Fast parallel bio-molecular solutions: the set-basis problem

Weng-Long Chang*

Department of Computer Science and Information Engineering,
National Kaohsiung University of Applied Sciences,
415 Chien Kung Road, 807 Kaohsiung, Taiwan, ROC
E-mail: changwl@cc.kuas.edu.tw
*Corresponding author

Michael Ho

Department of Information Management,
School of Information Technology,
Ming Chuan University,
5, Teh-Ming Rd., Gwei-Shan, 333 Taoyuan, Taiwan, ROC
E-mail: mhoincerritos@yahoo.com

Minyi Guo

School of Computer Science and Engineering,
University of Aizu, Aizu Wakamatsu City, 965 8580 Fukushima, Japan
E-mail: minyi@u-aizu.ac.jp

Chengfei Liu

Faculty of Information and Communication Technologies,
Swinburne University of Technology,
Melbourne, 3122 VIC, Australia
E-mail: cliu@swin.edu.au

Abstract: In the paper, it is demonstrated how to apply sticker in the sticker-based model for constructing solution space of DNA for the setbasis problem and how to apply DNA operations in the Adleman-Lipton model to solve that problem from solution space of sticker. Furthermore, this work shows the ability of DNA-based computing for resolving the NP-complete problems.

Keywords: biological parallel computing; DNA-based supercomputer; NP-complete problem; set-basis problem.

Reference to this paper should be made as follows: Chang, W-L., Ho, M., Guo, M. and Liu, C. (2006) 'Fast parallel bio-molecular solutions: the set-basis problem', *Int. J. Computational Science and Engineering*, Vol. 2, Nos. 1/2, pp.72–80.

Biographical notes: Weng-Long Chang received his PhD Degree in Computer Science and Information Engineering from the National Cheng Kung University, Taiwan in 1999. His research interests include molecular computing, and languages and compilers for parallel computing.

Michael Ho is an Associate Professor of Ming Chuan University with 25 years industrial/academic experiences in the computing field. He had worked as a Senior Software Engineer and Project Leader developing B2B/B2C/C2C web/multimedia applications and a Senior Database Administrator for SQL/Oracle/DB2 with many major US corporation LAN/WAN systems. He had ten years of college teaching/research experiences as an Associate/Assistant Professor at Southern Taiwan University of Technology/Central Missouri State University/the University of Texas. He earned a PhD in IS/CS with management /accounting minors from UT Austin. His research interests include computation theories, software engineering, multimedia, data mining, SOC, parallel/quantum/DNA computing.

Minyi Guo received his PhD Degree in Information Science from the University of Tsukuba, Japan in 1998. From 1998 to 2000, He had been a research scientist of NEC Soft, Ltd. Japan. He is currently an Associate Professor at the Department of Computer Software, The University of Aizu, Japan. He has served as General Chair, Program Committee and Organising Committee Chair for many international conferences. He is an Editor in Chief of the *Journal of Embedded Systems*. He is also on the editorial board of the *International Journal of High Performance Computing and Networking*, *Journal of Embedded Computing*, *Journal of Parallel and Distributed Scientific and Engineering Computing*, and *International Journal of Computer and Applications*. His research interests include parallel and distributed processing, parallelising compilers, data parallel languages, data mining, molecular computing and software engineering.

Chengfei Liu received his PhD Degree in Computer Science from the Nanjing University in 1988. He is currently an Associate Professor at the Faculty of Information and Communication Technologies, Swinburne University of Technology, Australia. He was a Senior Lecturer at the University of South Australia, a Lecturer at the University of Technology Sydney, and a Senior Research Scientist at DSTC in University of Queensland. He also held visiting positions at the IBM Silicon Valley Laboratory and the University of Aizu. His current research interests include advanced databases, XML data management and integration, advanced transaction models, workflows, and Web services.

1 Introduction

Nowadays, producing roughly 10^{18} DNA strands, that too in a test tube, through advances in molecular biology is possible (Sinden, 1994). Basic biological operations can be applied to simultaneously operate 10^{18} bits of information. This is to say that there are 10^{18} data processors to be parallelly executed. Hence, it is very clear that biological computing can provide a very similar parallelism for dealing with the problem in the real world.

Adleman (1994) wrote the first paper that DNA strands could be used to deal with solutions for an instance of the NP-complete Hamiltonian path problem (HPP). Lipton (1995) wrote the second paper that demonstrated that the Adleman techniques could be used to solve the NP-complete satisfiability (SAT) problem (the first NP-complete problem). Adleman and his coauthors (Roweis et al., 1999) proposed sticker for enhancing the Adleman-Lipton model.

In this paper, we use a sticker in the sticker based model for constructing a solution space of DNA for the setbasis problem. Simultaneously, we also apply DNA operations in the Adleman-Lipton model to develop a DNA algorithm. It is shown from the main result of the proposed DNA algorithm that the setbasis problem is resolved with biological operations in the Adleman-Lipton model from the solution space of the sticker. Furthermore, this work shows the ability of DNAbased computing for resolving the NP-complete problems.

The rest of this paper is organised as follows. In Section 2, the Adleman-Lipton model is introduced in detail and the comparison of the model with other models is given. Section 3 introduces a DNA algorithm for solving the setbasis problem from the solution space of the sticker in the Adleman-Lipton model. In Section 4, the experimental result of simulated DNA computing is discussed. Conclusions are drawn in Section 5.

2 DNA model of computation

In Subsection 2.1, a summary of DNA structure is given and the Adleman-Lipton model is described in detail. In Subsection 2.2, a comparison of the Adleman-Lipton model with other models is given.

2.1 The Adleman-Lipton model

A DNA (DeoxyriboNucleic Acid) is the molecule that plays the main role in DNA based computing (Paun et al., 1998). In the biochemical world of large and small molecules, polymers, and monomers, DNA is a polymer, which is strung together from monomers called deoxyribonucleotides. The monomers used for the construction of DNA are deoxyribonucleotides. Each deoxyribonucleotide contains three components: a sugar, a phosphate group, and a nitrogenous base. The sugar has five carbon atoms – for the sake of reference, there is a fixed numbering for them. Because the base also has carbons, to avoid confusion, the carbons of the sugar are numbered from 1' to 5' (rather than from 1 to 5). The phosphate group is attached to the 5' carbon, and the base is attached to the 1' carbon. Within the sugar structure there is a hydroxyl group attached to the 3' carbon.

Distinct nucleotides are detected only with their bases, which come in two sorts: purines and pyrimidines (Sinden, 1994; Paun et al., 1998). Purines include adenine and guanine, abbreviated *A* and *G*. Pyrimidines contain cytosine and thymine, abbreviated *C* and *T*. Because nucleotides are only distinguished from their bases, they are simply represented as *A*, *G*, *C*, or *T* nucleotides, depending upon the sort of base that they have. The structure of a nucleotide, cited from (Paun et al., 1998), is illustrated (in a very simplified way) in Figure 1. In Figure 1, B is one of the four possible bases (*A*, *G*, *C*, or *T*), *P* is the phosphate group, and the rest (the 'stick') is the sugar base (with its carbons enumerated 1' through 5').

Table 3 Sequences chosen to represent the two elements in *S* in Figure 2

Vertex	*5′ → 3′ DNA sequence*
x_2^0	CCTACCTCTCACCTT
x_1^0	*CCACATATCCATCCC*
x_2^1	*CATTACCTCTTTACT*
x_1^1	*CCCATCTTTCTTAAC*

Table 4 The energy for the binding of each probe to its corresponding region on a library strand

Vertex	*Enthalpy energy* (H)	*Entropy energy* (S)	*Free energy* (G)
x_2^0	109.3	278.4	26.2
x_1^0	110.9	278	28
x_2^1	106.7	279.7	23
x_1^1	112.1	288.8	25.8

The program simulates a mix and split combinatorial synthesis technique (Cukras et al., 1998) to synthesise the library strand to every possible subset. Those library strands are shown in Table 5 and, correspondingly, represent four possible subsets: φ, {1}, {2}, and {1, 2}. The program is also applied to figure out the average and standard deviation for the enthalpy, entropy and free energy over all probe/library strand interactions. The energy is shown in Table 6. The standard deviation for delta *G* is small because this is partially enforced by the constraint that there are 4, 5, or 6 Gs (the seventh constraint in Subsection 3.2) in the probe sequences.

Table 5 DNA sequences chosen represent all possible subsets

5' – CCT ACCTCTC ACCTTCC AC AT ATCC ATCCC – 3'
3' – GGATGGAGAGTGGAAGGTGTATAGGTAGGG –5'
5' – CCT ACCTCTC ACCTTCCC ATCTTTCTT AAC – 3'
3' – GGATGGAGAGTGGAAGGGTAGAAAGAATTG – 5'
5' – CATTACCTCTTTACTCC AC AT ATCC ATCCC – 3'
3' – GT AATGGAGAAATGAGGTGT AT AGGT AGGG – 5'
5' – CATTACCTCTTTACTCCC ATCTTTCTT AAC – 3'
3' – GTAATGGAGAAATGAGGGTAGAAAGAATTG – 5'

Table 6 The energy over all probe/library strand interactions

	Enthalpy energy (*H*)	*Entropy energy* (*S*)	*Free energy* (*G*)
Average	109.75	281.225	25.75
Standard deviation	2.02161	4.41807	1.79095

The Adleman program was employed for computing the distribution of the types of potential mishybridisations. The distribution of the types of potential mishybridisations is the absolute frequency of a probestrand match of length *k* from 0 to the bit length 15 (for DNA sequences) where probes are not supposed to match the strands. The distribution was, subsequently, 52, 102, 85, 95, 105, 109, 77, 36, 13, 12, 2, 0, 0, 0, 0 and 0. It is pointed out from the last five zeros that there are 0 occurrences where a probe matches a strand at 11, 12, 13, 14 or 15 places. This shows that the third constraint in subsection 3.2 has been satisfied. It is very clear that the number of matches peaks at five (109). That is to say that there are 109 occurrences where a probe matches a strand at five places.

It is indicated from the execution of Step 2 of simulation that the result generated by Step 2 was shown in Table 7. The goal of Step 3 is to find a set-basis from the result generated by Step 2. Hence, Step 3(a) of simulation, the set-basis was shown in Table 8. That is to say that the answer of the set-basis problem for the finite set *S* and the collection *C* in Figure 2 is {{1}, {2}}.

Table 7 DNA sequences generated by Step 2 represent legal subsets

5' – CCT ACCTCTC ACCTTCCC ATCTTTCTT AAC – 3'
5' – CATTACCTCTTTACTCC AC AT ATCC ATCCC – 3'

Table 8 DNA sequence represents the answer of the set-basis problem for the finite set *S* and the collection *C* in Figure 2

5' – CCTACCTCTCACCTTCCCATCTTTCTTAAC – 3'
5' – CATTACCTCTTTACTCCACATATCCATCCC – 3'

5 Conclusions

Applying splints constructs the solution space of the DNA sequence for solving the NPcomplete problem in the Adleman-Lipton and this is the reason that hybridisation has higher probabilities for errors. Adleman and his coauthors (Roweis et al., 1999) proposed a sticker to decrease probabilities of errors in hybridisation in the Adleman-Lipton model. In the proposed algorithm, the size of the solution space of the sticker is exponential. Hence, this is the limit to which we can resolve the size of the NPcomplete problem. The main result of the proposed algorithm shows that the set-basis problem is resolved with biological operations in the Adleman-Lipton model from solution space of sticker. Furthermore, this work demonstrates the ability of DNA based computing to solve NPcomplete problems.

References

Adleman, L.M. (1996) 'On constructing a molecular computer. DNA based computers', *DIMACS: Series in Discrete Mathematics and Theoretical Computer Science*, American Mathematical Society, pp.1–21, in the First Annual Meeting on DNA-based Computers.

Adleman, V. (1994) 'Molecular computation of solutions to combinatorial problems', *Science*, Vol. 266, 11th November, pp.1021–1024.

Amos, M. (1997) *DNA Computation*, PhD Thesis, Department of Computer Science, the University of Warwick, pp.29–38.

Arita, M., Suyama, A. and Hagiya, M. (1997) 'A heuristic approach for Hamiltonian path problem with molecules', *Proceedings of 2nd Genetic Programming (GP-97)*, pp.457–462.

Boneh, D., Dunworth, C., Lipton, R.J. and Sgall, J. (1996) 'On the computational power of DNA. In discrete applied mathematics, *Special Issue on Computational Molecular Biology*, Vol. 71, pp.79–94.

Braich, R.S., Johnson, C., Rothemund, P.W.K., Hwang, D., Chelyapov, N. and Adleman, L.M. (2000) 'Solution of a satisfiability problem on a gel-based DNA computer', *Proceedings of the 6th International Conference on DNA Computation in the Springer-Verlag Lecture Notes in Computer Science Series*, pp.27–42.

Chang, W-L. and Guo, M. (2002a) 'Solving the dominating-set problem in Adleman-Lipton's Model', *The Third International Conference on Parallel and Distributed Computing, Applications and Technologies*, Japan, pp.167–172.

Chang, W-L. and Guo, M. (2002b) 'Solving the Clique problem and the vertex cover problem in Adleman-Lipton's Model', *IASTED International Conference, Networks, Parallel and Distributed Processing, and Applications*, Japan, pp.431–436.

Chang, W-L. and Guo, M. (2002c) 'Solving NP-complete problem in the Adleman-Lipton Model', *The Proceedings of 2002 International Conference on Computer and Information Technology*, Japan, pp.157–162.

Chang, W-L. and Guo, M. (2002d) 'Resolving the 3-dimensional matching problem and the set packing problem in Adleman-Lipton's Model', *IASTED International Conference, Networks, Parallel and Distributed Processing, and Applications*, Japan, pp.455–460.

Guo, M., Ho, M. and Chang, W-L. (2004) 'Fast parallel molecular solution to the dominating-set problem on massively parallel bio-computing', *Parallel Computing*, Vol. 30, Nos. 9/10, September–October, pp.1109–1125.

Cormen, T.H., Leiserson, C.E. and Rivest, R.L. (1990) *Introduction to Algorithms*, MIT Press, Cambridge, Massachusetts.

Cukras, A.R., Faulhammer, D., Lipton, R.J. and Landweber, L.F. (1998) 'Chess games: a model for RNA-based computation', *Proceedings of the 4th DIMACS Meeting on DNA Based Computers*, held at the University of Pennsylvania, 16–19 June, pp.27–37.

Fu, B. (1997) '*Volume Bounded Molecular Computation*, *PhD Thesis*, Department of Computer Science, Yale University, http://fano.ics.uci.edu/cites/Document/Volume-Bounded-Molecular-Computation.html.

Garey, M.R. and Johnson, D.S. (1979) *Computer and Intractability*, Freeman, San Francisco, CA.

KalimMir (1996) 'A restricted genetic alphabet for DNA computing', in Baum, E.B. and Landweber, L.F. (Eds.): *DNA Based Computers II: DIMACS Workshop*, 10–12 June, *Vol.44 of DIMACS: Series in Discrete Mathematics and Theoretical Computer Science*, Providence, RI, 1998, pp.243–246.

Lipton, R.J. (1995) 'DNA solution of hard computational problems', *Science*, Vol. 268, pp.542–545.

Morimoto, N., Arita, M. and Suyama, A. (1999) 'Solid phase DNA solution to the Hamiltonian path problem', *DIMACS (Series in Discrete Mathematics and Theoretical Computer Science)*, Vol. 48, pp.93–206.

Narayanan, A. and Zorbala, S. (1998) 'DNA algorithms for computing shortest paths', in Koza, J.R. *et al*. (Eds.): *Genetic Programming 1998: Proceedings of the Third Annual Conference*, pp.718–724.

Paun, G., Rozenberg, G. and Salomaa, A. (1998) *DNA Computing: New Computing Paradigms*, Springer-Verlag, New York, ISBN: 3-540-64196-3.

Perez-Jimenez, M.J. and Sancho-Caparrini, F. (2001) 'Solving Knapsack problems in a sticker based model', *7th Annual Workshop on DNA Computing, DIMACS: Series in Discrete Mathematics and Theoretical Computer Science*, American Mathematical Society.

Quyang, Q., Kaplan, P.D., Liu, S. and Libchaber, A. (1997) 'DNA solution of the maximal clique problem', *Science*, Vol. 278, pp.446–449.

Roweis, S., Winfree, E., Burgoyne, R., Chelyapov, N.V., Goodman, M.F., Rothemund, P.W.K. and Adleman, L.M. (1999) 'A sticker based model for DNA computation', in Landweber, L. and Baum, E. (Eds.): *2nd Annual Workshop on DNA Computing*, Princeton University, *DIMACS: Series in Discrete Mathematics and Theoretical Computer Science*, American Mathematical Society, pp.1–29.

Shin, S-Y., Zhang, B-T. and Jun, S-S. (1999) 'Solving traveling salesman problems using molecular programming', *Proceedings of the 1999 Congress on Evolutionary Computation (CEC99)*, Vol. 2, pp.994–1000.

Sinden, R.R. (1994) *DNA Structure and Function*, Academic Press, ISBN 0126457506.