STUDIA UBB. CHEMIA, LV, 4, 2010

IMPLICATIONS OF SENSE/ANTISENSE NUCLEIC-ACID CODONS ON AMINO-ACID COUNTS

VLADIMIR R. ROSENFELD*, DOUGLAS J. KLEIN*

ABSTRACT. We study the amino-acid content of protein sequence factors translated from *codonic palindromes* of nucleotide sequences, which have each half comprised from an integer number of codons. Alternatively, our study may be viewed to seek consequences if sense & antisense translations for proteins originate with the two (complementary) strands of RNA.

Under either of these presuppositions, we conclude: the total number of aspartic-acid, asparagine, tyrosine, and histidine residues produced equals the total number of isoleucine, methionine, and valine residues produced. Further, we find a suite of inequalities on amino acid counts. Our results provide a rigorous consequence to a relation considered by Zull *et al*. Further, a "parity rule" of Chargaff *et al*. gives some support for a sense/antisense presumption.

Keywords: *nucleotide sequence, codonic palindrome, translation, parity rule, complementation*

INTRODUCTION

Nucleotide sequences of DNA (desoxyribonucleic acid) and RNA (ribonucleic acid) are constructed from four types of nucleotides denoted by characters **A**, **C**, **G**, and either **T** (for DNA) or **U** (for RNA). DNA consists of two complementary strands, with these four characters matched into two complementary pairs: **A** & **T** and **C** & **G**.

Here, we investigate the consequences of protein translation from both sense & antisense directions along nucleotide sequences. This might [1, 2], sometimes, arise from oppositely oriented translation along strands from complementary DNA strands. Or it can arise from a single RNA strand which is a "codonic" palindrome. It is natural to interrelate amino acids as to whether they have inverted nucleotide codons, and, indeed, such has already been done by Zull & coworkers [3, 4]. From this interrelation (conveniently expressible as a "graph", of vertices representing amino acids, and edges representing the relation), consequences then are sought. Zull & Smith [3] questioned whether 3 portions of this graph correspond to 3 classes of amino acids manifesting

* *Mathematical Chemistry Group, Department of Marine Sciences, Texas A&M University at Galveston, 200 Seawolf Parkway, Galveston, TX 77553-1675, USA, rosenfev@tamug.edu, vladimir_rosenfeld@yahoo.com, kleind@tamug.edu*

different secondary protein structure (α-helix, β-sheet, or random). This could only be a statistical correlation, as many amino acids occur in 2 or 3 types of secondary protein structures (though with different frequencies), and, indeed, Zull & Smith found only a (very) weak correlation. We developed a different type of consequence which however is rigorous, under either the condition of sense/antisense translation of complementary RNA strands or translation from a "codonic" palindrome. We found equal net weights for the frequencies of occurrence of amino acids in two subclasses comprising one of the (bipartite) fragments of the codon inversion graph. Either Chargaff's proposal [5] of forward (sense) & reverse (antisense) translations nucleic acid sequences or Zull's idea of codonic palindromes leads to a general sense/antisense reading of individual codons.

Given a nucleotide sequence, a later disjoint sequence is termed an *inverted repeat* if it consists of the complements of the first sequence in reverse order. The initial sequence and the later inverse repeat are together termed ([6], p. 76) a *complementary palindrome* – elsewhere often termed simply a "palindrome". Sometimes, the direct sequence and its inverted repeat both consist of an integer number of codons. Such a pair of palindromic sequences (or subsequences) consisting exactly of an integer number of codons is called a *codonic palindrome*. We represent the situation when there are s codons in each of RNA sequence by

$$a_1a_2 \cdot \cdot \cdot a_{3s-1}a_{3s} \cdot a_{3s}{}^*a_{3s-1}{}^* \cdot \cdot \cdot a_2{}^*a_1{}^*, \qquad (1)$$

where a_i & $a_i{}^*$ are two complementary nucleotides (say, **C** & **G**) of the nucleotide alphabet $\mathcal{A}$ = {**A**, **C**, **G**, **U**}. Note: the "codonic" condition on this (complementary) palindrome means the direct & inverted sequences each comprise an integer number of codons.

A more general notion allows "concatenation" of different codonic fragments of a codonic palindrome. The codons of a codoinic palindrome may be moved around to be placed in different positions, still preserving all codons, just in a different order. We term such a reassemblage a *codonic palindromic conglomerate*.

This allows consecutive codonic-palindromic loops (such as occur with introns), and this also accounts for nested loops (*i. e.*, loops of smaller size inserted into contour sequences of loops of larger size), it allows even multiply nested loops. We may imagine: at the first hypothetic stage, starting from a single giant codonic palindrome, with direct sequence t and inverted repeat u, each of which are to be broken up into codon subsequences, say as (t_1, t_2, . . . , t_m) and (u_s, u_{s-1}, . . . , u_1), with possibly different numbers of different-lengthed subsequences t_i & u_j from t & u; and at the second step, putting these different subsequences back together in an arbitrary order. The superpalindrome need not be biologically realized but rather just the intermixed codonic palindromic conglomerates.

Granted these ideas, we develop some formalism in the next section, so as to identify notable consequences on the numbers of amino acids formed within different selected groups, under the assumption that the RNA is a codonic palindromic conglomerate. Most of the formal discussion is not needed to understand the final biological consequences, which come in Propositions 4, 5, & 6.

FORMAL RESULTS

To manipulate nucleotide sequences, one may use three commuting operators: α standing for complementation (as indicated by (*) in (1)) of nucleotides in a string; β for inversion of the string, and the composition $\gamma = \alpha\beta = \beta\alpha$. We can formally rewrite (1) using γ:

$$a_1a_2\cdot\cdot\cdot a_{3s-1}a_{3s}\cdot\gamma(a_1a_2\cdot\cdot\cdot a_{3s-1}a_{3s}). \qquad (2)$$

Let $\mathcal{B} = \{b_1, b_2, \ldots, b_{21}\}$ be the set of 21 amino acids (where the 21st amino acid terminologically corresponds to the triple of stop codons). For a nonempty subset $S \subseteq \mathcal{B}$ of amino acids, denote by $C(S)$ the set of all codons for the amino acids from S. And let $\gamma C(S)$ denote the result of action of the operator γ on each codon in $C(S)$.

We investigate the consequences of a pair of subsets S_1 and S_2 of amino acids, for which $\gamma C(S_1) = C(S_2)$, or equivalently $\gamma C(S_2) = C(S_1)$, since γ is idempotent (*i. e.*, $\gamma^2 = 1$), as also are α and β.

Lemma 1. *Let T_1 & T_2 be two sets of amino acids such that $C(T1) = \gamma C(T2)$. Let $a = a_1a_2\cdot\cdot\cdot a_{3s-1}a_{3s}\cdot a_{3s}{}^*a_{3s-1}{}^*\cdot\cdot\cdot a_2{}^*a_1{}^*$ ($a_i, a_i{}^* \in \mathcal{A}$; $1 \le i \le 3s \ge 3$) be a codonic palindrome. Moreover, let l_j (res. r_j) (j = 1, 2) be the total number of occurrences of codons belonging to $C(T_j)$ in*

$$a_1a_2\cdot\cdot\cdot a_{3s-1}a_{3s} \text{ (res. } a_{3s}{}^*a_{3s-1}{}^*\cdot\cdot\cdot a_2{}^*a_1{}^*\text{). Then } l_1 = r_2 \text{ \& } r_1 = l_2.$$

Proof. Since $a = t\gamma(t)$, with $t = a_1a_2\cdot\cdot\cdot a_{3s-1}a_{3s}$ & $\gamma(t) = a_{3s}{}^*a_{3s-1}{}^*\cdot\cdot\cdot a_2{}^*a_1{}^*$, the numbers of "direct" and inverted codons in a are equal. Also, by construction, $C(T_1)$ and $C(T_2)$ are sets of mutually inverted codons, whence we immediately arrive at the proof.

Note: for any codon t representing a respective amino acid b_i, the corresponding codon $u = \gamma(t)$ always represents a distinct amino acid b_j. Therefore, γ induces a binary relation on the set $\mathcal{B}$ of all amino acids which can be represented thereon by a simple graph Γ, where amino acids b_i & b_j are adjacent (linked by an edge) if there exist a codon t of the former and a codon u of the latter which are interchanged by γ ($u = \gamma(t)$ & $t = \gamma(u)$). Important here are the connected components (maximal connected subgraphs of Γ). We immediately use these considerations in the following

Lemma 2. *Let T_1 &T_2 be two sets of amino acids with $C(T_1) = \gamma C(T_2)$. Then, for any $U_1 \subseteq T_1$ corresponding to a connected component of Γ, either U_1 is entirely in T_2 or else entirely external to T_2 (i. e., either $U_1 \cap T_2 = U_1$ or $U_1 \cap T_2 = \varnothing$).*

Proof. Associate to the union $T_u = T_1 \cup T_2$ a graph H whose vertex set is T_u, and two vertices i & j are adjacent in H if there exist codons t_i & t_j such that $t_i = \gamma(t_j)$. Now, attach exactly one self-loop to every vertex of H to obtain a derivative graph $\hat{H}$ having the same connectivity components. Clearly, $\hat{H}$ is an equivalence relation on T_u where any pair of vertices i and j are equivalent iff these belong to one connected component. Indeed, three conditions are satisfied: (i) reflexivity, as guaranteed by 'self-connectivity' of every vertex having an attached self-loop; (ii) symmetry, since $t_i = \gamma(t_j) \Leftrightarrow t_j = \gamma(t_i)$; and (iii) transitivity, as follows from the connectivity within a component. Evidently, in our hypothesis, U_1 is a single equivalence class of vertices of T_u, while T_2 is the union of a number of equivalence classes of vertices thereof. Since two equivalence classes of objects either coincide or share no element, U_1 is either included as one such class in T_2 or intersects with no equivalence class of vertices comprising T_2. This completes the proof.

Corollary 2.1. *Let T_1 &T_2 be two sets with $T_1 \neq T_2$ and $C(T_1) = \gamma C(T_2)$. Then, if T_1 & T_2 are minimal, they are disjoint*.

Proof. This uses reasoning similar to Lemma 2. Namely, minimal sets T_1 and T_2 are both equivalence classes of $T_u = T_1 \cup T_2$. Since $T_1 \neq T_2$, we immediately arrive at the proof.

Corollary 2.2. *Let $a = a_1a_2 \cdot \cdot \cdot a_{3s-1}a_{3s} \cdot a_{3s}{}^*a_{3s-1}{}^* \cdot \cdot \cdot a_2{}^*a_1{}^*$ ($a_i, a_i{}^* \in \mathcal{A}$; $1 \leq i \leq 3s \geq 3$) be a codonic palindrome. Moreover, let n_j (j = 1, 2) be the total number of occurrences of codons belonging to $C(T_j)$ in $a = a_1a_2 \cdot \cdot \cdot a_{3s-1}a_{3s}$ (res. $a^* = a_{3s}{}^*a_{3s-1}{}^* \cdot \cdot \cdot a_2{}^*a_1{}^*$). Then $n_1 = n_2$.*

Proof. Note that $n_j = l_j + r_j$ (j = 1, 2). By virtue of the equalities $l_1 = r_2$ and $r_1 = l_2$ demonstrated in Lemma 1, the proof is immediate.

Proposition 3. *In a codonic palindromic conglomerate, there are equal amounts n_1 and n_2 of amino acids from respective minimal subsets T_1 and T_2, as in Corollary 2.1.*

Proof. The initial codonic superpalindrome has $n_1 = n_2$, by Corollary 2.2. But breaking up into codons and rearranging all the various codons does not change the numbers of the different codons, so that one still has $n_1 = n_2$.

Next, we frame these results more biologically.

AMINO-ACID COUNTS

The relation γ which acts on a nucleic acid string to reverse & complement it leads to a relation between amino acids: if an amino acid has codon $u = a_1a_2a_3$, then, it is related or linked to the amino acid with $\gamma(u) = u_3{}^*u_2{}^*u_1{}^*$.

This overall γ-relation is conveniently represented as a graph Γ where an edge occurs between the amino acids of codons u & $\gamma(u)$. Using the standard codons (*e. g.*, as in Ch. 13 of [7]), the graph Γ appears in Fig. 1 – also given by Zull *et al.* [4]. But now (following the results of our preceeding section) we seek a minimal pair of subsets S_1 & S_2 of amino acids for which $\gamma C(S_1) = C(S_2)$, and $\gamma C(S_2) = C(S_1)$, since $\gamma^2 = 1$. It turns out that in Γ there is a pair of such sets: S_1 = {D, N, T, H} & S_2 = {I, M, V}, where D, N, T, H, I, M, V denote aspartic acid, asparagine, tyrosine, histidine, isoleucine, methionine, and valine, consecutively. One sees that our sets S_1 & S_2 are mutually interconnected while being completely disconnected from the remaining vertices. The corresponding codon sets are $C(S_1)$ = {**GAU**, **GAC**; **AAU**, **AAC**; **UAU**, **UAC**; **CAU**, **CAC**} and $C(S_2)$ = {**AUU**, **AUC**, **AUA**; **AUG**; **GUU**, **GUC**, **GUA**, **GUG**}. Application of the operator γ to $C(S_1)$ gives $\gamma C(S_1)$ = {**AUC**, **GUC**, **AUU**, **GUU**, **AUA**, **GUA**, **AUG**, **GUG**}, which is just $C(S_2)$. Hence, also $\gamma C(S_2) = C(S_1)$. This is the only pair of minimal distinct sets S_1 & S_2 of amino acids having the described property in Γ (to transform quantitatively into each other under the operator γ). The remnant set $S_3 = B \setminus S_1 \cup S_2$ of amino acids gives a minimal set $C(S_3)$ of codons closed under the action of γ.

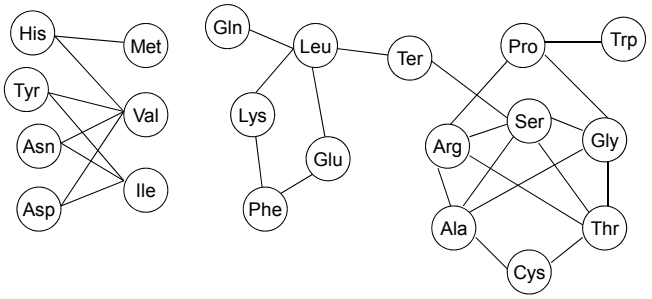

Figure 1: The graph Γ of γ-relations of amino acids; the left bipartite component displays the sets S_1 (the 4-site part: Hys, Tyr, Asn, Asp) and S_2 (the 3-site part: Met, Val, Ile), while the right component displays the set S_3 .

A codonic palindromic conglomerate merely conditions codons to occur in complementary pairs. Thence, allowing several codonic palindromes nested, or multiply nested, or interlinked in all kinds of ways. Instances of such objects can occur in introns. Recall that the mRNA of eukaryotes is obtained through *splicing* from pre-mRNA (precursor mRNA), which is similar to a portion of a strand of DNA. During splicing, relatively long factors called *introns* are removed from a pre-mRNA sequence. Most introns are 80 to 400 base pairs in size; though there also exist huge introns of length >10,000. While introns do not themselves participate in producing amino acids, it is of note that the intronic loops even of a very high degree are covered in the conditions of Proposition 3, where $n_1 = n_2$ is achieved. More explicitly for (T_1 &T_2 of Proposition 3 realized as) S_1 & S_2 in Fig. 1, with $\#_X$ being the number of amino acid moieties X produced, we arrive at a primary biological result:

Proposition 4. *In protein factors translated from codonic palindromic conglomerates, such as occur with various stem loops, numbers of amino-acid residues are related*

$$\#Asp + \#Asn + \#Tyr + \#His = \#Ile + \#Met + \#Val.$$

But, granted our codonic palindromic conglomerates, there are further (weaker) consequences, concerning inequalities on amino acid numbers. In particular, we have:

Proposition 5. *In protein factors translated from codonic palindromic conglomerates, there are inequalities on the numbers of different amino acids:*

$$\#Met \le \#His;$$
$$\#His \le \#Met + \#Val;$$
$$\#Ile \le \#Tyr + \#Asn + \#Asp;$$
$$\#Tyr + \#Asn + \#Asp \le \#Ile + \#Val;$$
$$\#Gln \le \#Leu;$$
$$\#Trp \le \#Pro;$$
$$\#Ter \le \#Leu + \#Ser;$$
$$\#Pro \le \#Trp + \#Arg + \#Gly;$$
$$\#Ser \le \#Ter + \#Arg + \#Gly + \#Ala + \#Thr$$
$$\#Cys \le \#Ala + \#Thr;$$
$$\#Ala + \#Thr \le \#S\ er + \#Arg + \#Gly + \#Cys;$$
$$\#Leu + \#Phe \le \#Gln + \#Lys + \#Glu + \#Ter;$$
$$\#Gln + \#Lys + \#Glu \le \#Phe + \#Leu;$$
$$\#Gln + \#Lys + \#Ter + \#Glu \le \#Leu + \#Phe + \#Ser;$$
$$\#Trp + \#Arg + \#Gly + \#Cys \le \#Pro + \#S\ er + \#Ala + \#Thr;$$
$$\#Pro + \#Ser + \#Cys \le \#Trp + \#Ter + \#Arg + \#Gln + \#Ala + \#Thr;$$

where the number #Ter of "stops" is conveniently identified to the number of different proteins.

Proof. Our proof begins with a transformation of G of Fig. 1 into a symmetric digraph Γ where each edge of G is converted into a pair of opposite directed arcs between the same two vertices (as originally connected by the replaced edge). We attach to every arc $i\,j$ of Γ a weight a_{ij} equal to the total multiplicity of all codons representing amino acid i which are transformed by the operator γ into codons of amino acid j. Next, we use a (common mathematical) definition that a subset I of vertices of G is *independent* if no two vertices of I are adjacent in G. Any independent subset I of amino acids (nontransformable one into another by γ) determines the set $J = N(I)$ of all amino acids adjacent to members of I. Evidently, the operator γ transforms all codons of amino acids from I into codons representing amino acids from J, but the converse is true if no two amino acids of J are adjacent in Γ (or Fig. 1). In general, there holds a (nonstrict) inequality interrelating the total numbers of codons of I transformed into codons of J and of codons of J itself, taking into account other possible transformations of codons of $N(I)$ – not into codons of I. Thus, we deduce

for the total numbers of codons in I and J that $|C(I)| \leq |C(J)|$. Hence, particular proofs for all cases considered in Proposition 3 follow, with different choices of independent I & neighbors $J = N(I)$ corresponding to the left & right sets of amino acids in each of these inequalities.

Presumably, these statements are most important when at least most of the RNA (or DNA) is comprised from codonic palindromic conglomerations. But, perhaps, most significantly Propositions 4 and 5 hold under the sense/ antisense circumstance proposal in [1] & [2], [5] and explored in [3] & [4]. That is:

Proposition 6. *If in place of the condition of codonic palindromic conglomerates in Propositions 4 & 5, the protein factors are translated from RNA, obtained from both (sense & antisense) DNA strands, then the conclusions 4 & 5 still hold.*

Proof. The two corresponding RNA strands may be viewed as a single codonic palindrome, say each of the strands being separated from one another by a hypothetical "stop" codon. Thence, Propositions 4 & 5 apply.

DISCUSSION

Comparison of proportions of amino acids as indicated by Propositions 4 & 5 are perhaps of practical interest. Clearly, 4 & 5 are most relevant when all or at least a major part of the RNA (or DNA) is comprised from codonic palindromes – in as much as the various indicated amino acids may be coded for in different amounts by the portion of the nucleotide chain outside the codonic palindromes. Chargaff & coworkers' "parity rule" [5, 8–10] is in general a little weaker than the hypothesis of 4 & 5, but still is supportive of it, for some selected species. Most significantly, our results on amino-acid counts apply fully if the sense/antisense hypothesis of [1, 2] is met. As such, our Proposition 6 offers a strong test of the occurrence of sense/antisense translations – such as we imagine though not a general occurrence, could be the situation for selected species.

Further, note that a "parity rule" of Chargaff and coworkers [5] suggests that, in a wide class of single strands of DNA, the numbers of **A**&**T** nucleotides match as also do the numbers of **C**&**G** nucleotides. (This seems to occur [8] especially for eubacterial and chloroplast DNA.) That is, granted the satisfaction of this Chargaff's rule, single DNA strands have met (in our formal nomenclature) a first condition for the whole DNA molecule to be a codonic palindromic conglomerate. A strengthening of this rule to say that complementary nucleotides fully "condense" into complementary
codons would then imply our result for a single strand of DNA.

Again, our ideas are related to Zull and coworkers [3, 4], though they look at the possibility of the graphic structure of Fig. 1 to be statistically manifested in secondary protein structures, whereas what we focus on is what might be termed "0-ary" structure (of amino acid counts). A further point is that our results (of Propositions 4, 5, 6) are robust to certain rare complications involving the rare alternative translation of a "stop" to some

other rare amino acid – and this may be seen not to hurt any of the inequalities in 5. For instance, the “stop” codon **UGA** can in certain mitochondria code for tryptophan and for selenocystein in certain Archaea. Also, this occurs because [12] the second stop codon **UAG** can code for pyrrolysine in Archaea and bacteria.

Besides, the (standard) mode of forming RNA loops, another hypothetic possibility might be imagined to form “reverse loops” (*i. e.*, helixlike loops) interconnecting between a directed sequence and a second sequence of nucleotides which, though complemented from the first sequence, is not reversed in direction along the strand. If such is imagined: namely, to occur (as has indeed been entertained as a possibility by Chargaff *et al.* [10]), one could then inquire about the numbers of different amino acids which arise from two so-related sequences. That is, one would inquire about the interrelated amino acids, considering our complementation operator α as interrelating the two nucleotide sequences – conglomerated or not. Then, the same sort of results found in our formal section apply, with γ replaced by α, now with reference to the α-graph of Fig. 2.

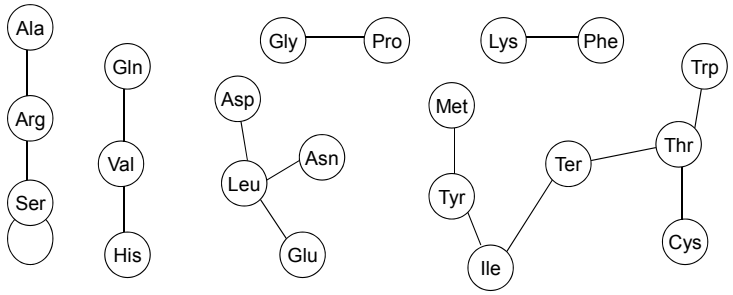

Figure 2: The graph of α-relations of amino acids.

With many bipartite components in this graph, this would evidently lead to a multiplicity of interrelations amongst numbers of various amino acids. For instance, this would imply that the amounts of glycine & and proline are the same (and also the amounts of lysine & phenylalanine) – seemingly, these equalities (and more) do not occur, so that the pairing between a direct sequence and a second sequence in a strand in the same direction, evidently, does not occur. The apparent reason must be that, *e. g.*, the pairing between **C** & **G** occurs only when the two nucleotides in making contact are oppositely oriented along a nucleotide chain, whence we might in fact distinguish the possibilities by $\vec{\mathbf{C}}\,\&\,\vec{\mathbf{G}}$ for nucleotides oriented in one “symparallel” direction along the chain, and $\overleftarrow{\mathbf{C}}\,\&\,\overleftarrow{\mathbf{G}}$ in the other direction along the chain – so that pairing occurs between antiparallel [10] $\vec{\mathbf{C}}\,\&\,\overleftarrow{\mathbf{G}}$ (or

between $\vec{\mathbf{G}}\&\overleftarrow{\mathbf{C}}$), but not between "symparallel" $\vec{\mathbf{C}}\&\vec{\mathbf{G}}$ (or $\overleftarrow{\mathbf{C}}\&\overleftarrow{\mathbf{G}}$). A similar comment applies for **A** & **U** (or **T**). That is, the conformational structure of each nucleotide is evidently different along the two different directions of a chain. Overall this evidently accounts for the fact that nucleotide sequences always form copies in antiparallel directions, rather than symparallel directions (with complementation). This, seemingly, is an evolutionarily selected (or) condition for faithful transcription.

CONCLUSION

Beyond the presumption of sense/antisense reading of codons, our exposition here arises from just very basic facts of molecular genetics. Under such (sense/antisense) conditions, we have found novel biological consequences enounced in Propositions 4, 5, and 6. Being rigorous consequences of these conditions, the amino-acid count relations may be used as tests for either Chargaff's sense/antisense hypothesis (in RNA) or for our codonic palindromic conglomerate condition (whence, then, Zull's hypothesis). That is, if our amino-acid conditions are not met, then this denies both Chargaff's and Zull's hypotheses. Finally, we may mention two other recent works [13, 14] which consider similar biological matters in a wider algebraic context.

ACKNOWLEDGMENTS

We thank Profs. Alexandru Balaban (Galveston) and James Zull (Cleveland) for their comments. Also, our thanks are addressed to Dr. Anton N. Ryzhov (Moscow, Russia) for his help in preparing the drawings. Support (through grant BD–0894) from the Welch Foundation of Houston, Texas, is acknowledged.

REFERENCES

1. L.B. Meckler, *Biofizika*, **1969**, *14*, 581.
2. R.G. Idlis, *Zhurnal Vses. Khim. Ob-va im. Mendeleeva*, **1980**, *25*, 431.
3. J.E. Zull and S.R. Smith, *Trends in Biochemical Sciences*, **1990**, *15*, 257.
4. J.E. Zull, R.C. Taylor, G.S. Michaels, and N.B. Rushforth, *Nucleic Acids Research*, **1994**, *22* (*16*), 3373.
5. R. Rudner, J.D. Karkas, and E. Chargaff, *Proc. Natl. Acad. Sci. USA*, **1968**, *60*, 921.
6. M.G. Sadovsky, J.A. Putintseva, and A.S. Shchepanovsky, *Theory Biosci*., **2008**, *127*, 69.
7. J.D.D. Watson, "*Molecular Biology of the Gene*", 3rd ed., W. A. Benjamin Inc., New York, **1976**.
8. C. Nikolaou and Y. Almirantis, *Gene*, **2006**, *381*, 34.

9. D.R. Forsdyke and J.R. Mortimer, *Gene*, **2000**, *261*, 127.
10. E. Chargaff, *Nature,* **1965**, *206*, 145.
11. S. Castellano, A.M. Andrés, E. Bosch, M. Bayes, R. Guido, and A.G. Clark, *Mol. Biol. Evol.*, **2009**, *26* (*9*), 2031.
12. G. Srinivasan, C.M. James, and J. A. Krzycki, *Science*, **2002**, *296* (*5572*), 1409.
13. V.R. Rosenfeld, *MATCH Commun. Math. Comput. Chem.*, **2006**, *56 (2)*, 281.
14. V.R. Rosenfeld., *MATCH Commun. Math. Comput. Chem.*, **2007**, *57 (1)*, 134.