*Acta Universitatis Latviensis, Biology, 2004, Vol. 676, pp. 97–105*

# Origin of globins and a mystery of myoglobin codon root symmetry

Nora Ieviņa, Gunārs Chipens

Department of Peptide Chemistry, Latvian Institute of Organic Synthesis, Aizkraukles 21, Rīga LV-1006, Latvia
*Corresponding author, E-mail: ievina@osi.lv

## Abstract

The amino acid sequence 49-94 of myoglobin involved in heam binding has 74 % symmetric codon roots – the second codon letters. The search for the reason of the symmetry revealed that the ancestor of globin genes was not formed by exon shuffling, but by multiplication of a 21-membered nucleotide and that it was a regular polynucleotide. The repeat unit pile of contemporary myoglobin codon root sequence posesses an inversion centre indicating that symmetry of codon roots evolved after the origin of globin gene. Its cause and possible biological functions is a mystery. A model of coding of the globin ancestor gene is suggested to be formed of identical repeat units separated by 23 introns.

**Key words:** multiplication of nucleotides, old ancestral introns, origin of introns, repeat units of genes and proteins.

## Introduction

The primary structures of angiotensin and bradykinin contain some symmetrically located amino acids posessing certain common physico-chemical and structural features. It is supposed that the partially symmetric structures of these tissue hormones evolved for optimal adaptation to cell receptor binding sites (Beddell et al. 1977). How widespread is the internal symmetry of peptide chains and what is it true functional significance are unknown.

The physico-chemical properties of amino acids correlate with the structures of amino acid codon roots – the second codon letters (Pelc 1965; Sjostrom, Wold 1985). During evolution, codon root structures are much more conservative and undergo less changes than amino acids. Complementarity of the codon roots A/T (U) and G/C determine the polar component of coded amino acid inter- and intramolecular interaction during protein folding and complexformation reactions (Chipens 1996). The codon root plays the most important role in the codon; its substitution is the most critical for protein structure and biological functions. Therefore, to study protein and gene internal symmetry we chose the method termed comparative amino acid codon root analysis (CAACRA; Chipens, Ievina 1994), and elaborated a new screening test – backward (reverse or invert) autoscanning of protein and gene sequences. This simple method revealed covered symmetry of myoglobin primary structure: in the middle part of the peptide chain, amino acids encoded by codons with identical codon roots (termed "common-root" amino acids) are located symmetrically. Search for the causes of this phenomenon led to a new theory of genes and intron emergence (Ievina, Chipens 2003), but the functions of myoglobin codon root dyad symmetry is still a mystery.

## Materials and methods

In accordance with the method of backward autoscanning, identical symbols (amino acids and/or codon roots) are counted and registered graphically or in a form of a table where the amino acid sequence of a peptide chain or a chain of gene exons (codon root sequence) is moved backward over themselves step-by-step i.e. symbol-by-symbol. At each step the overlapping symbols are compared, and identical symbols are counted and registered (Fig. 1A, B). The position of the symmetry axis shows the ordinal number of the step with the highest number of identical symbols.

The primary structures of genes and proteins were taken from the GenBank via Internet.

## Results and discussion

### *Cover symmetry of myoglobin*

The backward autoscanning graph of the sperm whale myoglobin central fragment His 49 - His 94 (since the main object of our studies is the regularity of gene and protein structures, we enumerate also myoglobin methionine in position 1) has a well expressed

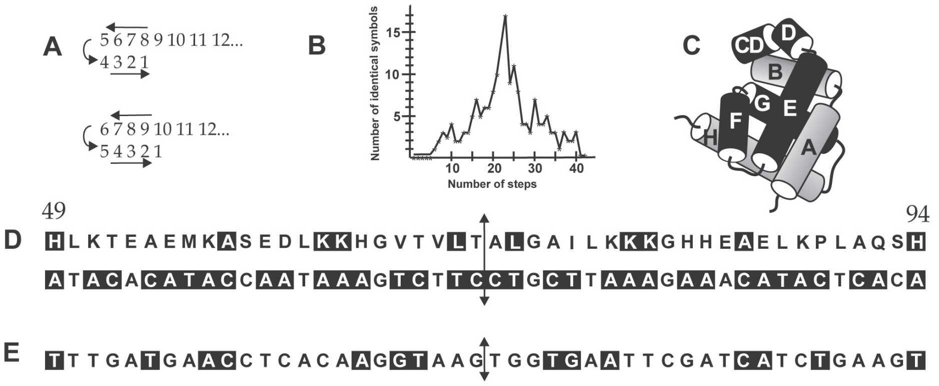
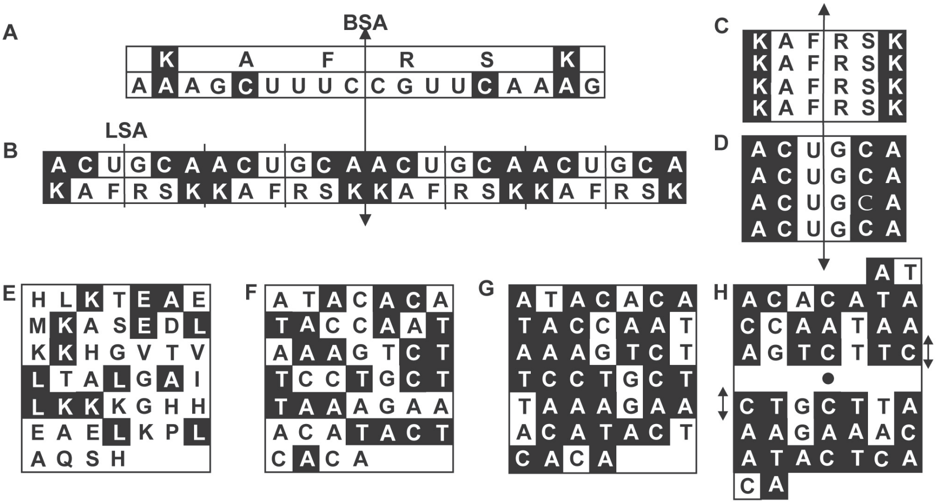

A
5 6 7 8 9 10 11 12...
4 3 2 1
6 7 8 9 10 11 12...
5 4 3 2 1
B
Number of identical symbols
15
10
5
10
20
30
40
Number of steps
C
CD
D
B
G
E
F
H
A
49
94
D
H L K T E A E M K A S E D L K K H G V T V L T A L G A I L K K G H H E A E L K P L A Q S H
A T A C A C A T A C C A A T A A A G T C T T C C T G C T T A A A G A A A C A T A C T C A C A
E
T T T G A T G A A C C T C A C A A G G T A A G T G G T G A A T T C G A T C A T C T G A A G T

**Fig. 1.** Analysis of the sperm whale myoglobin codon root symmetry. A, principle of backward autoscanning shown by two steps of analysis (the step 4 and 5) of a number sequence. Arrows indicate direction of chain movement. At each step opposite numbers (i.e. corresponding symbols of amino acids and/or their codon roots) are compared and the identical are counted and registered when a chain is moved backward over itself step-by-step, i.e. symbol by symbol. Parameters are correlated graphically with the ordinal numbers of steps. Maximums show local or basic positions of symmetry axes. B, a backward autoscanning graph of myoglobin fragment 49-94. C, location of amino acids having symmetric codon roots in myoglobin 3D-structure (shown by black colour). D, symmetric disposition of amino acids and amino acid codon roots (the second letters) revealed by analysis of sperm whale myoglobin fragment 49-94. Symmetric symbols are shown against a black background. Two head arrow denote basic symmetry axis. E, accidental positions of symmetric codon roots in randomized sequence of myoglobin 49-94.


A

BSA

K A F R S K

A A A G C U U U C C G U U C A A A G

B

LSA

A C U G C A A C U G C A A C U G C A A C U G C A

K A F R S K K A F R S K K A F R S K K A F R S K

C

K A F R S K

K A F R S K

K A F R S K

K A F R S K

D

A C U G C A

A C U G C A

A C U G C A

A C U G C A

E

H L K T E A E

M K A S E D L

K K H G V T V

L T A L G A I

L K K K G H H

E A E L K P L

A Q S H

F

A T A C A C A

T A C C A A T

A A A G T C T

T C C T G C T

T A A A G A A

A C A T A C T

C A C A

G

A T A C A C A

T A C C A A T

A A A G T C T

T C C T G C T

T A A A G A A

A C A T A C T

C A C A

H

A T

A C A C A T A

C C A A T A A

A G T C T T C

●

C T G C T T A

A A G A A A C

A T A C T C A

C A

**Fig. 2.** Repeat unit piles (RUP) of a model peptide and the central fragment of the sperm whale myoglobin 49-94. A, a model of a symmetric repeat unit containing four symmetric codon roots and two amino acids shown against a black background. B, a multimer of the model "A" repeat unit. A small part of a long multimer chain is shown containing four repeats. The local symmetry axes (LSA) crosses repeat unit boundaries as well as repeat unit centres. Symmetric symbols are shown against a black background. BSA, the basic symmetry axis. C, D, the model multimer of repeat unit piles (6RUP) of amino acid and codon root sequences are bilateraly symmetric (symmetric symbols are shown against a black background. E, F, repeat unit piles (7RUP) of amino acid and codon root sequences of the myoglobin central part 49-94 are not symmetric. Identical RUP symbols in vertical lines are shown against a black background. G, Symbols of myoglobin 49-94 codon roots which are mirror-symmetric in linear structure (Fig. 1D) are not mirror symmetric in 7RUP structure . H, the same structure of 7RUP as in "G" subdivided into two parts in accordance to the position of the basic symmetry axis (shown by small two-headed arrows between codon roots C (Thr71) and C (Ala72), see also Fig.1D) reveal symmetric structure with an inversion centre (shown as filled circle).

maximum (Fig. 1B) which corresponds to the basic symmetry axis (BSA) crossing the peptide chain between Thr 71 and Ala 72 (Fig. 1D). The analysed myoglobin fragment contains 74 % common-root amino acids (Table 1), among them 22 % identical amino acids. In the 3D-structure of myoglobin (Fig. 1C), this fragment forms α-helices D, E, F and G (partially) and takes part in haem binding (Kendrew et al. 1960). Accidental amino acid sequences with the same amino acid composition, generated by the repeated Monte Carlo simulation (Dorit et al. 1990) resulted in only 30 % symmetric codon roots (the highest value, obtained from 20 separate simulations; Fig. 1E). Thus, the symmetry of myoglobin codon root symbols is not accidental and it is interesting to understand how it arose.

Modelling experiments reveal that long symmetric polynucleotide chains can be formed by multiplication of oligonucleotides. If the nucleotide contains some symmetric elements (e.g. codon roots; Fig. 2A), then these elements after multiplication are distrubuted symmetrically along the full length of the polynucleotide chain, which in such case have many local symmetry axes (LSA) – between every two repeat units or

**Table 1.** Groups of potentially equifunctional, tantamount or related common-root amino acids

| Group name | Codon structure | Group composition |
|---|---|---|
| Adenine (A) group | NAN | D, E, H, N, Q, Y, K |
| Uracil (U) group | NUN | F, L, M, V, I |
| Cytosine (C) group | NCN | T, A, P, S |
| Guanine (G) group | NGN | C, G, R, S, W |

between repeat unit centres – and one as the main or basic symmetry axis (BSA) through the geometric centre of the chain (Fig. 2B). Discovery of myoglobin covert symmetry and modelling experiments of nucleic acid formation by oligonucleotide multiplication reactions indirectly pointed to the possibility that periodic nucleic acids could be a raw material for emergence of living mater (ancestors of genes) and that multiplication of nucleotides may be a universal common mechanism of genes origin. This concept has been generally accepted in our later studies (Ievina, Chipens 2003). However, it is necessary to note that covert symmetry of myoglobin represents a unique case, because it arose during evolution after formation of the globin gene ancestor by multiplication reactions of nucleotides.

### *A repeat unit of globins*

For investigation of myoglobin codon root symmetry, it was necessary to determine exact size of the globin repeat unit. For this purpose we use the method of common prime multipliers of gene numerical parameters, such as exon dimensions, intron co-ordinates, size of the gene coding part, etc. measured by numbers of nucleotides. Prime multipliers characterize the internal regularity of the numbers themselves. If the globin gene family has common ancestor, members of the family must have identical size of repeat units. The identity of repeat dimensions in turn must be reflected in exon sizes in cases when exon size and intron co-ordinates numerical values are determined by a whole number of repeats. In the given case we use the dimensions of three globin gene exons selected from a number of different globin exons, i.e. the human β-globin exon 3 (126nt), the soybean *Glycine max* symbiotic globin exon 2 (105nt) and the protozoa *Paramecium caudatum* β-globin exon 1 (189nt). The dimensions of all these exons have common internal regularity. The product of common prime multipliers (3 × 7, framed):

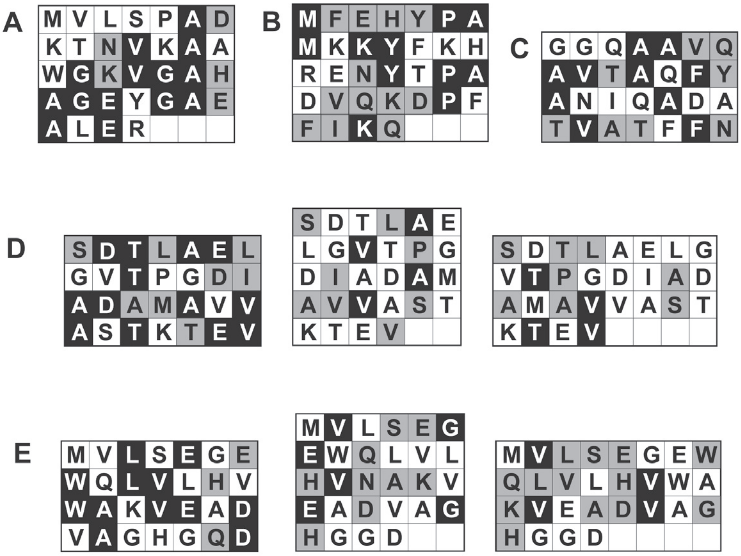
$$\begin{array}{cccccccc} 105 & = & & \boxed{3} & \times5\times & \boxed{7} & = & 5\times21 \\ 126 & = & 2\times & \boxed{3} & \times3\times & \boxed{7} & = & 6\times21 \\ 189 & = & 3\times & \boxed{3} & \times3\times & \boxed{7} & = & 9\times21 \end{array}$$

indicates that the potential size of the globin primary repeat unit is 7RU or 21nt/7aa. The number before the abbreviation of repeat unit (RU) indicates its size – the number of amino acids (aa) or codon roots in the repeat. Symbols 7RU and 21nt/7aa are equivalent.

In accordance to our new model of gene origin (Ievina, Chipens 2003), immediately after emergence of gene precursors, exons and introns were formed of repeat units (RU) that were identical in size and sequence. The coding parts of genes contained a whole number of RU, but introns possibly were located after every RU or micro-exon in the borders of repeats, termed the gene knot points. It is necessary to note that multiplication reactions from only symmetric exons with all introns in phase zero, but the real structures of contemporary genes indicate that during evolution introns can change their positions and phases and be partially or completely eliminated.

Intron positions in modern globin genes more or less correspond to the calculated size of the primary RU. Multiple alignement of 91 globins and globin-related proteins reveal 12 intron positions that are represented in seven genomic sequences. Among these, the soybean globin intron 71-0 (Stoltzfus et al. 1994) with the co-ordinates 210nt (intron co-ordinate is a nucleotide ordinal number of exon row just befor the intron) crosses the globin gene knot point and is a real candidate for an ancestral intron still sitting in the birth position (old "immobile" introns, OII). Multiple alignment of globins is a difficult problem because their sequences are very diverse and they can not be aligned reliably (Stoltzfus et al. 1994). More precise data can be obtained by comparison of the relative positions of introns in the frames of separate α-helices of globins. For example, the distance between introns B5/B6 and B12-2 is 20nt, between F3-1 and F9/10 – 21nt, between F8-1 and E14/E15 – 20nt (Hankeln et al. 1997). In these cases introns are separated by one repeat (21nt) with small deviations determined only by a intron phase change during evolution. This indicates the possibility that introns of the globin gene ancestor indeed were situated after every RU (a micro-exon) and that during evolution a massive loss of introns took place. Interestingly, the exons of several globin genes are formed of a whole number of repeats, which supports the above thesis, e.g., the 3'-terminal exons of myoglobin (the third exon, 49 codons, 7×7 RU) and soybean symbiotic globin $C_2$ (the fourth exon, 42 codons, 6×7 RU), the third exon of soybean globin $C_2$ (35 codons, 5×7 RU), the first exon of *Paramecium caudatum* β-globin (63 codons, 9×7 RU), etc.

The comparison of symmetric codon root sequences of model peptides formed of symmetric repeats (Fig. 2A, B) with the symmetric myoglobin codon root sequence (Fig. 1D) reveal some differences in the pattern of symmetric symbols. In a model of a repeat unit multimer, each repeat is symmetric and symmetric elements are repeated regularly. In myoglobin, codon root sequence does not possess such a regularity – the myoglobin sequence as a whole is symmetric. The difference is well seen in the form of repeat unit piles (RUP). For this purpose, peptide chains are cut into fragments corresponding to the repeat unit size and then are laid out horizontally in stacks to form RUP. Repeat unit piles (Fig. 2C, D) formed of a multimer chain containing symmetric repeats (Fig. 2B) are symmetric also relative to the symmetry axis going through the middle of the pile. In contrast to this, the 7RUP structure formed of the central part of sperm whale myoglobin is not mirror- or bilateraly symmetric, it does not contain isosteric symmetric amino acids or amino acid codon roots (Fig. 2E, F). This is better seen when codon root symbols, which are mirror-symmetric in a linear myoglobin codon root sequence (Fig. 1D), are shown in the 7RUP structure against a black background (Fig. 2G). Transformation of this pile structure by subdividing it into two parts (in accordance with the position of BSA) reveals a RUP possesing an inversion centre (Fig. 2H). At the inversion centre,any two points can interchange (in our case – any two symbols shown against the black background, because the codon root symmetry is not complete, but only 74 %) located at the same distance from the inversion centre along the line going through the centre. Thus, the highly expressed symmetry of myoglobin codon roots in the sequence of the central part of the chain (Fig. 1D) is not a direct result of chain formation by the multiplication reaction of a mirror-symmetric repeat, but is a consequence of evolution of the myoglobin sequence after the emergence of the gene ancestor. If the globin gene ancestor repeat unit itself and the corresponding product of multiplication has been bilaterally symmetric,



A

MVLSPAD
KTNVKAA
WGKVGAH
AGEYGAE
ALER

B

MFEHYPA
MKKYFKH
RENYTPA
DVQKDPF
FIKQ

C

GGQAAVQ
AVTAQFY
ANIQADA
TVATFFN

D

SDTLAEL
GVTPGDI
ADAMAVV
ASTKTEV

SDTLAE
LGVTPG
DIADAM
AVVAST
KTEV

SDTLAELG
VTPGDIAD
AMAVVAST
KTEV

E

MVLSEGE
WQLVLHV
WAKVEAD
VAGHGQD

MVLSEG
EWQLVL
HVNAKV
EADVAG
HGGD

MVLSEGEW
QLVLHVWA
KVEADVAG
HGGD

**Fig. 3.** Repeat unit piles 7RUP of different globin fragments. A, the human α-globin 1-32 (f = 59/44); B, the *Pseudoterranova decipiens* globin 51-82 (f = 69/34); C, the *Paramecium caudatum* globin 8-35 (f = 64/36); D, the *Chlamydomonas eugametos* globin 131-158 (7RUP f = 75/46; 6RUP f = 43/14; 8RUP f = 43/14); E, the sperm whale myoglobin 1-28 (7RUP f = 53/43; 6RUP f = 53/21; 8RUP f = 53/14). Identical symbols of amino acids in vertical lines are shown against a black background, and common-root amino acids – against a grey background.

then during evolution this symmetry was disrupted and excahnged by symmetry with the inversion centre. The driving force for this change may have been formation of structural peculiarities necessary to ensure some possible yet unknown important functional mechanism of myoglobin. In any case, if the structure of a protein is not necessary for a specific function, during evolution it is gradually disrupted by mutations.

Some parts of myoglobin and other primary structures of globin (Fig. 3) however have retained rudimentary regularity characteristic for ancestor globin repeat units containing seven symbols (codon roots or common-root amino acids). The regularity of repeat unit pile (RUP) structure may be characterized as the percentage ratio of common-root (CR) and identical (I) amino acids dominating in the vertical lines of RUP. This ratio we show as a fraction and it is termed the f-factor, f = CR/I (%). The f-factor usually is sensitive to small RU size changes, e.g., by one symbol. The Chlamydomonas eugametos globin sequence 131-158 in the form of 7RUP shows an f = 75/46 (there is a very small possibility to form such a high regularity by chance), but only f = 43/14 in the form of 6RUP or 8RUP (Fig. 3D). RUPs of myoglobin sequence 1-28 also show similar regularity changes (Fig. 3E).

### *A model of the globin gene organisation*

The framework of gene structural organisation is determined by the dimension of RU and disposition of the gene knot points, which are formed by borderlines of RU. In the globin exon row, the knot points are probably situated after each 21nt, and they indicate the potential positions of the splice sites of introns as well as the potential positions of local (including basic) symmetry axes. Theoretically, the basic symmetry axis (BSA) in the ancestor gene structure must cross the knot point in the centre of chain and coincide with the central intron, or this intron potential position if the intron itself has been lost during evolution. The psotion of myoglobin BSA determined by backward autoscanning (Fig. 1A, B) does not cross the middle point of the myoglobin peptide chain as a result of truncation during evolution of the gene 5'-terminus encoding signal peptide. Contrary to this, the COOH-terminal half of myoglobin peptide chain does not contain indels, because the size of the exon 3 is 49 codons (including the termination codon), that is 7×7RU, but the distance from the position of BSA to the 3'-terminus of the gene coding part also contains a whole number of 7RU, i.e. 84 codons (12×7) or 252nt (12×21 nt; Fig. 4C). If our concept of nucleotide multiplication reactions is correct, then the length of the globin ancestor gene coding part was 84×2 = 168 codons or 504nt (24×21 nt). We suppose that in this way we are virtuallu measuring the coding part of gene existing before divergence of plants and animals about 1,500 millions years ago (Anderson et al. 1996; Hardison 1996).

The calculated theoretical length of the globin gene ancestor coding part allows to design a model of globin structural organisation. The nucleotide chain containing 504nt

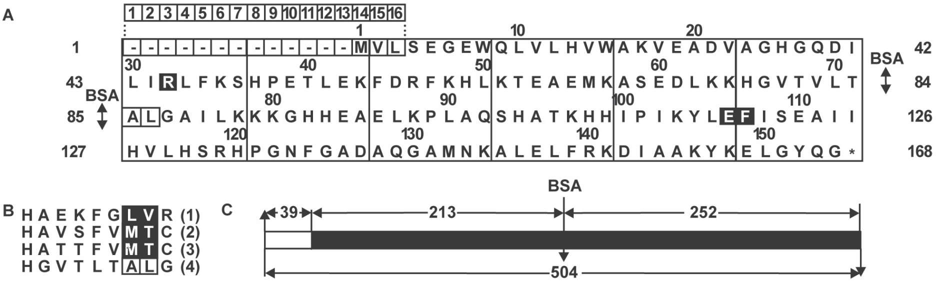

A
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16
1 10 20
1 - - - - - - - - - - - - - - - M V L S E G E W Q L V L H V W A K V E A D V A G H G Q D I 42
30 40 50 60 70
43 L I R L F K S H P E T L E K F D R F K H L K T E A E M K A S E D L K K H G V T V L T 84
BSA
BSA
80 90 100 110
85 A L G A I L K K K G H H E A E L K P L A Q S H A T K H H I P I K Y L E F I S E A I I 126
120 130 140 150
127 H V L H S R H P G N F G A D A Q G A M N K A L E L F R K D I A A K Y K E L G Y Q G * 168
B
H A E K F G L V R (1)
H A V S F V M T C (2)
H A T T F V M T C (3)
H G V T L T A L G (4)
C
BSA
39
213
252
504

**Fig. 4.** A model of structural organisation of the globin family protein ancestor. A, the sequence of sperm whale myoglobin is written in a frame of 7-membered repeat units grouped in four lines in such a way that the symmetry axis BSA of myoglobin coincide, with the geometric symmetry axis of the linear 168-membered peptide chain. The BSA is shown as two headed arrows between amino acid symbols in positions 84 and 85. The amino acid sequence has double numeration: in accordance to the model of globin ancestor structure (numbers in both flanks of lines) and natural myoglobin (above the symbols). The myoglobin intron N.1 in the phase 2 crossess the arginine 32, but intron N.2 crosses separate codons 106/107. Corresponding amino acid symbols are shown against a black background. The position of the central intron (lost in myoglobin structure) is between alanine 85 and leucine 86 (symbols are framed). The potential position of the myoglobin ancestor signal sequence 1-16 is shown separately (cells with numbers). An asterisk denotes the termination codon (position 168). B, determination of a potential central intron position in myoglobin by alignment of the E-helices of globins: the symbiotic *Glycine max* globin C2 62-70 (1); the *Glycine max* nonsymbiotic globin 69-77 (2); the *Parasponia andersonii* globin 70-78 (3), and myoglobin 65-73 (4). Intron positions (all in the phase-0) separating codons of corresponding amino acids are marked by a black colour. C, a sheme illustrating the determination of globin ancestor dimension using the revealed (Fig. 1) co-ordinates of the basic symmetry axis, BSA. Numbers show length, measured by numbers of nucleotides. The disposition of the sperm whale myoglobin sequence coding part is shown in black.

can be formed by 24 primary repeats 7RU (21nt/7aa). The calculated size of globin peptide chain encoded by 168 codons (504nt) accounts for 157 amino acids and a stop codon and is comparable with the size of natural globin peptide chains, e.g. the plant *Parasponia andersonii* globin (162aa), the midge *Chironomus thummi* globin encoded by the gene *ctt-XI* (167aa), the nematode *Ascaris* and *Pseudoterranova decipiens* globins (the first domain – 167aa) and others.

Writing of the sequence of the sperm whale myoglobin in the structural frames of the above described model in such a way that the basic symmetry axis BSA of myoglobin (between Thr71 and Ala72; Fig. 1) coincides with the geometric symmetry axis of the model (crossing the centre of the chain between codons 84/85) reveals an interesting picture (Fig. 4A): (i) the termination (stop) codon of the myoglobin gene coding part takes the extreme position of the model corresponding to the codon 168 (in accordance to our symmetry analysis and calculation data); (ii) the second (central) intron of the globin gene family is in a close neighbour position to myoglobin BSA. The central intron of myoglobin has been lost during evolution, but the alignment of myoglobin and plant globin sequences in the region of E-helices (Fig. 4B) indicates that the central intron of plant globin is only 3nt aside from myoglobin BSA; (iii) in the N-terminal part of the model there is a place for the globin signal peptide containing 16aa (Andersson et al. 1996). We suppose that the leader sequence of globins during evolution has been lost later (as in a case of myoglobin) or changed by mutations (as in a case of plant globins).

Up to 55 % of the mammalian genome consists of different kinds of repetitive and regular nucleic acids (satellites, mini- and micro-satellites, short and long interspersed elements, etc.; Cavalier-Smith 1985). The remaining half of non-repetitive DNA in several until up to now analysed genes after translation to codon root sequences also show residual regularity (Ievina, Chipens 2003). Consequently, possibly there is only one basic mechanism of nucleic acid (including the genome DNA) formation, in good accordance with the concept of biochemical universiality (Dayhoff 1972).

The discovery and study of internal covert symmetry and regularity of myoglobin for the first time bring to light the true mechanism of emergence of the very first gene ancestors by multiplication of nucleotides.

According to the exon theory of genes (Crick 1979; Gilbert 1987) the DNA sequence that codes for the globin ancestor did not begin to evolve as a single uninterrupted strech of DNA. Instead, it evolved from three distinct exons which already existed (evolved earlier) and were brought together in the genome by random shuffling. However, (i) the identical size of repeats of globin exons i.e. 7RU or 21nt/7aa; (ii) the identical internal regularity of several globin exons e.g. 5×7RU, 6×7RU, and 7×7RU; and (iii) the same regularity of soybean symbiotic globin $C_2$ "old immobile" intron co-ordinate (210 nt or 10×7RU) in our view do not support exon shuffling in this case.

## References

Andersson C.R., Jensen E.O., Llevellyn D.J., Dennis E.S., Peacock W.J. 1996. A new hemoglobin gene from soybean: A role for hemoglobin in all plants. *Proc. Natl. Acad. Sci. USA* 93: 5682–5687.

Beddell C.R., Sheppey G.C., Blundell T.L., Sasaki K., Dockerill S., Goodford P.J. 1977. Symmetrical features in polypeptide hormone-receptor interactions. *Int. J. Peptide Protein Res.* 9: 161–165.

Cavalier-Smith T. 1985. Evolutionary significance of genome size. Eucariotic gene numbers, noncoding DNA and genome size. In: Cavalier-Smith (ed) *The Evolution of Genome Size*. Wiley, London, pp. 1-104.

Chipens G. 1996. The second half of the genetic code. *Proc. Latv. Acad. Sci. Sect. B* 50: 151–172.

Chipens G., Ievina N. 1994. Comparative amino acid codon root analysis (CAACRA) of peptide chains. *Proc. Latv. Acad. Sci. Sect. B* 48: 50–54.

Crick F. 1979. Split genes and RNA splicing. *Science* 204: 204–271.

Dayhoff M.O., Eck R.V. 1972. Tracing biochemical evolution. In: Dayhoff M.O. (ed) *Atlas of Protein Sequence and Structure*. Natl. Biom. Res. Found., Washington, pp. 1–5.

Dorit R.L., Schoenbacher L., Gilbert W. 1990. How big is the universe of exons? *Science* 250: 1377–1382.

Gilbert W. 1987. The exon theory of genes. *Cold Spring Harbor Symp. Quant. Biol.* 52: 901–924.

Hankeln T., Ebersberger F.H., Schmidt M.J. 1997. A variable intron distribution in globin genes of *Chironomus*: evidence fro recent intron gain. *Gene* 205: 151–160.

Hardison R.C. 1996. A brief history of hemoglobins: Plant, animal, protist, and bacteria. *Proc. Natl. Acad. Sci. USA* 93: 5675–5679.

Ievina N., Chipens G. 2003. A new approach to study the origin of genes and introns. *Acta Univ. Latv.* 662: 67–79.

Kendrew J.C., Dickerson R.E., Strandberg B.E., Hart R.G., Davies D.R., Phillips D.C., Shore V.C. 1960. Structure of myoglobin. *Nature* 185: 422–427.

Pelc S.R. 1965. Correlation between coding-triplets and amino acids. *Nature* 207: 597–599.

Sjostrom M., Wold S. 1985. A multivariable study of the relationship between the genetic code and the physical-chemical properties of amino acids. *J. Mol. Evol.* 22: 272–277.

Stoltzfus A., Spencer D.F., Zuker M., Logsdon J.M.Jr., Doolitle W.F. 1994. Testing the exon theory of genes: the evidence from protein structure. *Science* 265: 202–207.

## Globīnu izcelšanās un mioglobīna kodona sakņu simetrijas noslēpums

Nora Ieviņa, Gunārs Čipēns

Peptīdu ķīmijas laboratorija, Latvijas Organiskās sintēzes institūts, Aizkraukles 21, Rīga LV-1006, Latvija

*Korespondējošais autors, E-pasts: ievina@osi.lv

### Kopsavilkums

Mioglobīna fragmentam, kas piedalās hēma saistīšanā, 74 % aminoskābju sekvences 49-94 kodē nukleotīdu tripleti, kuru kodonu saknes (otrie burti) ir simetriskas. Simetrijas cēloņu meklējumi ļāva secināt, ka globīna gēnu priekštecis neveidojās eksonu pārneses rezultātā, bet bija 21 locekļa nukleotīda multiplicēšanās galaprodukts – regulārs polinukleotīds. Mūsdienu mioglobīna kodona sakņu sekvences atkārtojuma vienību grēdai ir inversijas centrs, kas norāda, ka kodonu sakņu simetrija attīstījusies pēc globīnu gēnu rašanās. Simetrijas cēlonis un iespējamās bioloģiskās funkcijas pagaidām nav zināmas. Ir izveidots hipotētisks globīna gēnu priekšteča kodējošās daļas modelis, kas sastāv no identiskām atkārtojuma vienībām, kuras atdala 23 introni.