*Acta Universitatis Latviensis, 2006, Vol. 710, Biology, pp. 139–153*

# Internal regularity and quantization of gene parameters

**Nora Ieviņa*, Gunars Chipens, Ivars Kalvinsh**

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

**Abstract**

The origin and function of introns in eukaryotic cells has rised much attention since their discovery in 1977. However, none of the opposing theories is universally recognized today. We have elaborated a new approach and methodology to study the origin of genes and introns using analysis and comparison of gene numerical parametres and their internal regularity. Exon dimensions, intron coordinates, sizes of gene exon rows, etc. were investigated. Our aim was to demonstrate that the structures of primeval gene precursors originating at the very early stage of evolution were regular and periodic, and that this regularity was retained partially also in the structures of modern contemporary genes and the corresponding proteins. The ability to determine the size of the gene quantum and to quantize the gene numerical parameters for the most regular modern gene structures gave support for this new approach. In this article we describe new principles and a method of analysis of gene internal regularity, demonstrate the internal regularity of some gene segments, and determine the numerical values of their quanta.

**Key words:** gene quanta, *ras*-related genes, regularity of exon dimensions and intron coordinates, triosephosphate isomerase genes.

## Introduction

Two opposite theories well known as "introns early" (Gilbert 1978; Gilbert 1987; de Souza et al. 1988) and "introns late" (Palmer, Logsdon 1991; Logsdon 1998) have been elaborated to investigate gene and intron origin, but none of them is universally recognized today, as there is insufficient corroborative evidence. We attempted to elaborate and verify a new approach and method to study the origin of genes and introns using new methods of analysis of gene codon root and amino acid sequences (Chipens 1996; Ievina, Chipens 2004), as well as comparison of the internal regularity of gene parameters. We hypothesize that introns and exons are products of gene evolution originating at a definite geological period of time (Ievina, Chipens 2003; Ievina, Chipens 2004; Chipens et al. 2005a,b). We assume that gene precursors were highly regular periodic nucleic acids formed by replication of identical in size and sequence oligonucleotides, named repeat units (RU; Chipens, Ievina 1999). According to our model, exons and introns in gene structures originated from periodic nucleic acids after the emergence of the very first splicing machinery. Introns, in the absence of the constraints imposed by the coding function, as well as natural selection on the level of proteins, accumulated mutations without any limits during billions of years of evolution. Exons, contrary to introns, remained relatively conserved. If the hypothesis of oligonucleotide multiplication is correct, exons in modern

gene structures, according to the amino acid interaction code and the codon root code (Chipens 1996; Chipens, Ievina 2004a; Chipens, Ievina 2004b), likely retained some more or less expressed structural regularity.

Our present investigation was based on the concept of the translational symmetry, particularly the dimensionality and periodicity of lineland (Hargittai, Hargittai 1995). The aim of this research was to demonstrate that the regularity of gene ancestors partially was retained also in the structures of contemporary genes and the corresponding proteins. Support for these ideas was raised by the investigation of the amino acid interaction code (Chipens, Ievina 2004a; Chipens, Ievina 2004b), as well as the ability to quantize gene numerical parameters (Chipens et al. 2005a; Chipens et al 2005b). With the aim to determine elements of regularity in contemporary gene sequences in accordance with the model of gene origin by nucleotide repeat unit multiplication reactions (Ievina, Chipens 2003) in this article we focus on the dimensions of numerical parametres of gene and protein structural elements. In the course of our work new notions and definitions of biopolymer internal regularity as well as methods of their analysis were suggested and elaborated.

The theoretical basis of the investigation was the concept of the singular point and the translational symmetry, particularly the dimensionality and periodicity of lineland (Hargittai, Hargittai 1995), a one dimensional space group which describes the symmetry involving definite repetition or periodicity in one direction (Fig. 1). According to our model (Ievina, Chipens 2003; Chipens, Ievina 1999) gene precursors were formed by reactions of spontaneous saltatory replications of oligonucleotides or nucleotides. In analogous reactions of DNA satellite formation the number of repeat unit monomers in a nucleotide chain is very large [105-107 repeats, e.g. in *Xenopus laevis* or *Drosophila virilis* (Lewin 1995)] and correspond to the definition of lineland.

Simple translation symmetry is the most obvious symmetry element of the space groups. It brings the pattern into congruence with itself over and over again. The absence of a singular point leads to regularity expressed in infinite repetition which characterizes translation symmetry (Hargittai, Hargittai 1995). Real objects are not infinite. For symmetry consideration, it may be convenient to look only at some portion of the whole where the ends are not yet in sight and extend them in thought to infinity (Hargittai, Hargittai 1995). The concepts "lineland" and "translational symmetry" were the basis for elaborating a new term – gene "knot points" (Fig. 2) and a new method for investigation of biopolymer regularity – design and analysis of repeat unit piles (Fig.1 E, F; Chipens, Ievina 1999; Ievina, Chipens 2004). Analysis of gene knot points and intron coordinates led to the discovery of discreet values of gene parameters and the ability to calculate gene quanta Q.

## Materials and methods

We investigated the dimensions and numerical parameters of gene and protein structural elements and their internal regularity. Protein and gene parameters (exon length, intron coordinates, etc.) as well as sequences, were taken from the GenBank and SwissProt databases or literature. All parameters of genes were expressed using as a unit of measure one nucleotide (nt). Prime number (a prime number can only be divided exactly by itself or one) multipliers of the obtained parameters were calculated to determine their common

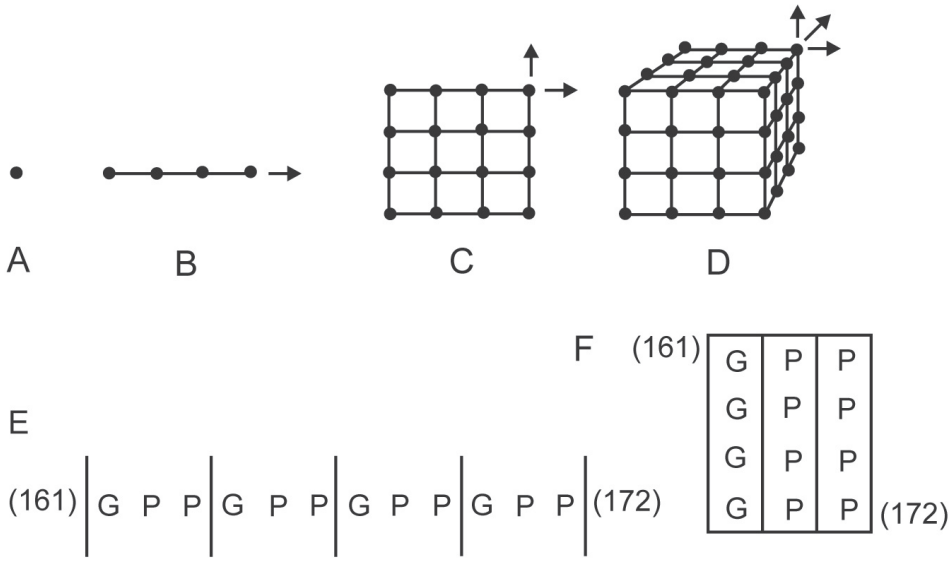


**Fig. 1.** Dimensionality and periodicity in point groups and space groups. A, pointland; B, lineland; C, flatland; D, spaceland (Hargittai, Hargittai 1995); E, a small fragment of human collagen *COL2A1* gene with amino acid sequence 161-172 (GenBank, accession Q14047) show translation symmetry (the repeat unit is 9 nt or 3 amino acids (GlyProPro); F, translation symmetry is demonstrated by repeat unit pile (RUP) structure (Ievina, Chipens 2004). Vertical lines of RUP contain identical amino acid symbols.

internal regularity (common identical prime multipliers, see an example Fig. 2 D) which in accordance with the model of lineland and gene knot points must be identical for regular segments of gene structure. Each numerical parameter of a gene, for example, the exon dimension or intron coordinate (the intron coordinate is the sum of preceeding exon lengths, expressed by the number of nt), can be represented as a product of prime number multipliers. If several numerical parameters of a gene have a set of common prime number multipliers then this phenomenon is named "common internal regularity". Theoretically, if a gene precursor could be analysed immediately after the multiplication reaction (before mutations and other factors change the precursor sequence) then the gene precursor structure would correspond to the structure of the lineland (Hargittai, Hargittai 1995) (Fig. 1 and 2). The most characteristic feature of the lineland-type structures is regularity expressed in infinite repetition which characterizes translational symmetry. Practically, such a structure can be observed in a multimer formed of identical in size and nucleotide sequence repeat units (Fig. 1 E, F). From the formal similarity of linelands and RU multimer linear chain structures we derrived a new term – the "gene knot point". Distances between two neighbour gene knot points determine the size of the repeat unit RU (nt) and the numerical value of the gene quantum Q (nt). In regular gene structures gene knot points determine the structural organization of the exon row. Distances in the atoms of crystals self-evidently are many times smaller than lengths of RU. However, for simplicity and to demonstrate the principal identity of atom location regularity in crystal facets and repeat unit regularity in gene ancestors and also in modern gene structures (partially), the RU

size in Fig. 2 is equalized to interatom distances in crystal cells.

The concept of gene RU and knot points has key significance in our working hypothesis and new method of gene analysis. It is supposed that gene splicing mechanisms measure exon length by whole numbers of codons by an unknown mechanism (Robberson et al. 1990) and, as we suppose, possibly also by whole numbers of RU.

## Results and discussion

### *The model of gene, exon and intron origin*

The general principle of the model of gene precursor formation is that nucleotides or fragments of nucleic acids were spontaneously multiplied laterally to generate a large number of identical copies termed repeat units (RU; Ievina, Chipens 2003; Chipens et

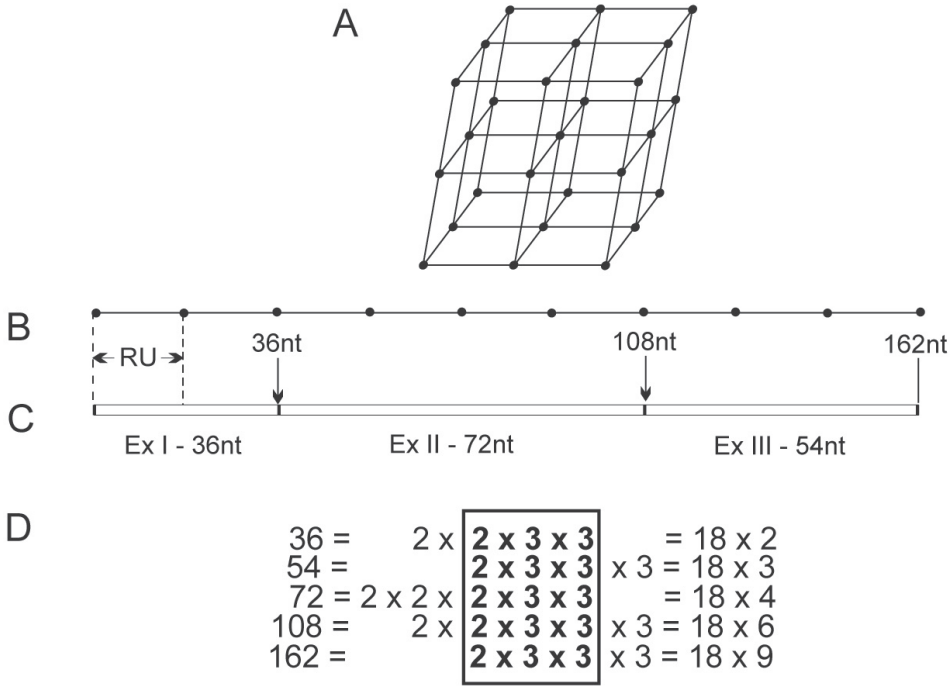


**Fig. 2.** Principles of calculation of common internal regularity of exon dimensions and intron coordinates. A, a threedimensional crystal cell structure (symmetry $G_3^3$ or spaceland); B, structure of lineland ($G_1^3$) derived from a crystal cell; C, scheme of a artificial highly regular model of gene structure. Exons are denoted by ordinal (Roman) numbers, the sizes (nt) of exons (Ex) are shown. Arrows topped with numbers (intron coordinates) indicate gene knot points where introns split the exon row. Repeat unit sizes correspond to the distance between two gene knot points. Exon dimensions are precisely multimers of repeat unit dimensions; D, calculation of prime number-multipliers and identification of common prime multipliers (framed). Common prime multipliers $2 \times 3 \times 3$ determine common internal regularity of the gene model numerical parameters. The product of common prime multipliers $2 \times 3 \times 3 = 18$ allows to determine the potential size of the gene quantum Q – i.e. the number of nucleotides in RU (18 nt), which also coresponds to the repeat unit dimension.

al. 2005a). Exons and introns arose simultaneously in early evolution from common precursors which were highly repetitive simple-structure nucleic acids. It is supposed that the formation of exons and introns was induced by the emergence and action of the very first splicing machinery. Devoid of the coding function, introns accumulated mutations without any limits, did it significantly faster than exons, and lost the regularity of nucleotide and codon root sequences. In accordance to the model intron positions in gene structures were not random – the birth positions of introns were located at internally regular distances between the repeat unit boundaries. These theses are confirmed by contemporary gene and protein structure analysis (Chipens et al. 2005a and this publication). It is necessary to note that reactions of nucleotide multiplication determined the formation of long open reading frames of gene precursors with symmetric exons and all introns in phase zero.

Immediately after the emergence of gene precursors exons and introns were formed of identical in size and sequence repeat units. However, mutations, indels and intron sliding in the course of evolution disrupted this sequence regularity. Introns gradually disappeared or even were lost at once completely. However, as we suppose, contemporary genes may have retained some or several introns in the birth positions. If so, the regularity of location of these introns must be identical with the sequence regularity of exons and exon-coded protein fragments. The latter in separate cases can be demonstrated after translation of gene and protein sequences into codon root symbols (the second codon letter; Ievina, Chipens 2003). However, the concept of gene knot points demonstrates for the first time the discrete structure of large segments of genes and introduces a new concept about gene quanta – the constant number of nucleotides in a given gene repeat unit.

### *Selection of proteins for studies*

Hundreds of G-protein coupled receptors initiate different intracellular signalling chain reactions by the G-protein nanomachine (Clapham 1996). Besides these G-proteins eucariotic cells contain a variety of other structurally related proteins of molecular mass around 20-30 kDa that function as monomers in different regulatory pathways through their capacity to bind and hydrolize GTP specifically (Burgoine 1989). This is the family of small G proteins (SGP), which consists of several subfamilies, such as Ras-, Ypt, Art and others (Ditmaier, Farby 1994). Among SGP, the p21 products of different *ras* genes have attracted the greatest attention, as mutated versions can cause malignant transformation of mammalian cells.

The Ypt proteins (a subfamily of ubiquitous eucariotic SGP) are structurally related to the *ras* gene products and share with them similar biochemical properties (Hanbruck et al. 1989). Several years ago (Ievina, Chipens 2003) we revealed in the basidomycete *Coprinus cinereus* ras gene an identical internal regularity of four neighbour side-by-side exons. Therefore, for analysis we chose other genes of the SGP family – a *ras*-related GTP binding protein encoded by the mice *YPT1* gene (Wichmann et al. 1989), a *ras*-like gene of the edible mushroom *Lentinus edode* (Hori et al. 1991), and the *ras*-related gene of the *Mucor racemosus* (Casale et al. 1990).

The main focus of investigation however was the triosephosphate isomerase (*TPI*) gene (Noltman 1972), as the exon theory arguments suggests that introns are as old as the genes themselves and that apparent correlation of many of the intron sites in plant, animal and fungal *TPI* genes is evidence of their assembly of ancient genes by exon shuffling (Gilbert et al. 1986; Marchionni, Gilbert 1986). The evolution of exon-intron structure of

eucariotic genes from the viewpoint of "introns early" theory is well described in review articles (Gilbert et al. 1977; Long et al. 1995; Fedorov et al. 2001). The data of Gilbert's team indicate that intron positions show non-random distribution in ancient genes (Long et al. 1995), and that only phase zero introns are correlated with structural elements or modules of ancient proteins (De Souza et al. 1998). These data support our concept that in early period of evolution all introns of gene ancestors were in the phase zero (Ievina, Chipens 2003).

Researchers representing the intron insertional theory sequenced the "new *TPI*" genes from three diverse eucariotes – the basidomycete *Coprinus cinereus*, the nematode *Caernorhabditis elegans*, and the insect *Heliothis virescens* (Logsdon et al. 1995). They revealed introns at seven novel positions that disrupt previously recognized gene/protein structural correlation revealed by the "introns early" school. They predicted that, when more *TPI* genes are sequenced, more intron positions would be discovered, and that these sites would fall randomly with respect to *TPI* gene protein structural elements or modules (Logsdon et al 1995).

One of the most studied *TPI* genes is that of chicken (Straus, Gilbert 1985; Banner et al. 1975). Therefore, to compare the results obtained by our model and to analyse the natural gene structure, our choice fell on this sequence. To test the possibilities of "mathematical modelling" of gene structures and to estimate the effects of virtual intron position transfer to the gene knot points (mainly by changing only the intron phase) we selected for analysis the *TPI* gene of *Coprinus cinereus* containing six exons (GenBank accession number U23079) and encoding a 251-amino-acid-residue-long peptide chain. This gene contains no intron in phase zero.

### *Internal regularity of small G proteins*

To determine the potential size of *YPT1* gene quantum and the dimensions of RU, we calculated the common prime multipliers of all exon dimensions and intron coordinates (Table 1 and 2). The obtained values allowed to select a set of common prime multipliers $2 \times 2 \times 3$ (Table 2, framed) and to determine their product – 12 nt. Almost all *YPT1* gene parameters (four intron coordinates and three exon dimensions) could be expressed as multiples of 12 nt. Thus we concluded that the RU size as well as the *YPT1* gene quantum Q is 12 nt. The analysis of this gene structure (Fig. 3 A) showed that most likely during evolution the first intron as a result of mutation had slid off from the gene knot point, i.e.

**Table 1.** The mouse *YPT1* gene exon dimensions according to Wichmann et al. (1989) and intron coordinates and phases calculated on the basis of exon sizes. The gene exon row structure is given in Fig. 3

| Exons ordinal No. | Exon size (nt) | Intron ordinal No. | Intron coordinate (nt) | Intron phase |
|---|---|---|---|---|
| 1 | 23 | 1 | 23 | 2 |
| 2 | 73 | 2 | 96 | 0 |
| 3 | 96 | 3 | 192 | 0 |
| 4 | 96 | 4 | 288 | 0 |
| 5 | 132 | 5 | 420 | 0 |
| 6 | 195 | - | - | - |

**Table 2.** Common prime number multipliers (framed) of exon dimensions and intron coordinates of the mouse *ras*-related gene *YPT1*

| Gene parameters | Dimensions (nt) | Prime multipliers | Parameters as multiples of the gene quantum |
|---|---|---|---|
| Exon III and IV dimensions | 96 | $2 \times 2 \times 2 \times \boxed{2 \times 2 \times 3}$ | $96 = 12 \times 8$ |
| Intron No. 2 coordinate | 96 | $2 \times 2 \times 2 \times \boxed{2 \times 2 \times 3}$ | $96 = 12 \times 8$ |
| Intron No. 3 coordinate | 192 | $2 \times 2 \times 2 \times 2 \times \boxed{2 \times 2 \times 3}$ | $192 = 12 \times 16$ |
| Intron No. 4 coordinate | 288 | $2 \times 2 \times 2 \times \boxed{2 \times 2 \times 3} \times 3$ | $288 = 12 \times 4$ |
| Exon V dimensions | 132 | $\boxed{2 \times 2 \times 3} \times 11$ | $132 = 12 \times 11$ |
| Intron No. 5 coordinate | 420 | $\boxed{2 \times 2 \times 3} \times 5 \times 7$ | $420 = 12 \times 35$ |

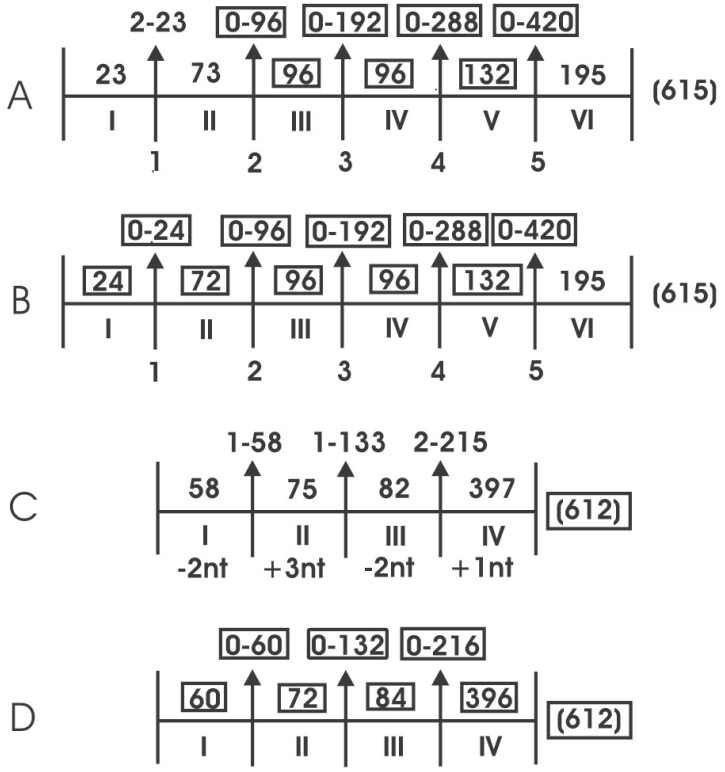


**Fig. 3.** Schemes of organisation of the mouse *YPT1* gene and *Mucor racemosus* gene *MRAS1* (shown not in scale). Arrows with ordinal numbers are topped with parameters characterizing intron phases and coordinates (nt). The horizontal middle line symbolizes the exon row nucleotide sequence. Exon lengths (nt) are shown above the line, ordinal (Roman) numbers – below. Numbers in brackets show exon row total length including initiation and termination codons. The framed and bold typed numbers and segments of schemes show gene parameters and elements, which can be expressed as gene quantum Q (12 nt) multiples. A, mouse gene *YPT1*; B, the same gene after virtual transfer of the first intron to zero position; C, *Mucor racemosus* gene *MRAS1* (GenBank accession M55175) and changes of exon dimensions (± nt). D, a model of a completely regular *MRAS1* gene.

was shifted upstream by one nucleotide. If the potential intron sliding was corrected, and the intron was virtually shifted back to the birth place, i.e. to the nearest knot point of the gene, then three internally regular structural parameters of the *YPT1* gene appeared (Fig. 3 B) – the length of the first and the second exon (correspondingly 2 × 12 = 24 nt and 6 × 12 = 72 nt) and the first intron coordinate (24 nt). Therefore, 10 numerical parameters of 11 of this gene were internally potentially regular and could be expressed as multiples of Q = 12 nt. The obtained results were unexpectedly good. But not all genes have retained such a high regularity of ancestors, e.g., none of the five *YPT1* intron positions matched those of the H-, K-, or N-ras genes (three introns) or those of the *R-ras* gene (five introns; Wichmann et al. 1989).

However, at the same time, several other genes (Casale et al. 1990) showed striking similarity to human *ras* genes, for example, the *MRAS1* gene of *Mucor racemosus* (GenBank accession M55175) showed the same identical internal regularity (Fig. 3 C, D). The revealed deviations of intron coordinates (1-2 nt) in this case most likely could be explained by changes of intron phases or virtual intron sliding (exact molecular mechanisms of intron phase changes are not yet known). Regardless the length of the exon row in the *MRAS1* gene is precisely internally regular (612 nt = 12 × 51).

The edible basidomycete (mushroom) *Lentinus edode* terminal 5'- and 3'-introns No. 1 (coordinate 12 nt) and No. 6 (coordinate 528 nt; 528 = 12 × 44) are precisely regular and can be quantized (Q = 12nt) indicating that gene structure, at least the nucleotide sequence 1-528 nt, is regular. However, introns No. 2 to No. 5 have slid off from the gene knot points by 1-2 nt (Table 3). Also, *YPT1* and *MRAS1* genes, and several other *YPT1* genes of the *RAS* family (e.g., *Chlamidomonas reinhardtii*, *Schizosaccharomyces pombe*, etc.; Chipens et al. 2005b) show partial regularity corresponding to the gene quantum 12 nt.

### *Elements of internal regularity of triosephosphateisomerase (TPI) genes*

Our viewpoint is that a stable principle of the new model of gene origin (Ievina, Chipens 2003) is the possibility to calculate theoretical sizes of exons and exon rows, as well as the potential intron positions in genes, if the size of the repeat unit is known. This allows to

**Table 3.** The *Lentinus edode ras*-like gene exon dimensions (GenBank D00742), calculated intron coordinates, and intron position deviations from the nearest gene knot point coordinates calculated using the gene quantum value Q = 12 nt

| No. | Exon coordinates (nt) | Exon dimensions (nt) | Intron coordinates (nt) | Intron phase | Calculated nearest gene knot point (nt) | Deviation of intron position from the knot point |
|---|---|---|---|---|---|---|
| 1 | 56 - 67 | 12 | 12 | 0 | 12 | 0 |
| 2 | 122 - 167 | 46 | 58 | 1 | 60 | -2 |
| 3 | 224 - 298 | 75 | 133 | 1 | 132 | -1 |
| 4 | 351 - 432 | 82 | 215 | 2 | 216 | -1 |
| 5 | 491 - 684 | 194 | 409 | 3 | 408 | 1 |
| 6 | 748 - 866 | 119 | 528 | 0 | 528 | 0 |
| 7 | 930 - 1055 | 123 | - | - | - | - |

**Table 4.** The chicken *TPI* gene exon row structure (GenBank 11941) and calculated intron coordinates. Gene parameters – multiples of *TPI* gene quantum (Q = 9 nt) are framed

| No. | Exon dimensions (nt) | Intron coordinates (nt) |
|---|---|---|
| 1 | 112 | 112 |
| 2 | 124 | 336 |
| 3 | 85 | 321 |
| 4 | 33 | 454 |
| 5 | 86 | 540 |
| 6 | 88 | 628 |
| 7 | 119 | 747 |

compare the calculated and the natural parameters of gene and protein structures. The first task was to determine the common prime multipliers of the well known chicken *TPI* gene (Straus, Gilbert 1985). The gene parameters, however, were very unregular (Table 4). None of the exon lengths contained a whole number of codons. We found only two internally regular parameters of the chicken *TPI* gene (Table 4): the intron No. 5 coordinate (540 nt) and the exon row length (as a sum of all exon dimensions) – 747 nt:

$$540 = 2 \times 2 \times 3 \times \boxed{3 \times 3} \times 5 \quad = 9 \times 60$$
$$747 = \qquad\qquad \boxed{3 \times 3} \times 83 \quad = 9 \times 83$$

The potential size of the *TPI* repeat unit and the gene quantum was 9 nt. The length of exons VI and VII between two internally regular gene parameters (intron No. 5 and the 3'-terminal nucleotide No. 747) also had to be internally regular. Virtual changes of exon VI and exon VII dimensions (88 nt + 2 and 119 nt – 2) indeed revealed two new internally regular parameters of the potential chicken *TPI* gene ancestor (Fig. 4): the exon VI dimension (90 nt) and the new intron No. 6 coordinate (630 nt). The common internal regularity of these elments was the same: 90 = 9 × 10 and 630 = 9 × 70. *TPI* genes of other species (Table 5) in separate regions showed the same size of gene quantum and RU. The parameters shown in Table 3 include exon dimensions, intron coordinates, and the exon row length.

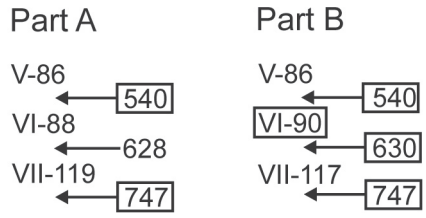


**Fig. 4.** The *Gallus gallus* gene intron No. 6 (coordinate 628 nt) split two neighbour exons – exon VI (dimension 88 nt) and exon VII (119 nt, Part A; see also Table 4). If the intron No. 6 coordinate is virtually changed by 2 nt, e.g., enlargeing exon VI, but at the same time shortening exon VII by 2 nt (without changing the length of the exon row, Part B), an internally regular unit of *TIM* gene is obtained, containing three new structural elements: exon VI (90 nt or 9 × 10), intron No. 6 in a new position (coordinate 630 or 9 × 70), and exon VII (117 nt or 9 × 13). Numerical values of regular parameters are framed.

**Table 5.** Examples of internally regular triosephosphate isomerase (*TPI*) gene structural parameters

| Species | GenBank accession | Q value | Internally regular parameter of the gene |
|---|---|---|---|
| *Drosophila melanogaster* | X57576 | 9 | Exon III dimension, 207 nt (9 × 23) |
| *Caneorhabditis elegans* | U23081 | 9 | Exon III dimension, 324 nt (9 × 36) |
| *Heliothis virescens* | U23080 | 9 | Exon IV dimension, 207 nt (9 × 23) |
| *Oryza sativa* | L04967 | 9 | Intron No. 4 coordinate 324 ny (9 × 36) |
| *Gallus gallus* | M11941 | 9 | Intron No. 5 coordinate 540 (9 × 60) |
| *Aspergillus nidulans* | | 9 | Total exon row length 747 nt (9 × 33) |
| *Heliothis virescens* | U23080 | 9 | Intron No. 3 coordinate 297 nt (9 × 33) |
| *Heliothis virescens* | U23080 | 9 | Total exon row length 504 nt (9 × 56) |
| *Coprinus cinereus* | U23079 | 9 | Total exon row length 756 nt (9 × 84) |

**Table 6.** A short description of the *Heliothis virescens* (tobacco budworm) *TPI* gene exon row structure and intron coordinates (GenBank accession U23080). *, Ordinal numbers of 3'-terminal stop codon nucleotide. **, Changes of intron coordinate dimensions (nt) relative to the natural gene parameters are shown in brackets. Numerical values of the gene parameters which are multiplies of the *TPI* gene family quantum (9 nt) are framed

| No. | Exon dimensions (nt) | Intron coordinates | Internally regular parameter of the gene |
|---|---|---|---|
| a, natural gene GenBank accession U28080 | | | |
| 1 | 75 | 75 | 0 |
| 2 | 112 | 187 | 1 |
| 3 | 110 | 297 | 0 |
| 4 | 207 | 504* | - |
| b, virtual gene model (the gene quantum 9 nt) | | | |
| 1 | 72 | 72 (-3)** | 0 |
| 2 | 117 | 189 (-2) | 0 |
| 3 | 108 | 297 (0) | 0 |
| 4 | 207 | 504 (0) | 0 |

Our analysis of *TPI* gene parameters demonstrated that the "new *TPI*" genes (Logsdon et al. 1995) have the same internal structural regularity as the "old *TPI*". For example, the *Heliothis virescens* gene contained a small internally regular fragment including the intron No. 3 coordinate (297 nt = 9 × 23), exon No. 4 length (207nt; 207 = 9 × 23) and the sum total of the exon row lengths (504 nt = 9 × 56; Table 6).

Particularly interesting seemed the "new *TPI*" gene of *Coprinus cinereus* (Logsdon et al. 1995). No one of the five introns of *Coprinus cinereus TPI* gene was in phase zero. Particularly this feature determined our choice to analyse the internal regularity of this gene (Table 7). The natural numerical parameters of exon lengths and intron coordinates of the *C. cinereus* gene showed no common internal regularity (Fig. 5 A). Virtual transfer

**Table 7.** Calculation of structural parameters of the *Coprinus cinereus* triosephosphate isomerase (*TPI*) gene. The exon lengths (nt) are in accordance with GenBank (accession U23079) data. *, this numerical parameter shows that the length of exon row (a sum total of all exon dimensions) – 756 is a multiple of the *TPI* gene quantum Q (9 × 84 = 756). Thus, the exon row contains a whole number of repeat units

| Exon ordinal number | The first and last nucleotides of exon | Exon length | Intron coordinate | Intron phase |
|---|---|---|---|---|
| I | 144-255 | 112 | 112 | 1 |
| II | 354-477 | 124 | 236 | 2 |
| III | 569-612 | 44 | 280 | 1 |
| IV | 676-874 | 199 | 479 | 2 |
| V | 957-1144 | 188 | 667 | 1 |
| VI | 1216-1304 | 89 | 756* | - |

of intron No. 5 (phase one) with coordinate 667 nt (Table 7) to the nearest supposed knot point (coordinate 666 nt = 9 × 74) at once revealed three internally regular parameters of the gene: the length of exons V (189 = 3 × 3 × 3 × 7) and VI (90 = 2 × 3 × 3 × 5), and the coordinate of intron No. 5 (666 = 2 × 3 × 3 × 37). The length of the exon row of the whole gene including the initiation and termination codons (252 codons = 756 nt; 756 = 2 × 2 × 3 × 3 × 7; Fig. 5 A, C) remained unchanged. The common internal regularity of *C. cinereus TPI* gene parameters was 3 × 3 nt and the gene quantum – 9 nt. These values did not differ from other *TPI* genes (Table 5). Slight correction of the gene parameters in accordance with the gene quantum value (shown in Fig. 5B) revealed that all of the numerical parameters of the gene in such a case were completely internally regular. The sum total of the exon dimensions before and after the virtual intron transfer to phase zero position was not changed (Fig. 5 A, C). The exon rows length corresponded to the whole number of repeat units (756 = 9 × 84). It is necessary to note that two neighbour exon dimensions (IV – 198 nt and V – 189 nt, Fig. 5 C) differed by 9 nt – that is one repeat unit. Comparison of the *Coprinus cinereus* gene-encoded amino acid sequence with intron positions crossing the

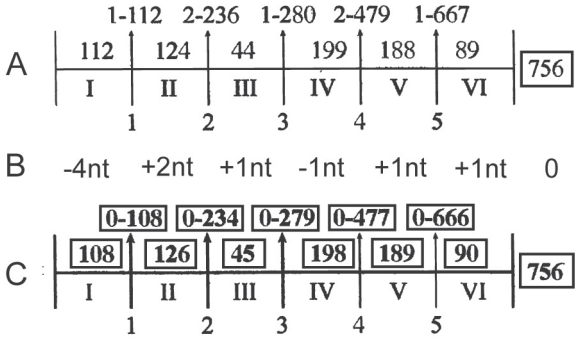


**Fig. 5.** Comparison of the *Coprinus cinereus TPI* gene numerical parameters. A, exon row of the natural gene; B, changes (± nt) of exon dimensions; C, *Coprinus cinereus TPI* gene after virtual transfer of all introns to phase zero position. Internally regular parameters of the gene are framed.

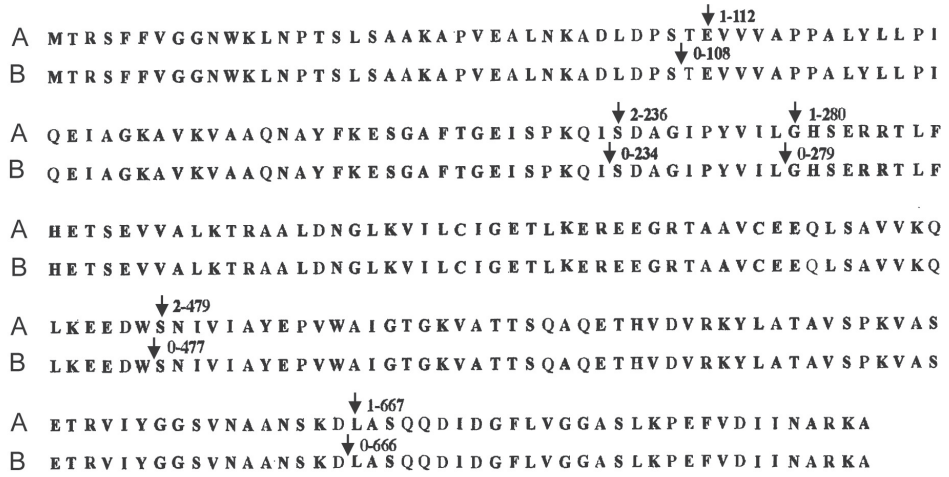


**Fig. 6.** Primary structure of the *Coprinus cinereus* TPI protein. A, all intron positions shown by arrows and characterised by phase and coordinate cross a codon of the corresponding amino acid; B, the duplicate of the same sequence. Intron positions are shown after their virtual transfer to the phase-zero position (i.e., between two neighbour codons). All coordinates are regular and can be expressed as multiples of the *TPI* gene quantum Q = 9 nt.

peptide chain before and after intron transfer to zero position is shown in Fig. 6 A, B.

In homologous genes we can often find introns in similar but not identical positions between genes separated by large evolutionary distances. In accordance to the exon theory of genes these positions represent the same original intron, possibly moved slightly in position (intron drift or sliding; Gilbert et al. 1997). Supporters of the introns late theory regard that introns can not slide and that intron location diversity in homologous genes is a result of intron loss and insertion or reinsertion in neighbour positions (Palmer, Logsdon 1991). Discrete values of gene intron coordinates and the quantization of gene numerical parameters once again show intron drift or sliding, most likely mainly by changing of intron phases. According to our point of view, assertions that introns whose coordinates in homologous genes differ by 1-2 nt were gained (inserted) independently, or that originally resided within an ancestral gene only one or two base pairs apart, lacks biochemical reasoning and are not logical. Interestingly, the analysis of intron positions splitting *TPI* amino acid sequences reveal the same Q values. Comparison of intron positions of amino acid sequences encoded by three different *TPI* genes – those of *Gallus*, *Heliothis* and *Zea* (alignment data from Logsdon et al. 1995) confirmed (Fig. 7), that the size of TPI family protein quanta and RU is 3aa (or 9 nt). The same parameter value was revealed by comparison of *Caernorhabditis* and *Aspergillus TPI* amino acid sequences – the distance between two phase zero (between codons) intron coordinates in the alignment of protein sequences was 9 amino acids or 3 × 3, i.e. 27 nt = 9 × 3 (Fig. 7).

To further elaborate the new theory and model of gene, intron, and exon origin by oligonucleotide multiplication reactions, we will focus on the similarity existing between the model of G13 (lineland) chrystallography and our models of nucleotide multiplication reactions and gene "knot points" – imaginary points in nucleotide multimers between

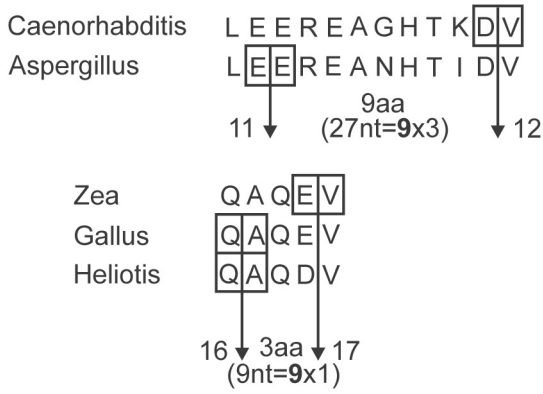


**Fig. 7.** Fragments of TPI amino acid sequences from genes containing known intron positions. Intron positions are indicated as boxes where they fall between codons (phase zero). Distances between two neighbour introns (shown in form of arrows) are indicated by the number of amino acids (aa). The numbers alongside the arrows correspond to intron designation (Logsdon et al. 1995).

identical in size and related in sequence repeat units. We have elaborated a method of determination of the internal regularity of gene numerical parameters – exon dimensions, intron coordinates, length of coding parts (all parameters are expressed by number of nucleotides, nt) by calculation of common factors and have demonstrated that in regular segments the parameters of gene structures are discrete and can be quantized. The gene quantum of small G protein subgroups *ras*, *MRAS* and *YPT* is 12 nt, and of triosephosphate isomerase genes – 9 nt.

## References

Banner D., Bloomer A., Petsko G., Philips D., Pogson C., Willson I. 1975. Structure of chicken muscle triosephosphate isomerase determined crystallographycally at 2.5 Å resolution. *Nature* 255: 609–614.

Burgoine R. 1989. Small GTP binding proteins. *Trends Biochem. Sci.* 14: 354–396.

Casale W., Mcconnell D., Wang S., Lee Y., Linz J. 1990. Expression of a gene family in the dimorfic fungus *Mucor racemosus* which exibits striking similarity to human ras genes. *Mol. Cell. Biol.* 10: 6654–6663.

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

Chipens G., Ievina N. 1999. Repeat units are the basic elements of gene and protein structural organisation. *Proc. Latvian Acad. Sci. Sect. B* 53: 54–55.

Chipens G., Ievina N. 2004a. The code of amino acid interactions. *Latvian J. Chem.* 2: 111–122.

Chipens G., Ievina N. 2004b. Basic language of the protein world. *Latvian J. Chem.* 4: 393–394.

Chipens G., Ievina N., Kalvinsh I. 2005a. A new theory of gene origin and quantization of aspartate aminotransferase gene parameters: mathematical modeling of modern gene structures. *Latvian J. Chem.* 4: 311–324.

Chipens G., Ievina N., Kalvinsh I. 2005b. Quantization of *ras*-related gene parameters. *Latvian J. Chem.* 4: 395–397.

Clapham D. 1996. The G-protein nanomachine. *Nature* 379: 297–299.

De Souza S., Long M., Klein R., Roy S., Lin S., Gilbert W. 1998. Toward a resolution of introns early/late debate: only phase zero introns are correlated with the structure of ancient proteins. *Proc.*

*Natl. Acad. Sci. USA* 95: 5094–5099.
Dietmaier W., Farby S. 1994. Analysis of introns in genes encoding small G-proteins. *Curr. Genet.* 26: 497–505.
Fedorov A., Cao X., Saxonov S., de Souza S., Roy S., Lin S., Gilbert W. 2001. Intron distribution difference for 276 ancient and 131 modern genes suggests the existence of ancient introns. *Proc. Natl. Acad. Sci. USA* 98: 13177–13182.
Gilbert W. 1978. Why genes in pieces? *Nature* 271: 501.
Gilbert W. 1987. The exon theory of genes. *Cold Spring Harbor Symp. Quant. Biol.* 52: 901–905.
Gilbert W., de Souza S.J., Long M. 1997. Origin of genes. *Proc. Natl. Acad. Sci. USA* 94: 7698–7703.
Gilbert W., Marchionni M., McKnight G. 1986. On the antiquity of introns. *Cell* 46: 151–154.
Haubruck H., Prange R., Vorgias C., Gallwitz D. 1989. The *ras*-related mouse Ypt1 protein can functionally replace *YPT1* gene product of yeast. *EMBO J.* 8: 1427–1432.
Hargittai I., Hargittai M. 1995. *Symmetry Through the Eyes of a Chemist.* Plenum Press, New York. 484 p.
Hori K., Kajiwara S., Saito T., Miyazava H., Katayose Y., Shishido K. 1991. Cloning, sequence analysis and transcriptional expression of a *ras* gene of the edible basidomycete *Lentinus edodes*. *Gene* 105: 91–96.
Ievina N., Chipens G. 2003. A new approach to study the origin of genes and introns. *Acta Univ. Latv.* 662: 67–79.
Ievina N., Chipens G. 2004. Origin of globins and a mystery of myoglobin codon root symmetry. *Acta Univ. Latv.* 676: 97–105.
Lewin B. 1994. *Genes V*. Oxford University Press, Oxford. 1926 p.
Logsdon J., Tyschenko M., Dixon C., Jafari J., Valker V., Palmer J. 1995. Seven newly discovered intron positions in the triose-phosphate isomerase gene: Evidence for the introns late theory. *Proc. Natl. Acad. Sci. USA* 92: 8507–8511.
Logsdon J.M.Jr. 1998. The recent origins of splicesomal introns revisited. *Curr. Opin. Genet. Dev.* 8: 637–648.
Long M., de Souza S., Gilbert W. 1995. Evolution of the intron-exon structure of the eucariotic genes. *Curr. Opin. Genet. Dev.* 5: 774–778.
Marchioni M., Gilbert W. 1986. The triosephosphate isomerase gene from maize: introns antedate the plant-animal divergence. *Cell* 46: 133–141.
Noltmann E.A. 1972. Triosephosphate isomerase. In: Boyer P.D. (ed) *The Enzymes*. Vol. 6. Academic Press, New York, pp. 328–340.
Palmer J.D., Logsdon J.M.Jr. 1991. The recent origin of introns. *Curr. Opin. Genet. Dev.* 1: 470–477.
Robberson B., Cote G., Berget S. 1990. Exon definition may facilitate splice site selection in RNAs with multiple exons. *Mol. Cell. Biol.* 10: 84–94.
Straus D., Gilbert W. 1985. Genetic engineering in the pre-cambrian: structure of the chicken triosephosphate isomerase gene. *Mol. Cell Biol.* 5: 3497–3506.
Wichmann H, Disela C., Hanbruck H.,Gallwitz D. 1989. Nucleotide sequence of the mouse *YPT1* gene encoding a *ras* related GTP binding protein. *Nucleic Acids Res.* 17: 6737.

## Gēnu parametru iekšējā regularitāte un kvantēšana

**Nora Ieviņa*, Gunārs Čipēns, Ivars Kalviņš**

Latvijas Organiskās sintēzes institūts, Aizkraukles 21, Rīga LV-1006, Latvija
*Korespondējošais autors, E-pasts: ievina@osi.lv

### Kopsavilkums

Par intronu rašanos un funkcionēšanu eikariotu šūnās ir uzdots ne mazums jautājumu jau kopš to atklāšanas 1977. gadā. Tomēr neviena no pretrunīgajām teorijām šodien nav vispārēji atzīta. Mēs esam izstrādājuši jaunu gēnu un intronu izcelsmes pētīšanas metodiku, izmantojot gēnu skaitlisko parametru salīdzināšanu un to iekšējās regularitātes analīzi. Pētot eksonu dimensijas, intronu koordinātes, gēnu eksonu rindu garumus utml., mūsu mērķis bija parādīt, ka agrīno gēnu priekšteču struktūras ir radušās agīnā evolūcijas stadijā, tās bijušas regulāras un periodiskas un, ka šī regularitāte ir daļēji saglabājusies arī mūsdienu moderno gēnu un atbilstošo proteīnu struktūrās. Iespējamība noteikt gēna kvanta lielumu un kvantēt gēna skaitliskos parametrus visregulārākajām mūsdienu gēnu struktūrām kalpo par pieradījumu jaunās pieejas pareizībai. Šajā darbā mēs izskaidrojam gēnu iekšējās regularitātes analīzes jaunos principus un analīžu metodiku, parādām dažu gēnu segmentu iekšējo regularitāti un nosakām to kvantu skaitliskās vērtības.