Asociación Argentina
de Mecánica Computacional

Mecánica Computacional Vol XXX, págs. 2853-2862 (artículo completo)
Oscar Möller, Javier W. Signorelli, Mario A. Storti (Eds.)
Rosario, Argentina, 1-4 Noviembre 2011

FROM UNFOLDED SEQUENCES TO STRUCTURAL MOTIFS

Blanca I. Niel[a], Walter A. Reartes[a] and Nélida B. Brignole[b,c]

[a]*Departamento de Matemática, Universidad Nacional del Sur, Av. L. N. Alem 1253, B8000CPB Bahía Blanca, Argentina, biniel@criba.edu.ar, reartes@uns.edu.ar, http://www.matematica.uns.edu.ar/*

[b,c]*Laboratorio de Investigación y Desarrollo en Computación Científica (LIDeCC), Universidad Nacional del Sur (DCIC-UNS), B8000CPB Bahía Blanca, Argentina, dybrigno@criba.edu.ar, http://www.lidecc.cs.uns.edu.ar/*

[c]*Planta Piloto de Ingeniera Química (UNS - CONICET), Cno La Carringanda Km.7, B.Blanca, Argentina, dybrigno@criba.edu.ar, http://www.plapiqui.edu.ar/*

Keywords: Homologous sequences, Biopolymer Folding, Base Pairing Contact Patterns, Repeats, Palindromes, Hairpin Loops.

Abstract. We consider that the biopolymer chain forms intramolecular contacts at the expense of losing conformational freedom. By means of a cohesive arithmetic method we have analysed the feasible base pairing contact patterns in nucleotide subsequences from untangled to knotted biopolymer chains. The procedure associates complementary integers to each of the two W-C base pairings of the four bases that integrate the homologous sequences. Given the number of the basic building blocks and the quantities of each of the nucleotides in the biopolymer, the unfolded sequence is obtained searching for the admissible Hamiltonian paths of different lengths under the penalisation of any zipping and/or stacking of base pairing processes. Finally, by comparing backward and forward readings in the obtained unfolded sequences we could individualize certain structural features of biological interest which are deployed in scientific papers about nucleic acids (J. P. Sheehy et. al., RNA, 16(2):417-429 (2010); A. Fernández, B. Niel and T. Burastero, Biophysical Chemistry, 74(2):89-98 (1998)).

1 INTRODUCTION

We consider that the biopolymer chain forms intramolecular contacts at the expense of losing conformational freedom. The *thermodynamic tenet* that "any feasible folding event for loop closure undergoes an enthalpic loss smaller than the entropic contribution ", e.g. Fernández, A., and Cendra, H. (1996) and the compilation of the *thermodynamic parameters* in oligonucleotides, e.g. Sheehy et al. (2010) allow us to deploy certain statements for a comprehensive analysis in single stranded biopolymer folding events. Furthermore, since *biopolymer folding is an expeditious process taking place within timescales incommensurably shorter than ergodic times* (Fernández, A., Niel, B., and Burastero, T., 1998) we assert three statements in order to replicate a least effort principle reflecting stepwise minimization of the conformational entropy cost with concurrent maximization of the base-pairing contacts for each folding event. By means of a cohesive arithmetic method we examine the feasible base pairing contact patterns (BPPs) in nucleotide subsequences from untangled to knotted oligonucleotides. For the four RNA nucleotides we adopt the standard notation **G**, **C**, **A**, **U**, where **G** = guanosine, **C** = cytosine, **A** = adenine and **U** = uracil, or eventually **T** = thymine for DNA residues. The Watson-Crick base pairing based upon the complementarity **G** − **C** and **A** − U (or eventually **A** − **T**). The procedure associates complementary integers to each of the two base pairings of the monomers that integrate the homologous sequences, (e.g. **G**$\sim \lfloor \frac{nt}{2} \rfloor + 1$ and **C**$\sim \lfloor \frac{nt}{2} \rfloor$, **A**$\sim \lfloor \frac{nt}{2} \rfloor + 2$ and **T**(**U**)$\sim \lfloor \frac{nt}{2} \rfloor - 1$; nt number of units). Given nt and the quantities of each of the monomers in the biopolymer, the unfolded sequence is obtained searching for the admissible Hamiltonian paths of order $nt - 1$ (or $nt - u_d$, in case of u_d discarded units) under the penalization of any zipping and/or stacking of elementary base pairing contacts. In other words, subsequences like -**GCGCGCGCGC**-, -**AUAUAUAUAUAU**-,-**GCAUGCAUGCAU**- are forbidden in the unravelled chain, because they beget by a kink or twisted turn the formation of zippers or stems of base-pairings (bps). On the contrary, in order to contrive a sequence that folds onto itself it is imposed the crucial definitions of cooperative and associative wrapping effects (Fernández, A., 2010). Herein, we propel the folding process under the following assumptions: the cooperative effect in the loop formations when the entropy loss associated to a loop closure of determined length is computed following the approach in Fernández, A., Niel, B., and Burastero, T. (1998). This involves the re-scaling of the effective loop lengths with the consequent short, medium and long term interactions during the folding stages. Furthermore, we impose that each base-pair, e.g. **AU**, **GC** as well as any of the specific quartets built by the four bases **GCAU**, and its cyclic permutations, shorten the end to end distance of the nucleotide sequence. Therefore, after the resolution of this special combinatorial problem of searching for the Watson-Crick complementary regions with antiparallel orientation, the base pairing contact matrix (BPPM) or the comparison of backward and forward readings in the obtained unfolded sequences allow us to individualize certain structural features of biological interest which are deployed in scientific papers about nucleic acids (Sheehy et al., 2010; Lu et al., 2006; Wieczorech, K. P. et al. , 2006; Kulinski, T. et al., 2003; Fernández, A., and Cendra, H., 1996; Zarrinkar, P. P., and Williamson, J. R., 1994).

We set in the context of this work in the preliminaries, Abstract and INTRODUCTION. §2 deploys THE PROPOSED METHODOLOGY. The main statements of the technique are at §2.1 "*A cohesive arithmetic algorithm*". While in §2.2 "*Matrices and Base Paring contact Patterns*" we choose BPPMs to sketch the oligonucleotide secondary structure motifs. §3 DIRECT APPLICATIONS TO OLIGONUCLEOTIDES shows unsophisticated applications of the procedure. Finally, we set forth the scope of our proposal in §4 CONCLUSIONS.

2 THE PROPOSED METHODOLOGY

We assume that the biopolymer built by four residues folds by cooperative bp contacts of simultaneous occurrence in distinctive domains of the macromolecule. Therefore, the secondary structure of oligonucleotides play a key role in the wrapping and folding process. The biophysical scenery of the chain is nearly the same of a polyelectrolyte in solution. The conformational space is resolved at the level of the transitions between base pairing contact patterns (BPPs). Every folding step is defined by a cooperative equilibrium between loop formation and base pairing (bp) stacking (stem). Consequently, each effective stage undegoes local minimization of the loss in conformational freedom and maximization of the number of intramolecular contacts. The rate determining step is the entropy loss (ΔS_{loop}) associated to a loop closure. The admissible loop motifs are: bulge, hairpin, eightshape or internal, and pseudoknot. Furthermore, the entropy change is a function of the number of the unpaired bases in the loop (Fernández, A., and Cendra, H., 1996). On the other hand, the released heat for the bp stackings is proportional to the number of the fastened bridges.

2.1 A cohesive arithmetic algorithm

Let n_t be the number of monomers that made up the oligonucleotide, and let $q_{\mathbf{G}}$, $q_{\mathbf{C}}$, $q_{\mathbf{A}}$ and $q_{\mathbf{U(T)}}$, denote the respective quantities of the bases **G**, **C**, **A** and **U**(**T**). The unfolded nucleotide sequence is built by breaking away the cooperative and complementary rules to deter the formation of any sequentially coiled structure. Therefore, we built the untangled maximal sequence by hampering the turns **AU**, **GC**, **UA**, **CG**, the tetraloop **GACU**, and its 23 permutational structural motifs; the kink **GGGC**, and each of its twisted motifs as well as their respective **CCCG** complementary quartets. We impose a cohesive arithmetic for each base-pairing, it is done to maintain the sequence ends nearby in a globular array. For example, if **G**, **C**, **A** and **U** are the nucleotides, "**the complementary associated integers** ", noted as $e(\mathbf{G})$, $e(\mathbf{C})$, $e(\mathbf{A})$, $e(\mathbf{U})$, are given after the selection of one of the following equations for specific $k_i \neq k_j$ integers:

$$\begin{aligned} e(\mathbf{C}) = k_j\,; \quad e(\mathbf{G}) = n_t - k_j \quad & 1 \leq k_j \leq \lfloor \frac{n_t}{2} \rfloor \\ k_j \neq k_i \quad & \\ e(\mathbf{A}) = k_i\,; \quad e(\mathbf{U}) = n_t - k_i \quad & 1 \leq k_i \leq \lfloor \frac{n_t}{2} \rfloor \end{aligned} \tag{1}$$

or

$$\begin{aligned} e(\mathbf{G}) = k_j\,; \quad e(\mathbf{C}) = n_t - k_j \quad & 1 \leq k_j \leq \lfloor \frac{n_t}{2} \rfloor \\ k_j \neq k_i \quad & \\ e(\mathbf{U}) = k_i\,; \quad e(\mathbf{A}) = n_t - k_i \quad & 1 \leq k_i \leq \lfloor \frac{n_t}{2} \rfloor \end{aligned} \tag{2}$$

First Main Statement: *The "**cohesive condition** ", **Eq. (1)** or **Eq. (2)**, establishes that the tetraloop* **GACU** *and its circular reorderings are always feasible tetraloop motifs independently of the whole numbers of monomers that build the nucleotide sequences. Normally* n_t *is odd, see Maizel et al. (1981).*

For example, the fundamental penalizations of **G**-**C** and **A**-**U** base pairing bindings for $k_j = \lfloor \frac{n_t}{2} \rfloor$, and $k_i = \lfloor \frac{n_t}{2} \rfloor - 1$, in Eq. (1), are the structural motifs in Table 2 at page 9.

Focusing on physical chemistry aspects, the "cohesive condition", i.e. Eq. 1 or Eq. 2, defines a complementary potential between anionic and cationic ligand bindings.

If $S = \{\cdots \mathbf{U}, \mathbf{G}, \mathbf{G}, \mathbf{A}, \cdots \mathbf{AG}, \cdots \mathbf{U}, \mathbf{C}, \mathbf{C}, \cdots \mathbf{C}, \cdots\}$ is an *arbitrary sequence* of an odd number n_t of residues, "the integer associated to the sequence S ", denoted $e(S)$ is given by $e(S) = \sum_{i=1}^{n_t} e(N_i)$ (modulo n), herein N_i is the nucleotide (residue) located at i^{th} position in S.

Let $S_m : \{N_1, N_2, \cdots, N_m\}$ be a finite sequence of residues N_i. A "*proper subsequence*" of S_m is every subsequence S_{jr} such that: $S_{jr} : N_j, N_{j+1}, \cdots, N_{j+r-1}, N_{j+r}$ for $1 \leq j \leq m$ and $r \geq 0$ with $j + r < m$ if $j = 1$, and $j + r \leq m$ if $j > 1$.

Second Main Statement: *"**the arithmetic algorithm**": An arbitrary sequence $S_{n_t} : \{N_1, N_2, \cdots, N_{n_t}\}$ of n_t residues N_i determines a closed unfolded sequence of order n_t, denoted by $(U_F)^{n_t}_{\circ}$, if and only if any proper subsequence has associated integer neither n_t nor a multiple of n_t and $e(S) \equiv 0$.*

Let $(U_F)^m_{\sim}$ stand for any open unfolded sequence of order m less than or equal to $n_t - 1$. Meanwhile, $L^m_{\circ}$ denotes a loop motif builds by m nucleotides.

The single stranded oligonucleotides fold and coil onto themselves until the final compact biological structure. The understanding of the folding process involves enormous widespread data of distinctive scientific works. Particularly, those papers devoted to the study of the secondary structure motifs of the kinetic intermediates in the folding and wrapping process. Herein, enormous efforts are strained in biological and thermodynamic parameter updated compilation (Sheehy et al., 2010); (Lu et al., 2006); (Fernández, A., 2010) as well as those consecrated to biotechnology or bioengineering developments (Fernández, A., 2010).
We focus on the secondary structure of single stranded oligonucleotides, released in different scientific works (Zarrinkar, P. P., and Williamson, J. R., 1994), (Kulinski, T. et al., 2003), since the rate-determining step is the synchronous loop and stem formations in a multidomain scenery. Consequently, "*the dominant folding pathway, the brachistochrone, at each stage minimized the entropy loss associated with loop closure with a synchronic maximization of the number of effective base-pairing contacts*" (Fernández, A., and Niel, B., 1997; Fernández, A., Niel, B., and Burastero, T., 1998). Then, the understanding of the biopolymer folding process begeted at a roughly level of the (BPPs) has equivalent explanations by base-pairing contact matrices (BPPMs) as transitions of quasi-equilibrium states in the conformational space of a given arbitrary primary sequence.

Third Main Statement: *"**Max. Order & Redundance vs Uniqueness**": It is given the quantities of **G** ($q_{\mathbf{G}}$), **C** ($q_{\mathbf{C}}$), **A**($q_{\mathbf{A}}$), and **U** ($q_{\mathbf{U}}$) of an arbitrary sequence of nucleotides the unfolded sequence could not exist at the maximum n_t length. Then, it is necessary the procedure searchs for the unfolded sequences of order $n_t - 1$ or $n_t - u_d$, u_d the number of discarded residues. Moreover, under the same relative quantities of nucleotides, even if they abide the Chargaff's rules, the arithmetic algorithm could bring forth various uncoiled sequences.*

2.2 Matrices and Base Paring contact Patterns

The completary integer dualities given in Table 2 at page 9 bring forth many loops and open unfolded nucleotide sequences[1], for specified residue quantities, e.g. $\mathbf{A}(\mathbf{C})_{n_t-3}$, $(\mathbf{A})_z(\mathbf{C})_{n_t-3z}$, $\mathbf{A}(\mathbf{C})_{n_t-4}\mathbf{AGG}$, $\mathbf{A}(\mathbf{GC})_{\lfloor \frac{n_t}{2} \rfloor -1}$, $\mathbf{AGA}(\mathbf{C})_{n_t-5}$, 5'$\sqcup(\mathbf{CA})_{\frac{n_t-1}{4}}\mathbf{GG}(\mathbf{UG})_{\frac{n_t-1}{4}-1}$3', 5'$(\mathbf{C})_{n_t-7}$ $\mathbf{CA\ CCUG}$3', 5'$\sqcup\mathbf{G\ G\ A}\ (\mathbf{C})_{n_t-7}\mathbf{AG\ G}$3', 5'$\sqcup(\mathbf{CA})_{\frac{n_t-3}{4}}\mathbf{CC}(\mathbf{UG})_{\frac{n_t-3}{4}}$3', 5'$(\mathbf{GU})_{\lfloor \frac{p}{2} \rfloor -i}$ $\mathbf{GG}\ (\mathbf{AC})_{\lfloor \frac{p}{2} \rfloor -(i+1)}\ \mathbf{A}\ (\mathbf{C})_{2l}\ \mathbf{A}\ (\mathbf{CA})_i\ \mathbf{CC}\ (\mathbf{UG})_i$3' for p odd, and $0 \leq i \leq \lfloor \frac{p}{2} \rfloor - 1$, $n_t = 2p + 1 + 2l$. In constrast, if p is even, and $0 \leq i \leq \frac{p}{2} - 1$ the unfolded sequences are: 5'$\sqcup(\mathbf{GU})_{\frac{p}{2}-i-1}\ \mathbf{CC}\ (\mathbf{AC})_{\frac{p}{2}-2-i}\ \mathbf{A}\ (\mathbf{C})_{2l}\mathbf{A}\ (\mathbf{CA})_i\ \mathbf{CC}\ (\mathbf{UG})_{i+1}$3' and 5'$\sqcup(\mathbf{GU})_{\frac{p}{2}-1-i}$ $\mathbf{GG}\ (\mathbf{AC})_{\frac{p}{2}-i-1}\ \mathbf{A}\ (\mathbf{C})_{2l}\mathbf{A}\ (\mathbf{CA})_i\ \mathbf{GG}\ (\mathbf{UG})_i$3', when $n_t = 2p + 1 + 2l$. Worthy of note is the fact that the various unfolded sequences of the last three examples exist whichever be the odd number n_t of residues. Clearly, these uncoiled sequences are independent of the quantity $q_{\mathbf{C}}$ if the number of unit n_t in the arbitrary sequences surpasses the minimal number of units $n_{t\min} = 2p + 3$, i.e. $n_t \geq 2p + 3$.

Alphabetic palindromes are recognized as diagonal lines in the lower left to upper right direction in a self comparative BPPM. An example is 5'$\sqcup\,\mathbf{GGA}(\mathbf{C})_{2l}\mathbf{AGG}$3', that is 5'$\sqcup\mathbf{GGA}(\mathbf{C})_l$ and $(\mathbf{C})_l\mathbf{AGG}$ starting at 2^{nd} and $(l+5)^{th}$ positions, respectively, see Table 1. In spite of the apparently no biological relevance this kind of motif in a self comparison matrix reveals the reversing matches in distinctive regions of the primary or secondary structures of the oligonucleotides.

	nt1	nt2	nt3	nt4	nt5	nt6	nt7	nt8	nt9
	$\sqcup$	G	G	A	C	C	A	G	G
$\sqcup$									
G		G	G					G	G
G		G	G					G	G
A				A			A		
C					C	C			
C					C	C			
A				A			A		
G		G	G					G	G
G		G	G					G	G

Table 1: Alphabetic palindrome $\mathbf{GGACCAGG}$ along the nonmatching diagonal.

The proposed algorithm complies with Maizel et al. (1981) rules in order to analyzed the BPP transitions. Specifically, these authors quoted: "...*Comparison of the sequence with its reversed complement reveals regions of self-complementarity that can be involved in the formation of secondary structure in single-stranded molecules*...".

An oligonucleotide is a palindrome if the reading from left to right and the reading from right to the left on its complementary results unit by unit the same, e.g. 5'$\cdots$ $\mathbf{ACCU}\sqcup\cdots\sqcup\mathbf{AGGU}$ $\cdots$3'. It is a palindrome, since its complementary sequence $\cdots\mathbf{UGGA}\sqcup\ \cdots\ \sqcup\mathbf{UCCA}\cdots$ read from the right to left coincides with the original one (Bachellier, S. et al. , 1999).

[1] A comma keeps apart distinct unfolded sequences. Furthermore, open unfolded sequences starts at 5'$\sqcup$ and ends at 3'.

3 DIRECT APPLICATIONS TO OLIGONUCLEOTIDES

In this section, we prompt the cohesive arithmetic procedure to comprehend certain features that appear in the kinetic intermediates in the biopolymer folding. It is an expeditious process (Fernández, A., and Niel, B., 1997). The folding and wrapping of nucleotide chains at the level of the base-pairing contact as quasi-states of the thermodynamic equilibrium should take place by cooperative effective base-pairing contacts in order to shorten the end to end unfolded chain distance (Fernández, A., 2010). Hence, at each step the folding process the cooperative and associative contacts update the real distance between scattered units, see Fernández, A., and Cendra, H. (1996). The orientation of the residues in long biopolyner sequences involves the existence of differentiated regions. Consequently, distinctive regions entail a necessary accompaniment of the cohesive arithmetics.

Table 2 shows algorithm renderings for the cohesive condition $k_j = \lfloor \frac{n_t}{2} \rfloor$ and $k_i = \lfloor \frac{n_t}{2} \rfloor - 1$, in Eq. (2) at page 3. Herein, a "worthy fact" is the loops independence of the odd number de units n_t in the studied nucleotide subsequence. The current complementary arithmetic exhibits that in the set of nucleotide sequences built by $q_{\mathbf{U}} = q_{\mathbf{C}} = q_{\mathbf{A}} = 3$, and $q_{\mathbf{G}} = 2$, **UCUC** and **GAGA** are antiparallel complementary subsequences. Furthermore, the unfolded sequence **UCUCAAGAGU** with base **C** discarded has a structural motif demanding greater entropy cost than the stacking of both antiparallel subsequences with the synchronous triloop formation in **UCUC(ACU)GAG A**. In this oligonucleotide **UCU CACUGAGA** the algorithm renders the kinetic intermediates: **UCUC(ACUG)AGA** and **UCUCA(CUGA)GA**. Both have feasible tetraloop formation, (between parenthesis). The first one with a stem of three ligands with an isolated base, **U**. The second one folds with one mismatch and two unpaired bases **UC**. It is easily discarded because its enthalpic change does not surpass the entropy cost. A caveat is reasonable in the equilibrium coexistence of **UCUC(ACU)GAGA** and **UCUC(A CUG)AGA**. This requires a fine comparison of the available thermodynamic data in the same environmental of replication conditions.

$n_t \sim$ odd	b-p	motifs	refs.
$\forall\ n_t$	**A U**	W-C	-
$\forall\ n_t$	**G C**	W-C	-
$\forall\ n_t$	**A C U G**	Tetraloop ($L^4_\circ$)	13^{th} Tab. 1 in Sheehy et al. (2010)
$\forall\ n_t$	**A G GG**	Tetraloop ($L^4_\circ$)	Tab. 1& 2 in Sheehy et al. (2010)
$n_t \geq 7$	**U C U C A C**	$L^6_\circ$	pg. 424 in Sheehy et al. (2010)
$n_t \geq 7$	**U G G A C C**	$L^6_\circ$	15^{th} Tab. 3 in Sheehy et al. (2010)
$n_t \geq 7$	**U G A G A G**	$L^6_\circ$	30^{th} Tab. 2 in Sheehy et al. (2010)
$n_t = 11$	**U C U C A A G A G U**	$L^{10}_\circ$	-

Table 2: Arithmetic cohesive motifs for P_5 triloop in sunYL-13 ribozyme, e.g. Fernández, A., Niel, B., and Burastero, T. (1998); Fernández, A., and Cendra, H. (1996).

Accordingly, a BPPM in Table 3 at page 7 illustrates the P_5 triloop region in sunYL-13 ribozyme, e.g. see Fernández, A., Niel, B., and Burastero, T. (1998), Fig. 3 at page 95; Fernández, A., and Cendra, H. (1996), Fig. 1 at pg. 338. In this secondary structure, ten of the fourteen labeled phosphate domains, are cooperative folding events feasible of been roled by the cohesive arithmetics in Table 2. P_3 and P_7 domains are engaged in a three dimensional motif, a pseudoknot, it is not under analysis in the present work.

Moreover, the domain labeled by P_8, 5'...CAACUCUAAGAGUUG...3', in sunYL-13 ribozyme is a quite similar study case to the structural motif P_5.

S	n_1	n_2	n_3	n_4	n_5	n_6	n_7	n_8	n_9	n_{10}	n_{11}
	U	C	U	C	A	C	U	G	A	G	A
U	U		U				U				*
C		C		C		C				*	
U	U		U				U		*		
C		C		C		C		*			
A					A		*		A		A
G						□		G		G	
U	U		U		*		U				
G				*				G		G	
A			*		A				A		A
G		*						G		G	
A	*				A				A		A

Table 3: P_5 triloop in sunYL-13 ribozyme, e.g. see Fernández, A., Niel, B., and Burastero, T. (1998); Fernández, A., and Cendra, H. (1996).

In this two paragraphs, we focus on the subsequence that starts up 13^{th} and ends at 31^{th} positions in the secondary structure of the L-21 Sca I ribozyme in Zarrinkar, P. P., and Williamson, J. R. (1994), Fig. 2 A, at page 919. A comparison of the first half of units of this oligonucleotide with its reversed complementary subsequence reveals a self-complementary region involved in the formation of a hairpin. Its BPPM is shown in Table 4. This matrix array doubled up at the asterisk line makes up the stem and loop zones distinguishable along the main diagonal. Specifically, this motif is unveiled by diagonal lines in the upper left to lower right direction CAGG beginning at position 14^{th} is self-complementary to CCUG beginning at position 24^{th}. There is a 2-fold symmetry on either side of the nonmatching diagonal that runs from the lower left to the upper right corners that is useful in recognizing the regions of self-complementarity. The two subsequences that could bind are identified by projecting vertical lines from the ends of the two diagonal segments to the horizontal (forward) sequence. Then, the loop is built by the nucleotides CAUGCA while its nearest neighbors are G and C that initiated the stacked base-pairing of the zipper (stem zone) of this secondary structure, see Table 4, at page 8.

The assume cohesive arithmetic algorithm prompts a re-scaling of distance amongst units since CAUGCA should be reduced to a coiled motif in any case because of the existence of elemental and/or stable tetraloops[2]: C(AU)(GC)A; C(AUGC)A; CA(UGCA). Furthermore, if the nearest neighbor are added in the analysis the re-scaling of distances is of similar order of magnitude, since: (GC)(AU)(GC)AC; (GC)A(U GCA)C; (G(C(AU)G)C)AC and (G C)A(U(G C)A)C. The same scenery is appliable to the P_{5c} domain, 5'...CCUUGC AAGG.... The algorithm entails a natural re-scaling of distance from the nested inward to outward bp contacts, i.e. 5'... (C(C(U(U(GC)A)A)G)G)... which amongs the feasible conformations it involves the zipping up of the bases.

[2]Here the bases between parenthesis stand for bp contacts.

In addition, Table 2 complementary arithmetic, strains the existence of the tetraloop motif **AGGG** and its four permutations, i.e. **GGGA**, **GAGG** and **GGAG**. Obviously, their respective complementary tetraloops are feasible motifs, e.g. **U**CCC. These quartets are more vitals in the secondary structures studied in the papers Kulinski, T. et al. (2003); De Gregorio, E. et al. (2006); Schnare1, M. N. et al. (1996). Precisely, in De Gregorio, E. et al. (2006), the secondary structure of Yersinia palindromic sequences have distinctive multidomains with hairpins sired by **GGGA** and **U**CCC tetraloops as essential features. Hence, the duality of the cohesive arithmetic enclosed in Table 2 could mimic therein the intramolecular bindings amongst units. Furthermore, the TAR hairpin structure of the HIV-1 and HIV-2 of extensive scientific interest could respond to this cohesive intramolecular attraction (Amann, R. I. et al., 1990; Wieczorech, K. P. et al. , 2006). Aims that we deal with in a subsequent work.

S	n_1	n_2	n_3	n_4	n_5	n_6	n_7	n_8	n_9	n_{10}	n_{11}	n_{12}	n_{13}	n_{14}	n_{15}
	⊔	C	A	G	G	C	A	U	G	C	A	C	C	U	G
⊔															
C		C				C				C		C	C		*
A			A				A				A			*	
G				G	G				G				*		G
G				G	G							*			G
U								U			*			U	
G				G	G					*					G
C		C				C			*	C		C	C		
A			A				A	*			A				
U							*	U						U	
G				G	G	*									G
C		C			*	C				C		C	C		
C		C		*		C				C		C	C		
U			*					U						U	
G		*		G	G										G

Table 4: BPPM of 1^{st} stem-loop in L-21 Sca I ribozyme: Zarrinkar, P. P., and Williamson, J. R. (1994).

In contrast, in Zarrinkar, P. P., and Williamson, J. R. (1994), ref. Fig. 2 at pg. 919, the secondary structure of the L-21 Sca I ribozyme has the hairpin loop labeled by P_{5b}, 5'...**G**[**G**(**GA A A**)**C**]**UU U**G...3', and the eightshaped motif at the end of the chain, both patterns hold the tetraloop first positioned in Table 1 of Sheehy et al. (2010), i.e. **GAAA**, with the nearest neighbor bp or closing base pairing **C**-**G**. In addition, P_{6b} stem loop 5'... **GUCAACAGA**(**UCU U**) **C**U...3' are domains compatible with the cohesive condition given for $k_j = \lfloor \frac{n}{2} \rfloor - 1$, and $k_i = \lfloor \frac{n}{2} \rfloor$ in Eq. (2), see Table 5. A multidomain scenery is compatible with this associative arithmetics. It makes evenly putative the existence of the tetraloops: **U** **CAG**, **GAAA**, **UUU**C, as well as their equivalent permutations. These feasible tetraloop formations and the elemental bp stackings have cooperative effects with the re-scaling of the distance amongst units. The first eightshaped motif (bulge plus a heptaloop) in the secondary structure of the L-21 Sca I ribozyme is a neat example. Precisely, in the oligonucleotide 5'...**AAA**CC**AA**U**AGA**U**UG**C**A**U **C**G**GUUU**... these cooperative events shorten the effective distance to 5'...**AAA**CC**AAGU GUUU**..., refers to the 10^{th} row in Table 5.

$n_t \sim$ odd	b-p	motifs	refs.
$\forall\ n_t$	A U	W-C	-
$\forall\ n_t$	G C	W-C	-
$\forall\ n_t$	ACUG	Tetraloop $(L_\circ^4)$	13^{th} Tab. 1 Sheehy et. al.
$\forall\ n_t$	GAAA	Tetraloop $(L_\circ^4)$	1^{st} Tab. 1 Sheehy et. al.
$\forall\ n_t$	UCUU	Tetraloop $(L_\circ^4)$	-
$n_t \geq 7$	GGACCU	$L_\circ^6$	-
$n_t \geq 7$	GAACUU	$L_\circ^6$	-
$n_t \geq 7$	$(AC)_3C$; $G(UG)_3$	$L_\circ^7$	-
$n_t \geq 7$	5'⊔CCAAGUG	$(U_F)_\sim^7$	-
$n_t \geq 9$	CCAAGUGU	$L_\circ^8$	-
$n_t \geq 9$	U GGAACCU	$L_\circ^8$	-
$n_t \geq 13$	5'⊔$(CU)_3$CCAAGG	$(U_F)_\sim^{12}$	-
$n_t \geq 13$	UGGAACCACAGU	$L_\circ^{12}$	-
$n_t \geq 15$	5'⊔CU GGAACCACAGAC	$(U_F)_\sim^{14}$	-

Table 5: Loops and untangled strings for $k_j = \lfloor \frac{n}{2} \rfloor - 1$, and $k_i = \lfloor \frac{n}{2} \rfloor$ in Eq. (2).

4 CONCLUSIONS

Biopolymer folding is an efficient process but of elusive theoretical underpinnings. Paramount theme in bioenginneering, therefore there are huge widespread scientific and technological information. Our approach roughly reflects certain cooperative events at the level of the base pairing contacts in oligonucleotides. Nevertheless, the environmental conditions for its validation could be differentiated. Consequently, we must struggle further on this aspects. The proposed algorithm could be used for the design of oligonucleotide probes once its overcome tough tests.

REFERENCES

Amann, R. I., Binder, B. J., Olson, R. J., Chisholm, S. W., Devereux, R., and Stahl, D. A. *Combination of 16S rRNA-Targeted Oligonucleotide Probes with Flow Cytometry for Analyzing Mixed Microbial Populations*. American Society for Microbiology, June 1990, Vol. 56, No. 6, p. 1919-1925.

Bachellier, S., Clement, J. M., and Hofnung, M. *Short palindromic repetitive DNA elements in enterobacteria: a survey*. Res. Microbiol. Elsevier, 150 (1999) 627-639.

Cendra, H., Fernández, A., and Reartes, W. *A geometric framework for polymer folding*. Journal of Mathematical Chemistry. Vol. 19, 1996, 331-336.

De Gregorio, E., Silvestro, G., Venditti, R., Carlomagno, M. S., and Di Nocera, P. P. *Structural Organization and Functional Properties of Miniature DNA Insertion Sequences in Yersiniae* . Journal of Bacteriology, November 2006, Vol. 188, No. 22, p. 7876-7884.

Fernández, A. *Transformative Concepts for Drug Design: Target Wrapping*. Springer-Verlag Berlin Heidelberg 2010. ISBN 978-3-642-11791-6.

Fernández, A., and Niel, B. *Folding pathways as brachistochrones.* Proceedings of the Fourth "Dr. Antonio A. R. Monteiro"Congress on Mathematics. Univ. Nac. del Sur, Bahía Blanca, 1997, 179-186.

Fernández, A., Niel, B., and Burastero, T. *The RNA folding problem: a variational problem within an adiabatic approximation*. Biophysical Chemistry. Vol. 74, April 1998, Printed in

U.S.A., 89-98.

Fernández, A., and Cendra, H. *In vitro RNA folding: the principle of sequential minimization of entropy loss at work*. Biophysical Chemistry. Vol. 58, No. 1, April 1996, Printed in U.S.A., 335-339.

Kulinski, T., Olejniczak, M., Huthoff, H., Bielecki, L., Pachulska-Wieczorek, K. , Das, A. T., Berkhout, B., and Adamiak, R. W. *The Apical Loop of the HIV-1 TAR RNA Hairpin Is Stabilized by a Cross-loop Base Pair*. The Journal of the Biological Chemistry. July 25, 2003. Vol. 278, No 40, 38892-39901.

Lu, Z. J., Turner, D. H., and Mathews, D, H. *A set of nearest neighbor parameters for predicting the enthalpy change of RNA secondary structure formation*. Nucleic Acids Research, 2006, Vol. 34, No.17, 4912-4924.

Maizel, J. R., Jr., and Lenk, R. P. *Enhanced graphic matrix analysis of nucleic acid and protein sequences*. Proc. Natl. Acad. Sci USA, Vol. 78, No. 12, 7665-7669, December 1981. Genetics.

Sheehy, J. P., Amber, R. D., and Znosko, B. M. *Thermodynamic characterization of naturally occurring RNA tetraloops*. RNA. January 4, 2010, 417-429. A publication of the RNA Society.

Schnare1, M. N., Damberger, S. H., Gray M. W., and R. Gutell R. R. *Comprehensive Comparison of Structural Characteristics in Eukaryotic Cytoplasmic Large Subunit (23 S-like) Ribosomal RNA*. Journal of Molecular Biology. 1996, 256, 701Ű719.

Zarrinkar, P. P., and Williamson, J. R. *Kinetic Intermediates in RNA Folding*. Science. Vol. 265, 12 August 1994.

Wieczorech, K. P., Purzycka, P. S., and Adamiak, R. W. *New, extended hairpin form of the TAR-2 RNA domain points to the structural polymorphism at the 5' end of the HIV-2 leader RNA.* Nucleic Acids Research. Vol 34, No. 10, pp. 2984-2997, April 2006.

Zhang, W., and Chen, S. J. *Master equation approach to finding the rate-limiting steps in biopolymer folding*. Journal of Chemical Physics. Vol. 118, No. 7, 15 February 2003.