

JELENKA B.  
SAVKOVIĆ-STEVAŃOVIĆ<sup>1</sup>  
DONALD R. FRANCESCETTI<sup>2</sup>  
MAX B. GARZON<sup>2</sup>  
R. DEATON<sup>2</sup>  
J.A. ROSE<sup>2</sup>  
R.C. MURPHY<sup>2</sup>

<sup>1</sup>Department of Chemical  
Engineering, Faculty of  
Technology and Metallurgy,  
Belgrade

<sup>2</sup>The Molecular Computing  
Group, the University of  
Memphis, Memphis, USA

SCIENTIFIC PAPER

516.644+577.1

## THE USE OF ARTIFICIAL NEURAL NETWORKS FOR PARAMETER DETERMINATION IN BIOLOGICAL SYSTEMS

*In this paper an algorithm was developed for DNA state simulation using hydrogen bonds between complementary pairs and stacking interactions between neighboring base pairs. In the aim to investigate the information transfer phenomenon between oligonucleotides, the hybridization reaction was modeled as a noisy channel in which error-free transmission of information occurs between perfect Watson-Crick complements, and errors occur if non-Watson-Crick base pairs are presented. The chemical potential of compounds was derived and a new group contribution method for DNA activity was developed. This method was implemented by using an artificial neural network. Parameters of different kinds of interactions and size differences between the molecules were predicted. An algorithm for DNA information processing, which involved hybridization equilibrium constants, activity coefficients of chain association and association constants, was developed. Melting profiles and local map stability prediction for sequenced DNA were illustrated. The amount of information that can be transmitted without error is bounded by the capacity of the channel. The main result in this paper is a new evolutionary algorithm which includes an artificial neural network model for DNA parameter determination and DNA state processing. The results of the communication depend on the hybridization reaction and reassociation, which also determine the number of molecules that can be interpreted as results without error. The computed results for hybridization energy were in a good agreement with the experimental data of other authors.*

The discovery that genetic information is coded along the length of a polymeric molecule composed of only 4 types of monomeric units is one of the major scientific achievements of this century. The informational content of DNA resides in the sequence in which these monomers, purine and pyrimidine deoxyribonucleotides, are ordered.

Along the length of the molecules, information is encoded in the specific sequences of nucleotides, the chemical building blocks of DNA—deoxyribonucleic acid. Specific sequential arrangements of nucleotides (oligos), these four chemical letters, encode the precise information in a gene. The molecule that conveys biological information from one generation to the next takes the three dimensional form of a double helix. If the entire library of genes stored within the microscopic nucleus of a single human cell were written in letters the size of those you are now reading, the information would fill more than a hundred books as large as this one. Thus the complex structural organization of an organism is specified by an inherited script conveying an enormous amount of coded information.

A particular sequence of nucleotides says the same thing to one organism as it does to another; differences between organisms reflect genetic programs of different nucleotide sequences.

The probability of mismatched hybridization is identified with the probability of error in the transmission of information [1]. When this is incorporated into the information theoretic analysis, the connection is made between the thermodynamics of the reaction, i.e. the free energy and species concentrations and the information transmission through DNA hybridization. The probabilities of different chemical species are related to

their relative occurrence in the reaction bath and quantified as their mole fractions. An initial population is chosen, randomly.

From the Shannon information theory [2], the amount of information that can be transmitted without error is bounded by the capacity of the hybridization channel. The capacity of the hybridization channel is determined by the probabilities of reactions between oligonucleotides, which are related to the change in Gibbs free energy for the reaction. Manipulation of the free energy distribution through the reaction conditions: temperature, reactant concentrations and base sequence may be a good method to control DNA [3–4].

The basic processing of a DNA based computation, as suggested by Adleman [5], is in the massive number of string comparisons that occur during the template matching reactions between DNA oligonucleotides. A small instance of the so called Hamiltonian path problem –HPP was encoded in molecules of DNA and solved in a test tube using standard methods of molecular biology which include hybridization, ligation, melting and annealing, restriction enzymes, polymerization chain reaction –PCR, nucleases, and repair enzymes [6]. Thus, a fundamental step in a DNA computation is the hybridization between oligonucleotides. Other proposals for DNA computation (Boolean logic based on DNA etc.) continue to rely on the mechanism of the template matching hybridization reaction [7–9]. Most assume that the hybridization between oligonucleotides occur error free [10–11].

The presence of massive numbers  $\sim 10^{12}$  of molecules representing each particular edge and vertex of the graph allowed for all possible molecular combinations to form simultaneously within their reaction bath. A typical test tube has huge numbers of molecules  $6.022 \times 10^{23}$  per mole [14].

The oligonucleotides are considered as symbols that are communicated through a noisy channel representing the hybridization. Therefore, the

Author address: J. B. Savković-Stevanović, Department of Chemical Engineering, Faculty of Technology and Metallurgy, Belgrade University, Karnegijeva 4, 11000 Belgrade, e-mail: savkovic@elab.tmf.bg.ac.yu  
Paper received: April 10, 2000  
Paper accepted: June 30, 2000

transmission of sequence information occurs either without an error, i.e. perfect complement base pair matching, or with an error. A complete molecule is a word composed of the oligo symbols. This analysis leads to an information theoretic capacity DNA based computation, and promises error free communication through appropriate coding of the DNA words.

In this paper the possibility of error in the series of hybridizations including DNA activity, stacking interactions and dissociation of complementary strands is examined. Prediction of melting profiles and local stability for sequenced DNA was illustrated. A new algorithm in DNA based computing, which involved chemical equilibrium constants, activity coefficients, helicity, chain association constants of complementary strands and the number of DNA sequences was developed.

The main contribution in this paper is the estimation of the activity coefficients for hydrogen bonds between complementary pairs on opposite strands and stacking interactions between neighboring base pairs. The chemical potential of compounds and functional groups were derived for DNA activity. An artificial neural network was used for activity coefficient prediction.

## DNA HYBRIDIZATION

The chemical reaction in which two single strands of DNA, oligonucleotides, are hydrogen bonded together [12–14] is a hybridization reaction as shown in Eq. (1) and Figure 1.

The heterocyclic bases purine and pyrimidine are the parent molecules of nucleosides and nucleotides. Nucleotides are ubiquitous in living cells, where they perform numerous key functions. Examples include incorporation, as their ribose (RNA) or deoxyribose (DNA) monophosphates, into nucleic acids, energy transduction etc. A nucleoside is composed of a purine or a pyrimidine base to which a sugar (ribose or

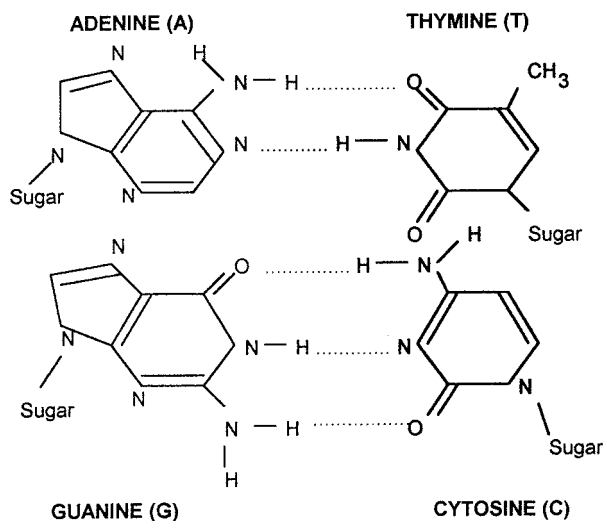


Figure 1. The DNA double strand

deoxyribose) is attached at the nitrogen. Nucleotides are phosphorylated on one or more of the hydroxyl groups of the sugar (ribose or deoxyribose). Figure 1 shows the DNA double strand.

The Watson–Crick model [10,11] assumes one strand of the DNA molecule has an A, the partner strand has a T. And G in one strand is always paired with a C in the complementary strand.

The binary reaction of hybridization is



The hybridization process is modeled as a noisy channel as shown in Figure 2. Figure 2 shows word formation of length  $n$  from input  $X$  and output  $Y$  through the hybridization.

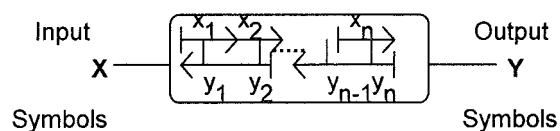


Figure 2. Word formation of length  $n$  from input  $X$  and output  $Y$  through the hybridization

The stoichiometric equation for two arbitrary oligonucleotides,  $x_i$  and  $x_j$  is given by Eq. (1) where  $\langle x_i x_j \rangle$  represents the hybridized oligonucleotides. There are many reactions like this.

The direction of the reaction is determined by the sign of the change in the Gibbs free energy

$$\Delta G = \Delta G^\circ + RT \log Q \quad (2)$$

where  $\Delta G^\circ$  is the free energy change under ideally dilute standard conditions of concentration and pressure,  $R$  is the gas constant, and  $T$  is the temperature.  $Q$  is defined as:

$$Q = \frac{|x_i x_j|}{|x_i||x_j|} \quad (3)$$

where  $||$  indicates mole fractions, and therefore,  $Q$  is dimensionless. The reaction will be driven towards chemical equilibrium, where the rates to the left and right side of the reaction are equal, and  $\Delta G = 0$ , translates to

$$\Delta G^\circ = RT \log K_{eq} \quad (4)$$

where  $K_{eq}$  is the equilibrium constant, which is given by

$$K_{eq} = \frac{|x_i x_j|_{eq}}{|x_i|_{eq} |x_j|_{eq}} \quad (5)$$

The Gibbs free energy is a function of temperature, pressure, and the amount, measured in moles, of each component in the system. It is assumed that the temperature and pressure are constant, and the amounts of the components are allowed to change infinitesimally, then the infinitesimal change in  $G$  is

$$dG = \sum_{i=1}^{NC} (\partial G / \partial n_i)_{T,P,n_{j \neq i}} dn_i \quad (6)$$

where  $n_i$  is the mole number of the  $i$ th of  $NC$  components in the system. The chemical potential of component  $i$  is defined as

$$\mu_i \equiv (\partial G / \partial n_i)_{T, P, n_{j \neq i}} \quad (7)$$

The condition for chemical equilibrium is

$$\sum_i \mu_i dn_i = 0 \quad (8)$$

For a general chemical reaction

$$0 = \sum_i \nu_i A_i \quad (9)$$

where the  $\nu_i$  are the stoichiometric coefficients, which are negative for reactants and positive for products. The change in the amount of species,  $A_i$ , in the reaction is proportional to the  $\nu_i$ ,  $dn_i = \nu_i d\xi$ , where the constant of proportionality is called the extent of the reaction,  $\xi$ . Substitution for  $dn_i$  into Eq. (8) produces another condition on chemical equilibrium,

$$\sum_i \nu_i \mu_i = 0 \quad (10)$$

The change in  $G$  with the extent of the reaction is

$$\frac{dG}{d\xi} = \sum_i \nu_i \mu_i \quad (11)$$

which is zero at equilibrium.

The free energy change can be written as

$$\frac{dG}{d\xi} = \Delta G^\circ + RT \log Q \quad (12)$$

where  $\Delta G^\circ$  is the free energy change under ideally dilute standard conditions, and  $RT \log Q$  is a correction term for non-standard conditions, with  $T$  the temperature and  $R$  the gas constant.

There are three chemical species spaces.  $X$  corresponds to the initial reactant space with mole fractions for oligos of  $x_i$ .  $Y$  is the space after the hybridization reactions, the final oligo reactant and hybridization product space, with mole fractions of the products  $\langle x_i x_j \rangle$ .  $U$  is the joint space where  $x_{ij}$  indicates oligo  $i$  in hybridization with oligo  $j$ . The correction term for the solution of the reacting oligos is

$$Q = \prod_i \alpha_i^{\nu_i} = \frac{x_{ij} \gamma_{ij}}{x_i x_j \gamma_i \gamma_j} \quad (13)$$

where  $\alpha_i$  is the activity of component  $i$ . The activity is defined as

$$\alpha = \exp \left[ \frac{(\mu_i - \mu_i^\circ)}{RT} \right] \quad (14)$$

where  $\mu_i^\circ$  is the chemical potential in the standard state of an ideally dilute solution [15]. Next, an activity coefficient is the ratio of the activity to the mole fraction,

$$\gamma = \alpha / x_i \quad (15)$$

where the mole fraction is  $x_i = (n_i) / (\sum_i n_i)$ . Through Eqs. (12) and (15), the free energy change is related to the concentrations of the reaction components. At equilibrium,  $dG/d\xi = 0$ , and the equilibrium constant is

$$K_{eq} = \prod_i (\alpha_{i,eq})^{\nu_i} = \exp(-\Delta G^\circ / RT) \quad (16)$$

where  $\alpha_{i,eq}$  are the equilibrium activities.

As seen from Eq. (12), the free energy change, and therefore, the directions of the reactions will depend on both the temperature and the concentrations of reactants and products. Exothermic reactions produce heat or the capacity for work. Endothermic reactions require heat or work input to proceed. Therefore, since exothermic reactions take the chemical system to a lower energy state, they will occur spontaneously, while endothermic reactions will not. The sign of  $\Delta G^\circ$  determines whether the reaction is spontaneous or not, and its direction (if  $\Delta G^\circ < 0$  the reaction proceeds to the right and if  $\Delta G^\circ > 0$  the reaction proceeds to the left).

$\Delta G^\circ$  also depends on whether the hybridization is a perfect Watson-Crick complement or not, and has been measured for different combinations of bases. Watson-Crick complement binding produces the largest decrease in free energy. Other bindings, however, also produce free energy decreases. The unbound state is considered as the reference energy of zero. In addition, the free energy change in a hybridization is greatly determined by the base stacking energies [16-17], and therefore is strongly dependent on the sequence of bases in the oligo encodings. Therefore, the variables that will influence the concentrations of the products of the hybridization reactions are reactant concentrations the base stacking sequence, oligo length, and reaction conditions.

## HELIX STABILITY

Under ordinary conditions, native DNA in an aqueous solution takes a double-stranded structure, known as the B-form [17]. This double stranded structure is maintained by two main forces: *hydrogen bonds* between complementary pairs (A||T and G||C pairs) on opposite strands, and *stacking interactions* between neighboring base pairs.

Since random coils have a larger degree of conformational freedom than the ordered double-stranded structure, the ordered structure is disrupted with an increase in temperature. The double-to-single strand transition represents thermal denaturation or melting.

Assume that we are considering a DNA chain with a known nucleotide sequence of  $KC$  base pairs long. The base pairs are numbered from left (5' to right (3') along one of the strands. In the base pair model a configuration of a DNA chain is represented by a vector  $c$ , consisting of  $KC$  elements, each of which takes 1 or 0 depending on the corresponding base pair being in the

bounded (closed) or unbounded (open) state. In the array of KC digits, the helical section is represented by a block of consecutive 1's, and the end coil by consecutive 0's at the end of the array, and on internal loop by consecutive 0's bounded by 1's at either side.

The free energy of the DNA in configuration  $c$  is related to  $Z_c$  by

$$G_c = -RT \ln(Z_c) \quad (17)$$

where  $R$  is the gas constant and  $T$  is the temperature in absolute units.

The fraction of molecules that has accomplished the transition is evaluated by

$$1 - \theta(T) = \frac{1}{1 + \exp(-\Delta G / RT)} \quad (18)$$

The probability of the chain being in configuration  $c$ , is  $Z_c/Z$ , where  $Z = \sum Z_c$ , and the summation is made for all the possible configurations but the all-zero-element configuration which corresponds to the fully denatured state.

The helical content or helicity,  $\theta$  can be calculated as

$$\theta = \sum N_c(1) \frac{Z_c}{Z} = \sum_{k=1}^{KC} p_k / KC \quad (19a)$$

where

$$p_k = \frac{\partial \ln(Z)}{\partial \ln(s_k)} \quad (19b)$$

and  $N_c$  the numbers of 1's in the elements of vectors  $c$ .

Similarly, the average number of helical + coiled sections,  $n_v$  is obtained by

$$n_v = \sum n_{v_c} \frac{Z_c}{Z} \quad (20)$$

The plot of  $1 - p_k$  against the site number  $k$  demonstrates the mean configuration of the molecules. A probability profile calculated within the helix-coil transition region is often called a theoretical denaturation map, since it directly corresponds to denaturation maps obtained by electron microscopic observations [18,19].

The stacking interactions make the dominant contributions to DNA stability. In this stacking model a configuration is represented by a vector with  $KC-1$  digits, of which 1 or 0 correspond to stacked and unstacked states, respectively. Locally unstable regions and frequently opening regions are located on some random sequence DNAs by calculating the probabilities of individual base pair doublets being in an unstacked state.

Positive correlations can be established between frequently opening regions and transcriptional promoters. For example, for 10 kinds of doublets,  $s_k$  takes one of 10 values,  $S_{MN}$  where  $M$  or  $N$  represents

one of four bases arranged in the order  $\begin{matrix} \uparrow N & \bar{N} \\ & \downarrow \\ M & \bar{M} \end{matrix}$ , an

arrow indicates 5' -3' direction and an upper bar indicates a complementary base. A stability parameter  $s_k$  is the equilibrium constant for the stacking model. According to the complementarity,  $S_{AA} = S_{TT}$ ,  $S_{GG} = S_{CC}$ ,  $S_{AG} = S_{CT}$ ,  $S_{GA} = S_{TC}$ ,  $S_{CA} = S_{TG}$ , and  $S_{AC} = S_{GT}$ . For the stacking model the site number  $k$  and loop size  $l$  must be reinterpreted,  $k=1,2,3,\dots$ ,  $KC-1$  is given to the doublet composed of the  $k$ th and  $(k+1)$  th base pairs, and  $l$  is the number of unstacked doublets or the number of unpaired base pairs +1.

At the final step of melting of DNA, the complementary strands are dissociated from each other. The separated single-stranded chains have translational and rotational freedoms different from associated strands.

Let  $\zeta$  be the fraction of total strands in single stranded form. If dissociation -association proceeds under equilibrium,

$$(Z_{\text{ext}}^I)^2 / Z_{\text{ext}}^{II} Z = 2C_0 \zeta^2 / (1 - \zeta) = 1/\beta_K Z \quad (21)$$

where  $Z_{\text{ext}}^I$  and  $Z_{\text{ext}}^{II}$  are the external partition functions for the single and duplex forms, respectively.  $\beta_K$  is the association constant for a chain length of  $KC$ ,  $Z$  the internal partition function for associated forms, and  $C_0$  the total strand concentration. For a single strand the internal partition function is taken to be unity as the reference. Solving Eq. (21) one obtains

$$\zeta = \sqrt{\left[ \left( \frac{1}{4\beta_K C_0 Z} + 1 \right)^2 - 1 \right]} - \frac{1}{4\beta_K C_0 Z} \quad (22)$$

The net fraction of doublets in the stacked state is now

$$\theta_{\text{net}} = (1 - \zeta)\theta \quad (23)$$

where  $\theta$  is the helical content or helicity.

The association constant for a chain length of  $KC$ ,  $\beta_K$  can be defined as

$$\beta_K = K(1 - \zeta) \quad (24)$$

## DNA ACTIVITY PREDICTION

The activity of a component in a reaction bath can be considered to be a correction to the concentration due to interactions and size differences between the molecules in the bath. Information about the different kinds of interactions must be available in some form in order to calculate the activity. A molecule can be considered to consist of different structural-functional groups, which will influence the physical and chemical behavior of a given molecule. Molecules belonging to the same class of compounds retain some common properties. This is the basis of group-contribution methods and it provides an easy way to classify molecules by their structural-functional groups.

Assuming that all information about the interactions between the molecules in the mixture can be related to the distribution of well defined structural-functional groups in the molecule and in the mixture [21], the activity and thereby also the activity coefficient can be calculated as a function of these group distributions and temperature only,

$$\gamma_i = a_i / x_i = M (\phi_{1i}, \phi_{2i}, \dots, \phi_{Ni}, \psi_1, \psi_2, \dots, \omega_N, T) \quad (25)$$

where

$$\phi_{ji} = \omega_{ji} x_{ji} / \sum_{k=1}^{NC} \omega_{jk} x_k \quad (26)$$

and

$$\psi_j = \sum_{k=1}^{NC} \omega_{jk} x_k / \sum_{i=1}^{NG} \sum_{k=1}^{NC} \omega_{jk} x_k \quad (27)$$

$\gamma_i$  is the activity coefficient,  $a_i$  is the activity and  $x_i$  is the mole fraction for component  $i$ ,  $\omega_{ij}$  is the number of groups  $j$  in component  $i$ .  $NC$  is the overall number of components and  $NG$  the overall number of structural-functional groups. The distribution of structural-functional groups which influence the physical and chemical behavior of DNA is defined by Eq. (27), (21-24). A new group contribution method is established. The following groups are considered:  $-\text{CH}_3$ ,  $-\text{CH}_3\text{CO}$ ,  $-\text{NH}_2$ , and  $-\text{PO}_3$ .

The activity coefficient can be calculated as a function of these group distributions and temperature.

$$\gamma_i = f (\phi_{\text{CH}_3}, \phi_{\text{CH}_3\text{CO}}, \phi_{\text{NH}_2}, \phi_{\text{PO}_3}, \psi_{\text{CH}_3}, \psi_{\text{CH}_3\text{CO}}, \psi_{\text{NH}_2}, \psi_{\text{PO}_3}, T) \quad (28)$$

If an analytical expression for this model is known, model parameters can be fitted. In this case there is no analytical expression and a new different approach is used by using an artificial neural net [27,28]. The neural net can be operated in two ways. Training and running. The training objective consists in modeling the functionalities between the input and output. Meanwhile, the running phase computes new activity coefficients.

## THE MUTUAL INFORMATION PROCESSING

The genetic information stored in the nucleotide sequence of DNA serves as a source of information for the synthesis of all protein molecules of the cell and organism, transcription, and it provides the information inherited by daughter cells, replication. The complementarity of the Watson and Crick double stranded model of DNA strongly suggests that replication of the DNA molecule occurs in a semiconservative manner.

Application of the Shannon information theory in DNA based computing is based on the necessity for the error free production of DNA molecules. Each DNA oligonucleotide  $|O_i\rangle$  (where  $|$  denotes the 5' and  $\rangle$  denotes the 3' ends of the DNA oligonucleotide,

respectively) of length  $n$ , or  $n$ -mer, is composed of two  $n/2$  -mers,  $p_i$  and  $q_i$ ,  $|O_i\rangle = |p_i q_i\rangle$ . In Adleman's original work (5), a double stranded DNA molecule is formed by  $p_i$ 's hybridizing to their Watson-Crick complements,  $\bar{p}_i$ , and  $q_i$ 's hybridizing with their Watson-Crick complements  $\bar{q}_i$ . Therefore, the oligos are strung together by  $O_i$ 's successively overlapping in either their  $p_i$ 's or  $q_i$ 's. An error is defined, for instance, as a hybridization of a  $p_i$  with a  $\bar{p}_i$ , where it is not the Watson-Crick complement  $\bar{p}_i$  of  $p_i$ . The same holds for the  $q_i$ 's. Many other hybridization products are possible that do not fit the model.

Define an alphabet of input symbols,  $X$  that contains all the oligos in our reaction bath. The basic symbol in the alphabet is an oligonucleotide together with its complement,  $|x, \bar{x}\rangle$ . For Adleman's application, the members of  $X$  would be the  $p_i$ 's and their complements  $\bar{p}_i$  and  $q_i$ 's and their complements  $\bar{q}_i$  (Figure 2). With no errors, an input/output symbol pair is a given oligo  $x_i$  and its complement  $\bar{x}_i$ . The set of words  $W$ , composed of symbols from  $X$  consists of all possible double stranded molecules that can be formed by hybridization of the elements of  $X$ . The probability distributions of the input and output symbols sets are  $p(x) = p(y)$ . The uncertainty in the input and output symbol sets is given by Shannon entropy

$$S(x) = -\langle \log p(x) \rangle = -\sum_{x \in X} p(x) \log p(x) \quad (29)$$

A complete graph can be chosen for the analysis. The completeness of the graph ensures that the consecutive input symbols chosen for hybridization are independent. Every possible combination for  $\bar{x}$  vertex encodings  $|x_i, \bar{x}_i\rangle$ , and edge encodings  $\langle \bar{x}_i, x_i |$ , is present and equally likely.

The probability that the chosen input symbol  $x$ , hybridizes with an output symbol  $y$ , is given by the conditional probability  $p(x|y)$ . The uncertainty in  $x$  given  $y$  is the conditional entropy, given by

$$S(x|y) = -\langle \log (p(x|y)) \rangle. \quad (30)$$

The amount of uncertainty in the output symbol  $y$ , or hybridization target of  $x$ , is  $S(y)$ , and the amount of uncertainty removed about  $y$  given knowledge of  $x$  is the mutual information between  $x$  and  $y$ ,  $I(x,y) = S(x) - S(x|y)$ . The joint entropy  $S(x,y)$ , is

$$S(x,y) = -\langle \log p(x,y) \rangle = -\sum_{x \in X} \sum_{y \in Y} p(x,y) \log p(x,y) \quad (31)$$

where  $p(x,y)$  is the joint probability distribution. The joint entropy is related to the conditional entropy by  $S(y,x) = S(y) + S(x|y)$ , which makes the mutual information

$$I(x,y) = \sum_{x \in X} \sum_{y \in Y} p(x,y) \log \frac{p(x,y)}{p(x)p(y)} = S(x) + S(y) - S(x,y). \quad (32)$$

In what follows,  $p(x)$  and  $p(y)$  are related to the mole fractions of the initial reactants supply, and final reaction products, respectively. The joint probability

$p(x,y)$  is identified with the different chemical environments, i.e. hybridized or unhybridized, in which the symbol oligos can occur. According to the well known results of Shannon (2) *the channel capacity (CC) of the hybridization is the maximum  $I(x,y)$  over all possible input distributions  $p(x)$ , and the number of sequences  $n_V$  long that can be distinguished with arbitrarily small probability of error is given by  $2^{n_V(CC)}$ .*

Interpreted in terms of DNA based computation, the above result says that the number of molecules, which consist of  $n$  oligonucleotides, that can be distinguished, is limited by the channel capacity of hybridization, which in turn is determined by the probability of hybridization among the set of oligos. Therefore, if  $p(x,y)$  or the probability of hybridization is known, then, the capacity of DNA based computation is known. In addition, another result of the information theory states that there exist codes that come arbitrarily close to *the limit of  $2^{n_V(CC)}$  sequences that can be transmitted without error.* Typically, this is done by adding controlled redundancy to the words for transmission.

Initial concentrations of reactants are supplied to the reactor at elevated temperature to prevent premature hybridization. The temperature is lowered and the reactions allowed to reach equilibrium, after which reaction products are removed and ligated for reading. In a real system, the bath of reactant oligos would correspond to an external supply of reactants, and the reaction products would be amplified and extracted by any of several techniques, including PCR and magnetic beads with target sequences attached. The bead extraction technique is part of Adleman's original technique (5), and has been proposed as part of a DNA based associative memory [29]. Both the beads and PCR use hybridization, and therefore, the analysis should also apply to them.

In a DNA computer, there are  $N$  binary reactions like Eq.(1). The probability of the reaction species is taken to be their relative occurrence before and after hybridization is allowed to take place. Volume concentrations, mole fractions, or activities could all be quantifiers of a chemical species relative occurrence. The probabilities are taken to be the mole fractions of the different species under consideration. An advantage of using the mole fraction is that they sum to one, just as a probability.

The Shannon entropy, conceptually, is related to the partitioning of the underlying probability space, with entropy increasing for finer and more uniform partitions. There are three spaces that are represented by  $S(x)$ ,  $S(y)$ , and  $S(x,y)$ , Eq.(31), corresponding to the initial oligo reactant space  $X$ , the final oligo reactant and hybridization product space  $Y$ , and the joint space  $U$ , between  $X$  and  $Y$ , which represents the different chemical environments in which each symbol oligo can occur. Let us denote the mole fractions of all species,

hybridized and unhybridized, in each of these spaces as a subscript lower case letter,  $x_i$  for space  $X$ ,  $y_i$  for space  $Y$ , and  $u_i$  for space  $U$ .

In terms of the mole fractions, the mutual information becomes

$$I(x,y) = -\sum_{x \in X} x \log x - \sum_{y \in Y} y \log y + \sum_{u \in U} u \log u \quad (33)$$

where the sums are taken over the respective spaces. Using the expression for the concentration factor, Eq.(13) in terms of mole fractions (which at equilibrium is the equilibrium constant) the mole fraction of products can be expressed as

$$x_{kl} = Q_{Xk} x_k = K_{kl} x_k x_l \gamma_k \gamma_l / \gamma_{kl} = K_{kl} (1/K_Y) x_k x_l \quad (34)$$

The mutual information is

$$\begin{aligned} I(x,y) = & -\sum_i^{MO} x_i \gamma_{x_i} \log x_i \gamma_{x_i} - \sum_i^{MO} y_i \gamma_{y_i} \log y_i \gamma_{y_i} + \\ & + \sum_k^{MO} \sum_{l \geq k}^{MO} K_{kl} (1/K_Y) y_k y_l \log K_{kl} (1/K_Y) y_k y_l \\ & - \sum_l^{MO} \frac{n^Y}{n^X} y_l \gamma_{y_l} \log \frac{n^Y}{n^X} y_l \gamma_{y_l} + \\ & + \sum_k^{MO} \sum_{l \geq k}^{MO} \frac{n^Y}{n^X} K_{kl} (1/K_Y) y_k y_l \log \frac{n^Y}{n^X} K_{kl} K_Y y_k y_l \end{aligned} \quad (35)$$

where MO is the number of initial symbol oligos. The total number of moles in space  $U$ , is  $n^U$ , the total number of moles in space  $X$  is  $n^X$ . The total number of moles in space  $Y$  is given by  $n^Y$ . The correction term,  $n^Y/n^X$  arises in order to express the mole fractions in space  $U$  in terms of the mole fractions of space  $Y$ . Equation (35) relates the information transfer among a bath of hybridizing oligonucleotides to their reaction thermodynamics through the concentration factors, and to the mole fractions of the chemical species corrected with activity coefficients. Information transfer will be affected by several factors. Uniform initial mole fractions will maximize  $S(x)$ , and thus  $I(x,y)$ . The free energy of formation for the hybridization will be determined by base pair sequence and will influence the information through its effect on the equilibrium constant.

## MELTING PROFILES AND STABILITY MAP CHARACTERIZATION

The double stranded structure of DNA is melted in solution by increasing the temperature or decreasing the salt concentration. The melting temperature  $T_m$  is influenced by the base composition of the DNA and by the salt concentration of the solution. DNA rich in G||C pairs, which have three hydrogen bonds, melts at a higher temperature than that rich in A||T pairs, which have two hydrogen bonds. Almost simultaneously, both the measuring system and the theory of DNA melting have undergone their maximal development [3, 4, 30–34]. Thus it has become possible to compare calculated with

measured profiles for known sequenced DNA species or restriction fragments. The stability  $\Delta G^\circ$  and the melting behavior  $\Delta H^\circ$  of any DNA duplex structure can be predicted from inspection of its primary sequence. The overall stability and the melting behavior of any DNA duplex structure can be predicted from the relative stability  $\Delta G^\circ$  and the temperature-dependent behavior ( $\Delta G^\circ$ ,  $\Delta C_p^\circ$  of each DNA nearest-neighbor [31,32]. The stability of each nearest-neighbor interaction at any temperature of interest can be calculated as

$$\Delta G^\circ = \Delta H^\circ - T\Delta S^\circ \quad (36)$$

The units of  $\Delta G^\circ$  are J/mole of interaction. Transition enthalpies of DNA oligomers can be predicted by the following equation

$$\Delta H_{\text{total}} = \Delta h_i + \sum_x \Delta h_x \quad (37)$$

assuming that the helix initiation enthalpy  $\Delta h_i$  equals zero.

Transition free energies of DNA oligomers can be calculated for any DNA duplex from its primary sequence

$$\Delta G_{\text{total}} = -(\Delta g_i + \Delta g_{\text{sym}}) + \sum_x \Delta g_x \quad (38)$$

where the symmetry term  $\Delta g_{\text{sym}}^*$  accounts for the entropic difference between a duplex formed from self-complementary sequences and  $\Delta g_i^{**}$  is the helix initiation free energy [4, 35, 36].

This capability should prove valuable in numerous applications, such as predicting the stability of a probe-gene complex, selecting optimal conditions for a hybridization experiment, deciding on the minimum length of a probe, predicting the influence of a specific transversion or transition on the stability of an affected DNA region, and predicting the relative stabilities of local domains within a DNA duplex. For selecting the DNA oligomer basis set, the following criteria were considered: the oligomers should be long enough so as to exhibit experimentally convenient melting temperatures; the oligomers should have G||C base pairs at the ends to minimize fraying of terminal A||T base pairs; the oligomers should exhibit CD spectra characteristic of the B conformation; the oligomers should exhibit "all or none" /two state/ melting behavior (in other words, the calorimetric and the van't Hoff transition enthalpies must be equal). The physical meaning of the parameters provides additional information about their dependence on sequence, loop size, or chain length.

The partition function  $Z_c$  for configuration  $c$  is obtained by a product of stability parameters  $s_k$ 's and

\*  $\Delta g_{\text{sym}}$  equals 1.68 KJ for a duplex formed from a self-complementary sequence whereas for a duplex formed from two complementary sequences,  $\Delta g_{\text{sym}}$  equals 0.

\*\*  $\Delta g_i$  assigns 20.92 KJ for duplexes containing G||C base pairs and 25.10 KJ for duplexes composed exclusively of A||T base pairs.

*loop weighting factors*  $\sigma_i$ 's assigned on the following rule: Assign  $s_k$  to the  $k$ th base pair if it is in the bounded state. Because A||T and G||C pairs differ in stability,  $s_k$  takes one of two values,  $S_A = S_T$  and  $S_G = S_C$ . Assign  $\sigma_i$  to an internal loop including  $l$  unbounded base pairs. And assign unity to an end coil as the reference of statistical weights.

The *stability parameter*  $s_k$  is the equilibrium for stacking of the  $(k+1)$ th base pair on the  $k$ th pair which is already stacked on the  $(k-1)$ th pair. When the  $k$ th and the  $(k+1)$ th bases on the sense strand are  $M$  and  $N$ , respectively,  $s_k$  is written in the form

$$s_k = s_{MN} = \exp [-(\Delta H_{MN} - T\Delta S_{MN})/RT]. \quad (39)$$

where the enthalpy  $\Delta H_{MN}$  and entropy  $\Delta S_{MN}$  are assumed to be constant. Hence, it may be reasonable to assume  $\Delta S_{MN} = \Delta S$  constant for all the doublets [37]. Then,  $T_{MN} = \Delta H_{MN}/\Delta S$  is the key parameter that

represents the measure of stability of the  $\begin{matrix} \uparrow N & \bar{N} \\ & \downarrow \\ M & M \end{matrix}$  doublet.

The *loop weighting factor* is usually to be considered of entropic origin,

$$\sigma_l = \sigma_0 \lambda_l \epsilon_l \quad (40)$$

where the loop entropy is defined as  $R \ln(\lambda_l)$ , the difference in electrostatic potential energies between the same sized internal loop and an end coil is  $-RT \ln(\epsilon_l)$ , and  $\sigma_0$  contains all the factors that are independent of  $l$  [38]. Since an internal loop has a smaller degree of conformational freedom and higher electrostatic potential energy than an end coil of the same size,  $\sigma_l$  is less than unity.

If a melted region on a DNA molecule is regarded as a Gaussian chain, the internal freedom of a circular chain is  $(2l)^{1.5}$  times smaller than that of a linear chain of the same size,  $\sigma_l$  of the form  $\sigma_l = \sigma_0 l^{-1.5}$ .

Consequently, in evaluating and proposing possible biological roles for specific sequences in naturally occurring DNAs, it would be extremely useful if one could predict the formation of particular secondary structural features along the polymer chain simply by inspecting the primary base sequence.

## THE METHOD FOR MELTING PROFILE AND STABILITY MAP CALCULATION

An efficient method for melting profile and stability map calculation on the basic model without any approximation was developed by Poland et.al. [39,40]. The probability of the  $k$ th doublet being in the stacked state,  $p_k$ , is obtained without first evaluating the partition function  $Z$  and its derivatives,  $\partial Z/\partial s_k$  Eq.(19).  $p_k$  is obtained through two series of recursion procedures using conditional probability  $P_k$ 's as temporary variables.  $P_k$  means the probability that the  $(k+1)$ th doublet is in the stacked state if it is known that the  $k$ th doublet is in the stacked state.

In Poland's method, the number of arithmetic steps needed for the recursion procedure amounts to the order of  $(KC)^2$ , since the probabilities for  $(KC)-k$  possible loops of different size must be evaluated in each step of recursion. This number of steps can be reduced to the order of  $(KC)^* l$ , provided that the loop weighted factor is appropriately approximated by a finite series of exponential functions.

Two quantities of interest, the helicity  $\theta$  and the segment number  $n_v$  are readily obtained if  $p_k$  and  $P_k$  are once evaluated by the recursion procedures.

$$\theta = \sum_1^{KC-1} p_k / [(KC) - 1] \quad (41)$$

$$n_v = 1 + p_{k-1} - p_1 + 2 \sum_1^{KC-2} p_k (1 - P_k) \quad (42)$$

The internal partition function  $Z$  is given by

$$Z = \left( p_1 \prod_1^{KC-2} P_k \right)^{-1} \quad (43)$$

In this method the probability profile is obtainable without any additional calculation steps.

#### THE METHOD FOR DNA ACTIVITY COEFFICIENT DETERMINATION

A neural net consists of a number of local information units, called neurons. This work will use the propagation algorithm for training a feedforward net. All neurons in a given layer will be connected to all neurons in the next layer, but no connections between neurons on the same layer will be allowed. The neuron will receive a number of inputs. The inputs identification for activity coefficient estimation was given by Eq.(28). These will be amplified or damped and the overall sum will be transformed by an activation function [27,28].

The basic model of a neuron is

$$y_i = f\left(\sum w_i u_i\right) \quad (44)$$

where  $y_i$  is the output from neuron  $j$ ,  $y_i$  is a function  $f$  of the added inputs  $u_i$  from neuron  $i$  on the previous layer multiplied by the weight  $w_i$ . A sigmoidal function is used as the activation function

$$f(u) = 1 / [1 + \exp(-u)]. \quad (45)$$

The network performance is a function of the topology, that is the number of neurons and the number of layers and also of the parameters/weights. The network can be operated in two ways, training and running. In the training process known values of both input and output are presented to the net. Based on the error in the calculated output, corrections to the weights are made. The training phase must allow the network to model the computational function for the activity coefficient

$$\gamma = f(\varphi_1, \varphi_2, \varphi_3, \varphi_4, \psi_1, \psi_2, \psi_3, \psi_4, T) \quad (46)$$

When the performance is satisfactory, the network can be used to calculate new output values based on new input values. In the running phase the neural network is set up to generate.

Using structural groups the activity coefficient function is given by Eq.(46). The artificial neuron network is set up to generate the function  $f$  with 9 neurons in the input layer, 1 neuron in the output layer using 10 parameters/weights. Several network topologies have been evaluated. A net with one hidden layer with five neurons ANN(9,5,1) using 54 parameters/weights shows the least error during training. The performance of the neural network for calculating activity coefficients has been tested an several examples. The sum of squared errors and the standard deviation on the error random sequence DNA are reported in Table 1.

Table 1. The sum of squared errors and the standard deviation on the error for calculating activity coefficients

Example	Model	Sum of squares error	Standard deviation
Training data	ANN(9,1)	17.7181	0.4217
	ANN(9,5,1)	2.6078	0.1736

#### THE ALGORITHM

The algorithm for DNA parameter determination and information processing is shown in Figure 3. A

*Step 1. Begin;*  
*Step 2. Input data;*  
*Step 3. Assign uniform initial oligos reactant mole fraction  $x_i$ , RND-Gaussian random number;*  
*Step 4. Calculation of equilibrium constant  $K$ , Eq.(5);*  
*Step 5. Free energy calculation under standard conditions  $\Delta G^0$ , Eq. (4);*  
*Step 6. Structural-functional group distribution parameter calculation  $\phi_{ij}, \psi_j$ , Eqs.(26) and (27);*  
*Step 7. Activity coefficient calculation  $\gamma_i$ , Eqs. (28), (45) and (46);*  
*Step 8. Activity calculation  $a_i$ , Eq.(15);*  
*Step 9. Calculation of free energy changes  $\Delta G$ , Eqs. (12) and (13);*  
*Step 10. Mutual information calculation  $I(x,y)$ , Eq.(35);*  
*Step 11. Calculation  $1-\theta$  and  $\theta$ , Eq. (18) and (19);*  
*Step 12. Calculation of segment number  $n_v$ , Eq.(42);*  
*Step 13. Dissociation kinetic constant calculation  $\beta_k$ , Eq.(24);*  
*Step 14. Calculation of single strands fraction  $\zeta$ , Eq.(22);*  
*Step 15. Calculation of information capacity storage;*  
*Step 16. New iteration,  $ii = ii+1$ ;*  
*Step 17. Calculation of the new corrected values  $K_{ii+1}$ ;*  
*Step 18. Repeat iteration from Step 3. to Step 17, while  $(K)_{ii+1} - (K)_{ii} < \zeta$  ( $\zeta$  is the convergence criteria);*  
*Step 19. Data output*  
*Step 20. Stop*

Figure 3. DNA algorithm for parameter determination and information processing

Gaussian random number generator was used for initial oligos in their reaction bath.

The programs were written in several subroutines which were devoted to manipulation of a nucleotide sequence encoded, into a compact format and tested on a PC computer. The output results and parameter values were generated in separate programs units and transferred through files. Consequently, a meaningful evaluation of the sequence dependent DNA structural preferences requires a DNA data base.

## RESULTS

The obtained results show good fit for the presented model including hydrogen bonds and stacking interactions. A new relation of doublet association and reassociation effects to the hybridization reaction was developed.

The new algorithm for DNA information processing which involves the equilibrium constant, activity coefficient, helicity and association constant is shown in Figure 3. The computer simulated results are shown in Figures 4, 5 and 6.

Figure 4 shows the equilibrium constant vs. temperature. The helicity and melting profiles are shown in Figure 5. Figure 5 shows the fraction of DNA strands in the stacked state vs. temperature. Association / reassociation is shown in Figure 6. Figure 6 shows the net fraction of doublets in the stacked state as a function of the association constant. The amount of information transmitted without error was in the range  $0.1E09-0.1E12$ . The obtained results for hybridization energy were in good agreement with the experimental

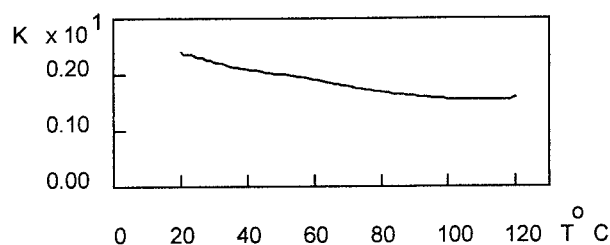


Figure 4. Hybridization reaction: Hybridization equilibrium constant  $K$  vs. temperature  $T$

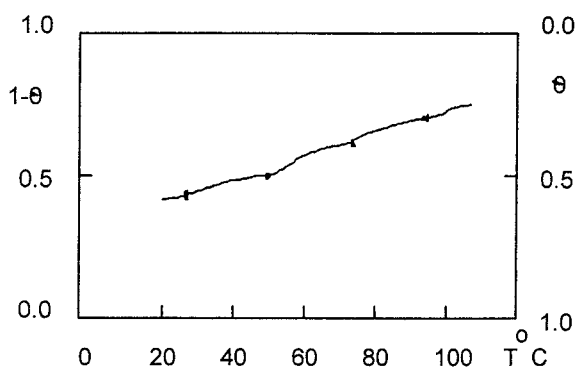


Figure 5. Helicity and melting profile: Helicity  $\theta$  and  $1-\theta$  as a function of reaction temperature  $T$

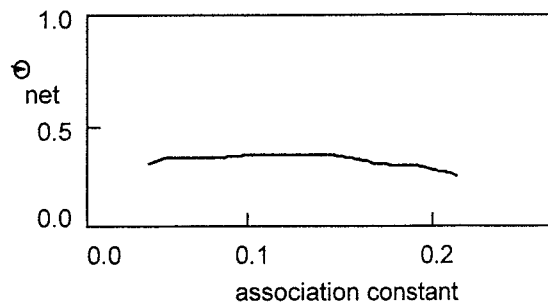


Figure 6. Association and reassociation: The net fraction of doublets in the stacked state  $\theta_{net}$  vs. association constant  $\beta_k$

data of other authors [4]. The results in Table 1 show that with no analytical expression for the activity coefficient, an artificial neural network can be successfully used to predict activity coefficient.

## CONCLUSIONS

The obtained relations give the effects of doublet reassociation and association to the hybridization reaction between complementary pairs. Hydrogen bonds between complementary pairs and stacking interactions between neighboring base pairs in DNA computation were successfully modeled. The information of different kinds of interactions and size differences between molecules in the strand were predicted by implementation of an artificial neural network. A group contribution method was developed for DNA activity predictions.

A new algorithm for information processing in DNA based computation which involved oligonucleotide hybridization, activity and stacking interactions was developed and tested.

The obtained results report complete thermodynamic characterization of all the nearest neighborhood interactions possible in the Watson-Crick DNA duplex structure. They demonstrated how these data could be used to predict the stability and melting behavior of any DNA duplex stability from the knowledge of its primary sequence.

DNA based computation represents real programmable matter, in which the biomolecules can be programmed to communicate useful information by appropriate manipulation of the thermodynamics, kinetics and chemistry. The results of the communication depend on the reaction, which also determines the ultimate capacity of a DNA based computer, or the number of molecules that can be interpreted as results without error.

## NOTATION

- CC – channel capacity
- $C_0$  – total strand concentration,  $\text{mol}/\text{m}^3$
- $G_0$  – free energy, J
- $G$  – free energy at standard condition, J
- $H_0$  – enthalpy, J
- $H$  – enthalpy at standard condition, J

- $I$  – information capacity, byte  
 $K$  – hybridization reaction constant  
 $KC$  – overall number of chains  
 $l$  – number of unstacked doublets  
 $MO$  – number of initial symbols of oligos  
 $n_i$  – number of moles  
 $n_v$  – average number of sections  
 $NC$  – overall number of component  
 $NG$  – number of functional groups  
 $P$  – pressure, 101325 Pa  
 $p$  – probability distribution ( $p(x)$ ,  $p(y)$  relate to the mole fractions of initial oligos)  
 $p_k$  – chain probability  
 $R$  – gas constant,  $J\ mol^{-1}\ K^{-1}$   
 $S$  – entropy,  $JK^{-1}$   
 $s_k$  – stand stability  
 $T$  – temperature, K  
 $U$  – join space  
 $u$  – fraction of species in space  $U$   
 $u'$  – neuron inputs  
 $w$  – weighting factor in neural network  
 $x$  – mole fraction, mol/mol  
 $X$  – input  
 $Y$  – output  
 $Z_{ext}^I$  – partition function for single strand  
 $Z_{ext}^{II}$  – partition function for duplex form  
 $Z_c$  – partition function in configuration c

#### Greek symbols

- $a$  – activity  
 $\beta_k$  – association constant  
 $\gamma$  – activity coefficient  
 $\mu$  – chemical potential  
 $1-\zeta$  – fraction of single stranded form  
 $\xi$  – convergence criteria  
 $\Phi$  – fractional parameter defined by Eq. (26)  
 $\omega$  – number of groups  $j$  in component  $i$   
 $\psi$  – functional parameter defined by Eq. (27)  
 $\sigma$  – loop weighting factor  
 $\theta$  – helicity

#### REFERENCES

- [1] H. Wernthges, G. Steger, D. Riesner, and H.J. Fritz, *Nucleic Acids Res.*, **14** (1986) 3773.
- [2] C.E. Shannon, *Bell Sys. Tech. Journal*, **27** (1948) 379.
- [3] R. Deaton, M. Garzon, R.C. Murphy, J.A. Rose, D.R. Franceschetti, and S.E. Stevens, Jr. *Physical Review Letters*, **80** (1998) 417.
- [4] K.J., Breslau, R. Frank, H. Blocker, and L. A. Marky, *Proc. Natl. Acad.Sci.*, **83** (1986) 3746.
- [5] L.M. Adleman, *Science*, **266** (1994) 1021.
- [6] L. M. Adleman, *DIMACS Series in Discrete Mathematics and Theoretical Computer Science*, **27** (1996) 1.
- [7] R.J. Lipton, *Science*, **268** (1995) 542.
- [8] DIMACS Proceedings of the First Annual Meeting on DNA Based Computers, Princeton University, 1995, DIMACS Proc. Series, AMS, Providence, RI, **27**, (1996).
- [9] DIMACS Proceedings of the Second Annual Meeting on DNA Based Computers, Princeton University, 1996, DIMACS Proc. Series, AMS, Providence, RI, **28** (1997)
- [10] F.H.C. Crick, *J. Mol. Biol.*, **19** (1966) 548.
- [11] J.D. Watson, N.H. Hopkins, J. W. Roberts, J.A. Steitz, and A. M. Weiner, *Molecular Biology Gene*, Menlo Park, CA, The Benjamin/Cummings Publishing Co., Inc. Fourt Ed. (1987)
- [12] D.A. Zarlring, E.P. Sena, and C.J. Green, *Process for nucleic acid hybridization and amplification*, US Patent # 5,223,414.
- [13] R. Rapley, *Mol. Biotechnology*, **2**, 295 (1994)12.
- [14] R. Deaton, R.C. Murphy, M. Garzon, D.R. Franceschetti, and S.E. Stevens Jr., *DIMACS Proceedings of the Second Annual Meeting on DNA Based Computers*, pp.159, (1996), Princeton University, DIMACS Proc. Series, AMS, Providence, RI, **28** (1997).
- [15] I.N. Levine, *Physical Chemistry*, New York, McGraw-Hill Book Com., Inc., (1995)
- [16] D.M. Crothers and B.H. Zimm, *J.Mol.Biol.*, **9** (1964) 1.
- [17] O. Gotoh, *Adv. Biophys.*, **16** (1983) 1.
- [18] A. Wada, S. Yabuki, and Y. Husimi, *CRC Crit. Rev. Biochem.*, **9** (1980) 87.
- [19] A.S. Borovik, Yu. A. Kalambet, Yu.L. Lyubchenko, V.T. Shitov, and Eu.I. Golovanov, *Nucl. Acids Res.*, **8** (1980) 4165.
- [20] J. Savković-Stevanović, M. Mišić-Vuković, G. Boncic-Caricic, B. Trisovic and S. Jezdic, *Sep.Sci. Techn.*, **27** (1992) 613.
- [21] P.A. Sharp, *Cell*, **42** (1985) 397.
- [22] J.R. Nevins, *Annu.Rev.Biochem.*, **52** (1983) 441.
- [23] B. Ruskin et.al., *Cell*, **38** (1984) 317.
- [24] J. Cordin et.al., *Science*, **209** (1994) 1406.
- [25] J. Darnell et.al. *Molecular Cell Biology*, Scientific American Books (1986).
- [26] D.J. Weatheral, *The New Genetics and Clinical Practice*, 2nd ed. Oxford Univ.Press (1986)
- [27] R. Peterson, A. Fredeslund and P. Rasmussen, *Computers chem. Engng.*, **18** (1994) 63.
- [28] J. Savković-Stevanović, *Computers Chem. Engng.*, **18** (1994) 1145.
- [29] E.B. Baum, *Science*, **268** (1995) 583.
- [30] A. Wada, S. Yabuki, and Y. Husimi, *CRC Crit. Rev. Biochem.*, **9** (1980) 87.
- [31] L.A. Marky and K.J. Breslau, *Biopolymers*, **21** (1982) 2185.
- [32] L.A. Marky, B. Buono, and K.J. Breslau, *Biophys. J.*, **45** (1984) 117a .
- [33] J.W. Nelson, F.H. Martin, and I. Jr. Tinoco, *Biopolymers*, **20** (1981) 2509.
- [34] S.M. Frerier, M. Petersheim, D.R. Hickey, and D.H. Turner, *J. Biomol. Struct. Dyn.* **1** (1984) 115.
- [35] O.C. Uhlenbeck, P.N. Borer B. Dengler and I. Tinoco Jr., *J. Mol. Biol.*, **73** (1973) 483.
- [36] H. Klump and T. Ackermann, *Biopolymers*, **10** (1971) 513.
- [37] O. Gotoh, A. Wada, and S.Yabuki, *Biopolymers*, **18** (1979) 805.
- [38] R.M. Pohl, *Eur. J. Biochem.*, **42** (1974) 495.
- [39] D. Poland, *Biopolymers*, **59** (1974) 1318.
- [40] M. Fixman and J.J. Freire, *Biopolymers*, **16** (1977) 2693.

## IZVOD

### KORIŠĆENJE NEURONSKIH MREŽA ZA ODREĐIVANJE PARAMETARA U BIOLOŠKIM SISTEMIMA

(Naučni rad)

Jelenka B. Savković-Stevanović<sup>1</sup>, Donald R. Franceschetti<sup>2</sup>, Max B. Garzon<sup>2</sup>, R. Deaton<sup>2</sup>, J.A. Rose<sup>2</sup>, R.C. Murphy<sup>2</sup>

<sup>1</sup>Tehnološko-metalurški fakultet, Beograd, Jugoslavija

<sup>2</sup>Univerzitet u Memfisu, SAD

U radu je razvijen algoritam za simuliranje stanja DNK-a koristeći vodonične mostove između komplementarnih parova i interakcija između susjednih parova osnovnih blokova DNK-a. U cilju istraživanja fenomena prenosa informacija između oligonukleotida modelovana je reakcija hibridizacije kao kanal sa šumom u kome se transmisija informacija između idealnih Votson-Krikovih kompleksa dešava bez greške, a kada to nisu Votson-Krikovi kompleksi greške se dešavaju. Izveden je hemijski potencijal komponenata i razvijen novi metod doprinosi funkcionalnih grupa za definisanje aktivnosti DNK-a. Ovaj metod implementiran je u algoritam pomoću jedne neuronske mreže. Izračunavani su parametri o različitim vrstama interakcija i različitim dimenzijama između molekula. Razvijen je algoritam za obradu informacija DNK-a koji obuhvata ravnotežne konstante hibridizacije, koeficijente aktivnosti asosovanih lanaca i konstante asocijacije. Ilustrovani su profil stapanja i lokalna stabilnost sekvencirane DNK-a. Količina informacija koja može biti transmitovana bez greške ograničena je kapacitetom posmatranog kanala. Glavni rezultat ovog rada je novi evolutivni algoritam koji uključuje model neuronske mreže za određivanje DNK-a parametara i obradu informacije o stanju DNK-a. Rezultati komunikacije zavise od reakcije hibridizacije i reasocijacije koje takođe određuju broj molekula koji mogu biti interpretirani kao rezultati bez greške. Proračunati rezultati energija hibridizacije bili su u dobrom slaganju sa eksperimentalnim podacima drugih autora.

Ključne reči: veštačka neuronska mreža za procenu parametara • hemijski potencijal • asocijacija • stabilnost heliksa • obrada informacija o stanju DNK-a •

Key words: An artificial neural network for parameter estimation • Chemical potential • Association • Helix stability • DNA information processing •