

## **Utilização do software MyGrid para adaptar uma aplicação de dinâmica molecular em uma Grade Computacional.**

Patrícia M. de Barros, Cleverson Veronez, Carla Osthoff  
LNCC/MCT - Laboratório Nacional de Computação Científica  
Av. Getulio Vargas, 333 - Quitandinha – Petrópolis/RJ  
CEP: 25651-070 - tel: (24) 2233-6179  
patricia@lncc.br, veronez@lncc.br, osthoff@lncc.br

Walfredo Cirne, Elizeu Santos-Neto, Lauro Costa  
DSC/UFCG  
walfredo@dsc.ufpb.br, elizeu@dsc.ufpb.br, lauro@dsc.ufpb.br

Pedro Pascutti, Alan Silva, Fabrício Silva  
IBCCF/UFRJ - NCE/UFRJ  
pascutti@biof.ufrj.br, alan@biof.ufrj.br, fasilva@nce.ufrj.br

### **Resumo**

Podemos definir os projetos de Computação em Grade, como sendo projetos que têm como objetivo, combinar múltiplas máquinas geograficamente dispersas como plataformas para a execução de computação distribuída. O software *MyGrid* é um projeto de Computação em Grade que provê um ambiente de execução global para permitir a execução remota de uma grande quantidade de tarefas em paralelo através das máquinas que o usuário tem acesso. Este projeto foi criado e está sendo desenvolvido pelo Departamento de Sistemas e Computação da Universidade Federal de Campina Grande. Dentre as aplicações que necessitam de grandes recursos computacionais, temos as aplicações na área de dinâmica molecular. Utilizamos para os nossos testes na Grade, o software *THOR* - que é um pacote de programas para modelagem e dinâmica molecular criado em Fortran 77. *THOR* está sendo desenvolvido e mantido pelo Laboratório de Física Biológica do Instituto de Biofísica Carlos Chagas Filho (IBCCF) da Universidade Federal do Rio de Janeiro, Brasil. O LNCC em colaboração com a UFCG, a Unisantos e o IBCCF estão com o objetivo de desenvolver um ambiente de execução de processos de dinâmica molecular em Grade. Este trabalho apresenta os testes preliminares implementados por este grupo em um sistema de Grade no Brasil para o estudo desta aplicação.

## Introdução

### Grade Computacional

Grades Computacionais são sistemas de suporte à execução de aplicações paralelas que acoplam recursos heterogêneos distribuídos, oferecendo acesso consistente e barato aos recursos independentemente de sua posição física. A tecnologia de Grades Computacionais possibilita agregar recursos computacionais variados e dispersos em um único "supercomputador virtual", acelerando grandemente a execução de várias aplicações paralelas. Grades se tornaram possíveis nos últimos anos devido a grande melhora em desempenho e redução em custo tanto de redes de computadores, quanto de microprocessadores. As vantagens de Grades Computacionais sobre plataformas para computação paralela tradicionais incluem: (i) integração de recursos computacionais geograficamente distribuídos, (ii) disponibilização de uma plataforma computacional distribuída de alto desempenho, (iii) maximização da utilização de recursos computacionais, e (iv) integração de grupos de pesquisadores.

Grades são construídos como um grupamento de serviços básicos independentes. Um aspecto essencial dos serviços de Grade é que estes estão disponíveis uniformemente através dos ambientes distribuídos que o compõe. Os serviços são agrupados em um sistema integrado, também chamado de middleware. Exemplos de ferramentas atuais de Grades incluem Globus [FOSTER98], Legion[LEWIS96], OpenGrid [CIRNE01], AppleS [BERMAN97].

### Dinâmica Molecular

Programas de simulação de dinâmica molecular são complexos e computacionalmente intensivos. Adaptar estes programas para execução na grade computacional não apenas reduz o tempo de execução das simulações, mas também torna possível a resolução de problemas consideravelmente maiores. Hoje em dia (na UFRJ), é possível simular moléculas com aproximadamente 20.000 átomos, enquanto diversos grupos de pesquisa no exterior já conseguem simular moléculas com 250.000 ou mais átomos. Um dos objetivos de pesquisa deste projeto é disponibilizar para os pesquisadores em biofísica molecular no Brasil uma capacidade de simulação equivalente aos mais avançados centros de pesquisa.

Aplicações de simulação de dinâmica molecular podem ser classificadas como irregulares e dinâmicas. Irregulares porque o número de operações por iteração pode variar para um mesmo passo, e dinâmicas porque o número de operações pode variar entre passos.

Utilizaremos em nosso trabalho o software de dinâmica molecular *THOR* [MOR 02], que é um pacote de programas para modelagem e dinâmica molecular criado em Fortran 77. *THOR* está sendo desenvolvido e mantido pelo Laboratório de Física Biológica do Instituto de Biofísica Carlos Chagas Filho (IBCCF) da Universidade Federal do Rio de Janeiro, Brasil.

## Software *MyGrid*

É uma ferramenta de suporte a Grade Computacional que permite a execução de aplicações *bag-of-tasks* em quaisquer máquinas que o usuário tenha acesso. Prevê um ambiente de execução global para permitir a execução remota de uma grande quantidade de tarefas em paralelo através das máquinas que o usuário tem acesso.

O software *MyGrid* [MYG 02] é um software de Grade desenvolvido em Java Versão 1.3. É um software de código aberto.

## Experimentos

Os testes foram efetuados na infra-estrutura da Grade, utilizando 6 domínios que passamos a descrever:

Domínio 1- LSD/UFCG, Domínio 2 - ApeLab/UCSD, Domínio 3 - Carcará/LNCC

Domínio 4 - NACAD/UFRJ, Domínio 5 - NCE/UFRJ, Domínio 6 - LCP/UFRJ

Temos então, um total de 6 domínios administrativos diferentes contribuindo para infra-estrutura.

Utilizamos *MyGrid* para executar *THOR* em 55 máquinas distribuídas pelos seguintes domínios administrativos: LSD/UFCG, NCE/UFRJ, APELab/UCSD, GRAIL/SDSC, LCP/UFRJ, LNCC e NACAD/UFRJ. Nossa execução do *THOR* foi composta de 60 tarefas independentes. Cada tarefa recebeu aproximadamente 3.3 MB de entrada e produziu 200 KB de saída. Cada tarefa executou em 4 minutos (na máquina mais rápida da nossa grade) a 33 minutos (na máquina mais lenta), isso com a CPU dedicada. Nós processamos em produção, sem nada dedicado. Conseguimos terminar as 60 tarefas em 38 minutos.

## Conclusões e Trabalhos Futuros

O trabalho apresentado foi realizado com arquivos de testes para entrada e saída de dados. O próximo passo, é utilizar arquivos de entrada reais da modelagem de inibidor/protease do HIV. Para a realização deste trabalho será necessário inicializar em uma grade computacional diversas aplicações de dinâmica molecular de simulação da interação inibidor/protease. Para este estudo, serão geradas diversas possibilidades de variantes de mutantes do HIV e de solventes inibidores para serem executados de forma independente em cada um dos nós disponíveis da Grade.

Cada simulação nesta aplicação é caracterizada por um tempo de processamento médio de 10 dias e pela transferência de arquivos de 10MB a 20MB de entrada, produzindo aproximadamente 100MB de saída. Nosso trabalho de pesquisa tem como objetivo final, desenvolver uma plataforma para modelagem da inibição da protease HIV. O grupo IBCCF

está estudando a protease do HIV-1 e uma família de seus mutantes, tanto livres como complexados com inibidores, por Dinâmica Molecular em solvente aquoso explícito.

A inibição da protease do HIV tem alcançado sucesso em tratamentos clínicos. Porém, uma dificuldade encontrada tem sido a rápida emergência da resistência a drogas pelo vírus. Diferentes cepas resistentes de HIV têm aparecido de acordo com a região do planeta. Por exemplo, no sul do Brasil há mutantes resistentes que possuem diferentes seqüências de aminoácidos na protease, em relação aos vírus mutantes resistentes de outras partes do país.

Além disso, uma previsão da Organização Mundial da Saúde é que em torno de noventa por cento das pessoas infectadas pelo vírus, neste início do século XXI, será do terceiro mundo. Assim, é necessário um aumento nos esforços teóricos e experimentais que possibilitem o desenvolvimento de novas drogas contra os mutantes regionais.

## Referências

- [ALL 01] ALLCOCK, W., Bester, J., Bresnahan, J., Chervenak, A., Foster, I., Kesselman, C., Meder, S., Nefedova, V., Quesnel, D., and Tuecke, S. Data Management and Transferring High-Performance Computational Grid Environments. *Parallel Computing* (2001).
- [BUY 02] BUYYA, R., Branson, K., Giddy, J. and Abramson, D., The Virtual Laboratory: Enabling On Demand Drug Design with the Worldwide Grid, *Concurrency and Computation: Practice and Experience* (2002).
- [CIR 01] CIRNE, W. and Marzullo, K., Open Grid: A User-Centric Approach for Grid Computing, nos Anais do 13th Symposium on Computer Architecture and High Performance Computing, (September 2001).
- [COL 95] COLLINS, J. R., Burt, S. K., and Erickson, J. W., Flap opening in HIV-1 Protease simulated by activated molecular dynamics, *Nature Struct. Biol.* 2, 334-338 (1995).
- [FOS 01] FOSTER, I. Kesselman, C. and Tuecke, S., The Anatomy of the Grid: Enabling Scalable Virtual Organizations, *Intl J. Supercomputer Applications* (15-3-2001).
- [MOR 02] Moret M.A., Bisch P.M., Mundim K.C. and Pascutti P.G., A new stochastic strategy to analyze helix folding. , *Mem Biophysical J*, 82 (3): 1123-1132.(2002).
- [MYG 02] <http://www.dsc.ufpb.br/mygrid/>
- [WAN 01] Wang & Koillman Computational study of protein specificity: the molecular basis of HIV-1 protease drug resistance. *P8(26)*, 14937-14942 (2001).