Xinyu Tang
http://parasol.tamu.edu/people/xinyut/

Statement of Research Interests

One of my major research interests, and the focus of my dissertation work, lies in the area of computational biology. My research in this area focuses on developing new computational techniques to study biological molecular motions, specifically RNA folding [4, 5, 8] and protein folding [9, 10]. My other research interests include the algorithmic aspects of robotic motion planning and their applications in computer animation and computer graphics. The best way to illustrate my work is through the descriptions and animations that can be found on my web page (url above).

Figure 1: Conformational changes of the N-terminal domain of calmodulin: (a) from calcium-free state (1CFD) to (d) bound state (1CLL). (b) and (c) are two transitional conformations from a folding pathway generated by our protein folding server: http://parasol.tamu.edu/foldingserver. You can watch the animation of the complete folding pathway in our protein folding server.

1 Computational Biology

In our work, we have developed a novel framework to simulate molecular motions. It is a general method that can be applied to study a range of molecular motions. Currently, we use it to study the protein [9,10] or RNA folding [4,5,8] process, i.e., the conformational change process (also referred to as the folding pathway) a protein/RNA molecule goes through as it reaches the so-called native state, the energetically most stable state.

It is critical that we better understand protein/RNA motion and the folding process for several reasons: understanding the folding process can provide insight into how to develop better structure prediction algorithms, treatments for diseases such as Alzheimer's and Mad Cow disease can be found by studying protein misfolding, and many biochemical processes are regulated by protein or RNA motion. Unfortunately, our understanding of protein/RNA folding and movement is still very limited. Experimental methods cannot operate at the time scales necessary to record protein folding and motions, and traditional simulation techniques such as Molecular Dynamics and Monte Carlo methods are computationally too expensive to simulate long enough time periods for anything other than small fragments of peptides (20 residues) or RNA (tens of residues). Moreover, such pathways cannot provide information about global folding kinetics.

In contrast, our method can extract thousands of microscopic folding pathways in a few hours on a desktop PC for much larger RNA/protein molecules (hundreds of residues) [8, 9] and can also reveal global folding kinetics such as population kinetics, folding rate and transition states [4, 5, 8, 9]. Our method is based on the successful probabilistic roadmap (PRM) method for motion planning, which we have adapted to construct a roadmap approximating the folding energy landscape. We develop two tools called Map-based Monte Carlo (MMC) and Map-based Master Equation (MME) to run simulations and analyses on the energy landscape [8, 9]. MME and MMC can be used to extract global properties and microscopic features of the folding process, respectively. The key advantage of our work over other computational techniques is that our approach is fast and efficient while bridging the gap between high-level folding events and low-level folding detail.

RNA folding. Many RNA functions (e.g., catalysis, plasmid copy number regulation, and gene expression regulation) were recently found to be active only during the folding process. Using our MMC and MME

techniques, we can study those functions by simulating the folding process.

We validate our method against other computational methods and known experimental data. The results indicate that our method can use small roadmaps to capture the major features of much larger (10 orders of magnitude) energy landscapes [5, 8]. We also show how our method can be used to study kinetics-based functions for some RNA with 200+ residues. Previously, those properties could only be studied for small RNA whose conformation space could be enumerated (e.g., RNA with 30-40 nucleotides) [4] or for RNA whose kinetics were restricted in some way. In one example, we study Cole1 RNA II (200 nucleotides), whose folding process can regulate the plasmid copy number, and this function is known to be associated with its folding rate. Using our MME technique, we compare simulated folding rates for ColE1 RNAII and its mutants against experimental rates. We show that we can compute the same relative folding order as seen in the experiment [8]. In another example, using the MMC method, we extract microscopic folding pathways to study RNA MS2 (135 nucleotides) whose folding can regulate the gene expression of phage MS2 maturation protein. This regulation works only when a subsequence SD is open during the folding process. In our simulation, we estimate the gene expression rate by measuring the SD opening probability. We predict the gene expression rates of wild-type RNA MS2 and three of its mutants and match them to the experiment. We show in Table 1 that we can compute the same relative functional rates as seen in experiments [8].

Mutant	Experimental Expression Rate (in order of magnitude)	Our Estimation Using Different t				
		t =0.2	t = 0.3	t = 0.4	t = 0.5	t = 0.6
SA	0.1	0.1	0.04	0.03	0.03	0.08
WT	1	1.0	1.0	1.0	1.0	1.0
U32C	1	2.1	1.8	1.4	0.8	1.2
CC3435AA	5	7.2	8.4	3.8	3.5	9.8

Table 1: An example of predicting gene expression rates of MS2 using our method. It compares expression rates between WT and three mutants of MS2. Column 2 is the relative experimental expression rate in order of magnitudes. Column 3-7 are our estimation of the relative experimental expression rates using different threshold parameter t. It shows that we can compute the same relative functional rates as seen in experiments.

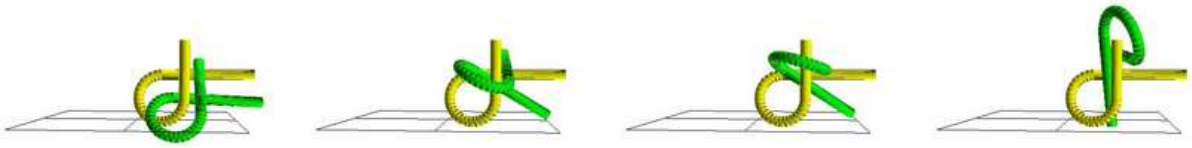
I plan to extend this research in several aspects. First, I plan to conduct a comprehensive study on a wide variety of RNA and predict some kinetics-related functional rates. I plan to collaborate with biochemists to perform some experiments to validate predictions by our simulations. I also plan to publish our simulation tools to help biochemists design mutants or experiments. Second, since there is huge amount of information in the energy landscape and the folding pathways, I plan to develop more analysis tools for our energy landscape and folding pathways. I will explore the use of some data mining and machine learning technologies to extract such information. For example, it has been found that the formation order of some "trigger" substructures is associated with the RNA folding results. Our tools provide an alternate and likely faster way to identify those "trigger" structures.

Protein folding. Compared to RNA folding, protein folding is widely considered to be harder since it involves a much larger energy landscape. For example, molecular dynamics simulation can take months to calculate a pathway on a supercomputer. In contrast, our technology can compute thousands of pathways and global folding kinetics in just a few hours on a desktop PC. We have obtained promising results for several small proteins (about 60-100 amino acids), and we have validated our pathways by comparing their secondary structure formation order with known experimental results, examining folding rates, and studying population kinetics. In addition, our recent work incorporating rigidity analysis [10] is able to capture the subtle folding differences between structurally similar proteins G, and two mutated forms of G, NuG1 and NuG2. As shown in Figure 1, we also illustrate how our technique can be used to study large-scale conformational changes in calmodulin, a 148 residue signaling protein known to undergo conformational changes. I also plan to develop more analysis and data mining tools to study protein kinetics. For example, we are currently adapting our MMC technique that we have applied to RNA for proteins — this will enable us to extract folding kinetics information from our folding pathways and correlate them to experimental results [9]. Using these techniques, I further plan to design protein mutants with desired folding kinetics

behaviours.

Our project has received great attention and has been reported in Genome Technology and has appeared on the BBC World News website. We have set up a protein folding server (http://parasol.tamu.edu/foldingserver) where users can submit their proteins so that we can simulate the folding process and generate animations and analysis results for the users. The server also includes a publicly available archive of protein motions.

2 Techniques and Other Applications of Robotic Motion Planning



(a) (b) (c) (d)

Figure 2: An solution our OBRRT algorithm found for a hard motion planning problem – alpha puzzle.

The motion planning problem consists of finding a valid (e.g., collision-free) path for an object from a start configuration to a goal configuration. It is a fundamental problem in many applications including robotics, animation and scientific computing. We actually encounter such problems in our daily actions whenever we move something (including ourselves) from one location to another. The complexity of motion planning depends on factors including properties or constraints of the robots and the complexity of the environment. Since most motion planning problems are NP-complete, deterministic solutions are infeasible in most cases. Recently, probabilistic methods such as Probabilistic Roadmap Method (PRM) or Rapidly-Exploring Random Tree (RRT) have been developed to solve many difficult problems. In our research, we develop new variants of probabilistic motion planning algorithms for robots with spatial constraints or in complex environments. We also apply those algorithms to a wide range of applications in robotics, bioinformatics and computer graphics and animation. In the future, I plan to continue to work in this area. I will expand our methods and apply them to more domains.

Motion Planning for Spatially Constrained Robots. Spatially constrained systems are widely involved in many applications in robotics and beyond, such as parallel robots, closed molecular chains, computer animation and grasping for single and multiple robots. Motion planning for such systems is particularly difficult due to additional *spatial constraints* placed on the robot, such as closure constraints for closed chains or requirements on end effector placement for articulated linkages. For example, the probability of randomly selecting a set of joint angles that satisfy the closure constraints is zero. We overcome this challenge by redefining the articulated linkage and its additional constraints into *reachable distance* space that enable us to sample the constraint surface directly [6,7]. Our method can be used to significantly improve the performance of sampling-based planners. We demonstrate its utility on a variety of spatially constrained systems including restricted end effector sampling for articulated linkages, on-line planning for drawing (or sculpting), and closed chain planning. In particular, we show that we can sample the constraint surface with complexity linear in the complexity of the robot system. Our method outperforms other randomized sampling methods such as PRM and RRT for sampling and planning for these spatially constrained systems.

Motion Planning in Complex Environments. Tree-based path planners have been shown to be well suited to solve various high-dimensional motion planning problems. However, they are known to have some difficulty in exploring complex environments with narrow passages, such as the alpha puzzle problem displayed in Figure 3. In this work, we present modifications that can be made to the Rapidly-Exploring Random Tree (RRT) path planning algorithm that allows it to explore narrow passages or difficult areas more effectively [3]. We show that both workspace obstacle information and C-space information can be used when deciding which direction to grow. The method includes many ways to grow the tree, some taking into account the obstacles in the environment. We found in our simulation results that using obstacle hints for directions to grow a tree for path planning can be beneficial, especially when exploring difficult areas.

Campus Navigator. In this project, we incorporate roadmap-based path planning techniques in a web-based route planner that covers the Texas A&M campus [2]. The goal is to allow users to quickly find directions to all TAMU buildings, departments, and major services. Transportation information (e.g., bus routes and parking lots) is incorporated to provide meaningful answers to users questions such as "How do I get from the Bright building to Reed Arena, taking an on-campus bus." We use a layered roadmap approach to compose multiple transportation methods into a single queryable roadmap. The user interface is implemented using Google Maps API.

Group Behaviors. Creating animations with complex and realistic group behaviors can be a difficult and time consuming task. This generally involves associating all possible behaviors of agents and associating the behaviors with all environmental events. In this work , we investigate methods to ease the process of producing realistic group behaviors [1]. More specifically, the main goal of this research is to simulate group behaviors by automatically combining a given set of simple composable behaviors for applications such as games, virtual reality, robotics and biological/ecological simulation. The result of this research is an easy to use, adaptive and flexible framework for simulating group behaviors.

(a) (b) (c) (d)

Figure 3: A sequence of images captured from the herding simulation with a group of shepherd controls the motion of a group of flock. We apply the idea of composable behaviors to compose shepherding behaviors from a set of simple primitive behaviors we call locomotions.

References

[1] J.-M. Lien, S. Rodriguez, X. Tang, J. Maffei, D. Corlette, A. Masciotra, and N. M. Amato. Composable group behaviors. Technical Report TR05-006, Parasol Lab, Dept. of Computer Science, Texas A&M University, Sep 2005.

[2] R. Pearce, B. Boyd, X. Tang, D. Haigler, A. Patel, and N. M. Amato. Supporting path planning queries incorporating multiple modes of transportation using layered roadmaps. Technical Report TR06-014, Parasol Lab, Dept. of Computer Science, Texas A&M University, Oct 2006.

[3] S. Rodriguez, X. Tang, J.-M. Lien, and N. M. Amato. An obstacle-based rapidly-exploring random tree. In *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, 2006.

[4] X. Tang, B. Kirkpatrick, S. Thomas, G. Song, and N. M. Amato. Using motion planning to study RNA folding kinetics. In *Proc. Int. Conf. Comput. Molecular Biology (RECOMB)*, pages 252–261, 2004.

[5] X. Tang, B. Kirkpatrick, S. Thomas, G. Song, and N. M. Amato. Using motion planning to study RNA folding kinetics. *J. Comput. Biol.*, 12(6):862–881, 2005. Special issue of Int. Conf. Comput. Molecular Biology (RECOMB) 2004.

[6] X. Tang, S. Thomas, and N. M. Amato. Effcient planning of spatially constrained robot using reachable distances. Technical Report TR07-001, Parasol Lab, Dept. of Computer Science, Texas A&M University, Jan 2007.

[7] X. Tang, S. Thomas, and N. M. Amato. Fast enforcement of closure constraints. In *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, 2007.

[8] X. Tang, S. Thomas, and N. M. Amato. Tools for simulating and analyzing rna folding kinetics. In *Proc. Int. Conf. Comput. Molecular Biology (RECOMB)*, 2007.

[9] L. Tapia, X. Tang, S. Thomas, and N. M. Amato. Roadmap-based methods for studying protein folding kinetics. Technical Report TR06-011, Parasol Lab, Dept. of Computer Science, Texas A&M University, Oct 2006.

[10] S. Thomas, X. Tang, L. Tapia, and N. M. Amato. Simulating protein motions with rigidity analysis. In *Proc. Int. Conf. Comput. Molecular Biology (RECOMB)*, pages 394–409, 2006.