

**В. А. Каргинов, С. Я. Головин, А. А. Бондарь, И. В. Морозов,
В. М. Блинов, Н. П. Мертвецов**

СРАВНИТЕЛЬНЫЙ АНАЛИЗ ПЕРВИЧНЫХ СТРУКТУР ДНК, КОДИРУЮЩИХ ПРООИОМЕЛАНКОРТИНЫ РЯДА ЖИВОТНЫХ И ЧЕЛОВЕКА

В работе представлен сравнительный компьютерный анализ секвенированных генов и известных нуклеотидных последовательностей, кодирующих проопиомеланокортины (ПОМК) крысы, мыши, норки, свиньи, быка, лягушки, лосося и человека. Рассмотрены эволюционные характеристики нуклеотидных последовательностей (консервативность, вариабельность отдельных участков). Сделано предположение, что различия в степени вариабельности отдельных доменов ПОМК могут объясняться особенностями структурной организации кодирующих эти домены участков ДНК (наличие в вариабельных участках прямых и инвертированных повторов, высокий процент GC-пар). Предложен механизм возникновения мутаций в вариабельных участках нуклеотидной последовательности, кодирующей ПОМК. Построено древо эволюционного родства ПОМК проанализированных видов животных.

Проопиомеланокортин (ПОМК) является белком-предшественником ад-решокартотропина (АСТН)*, β -липотропина (β -LPH)* и ряда элипсоидных пептидов. Продукты процессинга ПОМК влияют на такие физиологические функции, как защита организма от вредных воздействий, модуляция вегетативной нервной системы, нейро-эндокринные функции, поведенческие реакции [1].

Ген ПОМК является также удобной моделью для изучения структурно-функциональной организации и закономерностей экспрессии генов эукариот, регулируемых стероидными гормонами. Ранее нами [2—5] и другими авторами [6—15] были клонированы и секвенированы кДНК и фрагменты геномной ДНК, кодирующие ПОМК ряда животных и человека.

В настоящей работе представлен сравнительный анализ известных нуклеотидных последовательностей ДНК, кодирующих ПОМК семи видов животных и человека, для выявления особенностей их структурно-функциональной организации и построения древа эволюционного родства. Все расчеты осуществлялись на ПЭВМ «Apple II» при помощи специально разработанного нами пакета программ. При сравнении аминокислотных последовательностей степень сходства оценивали по [18]. Для построения древа эволюционного родства мРНК ПОМК использовали метод [19]. Вторичную структуру ДНК моделировали с применением подхода [24].

На первом этапе мы провели выравнивание аминокислотных последовательностей ПОМК, показавшее, что у всех восьми видов позвоночных биологически активные домены в белке-предшественнике расположены сходным образом и фланкированы с обоих концов парами положительно заряженных аминокислот Lys-Lys, Arg-Arg, Lys-Arg (рис. 1). Показано, что эти пары выполняют роль сайтов узнавания специфических протеаз, участвующих в процессинге ПОМК [23]. Из рис. 1 видно также, что степень консервативности аминокислотных последовательностей неравномерна по всей длине белка-предшественника и может сильно различаться на разных участках ПОМК, причем границы консервативных и вариабельных участков четко совпадают с таковыми биологически активных доменов ПОМК. Из этого следует, что эволюционные характеристики аминокислотной последовательности ПОМК (консервативность, вариабельность отдельных участков) коррелируют с ее функциональной топографией. Так, например, высокой консервативностью отличаются участки аминокислотной последовательности ПОМК, соответствующие α - и β -меланотропинам (β -MSH)*, β -эндорфину. Участок между кортикопипоподобным пептидом (CLIP)* и β -MSH, напротив,

* Здесь и далее отмеченные аббревиатуры даны аналогично указанным на рисунках.

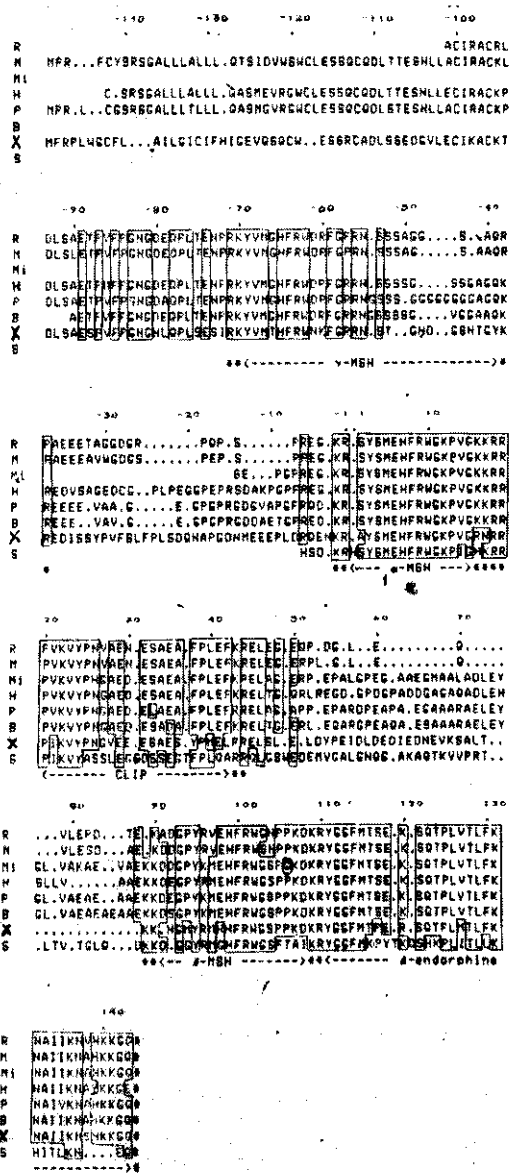
очень вариабелен и содержит многочисленные делеции, составившие у крысы и мыши почти половину длины этого района по сравнению с аналогичным районом ПОМК порки, человека и др.

Различия в вариабельности данных районов аминокислотной последовательности могут быть вызваны не только их отличающейся функциональной значимостью и влиянием естественного отбора, но и зависимостью частоты возникновения мутаций от структурных особенностей участков ДНК, кодирующих эти районы. С целью выявления таких особенностей нами было проведено выравнивание и сравнительный анализ нуклеотидных последовательностей, кодирующих ПОМК (рис. 2). Как и следовало ожидать, на уровне ДНК наблюдаются те же закономерности в распределении консервативных и вариабельных участков, что и на уровне аминокислотных последовательностей.

В процессе определения нуклеотидной последовательности кДНК ПОМК порки [5]

Рис. 1. Выровненные аминокислотные последовательности ПОМК крысы (*R*) [12], мыши (*M*) [13], порки (*Mi*) [5], человека (*H*) [3], свиньи (*P*) [9], быка (*B*) [4], лягушки (*X*) [15] и лосося (*S*) [14]. Нумерация аминокислотных остатков начинается с 1-го аминокислотного остатка α-MSH. Стрелками указано расположение биологически активных доменов ПОМК. Блоками выделены аминокислоты, идентичные не менее чем у шести проанализированных структур ПОМК.

Fig. 1. Alignment of the amino acid sequences of the rat (*R*) [12], mouse (*M*) [13], mink (*Mi*) [5], human (*H*) [3], pig (*P*) [9], bovine (*B*) [4], frog (*X*) [15] and salmon (*S*) [14] POMCs. Numbering of amino acid residues begins from the 1st amino acid residue α-MSH. The amino acids identical at least in six POMC sequences are boxed. Biologically active peptides POMCs are indicated by arrows



мы обратили внимание на значительное количество «компрессий» на радиоавтографах секвенирующих гелей в участках, соответствующих сильно вариабельному району; появление их обычно объясняют наличием шпилечных структур в анализируемой ДНК [20]. Для проверки этого предположения было проведено исследование структурных особенностей вариабельного участка ДНК, отличающих его от консервативных. Во-первых, он оказался насыщен прямыми повторами длиной 6—7 и более пар оснований (рис. 2) и, во-вторых, здесь наблюдается высокое содержание GC-пар (75—80 % против 60—65 % в среднем по другим районам). Известно, что прямые повторы играют определенную

роль в механизме возникновения делеций и дупликаций в процессе репликации и репарации ДНК [21]. При этом высокое содержание GC-пар должно увеличивать стабильность вторичных структур, необходимых для реализации этих механизмов.

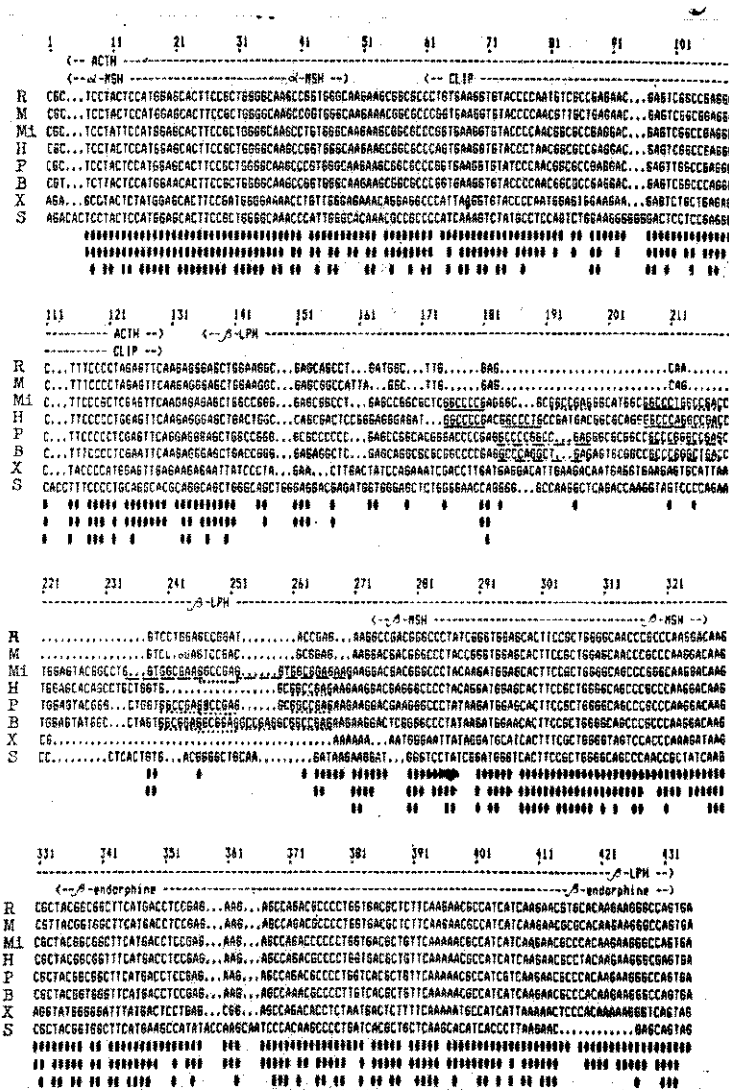


Рис. 2. Выравненные нуклеотидные последовательности ДНК, кодирующие ПОМК у рассматриваемых видов животных. Тремя звездочками отмечены позиции, для которых наблюдается совпадение нуклеотидов у всех восьми последовательностей; двумя — семь из восьми; одной — шесть нуклеотидов. Стрелками указаны участки ДНК, кодирующие биологически активные домены ПОМК

Fig. 2. Alignment of POMC nucleotide sequences of the species investigated: *** denote nucleotide positions identical for all eight sequences; ** — for seven sequences; * — for six sequences. Nucleotides coding for biologically active peptides POMC are indicated by arrows

Кроме того, в варибельном районе обнаружено значительное количество обратимых повторов, способных приводить к образованию шпильчатых структур, следствием чего может являться увеличение вероятности появления делеций, инсерций и точечных мутаций в соответствии с моделью Рипли [22].

Чтобы проверить последнее предположение, мы построили возможную вторичную структуру цепи ДНК, кодирующей ПОМК норки, и от-

метили на ней наиболее переменные основания (рис. 3, а). Видно, что переменный участок располагается в высокоструктурированном районе, причем наибольшее число отличий приходится на правую часть крестообразной структуры. Такая закономерность позволила нам предположить следующий возможный механизм возникновения мутаций

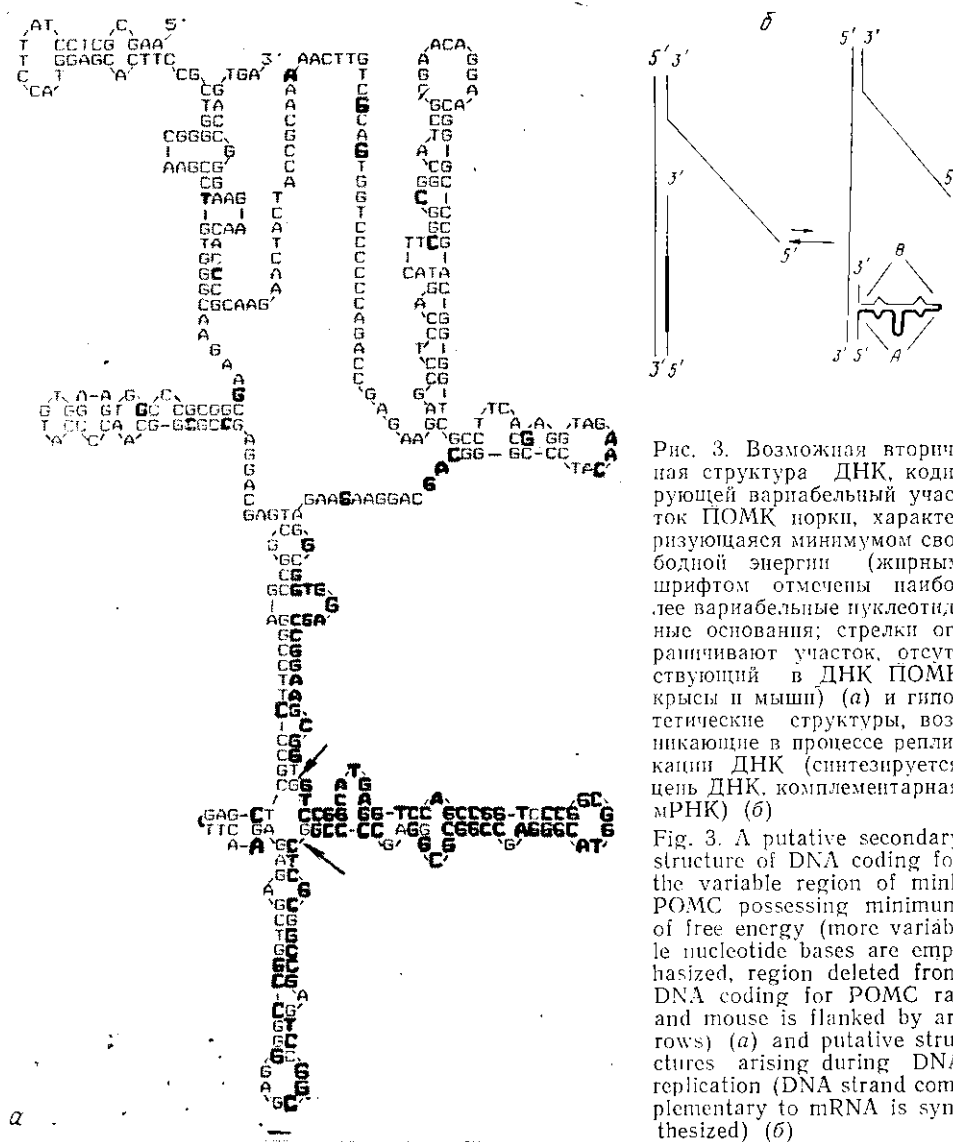


Рис. 3. Возможная вторичная структура ДНК, кодирующей переменный участок ПОМК норки, характеризующаяся минимумом свободной энергии (жирным шрифтом отмечены наиболее переменные нуклеотидные основания; стрелки ограничивают участок, отсутствующий в ДНК ПОМК крысы и мыши) (а) и гипотетические структуры, возникающие в процессе репликации ДНК (синтезируется цепь ДНК, комплементарная мРНК) (б)

Fig. 3. A putative secondary structure of DNA coding for the variable region of mink POMC possessing minimum of free energy (more variable nucleotide bases are emphasized, region deleted from DNA coding for POMC rat and mouse is flanked by arrows) (a) and putative structures arising during DNA replication (DNA strand complementary to mRNA is synthesized) (b)

(рис. 3, б). В соответствии с моделью Рипли [22] допускается образование в процессе репликации структур, подобных изображенной на рис. 3, б. Ферменты репарации, двигаясь в направлении $5' \rightarrow 3'$ вдоль новосинтезированной цепи, «проверяют» ее комплементарность материнской цепи, репарируя неспаренные участки. В районе А эти ферменты могут провести коррекцию нуклеотидной последовательности, приведя ее в соответствие с «псевдородительской» частично комплементарной цепью (районом В).

Таким образом, в районе А будут репарироваться участки, неп полностью комплементарные району В, следствием чего может быть возникновение мутаций в этих участках. В пользу возможности осуществления подобного механизма свидетельствует тот факт, что в ДНК, кодиру-

ющих ПОМК крысы и мыши, делегирован район, точно совпадающий с одной из шпилечных структур (рис. 2 и 3, а).

Можно предположить, что отмеченные нами структурные особенности ДНК и предложенный механизм могут реализоваться и в других случаях, когда наблюдается относительно высокая вариабельность определенных участков генома.

Для выяснения эволюционного родства рассматриваемого семейства мРНК ПОМК (крысы, мыши, норки, человека, быка, свиньи, лягушки

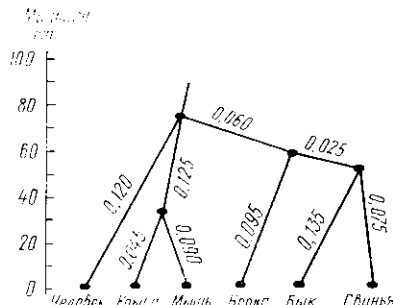


Рис. 4. Филогенетическое дерево эволюционного родства мРНК ПОМК млекопитающих и человека. Цифры на ребрах графа характеризуют эволюционное расстояние между двумя узлами графа — близкородственными ДНК ПОМК, которое приведено в числе синонимичных замен на синонимичный сайт

Fig. 4. A tree of evolutionary relations of the POMC mRNAs from mammals and a man. Figures on the tree indicating evolutionary distance between two species are given in the number of synonymous replacements per a synonymous site

и лосося) выбрали участок нуклеотидных последовательностей, кодирующий АСТН и β -LPH. Сравнение проводили по методу [19], анализируя все возможные пары мРНК. Корректный сравнительный анализ на более узких участках мРНК, кодирующих, например, α - и β -MSH, CLIP и т. д., оказался невозможным из-за ограничений метода. Высокий уровень отличий в синонимичных сайтах нуклеотидных последовательностей не позволил нам связать общими предками мРНК ПОМК лосося и лягушки *Xenopus laevis* с построенным филогенетическим графом эволюционного родства млекопитающих. Естественно, что при расширении видового разнообразия, можно будет восполнить недостающие звенья и предковые последовательности мРНК ПОМК, составить единое филогенетическое дерево и рассчитать первичную структуру общей предковой мРНК ПОМК. Филогенетическое дерево эволюционного родства мРНК ПОМК млекопитающих, отнормированное по известным палеонтологическим данным [23], представлено на рис. 4. Гипотетические вершины были вычислены методом минимизации отличий дли параллельных ветвей графа. Рассчитанные эволюционные расстояния между узлами графа отличаются в пределах погрешности метода. Определенная средняя скорость эволюции мРНК ПОМК составила при этом $(2.12 \pm 1.1) \cdot 10^{-9}$ синонимичных замен в год на синонимичный сайт.

СПИСОК ЛИТЕРАТУРЫ

1. *Endogenous opioids and related peptides: from molecular biology to clinical medicine* / Н. Imura, Y. Kato, Y. Nakai et al. // *J. Endocrinol.*— 1985.—107, N 1.— P. 147—157.
2. *Клонирование ДНК, комплементарной мРНК проопиомеланокортина (ПОМК) быка, и определение ее первичной структуры* / С. Я. Головин, Л. В. Мамаев, А. Б. Беклемнишев и др. // *Изв. Сиб. отд-ния АН СССР.*— 1985.— № 13.— С. 140—142.
3. *Синтез, клонирование и определение первичной структуры ДНК, комплементарной мРНК проопиомеланокортина из гипофиза человека* / С. Я. Головин, В. А. Каргинов, А. А. Бондарь и др. // *Биоорг. химия.*— 1987.—13, № 4.— С. 562—564.
4. *Клонирование ДНК, комплементарной мРНК проопиомеланокортина гипофиза быка, крысы и человека. Гормональная регуляция содержания мРНК проопиомеланокортина в гипофизе крысы* / Н. П. Мертвцов, С. Я. Головин, А. Б. Беклемнишев и др. // *Биохимия.*— 1987.—52, № 5.— С. 707—714.
5. *Синтез, клонирование и определение первичной структуры кДНК проопиомеланокортина из гипофиза норки (Mustela vison)* / С. Я. Головин, А. А. Бондарь, В. А. Каргинов и др. // *Биоорг. химия.*— 1988.—14, № 2.— С. 273—275.
6. *Nucleotide sequence of cloned cDNA for bovine corticotropin — β -lipotropin precursor* / S. Nakanishi, A. Inoue, T. Kita et al. // *Nature.*— 1979.—278, N 5703.— P. 423—427.

7. Chang A. C. Y., Cochet M., Cohen S. N. Structural organization of human genomic DNA, encoding the pro-opiomelanocortin peptide // Proc. Nat. Acad. Sci. USA.— 1980.—77, N 8.— P. 4890—4894.
8. Complete nucleotide sequence of the human corticotropin— β -lipotropin precursor gene / H. Takahashi, Y. Hakamata, Y. Watanabe et al. // Nucl. Acids Res.— 1983.— 11, N 19.— P. 6847—6858.
9. Complete structure of the porcine pro-opiomelanocortin mRNA derived from the nucleotide sequence of cloned cDNA / G. Boileau, L. Jeannotte, M. Cretien et al. // Ibid.— N 22.— P. 8063—8071.
10. Isolation and characterization of the mouse corticotropin— β -lipotropin precursor gene and related precursor gene and related pseudogene / M. Notake, T. Tobimatsu, Y. Watanabe et al. // FEBS Lett.— 1983.— 156, N 1.— P. 67—71.
11. Isolation and characterization of the bovine corticotropin— β -lipotropin precursor gene / S. Nakanishi, Y. Teranishi, Y. Watanabe et al. // Eur. J. Biochem.— 1981.— 115, N 3.— P. 429—438.
12. Structure of the rat pro-opiomelanocortin (POMC) gene / J. Drouin, M. Chamberland, J. Charron et al. // FEBS Lett.— 1985.— 193, N 1.— P. 54—58.
13. Uhler M., Herbert E. Complete amino acid sequence of mouse pro-opiomelanocortin derived from the nucleotide sequence of pro-opiomelanocortin cDNA // J. Biol. Chem.— 1983.— 258, N 1.— P. 257—261.
14. Somu G.-I., Kitahara N., Nishizawa T. Nucleotide sequence of a cloned cDNA for pro-opiomelanocortin precursor of chum salmon // Nucl. Acids Res.— 1984.— 12, N 21.— P. 8929—8941.
15. Martens G. J. M., Civelli O., Herbert E. Nucleotide sequence of cloned cDNA for pro-opiomelanocortin in the amphibian *Xenopus laevis* // J. Biol. Chem.— 1985.— 260, N 25.— P. 13685—13689.
16. Staden R. An interactive graphics program for comparing and aligning nucleic acids and amino acids sequences // Nucl. Acids Res.— 1982.— 10, N 9.— P. 2951—2961.
17. Li W.-H., Wu C.-L., Luo C. C. A new method for estimating synonymous and non-synonymous rates of nucleotides substitution considering the relative likelihood of nucleotide and codon changes // Mol. Biol. Evol.— 1985.— 2, N 2.— P. 150—174.
18. Краев А. С., Скрябин К. Г. Методы изучения первичных структур нуклеиновых кислот // Итоги науки и техники.— М.: ВИНТИ, 1980.— С. 141—197 (Молекуляр. биология, генет. инженерия; Т. 12; Ч. 2).
19. Analysis of spontaneous deletions and gene amplification in the *lac* region of *E. coli* / A. M. Albertini, M. P. Hofer, M. P. Calos et al. // Cold Spring Harbor Symp. Quant. Biol.— 1983.— 47, pt 2.— P. 841—850.
20. Ripley L. S., Glickman B. W. Unique self-complementarity of palindromic sequences provides DNA structural intermediates for mutation // Ibid.— P. 851—861.
21. Douglas J., Civelli O., Herbert E. Polypeptide gene expression: generation of diversity of neuroendocrine peptides // Ann. Rev. Biochem.— 1984.— 53.— P. 665—715.
22. Williams A. L., Tinoco I. A dynamic programming algorithm for finding alternative RNA secondary structures // Nucl. Acids Res.— 1986.— 14, N 1.— P. 299—315.
23. Жарких А. А., Ратнер В. А., Родион С. И. Теоретический анализ эволюции генов и белков.— Новосибирск: Наука, 1985.— С. 97—149.

Ин-т биоорг. химии Сиб. отд-ния АН СССР, Новосибирск

Получено 27.09.88

Ин-т клин. и эксперим. медицины

Сиб. отд-ния АМН СССР, Новосибирск

A COMPARATIVE ANALYSIS OF THE PRIMARY STRUCTURE OF DNAs CODING FOR PROOPIOMELANOCORTINS OF SEVEN ANIMAL SPECIES AND HUMAN

V. A. Karginov, S. Ya. Golovin, A. A. Bondar, I. V. Morozov, V. M. Blinov, N. P. Merivtsov

Institute of Bioorganic Chemistry,

Siberian Branch of the Academy of Sciences of the USSR, Novosibirsk;

Institute of Clinical and Experimental Medicine,

Siberian Branch of Academy of Medical Sciences of the USSR, Novosibirsk

Summary

A comparative computer analysis of rat, mouse, mink, bovine, pig, frog, salmon and human DNA coding for POMC has been carried out. The analysis has revealed conservative and variable POMC regions in the species examined. It is suggested that the differences in the variability extent of some domains within mink POMC may be due to the properties of the encoding DNA regions (direct and inverted repeats in variable regions, a high percentage of G—C pairs). A mechanism of mutations within the variable regions of nucleotide mRNA POMC sequences is suggested. A tree of evolutionary relations between the mammalian POMC species examined is constructed.