Design of DNA Coding Based on Immune Clone Selection Algorithm

Yourui Huang
School of Electrical and Information Engineering, Anhui University of Science and Technology, Huainan. China
E-mail: hyr628@163.com

Liguo Qu and Chaoli Tang
School of Electrical and Information Engineering, Anhui University of Science and Technology, Huainan. China
E-mail: {qlg77, chaolitang}@163.com

***Abstract*—In DNA computing, the design of DNA code is the most important step. The definition of DNA code, the constraints of DNA encoding and the corresponding formula are firstly described in this article, and then the design of DNA encoding based on immune clone algorithm is introduced. Simulation experiments show that DNA molecular chains that satisfy specific constraints and the corresponding thermodynamic characteristics can be produced based on immune clone selection algorithm. Using immune clone selection algorithm can improves the efficiency of encoding by, and using a series of constraints can produce very stable DNA sequences. It laid a good foundation for DNA computing.**

***Index Terms*—Immune clone selection algorithm, DNA code, constraint condition, thermodynamics**

I. INTRODUCION

In 1987, Head firstly formalized the operation of DNA restructuring, introduced the concept of cutting system (H system) and researched the generative ability of language of this cutting system, these can be seen as the open symbol of DNA computing theory. In 1994, Adleman published the first experiment thesis about DNA computing on Science magazine, the basic content is taking DNA molecule as calculated medium, using modern molecular techniques as a means, it successfully solved the problem of Hamiltonian path with 7 vertices. Experimental results demonstrate that the possibilities of calculation based on the molecular level, DNA computing show up a lot of advantages in solving the NP complete problem, that advantage directly make DNA computation become a big hot research topic [1, 2, 3, 4].

DNA computation process is divided into 3 stages: (1) Stage of encode: encoding the problem to be solved into a collection of DNA molecules, namely as input. (2) Stage of computing: producing the solution space by carry through all kinds of possible biochemical reactions between DNA molecules, this stage also is biochemical reaction process. (3) Stage of extracting candidate solutions: in the solution space after calculating, solutions that meet specific conditions can be extracted by using molecular biological techniques, such as PCR technology, molecular purification, gel electrophoresis, magnetic separation and so on. This stage is process of identifies the target solution [5, 6, 7]. Summing up the above, it is clear that the accuracy of biochemical reactions will affect the results of DNA computing at large extent, and one of the key factors of affecting the biochemical reaction is the design of DNA encoding, so it is vital important to design DNA code by using appropriate bionic.

In this article, there have come up with a design of DNA encoding based on immune clone selection algorithm, Combining the characteristics of particular constraint formulas with artificial immune clone selection algorithm, result in the base distribution of DNA sequence are more uniform and have greater thermodynamic stability, it is helpful to avoid false positives in biochemical reactions. Thus immune clone selection algorithm not only maintained individual diversity and avoid fall into locally optimal solutions in evolutionary, iterative process, but also can improve local search capabilities and speed up the evolution and convergence rate by using memory unit.

II DEFINITION OF DNA ENCODING AND SPECIFIC CONSTRAINTS

A. Definition of DNA Encoding

The basic idea of DNA computing is: encode candidate solutions of practical problems into DNA chains, find optimal solutions of practical problems by proceeding biochemical reactions, so the key issue of whole DNA computing is research of design of DNA code. The higher of code quality is, the more reliable of results, the larger of the amount of code, the more of data that can be handled [8, 9].

The formalized language of encoding problem can be expressed as [10, 11, 12, 13]: on the alphabet {A, C, G, T} of a DNA molecule, there is a collection S of DNA molecules which encoding length is n, find a subset $C \subseteq S$, Allows for $\forall s_i, s_j \in C$ to meet:

doi:10.4304/jsw.7.10.2169-2176

$$\tau(s_i, s_j) \geq k. \quad (1)$$

Where, τ is the constraints condition to evaluate the nature of DNA encoding, such as the constraint function of Hamming distance, the constraint function of similarity, the constraint function of GC content and so on. K is a positive integer.

B Specific Constraints to Evaluate Quality of DNA Encoding

(1) The constraint function of similarity

The constraint function of similarity is calculate the similarity of DNA sequence of same two directions (5' to 3') under sliding matching, make DNA sequences as dissimilarity as possible. The definition of similarity is shown as follows [14, 15, 16]:

$$F_{similarity}(\Sigma) = \sum_{i=1}^{m} f_{similarity}(x_i). \quad (2)$$

$$f_{similarity}(x_i) = \sum_{\substack{1 \leq j \leq m \\ j \neq i}} \max_{0 \leq g \leq n} \max_{0 \leq k \leq n+g-1} S(x_i(-)^g x_i, \sigma^k(x_j)). \quad (3)$$

Where, $(-)^g$ show there are g gaps, $\sigma^k(x_j)$ show DNA sequence x_j right shift k steps, $S(x_i(-)^g x_i, \sigma^k(x_j))$ is the number of same characters in corresponding position of two DNA sequences $x_i(-)^g x_i$ and $\sigma^k(x_j)$.

(2) The constraint function of H_measure

The definition of H_measure is similar to the definition of similarity, the only difference is that H_measure is the opposite direction comparison of same DNA sequences $\sigma^k(x_j)$, while the definition of similarity is the same direction comparison of two DNA sequences. To given two DNA sequences, H_measure measure the extent of crisscrossing through calculated the logarithmic of Watson-Crick complementary base-pair. Like calculate similarity, slide displacement need to be considered, the calculated formula is defined as [17]:

$$F_{H\_measure}(\Sigma) = \sum_{i=1}^{m} f_{H\_measure}(x_i). \quad (4)$$

$$f_{H\_measure}(x_i) = \sum_{\substack{1 \leq j \leq m \\ j \neq i}} \max_{0 \leq g \leq n} \max_{0 \leq k \leq n+g-1} C(x_i(-)^g x_i, \sigma^k(x_j^R)). \quad (5)$$

Where $C(x_i(-)^g x_i, \sigma^k(x_j^R))$ is the number of same characters in corresponding position of two DNA sequence $x_i(-)^g x_i$ and $\sigma^k(x_j^R)$.

(3) The constraint function of GC content

The definition of GC content is the percentage of basic group G and C in a DNA sequence, GC content is not only in maintaining the stability of chemical properties of the sequence has a very important role, but also guarantee melting temperature T_m of DNA molecules in the reaction remain the same. Thus effectively avoid the occurrence of non-specific hybridization [18]. The constraint function of GC content is show as follows:

$$F_{GC}(\Sigma) = \sum_{i=1}^{m} f_{GC}(x_i). \quad (6)$$

$$f_{GC}(x_i) = [GC_{gen}(x_i) - GC_{tar}(x_i)]^2. \quad (7)$$

Where $GC_{tar}(x_i)$ is the expected GC content of DNA sequence x_i, $GC_{gen}(x_i)$ is the actual GC content of DNA sequence produced by algorithm, generally speaking, the value of $GC_{tar}(x_i)$ is in [40, 60].

(4) The constraint function of melting temperature T_m

The definition of melting temperature is temperature of when 50% DNA molecule appear denaturation in DNA double-stranded, melting temperature is a important parameter which determine efficiency of reaction. In order to effectively lower the occurrence of unmatching double-strands, the melting temperature T_m of DNA sequence that participate reaction should be as same as possible had asked to participate in reactions of DNA sequences as far as possible the same.

At present, there are many methods that can be used to calculate melting temperature T_m, such as nearest neighbor method, GC content, and so on [19]. In this paper, mainly uses nearest neighbor thermodynamic method to calculate melting temperature T_m, the calculation formula is shown as follows:

$$F_{Tm}(\Sigma) = \sum_{i=1}^{m} f_{Tm}(x_i). \quad (8)$$

$$f_{Tm}(x_i) = [Tm_{tar}(x_i) - Tm_{gen}(x_i)]^2. \quad (9)$$

$$Tm_{gen} = \frac{\Delta H^*}{\Delta S^* + R\ln(C_T/\alpha)} - 237.15. \quad (10)$$

Where, Tm_{tar} is target melting temperature; $Tm_{gen}(x_i)$ is the melting temperature of generated DNA sequence; R is the molar gas constant, the value is 1.987 CAL/MOL/k; C_T is the salt concentration of nucleic acid solution; ΔS and ΔH take a series of complex oligonucleotide as model, they respective are enthalpy change and entropy change; The value of α is 4.

(5) The constraint function of continuity

In DNA sequence, continuously same basic group may result in the production of unwanted biological structure.

Continuity is usually used to depict continuous degrees of the same base in DNA sequences. The calculated formula is shown as follows [20]:

$$F_{Con}(\Sigma)=\sum_{i=1}^{i=m} f_{Con}(x_i). \tag{11}$$

$$f_{Con}(x_i)=\sum_{i=1}^{n-t+1}\sum_{\alpha\in\wedge} T(C_\alpha(x,i),t)^2. \tag{12}$$

$$C_\alpha(x,i)=\begin{cases} c, \exists c, s.t.\ x_i\neq\alpha, x_{i+j}=\alpha, (1\le j\le c), x_{i+c+1}\neq\alpha \\ 0, else \end{cases}. \tag{13}$$

$$T(i,j)=\begin{cases} i, & i>j \\ 0, & else \end{cases}. \tag{14}$$

Where, $C_\alpha(x,i)$ is the number of continuous appear basic group α in the i-th position of DNA sequence x; t is a continuous threshold value, in this paper, the value is 2.

(6) The constraint function of hairpin

Hairpin structure is circular structure which formed by self-complementary of the DNA sequence, due to hairpin structure hybridizes with itself, hairpin structure is not allowed in many experiments. In order to avoid they hybridize and make Watson-crick pairing method more efficient, hairpin structure must be avoided. The calculated formula of hairpin is shown as follows:

$$F_{Hairpin}(\Sigma)=\sum_{i=1}^{m} f_{Hairpin}(x_i). \tag{15}$$

$$f_{Hairpin}(x_i)=\sum_{r}^{(n-2,pinien)}\ \sum_{c=pinlen+[r/2]}^{(n-pinlen-[r/2])} Hairpin(x_i,c). \tag{16}$$

Where r is the minimum length of loop in hairpin structure, pinlen is the minimum length of stalk in hairpin structure. To DNA sequence x_i, hairpin structure is formed by folding at the location c, If the number of complementary bases in the stem of hairpin structure is more than half of the length of stem, then the value of $Hairpin(x_i,c)$ is 1, else is 0.

In this text, these six functions need to be optimized simultaneously. It is clearly that the optimized problem is multi-objective problem. Generally speaking, GC content and melting temperature are often seen as constraints rather than objective functions, so the fitness function that evaluate the quality of DNA encoding can be seen as the weighted constraint functions sum except GC content and melting temperature. That is convert multi-objective optimization problem to single-objective optimization problems, and simplifies the optimization problem. Fitness function construct is shown as follows:

$$\begin{aligned} &Fitness=\sum_i w_i F_i(\Sigma) \\ &\quad i\in\{Similarity, H\_measure, Con, Hairpin\} \\ &Subject\ to \\ &\quad \{F_{GC}(\Sigma),\ F_{Tm}(\Sigma)\}. \end{aligned} \tag{17}$$

Where $F_i(\Sigma)$ is the constraint function above mentioned, w_i is the weight of every constraint function, For the sake of simplicity, the value of each weight is set to 1.

III IMMUNE CLONE SELECTION ALGORITHM

A The Principle of Immune Clone Selection Algorithm

The concept of immune system was presented based on biological immune system. The operation mechanism of immune system was summarized as follows:

(1) Immune Response: Immune cells recognize antigens and antibodies, and then they were activated, divided and proliferated new antibodies so as to achieve the purpose of eradicate antigens. Immune cells can dynamically adapt to the changing of external environment.

(2) Immune Memory: Biological immune system has two types of immune responses: initial immune response and secondary immune responses. When antigen invades organisms for the first time is the initial immune response, in this process, the immune system recognize and learn antigens, it produce memory cells by using clone selection mechanism. When the same antigens invade the body again will trigger secondary immune response. Through wake up memory cells to produce large amounts of antibodies to perish antigens in a shorter period of time, repeat this process the system can strengthen the immune system's memory capacity by produce memory cells which can fight known antigens

(3) Immune Regulation: During Immune responses between immune cells, between antigens and antibodies, between antibodies and antibody will form a dynamic balance system of interaction and mutual influence. When Antigens invade immune system, through a dynamic and real-time regulation of the immune system new immune balance can be achieved. Even though there are not antigens invasion, through the effect of mutual promotion and inhibition between antibodies, immune system also can adjusts itself to maintain dynamic equilibrium between the immune system.

The process of clone selection refer to the concept of affinity maturation, that is under the clone selection mechanism those individuals which have low affinity with antigen through after a series of reproduce, proliferation and mutation actions its affinity increased gradually and achieved mature finally. Theory of clone

selection was realized by using the operator of genetic, the operator of cross, and the operator of mutation and group's control mechanism.

Immune clone selection algorithm (the initials is ICSA) is a new kind of bionic intelligent algorithm and a model of the natural immune system, The algorithm has the properties that fast converge to the global optimization, which introduction the mechanism of affinity maturation, clone and memory based on the traditional bionic evolutionary algorithm and use the corresponding operators guarantee itself converge to the global optimization.

In immune algorithm, the objective function and the constraints of practical problems as antigen, the viable solutions of problem as antibody and the value of objective function of viable solutions as the affinity between antibody and Antigen. Immune algorithm always give priority to select those individual who has good affinity with antigen as descendant antibodies to achieve the aim of eradication antigen by preserving and reproducing those antibody who has good fitness

B DNA Coding Algorithm based on ICSA

Take solution of practical problems as antigen, and candidate solutions as antibody, then fitness values of feasible solutions is corresponding to the degree of affinity of antibodies and antigens. Only those cells that recognize antigens will be fragmentation, reproduction, multiplication, size of the selected cells depends on the degree of affinity, the greater affinity, the higher probability of being selected. Optimal solution must is the antibody which affinity with antigen is the largest.

In DNA computing, the quality of DNA encoding is vital important to result of calculation, measure the quality of DNA coding is good or bad depend on whether the DNA encoding are met constraint conditions or not. In this paper, take GC content and melting temperature as constraint condition, and (17) as fitness function of every function. According to the value of fitness, select the corresponding DNA sequence copy, variation. The basic steps of design of DNA encoding based on ICSA are shown as follows:

Step 1 code candidate solutions of actual problem into DNA sequence, take quaternary to encode candidate solutions, that is 0 represents A, 1 represents C, 2 represents G, 3 represents T. Group of DNA sequences have certain length and quantity are recorded as Ab, the group size is recorded as N.

Step 2 According to the (19) to calculate the fitness value of each DNA sequence in the group, the fitness value namely accessibility f.

Step 3 In above mentioned DNA sequence, according to the value of affinity select n DNA sequences in proper order to constitute new DNA sequence population Ab_n.

Step 4 copy new DNA sequence group Ab_n, new DNA sequence group is produced and recorded as C, in the process of copying. To DNA sequence, the higher affinity of DNA sequence is, the more offspring is produced by copying.

Step 5 Mutate the new DNA sequence C and mature DNA sequence group C' is produced, the formula of mutation probability is shown as follows:

$$p\{T_m(x)=Y\}=p_m^{H(X,Y)}(1-p_m^{L-H(X,Y)}). \quad (18)$$

Where $p_m>0$ is the mutation probability of antibody gene; H(X，Y) is the Hamming distance of antibody x and y, that is:

$$H(X,Y)=\sum_{i=1}^{L}\left|X^i-Y^i\right|. \quad (19)$$

Step 6 Calculate the affinity value f ' of DNA sequence in group C'.

Step 7 According to the affinity value f ', select n DNA sequences which affinity value is the highest in group C' to constitute new DNA sequence group Ab_n, then put Ab_n into the initial DNA sequence group Ab.

Step 8 Randomly generate m new DNA sequences to constitute new DNA sequence group Ab_m and instead of individuals which affinity value are the lowest in initial group Ab. Repeat step 2 to step 8.

The flowchart of entire DNA encoding based on immune clone selection algorithm is shown in the following figure:

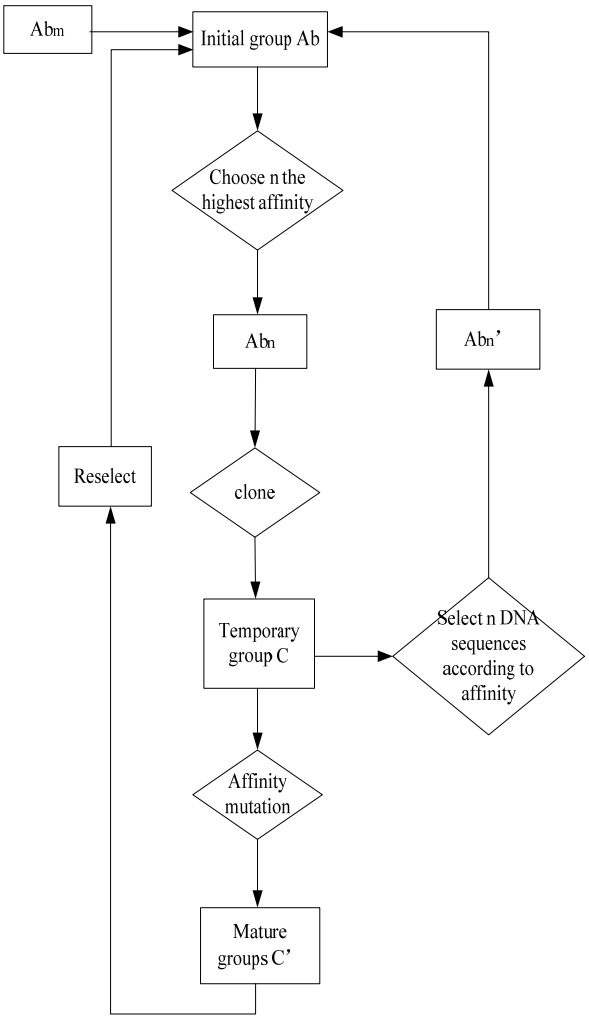

Fig. 1 The flowchart of entire DNA encoding based on immune clone select algorithm

IV EXPERIMENTAL RESULTS AND ANALYSIS

DNA sequences produced by ICSA take full account of the impact of H_measure, similarity, hairpin and continuity, and at the same time meet the constraints of GC content and melting temperature.

In this text, the range of melting temperature change is 30° ~ 80°, the range of GC content change is 30% ~ 80%. 7 DNA sequences of meeting all kinds of constraints are produced based on ICSA and are listed in table I.

TABLE I
SEVEN DNA SEQUENCES DESIGNED WITH ICSA WITH 20-MER LENGTH

DNA sequence	Melting temperature	GC content	H_measure	Similarity	Hairpin	Continuity
CGATTTTACTGGTGTGTGGC	42.46	50	40	43	0	16
TTCTCCAATCGCAGGACTT	42.55	45	43	42	0	0
GATTAGGAGTAGGTCCAGCC	43.65	55	44	47	0	0
GATTAGGAGTAGGTCCAGCC	41.68	50	43	42	0	0
GAGTGCTAACAGAAACCGTAC	39.94	40	42	44	0	18
TGAGATCCTCGTTCGCCATG	47.00	60	42	44	0	0
TCGCTATGCCATCCTCTTGT	34.56	50	44	44	0	0

It clearly can been seen from the above table, seven DNA sequences are meet all kinds of constraints. Melting temperature and GC content also are in the appropriate range of constraint, so the seven DNA sequences designed with ICSA are the candidate solutions of practical problem, then select optimal DNA sequence by calculating the fitness value of DNA sequences according to (17), the optimal DNA sequence namely the optimal solution of practical problem. In table II, there is a comparison between DNA encoding algorithms based on ICSA with algorithm of Deaton et al. Specific content is shown in the following table:

TABLE II
THE COMPARISON OF DNA SEQUENCE OBTAINED FROM ICSA AND ALGORITHM OF DEATON ET AL

DNA sequence(5'-3')	continuity	Hairpin	H_measure	similarity	Melting temperate	GCcontent
DNA sequence designed with ICSA						
AACAATGAATGGGCAGGAGT	9	3	54	56	52.90	45
CAGGACTAAACAATTCCAAA	18	3	53	60	46.90	35
CACATTACGCCAAGGATACC	0	0	54	53	52.25	50
GACCGCAAGACAGAAGAGAA	0	0	48	61	53.36	50
ACCGACGTCCGTAACTGACC	0	0	59	54	57.70	60
ACATGAGATCAACCTGCGCA	0	0	54	56	55.60	50
TAAGAGAATGCCAGAATAAG	0	0	50	60	45.50	35
DNA sequence designed with Deaton et al algorithm						
ATAGAGTGGATAGTTCTGGG	9	3	55	57	52.65	45
CATTGGCGGCGCGTAGGCTT	0	0	61	50	69.20	65
CTTGTGACCGCTTCTGGGGA	16	0	60	56	60.85	60
GAAAAAGGACCAAAAGAGAG	41	0	57	45	52.71	40
GATGGTGGTTAGAGAAGTGG	0	0	54	54	55.30	50
TGTATCTCGTTTTAACATCC	16	4	61	50	48.44	35
TTGTAAGCCTACTGCGTGAC	0	3	66	56	56.70	50

It clearly can been seen from the table II, DNA sequences designed with ICSA are superior to those designed with algorithm of Deaton et al in continuity, H_measure, hairpin and similarity. Because DNA sequences designed with ICSA have lower hairpin structure, continuity, H_measure, so this method can effectively reduce the probability of non-specific hybridization, improve quality of DNA encoding, and eventually improve the efficiency of DNA computing. In this text, the range of melting temperature of DNA sequence is 45.5851-57.7230, but the range of melting temperature in Ref. [22] is 48.4451-69.2009. It is clear that the scope of the former is narrower than the latter, so the melting temperature of DNA sequences designed with ICSA is more uniform, that is these DNA sequences have better combination property.

The comparison of table II is displayed in the form of histogram in Fig. 2, the histogram makes the comparison more clear. In the Fig. 2, it is clearly can been seen that blue bar represents the average of objective function in the ICSA, red bar represents the average of objective function in the algorithm of Deaton et al. In Fig. 1, it is clear that ICSA have more advantages than the algorithm of Deaton et al in continuity, hairpin, H_measure and similarity. The example fully demonstrates that the design strategy of DNA encoding based on ICSA is effectiveness.

At last, there is a comparison between the design strategy of DNA encoding based on ICSA with the design strategy of DNA encoding based on multi-objective particle swarm optimization (MOPSO) [24], genetic algorithm (GA) [23] and multi-objective evolution algorithm (NACST/Seq) [12]. In table III the value of constraint function of DNA sequences designed with various algorithms is shown in detail, the comparison results are displayed in the form of histogram and shown in the Fig. 3.

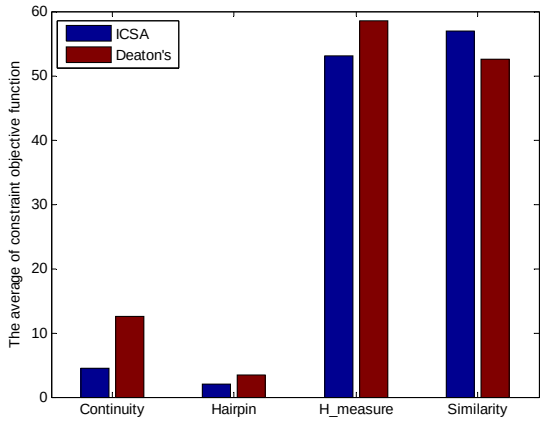

Fig. 2 Graphical comparison for average objective value of ICSA and algorithm of Deaton et al

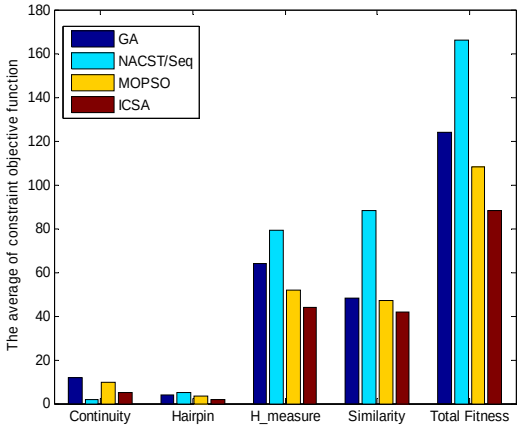

Fig. 3 Graphical comparison for average objective value of ICSA and others algorithm

TABLE III
COMPARISON OF DNA SEQUENCES DESIGNED WITH ICSA AND OTHERS ALGORITHM

DNA sequences (5'-3')	continuity	hairpin	H_measure	similarity
DNA sequences designed with ICSA				
CGATTTTACTGGTGTGTGGC	16	0	40	43
TTCTCCAATACGCAGGACTT	0	0	43	42
GATTAGGAGTAGGTCCAGCC	0	0	44	47
GAGTGCTAACAGAACCGTAC	0	0	43	42
ATTTGGGAATTGATGTTGCT	18	0	42	44
TGAGATCCTCGTTCGCCATG	0	0	42	44
TCGCTATGCCATCCTCTTGT	0	0	44	44
DNA sequences designed with GA				
ATAGAGTGGATAGTTCTGGG	9	0	57	55

CATTGGCGGCGCGTAGGCTT	0	0	65	44
CTTGTGACCGCTTCTGGGGA	16	0	67	60
GAAAAAGGACCAAAAGAGAG	41	0	58	40
GATGGTGCTTAGAGAAGTGG	0	0	54	51
TGTATCTCGTTTTAACATCC	16	4	72	41
TTGTAAGCCTACTGCGTGAC	0	0	71	47
DNA sequences designed with NACST				
CTCTTCATCCACCTCTTCTC	0	0	70	95
CTCTCATCTCTCCGTTCTTC	0	0	60	99
TATCCTGTGGTGTCCTTCCT	0	0	76	86
ATTCTGTTCCGTTGCGTGTC	0	0	82	94
TCTCTTACGTTGGTTGGCTG	0	0	86	89
GTATTCCAAGCGTCCGTGTT	0	0	91	88
AAACCTCCACCAACACACCA	9	0	89	73
DNA sequences designed with MPSO				
CATCAGCCGGACTCGTCAGT	0	0	47	45
AGATCGCATGTAAAGGAGTG	9	0	51	45
AAAGCAGGGTGTATCAGTCA	18	0	47	47
TACAGGCGCTAATTAGCTCC	0	0	67	43
GCGGACCCAACACATATGAG	9	0	51	48
ATCATCATTTCATGGGGCAA	25	0	52	43
GGGATCGACGTATATTAACG	9	0	49	49

Fig. 3 demonstrate that the reliability of DNA encoding algorithm based on ICSA is higher than other DNA encoding algorithm, such as genetic algorithm which proposed by Deaton et al, multi-objective evolution algorithm (NACST/Seq) which proposed by Shin et al, and so on. It clearly can been seen from Fig. 3 that ICSA is much better than other algorithm above mentioned on the performance of other constraints function except for continuity. However the continuity of DNA sequence designed with NACST is the minimum, the total fitness value of DNA sequence designed with ICSA is the minimum. So in the process of optimizing DNA sequence, ICSA can obtain desirable results, it fully proves that the policy proposed in this article is effective and useful.

V. CONCLUSIONS

In the process of designing DNA encoding, ICSA has successfully completed the optimization for objective function. Studies have shown that ICSA is a very reliable algorithm, and is more convenient and effective in implementation when compared with the other existing algorithms. Using this coding strategy can effectively improve quality of DNA encoding, this ensures the reliability of DNA computing. Generally speaking, immune clone selection algorithm is effective in solving the question of designing DNA sequence.

It also provides a new solution for designing DNA sequence and broadens the area of application.

ACKNOWLEDGMENT

The work is supported by the National Natural Science Foundation of China (61073101) and Anhui Provincial University Key Project of Natural Science (KJ2011A095)

REFERENCES

[1] L. Adleman, “Molecualr computation of solution to combinatorial problems,” Science, vol. 266, pp. 1021-1024, 1994.

[2] K. Sakamoto, H. Gouzu, K. Komiya, “Molecular computation by DNA hairpin formation,” *Science*, vol. 288, pp. 1223-1226, 2000.

[3] C. H. Bennett, “On constructing a molecular computer,” *IBM Journal of Research and Development*, vol. 17, pp. 525-532, 1973.

[4] M. Conrad, “On design principles for a molecular computer,” *Communications of the ACM*, vol. 28, pp. 464-480, 1985.

[5] R. Lipton, “DNA solution of hard computational problems,” *Science*, vol. 268, pp. 542-545, 1995.

[6] N. Jonoska, “Trends in computing with DNA,” *Journal of Computer Science and Technology*, vol. 19, pp. 98-113, 2004.

[7] T. Head, “Formal language theory and DNA: an analysis of the generative capacity of specific recombinant

behaviors," *Bulletin of Mathematical Biology*, vol. 49, pp. 737-759, 1987.
[8] Q. Liu, L. Wang, A. Frutos, "DNA computing on surface," *Nature*, vol. 403, pp. 175-178, 2000.
[9] H. Y. Wu, "An improved surface-based method for DNA computation," *Biosystems*, vol. 59, pp. 1-5, 2001.
[10] T. Head, G. Rozenberg, R. B. Bladergroen, "Computing with DNA by operating on plasmids," *Biosystems*, vol. 57, pp. 87-93, 2000.
[11] Y. Benenson, B. Gil, Ben-Dorl Uri, "An autonomous molecular computer for logical control of gene expression," *Nature*, vol. 429, pp. 423-429, 2004.
[12] S. Y. Shin, I. H. Lee, D. Kim, and B. T. Zhang, "Multiobjective evolutionary optimization of DNA sequence for reliable DNA computing," *IEEE Transactions on Evolutionary Computation*, vol. 9, pp. 143-158, 2005.
[13] S. B. Ravinderjit, C. Nickolas, J. Cliff, "Solution of a 20-variable 3-SAT problem on a DNA computer," *Science*, vol. 296, pp. 499-502, 2002.
[14] Q. Ouyang, P. D. Kaplan, S. Liu, "DNA solution of the maximal clique problem," *Science*, vol. 278, pp. 446-449, 1997.
[15] Z. X. Yin, F. Y. Zhang, J. Xu, "A Chinese postman problem based on DNA computing," *Journal of Chemical Information and Computer Sciences*, vol. 42, pp. 222-224, 2002.
[16] Z. X. Yin, J. Z. Cui, X. L. Shi, "The semi-automated DNA computing model for solcing the satisfiability problem," *Intermational Journal of Computer Science and Network Security*, vol. 6, pp. 163-167, 2006.
[17] M. S. Muhammad, K. V. Selvan, S. M. W. Masra, Z. Ibrahim, A. F. Z. Abidin, "An improved binary particle swarm optimization algorithm for DNA encoding enhancement," 2011 *IEEE Symposium on Digital Object*, pp. 1-8, 2011.
[18] G. Z. Cui, X. G. Li, "The Optimization of DNA Encodings Based on Modified PSO/GA Algorithm," *2010 International Conference On Computer Design And Appliations*, vol. 1, pp. 609-614, 2010.
[19] H. Y. Zhang, X. Y. Liu, "Improved Genetic Algorithm for Designing DNA Sequences," *2009 Second International Symposium on Electronic Commerce and Security*, pp. 514-518, 2009.
[20] [20] J. H. Xiao, C. Zhen, "DNA Sequences Optimization Based on Gravitational Search Algorithm for Reliable DNA computing," *2011 Sixth International Conference on Bio-Inspired Computing: Theories and Applications*, pp. 103~107, 2011.
[21] T. B. Kurniawan, N. K. Khalid, Z. Ibrahim, M. S. Z. Abidin, M. Khalid, "Sequence Design for Direct-proportional Length-Based DNA Computing using populationa-Based Ant Colony Optimization," *ICROS-SICE International Joint Conference 2009*, pp. 1486-1491, 2009.
[22] R. Deaton, R. C. Murphy, M. Garzon, "Good encoding for DNA-based solutions to combinatorial problems," *Proceedings of 2nd Annual DIMACS Meeting on DNA Based Computers*, pp. 131-140, 1996.
[23] L. Gao, J. Xu, "DNA solution of vertex cover problem based on sticker model," *Chinese Journal of Electronics*, vol. 11, pp. 280-284, 2002.
[24] S. Zhou, Q. Zhang, J. Zhao, and J. Li, "DNA encoding based on multiobjective particle swarm," *Journal of Computational and Theoretical Nanoscience*, vol. 4, pp. 1249-1252, 2007.

Yourui Huang was born in 1971. He is a Ph.D. and a professor in Anhui University of Science and Technology, Huainan city, Anhui Province, P.R.China. His research interests are intelligent information process and wireless sensor network.

Liguo Qu was born in 1979.He is a lecturer in Anhui University of Science and Technology, Huainan city, Anhui Province, P.R.China. His research interests is wireless sensor network.

Chaoli Tang was born in 1980.He is a lecturer in Anhui University of Science and Technology, Huainan city, Anhui Province, P.R.China. His research interests is Intelligent information processing.