Dynamic Neighborhood Searches for Thermodynamically Designing DNA Sequence[⋆]

Suguru Kawashimo, Hirotaka Ono, Kunihiko Sadakane,
and Masafumi Yamashita

Dept. of Computer Science and Communication Engineering, Kyushu University,
744 Motooka, Nishi-ku, Fukuoka, Fukuoka 819-0395, Japan
kawa@tcslab.csce.kyushu-u.ac.jp, {ono,sada,mak}@csce.kyushu-u.ac.jp

Abstract. We present a local search based algorithm designing DNA short-sequence sets satisfying thermodynamical constraints about minimum free energy (MFE) criteria. In DNA12, Kawashimo et al. propose a dynamic neighborhood search algorithm for the sequence design under hamming distance based constraints, where an efficient search is achieved by dynamically controlling the neighborhood structures. Different from the hamming distance based constraints, the thermodynamical constraints are generally difficult to handle in local-search type algorithms. This is because they require a large number of evaluations of MFE to find an improved solution, but the definition of MFE itself contains time-consuming computation. In this paper, we introduce techniques to reduce such time-consuming evaluations of MFE, by which the proposed dynamic neighborhood search strategy become applicable to the thermodynamical constraints in practice. In computational experiments, our algorithm succeeded in generating better sequence sets for many constraints than exiting methods.

Keywords: DNA Sequence Design Algorithm, Local Search, Statistical Thermodynamical Constraints.

1 Introduction

Designing DNA sequence sets is a fundamental issue in the fields of nanotechnology and nanocomputing, e.g., Adleman's DNA solution for the Hamiltonian path [1], DNA tiling with its self-assemble [22], hairpin-based state machine [10] and so on. One aspect of DNA computing / technology is to control the DNA molecules reactions. For a robust "computation", it is desirable that DNA molecules react only in expected ways, because unexpected secondary structures of DNA sequences may cause error, for example. Sequence design is an approach for a robust computation by designing DNA sequences that satisfy some constraints to avoid unexpected molecular reactions. Since expected or unexpected reactions depend on the applications or the purposes, several representative constraints

[⋆] This research partly received financial support from Scientific research fund of Ministry of Education, Culture, Sports, Science and Technology.

M.H. Garzon and H. Yan (Eds.): DNA 13, LNCS 4848, pp. 130–139, 2008.

are usually considered as below mentioned. Another requirement for DNA sequence sets is to be large. This is because designed DNA sequences are used as elemental components of computation; the amount of resources on DNA computation is proportional to the size of a sequence set. In summary, systematic methodologies are required to design large set of sequences, which satisfy certain types of constraints.

In the sequence design, constraints are introduced to prohibit unexpected secondary structures of DNA sequences, and several types of prohibition are proposed. Roughly speaking, the types of prohibitions are classified into combinatorial types and thermodynamical types. Combinatorial constraints are based on the idea that base conjugations of DNA sequences are regarded as a kind of combinatorial pattern matching, while the thermodynamical constraints are based on the thermodynamical property of the molecular reaction mechanism, in which conformations of small (resp., large) Gibbs standard free energies tend to be stable (resp., unstable). Although the thermodynamical ones seem to be more sophisticated, many algorithmic studies of the sequence design have treated combinatorial constraints due to their simplicity. Also the combinatorial properties help to bring efficient algorithms from combinatorics or combinatorial optimization fields [3,4,19,20]. On the other hand, there are few studies under thermodynamical constraints from the combinatorial algorithmic point of view; one example that the authors know is a Stochastic Local Search method by Tulpan et al. [21].

In this paper, we consider DNA sequence design algorithm under thermodynamical constraints from the viewpoint of the combinatorial optimization. More precisely, we propose a local-search type algorithm for the DNA sequence design under thermodynamical ones. A local search is a method to find a good solution by replacing a current solution with a better (improving) solution in its neighborhood until no better solution is found. In DNA12, the authors proposed a dynamic neighborhood search algorithm for DNA sequence design problem [11], which targets on short-sequence sets under combinatorial constraints. The algorithm is equipped with high search performance by changing the neighborhood structures dynamically. The computational experiments show a good design power of the algorithm; it succeeded in generating better sequence sets than exiting methods [3,4,19]. Also by the nature of local search methods, it has a good flexibility; we can finely adjust the constraints. Therefore, we attempt to implement the idea of our previous algorithm for the thermodynamical constraints, especially *Minimum Free Energy* (MFE, for short) constraints, in this paper.

However, such an implementation is nontrivial in general. The Gibbs standard free energy is an energy value associated with the conformation of a sequence or sequences given, and the MFE is the minimum value among free energies of all the possible structures. Namely, the definition of MFE itself contains a time-consuming calculation, and in fact its time complexity is $O(n^3)$ time where n is the length of a given sequence. That is, a large amount of evaluations of MFE values are not practical, which implies that a local-search type algorithms are not suitable since they need to repeatedly evaluate many solution values. A main contribution of this paper is to overcome this difficulty; we present two techniques to circumvent the heavy calculations. One is to realize an effective neighborhood search. For this purpose, we store extra data among bases of DNA sequences, by which we can find a base involved with the violation for MFE constraints. The other is to realize an efficient evaluation of MFEs. In neighborhood searches, most of neighbor solutions are apparently worse, and only a few of them are candidates of improving solutions. For screening such apparently worse solutions, we introduce a preprocessing phase in the search; instead of applying $O(n^3)$ time MFE calculation, we utilize an approximate calculation of the MFE.

By these techniques, our search framework introduced in [11] becomes applicable to the MFE-based constraints in practice. In order to see the performance of our approach, we conduct computational experiments for various settings of MFE constraints. The results show that we succeeded in designing a large set of sequences for many case. One virtue of our algorithm is that it is a practical local search: It is quite flexible and is easy to introduce a new constraint. Moreover, if a non-local-search type algorithm finds a (good) sequence set, then we may obtain an even better solution by applying our algorithm to the solution.

1.1 Related Work

Many studies consider the thermodynamical natures of DNA computing from various points of view (e.g., [14,16]), and the thermodynamical qualities of sequence sets are also discussed in several papers. Especially, Dirks et al. [7] discuss various thermodynamical criteria of designing secondary structures, and Rose et al. [15] propose a statistical thermodynamic error model in DNA computing.

Tulpan et al. succeeded in designing sequence sets under very complicated thermodynamical constraints by Stochastic Local Search method [21], though the running time is not clear because they evaluated the search time except the calculation of energy values in their experiments. They also proposed new thermodynamical constraints. One advantage of their method is that they can treat complicated constraints as well as ours, since it is a local-search type algorithm. Garzon et al. also designed sequence sets [9]. They designed sequence set under combinatorial constraint as preprocessing, and remove thermodynamically violated sequences from the set obtained in preprocessing by the reduction to the minimum vertex cover problem (actually, they consider the maximum independent set). However the minimum vertex cover problem itself is known to be NP-hard. Tanaka et al. used random-generation based method [18]. To reduce the calculating-time of evaluation, they proposed approximate method of calculating MFE by the greedy manner.

The remainder of the paper is organized as follows: Section 2 gives preliminaries of the paper, thermodynamical constraints, and basic definitions for local search. Section 3 discusses how the heavy MFE calculations can be embedded into our search framework. Section 4 shows the results of computational experiments, and then Section 5 concludes the paper.

2 Preliminaries

2.1 Definitions and MFE Constraints

A DNA sequences s is a string over {A,T,C,G}. A DNA sequence or sequences form *secondary structures* by the Watson-Crick property, which are also called *conformations*. Each conformation of a sequence (or sequences) has a Gibbs standard *free energy*. The *Minimum Free Energy* (MFE, for short) of a sequence (resp., sequences) is the minimum value among free energies of all possible conformations of a sequence (resp., sequences). It is known that a conformation with a small Gibbs standard free energy is more stable than ones with larger Gibbs standard free energies. The Gibbs standard free energy values are measured through actual experiments and we can compute the value for one conformation in linear time of the length of the sequence.

Let s, s' be DNA sequences of length n, then $s, s' \in \{\mathtt{A}, \mathtt{T}, \mathtt{G}, \mathtt{C}\}^n$. Sequences are represented by $s = s_1 s_2 \cdots s_n$, and $s' = s'_1 s'_2 \cdots s'_n$. In these representations, the left end of a sequence corresponds to $5'$ end of a DNA sequence. In addition, $wcc(s)$ denotes the Watson-Crick complement sequence of DNA sequence s, here, $wcc(s)$ is the sequence which reverse s and replaced each A in s by T and vice versa, replaced each G in s by C and vice versa.

Let S be the sequence set. In the context of the sequence design problems, we let "hybridization" refer to "the phenomenon that a sequence in S forms completely hydrogen bonds with its complement sequence", and "miss-hybridization" refer to "conformations which are not hybridization". The constraints described below are introduced in order to avoid miss-hybridization.

The MFE between s and s' is represented by $\Delta G(s, s')$ which can be calculated $O(n^3)$-time by the dynamic programming [2,12,23].

Let $wcc(S) = \{wcc(s) | s \in S\}$. Given threshold parameters t_{ww}, t_{wc}, and t_{cc}, we define the following constraints based on the MFE measure:

> **Word-Word Constraint:** for all pairs of s, s' in S, $\Delta G(s, s') \geq t_{ww}$. That is, $\Delta G_{ww}(S) \overset{\text{def}}{=} \min_{s,s' \in S}\{\Delta G(s, s')\} \geq t_{ww}$.
>
> **Word-Complement Constraint:** for all pairs of s in S, s' in $wcc(S)$, and $s \neq wcc(s')$, $\Delta G(s, s') \geq t_{wc}$.
> That is, $\Delta G_{wc}(S) \overset{\text{def}}{=} \min_{s \in S, s' \in wcc(S), s \neq wcc(s')}\{\Delta G(s, s')\} \geq t_{wc}$.
>
> **Complement-Complement Constraint:** for all pairs of s, s' in $wcc(S)$, $\Delta G(s, s') \geq t_{cc}$. That is, $\Delta G_{cc}(S) \overset{\text{def}}{=} \min_{s,s' \in wcc(S)}\{\Delta G(s, s')\} \geq t_{cc}$.

Note that, in these constraints, self reactions of one sequence are under consideration. On the other hand, we do not concern with pseudo-knots.

In this paper, we adopt only three constraints for the sequence design, following the work by Garzon et al. [9]. This does not mean that our algorithm is specified to these constraints, and it is applicable to many other criteria based on MFE (e.g. energy gap [21]). For other criteria, such as melting temperature and DNA error rate [15], though we may need careful adjustments, it is also applicable.

By using these, our problem is described as "*find* S *such that* $\Delta G_{ww}(S) \geq t_{ww}$, $\Delta G_{wc}(S) \geq t_{wc}$ *and* $\Delta G_{cc}(S) \geq t_{cc}$ *for large* t_{ww}, t_{wc}, *and* t_{cc}".

2.2 Local Search, Neighborhood and Objective Functions

A local search is a method to find a solution by replacing a current solution with a solution which has better *objective function* value in its *neighborhood* until no better solution is found. In DNA12, we proposed a local search based algorithm for DNA sequence design problem under combinatorial constraints. In this paper, we apply this algorithm for thermodynamical constraints. We hope interested readers refer to [11], in which more details about our algorithm can be found[1].

We define the neighborhood of S (we represent it as $N(S)$) for the local search as follows: sequence sets obtained by flipping 1 base of a sequence belonging to S. Due to the simplicity of the definition, we can flexibly apply it to various constraints.

In this problem, we need to design the set such as $\Delta G_{ww}(S) \geq t_{ww}$, $\Delta G_{wc}(S)$ $\geq t_{wc}$, and $\Delta G_{cc}(S) \geq t_{cc}$. Therefore, when we take together these constraints, the objective function is described as follows:

$$\begin{aligned}\Delta G_{min}(S) \overset{\text{def}}{=} \ & \min\{\Delta G_{ww}(S) - t_{ww}, 0\} + \\ & \quad \min\{\Delta G_{wc}(S) - t_{wc}, 0\} + \\ & \quad\quad \min\{\Delta G_{cc}(S) - t_{cc}, 0\}. \end{aligned} \tag{1}$$

By definition, $\Delta G_{min}(S) = 0$ means that it satisfies the constraints, and it takes $O(m^2 n^3)$ time to evaluate $\Delta G_{min}(S)$.

3 Techniques to Reduce MFE Evaluations

In the local search, to determine if the neighbor solution is an improving solution or not, its solution value should be calculated. This operation is executed many times, since the size of neighborhood is usually very large. As mentioned above it takes $O(m^2 n^3)$ time to evaluate one solution, but that running time can be reduced in our neighborhood search, because all pairs of sequences for S and all pairs of sequences for $S' \in N(S)$ are overlapping. By reusing the calculation of $\Delta G_{min}(S)$, the calculation of $\Delta G_{min}(S')$ for $S' \in N(S)$ can be done in $O(mn^3)$ time.

However, it is still too time-consuming. That is, naive local search type algorithms may not work well. In this section, we explain two techniques by which we skip such a large amount calculations. One is a device to effectively check neighbor solutions, and the other is to screen bad solutions without calculating the exact $\Delta G(s, s')$.

[1] In this paper, we use the new framework which is simplified and improved from the previous one.

3.1 Effective Neighborhood Search

In the neighborhood search, we need to evaluate ΔG_{min} of neighborhood solutions to determine if we move to the solution or not. This means that evaluating ΔG_{min} for worse solutions is wasting time; by effectively finding an improving solution we can reduce the calculation of ΔG_{min} values. In this subsection, we explain how to realize a fast discovery of improving solutions. More concretely, we define a good order of checking the neighbor solutions, in which solutions to be likely improvements have high priorities.

To define the ordering, we use an array *min_related* as counters for bases in S; *min_related* is on all the bases in S, and *min_related*(s_i) for a base s_i of $s \in S$ stores the number of occurrences of base s_i for $\Delta G_x(S)$ where $x \in \{ww, wc, cc\}$. The idea itself was introduced in the previous work [11], but it is extended from the previous one. Here, an "occurrence of base s_i for $\Delta G_x(S)$" means the following two conditions are satisfied: (i) s containing the base s_i and another sequence s' have the MFE value equal to $\Delta G_x(S)$ and (ii) in the MFE structure of the $\Delta G_x(S)$, s_i forms hydrogen bonds. If a base has a large value of *min_related*, the base may be critical for $\Delta G_{min}(S)$, therefore flipping such a base probably improves the solution value. On the other hand, flipping bases with *min_related* $= 0$ does not change the solution value by definition. Therefore, we define the search order of neighbor solutions in $N(S)$ according to *min_related* values of the descending order. In case of ties, i.e., some bases have a same value of *min_related*, we use another array *bond_related* to determine the order. The *bond_related* on all the bases similarly stores the number of occurrences of a base for hydrogen bonds about not MFE-structures of $\Delta G_{min}(S)$ but all MFE-structures. By a similar argument, we define the search order for ties in *min_related* according to *bond_related* of the descending order.

Table 1 shows results of preliminary computational experiments concerning the effectively of *min_related* and *bond_related*. This result shows that the ordering based on *min_related* and *bond_related* apparently realizes an effective search.

Table 1. Result of the preliminary experiments for *min_related* and *bond_related*

n	m	τ	time; with *min_related*	time; random order
			average / standard deviation	average / standard deviation
8	30	-6.0(kcal/mol)	1.87(sec)/ 0.81(sec)	19.34(sec)/ 11.03(sec)
12	50	-10.0(kcal/mol)	5.79(sec)/ 4.54(sec)	120.61(sec)/116.17(sec)
15	20	-6.0(kcal/mol)	29.65(sec)/19.91(sec)	137.47(sec)/ 35.52(sec)
16	30	-8.0(kcal/mol)	34.31(sec)/ 9.40(sec)	255.44(sec)/ 87.65(sec)
20	40	-6.0(kcal/mol)	29.00(sec)/16.77(sec)	480.69(sec)/270.04(sec)

Give a length n, a size m, and τ. Randomly generate initial set which has m sequences, and apply our algorithm to improve the set until it satisfies $\Delta G_{ww}(S) \geq \tau$, $\Delta G_{wc}(S) \geq \tau$, and $\Delta G_{cc}(S) \geq \tau$. We measure the running-time to satisfy the constraint with *min_related* and *bond_related* or random order. We perform 50 trials for each condition.

3.2 Efficient Evaluation of MFEs

In neighborhood structures of the search, a good solution has a few good neighbor solutions and many worse neighbor solutions in general. This means that we need to check many neighbor solutions with worse solution values to find a neighbor solution with a better solution value. This means that the total evaluation-time is mainly occupied by evaluations of worse solutions; if we can quickly reject such bad solutions, we may greatly reduce the total evaluation-time. In this subsection, we explain how to screen bad solutions efficiently. We introduce a preprocessing phase that computes an approximate MFE values. A similar approach is also used in [18], but in a little different context[2].

We define the approximate MFE as the minimum Gibbs standard free energy under the restriction in which self reaction in one sequence is forbidden and the size of loop is bounded[3]. The approximate MFE is denoted by $\Delta G_{app}(s, s')$. This value itself can be computed in $O(ln^2)$ time by dynamic programming in theory, where l is the maximum-loop-size. Clearly $\Delta G_{app}(s, s') \geq \Delta G(s, s')$ holds.

We define $\Delta G_{app}(S)$ as an approximate of $\Delta G_{min}(S)$ (equation (1)) in which $\Delta G_{app}(s, s')$ is used instead of $\Delta G(s, s')$. To check if a neighbor solution S_{new} is an improvement from S_{old}, we perform the following operation using $\Delta G_{app}(S)$ in the preprocessing phase:

(1) Calculate $\Delta G_{app}(S_{new})$. If $\Delta G_{app}(S_{new}) \geq \Delta G_{min}(S_{old})$, we can determine that S_{new} is not an improvement of S_{old}. (End this routine.) Otherwise, go to (2).

(2) Calculate $\Delta G_{min}(S_{new})$. If $\Delta G_{min}(S_{new}) \geq \Delta G_{min}(S_{old})$, we determine that S_{new} is not an improvement of S_{old}. Otherwise, we determine that S_{new} is an improvement. End this routine.

The performance of this operation depends on the approximation quality of $\Delta G_{app}(s, s')$. If we set sufficiently large maximum-loop-size l, then the approximation quality is good enough but it takes large time, since the value can be computed in $O(ln^2)$ time. Therefore, to see the quality of $\Delta G_{app}(s, s')$ for a small l, we perform preliminary computational experiments. In the experiments, we randomly generate 10000 pairs of sequences, and calculate $\Delta G(s, s')$ and $\Delta G_{app}(s, s')$ for each pair. Table 2 shows the results.

As shown in this table, the calculating-times of $\Delta G_{app}(s, s')$ are much faster than $\Delta G(s, s')$, while the approximation ratios are good for short lengths. This might be because short sequences hardly take self reactions. In particular, the calculating-times of $\Delta G_{app}(s, s')$ are 1/4 to 1/5 compared with that of $\Delta G(s, s')$, while the ratio of $\Delta G_{app}(s, s') = \Delta G(s, s')$ are very high especially for small lengths. This result if preferable when we design sets of short sequences. Therefore, we adopt the screening phase by $\Delta G_{app}(s, s')$ in our search strategy.

[2] In [18], they introduce a notion of "degree" k, and an $O(kn^2)$ time greedy algorithm for the approximation is proposed.

[3] The function `pairfold_mfe_nointra` included `PairFold` package [2] can calculate this.

Table 2. Preliminary Experiments for $\Delta G_{app}(s, s')$

n	match ratio	$\Delta G_{app}(s, s') - \Delta G(s, s')$ average / standard deviation	time of $\Delta G(s, s')$	time of $\Delta G_{app}(s, s')$
8	99.91%	0.00043(kcal/mol) / 0.01646(kcal/mol)	1.69(sec)	0.45(sec)
12	97.33%	0.01472(kcal/mol) / 0.11472(kcal/mol)	6.50(sec)	1.25(sec)
15	91.70%	0.05370(kcal/mol) / 0.14155(kcal/mol)	13.61(sec)	2.81(sec)
16	88.52%	0.07915(kcal/mol) / 0.30228(kcal/mol)	14.06(sec)	3.24(sec)
20	68.50%	0.30268(kcal/mol) / 0.66313(kcal/mol)	30.7(sec)	6.30(sec)

We set the parameters (n,l)=(8,2),(12,2),(15,4),(16,5),(20,6) in this experiment. Column "match ratio" shows the ratio of pairs satisfying $\Delta G_{app}(s, s') = \Delta G(s, s')$.

4 Computational Experiments

We implement the algorithm, and perform computational experiments. We use `PairFold` package [2] for calculation of MFEs. The setting temperature is 37°C. The cpu-times of experiments are between 2 hours and 24 hours.

We compare our results with Garzon et al. [9], also we compare with Deaton et al. [6][4]. Penchovsky et al. [13], Shortreed et al. [17], Braich et al. [5] and Faulhammer et al. [8].

Table 3. Results of Computational Experiments

No	n	$\Delta G_{ww}(S)$	$\Delta G_{wc}(S)$	$\Delta G_{cc}(S)$	size of ours	size of compared set
1	8	-3.3	-5.4	-3.9	**237**	(Garzon)132
2	8	-5.3	-6.5	-5.5	**233**	(Garzon)173
3	12	-3.5	-9.3	-4.5	**152**	(Shortreed)64
4	12	-5.9	-9.9	-5.9	321	(Garzon)**617**
5	12	-9.2	-11.2	-10.0	689	(Garzon)**1424**
6	15	-4.3	-8.3	-4.3	**80**	(Braich)40
7	15	-3.7	-10.4	-4.5	**85**	(Faulhammer)20
8	15	-6.0	-14.9	-7.3	**92**	(Garzon)42
9	15	-12.3	-15.3	-12.3	**224**	(Garzon)96
10	16	-7.5	-8.1	-8.5	**141**	(Shortreed)64
11	16	-1.5	-8.7	-3.9	**53**	(Penchovsky)24
12	20	-7.7	-7.2	-10.2	**88**	(Deaton)40

Table 3 shows the result. Column "size of ours" (resp., "size of compared set") shows size of sets obtained by our method (resp., the size of sets reported in [5,6,8,9,13,17]). Columns $\Delta G_{ww}(S)$, $\Delta G_{wc}(S)$, and $\Delta G_{cc}(S)$ represent parameters used in the experiment, which are obtained by the compared method[5]. For example, in No.1, we design S such as $\Delta G_{ww}(S) = -3.3$, $\Delta G_{wc}(S) = -5.4$,

[4] Deaton's set includes nucleotide "h", we treat "h" as "g".

[5] Except Garzon's set, values are calculated by us with published `PairFold` package.

$\Delta G_{cc}(S) = -3.9$, which are based on the result of Garzon's set with size 132. Sequence sets with greater sizes are better if the values of $\Delta G_x(S)$ are same.

For the cases of No.1–3, and 6–12, our sets have the same $\Delta G_x(S)$ values as the sets generated by the existing methods, however sizes of these are greater. That is to say, in spite that the sets generated by our method are larger, our sets cause "miss-hybridization" as well as compared sets. This implies that our method can design good sets and efficient for thermodynamical constraints.

Only for the cases of No.4 and 5, although these have same $\Delta G_x(S)$ values, the sizes of sets generated by our method are smaller than these generated by Garzon et al. Thus, we consider that our method is suitable for designing short sequence sets which are relatively small. However our method does not lose its worth even for the longer sequences, because we can treat the set generated by Garzon et al. as the initial set and may improve it.

5 Conclusion

In this paper, we present a local search type algorithm for short-sequence design under the thermodynamical constraints. Since local search type algorithms are not practical under the thermodynamical constraints due to time-consuming operations, we propose two thechniques for efficient computations. One of the techniques is the effective order in neighborhood search, and the other is a bounding technique to skip the search for apparently bad solution, by the preprocessing phase with approximate MFE. In the computational experiments, we succeeded in designing better sequence sets than the existing methods in the case where the sizes of sets are relatively small. Also for larger sets, our method is easy to be combined with non-local search methods such as [9].

As future work, further reduction of computational time is considered. For example, recalculation of MFE values for a new solution is time-consuming, but it can be reduced because the difference between the new and the previous solutions is very small; many internal calculations for MFE values can be reused. Applying our method to more complicated constraints, such as hairpin state machine [10] which have properties of "sequence set design" and "reverse folding problem", is another interesting issue.

References

1. Adleman, L.: Molecular Computation of Solutions to Combinatorial Problems. Science 266(5187), 1021–1024 (1994)
2. Andronescu, M., Zhang, Z., Condon, A.: Secondary Structure Prediction of Interacting RNA Molecules. J. of Molecular Biology 345(5), 987–1001 (2005), `www.rnasoft.ca/download.html`
3. Arita, M., Kobayashi, S.: DNA Sequence Design Using Templates. New Generation Computing 20(3), 263–273 (2002)
4. Asahiro, Y.: Simple Greedy Methods for DNA Word Design. Proc. of 9th World Multi-Conference on Systemics, Cybernetics and Informatics 3, 186–191 (2005)

5. Braich, R., Chelyapov, N., Johnson, C., Rothemund, P., Adleman, L.: Solution of a 20-Variable 3-SAT Problem on a DNA Computer. Science 296(5567), 499–502 (2002)
6. Deaton, R., Kim, J., Chen, J.: Design and test of noncrosshybridizing oligonucleotide building blocks for DNA computers and nanostructures. Applied Physics Letters 82(8), 1305–1307 (2003)
7. Dirks, R., Lin, M., Winfree, E., Pierce, N.: Paradigms for computational nucleic acid design. Nucleic Acids Research 32(4), 1392–1403 (2004)
8. Faulhammer, D., Cukras, A., Lipton, R., Landweber, L.: Molecular computation: RNA solutions to chess problems. Proc. of the National Academy of Sciences of the United States of America 97(4), 1385–1389 (2000)
9. Garzon, M., Phan, V., Roy, S., Neel, A.: In Search of Optimal Codes for DNA Computing. In: Mao, C., Yokomori, T. (eds.) DNA Computing. LNCS, vol. 4287, pp. 143–156. Springer, Heidelberg (2006)
10. Kameda, A., Yamamoto, M., Ohuchi, A., Yaegashi, S., Hagiya, M.: Unravel Four Hairpins! In: Mao, C., Yokomori, T. (eds.) DNA Computing. LNCS, vol. 4287, pp. 381–392. Springer, Heidelberg (2006)
11. Kawshimo, S., Ono, H., Sadakane, K., Yamashita, M.: DNA Sequence Design by Dynamic Neighborhood Searches. In: Mao, C., Yokomori, T. (eds.) DNA Computing. LNCS, vol. 4287, pp. 157–171. Springer, Heidelberg (2006)
12. Lyngsø, R., Zuker, M., Pedersen, C.: Fast evaluation of internal loops in RNA secondary structure prediction. Bioinfomatics 15, 440–445 (1999)
13. Penchovsky, R., Ackermann, J.: DNA Library Design for Molecular Computation. J. of Computational Biology 10(2), 215–229 (2003)
14. Reif, J., Sahu, S., Yin, P.: Complexity of Graph Self-assembly in Accretive Systems and Self-destructible Systems. In: Carbone, A., Pierce, N.A. (eds.) DNA Computing. LNCS, vol. 3892, pp. 257–274. Springer, Heidelberg (2006)
15. Rose, J., Deaton, R., Suyama, A.: Statistical thermodynamic analysis and design of DNA-based computers. Natural Computing 3, 443–459 (2004)
16. Shiozaki, M., Ono, H., Sadakane, K., Yamashita, M.: A Probabilistic Model of DNA Conformational Change. In: Mao, C., Yokomori, T. (eds.) DNA Computing. LNCS, vol. 4287, pp. 274–285. Springer, Heidelberg (2006)
17. Shorteed, M., Chang, S., Hong, D., Phillips, M., Campion, B., Tulpan, D., Andronescu, M., Condon, A., Hoos, H., Smith, L.: A thermodynamic approach to designing struct-free combinatorial DNA word set. Nucleic Acids Research 33(15), 4965–4977 (2005)
18. Tanaka, F., Kameda, A., Yamamoto, M., Ohuchi, A.: Design of nucleic acid sequences for DNA computing based on a thermodynamic approach. Nucleic Acids Research 33(3), 903–911 (2005)
19. Tulpan, D., Hoos, H., Condon, A.: Stochastic Local Search Algorithms for DNA Word Design. In: Hagiya, M., Ohuchi, A. (eds.) DNA Computing. LNCS, vol. 2568, pp. 229–241. Springer, Heidelberg (2003)
20. Tulpan, D., Hoos, H.: Hybrid Randomized Neighborhoods Improve Stochastic Local Search for DNA Code Design. In: Xiang, Y., Chaib-draa, B. (eds.) Advances in Artificial Intelligence. LNCS (LNAI), vol. 2671, pp. 418–433. Springer, Heidelberg (2003)
21. Tulpan, D., Andronescu, M., Changf, S., Shortreed, M., Condon, A., Hoos, H., Smith, L.: Thermodynamically based DNA strand design. Nucleic Acids Research 33(15), 4951–4964 (2005)
22. Winfree, E., Liu, F., Wenzler, L., Seeman, N.: Design and self-assembly of DNA crystals. Nature 394, 539–544 (1998)
23. Zuker, M., Stiegler, P.: Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. Nucleic Acids Research 9, 133–148 (1981)