

## **Tutorial Proposal for ISMB**

### **Title: A massively parallel high performance computing environment for Computational Biology**

**General Abstract:** Computation is playing an ever increasing and vital role in biology creating demand for new machines. Vendors strive to meet demands with advanced computer architectures such as IBM's Blue Gene machine. In this tutorial, we will give an overview of the Blue Gene architecture, both the hardware and software architecture. We will emphasize the key features that allow thousands of processors to work together on a user's problem. We will present the programming model used on Blue Gene. We will explain ways to take advantage of the Blue Gene nodes and their associated networks. We hope to provide a foundation for attendees to begin to think about problems and how to design and implement them so they will scale out and take full advantage of the computational power in Blue Gene.

Through the computational power of Blue Gene, scientists will tackle problems that to date they had not considered. This will happen in some ways we can not today predict. We will discuss alternative approaches to design of computation of the problem. We will also show by example some problems that one might not think would be suitable for the Blue Gene architecture. We will discuss in these examples which have been run on Blue Gene systems, actual performance results. More importantly, we will try to point how having unprecedented number of processors changes how one approaches the computational problem.

The tutorial will proceed to go in depth on one or two application codes. In this way, we hope to provide the attendees with a good understanding that while Blue Gene is about using a massive number of processors, the message passing programming model through MPI should be very familiar. We plan to have a couple of examples prepared with final selection of specific examples to be discussed determined by the majority of the audience.

Armed with this basic knowledge, the attendees will have an opportunity to try out some code examples on a Blue Gene system. Depending on the partition size, we demonstrate some simple scaling up to the number of processors available.

#### **Detailed Outline:**

The tutorial will be split into three major session with a demonstration or hands-on session at the end.

- General overview of Blue Gene, what Blue Gene is about – lots of processors. Some detail on the hardware architecture, detailing the compute core comprising the Blue Gene compute nodes, the I/O nodes, front-end node and service nodes. The five different networks will be discussed and the use of these networks will be described.

- The Blue Gene programming model will be discussed. Some discussion will be given on making use of MPI, such as using subcommunicators, to partition a problem to fit in memory.
- An overview of performance will be given highlighting optimization techniques to achieve better performance with a discussion of/overview of a wide range of applications such as Adaptive Mesh Refinement, nearest neighbor, tree communications etc. How to find the best mapping for optimal communications will be discussed.
- A detailed discussion on techniques to do really large scaling of problems. Examples of problems that scale will be discussed with some in depth analysis to understand how to apply these techniques to other problems.
- Several demonstrations and possibly an opportunity for some hands-on work with selected codes on a Blue Gene machine will be provided.