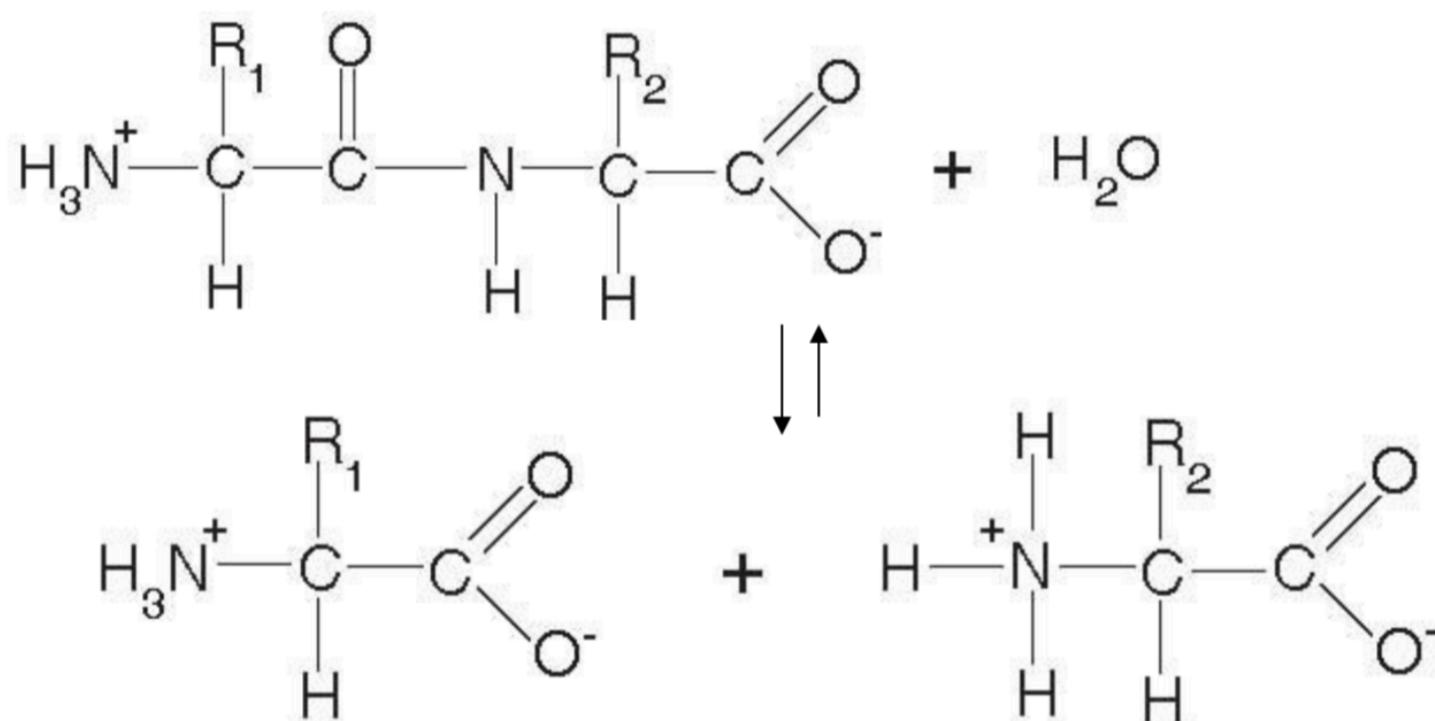
slobodno šetali sudom pre nego što se desila reakcija. Entropija je pre reakcije bila ogromna, a nakon reakcije, nakon što se molekuli poslažu u strukturu, mnogo manja. Sa aspekta entropije proces nastajanja vode je spontan, jer je entalpija pobedila entropiju. To se može videti i na grafiku ispod,  $\Delta H$  i  $\Delta G$  "vuku" na negativnu stranu, dok entropija ( $-T * \Delta S$ ) "vuče" na pozitivnu stranu. Zbog toga što je  $\Delta G$  manje od nule, zaključuje se da je proces spontan.

Posmatrajmo drugi način da proces bude spontan, a to je da entropija pobedi nepovoljnu entalpiju. Posmatrajmo drugu polovicu gornje slike. Kod kristala natrijum-hlorida, joni natrijuma (plave boje) i joni hlorata (zelene boje) grade veoma jake jonske veze. Kad se rastvore u vodi grade slabe interakcije, ali ne grade međusobno jonske veze. Što se entalpije tiče ovaj proces gde su interakcije jače joj više odgovara. Entropija nakon reakcije dosta raste, pa posmatranjem grafika ispod reakcija može se videti da se entalpija promenila nepovoljno, a entropija promenila povoljno i pobedila entalpiju. Proces rastvaranja kristalne rešetke neke soli u vodi je spontan zato što se povećava entropija sistema.

Svaki proces se može posmatrati i kroz ne-hemijski aspekt. Ako bi kuća stajala na nekoj planini, i nju niko ne popravlja, a izložena je raznim spoljašnjim faktorima: pada kiša, duva vetar itd., ne uzimajući vremensku skalu u obzir, ta kuća će vremenom promeniti svoj oblik. Početno stanje je kuća, a završno stanje je pojedinačno kamenje na nekoj livadi. Entropija je porasla, jer su se pojedinačne cigle rastavile, a entalpija je povećana na početku, jer postoji cement među ciglama koji se vremenom razgradio.

## PEPTIDNA VEZA

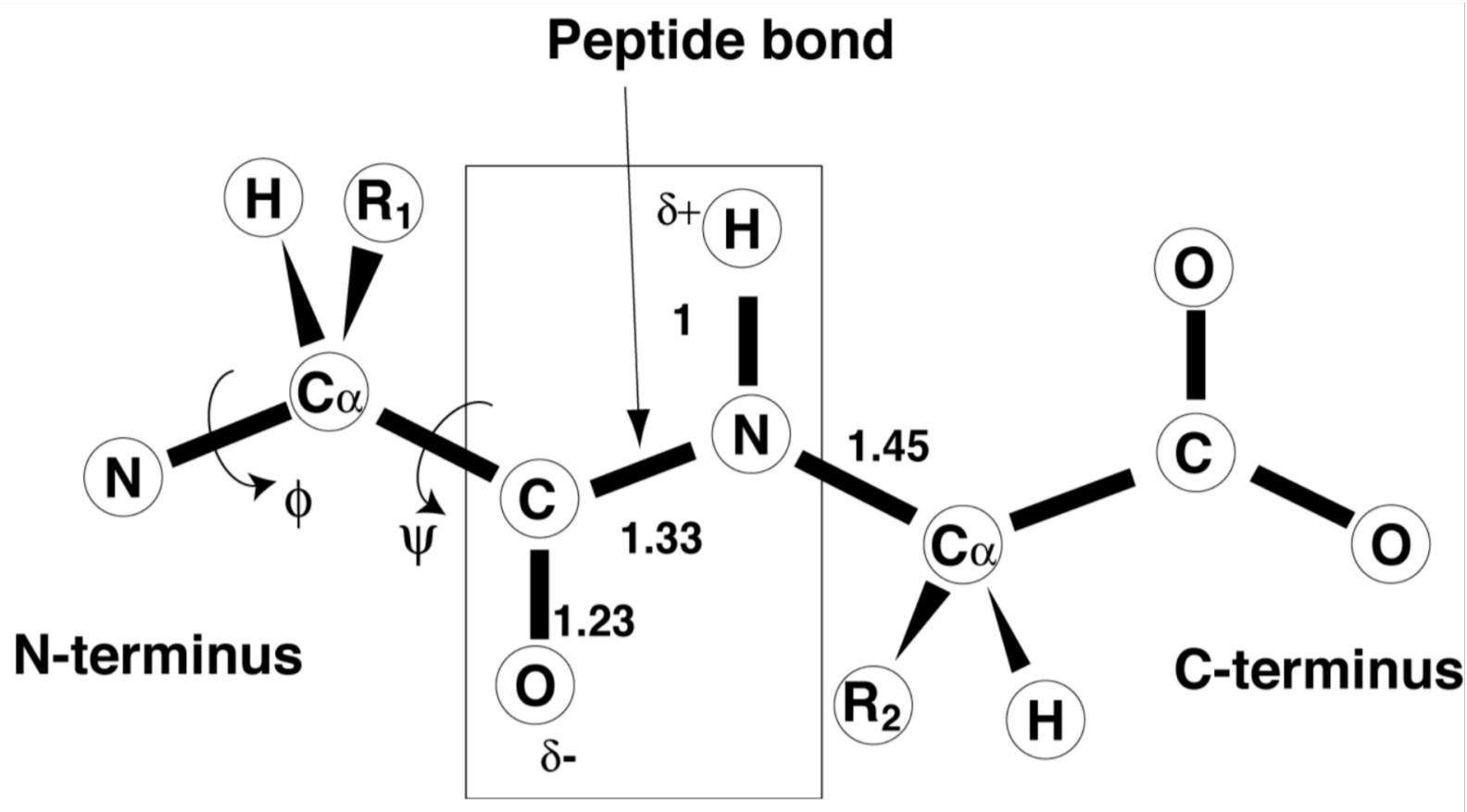
Na narednoj slici nalaze se dve aminokiseline povezane peptidnom vezom. Ostatak prve i druge aminokiseline daju vodu, a veza koja se stvara između C i N naziva se peptidnom vezom. Peptidna veza u vodenoj sredini može da ishidrolizuje na pojedinačne aminokiseline, dok se voda kao rastvarač ne uzima u obzir. Entropija vuče ovaj proces ka hidrolizi peptidne veze. Na osnovu termodinamike, proteini ne bi smeli da postoje u vodenoj sredini, ali oni ipak postoje, jer je ta reakcija beskrajno spora. Naš životni vek je duplo kraći od hidrolize peptidne veze. Dakle, proces hidrolize je spontan, samo je jako spor.



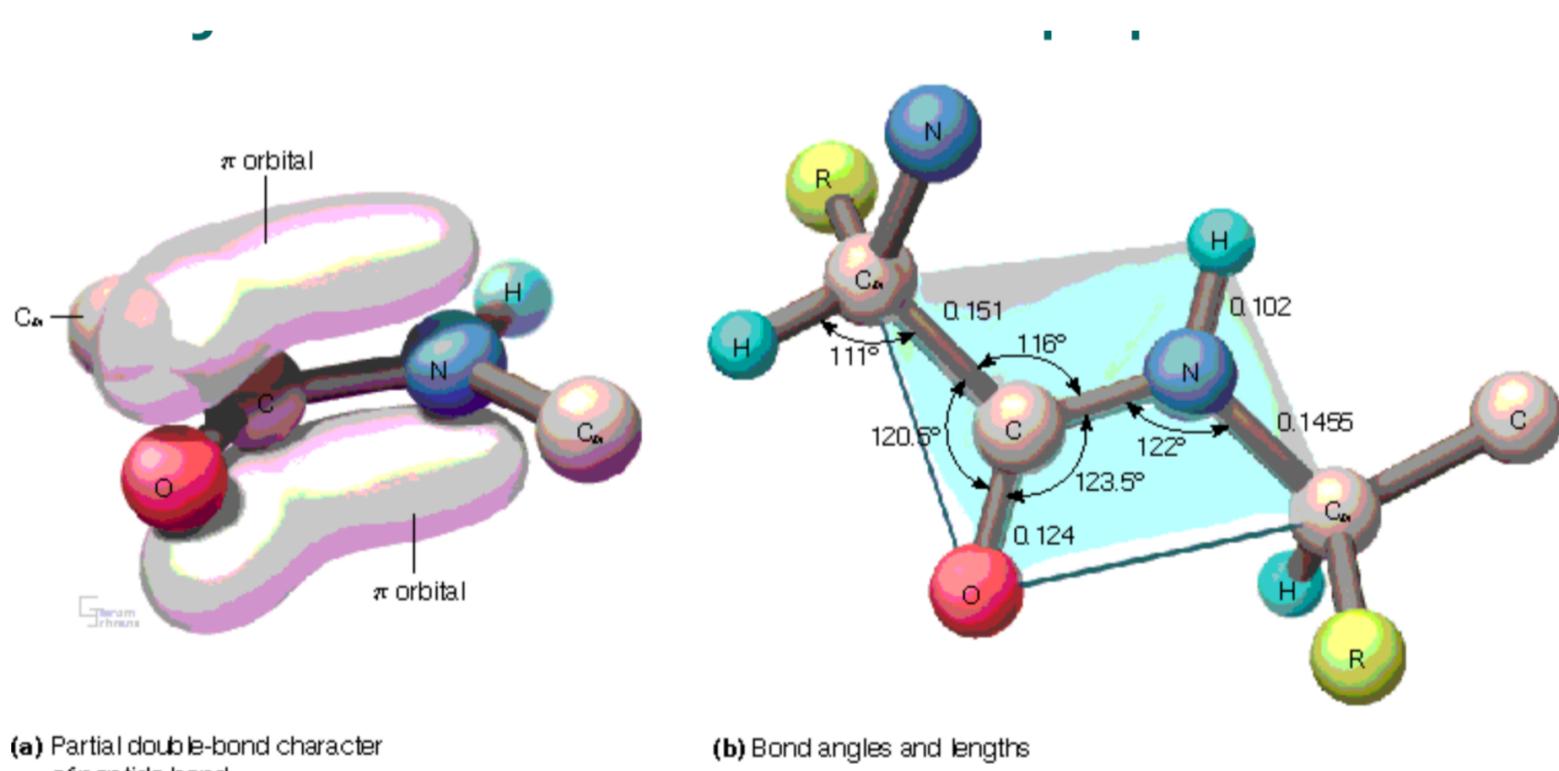
Kad god je neki proces spontan on se može ubrzati. Sinteza peptidne veze mora da bude jako komplikovan i energetski skup proces za ćeliju, jer pojedinačne aminokiseline nikad neće dati peptidnu vezu. Proces je energetski skup za ćeliju u smislu da je potrebno mnogo učesnika kao što su ribozom, transportna RNK, itd. što sve zajedno troši mnogo energetske valute ćelije - ATP-a. Svi naši enzimi zaslužni su za ubrzavanje reakcija, na primer pri varenju (razlaganje mesa, odnosno proteina). Sinteza je skupa i komplikovana.

Reakcija hidrolize je pomerena u desno, jer raste entropija. Kako opstaje? Reakcija je spora. Peptidna veza je termodinamički nestabilna, volela bi da hidrolizuje, ali je kinetički inertna, sporo se pokreće. Postoje načini da se reakcija ubrza - katalizuje. Na primer, čišćenje odvoda se vrši prahom koji predstavlja jaku bazu, nakon 15 minuta (koliko je potrebno da izhidrolizuje) pušta se voda i proces je završen.

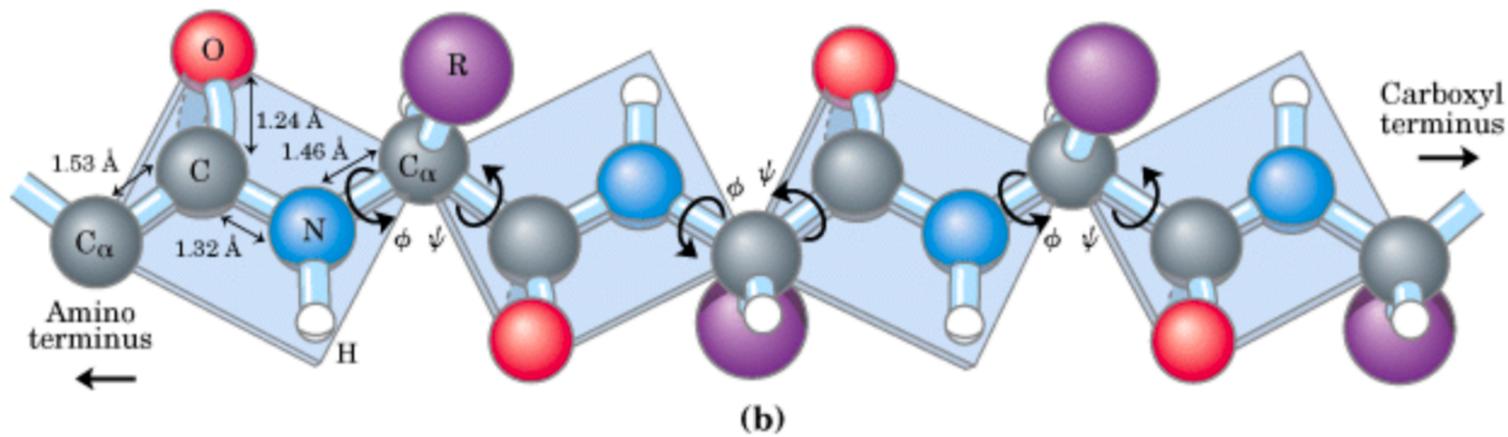
## Strukturni aspekti peptidne veze



Po konvenciji protein počinje *N* krajem. Niz  $N - C\alpha - C - N - C\alpha - C - O$  predstavlja polipeptidnu kičmu. Na  $C\alpha$  atomu biće  $R$  ostatak. Peptidnom vezom smatramo  $C - N$  i supstituente na njima, na slici oni su uokvireni. Peptidna veza ne dozvoljava obrtanje jer na azotu postoji elektronski par i taj par sa gornjim i donjim elektronima pravi oblak koji onemogućava rotaciju.

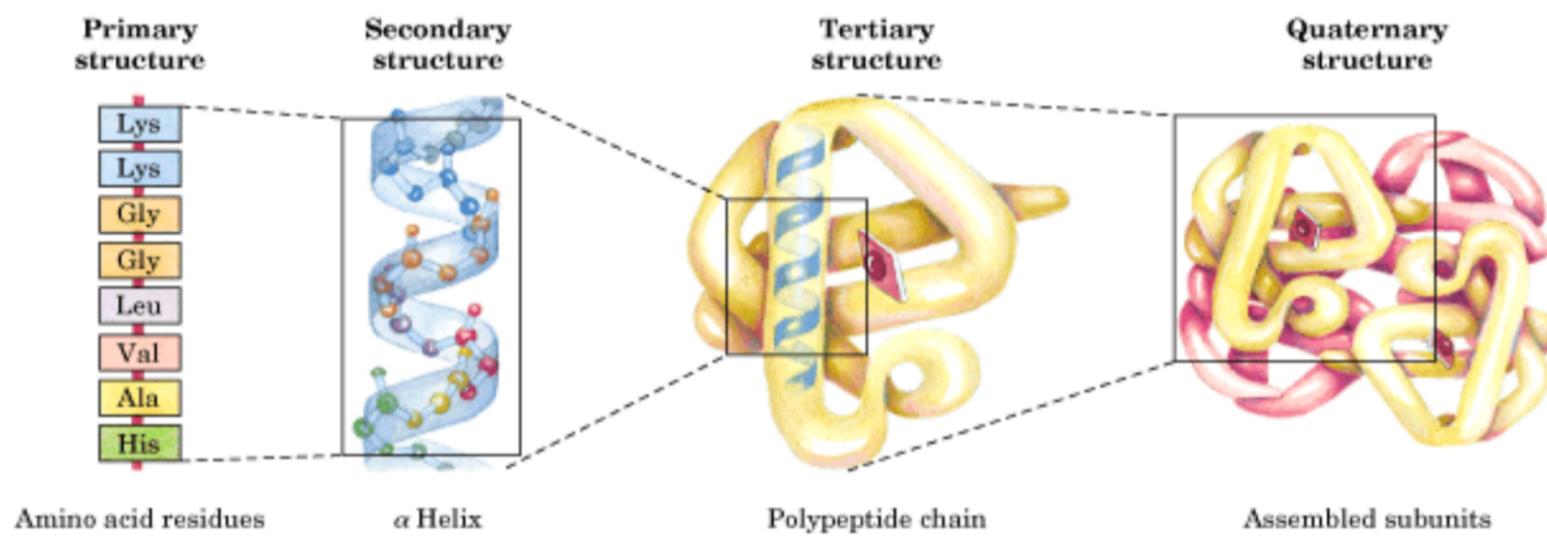


Za peptidnu vezu kažemo da ima **delimično dvostruki karakter**. To znači da oko cele peptidne veze ova 4 atoma će morati da budu u ravni i oni su zapravo organizovani kao na nekoj pločici. Kaže se da je peptidna veza **planarna**, što je bitno za naredna gužvanja polipeptidnog niza prilikom zauzimanja viših nivoa struktura. Polipeptidni niz ne može da se gužva na bilo koji način, već se poput ogrlice sa pločicama uvija gde jedino između tih pločica može biti zaokretanja.



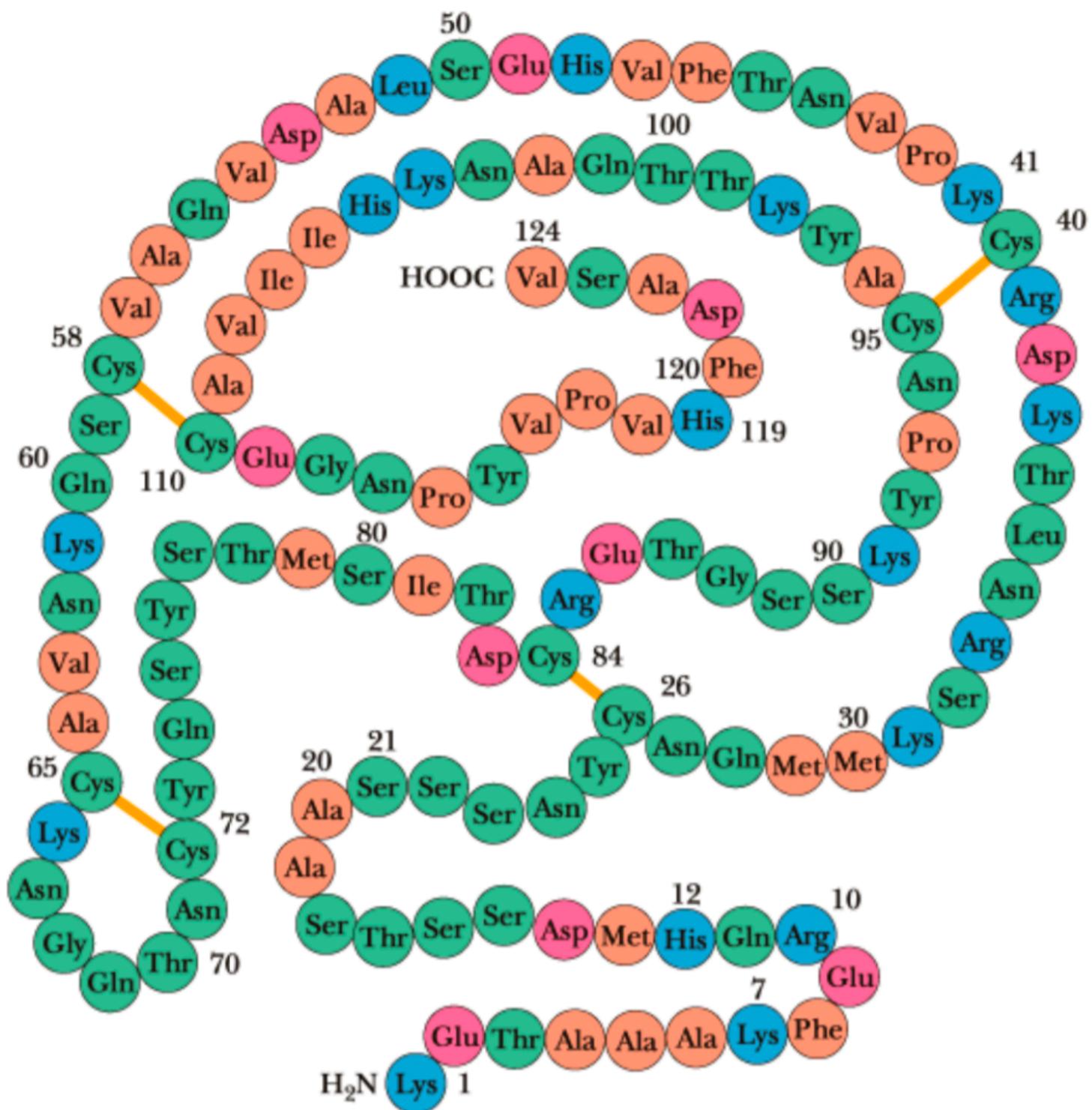
# NIVOI STRUKTURE PROTEINA

Primarna struktura je sama sekvenca koja čini protein. Sekundarna će biti objašnjena detaljnije u nastavku. Tercijarna je trodimenzionalni prikaz strukture proteina, odnosno konformacija. Kvaternarna je tercijarnih struktura zajedno kako bi se dobilo neko više funkcionalno svojstvo.



Primarna struktura se naziva još i kovalentna struktura, ne uzimaju se u obzir nikakve interakcije već samo kojim redosledom su aminokiseline povezane peptidnom vezom. Taj redosled je kodiran na genu. U primarnu strukturu, pored samog redosleda aminokiselina koji je genski determinisan, ulazi i položaj disulfidnih mostova.

Primer: Na slici ispod prikazana je jedna sekvenca aminokiselina. Posmatranjem cisteina  $R - SH$ , gde se  $SH$  naziva i *tiolna grupa* koja može sa drugim cisteinom reagovati uz kiseonik iz vazduha i daje vodu - u pitanju je oksidoredukcija. Ovom reakcijom dobija se  $R - S - S - R'$  koji nazivamo *cistin*. Veza (disulfidni most) koja se stvara je kovalentna i veoma jaka i nastaje reakcijom dva bočna niza iz dve aminokiseline cisteina. U proteinu na slici postoji četiri disulfidna mosta, koja povezuju osam cisteina: 26. je vezan sa 84., 26. nije vezan sa 65., itd. Pravilan položaj disulfidnih mostova nije genski kodiran, već je kodiran uvijanjem, ali nije jasno kako. Ono što se zna je da kada se polipeptidni niz adekvatno uvije dva cisteina će doći u dovoljnu blizinu da sa kiseonikom odreaguju dajući disulfidni most. Cistein je svrstan u polarne aminokiseline, a cistin je ekstremno nepolaran, jer su se izgubili  $H$ .



## ODREĐIVANJE SEKVENCE AMINOKISELINA

Podaci koji se nalaze u bazama podataka su dobijeni sekvenciranjem proteina ili sekvenciranjem gena. U UniProt bazi podataka svetlijom zvezdicom su označeni tzv. *hipotetički proteini* za koje se ne zna ništa na realnom nivou. Za ove proteine se tokom sekvenciranja genoma uvidelo da postoji tzv. *otvoreni okvir čitanja* - gen, pa postoji genetski potencijal da se napravi protein te sekvence. Za njih se može proveravati da li su srodnii nečemu upoređivanjem sekvenci na različite načine, ali oni nikad nisu detektovani. U UniProt bazi postoji i zlatna zvezdica kojom su označeni proteini koji su detektovani na eksperimentalnom nivou i za njih postoji validna literatura sa objašnjenjima uloge, itd. Frederik Sanger je prvi sekvencirao proteine ikad.

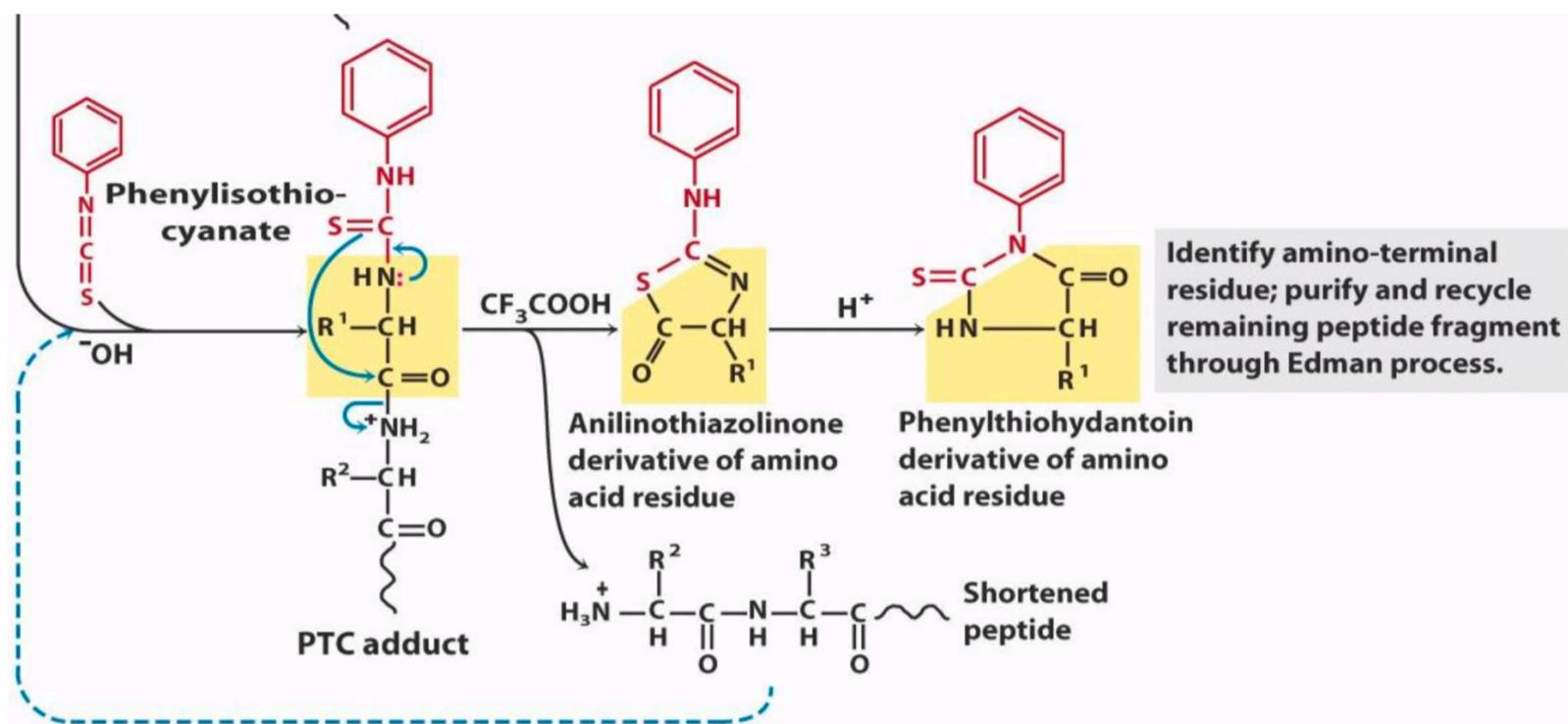
Određivanje aminokiselinske sekvence u 8 koraka:

1. Odredi/identifikuj N- i C- terminalne aminokiseline,
2. Ako protein ima više nizova razdvojih ih,
3. Redukuj disulfidne mostove i zaštititi SH grupe,
4. Odredi aminokiselinski sastav svakog niza,
5. Sekvenciraj niz sa N-kraja,
6. Fragmentiši niz enzimatski i/ili hemijski i sekvenciraj svaki od fragmenata,
7. Rekonstituiši fragmente,
8. Odredi položaj disulfidnih mostova.

Iz navedenog opisa određivanja sekvence vidi se koliko je taj postupak eksperimentalno zahtevan i ograničava razlikovanje asparagina i glutamina od aspartata i glutamata zbog nesavršenosti procesa. Svaki od eksperimentalnih procesa koji su punili neku bazu podataka ima svoja ograničenja.

Jako je teško tokom sekvenciranja odrediti položaj disulfidnih mostova, čak i eksperimentalno. **Edmanova degradacija** je proces koji je jednostavniji i ima mnogo manje grešaka (pravi razliku između asparagina i glutamina i aspartata i glutamata), ali ni on nije savršen - ograničen je na samo kraće peptidne segmente (do 20-30 aminokiselina). To znači da bi ceo niz morao da se fragmentiše na dovoljno male fragmente da svaki od njih pojedinačno bude sekvenciran. Potom, javlja se problem redosleda segmenata, mora se isti

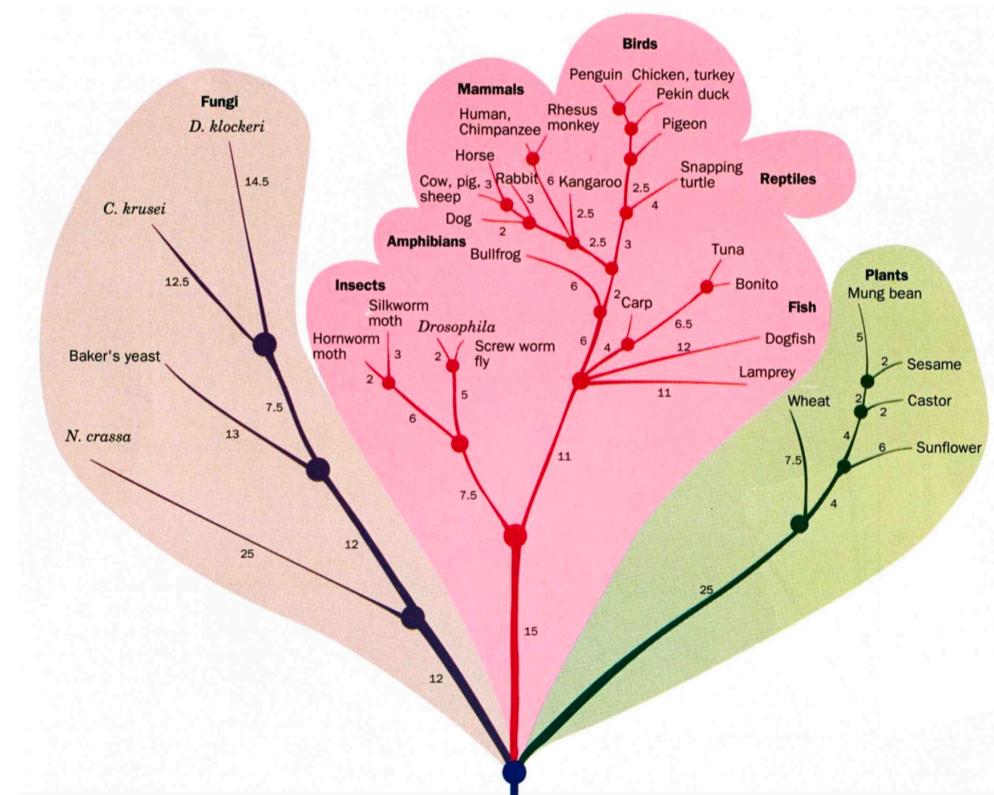
uzorak prelići u dve čašice i fragmentisati na dva različita načina, pa sekvencirati i jedne i druge fragmente i onda u programu tražiti preklapajuće fragmente kako bi mogli da se poređaju. Često se zbog toga sirove niske u bazama podataka nalaze u dva rezitorijuma, što znači da nije urađeno asembliranje fragmenata.



Primarna struktura je bitna jer određuje više nivo struktura. Zna se da aminokiselinska sekvenca sama zna kako da se uvije, ali se ne zna kako zna. Primarna struktura se koristi za predviđanje strukture proteina, međutim, pouzdanost te predikcije je poput horoskopa.

Sama sekvenca može služiti za praćenje evolucije nekog polipeptida kroz različite organizme kako bi se videlo šta se u njemu promenilo, takođe, ima primene u medicini za **monogenske bolesti** - one kod kojih je mutacija jednog proteina odgovorna za razvoj neke bolesti. Osobina boja očiju je toliko multgenska da mi ne znamo kako je genski determinisana, dok, na primer, srpska anemija u potpunosti zavisi od jedne aminokiselinske zamene na beta lancu hemoglobina. Skoro svi testovi koji se rade trudnicama služe za skrining na monogenske bolesti.

Evolucija se dešava tako što se u sekvenci DNA, sama od sebe desi neka promena, koja može da bude ubrzana ako se doda neki mutagen ili zračenje. Ta promena se odražava na sekvencu proteina, a možda i na strukturu proteina. Ako se odrazila na strukturu proteina, ona može, ali ne mora, da se odrazi na funkciju proteina. Ako se



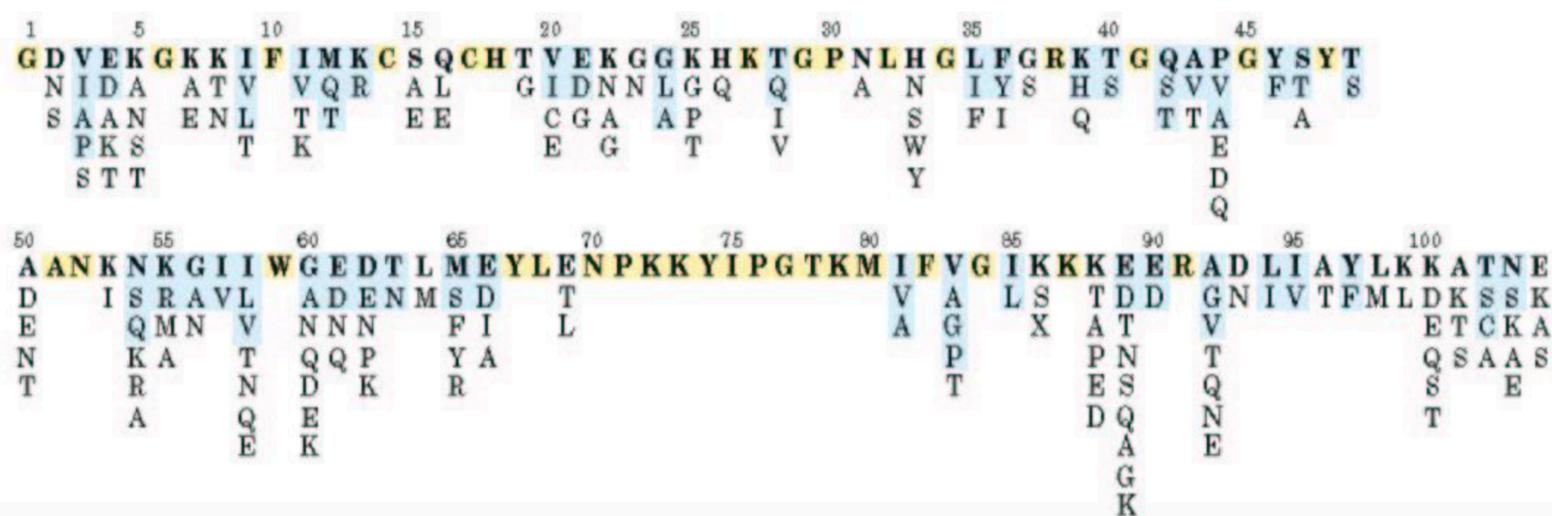
odrazila na funkciju proteina, postoje dve opcije:

1. Protein gubi funkciju i taj organizam neće zaživeti. Selekcija se brine o tome.
2. Ako je nastala poboljšana funkcija, opet će selekcija to da sredi tako što će kroz naredne generacije preostati samo njegovi potomci, jer su bolji, jači i prihvatljiviji partneri od ostalih.

Promene se na sekvenci dešavaju *potpuno spontano* na ma kojem mestu.

### ■ invajirantni ostaci

### ■ konzervativne izmene

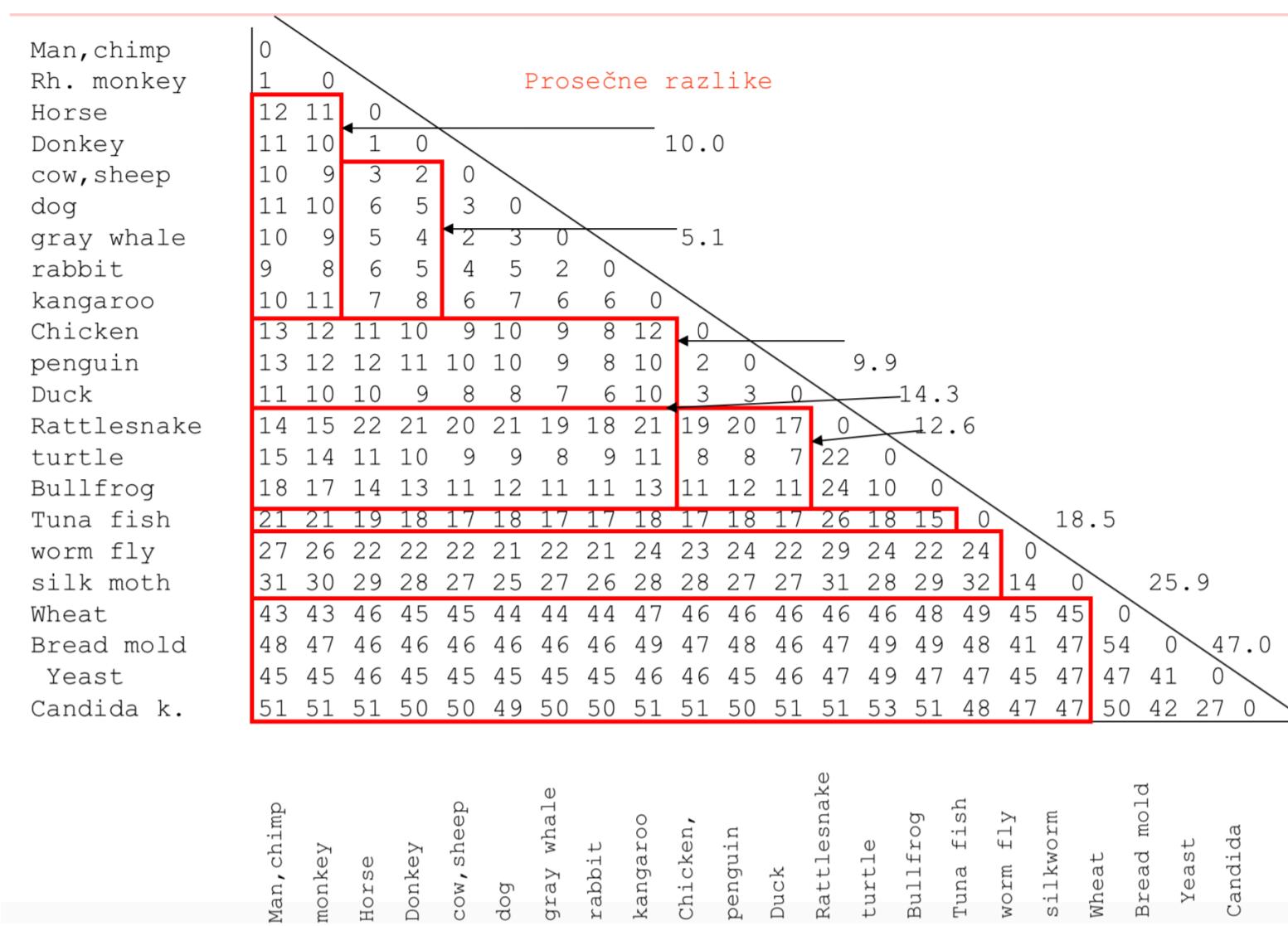


Kada se posmatra evolutivno stablo *cytochrome c*, jednog od proteina koji je najzastupljeniji u živom svetu, jer je zaslužan za naš aeroban život, plavim su obeležene promene, dok su žutom obeležene aminokiseline koje se nisu menjale. Uočavaju se regioni koji su super-variabilni, a uočavaju se i regioni koji se uopšte nisu menjali. Posmatranjem tog proteina kroz više organizama, kaže se da je u pitanju homologi protein - radi istu stvar u različitim organizmima, a ima isto poreklo.

Za neki homologi protein prema starom parametru važi da se razlikuje od organizma do organizma do 30%, sada se zna da to nema smisla (ali je ova informacija bitna zbog literature). **Homologi proteini su oni koji imaju neku sličnost u sekvenci (bez kvantitativnog sečenja da mora da bude određena vrednost), sličnost u načinu uvijanja i sličnost u funkcijama.** Kada se primete invarijantni regioni, to ne znači da gen nije imao mutaciju u tim regionima, već da te mutacije nisu evolutivno prihvatanе.

Invarijantni ostaci reflektuju hemijske neophodnosti, ako se tu nešto promeni protein ili neće moći da se uvije ili neće moći da obavi svoju funkciju. Mogu postojati i konzervativne izmene - zamene sličnim, koje su često dozvoljene, aspartat glutamatom, serin treoninom,... Na taj način se može kreirati filogenetsko stablo - odnos među organizmima koji produkuju protein. Svaka tačka granjanja ukazuje na zajedničkog pretka. **Relativno evoluciono rastojanje**

među tačkama grananja se izražava u broju izmena na 100 aminokiselinskih ostataka u proteinu. **PAM** - *percentage of accepted mutations*



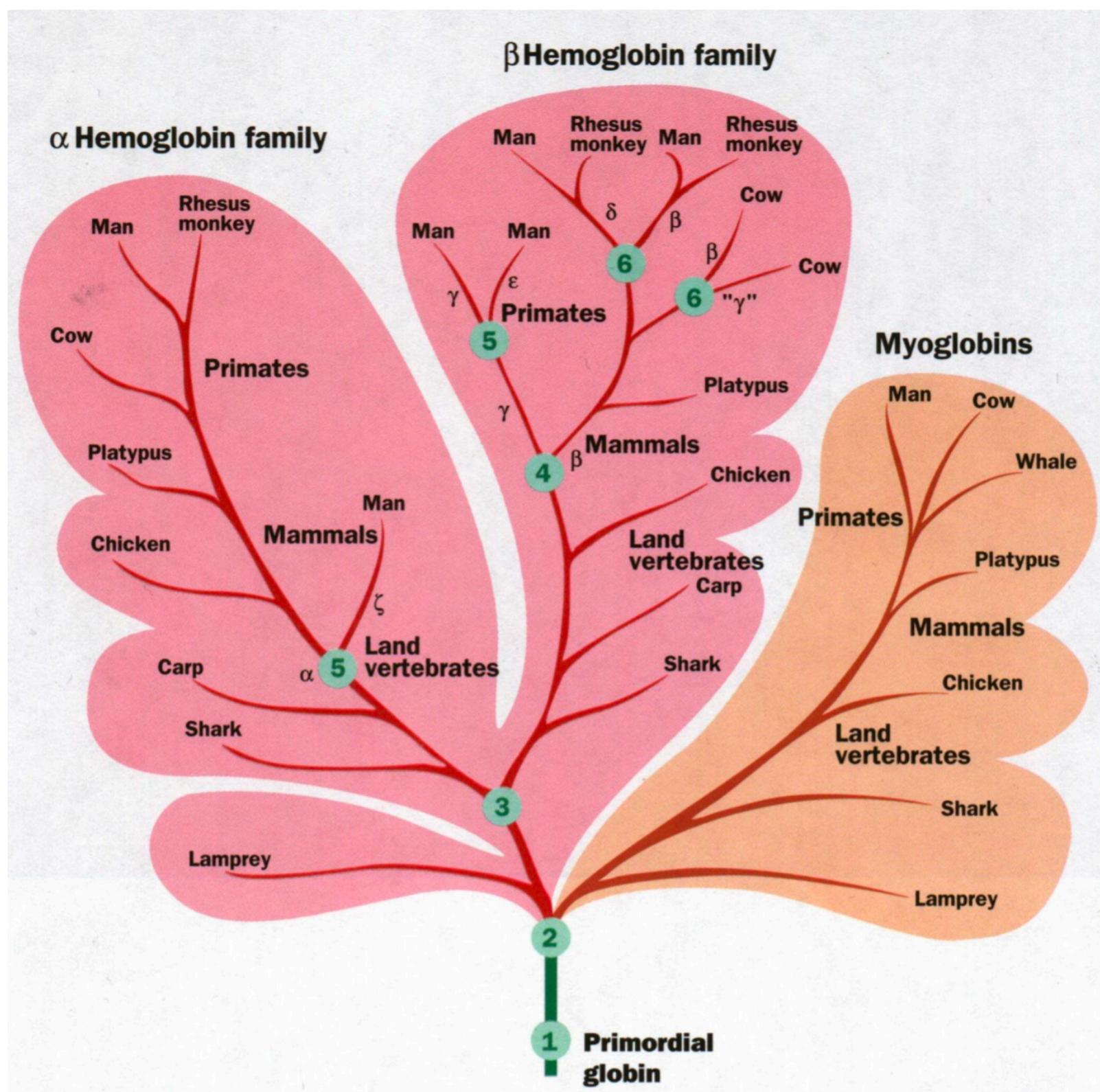
Stablo citohroma oslikano je na slici iznad.

## VELIKE DUPLIKACIJE

Postoje mobilni genski elementi koji se nazivaju *transpozoni*. Oni su evolutivno bili verovatno retrovirusi. Transpozoni kopiraju segmente gena i prenose ih na druga mesta, taj segment je tada udvojen i oni mogu paralelno da evoluiraju. Original će davati funkcionalni protein koji je neophodan organizmu, a kopija će imati mnogobrojne mutacije (ali organizam neće zbog tih mutacija umirati). Usled mutacija može da se desi da se razvije protein koji će imati sasvim drugačiju funkciju.

Postoji dosta primera koji su nastali u n-tosnručavanjem gena, homologi proteini se mogu posmatrati u okviru istog organizma. Primer je globinska familija, gde hemoglobin transportuje kiseonik sa mesta gde ga ima mnogo - u plućima - ka perifernim delovima. Mioglobin se eksprimira u perifernim tkivima, on skladišti kiseonik u perifernim delovima. To je veoma korisno ako bi se, na primer, bežalo od medveda u šumi, tada bi dovođenje kiseonika isključivo iz pluća uz pomoć hemoglobina bilo previše sporo, a mioglobin

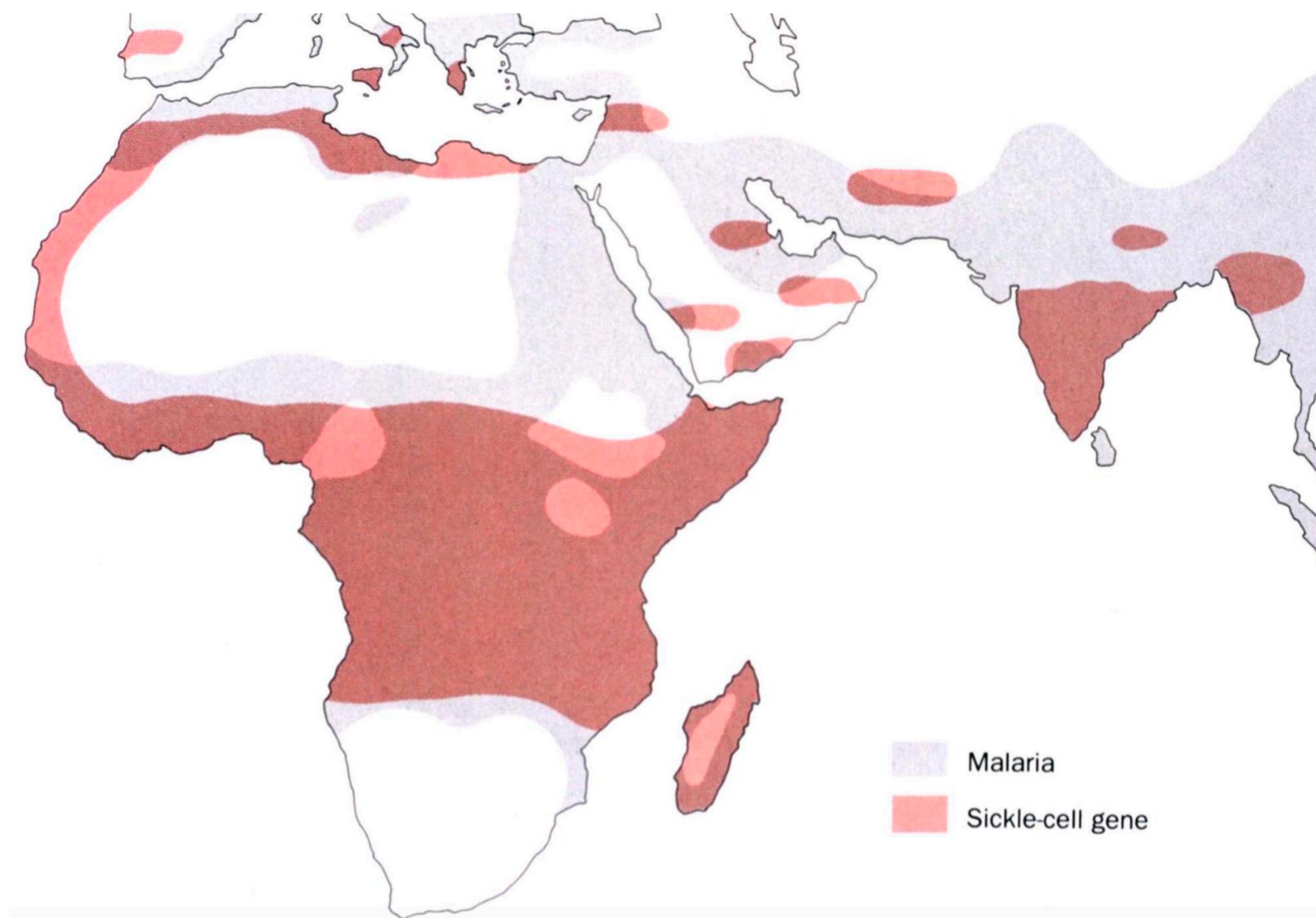
obezbeđuje da rezerve kiseonika već postoje u mišićima. Daljim mutacijama su se razvili alfa i beta-hemoglobini, itd.



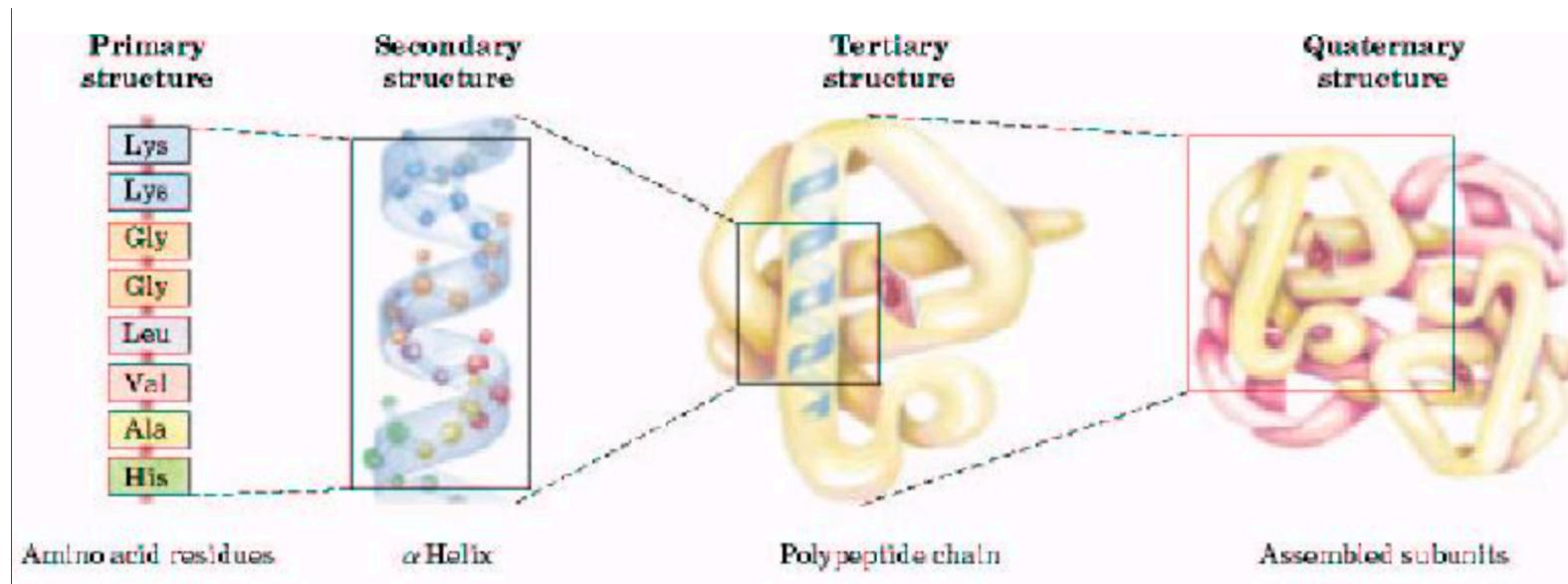
Primarna struktura je bitna i zbog monogenskih bolesti kao što je već pomenuta srpska anemija. Na beta lancu hemoglobina se desi jedna jedina promena glutamat u valin. Pitanje je zašto to nije do sada izbačeno evolucijom, zašto i dalje postoji u populaciji? Odgovor leži u tome što ljudi sa srpskom anemijom ne mogu da dobiju malariju. Normalni hemoglobini stoje u eritrocitu i mogu da se pomjeraju, a srpski hemoglobin sadrži rupicu - hidrofobni džep - koja odgovara hidrofobnom valinu. Kod ljudi koji imaju srpsku anemiju u eritrocitima pojedinačni hemoglobini grade neke agregate umesto da budu slobodni molekuli što rezultuje time da su im eritrociti krtiji nego što su kod normalnih ljudi.

Malarični komarac koji nosi plazmodijum, koji se razmnožava u ljudskim eritrocitima, pravi probleme ljudima sa normalnim eritrocitima, ali u srpastim ne može da pravi problem.

Na slici ispod vidi se mapa rasprostranjenosti srpaste anemije.



# STRUKTURE PROTEINA



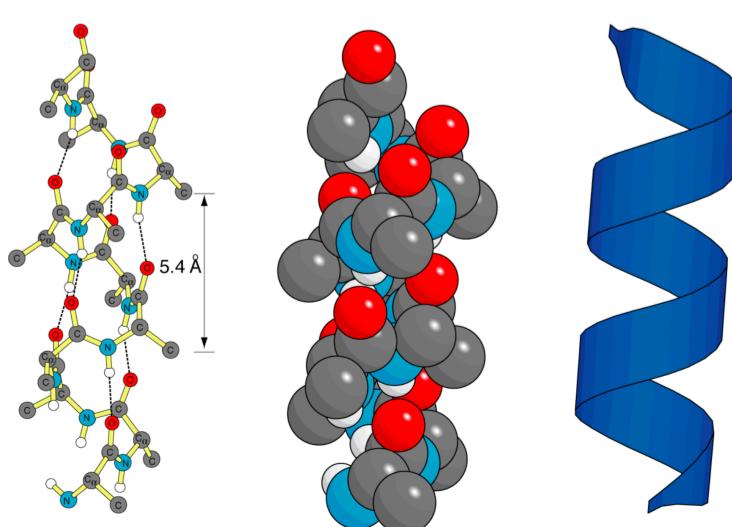
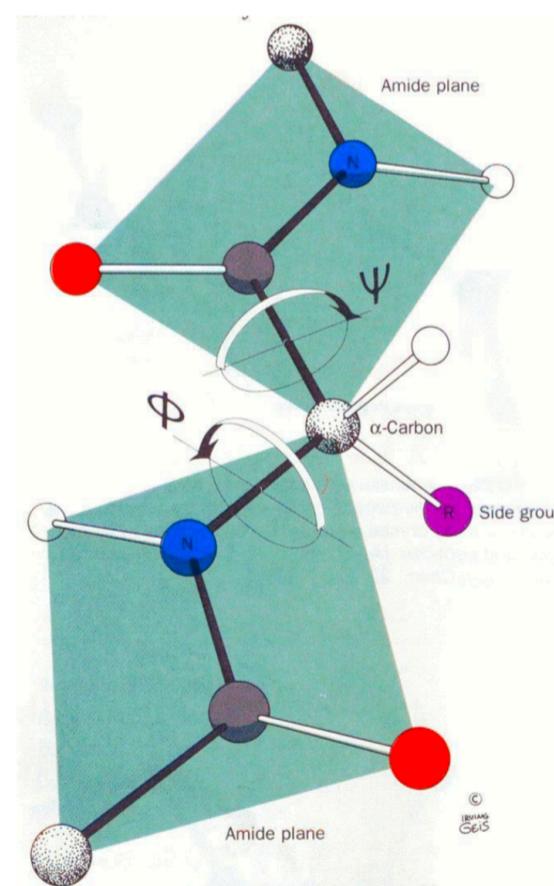
Podsetimo se najpre da je primarna struktura poredak aminokiselina, a da je tercijarna 3-D struktura. Peptidna veza ima dvostruki karakter, usled toga je planarna, pa se na slici desno može videti kako izgledaju dve peptidne veze. Posmatranjem, uočava se da je vodonik jako mali. Pitanje je šta se dešava, ako je sve drugo u ravni, oko

$C - \alpha$  atoma? Imamo tetraedar. Uglovi  $\Phi$  i  $\Psi$  predstavljaju koliki je zaokret. Uočeno je da mogu različito stajati, pa je čovek "vrteo" modele kako bi video koja su sve uvijanja moguća i došao do zaključka da koju god kombinaciju aminokiselina imali (osim dva glicina - što je i nemoguće), gde god postoji realan  $R$ -ostatak koji nije glicin, na primer alanin, suštinski postoje tri opcije:

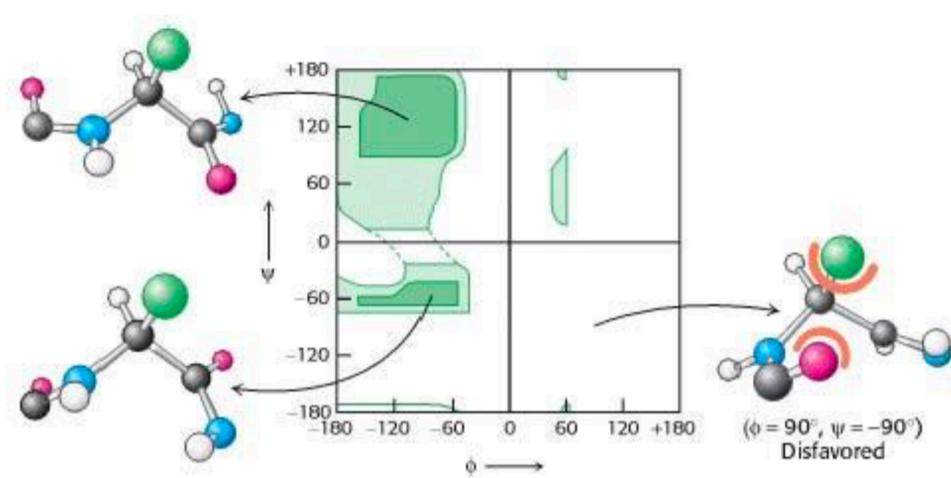
1. Dve peptidne veze su sa iste strane, a dva  $R$ -ostatka su na različitim stranama (isto samo sa drugim znakom -60 i +60) - maksimalno istegnuto

2. U prirodi je nađen samo desni heliks,

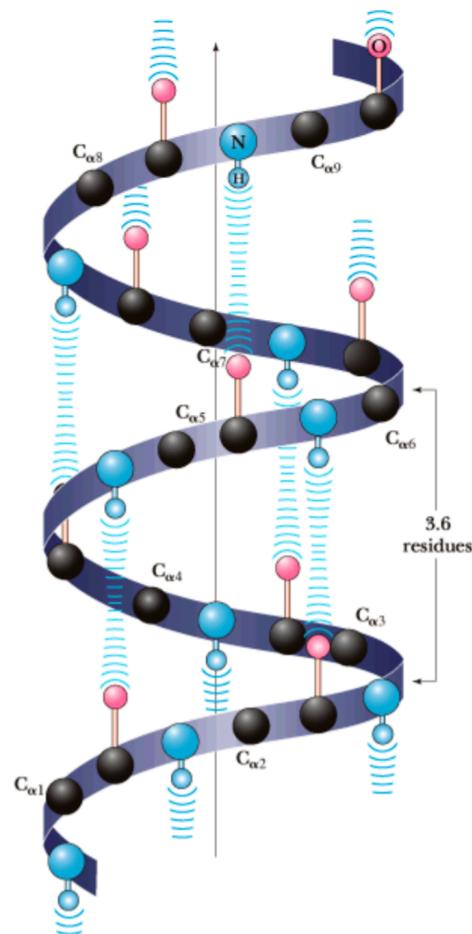
3. Levi je teorijski moguć, ali u prirodnim proteinima ga nema.



Sekundarne strukture su nužna posledica toga da u polipeptidnoj kičmi mogu da postoje ili



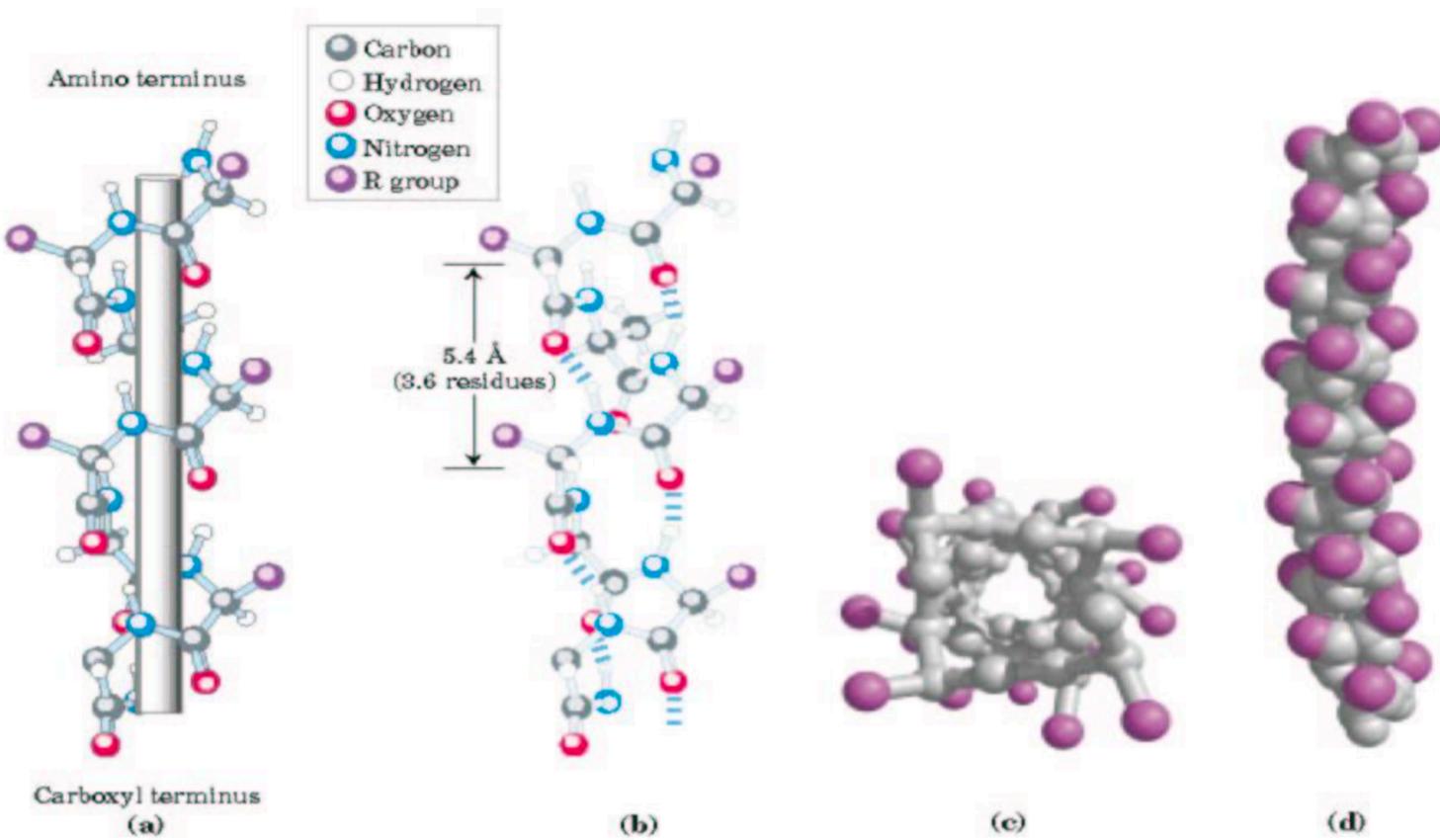
jedni ili drugi uglovi. Na slici levo nalazi se Ramachandranov dijagram, a na slici desno, nalazi se prikaz desnog  $\alpha$ -heliksa.



Posmatranjem samo polipeptidne kičme (zanemarujući  $R$ -ostatke), kažemo da će polipeptidna kičma da se ponaša tako da  $\Phi$  i  $\Psi$  uglovi oko svakog  $C - \alpha$ -atoma budu u dozvoljenim oblastima Ramachandranovog dijagrama. Posmatramo  $C - \alpha$  atome sa kojih štrče  $R$  ostaci (iako oni nisu prikazani na slici). Kada pričamo o sekundarnim strukturama, govorimo o lokalnim savijanjima polipeptidne kičme, zanemarujući  $R$ -ostatke. Sekundarne strukture su stabilizovane vodoničnim vezama (koje nastaju između  $O$  i  $H$ ), opet posmatramo samo polipeptidnu kičmu (u kojoj su peptidne veze) i zanemarujemo  $R$ -ostatke. Postoje peptidne  $N - H$  grupe orijentisane na dole, peptidne  $C = O$  veze su orijentisane na gore i one su kolinearne. To znači da sva 4 atoma stoje na jednoj pravoj i onda mogu da ostvare vodoničnu vezu. **Savijanje polipeptidne kičme uz zanemarivanje  $R$ -ostataka gde članovi peptidnih veza grade dodatne vodonične veze - definicija sekundarne strukture.**

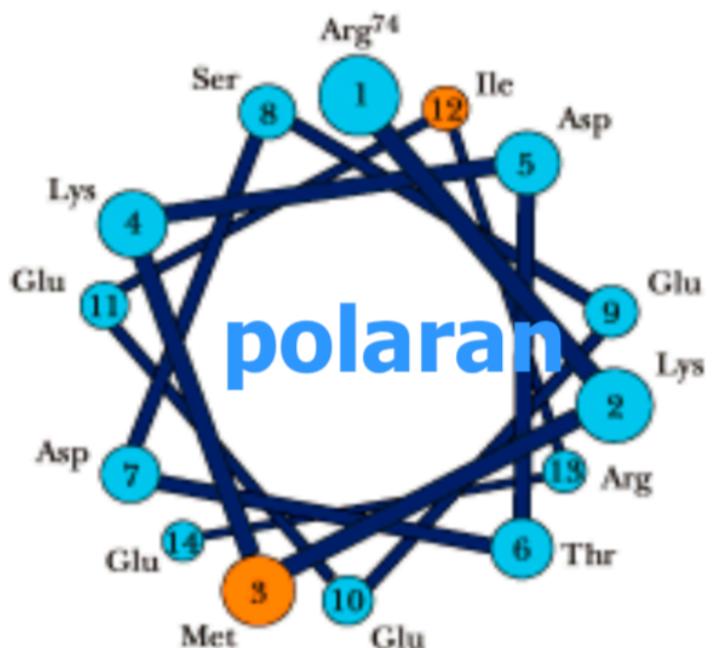
## $\alpha$ -heliks

$\alpha$ -heliks se može prikazati na više različitih načina u zavisnosti od modela. Prva slika nije realan prikaz, heliks je kao valjak sa kog štrče  $R$ -ostaci. Unutrašnjost heliksa **nije šuplja**. Poslednji prikaz je tzv. kalupni model na kome se vidi da je sve ispunjeno.  $R$ -ostaci su obojeni ljubičastom bojom.

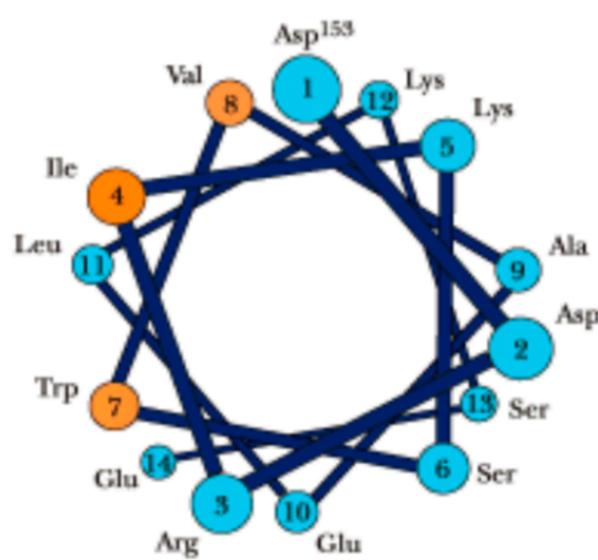
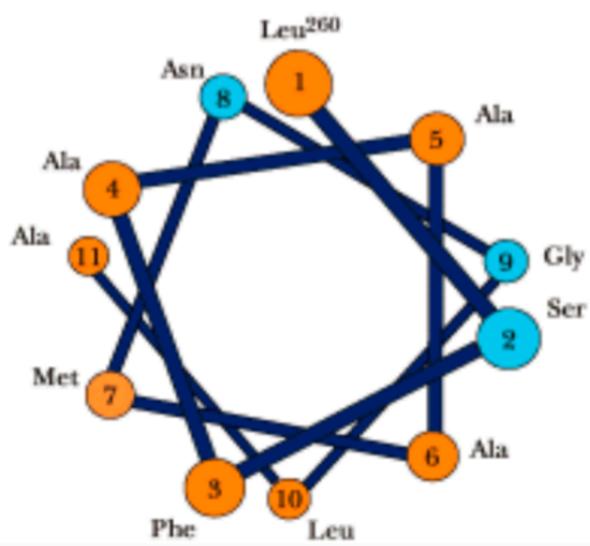


U zavisnosti od toga kakav je  $\alpha$ -heliks, može biti:

1. Polaran
2. Nepolaran
3. Amfipatičan



**nepolaran**

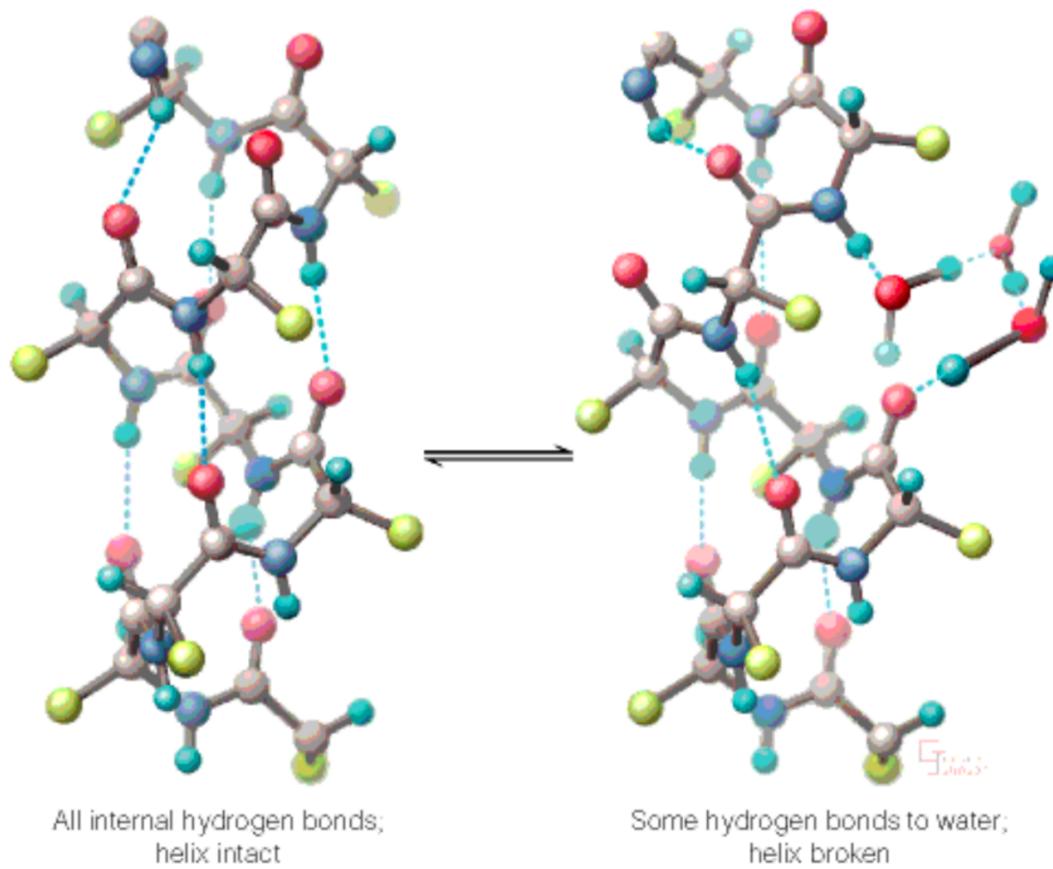
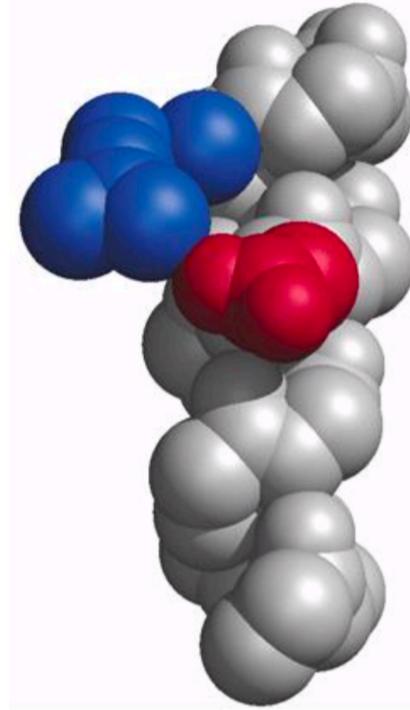


**amfipatični**

Ako je  $\alpha$ -heliks na površini proteina onda je polaran, može da bude nepolaran skroz ako je  $\alpha$ -heliks uronjen unutra i može da bude amfipatičan (pola-pola) ako je polu-uronjen u protein.

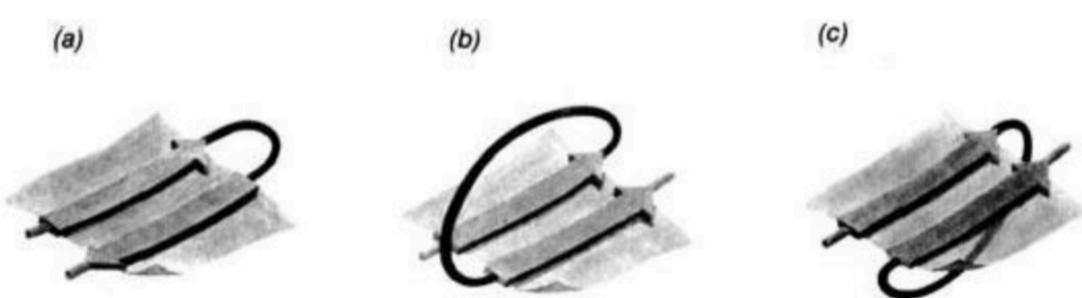
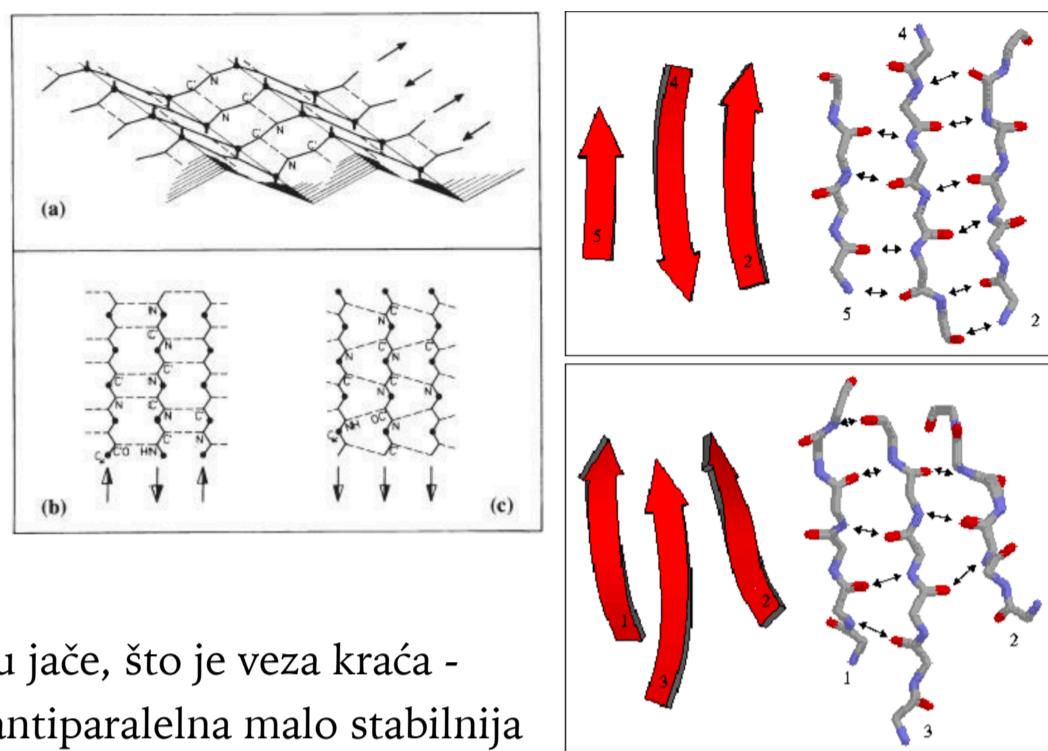
Pitanje je da li  $\alpha$ -heliks van tercijarne strukture uopšte može da postoji? Ako bi jedino stabilizovale vodonične veze, a one ne mogu da se dese u vodenoj sredini zbog kompeticije sa okolnom vodom, onda će ih raskidati, to znači da  $\alpha$ -heliks nije stabilan van tercijarne strukture. Stabilan je u tercijarnoj zato što će njegov deo da bude uronjen u hidrofobno okruženje i onda će u takvom okruženju vodonične veze moći da stabilizuju heliks.

$\alpha$ -heliks može biti **destabilizovan** - ako su R-ostaci koji su sa iste strane prvi i četvrti u nizu (okret heliksa je 3.6), to znači da su geometrijski blizu jedan drugog. Ako su voluminozni, tako da heliks ne može da se spakuje onda on neće moći da se formira.



## Paralelna i antiparalelna $\beta$ -ploča

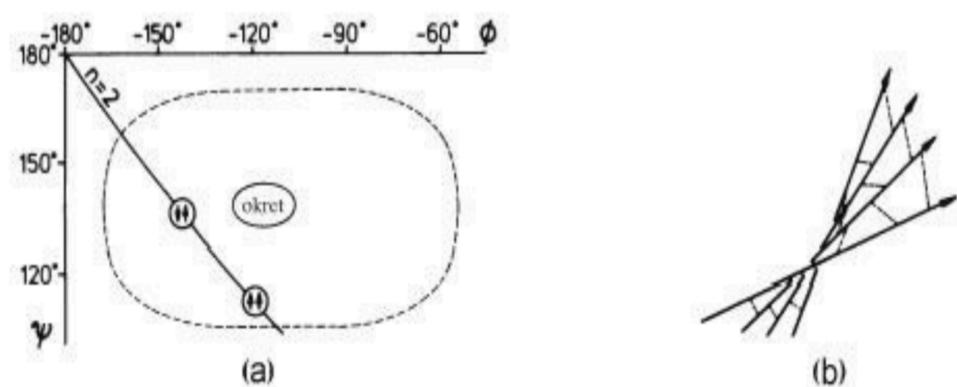
Kada postoji jedan maksimalno istegnuti niz i priđe mu drugi maksimalno istegnuti niz i  $C - O$  i  $N - H$  iz respektivnih nizova izgrade vodonične veze onda je to **antiparalelna  $\beta$ -ploča**. Isto je stabilizovana kao i  $\alpha$ -heliks. Kod antiparalelne  $\beta$ -ploče, vodonične veze su kraće nego kod paralelnih. To znači da su jače, što je veza kraća - jača je. Iz toga se zaključuje da je antiparalelna malo stabilnija od **paralelne  $\beta$ -ploče**.



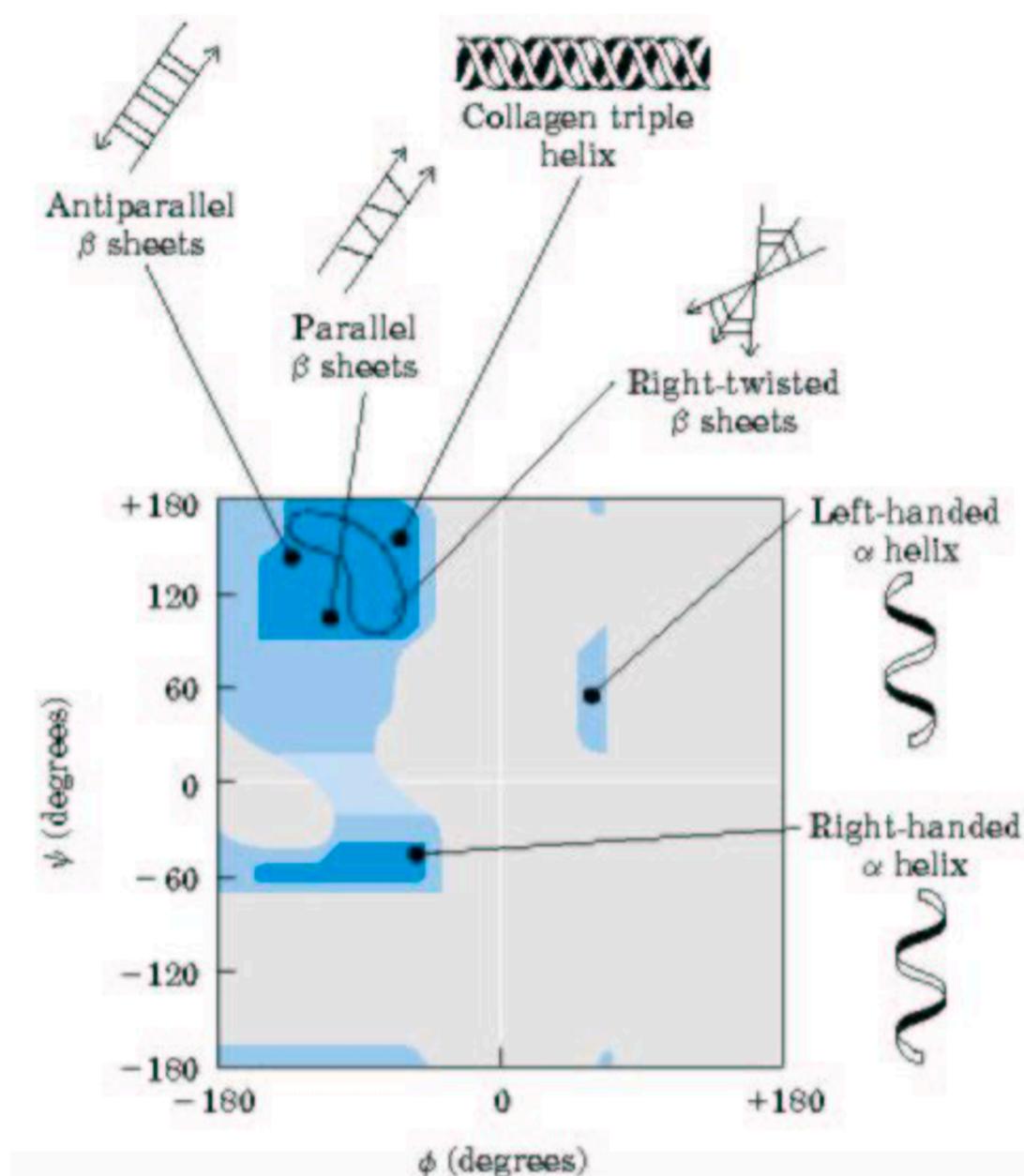
Slika 4.17: Povezivanje susednih nizova u  $\beta$ -pločici: (a) povezivanje susednih antiparalelnih peptidnih nizova vezom tipa ukosnice; (b) desna petlja medju paralelnim nizovima i (c) leva petlja medju paralelnim nizovima. [Na osnovu J.S.Richards, *Adv. Protein. Chem.*, 34, 290 (1981).]

$\beta$ -ploče nađene u molekulima proteina se izvijaju u desno, kako bi se struktura stabilizovala uz pomoć kraćih H veza.

Pored svega navedenog postoji i  $\beta$ -zavijutak - kao  $\alpha$ -heliks ali nema pun okret. Sve ostalo što postoji je kao  $\alpha$ -heliks, ali se uglovi razlikuju. Često se ovakve varijante nazivaju i petljama.



Prikaz na Ramachandranovom dijagramu svih navedenih struktura.



# TERCIJARNA STRUKTURA

Primarna struktura se može odrediti sekvenciranjem proteina, ali sada je određivanje automatizovano i mnogo brže vrši sekvenciranje gena ili celih genoma. Potom se, prevođenjem preko tablice, dobija primarna struktura proteina. Zbog brzine ovih procesa baza primarnih struktura raste eksponencijalno. Sa druge strane, baza podataka tercijarnih struktura se određuje ili kristalografski - X-Ray ili NMR-om, koji ima svoje mane. Metoda je eksperimentalno jako zahtevna i ne može se mnogo ubrzati. **Baza primarnih struktura raste eksponencijalno, dok baza tercijarnih struktura raste linearno.** Postoji ogromna diskrepanca između podataka u bazi primarnih i tercijarnih struktura.

Primeri raznih proteina i podela prema funkciji može se videti u tabeli ispod. Jedino strukturni proteini nemaju tercijarnu strukturu, nego su prosta asocijacija sekundarnih struktura. Ovi proteini su fibrilarni - vlaknasti, dok su svi ostali globularni. Mi najčešće govorimo o globularnim proteinima.

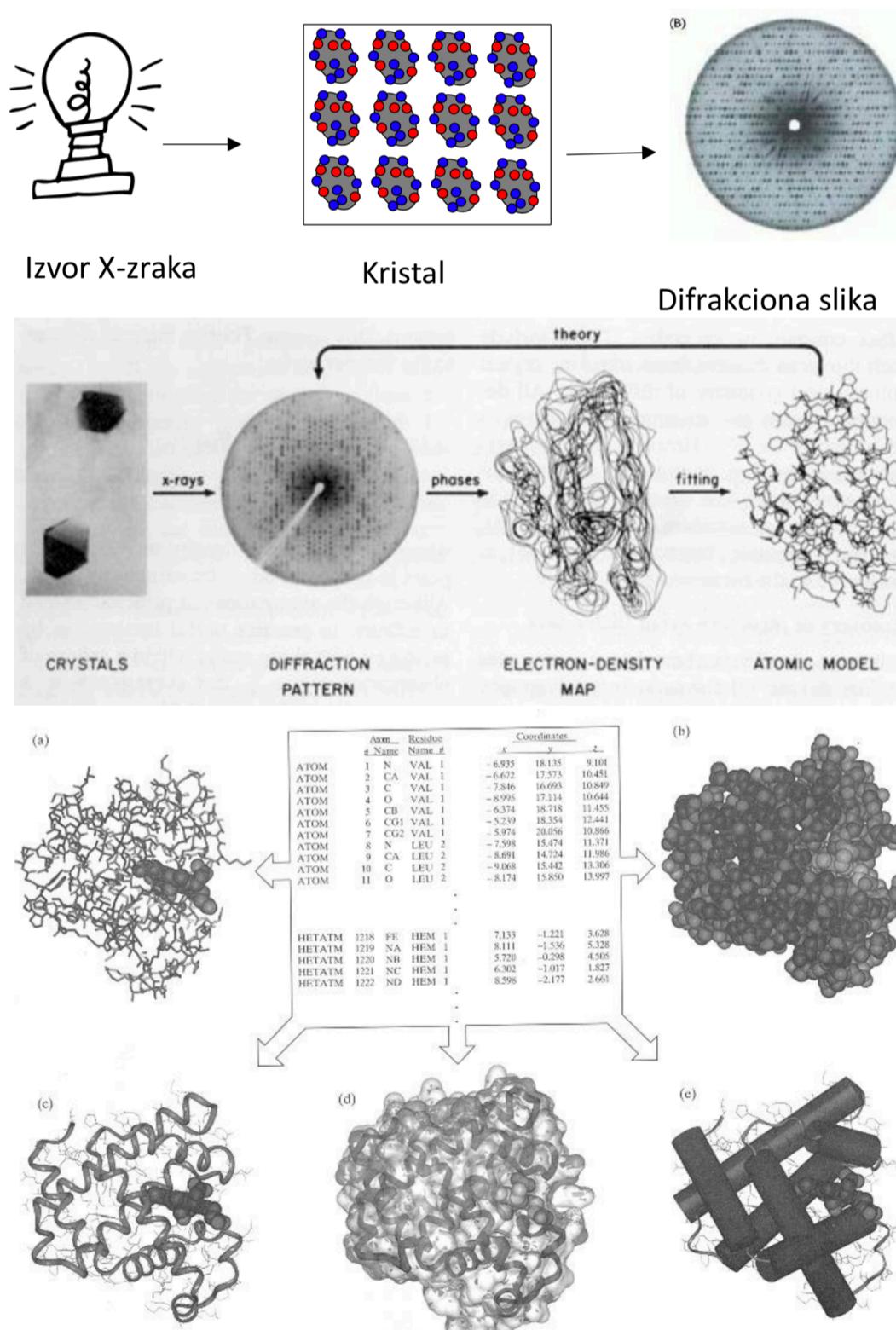
NAZIV PROTEINA	AKTIVNOST / FUNKCIJA / NALAZENJE
Enzimi	kataliza svih reakcija u zivim sistemima
Transportni proteini hemoglobin mioglobin serum albumin	transport kiseonika i ugljendioksida transport kiseonika u misicima transport masnih kiselina, steroida, lekova...
Kontraktilni proteini miozin aktin	pokretljivost misica
Zastitni proteini imunoglobulini fibrinogen trombin	stvaraju kompleks sa stranim telom prekursor fibrina pri zgrusavanju krvi komponenta u zgrusavanju krvi
Hormoni insulin hormoni rasta	regulise metabolizam glukoze stimulisu rast
Rezervni proteini ovalbumin kazein gliadin	rezerva aminokiselina za mladu jedinku Jaje Mleko Psenica
Strukturni proteini α-keratin fibroin kolagen	kosa, koza, krvno, nokti svila, paukova mreza vezivno tkivo

# KAKO SE ODREĐUJE TERCIJARNA STRUKTURA?

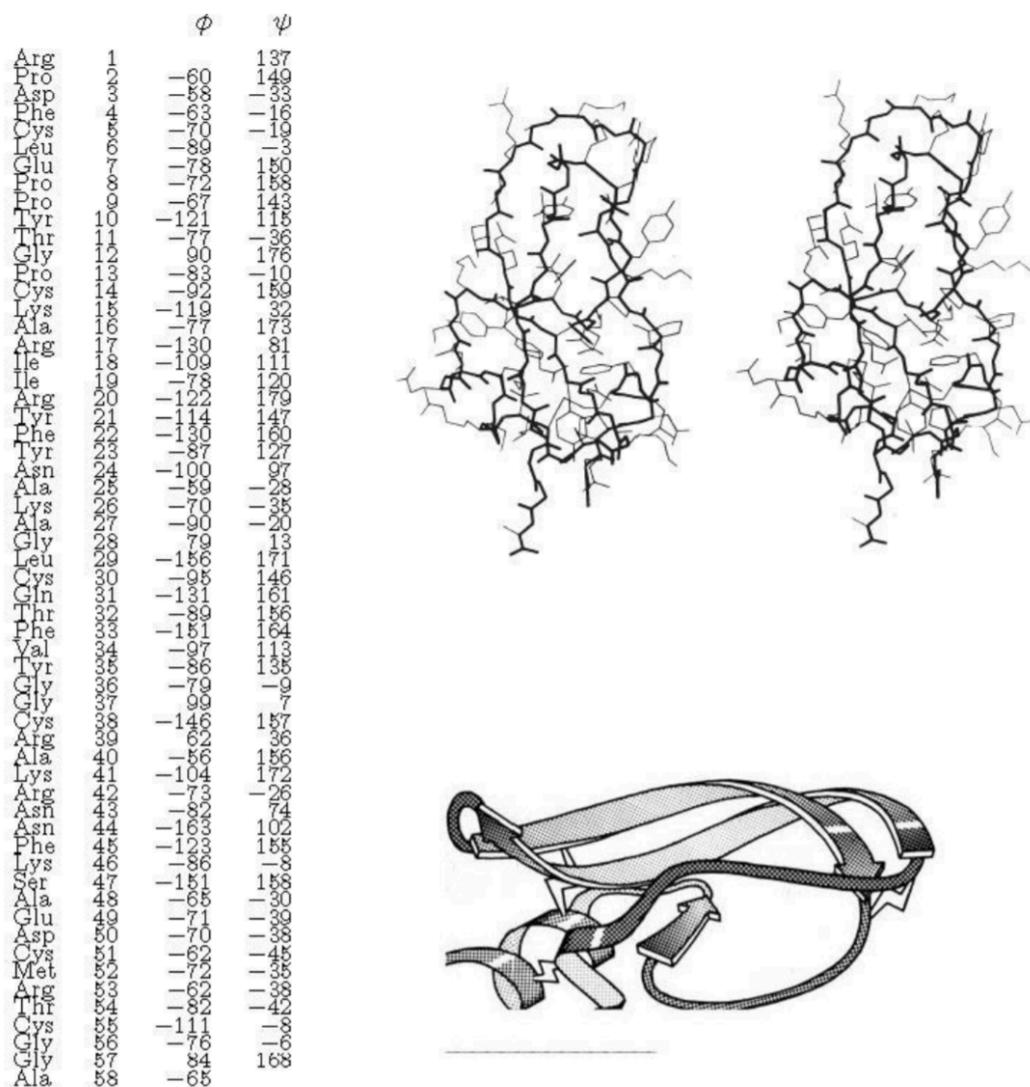
Kristalografski ili NMR-om. Predviđanje tercijarne strukture nije isto što i određivanje. Određivanje je definitivni, eksperimentalno dokazani način, ovi proteini se nalaze u PDB bazi podataka. Predviđanje je kao "horoskop", predviđena struktura ne mora imati veze sa realnom strukturom, jer još uvek nismo stigli do toga da možemo realno da predviđamo.

Rentgenska strukturalna analiza proteina se sastoji iz nekoliko koraka:

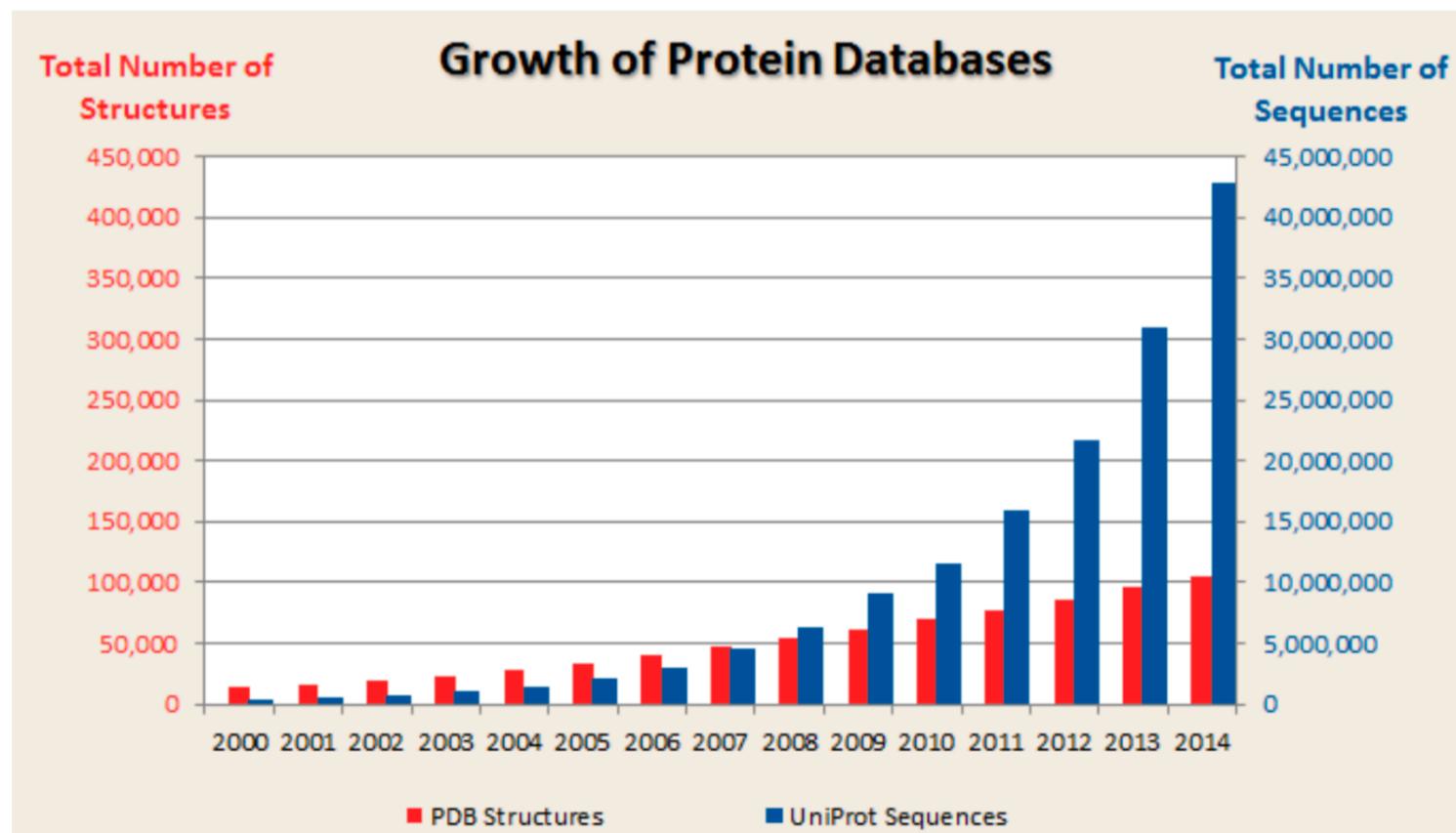
1. Kroz kristal proteina se propuste X-zraci
2. X-zrak kada dođe do nekog atoma koji nije vodonik, odbije se
3. Zrak pada na ploču (poput fotografске ploče) i gde god je zrak pao javiće se tačka
4. Iz tih podataka se računa 3-D struktura

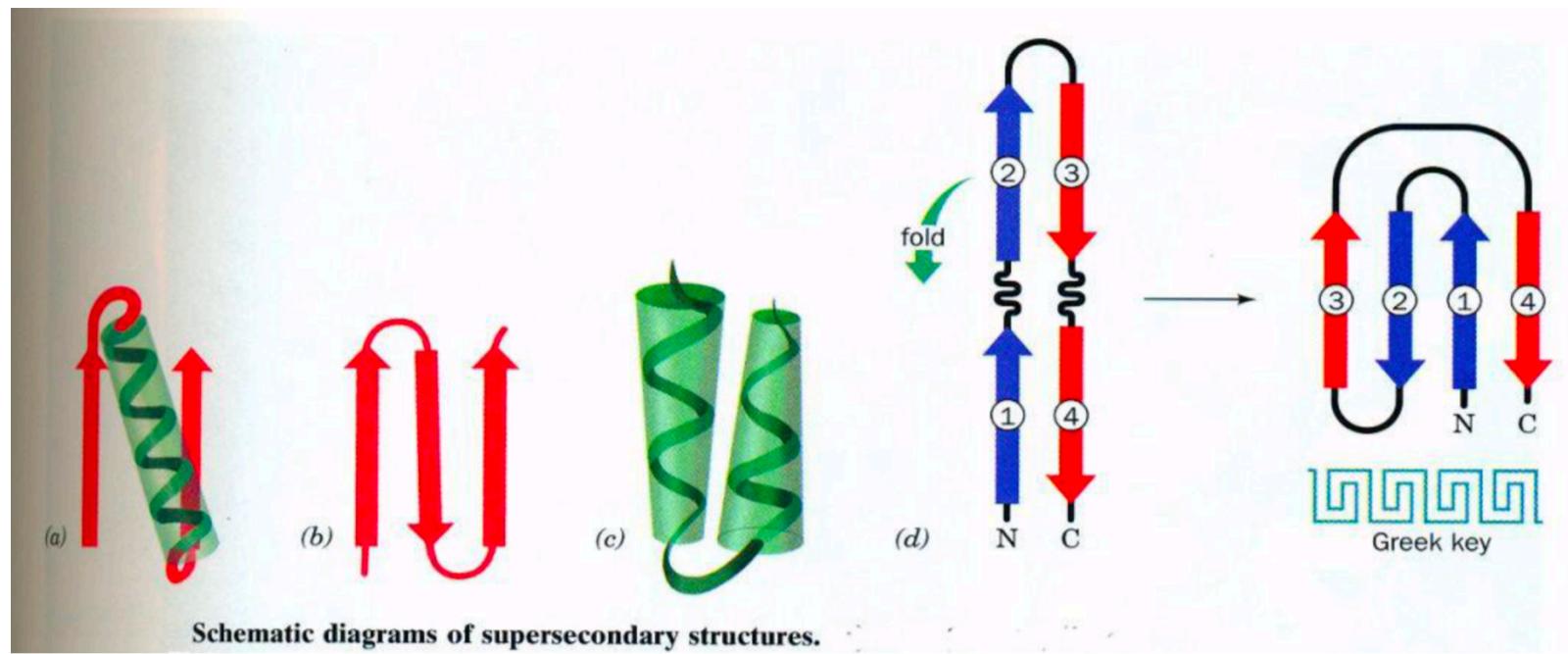


Na slici iznad se vidi kako se tumače dobijeni podaci iz baze, gde je za svaki atom data X, Y i Z koordinata. Na slici ispod je prikazan jedan od starijih načina za prikazivanje gde se na osnovu datih aminokiselina i uglova može proračunati gde se nalaze ostali atomi, osim vodonika.



U narednoj tabeli prikazano je povećanje u broju unosa za UniProt i PDB baze podataka. Sa slike se jasno vide eksponencijalan rast jedne i linearni rast druge.



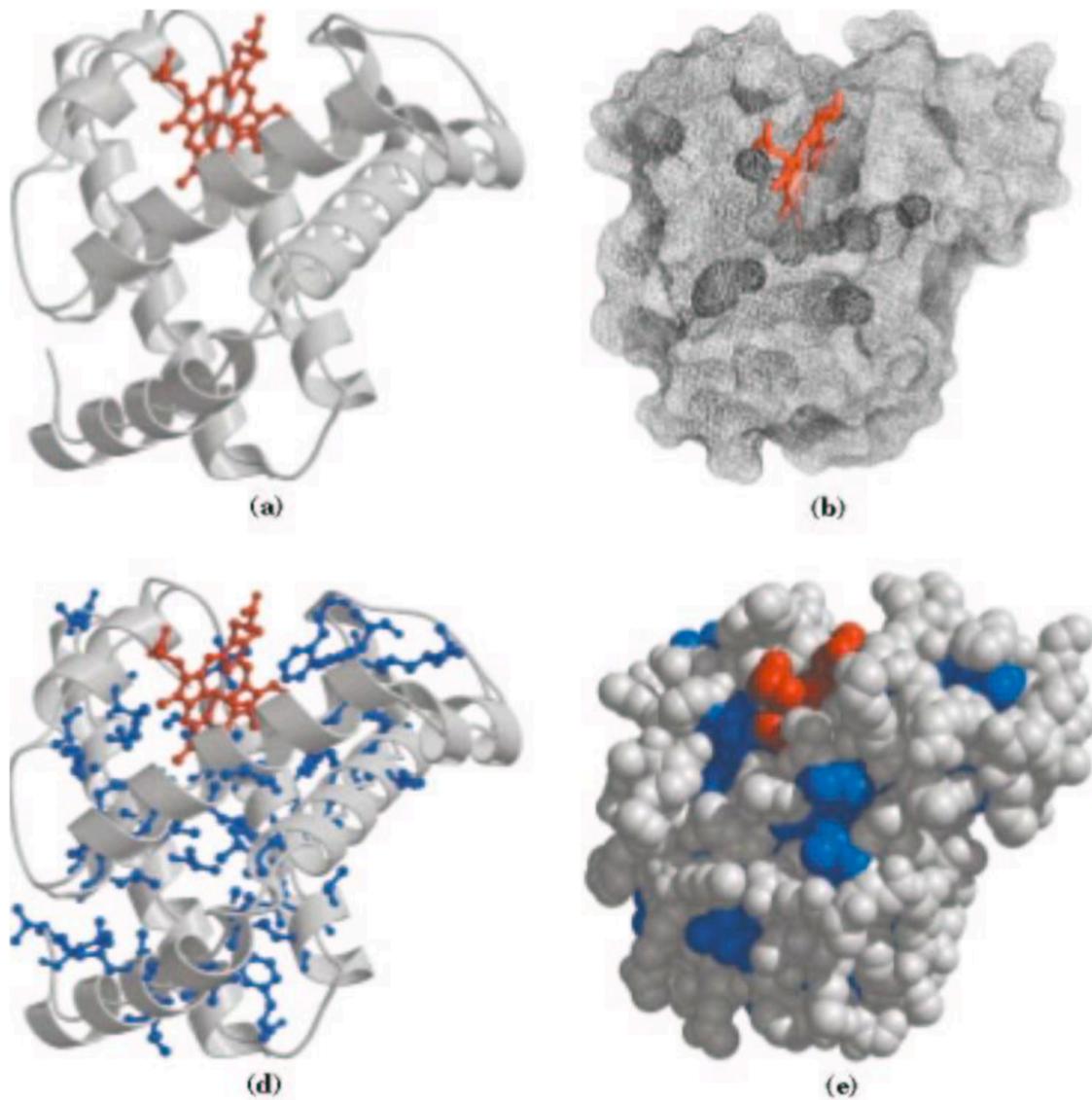


## Tercijarna struktura mioglobina

Kada posmatramo jedan protein, sve što važi za njega važi i za sve ostale.

Postoji više različitih načina da posmatramo jedan protein, ali je važno razumeti da nema šupljina iako tako deluje na prikazu. Na slici ispod je prikazan mioglobin na četiri različita načina. Molekul je jako

kompaktan - gusto spakovan. Ono što je tamno u drugom prikazu to je nepolarno (tamno plavom bojom) i oni su uglavnom okrenuti na unutra, a polarne su obrnute na spolja - ka vodi. Unutrašnjost nepolarna, spoljašnjost polarna.

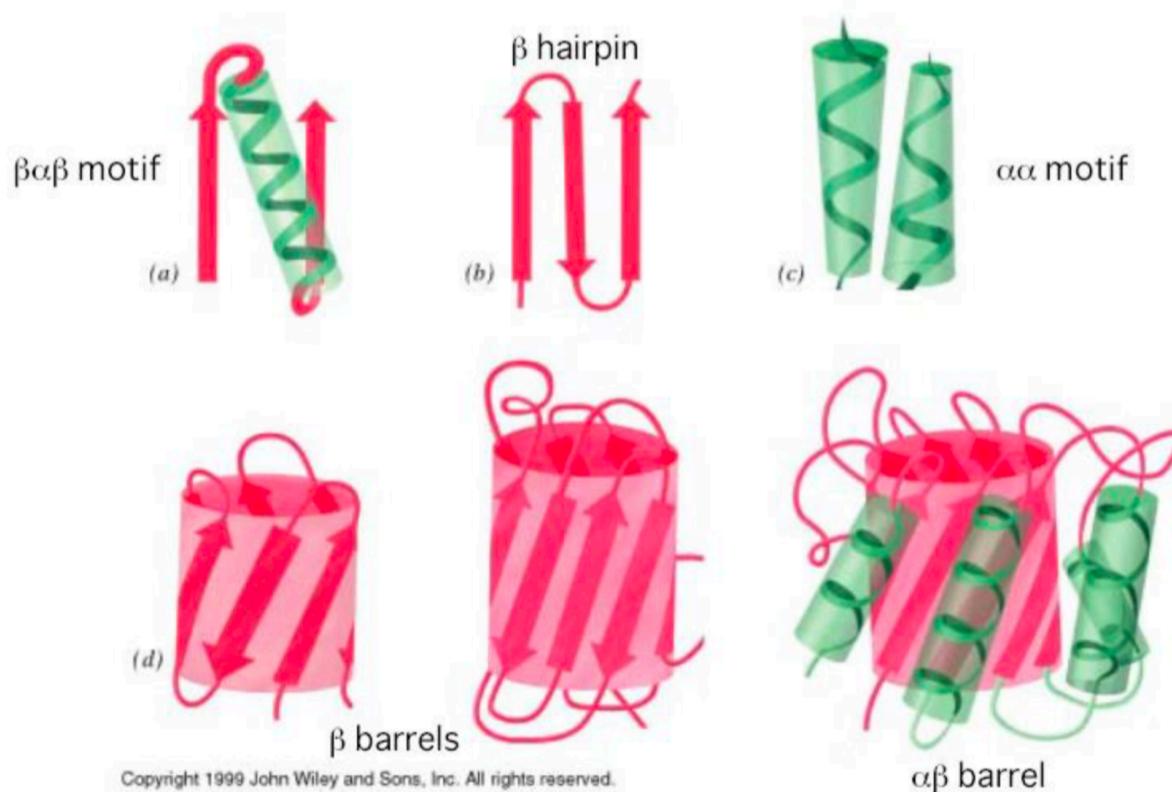


## SUPERSEKUNDARNE STRUKTURE ILI MOTIVI UVIJANJA

Motivi, drugačije nazvani i supersekundarne strukture, su jednostavne kombinacije sekundarnih struktura u određenom geometrijskom rasporedu. Posmatranjem slike ispod, mogu se uočiti tipovi supersekundarnih struktura, to su:

1.  $\alpha\alpha$
2.  $\beta\beta$
3.  $\beta\alpha\beta$

Ako je  $\beta$ -pločica dugačka, ona može graditi tzv.  $\beta$ -bure. Tako što su R-ostaci ispod i iznad ravni pločice, tada će oni popuniti unutrašnjost bureta i ako se ono nagradilo od paralelnih  $\alpha$  onda je to  $\alpha\beta$  bure.



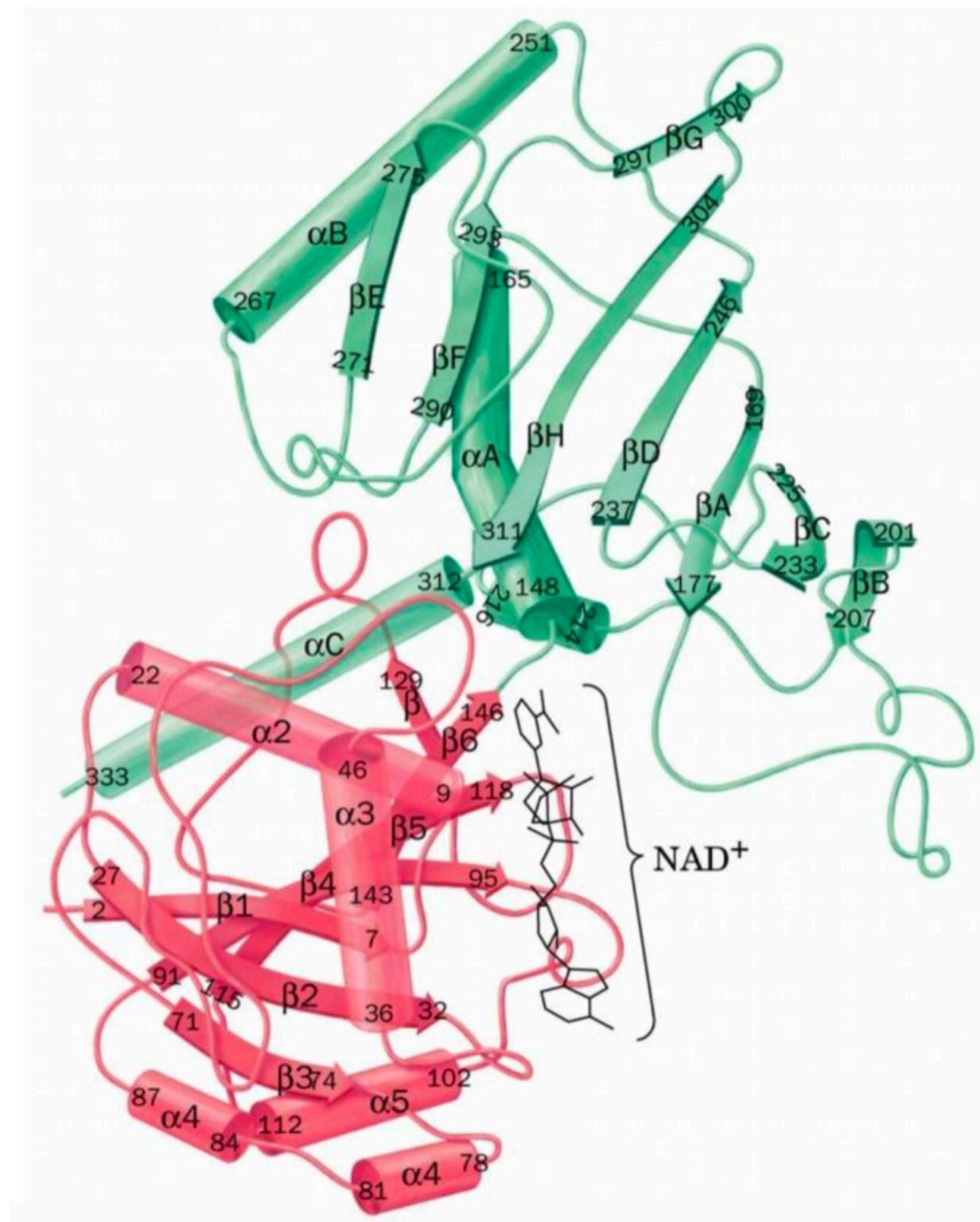
Kada se opisuje neki  $\alpha\alpha$  motiv, tada heliksi stoje pod uglom od 20 stepeni, zbog R-ostataka (1-4-7-11,...). Drugi heliks mora biti malo zakriviljen u odnosu na prvi, kako bi R-ostaci interagovali.

## Domeni

Domeni su poreklom nekada bili nezavisni proteini, a domen je jedna (od više) globula na istom nizu. Usled duplikacije njihovih gena desilo se da je npr. crveni domen iskopiran odnekud i paste-ovan odmah nakon zelenog domena i nakon toga je organizam krenuo da kreira ovakve proteine. Tada se dešava da se svaki od njih uvija nezavisno od onog drugog, ali su povezani peptidnom vezom. To je za neke proteine ispalo odlično. Oksidoreduktaze vezuju dva molekula M1 i M2. M1 treba da oksiduje (uzimaju mu se elektroni), a M2 treba da redukuje (daju mu se elektroni).

Bisupstratni je protein I radi dvostruku reakciju uzima sa jednog dela i daje drugom. Jedan domen prepozna jedan supstrat, drugi domen prepozna drugi supstrat i onda se dešava reakcija između njih. Ti domeni su nekada bili odvojeni proteini, ali su zajedno mnogo efikasniji.

Domeni su strukturni i/ili funkcionalni moduli u okviru jednog proteina, ti moduli su nekada bili nezavisni.



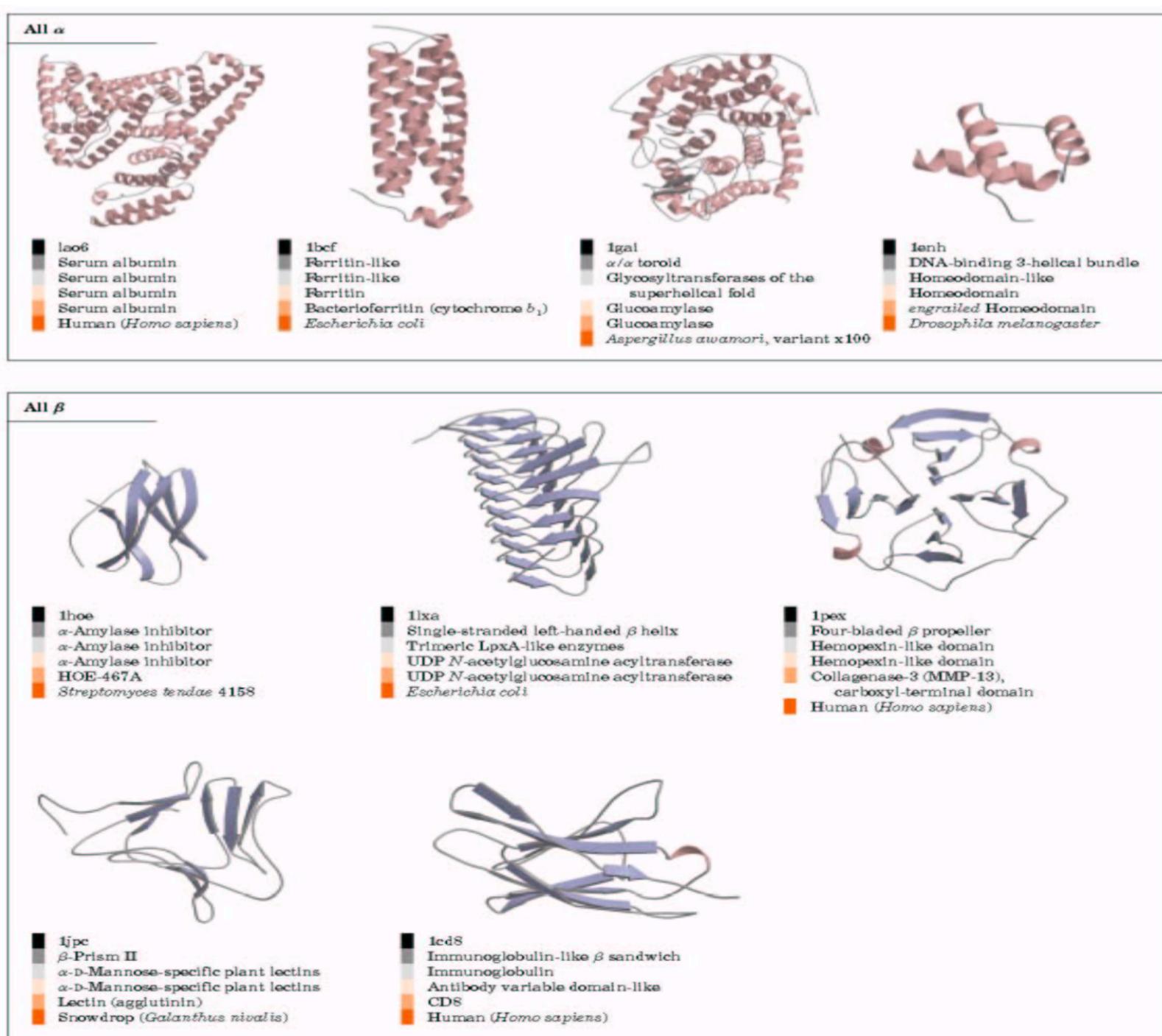
Na osnovu strukture domena, proteini se grupišu u familije, zato što su to evolutivno sve bili nezavisni proteini. Proteini se grupišu u familije na osnovu dva kriterijuma:

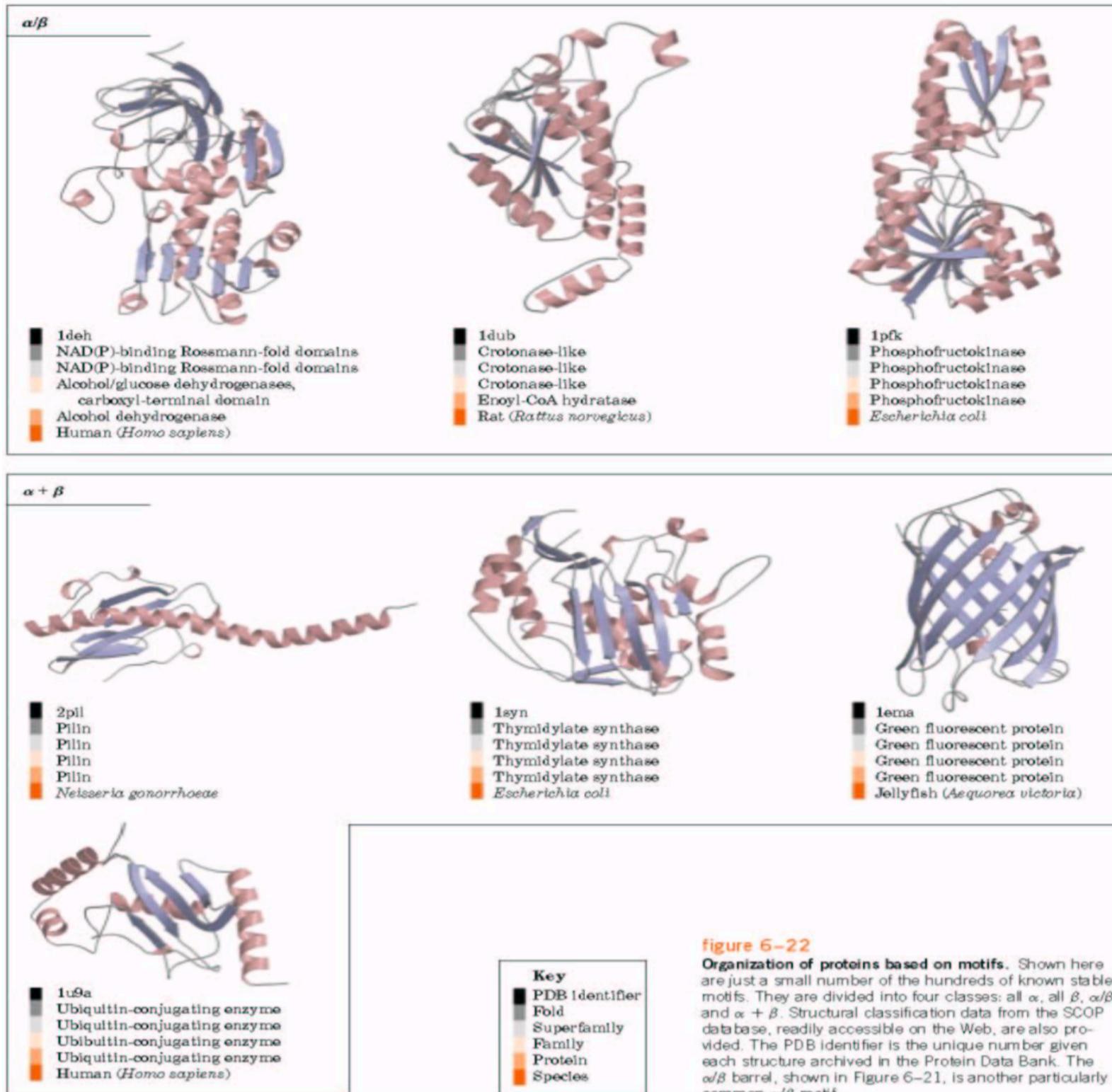
1. (stari kriterijum) proteini imaju  $> 30\%$  identičnih ostataka u aminokiselinskoj sekvenци.
2. Proteini imaju slične 3-D strukture i funkcije.

Svi proteini se klasifikuju u: klase, načine uvijanja, superfamilije, familije, proteinske domene, vrste i domene.

Svi proteini (domeni) se klasifikuju u 4 najznačajnije klase:

1. Samo  $\alpha$
2. Samo  $\beta$  - antiparalelne beta pločice
3.  $\alpha/\beta$
4.  $\alpha + \beta$





**figure 6–22**

**Organization of proteins based on motifs.** Shown here are just a small number of the hundreds of known stable motifs. They are divided into four classes: all  $\alpha$ , all  $\beta$ ,  $\alpha/\beta$ , and  $\alpha + \beta$ . Structural classification data from the SCOP database, readily accessible on the Web, are also provided. The PDB identifier is the unique number given each structure archived in the Protein Data Bank. The  $\alpha/\beta$  barrel, shown in Figure 6–21, is another particularly common  $\alpha/\beta$  motif.

