Maze Exploration with Molecular-Scale Walkers
Extended Abstract

Darko Stefanovic[1,2]

[1] Department of Computer Science, University of New Mexico
[2] Center for Biomedical Engineering, University of New Mexico
Albuquerque, NM 87131, United States

Abstract. Molecular spiders are nanoscale walkers made with catalytic DNA legs attached to a rigid body. They move in a matrix of DNA substrates, cleaving them and leaving behind product DNA strands. Unlike a self-avoiding walker, a spider is able to revisit the products. However, the legs cleave and detach from substrates more slowly than they detach from products. This difference in residence time and the presence of multiple legs make a spider move differently from an ordinary random walker. The number of legs, and their lengths, can be varied, and this defines the spider's local gait, which affects its behavior in global tasks. In this work we define an abstract model of molecular spiders, and within it we study the efficiency of maze exploration as a function of the spider structure. For a fixed geometry, there is an optimal setting of chemical kinetics parameters that minimizes the mean time to traverse a maze.

Keywords: molecular walkers, maze search, DNA computing

1 Introduction

The tasks of spatial search and exploration are among the hallmarks of intelligent behavior. Among such tasks, exploring mazes or labyrinths has a glorious history, from the myth of Theseus to the recent maze explorations by physarum [1] and fungi [2]. But, is it possible for even simpler, not quite living agents, also to explore mazes?

The agents we will use are synthetic nanoscale walkers called molecular spiders [3, 4], which move using a mechanism of multivalent chemical interactions of their multiple legs with the environment (Section 2), in which the legs catalytically convert substrates to products, thereby extracting chemical energy from the environment. Molecular spiders may find use in biomedical applications, such as searching for clinically relevant targets on the surface of a cell; exploring mazes is an abstraction of such walks in unknown environments with obstacles. Ahead of constructing mazes in the laboratory, here we present simulation-based results on the speed of maze exploration by molecular spiders. Spiders and their targets are simulated on a two-dimensional grid of chemical sites (Section 3), corresponding to the DNA origami [5, 6] on which our mazes will be self-assembled.

Previous studies have shown that, at a sufficiently high level of abstraction, the main kinetic parameter governing the walking behavior of a molecular spider is the ratio r between the residence time of a spider's leg on a previously visited site (i.e.,

a spent product) and the residence time on a new site (i.e., a fresh substrate) [7–10]. When $r = 1$, ordinary diffusion ensues; when $r < 1$, the motion exhibits an increasingly strong superdiffusive transient as r decreases. Here we find that the chemical kinetics influences the efficiency of maze exploration as well. The dependence is strong: for two-legged walkers in our test maze, a clear optimum r value emerges at around 1:100 product-to-substrate residence time ratio (Section 4). This behavior is a consequence of the interplay between the multivalency and the kinetic bias (Section 5): if catalysis is absent ($r = 1$), two-legged walkers are slower than one-legged spiders (ordinary random walkers); with catalysis, one-legged spiders are only ever slowed down by the longer residence time on substrates, whereas two-legged spiders are faster for a range of values $r < 1$ thanks to the emergent bias towards substrates [8, 9]. This bias may be useful in a maze: a spider emerging for the first time from a dead end faces a T-junction, at which it will be biased towards the unexplored path rather than the path whence it originally came.

2 Molecular Spiders

Translational molecular motors, made of proteins, are nature's solution to the problem of efficient molecular cargo transport across the large diameter of a cell. They walk directionally along self-assembled, directional filaments and microtubules [11–17]. Recent advances in single-molecule chemistry have led to the development of *synthetic* molecular motors [18, 19], including molecular assemblies that walk over surfaces, following fabricated or self-assembled tracks [4, 20–27]. Among them are *molecular spiders*, autonomous synthetic molecular motors based on catalytic DNA [3, 4].

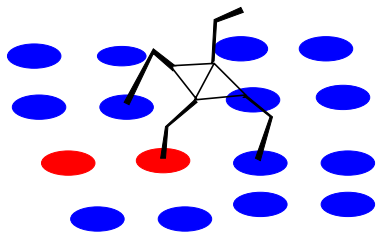

Fig. 1: Molecular spider walking across a substrate-laden surface: "blue" substrates are transformed into "red" products.

A molecular spider has a rigid inert body and several flexible enzymatic legs (Figure 1). We have made spiders with up to six legs, using a streptavidin or streptavidin dimer scaffold for the body [3]. Each leg is a deoxyribozyme—an enzymatic sequence of single-stranded DNA that can bind to and cleave a complementary strand of a DNA substrate [28]. The hip joint between the body and a leg is a flexible biotin linkage. When a molecular spider is placed on a surface coated with the single-stranded DNA

substrate, its legs bind to the substrate. A bound leg can either detach from the substrate without modifying it, or it can catalyze the cleavage of the substrate, creating two product strands. Upon cleavage the two product strands eventually dissociate from the enzyme leg. The lower product remains bound to the surface. Because the lower product is complementary to the lower part of the spider's leg, there is a residual binding of the leg to the product, i.e., it is possible for a spider's legs to visit products; however, this binding is typically much weaker than the leg-substrate binding and thus much shorter-lived.

The small scale of the spiders makes their direct observation difficult and necessitates detailed modelling work, which several groups have undertaken through mathematical analysis and computer-based simulations [7–10, 29–31]. The motion of a spiders can be described as a special type of random walk. Their asymptotic behavior is diffusive, just as with ordinary random walkers. Surprisingly, however, superdiffusive behavior is observed in the transient, and it lasts for significant amounts of time, over which a spider covers significant distances. This has profound implications for practical applications, including for maze exploration. We showed [9] that the residence-time bias between substrates and products causes this behavior through a micromechanism of the spider's switching between two states—being on the boundary of the area of fresh substrates, and being in the area of already cleaved products. A spider on the boundary extracts chemical energy from the landscape, moves preferentially towards fresh substrates, and thus carries the boundary along. A spider that has left the boundary and stepped back into the products simply diffuses. Eventually this behavior dominates.

By varying the spiders' geometry and chemical characteristics, we can optimize how they walk. The next question is what they can do—for instance, they may transport molecular cargo, search and capture targets, or follow predesigned tracks. Here we look at spiders following complex branching tracks, i.e., mazes. We describe models of future experiments in which spiders released at one end of a maze (self-assembled using DNA origami techniques) race to the goal at the opposite end. We are interested in the end-to-end time. Alternatively, we could ask how much time it takes to explore the entire maze—we defer this question to future work.

3 Abstract Model of Molecular Spider Motion for Maze Exploration

Spider Motion. In our model, a k-legged spider walks on a square lattice of chemical sites. At all times, all legs are attached to sites, i.e., the reattachment process is complete and infinitely fast—this is expedient, but a gross simplification that ignores the kinetics of reattachment, which we treat elsewhere [10]. No more than one leg can be attached to a site at a time (complete exclusion). A leg detaches from a site according to a Poisson process, i.e., according to first-order chemical kinetics that describe the dissociation of the leg DNA strand and the product DNA strand. For a leg visiting a substrate, we also fold the kinetics of catalysis into a single detachment rate. Thus, a leg detaches at rate 1 from a product, and at rate r from a substrate. When a leg detaches from a substrate, that substrate is transformed into a product. Once detached, the leg immediately attaches to

another site. The new site is chosen at random from among a set of feasible locations, and all of those are taken as equally probable.

The model admits a variety of options for the set of feasible new locations. When a physical spider's leg moves, it remains attached to the body, and while the body can move, it can only do so within the constraints imposed by the remaining legs, still attached to the surface. This can be expressed in the model as a constraint that the new locations must be within a certain distance S from each remaining leg's position. Furthermore, either the kinetics of binding to a new site or the diffusion needed to reach it can be the limiting factor for reattachment, and to express the latter case the model can restrict the feasible new sites to be within a certain distance R from the old site (the site from which the leg detached). A physical spider's leg can always reattach to the site it just left, but in a model such unproductive steps can be ruled out to expedite simulations. In this paper we take $k = 2$, $S = 2$, and $R = 1$, measuring distances using the Euclidean metric. In other words, we consider two-legged spiders with nearest-neighbor hopping.

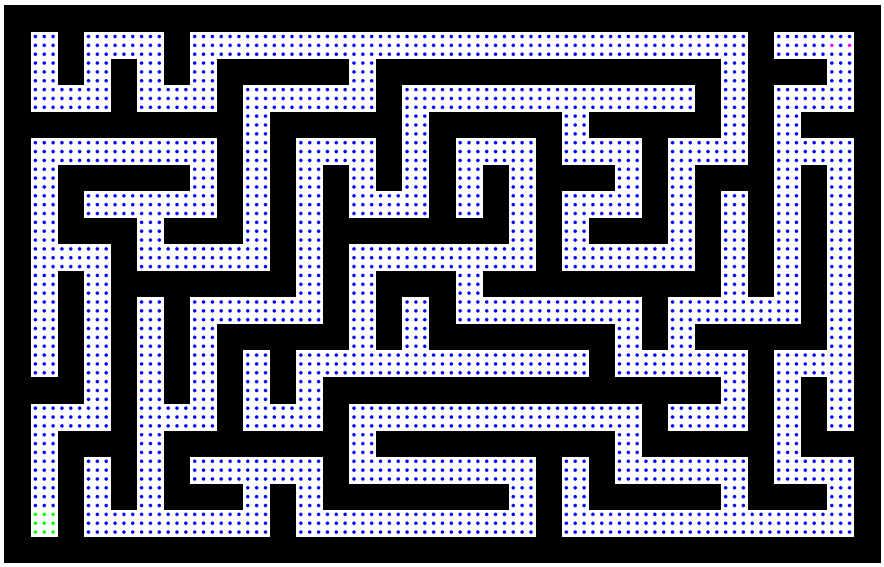

Fig. 2: The maze used for the exploration task. Black walls are areas on the DNA origami tile without exposed single-stranded DNA, and assumed inaccessible to the spider. Each blue dot is a chemical site, initially a substrate. The spider starts in the top right corner with its two legs at the two purple sites. Its goal is the bottom left corner, where it can be trapped by uncleavable substrates, shown in green. The shortest path to the goal is of length 197 maze tiles, or 591 inter-substrate spaces.

Maze Exploration. Onto a grid of chemical substrates we superimpose a maze grid with a pitch of three, i.e., the tracks in the maze are three substrates wide, and the

walls of the maze are also three sites wide, but empty. We generate random, tile-based (with walls of finite thickness), perfect (loop-free) mazes. All results below pertain to one particular randomly generated maze, shown in Figure 2. The size of this maze is chosen to correspond to realistic mazes we can build in the future. We will use the new generation of DNA origami,[3] which can measure 200 nm by 300 nm. With three-substrate-wide walking tracks and walls, the maze grid can be 21 by 31, large enough for non-trivial mazes.

4 Simulation Results

Together the spider state and the state of the entire surface represent the state of the continuous-time Markov process for our model. We use the Kinetic Monte Carlo method [32] to simulate multiple trajectories of this Markov process. The simulation stops when the spider is trapped at the goal, and we record the simulated time as an observation of a first-passage-time random variable, viz., the time it took the spider to cross the maze from the start corner to the goal corner. The results below will show the mean first passage time, as well as the probability distribution of first passage times, estimated from our traces.

The space of possible spider configurations is large. Here we only use the canonical two-legged spider with nearest-neighbor hopping, i.e., $k = 2, S = 2, R = 1$, and a one-legged walker (normally we do not call such walkers spiders).

Before we examine the statistics, we illustrate one random trajectory of a two-legged spider on the test maze. Figure 3 shows how many times each chemical site was visited. The start and the large, branched dead end in the middle of the maze were the most heavily visited areas.

In Figure 4 we show the mean time to the goal for two-legged spiders as a function of the kinetic rate r. We vary the kinetic parameter r over 12 orders of magnitude. At $r = 1$ there is no distinction between substrates and products, and the motion degenerates into a bipedal random walk [7]. Reducing r, we introduce catalysis and make substrates "stickier", so a leg spends more time visiting a substrate. Although this may at first seem counterintuitive, this improves the overall performance, up to a critical value of $r \approx 0.01$. This is in agreement with our previous observations of the superdiffusive transient in 1D spiders [9]. At $r \approx 0.01$, the shortest mean time is obtained: if we are free to design the chemical kinetics of the legs for a fixed two-legged geometry, we should aim for this ratio. Below 0.01, the time is dominated by new substrate visits and the mean time to the goal scales as $1/x$. Usually it is assumed that $r \leqslant 1$, but it is possible (for spiders, or other walkers with different chemical structure) to have the products be stickier than substrates, $r > 1$. We simulated this case as well. Performance worsens as r is increased above 1, but it approaches an asymptote; as $r \to \infty$, substrates are immediately converted to products, and this has the effect of repelling the walker backwards. In the range $0.01 < r < 10$, we empirically estimate that the mean time to the goal scales as $r^{1/4}$; this dependence calls for further analytical studies.

To explore the utility of spiders as multivalent random walkers, we compare them with one-legged walkers. In Figure 5, the performance of one-legged spiders uniformly

[3] Thomas LaBean, personal communication, 2012.

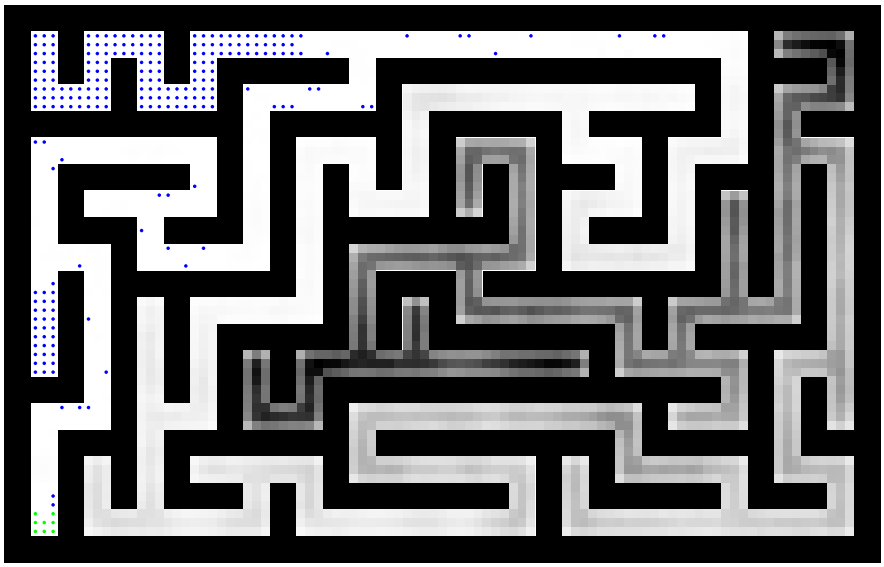

Fig. 3: A typical exploration trajectory. The number of visits to a site is shown by shading: darker sites were visited more times. Some unvisited substrates remain (blue dots).

improves as r increases; this is similar to the results of the study of one-legged spiders in one dimension [8]. Thus, the presence of memory on the surface in the form of kinetically differing substrates and products does not improve the performance of a monovalent random walker. The plots answer a likely design question: if we have perfected a design for the legs and therefore fixed their chemical kinetics r, how should we decide how many legs to attach to the spider body? If the choice is between one and two legs, a two-legged design is advantageous whenever $r \lesssim 0.1$. The magnitude of the advantage is shown in Figure 6.

4.1 Search time distribution

The mean first passage times to the goal tell only one part of the story. The complete probability distribution of first passage times displays high variance. For two-legged spiders, and one particular value $r = 0.01$, we show a histogram of the distribution in Figure 7. The mode of the distribution is well below its mean. Similar fairly long tails obtain for other parameters. This will be important to keep in mind in the design of applications and laboratory experiments using molecular walkers, whenever the task is of a first-passage-time flavor.

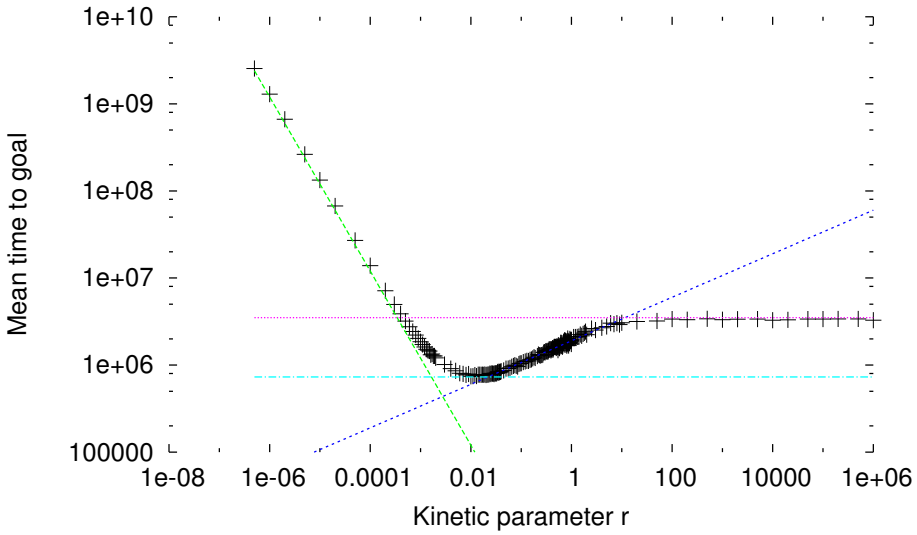

Fig. 4: Mean first passage time to the goal: two-legged spiders ($k = 2$). Note the log-log scale. The lines drawn in are visually estimated asymptotes (for < 0.001, $1200 \times r^{-1}$; for $0.05 < r < 10$, $1.9 \times 10^6 \times r^{0.25}$; and for $r \to \infty$, 3.5×10^6) and the minimum mean first passage time of 7.3×10^5, obtained at $r = 0.01$.

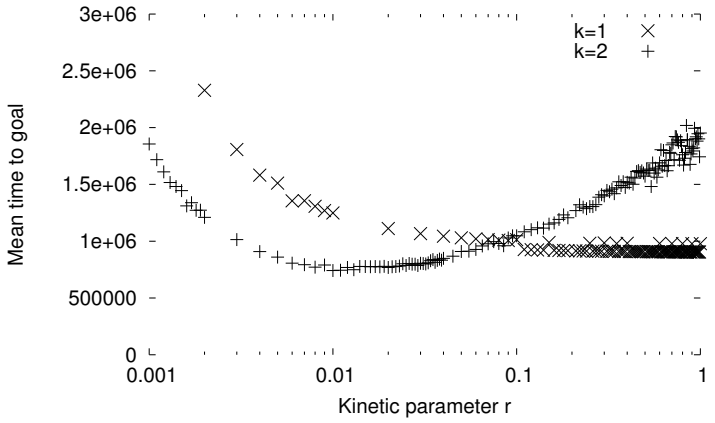

Fig. 5: Mean first passage time to the goal for one- and two-legged spiders, focusing on the (chemically most plausible) range of r values between 0.001 and 1.

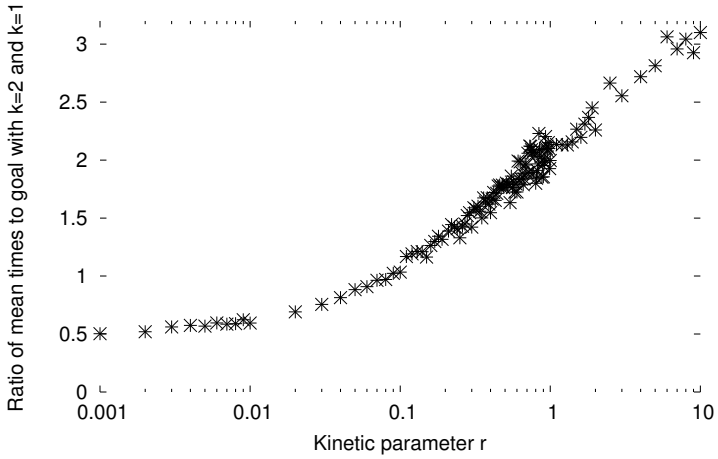

Fig. 6: Advantage of multivalency. Shown is the ratio of mean first passage time for two- and for one-legged spiders.

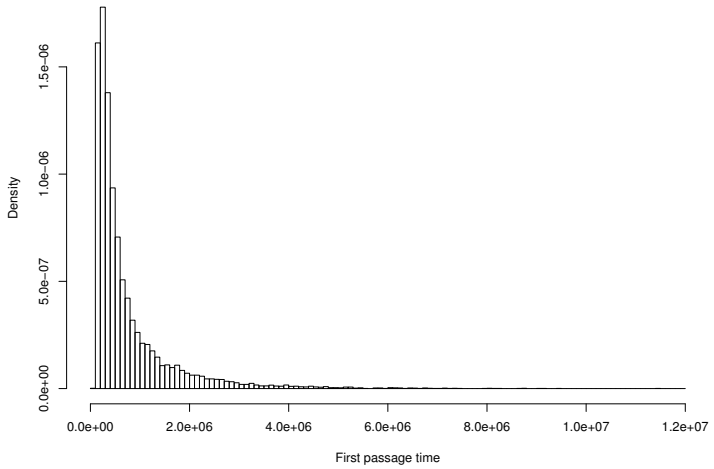

Fig. 7: Probability distribution of first passage times for $k = 2$ and $r = 0.01$. Histogram based on 14395 sample trajectories.

4.2 The effect of the gait

Note added in print. It is possible to improve search times further by changing the structure of the walker. More legs can be added, but the basic model above turns into a shuffling gait, or even blockage, if too many legs are added. One way to resolve the issue is by lengthening the legs. Another way is to use the "quick" spiders model, in which the moving leg samples a space independent of its previous position, and only constrained by the attachment points of the remaining legs—this model is more physically realistic, reflecting the separation of time-scales between the fast physical equilibrium of the spider's body and legs, and the slower chemical processes of DNA hybridization and catalysis. Here we show the outcome with quick spiders having four legs, and leg lengths 1.35, i.e., $k = 4$, $S = 2.7$, $R = \infty$; this configuration corresponds to the laboratory experiments with NICK 4.4A molecular spiders [3]. Now the multivalency is beneficial as soon as $r \lesssim 0.5$, and the optimum r is 0.05. Serendipitously, this value is close to 0.04, which is our estimate of the r parameter from our measurements of the NICK 4.4A spiders.

5 Discussion

The simulations used a fixed and relatively small maze. Its size, of course, was chosen to be physically realistic, but we expect that varying the size will influence the optimizing r values, because a larger or smaller portion of the superdiffusive transient will be exploited. It remains to be seen if, under laboratory conditions, it will be possible

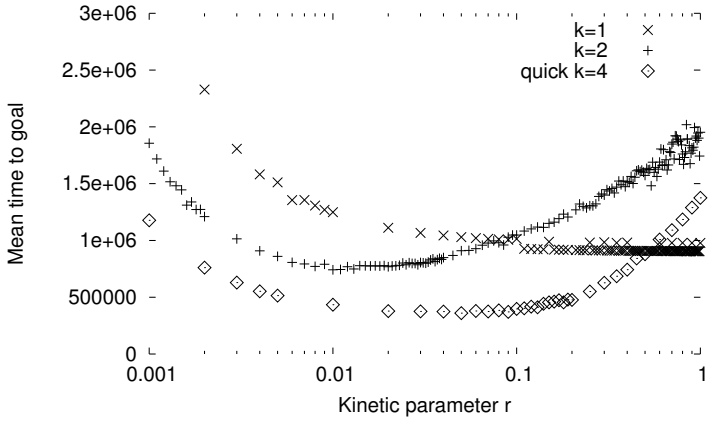

Fig. 8: Searching with four-legged "quick" spiders. (One- and two-legged spiders are left in for comparison.)

to exploit the superdiffusive transient or if, at the scale dictated by such conditions, the motion will be dominated by the diffusive asymptotics. Even in the latter case, however, the effective diffusion coefficient will be governed by the details of the gait; therefore, it will be useful to explore a variety of randomly generated mazes, including non-perfect ones (with loops), and differing track widths, to understand the interplay between maze geometry and the spider kinematic constraints such as k, S, and R.

Elsewhere we have studied the behavior of multiple molecular spiders released at a point source on an infinite 1D lattice [30]. Multiple spiders injected at the start and absorbed at the goal may search more efficiently than a single spider, and we intend to study this scenario as well.

A number of assumptions had to made to reduce the complex interactions of physical molecular spiders on DNA origami tiles, which have not yet been fully experimentally characterized and understood, to a tractable mathematical model amenable to efficient computer simulation and parameter space exploration. We will use finer-grain models [10] to refine these initial results.

Acknowledgments. I am grateful to Paul Krapivsky, Mark Olah, Oleg Semenov, and Milan Stojanovic for many inspirational conversations, and to Steven Taylor for the detailed kinetic characterization of molecular spiders' legs. I thank the conference reviewers for their incisive comments. This material is based upon work supported by the National Science Foundation under grants 0829896 and 1028238.

References

1. Nakagaki, T., Yamada, H., Tóth, A.: Maze-solving by an amoeboid organism. Nature **407** (2000) 470
2. Nicolau, D.V., Nicolau, D.V., Jr., Solana, G., Hanson, K.L., Filipponi, L., Wang, L., Lee, A.P.: Molecular motors-based micro- and nano-biocomputation devices. Microelectronic Engineering **83**(4-9) (2006) 1582–1588
3. Pei, R., Taylor, S.K., Stefanovic, D., Rudchenko, S., Mitchell, T.E., Stojanovic, M.N.: Behavior of polycatalytic assemblies in a substrate-displaying matrix. Journal of the American Chemical Society **128**(39) (2006) 12693–12699
4. Lund, K., Manzo, A.J., Dabby, N., Michelotti, N., Johnson-Buck, A., Nangreave, J., Taylor, S., Pei, R., Stojanovic, M.N., Walter, N.G., Winfree, E., Yan, H.: Molecular robots guided by prescriptive landscapes. Nature **465** (May 2010) 206–210
5. Rothemund, P.W.K.: Folding DNA to create nanoscale shapes and patterns. Nature **440** (2006) 297–302
6. Pinheiro, A.V., Han, D., Shih, W.M., Yan, H.: Challenges and opportunities for structural DNA nanotechnology. Nature Nanotechnology **6** (2011) 763–772
7. Antal, T., Krapivsky, P.L., Mallick, K.: Molecular spiders in one dimension. Journal of Statistical Mechanics: Theory and Experiment **2007**(08) (2007) P08027
8. Antal, T., Krapivsky, P.L.: Molecular spiders with memory. Physical Review E **76**(2) (2007) 021121
9. Semenov, O., Olah, M.J., Stefanovic, D.: Mechanism of diffusive transport in molecular spider models. Physical Review E **83**(2) (February 2011) 021117
10. Olah, M.J., Stefanovic, D.: Multivalent random walkers — a model for deoxyribozyme walkers. In: DNA 17: Proceedings of The Seventeenth International Meeting on DNA Computing and Molecular Programming. Volume 6397 of Lecture Notes in Computer Science., Springer (2011) 160–174
11. Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., Walter, P.: Molecular Biology of The Cell. 4th edn. Garland Science (2002)
12. Kolomeisky, A.B., Fisher, M.E.: Molecular motors: A theorist's perspective. Annual Review of Physical Chemistry **58** (2007) 675–95
13. Bier, M.: The energetics, chemistry, and mechanics of a processive motor protein. BioSystems **93** (2008) 23–28
14. Astumian, R.D.: Thermodynamics and kinetics of molecular motors. Biophysical Journal **98** (June 2010) 2401–2409
15. Jamison, D.K., Driver, J.W., Rogers, A.R., Constantinou, P.E., Diehl, M.R.: Two kinesins transport cargo primarily via the action of one motor: Implications for intracellular transport. Biophysical Journal **99** (November 2010) 2967–2977
16. Lipowsky, R., Beeg, J., Dimova, R., Klumpp, S., Müller, M.K.I.: Cooperative behavior of molecular motors: Cargo transport and traffic phenomena. Physica E **42** (2010) 649–661
17. Driver, J.W., Jamison, D.K., Uppulury, K., Rogers, A.R., Kolomeisky, A.B., Diehl, M.R.: Productive cooperation among processive motors depends inversely on their mechanochemical efficiency. Biophysical Journal **101** (July 2011) 386–395
18. Kay, E.R., Leigh, D.A., Zerbetto, F.: Synthetic molecular motors and mechanical machines. Angew. Chem. Int. Ed. **46** (2007) 72–191
19. Hugel, T., Lumme, C.: Bio-inspired novel design principles for artificial molecular motors. Current Opinion in Biotechnology **21**(5) (2010) 683 – 689
20. Yurke, B., Turberfield, A.J., Mills, Jr., A.P., Simmel, F.C., Neumann, J.L.: A DNA-fuelled molecular machine made of DNA. Nature **406** (August 2000) 605–608

21. Shirai, Y., Osgood, A.J., Zhao, Y., Kelly, K.F., Tour, J.M.: Directional control in thermally driven single-molecule nanocars. Nano Letters **5**(11) (2005) 2330–2334
22. Tian, Y., He, Y., Chen, Y., Yin, P., Mao, C.: A DNAzyme that walks processively and autonomously along a one-dimensional track. Angew. Chem. Int. Ed. **44** (2005) 4355–4358
23. Venkataraman, S., Dirks, R.M., Rothemund, P.W.K., Winfree, E., Pierce, N.A.: An autonomous polymerization motor powered by DNA hybridization. Nature Nanotechnology **2** (2007) 490–494
24. Green, S.J., Bath, J., Turberfield, A.J.: Coordinated chemomechanical cycles: A mechanism for autonomous molecular motion. Physical Review Letters **101** (2008) 238101–+
25. Omabegho, T., Sha, R., Seeman, N.C.: A bipedal DNA brownian motor with coordinated legs. Science **324** (April 2009) 67–71
26. Bath, J., Green, S.J., Allen, K.E., Turberfield, A.J.: Mechanism for a directional, processive, and reversible DNA motor. Small **5**(13) (2009) 1513–1516
27. Gu, H., Chao, J., Xiao, S.J., Seeman, N.C.: A proximity-based programmable DNA nanoscale assembly line. Nature **465** (May 2010) 202–206
28. Santoro, S.W., Joyce, G.F.: A general purpose RNA-cleaving DNA enzyme. Proceedings of the National Academy of Sciences of the USA (PNAS) **94** (April 1997) 4262–4266
29. Samii, L., Linke, H., Zuckermann, M.J., Forde, N.R.: Biased motion and molecular motor properties of bipedal spiders. Physical Review E **81**(2) (Feb 2010) 021106–+
30. Semenov, O., Olah, M.J., Stefanovic, D.: Multiple molecular spiders with a single localized source—the one-dimensional case. In: DNA 17: Proceedings of The Seventeenth International Meeting on DNA Computing and Molecular Programming. Volume 6397 of Lecture Notes in Computer Science., Springer (2011) 204–216
31. Samii, L., Blab, G.A., Bromley, E.H.C., Linke, H., Curmi, P.M.G., Zuckermann, M.J., Forde, N.R.: Time-dependent motor properties of multipedal molecular spiders. Phys. Rev. E **84** (Sep 2011) 031111
32. Bortz, A.B., Kalos, M.H., Lebowitz, J.L.: A new algorithm for Monte Carlo simulation of Ising spin systems. Journal of Computational Physics **17**(1) (1975) 10–18