

Das Ausmaß asymmetrischer Basenverteilung in eukaryotischen Genen

R. Piechocki, W. Berg und A. Bergmann

Wissenschaftsbereich Genetik der Sektion Biowissenschaften, Martin-Luther-Universität, Halle (Saale), DDR
Deutsche Akademie der Naturforscher Leopoldina, Archiv, Halle (Saale), DDR
Abteilung Analyse und Programmierung, Organisations- und Rechenzentrum der Martin-Luther-Universität, Halle (Saale), DDR

The Extent of Asymmetric Distribution of DNA-Bases in Genes

Summary. The correlation between the evolutionary rates of proteins and frequencies of DNA-bases in the first and in the second position of corresponding codons was investigated.

Adenine in the first and guanine in the second position showed the strongest positive correlation with evolutionary rate whereas cytosine in the first as well as in the second position showed a strong negative correlation. The correlation can not be explained by a significant change in GC-content but rather by the asymmetric distribution of the base pairs. The extent of pyrimidine/purine asymmetry was assessed quantitatively.

Variations in frequencies of bases lead to changed frequencies of neighbouring bases and thus to changed interactions between adjacent bases in the genes. Slowly evolving proteins are coded by genes with a maximum of thermodynamic interactions between the adjacent bases in the codogeneous chain as well as in the complementary chain. The genes for highly evolving proteins are characterized by minima interactions between adjacent bases.

The possible relations between asymmetry and mutability of genes are discussed.

Zusammenfassung. Die Korrelation zwischen den Evolutionsraten der Proteine und den Basenhäufigkeiten in den 1. und 2. Positionen der zugehörigen DNS-Codonen wurde untersucht.

Adenin in der 1. und Guanin in der 2. Triplettsposition weisen die stärksten positiven Korrelationen mit den Evolutionsraten auf, während Cytosin sowohl in der 1. als auch in der 2. Position stark negativ korreliert. Die Korrelationen werden nicht durch verschiedene GC-Gehalte, sondern durch asymmetrische Verteilungen der Basen verursacht. Die dafür verantwortliche Pyrimidin/Purin-Asymmetrie wurde quantitativ charakterisiert.

Veränderungen in den Basenhäufigkeiten führen zu veränderten Nachbarschaftshäufigkeiten und dadurch zu unterschiedlich starken Basenwechselwirkungen innerhalb der Gene: Proteine mit einer geringen Evolutionsrate werden durch Gene codiert, deren Wechselwirkungsenergien in erster Näherung sowohl im codogenen als auch im komplementären Strang maximiert sind. Die Gene von Proteinen mit hohen Evolutionsraten sind durch geringere Wechselwirkungen zwischen benachbarten Basen charakterisiert.

Die möglichen Beziehungen zwischen Asymmetrie und Mutabilität der Gene werden diskutiert.

1. Einleitung

Seitdem das Material der Evolution bis zum molekularen Niveau analysiert werden kann, ist die Diskussion um "selektionsneutrale" Mutationen nicht mehr verstummt. Einleuchtenden Argumentationen für deren Existenz (zuerst bei Kimura 1968a; King und Jukes 1969; Arnheim und Taylor 1969; Übersicht bei Kimura und Ohta 1974) stehen detaillierte Kritiken gegenüber (Richmond 1970; Clarke 1970). Die Gründe für die Existenz und Bedeutung "selektionsneutraler" Mutationen sind auf den ersten Blick stichhaltig: 1. Nach der Entzifferung des genetischen Codes wurde gezeigt, daß durch die Degeneration des

Codes ein Großteil aller möglichen Mutationen nicht zu einer Veränderung von Aminosäuresequenzen der Proteine führt (Kimura 1968b).

2. Proteinpolymorphismen sowie Vergleiche der Proteinevolutionsraten machten deutlich, daß an sehr vielen Positionen in Proteinen Aminosäuren ohne Wirkung auf Struktur und Funktion der Proteine ausgetauscht sein können (Arnheim und Taylor 1969; Kimura und Ohta 1971).

Wenn aber ein mehr oder weniger großer Teil der Nukleotidsubstitutionen sich nicht oder nicht wesentlich auf die Struktur und Funktion der Proteine auswirkt, dann kann die Geschwindigkeit und Richtung der Evolution nicht so sehr durch Selektion bestimmt

werden, sondern - mindestens auch - durch die Fixierung solcher "selektionsneutraler" Mutationen (Laird, McCaughy und McCarthy 1969).

Kimura (1968a) entwickelte aus diesen Ansätzen seine "neutral mutation - random drift hypothesis", wofür King und Jukes (1969) den mehr provokativen Ausdruck "Non-Darwinian Evolution" wählten.

Ein Argument gegen die Neutralität der "selektionsneutralen" Mutationen war, daß Nukleotidveränderungen, die keinen Einfluß auf Struktur und Funktion der Proteine haben, durch die veränderten Basensequenzen in den Genen selbst durchaus einen wesentlichen Effekt haben könnten (Richmond 1970; Clarke 1970). Nach dem heutigen Wissen kommen dafür sowohl die Wechselwirkungen von DNS und Proteinen als auch die Transkription, Translation und Mutabilität der einzelnen Gene in Frage.

Es ist wissenschaftshistorisch interessant, daß zu Beginn der molekularbiologischen Diskussion um "neutrale" Mutationen experimentelle und theoretische Daten über die Mutabilität von Genen und über die Wechselwirkungen zwischen benachbarten Basen bekannt waren, die der Theorienbildung durchaus ein anderes Gesicht hätten geben können. Benzer und Freese (1958) und Benzer (1961) bewiesen, daß Nukleotidsubstitutionen die Mutabilität benachbarter Basensequenzen entscheidend verändern können. Anfang der 60er Jahre wurde mittels quantenmechanischer Berechnungen gezeigt, daß veränderte Nachbarschaftsbeziehungen zu unterschiedlichen Wechselwirkungsenergien der Gene führen (DeVoe und Tinoco 1962; Hoffmann und Ladik 1964). Diese sind in Dinukleotiden dann am größten, wenn gleiche Basenkomponenten übereinanderliegen, z.B. AA oder TT. Sie sind wesentlich kleiner, wenn sich eine Pyrimidin- unter einer Purin-Base befindet (Ladik 1972). Wenn auch die quantenmechanischen Berechnungen solcher komplizierten Moleküle quantitativ nicht sehr zuverlässig sind, beweisen sie mindestens qualitativ einen gewichtigen Unterschied der Wechselwirkungsenergie bei verschiedener Anordnung der Basen längs der Kette und damit eine unterschiedliche Stabilität verschiedener Nukleotidtexte (Wolkenstein 1969).

In der vorliegenden Arbeit werden die Verschiedenheiten im Asymmetriegrad der Basenverteilungen eukaryotischer Nukleotidtexte auf Grund bekannten experimentellen Materials beschrieben. Mittels stati-

stischer Analysen wurde versucht abzuschätzen, ob es sich - wie nach der "neutral mutation-random drift hypothesis" zu erwarten - um zufällige Unterschiede handelt oder um systematische. Es ergab sich, daß die Asymmetrie der Basenverteilungen in den Gensequenzen mit der Abnahme der Evolutionsgeschwindigkeiten der Proteine stark zunimmt. Da diese systematischen Veränderungen der Basenhäufigkeit veränderte Nachbarschaftsbeziehungen zur Folge haben, lassen sich konkrete Schlußfolgerungen für die Stabilität verschiedener Gene ziehen.

2. Definition und quantitative Charakterisierung der "Asymmetrie" von Basenverteilungen

Über die Variabilität des GC (Guanin + Cytosin)-Gehaltes in verschiedenen Organismengruppen liegen mehrere Untersuchungen vor (Sueoka 1961; Singer und Ames 1970). Da G und C des Einzelstranges in der Doppelhelix als GC bzw. CG Paare vorliegen, ist der Anteil an G+C stets in beiden Strängen gleich, d.h. symmetrisch. Er ist nur variabel als Gesamt-GC-Gehalt der DNS (zwischen 25 und 75%). Entsprechendes gilt für Adenin und Thymin (AT).

Bei jeder beliebigen Basenhäufigkeitsverteilung ist somit die Summe von A+T bzw. G+C im Doppel-

symmetrische a: Verteilung				asymmetrische b: Verteilung			
T	A			G	C		
T	A			A	T		
G	C			C	G		
C	G			A	T		
A	T			G	C		
C	G			A	T		
A	T			G	C		
T	A			T	A		
G	C			C	G		
A	T			A	T		
C	G			A	T		
G	C			G	C		
C	G			G	C		
G	C			C	G		

	d	c	k	d	c	k
$\frac{G+C}{A+T}$	1.3	1.3	1.3	1.3	1.3	1.3
$\frac{A+C}{G+T}$	1	1	1	1	1.3	0.75
$\frac{C+T}{G+A}$	1	1	1	1	0.4	2.5

Abb. 1. Darstellung symmetrischer und asymmetrischer Basenverteilungen anhand zweier fiktiver DNS-Segmente (je 14 Basenpaare; 6 AT und 8 GC) d = doppelsträngige DNS; c = codogener Strang; k = komplementärer Strang

strang stets identisch mit den Summen in jedem der beiden Einzelstränge.

Bezüglich der Häufigkeit einzelner Basen können aber codogener und komplementärer Strang asymmetrisch sein (Abb. 1). Bei gleicher Häufigkeit im Doppelstrang können die Einzelstränge von einer idealen Gleichverteilung (symmetrisch, Beispiel a) mehr oder weniger stark abweichen, so daß der codogene Strang z.B. purinreich wird (Beispiel b). Art und Ausmaß der Asymmetrie läßt sich durch die Quotienten $C+T/G+A$ und $A+C/G+T$ quantitativ beschreiben. Bei einer symmetrischen Verteilung der Basen (Beispiel a) sind beide Quotienten sowohl im Doppelstrang als auch in den Einzelsträngen stets 1. Ein Abweichen vom Wert 1 im codogenen und damit auch im komplementären Strang ist ein exaktes Maß für asymmetrische Basenverteilung im Gen. Überwiegt die Pyrimidin/Purin-Asymmetrie, dann weicht $C+T/G+A$ stärker von 1 ab als $A+C/G+T$, überwiegt die 6-Aminobasen/6-Ketobasen-Asymmetrie, dann ist es umgekehrt.

3. Korrelationen zwischen Basenhäufigkeiten und Evolutionsraten

In einer früheren Arbeit konnte bereits festgestellt werden, daß zwischen den geschätzten durchschnittlichen DNS-Basenhäufigkeiten in Genen und den Evolutionsraten verschiedener Proteinfamilien strenge Korrelationen existieren (Piechocki 1976).

Dabei wurde von den bekannten Aminosäuresequenzen ausgegangen (Dayhoff 1972). Die Schätzung der Basenhäufigkeit basierte wegen der Degeneration des Codes auf den Durchschnittshäufigkeiten der Basen in den jeweils möglichen Codonen. Für die ersten beiden Triplettspositionen ist mit Ausnahme von Leucin, Arginin und Serin in der 1. Position sowie Serin in der 2. Position eine eindeutige Zuordnung der DNS-Basen möglich. Bei den mehrdeutigen Zuordnungsmöglichkeiten in den genannten 1. und 2. und praktisch allen 3. Positionen wurde der Durchschnittswert bei Annahme gleicher Häufigkeit der 6,4 bzw. 2 Codonen verwendet, eine dort mögliche Selektion also vorerst nicht berücksichtigt.

In dieser Arbeit wird das Ergebnis einer detaillierteren Analyse mitgeteilt, bei der einerseits mehr Pro-

Tabelle 1. Übersicht über die für die statistischen Analysen verwendeten Proteinfamilien mit den von Dayhoff (1972) berechneten Evolutionsraten (n = Anzahl der akzeptierten Punktmutationen bei Proteinen, bezogen auf 100 Aminosäurereste und 100 Millionen Jahre)

Proteine	n
Histon IV	0.06
Glycerinaldehyd-3-Phosphat-Dehydrogenase	2
Cytochrom c	3
Insulin	4
Trypsinogen	5
Encephalitogenes Protein der Myelin-Membran	7
Tierisches Lysozym	10
Pancreas-trypsin-Inhibitor	11
Myoglobin	13
Hämoglobin-Ketten	14
Lactalbumin	25
Immunglobuline	32
Kappa-Kette C-Region	39
Kappa-Kette V-Region	33
Gamma-Kette C-Region	31
Gamma-Kette V-Region	27
Pancreas-Ribonuclease	33
Wachstumshormon	37

teinfamilien und andererseits wegen der fast eindeutigen Zuordnungsmöglichkeiten nur die 1. und 2. Triplettspositionen berücksichtigt worden sind. Soweit möglich wurden von jeder der in Tab. 1 aufgeführten Proteinfamilien mindestens 5 Vertreter für die statistische Analyse benutzt.

Das Korrelationsverhalten der DNS-Basen-Häufigkeit mit den Evolutionsraten wurde sowohl getrennt für 1. bzw. 2. Triplettspositionen als auch zusammen für 1. + 2. Positionen untersucht. Dadurch wurde deutlich, mit welcher Gewichtigkeit die einzelnen Positionen in den Codonen zum Korrelationsverhalten beitragen.

In Abb. 2 sind für die 4 Basen die Ergebnisse der Analysen dargestellt. Tabelle 2 enthält die zugehörigen Schnittpunkte der Geraden mit der Ordinate (a), die Regressionskoeffizienten ($b_{y/x}$) sowie die Korrelationskoeffizienten (r).

Aus Tab. 3 ist zu entnehmen, daß von den 6 möglichen Zweierkombinationen der Basen die Purine bzw. die Pyrimidine das stärkste Korrelationsverhalten zeigen.

Auf Grund der detaillierten Analyse des Korrelationsverhaltens der Basen in der 1. und 2. Triplettsposition kann gezeigt werden, daß die stärksten positiven Korrelationstendenzen Adenin in der 1. und Gua-

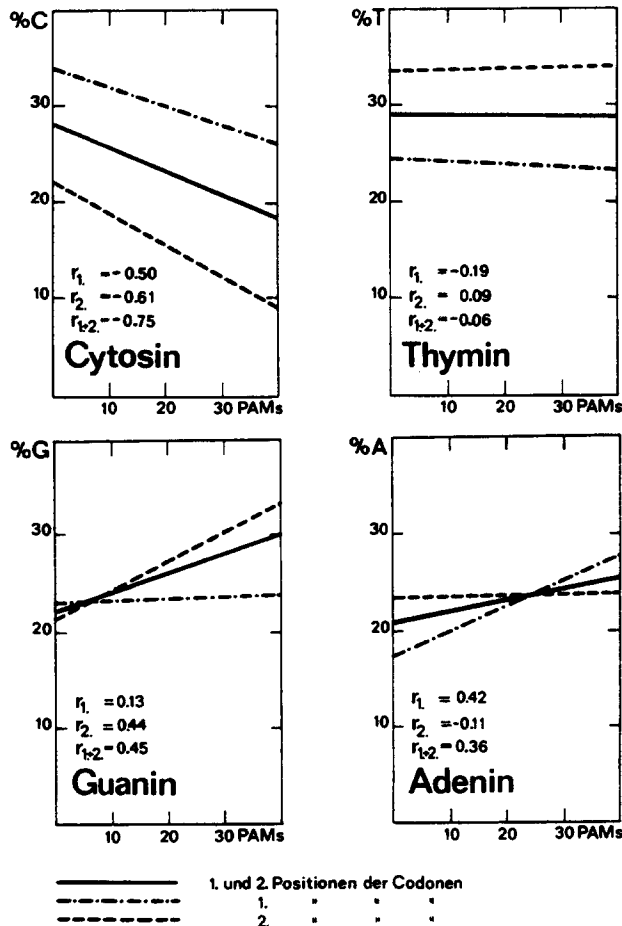


Abb. 2. Korrelation zwischen den Evolutionsraten der Proteine (PAMs) und den Häufigkeiten der 4 DNS-Basen in den entsprechenden Genen. PAM = Anzahl der akzeptierten Punktmutationen bei Proteinen, bezogen auf 100 Aminosäurereste und 100 Millionen Jahre; r = Korrelationskoeffizient

Tabelle 2. Quantitative Charakterisierung der Korrelation der DNS-Basen-Häufigkeiten mit den Evolutionsraten der Genprodukte

DNS-Base	Triplet-Position	a	$b_{y/x}$	r
Cytosin	1.	33,96	-0,174	-0,497
	2.	22,38	-0,311	-0,613
	1.+2.	28,18	-0,242	-0,747
Thymin	1.	25,82	-0,082	-0,187
	2.	32,19	+0,035	+0,092
	1.+2.	29,03	-0,023	-0,064
Adenin	1.	17,49	+0,202	+0,421
	2.	23,49	-0,003	-0,112
	1.+2.	20,47	+0,101	+0,364
Guanin	1.	22,71	+0,050	+0,132
	2.	21,88	+0,276	+0,447
	1.+2.	22,30	+0,164	+0,454

a - Schnittpunkt mit der y-Achse,
 $b_{y/x}$ - Regressionskoeffizient,
 r - Korrelationskoeffizient

Tabelle 3. Quantitative Charakterisierung der Korrelation der 6 möglichen Basenkombinationen mit den Evolutionsraten der Genprodukte (Symbole wie in Tab. 2)

DNS-Basen		a	$b_{y/x}$	r
Purine	G + A	42,78	+0,262	+0,592
Pyrimidine	C + T	57,19	-0,261	-0,594
6-Ketobasen	G + T	51,33	+0,129	+0,467
6-Aminobasen	A + C	48,65	-0,129	-0,468
AT-Basenpaare	A + T	49,54	+0,084	+0,203
GC-Basenpaare	G + C	50,48	-0,086	-0,211

Tabelle 4. Quantitative Charakterisierung des unterschiedlichen Korrelationsverhaltens von Adenin und Guanin in den ersten und zweiten Positionen der Codonen (Symbole wie in Tab. 2)

DNS-Basen in		a	$b_{y/x}$	r
1. Position	2. Position			
A + G		40,19	+0,252	+0,440
	A + G	45,38	+0,272	+0,509
A	G	39,38	+0,477	+0,531
G	A	46,21	+0,046	+0,083

nin in der 2. Position zeigen, was allein auf Kosten von Cytosin geschieht, das sowohl in der 1. als auch in der 2. Position stark negativ korreliert (Tab. 2). Um diese auffallende Tatsache weiter zu belegen, wurde das Korrelationsverhalten der Kombination von A + G in den ersten beiden Triplettspositionen untersucht (Tab. 4). Daraus geht hervor, daß in der ersten Position Adenin und in der zweiten Position Guanin antagonistisch zum Cytosin wirken.

Diese Kombination korreliert am stärksten mit den Evolutionsraten der Proteine, während die umgekehrte Kombination (1. Position G + 2. Position A) überhaupt keine systematischen Veränderungen anzeigt (vgl. die beiden letzten Zeilen von Tab. 4).

4. Art und Ausmaß der Asymmetrie

In Abb. 1 wurde demonstriert, daß die Quotienten $C + T / G + A$ und $A + C / G + T$ geeignet sind, das Ausmaß der Asymmetrie zu charakterisieren. Die Abweichungen von 1, die für den codogenen Strang berechnet wurden, sind quantitativer Ausdruck asymmetrischer Basenverteilung.

In Abb.3 ist für die Basenhäufigkeit der 1. + 2. Positionen das Ausmaß der Asymmetrie dargestellt. Für den Quotienten $C+T/G+A$ ergeben sich systematische Abweichungen von einer Zufallsverteilung, während sich die Werte für den Quotienten $A+C/G+T$ um den für eine zufällige Verteilung zu erwartenden Wert 1 gruppieren.

Um die systematischen Abweichungen zu quantifizieren, wurde mit Hilfe eines spezifischen Iterationsverfahrens der Punktwolke eine Exponentialfunktion $y = a_0 + a_1 e^{\gamma x}$ nach der Methode der kleinsten Quadrate angepaßt.

Daraus ist abzuleiten, daß für die beobachteten Korrelationen die Pyrimidin/Purin-Asymmetrie verantwortlich ist.

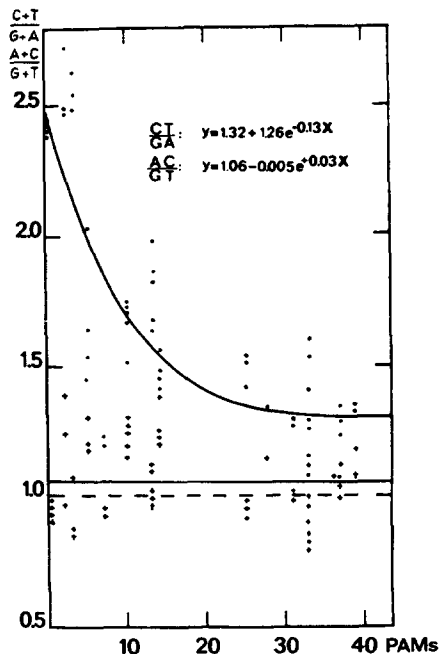


Abb.3. Darstellung der Beziehung zwischen Evolutionsraten der Proteine (PAMs, vgl. Abb.2) und dem Ausmaß der Asymmetrie in den entsprechenden Genen. CT/GA für Pyrimidin/Purin-Asymmetrie (Punkte und durchgezogene Kurve); AC/GT für 6-Amino/6-Ketobasen-Asymmetrie (Kreuze und gestrichelte Linie)

5. Diskussion

Es konnte gezeigt werden, daß zwischen den Evolutionsraten von Proteinen und dem Ausmaß asymmetrischer Basenverteilung in den entsprechenden Genen ein systematischer Zusammenhang besteht. Je niedri-

ger die Evolutionsraten der Proteine sind, um so stärker ist die Asymmetrie ausgeprägt. Die Proteine mit geringer Evolutionsrate werden von Genen codiert, die relativ viel C in den 1. und 2. Triplettspositionen enthalten, welches mit zunehmender Evolutionsrate in der 1. Position vorwiegend durch A, in der 2. Position vorwiegend durch G ersetzt ist. Dieses Phänomen ist gleichbedeutend mit einer Abnahme gleichartiger Basen in der 1. und 2. Position: während die Gene niedrig evolvierender Proteine häufiger C C (und im komplementären Strang G G) aufweisen, sind diejenigen hoch evolvierender Proteine häufiger durch A in der 1. und G in der 2. Position ausgezeichnet.

Mittels quantitativer Daten über die energetischen Wechselwirkungen bei unterschiedlichen DNS-Basen-Nachbarschaften in Dinukleotiden (Wolkenstein 1969) lassen sich in erster Näherung die Auswirkungen auf die Basenwechselwirkungen in diesen Genen berechnen. Die exakten Werte sollen in einer folgenden Arbeit mitgeteilt werden. Aus ihnen ergibt sich, daß die unterschiedlichen Korrelationstendenzen der einzelnen Basen bei Proteinen mit niedriger Evolutionsrate zu maximalen, bei solchen mit hoher Evolutionsrate zu minimalen Basen-Wechselwirkungen in den korrespondierenden Genen führen. Ursache dafür ist die hier beschriebene Tatsache, daß mit zunehmender Evolutionsrate die Wahrscheinlichkeit für gleichartige Basen in 1. und 2. Position abnimmt (vgl. Einleitung). Dabei muß berücksichtigt werden, daß sich diese erheblichen Unterschiede ohne die 3. Triplettsposition, also nur mit einem Drittel der unmittelbaren Nachbarschaftsbeziehungen, ergeben.

In der Einleitung wurde darauf hingewiesen, daß der Nachweis der Wirkungslosigkeit einer Mutation auf Struktur und Funktion von Proteinen allein nicht ausreicht, sie für selektionsneutral zu erklären, solange der Einfluß auf die Interaktion von DNS und Proteinen, auf die Transkriptionseffektivität, die Translation und die Mutabilität der Gene nicht bekannt ist. Mit anderer Methodik wurde kürzlich gezeigt, daß auch in den DNS-Sequenzen der Gene für Cytochrom c die Pyrimidin/Purin-Asymmetrie mit steigender Organisationshöhe der Organismen zunimmt (Mazin 1975). Eine ausführliche Diskussion über die Bedeutung der aufgefundenen Asymmetrien für eine oder mehrere dieser Ebenen erscheint ohne Kenntnis der vollständigen DNS-Sequenzen verfrüht. Die Problematik soll aber

wenigstens an möglichen Konsequenzen für die Mutabilität der Gene skizziert werden.

Daß verschiedene Gene unterschiedlich stark mutieren (Stadler 1930), daß sogar verschiedene Allele eines Genes bedingt durch ihre unterschiedliche Genstruktur unterschiedliche Mutationsraten aufweisen (Timoféeff-Ressovsky 1932), wurde bereits zu Beginn der experimentellen Mutationsforschung demonstriert. Dreißig Jahre später konnte auf molekularem Niveau gezeigt werden, daß die Genstruktur einen entscheidenden Einfluß auf die Mutabilität haben kann. Koch (1971) bewies experimentell, daß eine Nukleotidsubstitution die Mutabilität der Nachbarbase um ein Vielfaches (bis zu 23fach) verändern kann.

Es sollte deshalb unterschieden werden, daß eine Mutation (Basensubstitution)

a) funktionsneutral sein kann, wenn sie zum Austausch einer Aminosäure führt, die Gesamtstruktur und Funktion des Proteins aber nicht nachweisbar verändert,

b) strukturneutral sein kann, wenn sie zu einem Codon für die gleiche Aminosäure führt.

In keinem Fall ist aber bis dahin erwiesen, daß sie wirklich

c) selektionsneutral ist.

Zumindest eine grobe Selektion der Mutabilität muß a priori erwartet werden, da sowohl zu hohe Mutabilität (wegen der Mutationsbelastung der Population) als auch zu niedrige (wegen eventueller Beeinträchtigung der genetischen Variabilität) nachteilig wäre (Dobzhansky 1970).

Dennoch wird die Mutabilität nicht als ratendeterminierender Faktor der Evolution angesehen. Unterschiedliche Evolutionsraten von Proteinen werden ausschließlich durch die Selektion am Phänotyp verstanden und zwar so, daß Veränderungen an sehr alten und funktionsoptimierten Proteinen fast nie, an jungen und nicht mit der gesamten Aminosäurekette funktionsspezifischen Proteinen aber häufiger toleriert werden können.

Die oben beschriebenen systematischen Veränderungen im Ausmaß der Asymmetrie in den Basensequenzen lassen sich aber dadurch nicht erklären. Zwar wären unterschiedliche Asymmetriegrade in Genen verschiedener Proteine durch zufällige Fixierung funktions- oder strukturneutraler Mutationen möglich, sie wären dann aber auch zufällig über das ge-

samte Spektrum der Proteinfamilien verteilt zu erwarten, nicht in Korrelation mit den Evolutionsraten.

Die Gene von Histonen, Glycerinaldehyd-3-Phosphat-Dehydrogenase und Cytochrom c weisen die stärksten Asymmetrien und dadurch eine Maximierung der Wechselwirkungen zwischen den Basen auf. Dafür können zwei Hypothesen diskutiert werden:

1. Selektion auf niedrige Mutabilität. Diese drei Proteine sind an biochemischen Reaktionen beteiligt, die für alle Eukaryoten nahezu identisch und essenziell sind. Es handelt sich um sehr alte Proteine, die offenbar in Funktion und Struktur weitgehend optimiert sind. Daher ist es äußerst unwahrscheinlich, daß neu auftretende Mutationen in den entsprechenden Genen zu einer weiteren Optimierung beitragen können. Bestimmte Prozesse der Genregulation, der Atmungskette und des Citratzyklus sind so fundamental wichtig und angepaßt, daß jede neue Mutation, die diese Funktionen betrifft, zu einer Belastung der Population führen würde. Eine Reduktion der spontanen Mutabilität durch selektierte Maximierung der Wechselwirkungen zwischen benachbarten Basen könnte zu einer Verminderung der Mutationslast einer Population führen.

2. Präferentielle Mutabilität. Es ist zu erwarten, daß die Stärke der Wechselwirkungen benachbarter Basen deren Austauschbarkeit beeinflußt. Das hätte zur Folge, daß sich durch Mutationen über lange Zeiträume im Rahmen des biochemisch-funktionell Möglichen (phänotypische Selektion) immer stabilere Basenzusammensetzungen ergeben, solange keine Selektion dagegen wirkt. Das wäre um so stärker ausgeprägt, je älter die Proteine sind.

Besonders Hypothese (2) enthält relativ strenge Erwartungen für die Basen der 3. Triplettspositionen, deren bedingte Wahrscheinlichkeiten im Zusammenhang mit den vorausgehenden und folgenden (2. und 1.) Basen aber erst an vollständig sequenzierten Genen überprüft werden können.

Welche theoretischen Deutungen diese asymmetrischen Basenverteilungen schließlich auch erfahren werden, sie stützen die Aussage (Piechocki 1976), daß die Häufigkeit der einzelnen Aminosäuren in den Proteinen nicht und auch nicht vorwiegend als Zufallsverteilungen entsprechend der Anzahl ihrer Codonen zu verstehen ist. Daß diesbezügliche Kalkulationen zu solchem Ergebnis kamen und es als Argument für

Theorien "neutraler" Mutationen verwendet wurde, erklärt sich daraus, daß Proteine mit niedriger und hoher Evolutionsrate entgegengesetzt von einer Zufallsverteilung abweichen und somit der Durchschnittswert aller Proteinfamilien zusammengenommen eine solche vortäuscht (vgl. King und Jukes 1969). Bei umfangreicheren Analysen lassen sich signifikante Abweichungen einzelner Aminosäuren auch durch die Mittelwertbildung über statistisch heterogene Proteinfamilien nicht mehr vermeiden (vgl. Jukes et al. 1975), ohne allerdings den zugrunde liegenden systematischen Zusammenhang erkennen zu lassen.

Literatur

- Arnheim, N.; Taylor, Ch.E.: Non-Darwinian evolution: Consequences for neutral allelic variation. *Nature* **223**, 900-903 (1969)
- Benzer, S.: On the topography of the genetic fine structure. *Proc. Natl. Acad. Sci. U.S.* **47**, 403-415 (1961)
- Benzer, S.; Freese, E.: Induction of specific mutations with 5-bromouracil. *Proc. Natl. Acad. Sci. U.S.* **44**, 112-119 (1958)
- Clarke, B.: Darwinian evolution of proteins. *Science* **168**, 1009-1011 (1970)
- Dayhoff, M.O. (Ed.): *Atlas of Protein Sequence and Structure*. Vol. 5. Washington D.C.: National Biomedical Research Foundation 1972
- Dobzhansky, T.: *Genetics of the Evolutionary Process*. New York, London: Columbia University Press 1970
- Hoffmann, T.A.; Ladik, J. (1964) zit. nach Ladik 1972
- Jukes, T.H.; Holmquist, R.; Moise, H.: Amino acid composition of proteins: Selection against the genetic code. *Science* **189**, 50-51 (1975)
- Kimura, M.: Evolutionary rate at the molecular level. *Nature* **217**, 624-626 (1968a)
- Kimura, M.: Genetic variability maintained in a finite population due to mutational production of neutral and nearly neutral isoalleles. *Genet. Res.* **11**, 247-269 (1968b)
- Kimura, M.; Ohta, T.: Protein polymorphism as a phase of molecular evolution. *Nature* **229**, 467-469 (1971)
- Kimura, M.; Ohta, T.: On some principles governing molecular evolution. *Proc. Natl. Acad. Sci. U.S.* **71**, 2848-2852 (1974)
- King, J.L.; Jukes, T.H.: Non-Darwinian evolution. *Science* **164**, 788-798 (1969)
- Koch, R.E.: The influence of neighboring base pairs upon base-pair substitution mutation rates. *Proc. Natl. Acad. Sci. U.S.* **68**, 773-776 (1971)
- Ladik, J.: *Quantenbiochemie für Chemiker und Biologen*. Stuttgart: F. Enke 1972
- Laird, Ch.D.; McConaughy, B.L.; McCarthy, B.J.: Rate of fixation of nucleotide substitutions in evolution. *Nature* **224**, 149-154 (1969)
- Mazin, A.L.: Evolution of DNA structure: Direction, mechanism, rate. (russ.) *Molekuljarnaja Biologija* **9** (2), 252-274 (1975)
- Piechocki, R.: Correlation between estimated average frequencies of DNA-bases of genes and evolutionary rates of corresponding proteins. *Studia biophysica* **58**, 51-62 (1976)
- Richmond, R.C.: Non-Darwinian evolution: A critique. *Nature* **225**, 1025-1028 (1970)
- Singer, C.E.; Ames, B.N.: Sunlight ultraviolet and bacterial DNA base ratios. *Science* **170**, 822-826 (1970)
- Stadler, L.J.: The frequency of mutation of specific genes in maize. (Abstract) *Anat. Rec.* **47**, 381 (1930)
- Sueoka, N.: Correlation between base composition of deoxyribonucleic acid and amino acid composition of protein. *Proc. Natl. Acad. Sci. U.S.* **47**, 1141-1149 (1961)
- Timoféeff-Ressovsky, N.W.: Verschiedenheit der "normalen" Allele der white-Serie aus zwei geographisch getrennten Populationen von *Drosophila melanogaster*. *Biol. Zbl.* **52**, 468-476 (1932)
- DeVoe, H.; Tinoco, I. (1962) zit. nach Wolkenstein 1969
- Wolkenstein, N.W.: *Moleküle und Leben*. Leipzig: G. Thieme 1969

Received 31.1.77

Communicated by R. Hagemann

R. Piechocki
Wissenschaftsbereich Genetik
Sektion Biowissenschaften
Martin-Luther-Universität
Halle-Wittenberg
DDR-401 Halle/Saale
Domplatz 1
German Dem. Rep.