
3. Noțiuni de biochimie

În acest capitol vom trece în revistă structura și proprietățile a două mari clase de molecule ce constituie de fapt obiect de studiu pentru bioinformatică:.

- proteinele, compuse din aminoacizi
- acizi nucleici, compuși din nucleotide

3.1. Aminoacizii

3.1.1. Structură generală

A. Aminoacizii (AA) sunt molecule organice care conțin o grupare amino ($-\text{NH}_2$) și o grupare acidă carboxil ($-\text{COOH}$).

B. Formula generală este $\text{H}_2\text{N}-\text{CH}-\text{R}-\text{COOH}$. Atomul de carbon primar (numit și C_α) are cele 4 valențe ocupate astfel: una care leagă gruparea amino, a doua leagă gruparea carboxil, a treia leagă un proton ($-\text{H}$) iar a patra leagă o grupare numită radical ($-\text{R}$). Aminoacizii diferă între ei prin acest radical (figura 3.1.1.a).

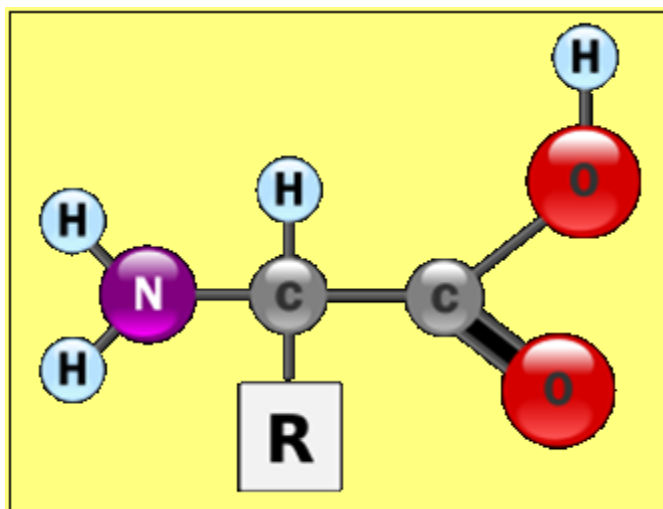


Fig. 3.1.1.a. Formula generală a aminoacizilor

C. În materia vie întâlnim doar 20 AA, aceiași atât în regnul vegetal cât și animal, la toată scara evolutivă, demonstrând elocvent unitatea materiei vii pe pământ. Însă numărul practic infinit al combinațiilor posibile, acoperă cu prisosință orice paletă imaginativă, explicând toate diversitățile, la fel cum cu numărul restrâns al literelor alfabetului se pot crea orice cuvinte.

D. În figura 3.1.1.b sunt prezentate structurile moleculare în forma lor uzuală.

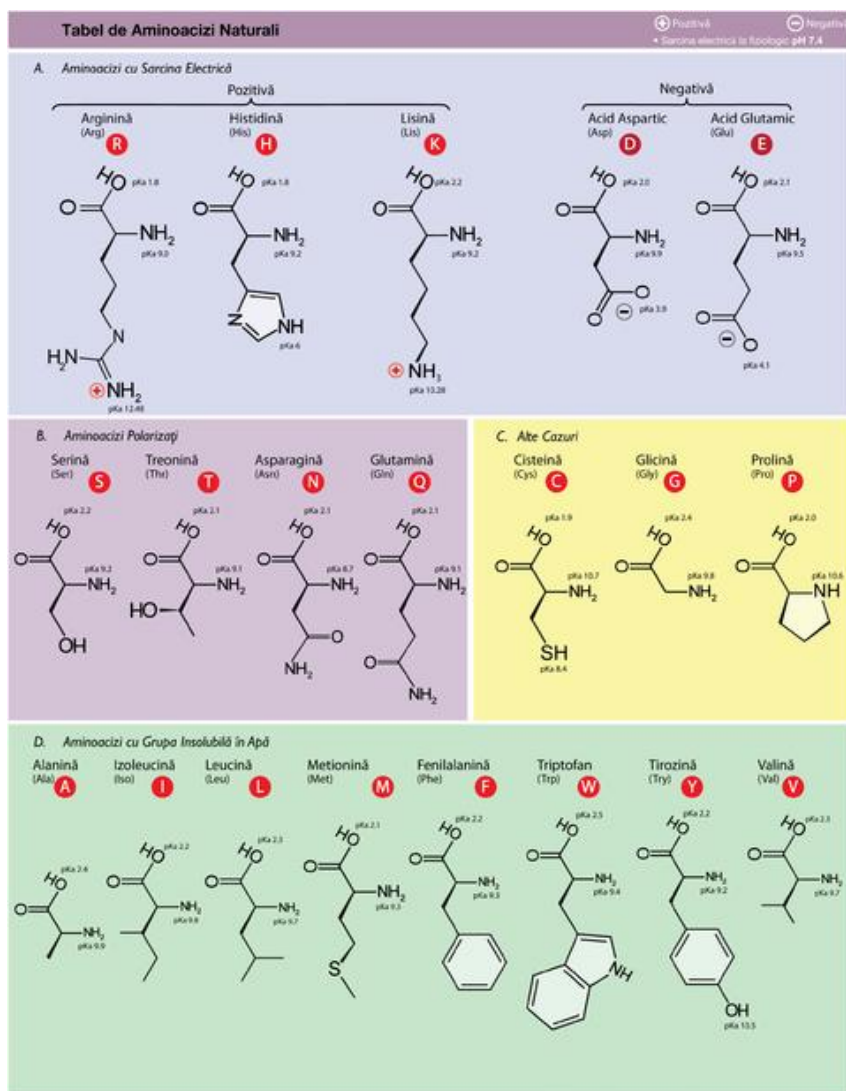


Fig. 3.1.1.b. Structurile moleculare ale aminoacizilor

3.1.2. Formele ionice ale aminoacizilor

A. Gruparea carboxil are caracter acid, putând elibera un proton și trecând în forma COO^- , cu sarcină negativă.

B. Gruparea amino are ca caracter bazic, putând accepta un proton și trecând în forma NH_3^+ , cu sarcină pozitivă.

C. Având caracter dublu, atât acid cât și alcalin, spunem că are un caracter amfoter.

D. Formele ionice care se pot forma, fiind atât pozitive cât și negative, cu posibilitatea ca o moleculă să prezinte simultan ambele sarcini, poartă denumirea de “zwitter-ioni”, și sunt prezentați în figura 3.1.2.a.

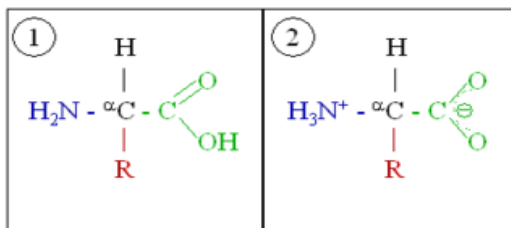


Fig. 3.1.2.a. Forme ionice ale aminoacizilor

E. Sarcina pe care o are un aminoacid la un moment dat depinde de pH-ul mediului. În mediu acid, unde avem o abundență de protoni H⁺, gruparea carboxil nu disociază, în schimb gruparea amino se va găsi în formă ionică, iar aminoacidul va avea sarcină pozitivă. În schimb în mediu bazic, gruparea amino va fi neutră, iar cea carboxil ionizată, având o sarcină globală negativă.

F. Pentru fiecare aminoacid există o anumită valoare a pH-ului, numită *punct izoelectric* pentru care gradul de disociere ca acid este egal cu cel de disociere ca bază, numărul sarcinilor pozitive este egal cu cel al sarcinilor negative, iar molecula în ansamblu este neutră.

G. Dependența sarcinii de pH stă la baza unei metode de separare a proteinelor numită *electroforeză*. O bandă de hârtie de filtru umezită în soluție salină are marginile în contact cu doi electrozi între care se aplică o diferență de potențial de ordinul sutelor de volți. Pe bandă se depune o picătură din soluția unei proteine. În prezența câmpului electric moleculele se deplasează (migreză) în funcție de sarcina lor. Viteza de deplasare este invers proporțională cu masa moleculară. Această deplasare datorată câmpului electric se numește *electroforeză*.

3.1.3. Proprietățile aminoacizilor

A. Abundență

Cei 20 AA nu au o răspândire uniformă în proteinele din natură. Abundența lor (în procente) este prezentată în tabelul 3.1.3.a.

B. Izomerie structurală

Dacă se trasează un plan imaginar creat de carbonul α, carbonul din carboxil și azotul din amino, atunci pentru fiecare AA avem două posibilități de poziționare a radicalului R - fie de partea stângă, fie de partea dreaptă a planului, cu alte cuvinte avem 2 izomeri sterici, care poartă numele S și R, iar proprietatea de a prezenta izomeri sterici se numește *chiralitate*.

Aminoacizii din proteinele din natură sunt cu toții izomeri S, acceptând cisteina (Cys) care este R și glicina (Gly) care este nonchirală.

C. Izomerie optică

Compușii care prezintă izomeri structurali, prezintă de obicei și izomerie optică. O serie de substanțe (numite „optice active”) au proprietatea de a roti planul luminii polarizate când această lumină le traversează. În funcție de sensul în care este rotit planul luminii polarizate substanțele pot fi D (dextrogire) - care rotesc planul la dreapta și L (levogire) care rotesc planul la stânga. În aminoacizii de sinteză cele două forme se produc în proporții egale. În aminoacizii naturali întâlnim numai forma L. Fenomenul încă nu are o explicație unanim acceptată.

Tabel 3.1.3.a. Răspândirea aminoacizilor

Denumirea (Residue)	cod 3-litere	cod 1 literă code	Abundență />(%) E.C.
Alanină	ALA	A	13.0
Arginină	ARG	R	5.3
Asparagină	ASN	N	9.9
Aspartat	ASP	D	9.9
Cisteină	CYS	C	1.8
Acid glutamic	GLU	E	10.8
Glutamină	GLN	Q	10.8
Glicină	GLY	G	7.8
Histidină	HIS	H	0.7
Isoleucină	ILE	I	4.4
Leucină	LEU	L	7.8
Lizină	LYS	K	7.0
Metionină	MET	M	3.8
Fenilalanină	PHE	F	3.3
Prolină	PRO	P	4.6
Serină	SER	S	6.0
Treonină	THR	T	4.6
Triptofan	TRP	W	1.0
Tirosină	TYR	Y	2.2
Valină	VAL	V	6.0

D. Aminoacizii esențiali

Majoritatea aminoacizilor pot fi sintetizați în organismul uman. Totuși există unii AA care nu pot fi sintetizați – aceștia poartă numele de AA esențiali. Ei sunt în număr de 9: Val, Leu, Ile, Lys, Thr, Met, His, Trp, Phe. AA sunt sintetizați de plante și sunt necesari în aportul alimentar, ei se găsesc și în produse animale.

E. Notatii

Se folosesc frecvent niște notații prescurtate, formate din 3 litere convenabile pentru recunoașterea denumirii lor. Totuși în analiza secvențială este preferată notația simbolică cu câte 1 literă.

3.1.4. Hidrofobicitatea

A. Grupările din radicali R sunt foarte variate, cu proprietăți diferite, ceea ce conferă aminoacizilor – și prin ei, proteinelor – o paletă largă de proprietăți. În funcție de aceste grupări putem face o clasificare a AA (figura 3.1.4.a)

- AA cu grupări nepolare: Gly, Ala, Val, Leu, Ile, Phe
- AA cu grupări polare neîncărcate electric: Ser, Thr, Cys, Met, Asn, Gln, Tyr, Trp, Pro
- AA cu grupări polare cu caracter acid: Asp, Glu
- AA cu grupări polare cu caracter bazic: Arg, Lys

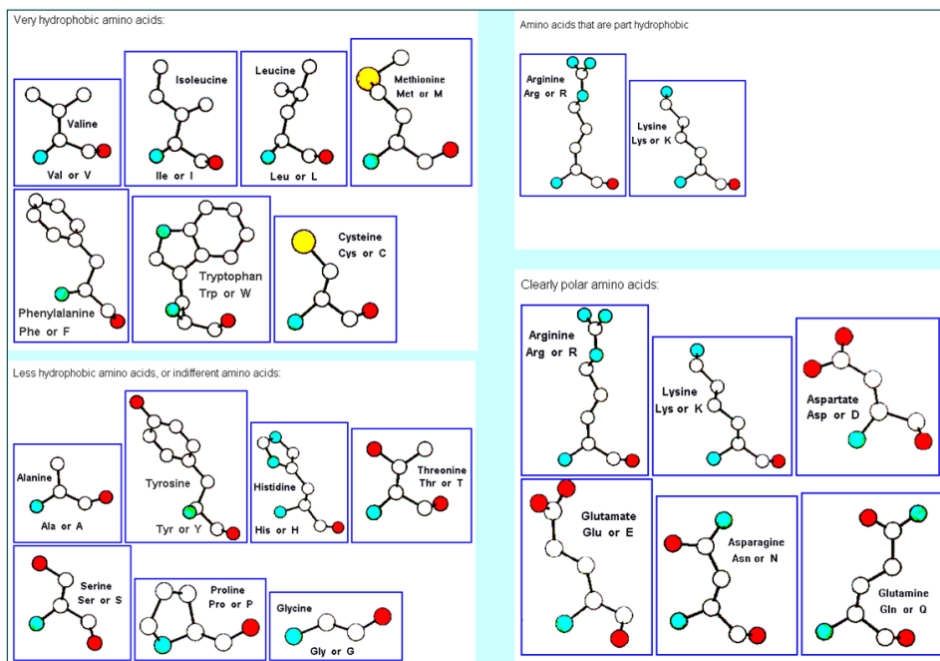


Fig. 3.1.4.a. Clasificarea aminoacizilor

B. Polaritatea grupărilor determină și solubilitatea lor în apă. Astfel AA cu grupări polare sunt hidrofobi, cei cu sarcină sunt hidrofilii, iar cei cu grupări polare neîncărcate electric sunt relativ solubili prin punțile de hidrogen care se pot forma.

C. Scări de hidrofobicitate

Evaluarea gradului de hidrofobicitate al unei molecule este dificilă, fiind cel mai adesea bazată pe partiția moleculei între diferiți solvenți. În funcție de solvenții folosiți s-au stabilit diverse scări care ierarhizează aminoacizii figura 3.1.4.b, dreapta.

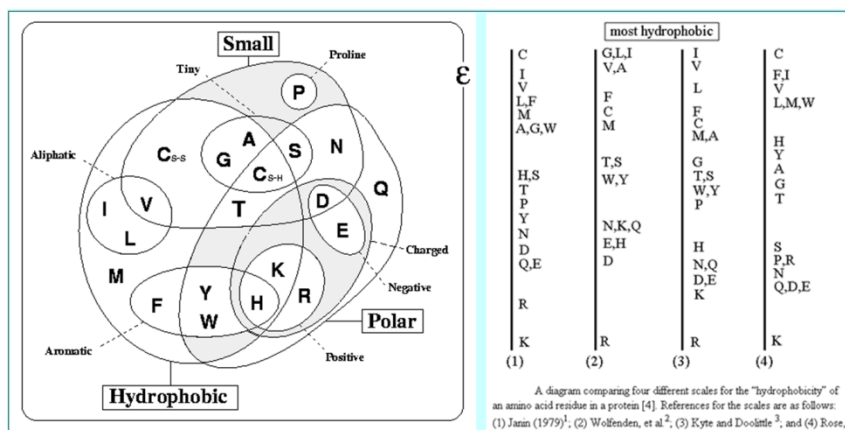


Fig. 3.1.4.b. Hidrofobicitatea aminoacizilor

D. Dacă ținem cont de dimensiunea moleculei, pe lângă hidrofobicitate și caracterul polar/nepolar sau sarcina moleculei la pH fiziologic, putem prezenta schematic o diagramă care sintetizează toate aceste proprietăți figura 3.1.4.b stânga.

E. Pentru a avea o imagine mai detaliată asupra proprietăților aminoacizilor, să mai adăugăm determinările unor proprietăți fizice: densitate, volum, suprafață, constante de echilibru de disociere a grupărilor din radical (unde este cazul). Aceste proprietăți sunt sintetizate în tabelul 3.1.4.

Tabelul 3.1.4. Proprietățile aminoacizilor solubilitate, densitate, volum, suprafață și pK

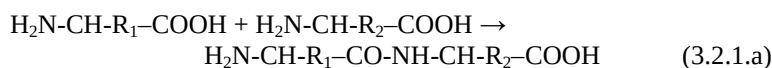
Name	Solubility (g/100g, 25 °C)	Crystal Density (g/ml)	pI at 25 °C	Residue Volume	Surface Area	Side Chain pKa
Alanine	16.65	1.401	6.107	A	88.6	-
Arginine	15	1.1	10.76	R	173.4	~12
Aspartic Acid	0.778	1.66	2.98	D	111.1	4.5
Asparagine	3.53	1.54	-	N	114.1	-
Cysteine	very	-	5.02	C	108.5	9.1-9.5
Glutamic Acid	0.864	1.460	3.08	E	138.4	4.6
Glutamine	2.5	-	-	Q	143.8	-
Glycine	24.99	1.607	6.064	G	60.1	-
Histidine	4.19	-	7.64	H	153.2	6.2
Isoleucine	4.117	-	6.038	I	166.7	-
Leucine	2.426	1.191	6.036	L	166.7	-
Lysine	very	-	9.47	K	168.6	10.4
Methionine	3.381	1.340	5.74	M	162.9	-
Phenylalanine	2.965	-	5.91	F	189.9	-
Proline	162.3	-	6.3	P	112.7	-
Serine	5.023	1.537	5.68	S	89.0	-
Threonine	very	-	-	T	116.1	-
Tryptophan	1.136	-	5.88	W	227.8	-
Tyrosine	0.0453	1.456	5.63	Y	193.6	9.7
Valine	8.85	1.230	6.002	V	140.0	-

a-amino pKa = 6.8 - 7.9, a-carboxyl pKa = 3.5 - 4.3

3.2. Proteine

3.2.1. Legătura peptidică

A. Doi aminoacizi se pot lega chimic prin reacția între gruparea carboxil a unui AA cu gruparea amino a celuilalt, cu eliminarea unei molecule de apă:



B. Legătura formată se numește legătură peptidică. Reacția este bine ilustrată în figura 3.2.1.a, iar proprietățile legăturii peptidice sunt prezentate în figura 3.2.1.b.

C. Ne putem ușor imagina că dipeptidul format, având un capăt amino și un capăt carboxil, poate la rândul său forma o nouă legătură peptidică și așa mai departe. O astfel de moleculă se numește polipeptid sau proteină. Pentru un număr mic de AA în structură se pot folosi prefixele corespunzătoare: dipeptid, tri~, tetra~, penta~ ș.a.m.d.

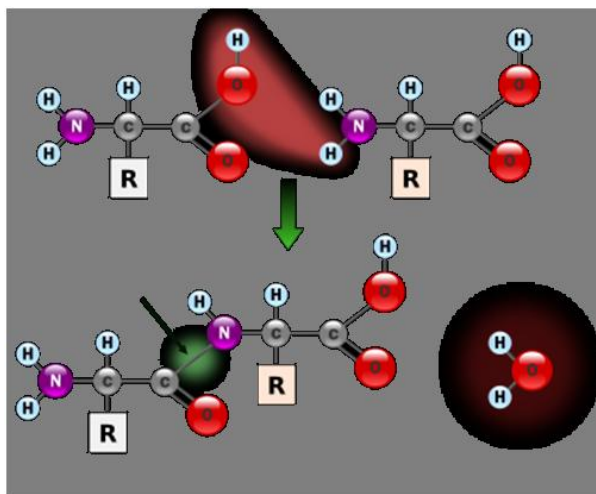


Fig. 3.2.1.a. Formarea legăturii peptidice

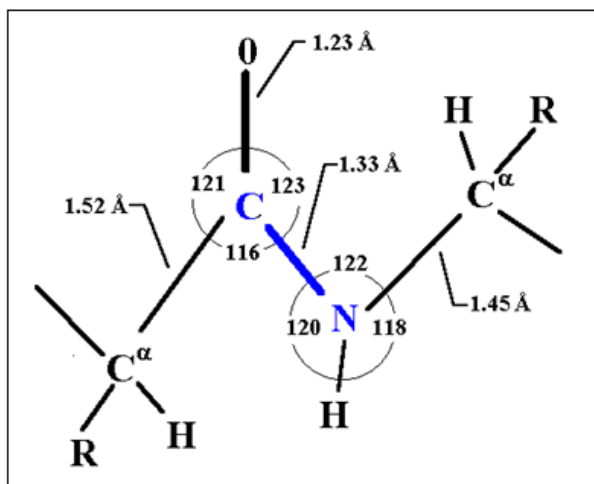


Fig. 3.2.1.b. Proprietățile legăturii peptidice

3.2.2. Structura primară a proteinelor

A. O primă caracteristică, de importanță majoră în bioinformatică este reprezentată de succesiunea de AA care formează molecula proteică.

B. Această succesiune poartă numele de „structură primară” a proteinelor. Pozițiile aminoacizilor se numerează începând de la capătul cu gruparea amino liberă.

C. Determinarea structurii primare a proteinelor

Există mai multe metode de determinare a structurii primare a proteinelor, pe care nu le vom trata în detaliu ci doar le vom enumera aici:

- a) metode biochimice de determinare a secvenței, cuprinzând:
 - degradarea Edman
 - fracționarea prin digestie cu enzime proteolitice
 - spectrometrie de masă

- hidroliză în acizi (doar determinarea cantităților de aminoacizi)
- b) deducerea din secvența genelor

D. Exemple de structuri primare

Una dintre primele structuri primare determinate a fost cea a insulinei, prezentată în figura 3.2.2.a, de către Sanger în 1951, pentru care i s-a acordat premiul Nobel.

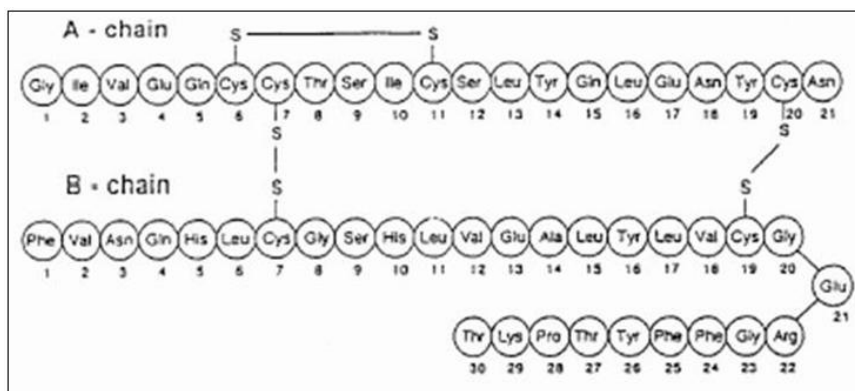


Fig. 3.2.2.a. Structura primară a proteinelor: insulina

3.2.3. Structura secundară a proteinelor

A. O moleculă proteică poate avea de la câteva zeci de aminoacizi până la câteva sute. Prin proprietățile pe care le au radicalii aminoacizilor din lanț, se formează adesea niște structuri spațiale specifice ușor de recunoscut. Corey și Pauling (1943) au fost primii care au semnalat aceste structuri denumite generic „structura secundară” a proteinelor.

B. Structura - helix

Cea mai bine cunoscută formă specifică este cea de helix pe care o întâlnim frecvent la proteine. Unii aminoacizi, mai ales Met, Ala, Leu, Glu și Lys se dispun în forma unor spirale ce s-ar înfășura pe un cilindru virtual, numit - helix. Alți AA însă nu se dispun în helix, apariția lor (de exemplu Gly și Pro) generând discontinuitatea spiralei, de aceea se și numesc „helix breakers”. Procentul de AA angajați în structurile secundare variază. În tabelul 3.2.3.a. sunt trecute proporțiile de AA cuprinși în structuri de tip - helix.

Tabelul 3.2.3.a. Procent de AA în elice

Proteina	% elice α
Mioglobina	70
Insulina	38
Ovalbumina	31
Serumalbumina	46
Pepsina	31
Ribonucleaza	16
Chimotripsina	15

C. Să trecem în revistă câteva proprietăți ale unui α -helix (figura 3.2.3.a)

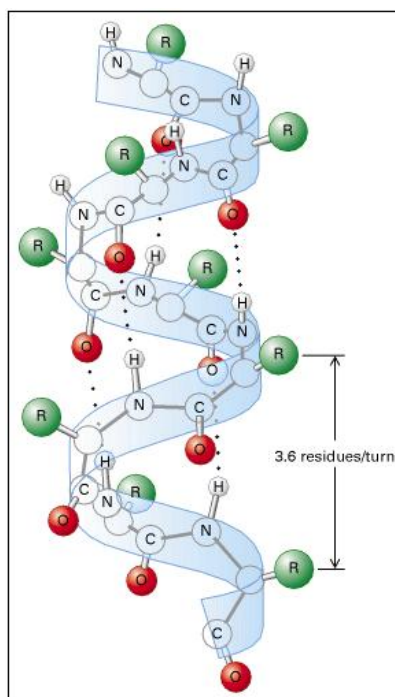


Fig. 3.2.3.a. Structura α helix a proteinelor

- o spirală cuprinde 3,6 AA
- pasul spirei 5,21 Å
- forma spiralei ca un șurub cu pasul spre dreapta
- catenele (radicalii R) sunt orientate spre exteriorul cilindrului.

D. Determinarea structurii secundare

Cea mai potrivită metodă este difracția cu raze X, (utilizată de Perutz și Kendrew pentru structura mioglobinei), radicalii X având lungimea de undă de ordinul de mărime al distanțelor interatomice ($\sim$ Å). O limitare a metodei este că poate fi folosită numai pe cristale, deci moleculele care nu pot fi aduse în formă cristalină nu pot fi studiate prin această metodă.

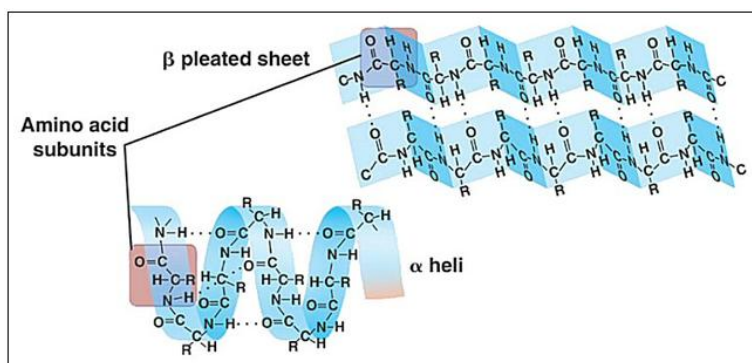


Fig. 3.2.3.b. Structura pliuri β a proteinelor

E. Fâșiile β

O altă formă întâlnită ca structură secundară este reprezentată de „fâșiile β ” sau „pliurile β ” (figura 3.2.3.b). Alternanța unor AA care generează alternanța orientării valențelor duce la formarea unor regiuni cu formă specifică de panglică pliată. Aceste fâșii apar în special în regiunile cu AA cromatici (Trp, Tyr și Phe). De asemenea, forma β -C este generată în zonele cu Ile, Val, Thr.

3.2.4. Structura terțiară a proteinelor

A. Pe lângă structura secundară, între diferitele părți ale unei molecule proteice pot apărea interacțiuni:

- legături între părți prin punți de hidrogen, legături disulfidice sau legături Van der Waals
- plieri, încovoieri.

Toate acestea generează o structură tridimensională complexă.

3.2.5. Structura cuaternară

A. În cazul moleculelor foarte mari se pot realiza chiar legături între mai multe unități, formând așa numita structură cuaternară. Procesul acesta este catalizat de enzime specifice numite „holoenzime”.

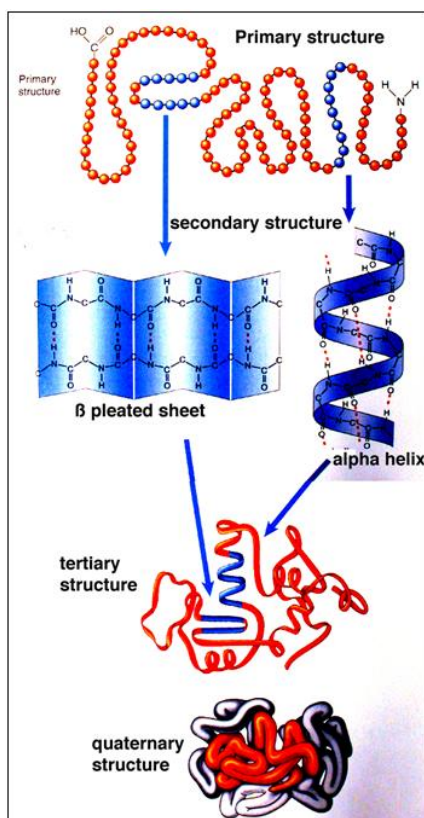


Fig. 3.2.5. Structura proteinelor – viziune de ansamblu

B. Exemple

Structură cuaternară întâlnim în special la proteinele globulare. Un exemplu hemoglobina, care are 4 subunități, ”molecula” fiind de fapt un tetramer.

În aceeași clasă mai putem include ADN polimeraza precum și proteinele componente ale canalelor ionice.

În figura 3.2.5 este prezentată o sinteză privind structura proteinelor.

3.3. Componentele acizilor nucleici

3.3.1. Componentele unui nucleotid

Acizii nucleici au și ei o structură secvențială, asemănătoare întrucâtva cu cea a proteinelor. Unitatea de secvență, echivalentă aminoacizilor, se numește „nucleotid”. Un nucleotid are 3 componente principale:

- o pentoză
- un acid fosforic
- o bază azotată

3.3.2. Pentozele din acizii nucleici

A. În acizii nucleici întâlnim două tipuri de pentoze:

- riboza - în acidul ribonucleic (ARN)
- de(z)oxiriboza - în acidul de(z)oxiribonucleic ADN

B. Structura chimică și numerotarea atomilor sunt prezentate în figura (3.2.2).

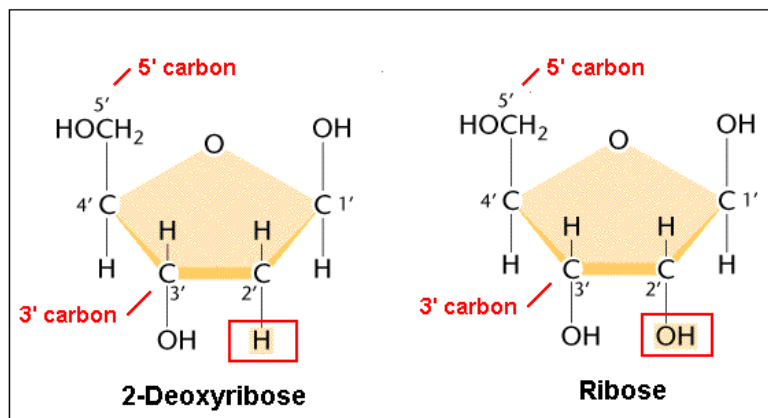


Fig. 3.2.2. Structura ribozei și dezoxiribozei

Observăm că atomii C_1' - C_4' și oxigenul formează un plan, în timp ce carbonul C_5' este înafara planului.

De asemenea, gruparea -OH de la carbonul $1'$ este de aceeași parte cu carbonul $5'$, în timp ce grupările -OH din pozițiile $2'$ și $3'$ sunt de cealaltă parte a planului.

C. În moleculele de dezoxiriboză lipsește gruparea -OH din poziția $2'$.

3.3.3. Bazele azotate

A. Fiecare tip de acid nucleic conține 4 tipuri de baze azotate: două baze purinice și două baze pirimidinice.

Bazele purinice sunt adenina (A) și guanina (G), iar cele pirimidinice sunt citozina (C) și Timina (T) sau uracilul (U). În timp ce bazele A, G și C se găsesc și în ADN și în ARN, cea de a patra este diferită în ADN întâlnim timina, în timp ce în ARN întâlnim uracilul.

B. Numerotarea atomilor din bazele azotate se face numai pentru atomii din cicluri și este redată împreună cu structura moleculelor de purină și pirimidină, precum și bazele azotate derivate: A și G, respectiv C, T și U în figura 3.3.3.

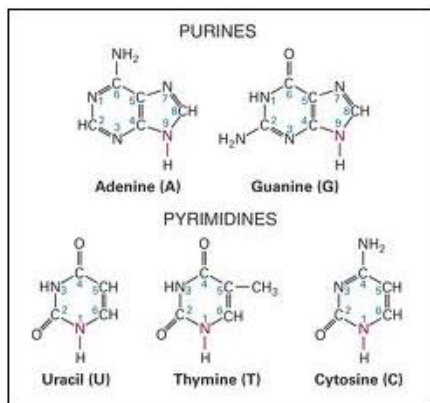


Fig. 3.3.3. Bazele azotate din acizii nucleici

3.3.4. Nucleozide

A. Bazele azotate se leagă de pentoze printr-o legătură glicozidică, formând o structură numită „nucleozid”. Denumirea nucleozidelor pornește de la numele bazei azotate; de ex de la adenină se formează adenzina (prin cuplarea de riboză) sau dezoxiadenozina prin legarea de (prin legarea de dezoxiriboză; similar avem guanină/guanozină, dezoxiguanozină, citozină/citidină/dezoxicitidină, timină/timidină (cu dezoxiriboză), respectiv uracil/uridină (cu riboză).

Legarea bazelor azotate se face întotdeauna la carbonul 1 al pentozei. Bazele purinice se leagă prin azotul din poziția 9, iar cele pirimidinice prin azotul din poziția 1.

B. Structura nucleozidelor este prezentată în figura 3.3.4.

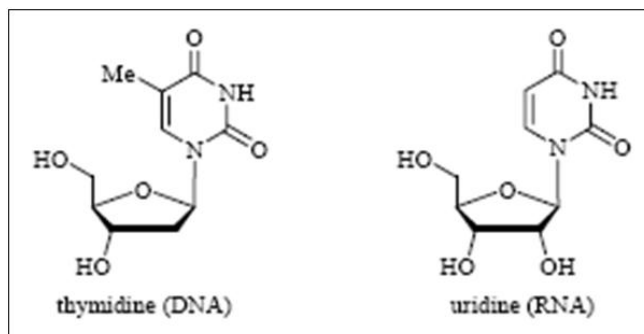


Fig. 3.3.4. Structura nucleozidelor

3.3.5. Nucleotide

Pentru formarea lanțurilor acizilor nucleici, nucleozidele sunt fosforilate. Cuplarea moleculei de acid fosforic se face la carbonul 5' al pentozei, structura nou formată fiind numită „nucleotid”. Deși nucleotidele, ca entități separate au denumiri specifice (acid adenilic, guanilic, citidilic sau timidilic, uneori sub formă de radical adenilat, guanilat, citidilat, timidilat) în cazul precizării unui nucleotid într-o secvență de acid nucleic, se folosește doar numele bazei azotate din componența sa. În figura 3.3.5. este redată structura nucleotidelor.

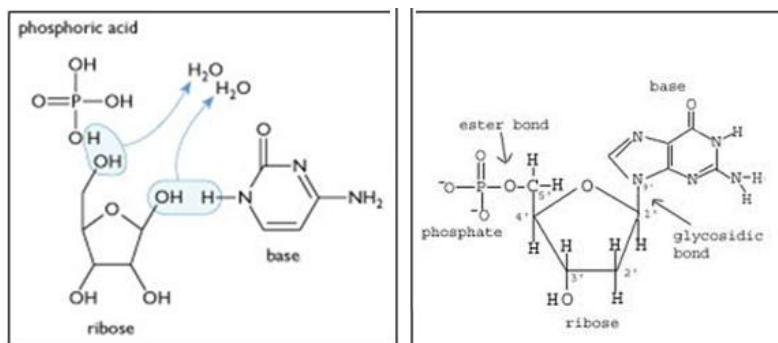


Fig. 3.3.5. Structura nucleotidelor

3.4. Structura acizilor nucleici

3.4.1. Formarea lanțului polinucleotidic

A. Nucleotidele se pot lega unele de altele prin eliminarea unei molecule de apă între o grupare - OH a fosforului legat de carbonul 5' și gruparea OH a carbonului 3' din pentoză (figura 3.4.1.a).

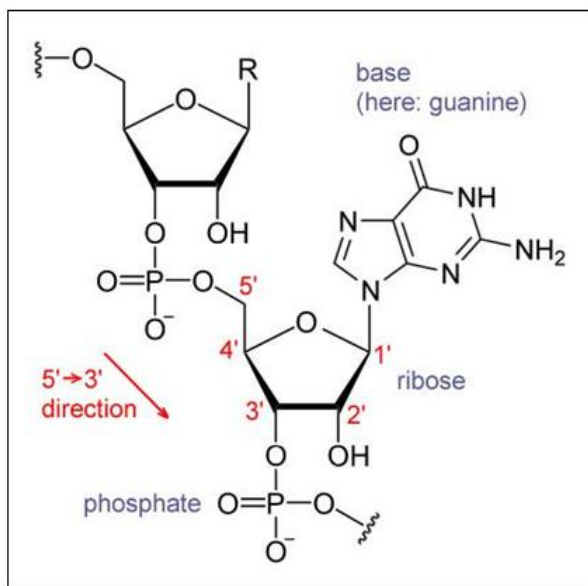


Fig. 3.4.1.a. Formarea unui dinucleotid

B. Prin repetarea procesului se poate obține un lanț care să cuprindă de la câteva zeci până la câteva mii de nucleotide (figura 3.4.1.b).

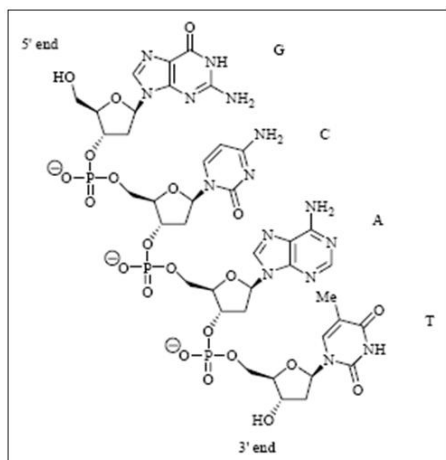


Fig. 3.4.1.b. Formarea lanțului polinucleotidic GCAT

3.4.2. Structura primară a acizilor nucleici

A. Toate lanțurile de acizi nucleici au două capete libere, unul la atomul 5', altul la atomul 3'. O secvență de lanț al unui acid nucleic se citește în ordinea 5' → 3'

B. Secvența bazelor azotate într-un lanț de acid nucleic poartă denumirea de structură primară a acizilor nucleici (figura 3.4.2).

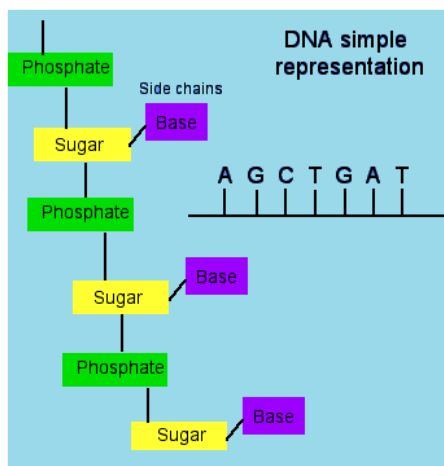


Fig. 3.4.2. Structura primară a unei secvențe ADN

3.4.3. Structura secundară a acizilor nucleici

A. Descifrarea structurii acizilor nucleici a reprezentat un pas important în înțelegerea unui întreg șir de fenomene din biologie în general și din genetică în special. Acumularea unor date experimentale privind constanța rapoartelor A/T și G/C, sau considerente teoretice privind structura tridimensională a moleculelor, au condus în cele

din urmă la stabilirea modelului „dublu-helix” al moleculei de ADN de către Watson și Crick în 1953, laureați ai premiului Nobel.

B. Moleculele de ADN formează o dublă elice înfășurată pe un cilindru virtual, având orientate spre interior bazele azotate. Între două baze azotate din cele două lanțuri se stabilesc niște punți de hidrogen. Prin configurația spațială a norului electronic sunt posibile doar perechi între o adenină de pe un lanț cu o timină pe lanțul opus (prin 2 punți de hidrogen) sau între o guanină și o citozină (prin 3 punți de hidrogen), așa cum se poate vedea în figura 3.4.3.a.

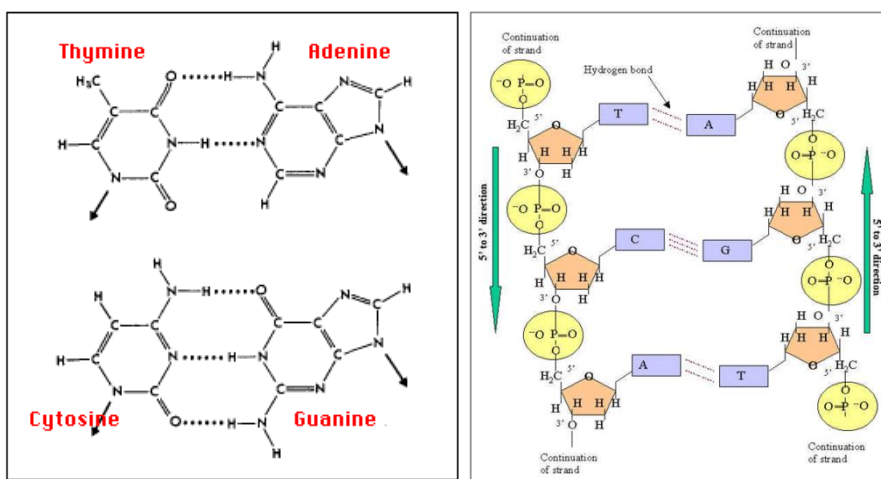


Fig. 3.4.3.a. Perechile de baze azotate în ADN

C. Proprietățile geometrice ale dublului helix au fost ulterior confirmate experimental și sunt redată sintetic în figura 3.4.3.b.

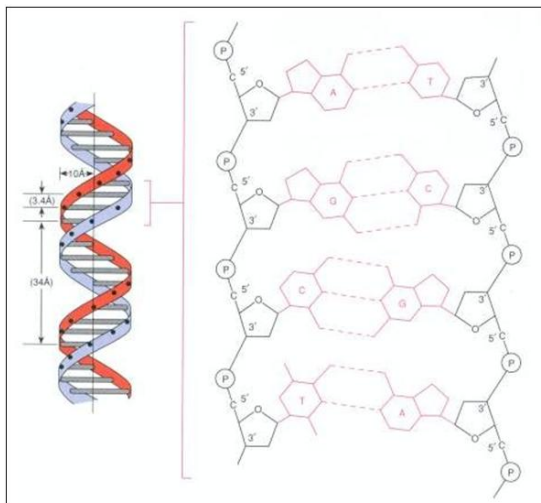


Fig. 3.4.3.b. Dublul helix ADN

D. Prezența legăturilor de hidrogen conferă o bună stabilitate moleculelor de acizi nucleici.

În moleculele de ARN, în locul timinei găsim uracilul.