

JEDAN NOVI ASPEKT TUMAČENJA EVOLUCIONOG PROCESA

— Eksponencijalni karakter evolucije proteina —

A NEW ASPECT OF INTERPRETATION OF EVOLUTIONARY PROCESS

— The exponential character of protein evolution —

M. M. RAKOČEVIĆ

Jugoslovenski bibliografski institut,
11000 Beograd

JEDAN NOVI ASPEKT TUMAČENJA EVOLUCIONOG PROCESA

— Eksponencijalni karakter evolucije proteina —

M. M. RAKOČEVIĆ

Jugoslovenski bibliografski institut,
11000 Beograd

M. M. Rakočević (1985): *A new aspect of interpretation of evolutionary process. — The exponential character of protein evolution.* Arh. biol. nauka, Beograd, 37 (1—4), 1—10, 1985.

Subject of this investigation is the framework of theoretical biology and biochemistry. It gives a new method (aspect) to interpret the molecular evolutionary process by the application of information and system theory. Proceeding from works and experimental results of King and Jukes (1969), Dickerson (1972), Nei and Chakraborty (1976), the author has given a proof for the exponential evolution process of the proteins.

UDC 575.8.01:519.72

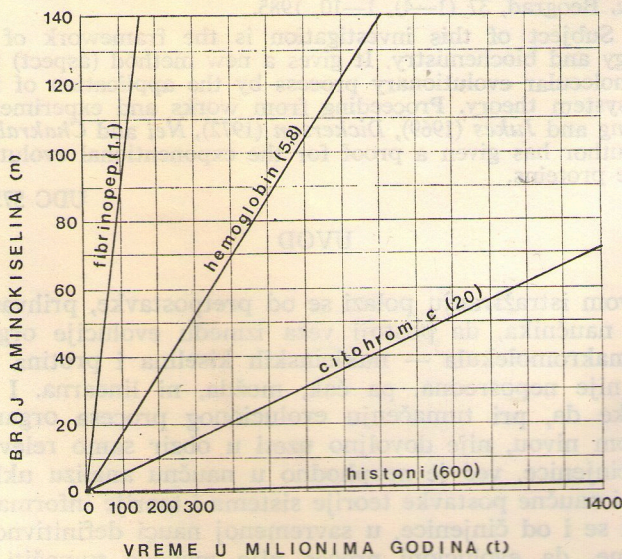
UVOD

U ovom istraživanju polazi se od pretpostavke, prihvaćene inače od većina naučnika, da postoji veza između evolucije organizama i evolucije makromolekula — nukleinskih kiselina i protina — uprkos tome što nije neposredna, pa čak, možda, ni linearna. I od druge pretpostavke da, pri tumačenju evolucionog procesa organizama na molekularskom nivou, nije dovoljno uzeti u obzir samo relevantne biohemijske činjenice, već je neophodno u naučnu analizu uključiti metodologiju i naučne postavke teorije sistema i teorije informacija. Osim toga polazi se i od činjenice, u savremenoj nauci definitivno utvrđene i prihvaćene, da evolucionni proces nije moguće tumačiti na višim hijerarhijskim morfološkim nivoima, ukoliko prethodno nije izučen na molekularnom nivou. Predmet istraživanja u ovom radu zapravo i jeste tumačenje evolucionog procesa na molekularskom nivou kroz primenu teorije sistema i teorije informacija.

Pre navođenja naučne argumentacije u prilog opravdanosti takve primene, smatramo da je neophodno najpre navesti nekoliko osnovnih postavki utvrđenih u savremenoj nauci o organskoj evoluciji. Naime, najveći broj naučnika smatra evoluciju utvrđenom činjenicom čiji su osnovni faktori mutabilnost i selekcija, tj. promenljivost i prirodno odabiranje. Primarni izvor varijabilnosti proteina, pa time i organizama je mutacioni proces koji se odigrava na sledeći način: U svakoj ćeliji višećelijskih organizama nalazi se veliki broj hromozoma; u svakom hromozomu više hiljada gena. Većina gena javlja se

u više alelnih oblika i ta raznovrsnost genskih alela, koja je rezultat mutacija, jednim delom je uzrok genetičke varijabilnosti. Mutacije same, međutim, nisu dovoljne; ne mogu obezbediti dovoljan iznos individualne raznolikosti i tu ulogu preuzimaju brojne rekombinacije.

Hemijskom analizom primarne strukture proteina utvrđujemo sastav i redosled aminokiselina u njima koji je korespondiran odgovarajućem redosledu nukleotida u molekulima DNK i RNK. Istraživanja Margolijaša i saradnika (Magoliaš 1972), ukazala su na zanimljivu činjenicu da je zavisnost proteklog evolucionog vremena i broja zamenjenih aminokiselina u homologim proteinima skoro u potpunosti linearna, mada su, pri tome, vremenske jedinice različite. Prema Dickersonu (1972) vremenska jedinica za fibrinopeptide iznosi 1,1 miliona godina, za hemoglobin 5,8; citohrom c 20, i za histone 600 miliona godina (Sl. 1). Dakle, za jednu takvu vremensku jedinicu izvrši se promena (zmena) jedne aminokiseline u proteinu.



Sl. 1. Evolucija proteina u funkciji vremena
Fig. 1. Evolution of proteins in function of time

Preliminarna razmatranja naučnog problema

U cilju tumačenja eksperimentom uočene linearne zavisnosti, postavlja se pitanje da li su fizičko-hemijske eksperimentalne činjenice dovoljne za objašnjenje promene redosleda aminokiselina u homologim proteinima različitih organizama u određenom evolucionom periodu? Sva dosadašnja naučna razmatranja pokazuju da nisu. Fizičko-hemijske karakteristike molekula su samo jedna grupa faktora koji uslovljavaju jedan, a ne neki drugi redosled. Međutim, pored njih deluju i brojni drugi faktori, kako na molekulskom, tako i

na ćelijskom i višim nivoima organizacije, koji dovode do promena, do mutacija. Definitivni ishod svih tih uticaja i svih promena jeste proces stalne, neprekidne evolucije genetičkog nukleotidnog i aminokiselinskog jezika.

Kad bi samo fizičko-hemijske karakteristike delovale, ova pravilnost se ne bi mogla očekivati iz jednostavnog razloga što se neke aminokiseline međusobno zamenjuju lakše, a neke teže. Veća je verovatnoća da se slična aminokiselina sličnom, negoli nesličnom zameni. Pa, ipak dešavaju se i ovi manje verovatni događaji i to sa istim utroškom vremena! Logično bi, međutim, bilo očekivati da će za manje verovatan događaj vremenska jedinica biti duža. Pa otkuda, onda, potiče linearnost koja se potvrđuje eksperimentom? U odgovoru na ovo pitanje, pokazaćemo, međutim, da je ta linearnost prividna, a ceo naučni problem i nastalu protivurečnost je moguće razrešiti jedino sistemskim pristupom i primenom teorije informacija.

Za dokaz polazimo od činjenice da je deoba ćelija slučajan eksponencijalni proces iz čega sledi i da je nastanak novih mutanata takođe eksponencijalnog karaktera (Kaplan 1978). S obzirom na to da sve mutacije počinju kao „tačkaste mutacije“, to znači da i promene nukleotida u molekulu DNK nastaju po eksponencijalnom zakonu. Sledeće pitanje je da li postoji linearna zavisnost i između tačkastih mutacija i broja zamenjenih aminoselina u proteinima? Čitav naučni problem shematski se ovako može predstaviti:

tačkaste mutacije —————→ $N_1 : t$

eksponencijalna zavisnost

zavisnost
linearna

broj zamena aminokiselina —————→ $A : t$

?

Ako su N_1 i A međusobno linearno zavisni, a, s druge strane, N_1 i t eksponencijalno, sledi da je i veza A sa t eksponencijalna! Linearnu zavisnost N_1 i A dokazali su eksperimentalno King i Jukes (1969), a zatim Nei i Chakraborty (1976). Time, međutim, problem nije rešen jer eksperimenat „dokazuje“ i jedno i drugo: da je veza između proteklog vremena i broja zamenjenih aminokiselina — linearna, a zatim da je ta zavisnost — eksponencijalna!?

Razlog tome što u nauci do sada nije posvećena pažnja ovom problemu (da li je proces evolucije proteina uniformnog ili eksponencijalnog karaktera) verovatno se nalazi u činjenici da je odgovor tražen samo u fizičko-hemijskim karakteristikama samih molekula, a nije posmatran proces kao takav, u svom univerzalnijem značenju. Ni Kaplan nije postavio to pitanje; on, i kada ulazi u detalje, raspravlja samo pitanja umnožavanja jedinki „po eksponencijalnom zakonu“, odnosno „geometrijskom progresijom“. Shulz i Schirmer

(1982), u svojoj inače izvanrednoj monografiji: „Principles of protein structure”, (koja je prva te vrste u svetu sa pretenzijom da obuhvati sva pitanja građe proteina i njihovih bioloških funkcija) ovo pitanje takođe skoro da i ne pominju. Doduše, na jednom jedinom mestu (op. cit. str. 212) upozoravaju da „pri upoređivanju homologih proteina koji vrše razne funkcije” ne vodi se dovoljno računa i „ne primećuje se stalnost brzine fiksacije aminokiselinskih zamena”. Naučnim problemom vremenskih jedinica u evoluciji proteina, ipak detaljnije se bavio Dickerson (1972). Ali on prvenstveno nastoji da objasni otkud različite dužine vremenskih jedinica za različite proteine, a samo na jednom mestu pokušava odgovoriti na pitanje otkud uniformnost: „Posmatrana uniformna brzina promene u citohromu c prosto znači da je biohemija respiratornog aparata mitohondrija izolovana od prirodne selekcije ... Selekcioni pritisci postaju usaglašeni na molekulskom nivou duž vremenskog razmaka od miliona godina”. Ako bi tako bilo, onda je to u suprotnosti sa njegovim tumačenjem različitih vremenskih jedinica kod različitih proteina, jer tada ne dopušta usaglašavanje selekcionih pritisaka na „molekulskom nivou”. S druge strane, evidentno je da on pokušava naći odgovor samo u okviru fizično-hemijskih karakteristika molekula i ne uviđa univerzalnu statističku i kibernetičku sistemsko-informacionu suštinu koja prvenstveno diriguje procesom uniformnosti.

Metodologija istraživanja

U istraživanju problema uniformnosti — eksponencijalnosti evolucije proteina primenili smo uobičajenu metodologiju koja se koristi u teoriji sistema i teoriji informacija. Pri tome smo skup od 20 aminokiselina, njihovih osobina i međusobnih odnosa, smatrali sistemom u smislu Bertalamffy-a (1949) i Mesarovića (1964; 1968). Osim sistemske analize primenili smo i statističku analizu na način kako se to čini u teorijskoj fizici imajući u vidu multifaktorijalni karakter evolucionog procesa.

Naučna argumentacija eksponencijalnog karaktera evolucije proteina

Da bismo stvorili uslove za interpretaciju informacionog pristupa u izučavanju evolucije proteina, neophodno je najpre ukazati na činjenicu da postoji čitav niz pojava u prirodi kao što su proticaje svetlosti kroz homogenu sredinu, protok elektrona kroz različite medije, monomolekulske hemijske reakcije, radioaktivno raspadanje, rast bakterija itd., za koje važi eksponencijalni zakon:

$$N = N_0 e^{-kt}$$

Odgovor na pitanje zašto je to tako nalazi se, kako smo dokazali u našem prethodnom radu (Rakočević 1985), u zajedničkoj, adekvatnoj i istovetnoj statističkoj suštini koja se ispoljava kroz

različitošć jedinki mase, sadržanoj u svim tim pojavama. Drugim rećima, važi isti matematićki model i odnos.* U istraživanju smo zatim postavili pitanje da li se i promene u genetićkom jeziku, nukleotidnom i aminokiselinskom, mogu svesti na isti eksponencijalni izraz, pri ćemu je, na primer, N_0 broj aminokiselina u molekulu nekog proteina na „poćetku“ evolucije, N broj aminokiselina zahvaćenih promenom, t proteklo vreme i k koeficijenat proporcionalnosti (konstanta)?

Coutelle i Hofacker (1982) nisu neposredno izućavali ovo pitanje, ali njihovi rezultati posredno se mogu dovesti u vezu sa njim. Naime, pri pokućsaju da odrede kolićinu informacije koja se dobija pri promenama sekvenci proteina u toku evolucije, vrćili su izraćunavanje preko promene entropije ΔS i verovatnoća za pojedine proteinske sastave, polazeći od sledećeg izraza:

$$N! / \prod_{i=1}^{20} N_i!$$

koji predstavlja broj razlićitih sekvenci istog sastava. Izraz je istovetan sa poznatim izrazom za skup ćestica idealnog gasa, što znaći da i ovi istraživaći polaze od primene teorije informacija i statistićeke fizike u izućavanju genetićkih fenomena na molekulskom nivou.

Osim ovog moguć je mećutim, joć generalniji pristup: „Razmotrimo verovatnoću multiplikacije sekvenci aminokiselina u nizu od 100 jedinica. S obzirom da se svaka od 20 aminokiselina moće naći u svakoj poziciji, broj mogućih razlićitih sekvenci iznosi 20^{100} “ (Dayhoff 1969). Navodeći izraćunavanje Meler-a iz 1958. godine, prema kojima u humanoj polnoj ćeliji ima oko 4×10^9 nukleotidnih parova iz ćega dalje, „proizilazi da moguć broj genetićkih darova iznosi $10^{2\ 4000\ 000\ 000}$ “, Teodosijus Dobžanski, podsećajući da ukupan broj atoma u Kosmosu ne prelazi 10^{76} , zakljućuje: „Broj genskih obrazaca potencijalno je beskonaćan, ali ... je tek beskrajno malo onih koji se uopće mogu ostvariti“ (Dobžanski 1982). Slićan primer navodi i Eigen (1973) i u rezultatu dobija 10^{130} . Konstatujući da je to mnogo, Eigen upravo u toj ćinjenici vidi delovanje prirodnog odabiranja i mogućnost primene teorije informacija.

Ako bismo ovu vrstu analize primenili na molekul citohroma c, kako smo i ućinili u naćem istraživanju (Rakoćević 1985), imali bismo 20^{104} mogućih razlićitih sekvenci. I kad bi svi „ishodi“ bili jednako verovatni, nema razloga da ne pretpostavimo, saglasno ergodićkoj teoriji, da bi se, u dovoljno dugom vremenskom periodu, svi i realizovali. Mećutim, s obzirom na ćinjenicu da na stvaranje sekvence aminokiselina i promene u njoj utiće ogromno mnoćstvo faktora, meću njima i fizićko-hemijske karakteristike samih molekula — nukleotida i aminokiselina, razumljivo je da će u toku evolucije sekvence i promene na njima doćićei relativno mali broj i to sa

*) Reć je o izomorfizmu procesa, a ne objekata!

različitim brzinama za pojedine proteine. Ali to, naravno, ništa ne menja suštinu sistemsko-informacionog pristupa prema kome se uvek polazi od verovatnoće „a priori” — od potencijalno velikog broja mogućnosti. Pri tome, u jednom misaonom eksperimentu za izračunavanje informacionog sadržaja, koji sobom nose tih 20^{104} mogućnosti, ne polazimo od informacije Hartly-evog tipa, koja nužno pretpostavlja jednaku verovatnost, već Shannon-ovog:

$$\Delta S = - \sum p_i \ln p_i$$

gde p ($i = 1, 2, 3, \dots, n$) predstavlja verovatnoću pojavljivanja svake pojedine sekvence. Kada, međutim, određujemo verovatnoću „a posteriori”, kao frekvenciju sekvenci realizovanih u proteinima svih organizama, na osnovu rezultata biohemijske analize, tada pokušavamo da utvrdimo i faktore koji su doveli pre do jednog nego do nekog drugog sekvencionog aminokiselinskog zapisa. Pri tome treba imati u vidu činjenicu da iako evolucija ide samo progresivno napred, postoji mogućnost ponavljanja nekih sekvenci; sa nekoliko uzastopnih promena može ponovo nastati sekvenca koja je prethodno već bila realizovana.

Sa ovim razmatranjima još uvek smo daleko od mogućnosti procenjivanja i dokazivanja eventualnog linearnog ili eksponencijalnog odnosa između broja zamenjenih aminokiselina, tj. promene informacionog sadržaja (entropije) i proteklog vremena u evolucionom hodu. Potrebno je uvesti još nekoliko pretpostavki, dovoljno očiglednih da ih posebno ne treba dokazivati. Proces promene (zamene) aminokiselina u proteinskim nizovima, može se smatrati događajem Bernulijevoj tipa. Uprkos tome što ima 20 aminokiselina, čitav problem se može svesti na postavljanje pitanja: ima li promene u sekvenci, ili nema? Dakle tipično za Bernulijevu šemu — događaj A , i događaj \bar{A} (\bar{A})! Otuda,

$$p^k \cdot q^{n-k} = 1$$

Naravno, odavde, zatim, nije teško izvršiti izračunavanje tako što će se uzeti u obzir svih 20 aminokiselina sa verovatnoćama promene (zamene) svake od njih, po zakonima multinomijalne raspodele. S druge strane, kako su svi ovi događaji „piktualnog” karaktera, ima smisla postaviti i pitanje da li ovde možda važi i Poisson-ova raspodela? Ali odgovor ćemo moći da dobijemo tek kada budemo raspolagali sa daleko više činjenica o genetičkom jeziku nego što nauka danas raspolaze.

Međutim, pre nastavljanja istraživanja u tom smeru, već sada je moguće ukazati na vezu procesa aminokiselinskih zamena sa procesima Markova. Naime, poznato je da se DNK može smatrati Markovljevim lancem sa memorijom drugog reda. A, s obzirom na to da ne može biti promena u aminokiselinskim sekvencama bez promene u molekulu DNK, sledi da je i proces zamene aminokiselina odnosno evolucija proteina takođe Markovljev proces.

Kao i kod lanaca Markova, za procese Markova takođe važe jednačine Čepmen-Kolmogorova:

$$p_{ij}(s+t) = \sum_k p_{ik}(s) p_{kj}(t), s, t \geq 0; i, j = 1, 2, \dots$$

Uvodeći matricu verovatnoća prelaza

$$P(t) = \| p_{ij}(t) \|, i, j = 1, 2, \dots, t \geq 0; P(0) = I$$

gde je I jedinična matrica, dobijamo

$$P(s+t) = P(s) P(t); s, t \geq 0$$

Ako sada želimo da saznamo vreme boravka sistema u stanju x , uzimamo da je ono slučajna promenljiva τ_i ($P\{\tau_i \geq 0\} = 1$) i predstavlja dužinu vremenskog intervala od trenutka kada sistem registrujemo u stanju x_i do slučajnog trenutka kada sistem prvi put napušta stanje x_i (Ivković 1976). Ako je početni trenutak u stanju $x_i, t = 0$, onda raspodelu za τ_i tražimo u obliku

$$G_i(t) = P\{\tau_i \geq t\}, t \geq 0$$

Otuda sledi:

$$G_i(s+t) = G_i(s) G_i(t); s, t \geq 0$$

Kako je ovo funkcionalna jednačina, jedinstveno rešenje koje zadovoljava uslov

$$0 \leq G_i(t) \leq 1; t \geq 0$$

(s obzirom da je $G(t)$ verovatnoća) je

$$G_i(t) = e^{-\lambda_i t}, t \geq 0$$

gde je λ_i neka pozitivna konstanta.

Ako važe pretpostavke da je evolucija proteina Markovljev proces dobija se, dakle, eksponencijalni izraz istog oblika koji smo imali za sve analizirane prirodne fenomene u kojima se promena količine u toku vremena karakteriše eksponencijalnim zakonom. Prema tome informacioni sadržaj (dobitak) ΔS i t međusobno su u linearnoj zavisnosti. Otkud, onda, eksperimentalne činjenice za koje izgleda kao da su direktno suprotne ovom zaključku, tj. da je broj izmena (zamen) aminokisilina u proteinima direktno proporcionalan proteklom vremenu, odnosno da je zavisnost *linearna* a ne *eksponencijalna*? Na prvi pogled ovdje ne mora ništa izgledati čudno. U pitanju je odnos između broja zamenjenih aminokiselina i proteklog vremena, a u pristupu preko Markovljevih procesa razmatrali smo samo odnos između promene količine informacije (entropije) ΔS i vremena t . Međutim, imajući u vidu sve što je rečeno o sekvencama kao ishodima jednog opšteg opita realizacije bar jedne od 20^{104} mogućih sekvenci (ako je u pitanju citohrom c), onda je jasno da svaka fiksacija promene aminokiseline u sekvenci, predstavlja ne samo promenu (zamenu) jedne aminokiseline, već i završetak jednog celog ciklusa događaja, jednog „statističkog obrta”, na isti način kako se to dešava i u svim drugim eksponencijalnim procesima koje smo analizirali. Prema tome, polazeći iz početnog stanja W_1 sa x mogućnosti u realnim uslovima (diktiranim uticajem mnoštva faktora) od ukupno 20^{104} teorijskih i idealnih, i dolazeći u novo stanje W_2 , sa y mogućnosti, ostvari se jedan pun *statistički obrt* sa informacionom dobiti ΔS čiji je „dar”, kako bi rekao Dobžanski, i — jedna pro-

menjena aminokiselina u sekvenci. Tu informacionu dobit možemo, onda, iskazati na uobičajeni način:

$$S_1 = k \ln W_1 \text{ (odnosno } S_1 = - \sum p_i \ln p_i \text{)}$$

$$S_2 = k \ln W_2 \text{ (odnosno } S_2 = - \sum p_j \ln p_j \text{)}$$

$$\Delta S = k \ln \frac{W_2}{W_1}; W_2 = W_1 \cdot e^{1/k\Delta S}$$

Prema tome, gore pomenuta protivurečnost je prividna i nećemo pogrešiti ako ordinatu na Sl. 1, umesto sa n (broj promenjenih aminokiselina), obeležimo sa ΔS — informacionom dobiti (broj statističkih obrta, odnosno broj „generacija“). Na ovaj način dolazimo do zaključka da je evolucioni proces, tj. promena (zamena) aminokiselina u toku vremena eksponencijalan, a ne linearan, kako bi to moglo proizilaziti iz „bukvalnog“ čitanja eksperimentalnih rezultata.

LITERATURA

- Bertalanffy, L. (1949): Zu einer allgemeinen Systemlehre, *Biologia Generalis*, **19** 114—129.
- Bertalanffy, L., Beier, W., Laue, R. (1975): *Biophysik des Fließgleichgewichts*, Vieweg, Braunschweig.
- Coutelle, R., Hofacker, G. L. (1982): Influence of selective processes on the amino acid compositions of proteins, *J. Mol. Biology*, **95**, 4.
- Dayhoff, M. O. (1969): Computer analysis of protein evolution, *Scientific American*, **1**, 221.
- Dayhoff, M. O. (1972): *Atlas of Protein Sequences and Structure*, 5. Georgetown University Medical Center, Washington, D. C.: National Biochemical Research Foundation.
- Dickerson, R. E. (1972): The structure and history of an ancient protein, *Scientific American*, **22**, 189.
- Dobzhansky, Th. (1953): *Genetics and the origin of Species*, New York.
- Dobzhansky Th. (1982): *Evolucija čovečanstva*, Nolit, Beograd.
- Eigen, M. (1973): *Samoorganizacija materii i evolucija bioloških makromolekyl*, Moskva.
- Firraro, S. (1983): The theory of self-organizing systems in physics. *Biology and psychology CENS*, Milano.
- Ivković, A. Z. *Matematička statistika*, Naučna knjiga, Beograd.
- Kaplan, R. W. (1967): *Probleme der Lebensentstehung und der frühesten Evolution*, Stuttgart
- Kaplan, R. W. (1978): *Der Ursprung des Lebens*, Stuttgart.
- King, J. L., Jukes T. H. (1969): Non-Darwinian evolution. *Science*, **164**, 788.
- King, J. L. (1976): Progress in the neutral mutation-random drift controversy, *Fed. Proc.*, **35**, 2087.
- Kolmogorov, A. N. (1956): *Foundations of the theory of probability*, New York.
- Margoliash, (1972): citirano prema Dickersonu.
- Mesarović, M. D. (1964): *Foundations for a general systems theory*, New York.
- Mesarović, M. D. (1968): *Systems theory and biology*, New York.

- Nei, M., Chakraborty, R. (1976): *J. Mol. Evol.*, **7**, 313—323.
- Rakočević, M. (1985): *Doktorska disertacija*, Univerzitet u Beogradu.
- Shannon, C. E., Weaver, W. (1949): *The mathematical theory of communication*, Univ. of Illinois, Urbana.
- Shulz G. E., Schirmer R. H. (1979): *Principles of Protein Structure*, New York
- Wiener, N. (1961): *Cybernetics*, MIT press, Cambridge.

A NEW ASPECT OF INTERPRETATION OF EVOLUTIONARY PROCESS
— The exponential character of protein evolution —

M. M. RAKOČEVIĆ
Jugoslovenski bibliografski institut,
11000 Beograd

The subject of this work is a system-informational approach to the investigation of protein evolution. Proceeding from the works of Margoliash (1972) and Dickerson (1972), the author has shown that the amino acid sequences from various species are different and that the degree of difference corresponds quite well with the distance that separates the two species on the evolutionary tree. This indicates that the evolutionary process of the proteins is uniform. How can it be? The author has shown that this uniform process is an illusion. Out of this results a contradiction between works and experimental results of Margoliash and Dickerson (1972) on one hand and King and Jukes (1969), Nei and Chakraborty (1976) on the other hand. In this work the author has shown that the contradiction can disappear, if we assume that the evolutionary process of the protein is exponential, not uniform. In order to give the proofs for his thesis, the author has been analyzed the evolution of proteins as a random process by the application of Markov chain models and the information gain (Coutelle and Hofacker 1982).