Virtual Polymerase Chain Reaction

The polymerase chain reaction (PCR) stands among the keystone technologies for analysis of biological sequence data. PCR is used to amplify DNA, to generate many copies from as little as a single template. This is essential, for example, in processing forensic DNA samples, pathogen detection in clinical or biothreat surveillance applications, and medical genotyping for diagnosis and treatment of disease. It is used in virtually every laboratory doing molecular, cellular, genetic, ecologic, forensic, or medical research. Despite its ubiquity, we lack the precise predictive capability that would enable detailed optimization of PCR reaction dynamics. We are developing Virtual PCR (vPCR) software, an *in silico* method to model the kinetic, thermodynamic, and biological processes of PCR reactions.

Project Goals

The challenges that we plan to address using vPCR include:

1. computational optimization of signatures for pathogen detection, for a significant savings in cost and time over purely empirical assay optimization; and
2. assessment of signatures for forensic discrimination of closely related sequences.

The result of this project, a suite of programs that predict PCR products as a function of reaction conditions and sequences, will be used to address outstanding questions in pathogen detection and forensics at LLNL. vPCR should enable scientists to optimize PCR protocols in terms of time, temperature, ion concentration, and primer sequences and concentrations, and to estimate products and error rates in advance of performing experiments. Our proposed capabilities are well ahead of all currently available technologies, which do not model nonequilibrium kinetics, polymerase extension, or predict multiple or undesired PCR products. Licensing opportunities will be explored, and a publication is being prepared. A provisional and a full patent application have been filed.

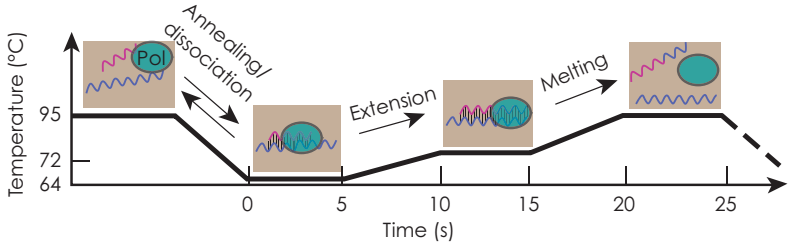

Figure 1. The diagram of a single thermocycle. We model four types of reactions: annealing, dissociation, polymerase extension, and melting. Although each of these predominantly occurs in a specific temperature regime, they are in constant competition throughout the thermocycle.

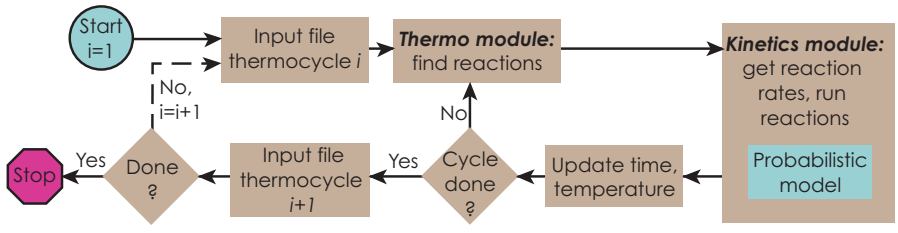

Figure 2. Simplified code overview.

For more information contact
Shea N. Gardner
(925) 422-4317
gardner26@llnl.gov

Relevance to LLNL Mission

vPCR supports LLNL missions in homeland security, Genomes to Life (GTL), and human health. Any field that uses PCR, including bioforensics, biodetection, basic research in GTL, and disease research will benefit.

FY2005 Accomplishments and Results

We have written prototype code for much of the vPCR software, and we have drafted a manuscript. The code incorporates simulation during the temperature ramps as well as the constant-temperature soaks; competition between multiple (dozens to hundreds) of potential reaction pathways, including both hybridization and denaturation; and several alternative formulations of stochastic kinetic simulation algorithms that we have compared and optimized to balance speed and accuracy to handle the complexity of our library of test cases.

In addition, we model temperature-dependent extension of DNA and decay of the Taq polymerase enzyme, and track all reaction particles and their concentrations through time. We have incorporated more efficient data structures, a speedier simulation algorithm, and reasonable approximations to improve the capacity to handle more complex reactions than could the prototype version of the code. Combined, these improvements in speed give running times that are two orders of magnitude faster. We are in the process of including hairpin kinetics of DNA secondary structure in the simulations. Experiments have been performed, and the data are now being analyzed.

Figures 1 through 4 show overviews of our modeling activities and sample results. The code is documented on the web at https://kpath.llnl.gov/vpcr/docs/vPCR_API/.

Related References

1. Allawi, H. T., and J. Santalucia, Jr., "Thermodynamics and NMR of Internal GT Mismatches in DNA," *Biochem.*, **36**, pp. 10581-10594, 1997.
2. Gillespie, D., and L. R. Petzold, "Improved Leap-Size Selection for Accelerated Stochastic Simulation," *J. Chem. Phys.*, **119**, pp. 8229-8234, 2003.
3. Cao, Y., D. T. Gillespie, and L. R. Petzold, "The Slow-Scale Stochastic Simulation Algorithm," *J. Chem. Phys.* **122**, 2005.
4. Maheshri, N., and D. V. Schaffer, "Computational and Experimental Analysis of DNA Shuffling," *Proc. Natl. Acad. Sci. USA*, **100**, pp. 3071-3076, 2003.
5. Wetmur, J. G., "DNA Probes: Applications of the Principles of Nucleic Acid Hybridization," *Crit. Rev. Biochem. Mol. Biol.*, **26**, pp. 227-59, 1991.

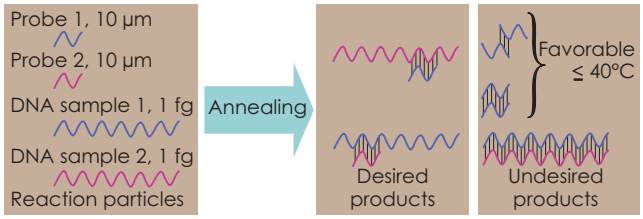

Figure 3. Example competing annealing reactions. At lower soak temperatures, annealing reactions are generally more favorable, opening pathways to undesired products.

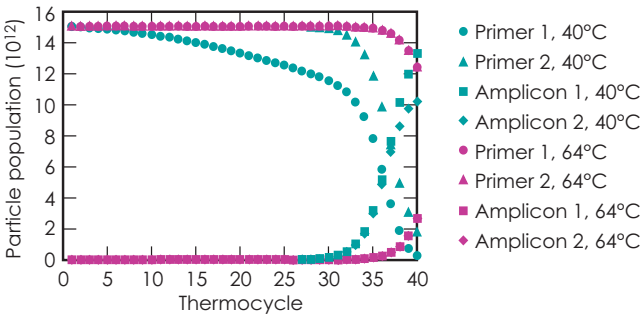

Figure 4. The system diagramed in Fig. 3 is simulated through 40 thermocycles, at soak temperatures of 40°C and 64 °C. Note that higher concentrations of desired amplicons are output at 40 °C, where pathways to undesired products also exist. Lowering the soak temperature increases the rates of all annealing reactions, which in this case outweighs the addition of undesired reactions. Optimizing the annealing temperature requires finding a balance between generating detectable yield and amplification of undesired products.

FY2006 Proposed Work

In FY2006, we plan to parallelize various portions of the code for running on the LC machines; complete experimental data analysis and compare it to predictions from the software; submit our first manuscript and prepare a second that includes the results of our work; continue the process of debugging and optimization, generating additional test cases; and complete the incorporation of hairpin kinetics.