Primeiro Curso STING Millennium Suite Chemogenomics: Ferramentas para Analisar Macromoleculares e Aplicações em Chemogenomics

Paula R. Kuser, Jair L. de Siqueira Neto, Jorge H. Fernandez, Roberto H. Higa, Christian Baudet, Goran Neshich

Ministério da Agricultura,
Pecuária e Abastecimento

Documentos

Dezembro, 2002 26

ISSN 1677-9274

Primeiro Curso *STING Millennium Suite Chemogenomics*: Ferramentas para Analisar Macromoleculares e Aplicações em *Chemogenomics*

Embrapa

ISSN 1677-9274
Dezembro, 2002

Empresa Brasileira de Pesquisa Agropecuária
Embrapa Informática Agropecuária
Ministério da Agricultura, Pecuária e Abastecimento

Documentos 26

Primeiro Curso *STING Millennium Suite Chemogenomics*: Ferramentas para Analisar Macromoleculares e Aplicações em *Chemogenomics*

Paula Kuser Falcão
Jair Lage de Siqueira Neto
Jorge Hernandez Fernandez
Roberto Hiroshi Higa
Christian Baudet
Goran Neshich

Campinas, SP
2002

Embrapa Informática Agropecuária
Área de Comunicação e Negócios (ACN)
Av. André Tosello, 209
Cidade Universitária "Zeferino Vaz" – Barão Geraldo
Caixa Postal 6041
13083-970 – Campinas, SP
Telefone (19) 3789-5743 - Fax (19) 3289-9594
URL: http://www.cnptia.embrapa.br
e-mail: sac@cnptia.embrapa.br

Supervisor editorial**:** *Ivanilde Dispato*
Normalização bibliográfica: *Marcia Izabel Fugisawa Souza*
Capa: *Intermídia Produções Gráficas*
Editoração eletrônica: *Intermídia Produções Gráficas*

1ª. edição
on-line - 2002

Primeiro curso *STING Millennium Suite Chemogenomis* : ferramentas para analisar estruturas macromoleculares e aplicações em *chemogenomics* / Paula Kuser Falcão... [et al.]. — Campinas : Embrapa Informática Agropecuária, 2002.
26 p. : il. — (Documentos / Embrapa Informática Agropecuária ; 26).

ISSN 1677-9274

1. Bioinformática. I. Falcão, Paula Kuser. II. Série.

CDD – 570.285 (21st ed.)

Autores

Paula Kuser Falcão
Ph.D. em Física Aplicada, Cristalografia de Proteínas, Pesquisadora da Embrapa Informática Agropecuária, Caixa Postal 6041, Barão Geraldo - 13083-970 - Campinas, SP.
e-mail: paula@cnptia.embrapa.br

Jair Lage de Siqueira Neto
Estudante de Ciências Biológicas, Estagiário da Informática Agropecuária, Caixa Postal 6041, Barão Geraldo - 13083-970 - Campinas, SP.
e-mail: jair@cnptia.embrapa.br

Jorge Hernandez Fernandez
Ph.D. em Biologia Molecular, Colaborador Científico do Núcleo de Bioinformática Estrutural da Embrapa Informática Agropecuária, Caixa Postal 6041, Barão Geraldo - 13083-970 - Campinas, SP.
e-mail: jorgeh@cnptia.embrapa.br

Roberto Hiroshi Higa
M.Sc. em Engenharia Elétrica, Pesquisador da Embrapa Informática Agropecuária, Caixa Postal 6041, Barão Geraldo - 13083-970 - Campinas, SP.
e-mail: roberto@cnptia.embrapa.br

Christian Baudet
Estudante de Engenharia da Computação, Estagiário da Embrapa Informática Agropecuária, Caixa Postal 6041, Barão Geraldo - 13083-970 - Campinas, SP.
e-mail: christian@cnptia.embrapa.br

Goran Neshich
Ph.D. em Biofísica, Pesquisador da Embrapa Informática Agropecuária, Caixa Postal 6041, Barão Geraldo - 13083-970 - Campinas, SP.
e-mail: neshich@cnptia.embrapa.br

Apresentação

Tanto o conhecimento do funcionamento das ferramentas de bioinformática, quanto os problemas de biologia que podem ser estudados ou resolvidos por meio de "softwares" utilizados em computadores, são uma nova fonte de pesquisa e trabalho para biólogos interessados em identificar novos genes ou novos padrões e para matemáticos e cientistas da área de computação que desejam desenvolver e aplicar novos algoritmos computacionais a problemas da biologia.

O software *Sting Millennium Suite* (SMS) é uma ferramenta importante para o pesquisador que trabalha com estruturas de proteínas. As análises obtidas com a utilização do programa podem dar subsídios para entender questões relativas ao funcionamento dessas moléculas, levando em conta a relação estrutura/função.

Para contribuir com a formação de recursos humanos nesta nova área promissora e moderna da bioinformática, o Núcleo de Bioinformática Estrutural da Embrapa Informática Agropecuária ofereceu o "Primeiro Curso STING *Millennium Suite Chemogenomics*: ferramentas para analisar estruturas macromoleculares e aplicações em *chemogenomics*". O curso permitiu o domínio das ferramentas e técnicas de bioinformática oferecidas pelo SMS, assim como adquirir um conhecimento multidisciplinar em Biologia Molecular Computacional. O curso forneceu diploma e conteve em sua organização, seções teóricas e práticas.

José Gilberto Jardine
Chefe-Geral

Sumário

Primeiro Curso *STING Millennium Suite Chemogenomics*: Ferramentas para Analisar Macromoleculares e Aplicações em *Chemogenomics*

Paula Kuser Falcão
Jair Lage de Siqueira Neto
Jorge Hernandez Fernandez
Roberto Hiroshi Higa
Christian Baudet
Goran Neshich

Introdução

O "Primeiro Curso *STING Millennium Suíte Chemogenomics*" ocorreu no Núcleo de Bioinformática Estrutural da Embrapa Informática Agropecuária, localizada em Campinas – SP, nos dias 29 e 30 de outubro de 2002. O curso foi constituído por palestras ministradas por pesquisadores da Embrapa ou colaboradores, e por atividades práticas desenvolvidas nos computadores do Núcleo de Bioinformática.

O objetivo do curso foi familiarizar pesquisadores de áreas de biotecnologia, como genética, bioquímica, biologia molecular e bioinformática, com as novas ferramentas disponíveis para a análise de estruturas macromoleculares, juntamente com a explicação de novas abordagens e metodologias para o estudo das macromoléculas.

Com o grande aumento de projetos genoma, que tem como objetivo identificar todos os gens de um organismo, identificar as sequências de DNA e as proteínas que são codificados por estes, há uma grande procura por programas que consigam extrair informações importantes de

estruturas de proteínas aumentando o conhecimento das mesmas. Existem atualmente centenas de projetos genoma em andamento em laboratórios de todo o mundo. Estes projetos trouxeram a público milhares de seqüências de proteínas com função ainda desconhecida. O estudo e análise de macromoléculas com estrutura já desvendada pode ajudar na determinação de proteínas homólogas.

O público-alvo do curso foi constituído por grupos de pesquisadores participantes do projeto " Aplicações de BioInformática com alta demanda por CPUs na era pós-genoma (BIpósG)", financiado pela agência Financiadora de Estudos e Projetos ligada ao Ministério da Ciência e Tecnologia (FINEP/MCT).

Os aspectos gerais abordados no curso foram: busca por similaridade de seqüências, dinâmica e estrutura de macromoléculas, características de superfície determinando interações moleculares e *chemogenomics.*

A busca por similaridade de seqüências de aminoácidos é realizada com o objetivo de identificar relações de evolução, estruturais e funcionais entre as seqüências que estão sendo estudadas. Esta busca é feita através de um alinhamento de seqüências, que é um processo computacional com o objetivo de dispor seqüências alinhadas de forma a ressaltar as similaridades entre elas e que permita dizer se as seqüências tem uma similaridade suficiente que justifique inferir que elas são homólogas. Embora os termos similaridade e homologia são muitas vezes confundidos, nesta situação eles possuem significados diferentes. Similaridade é uma quantidade que pode ser expressa em por exemplo, porcentagem de identidade. Homologia, por outro lado, se refere a uma conclusão tirada desses dados que dois gens tem uma história evolucionária comum.

A relação estrutura-função é muito significativa em proteínas (ou macromoléculas). Para entender o funcionamento dos sistemas biológicos e os mecanismos de ação da vida, é necessário entender como as proteínas funcionam, e conseqüentemente, precisa-se conhecer e ter ferramentas para analisar suas estruturas. Estudar a dinâmica das proteínas significa estudar as energias envolvidas na estabilidade da estrutura tridimensional da proteína ou em alterações da estrutura.

A importância de estudar-se as superfícies das proteínas advém do fato de que a maioria das interações entre proteína-proteína, proteína-DNA, proteína-ligante, ocorrem nas superfícies.

Chemogenomics, além de ser um termo muito recente, é uma nova área de estudo dentro da química-genômica. Nesta ciência, pequenas moléculas já bem conhecidas de uma determinada família são utilizadas para elucidar a função e o papel biológico de outro membro da mesma família cuja função ainda é desconhecida. Este método genômico-químico pode representar uma nova maneira de identificar moléculas alvo e assim acelerar dramaticamente o processo de desenvolvimento de novos fármacos.

O curso foi dividido em três etapas:

1) Estudo de propriedades da estrutura de proteínas: proteínas e ácidos nucléicos são os maiores responsáveis por todas as reações que ocorrem nas vias bioquímicas dos seres vivos. Cada proteína têm sua própria função e o estudo das estruturas das proteínas é importante em vários aspectos. As estruturas confirmam as mudanças de evolução que ocorrem em espécies relacionadas, através de mutações randômicas. Uma vez que estas mutações podem levar a desordens genéticas e doenças no nível molecular, um entendimento claro da natureza dessas doenças depende da determinação precisa de sua estrutura. Além disso, quando se conhece bem a estrutura de uma enzima, um inibidor adequado pode ser desenhado, criando um novo fármaco. Esta etapa foi constituída de apresentações orais, incluindo uma palestra introdutória. Em seguida, foi realizada uma atividade prática, onde os participantes puderam vivenciar toda a teoria apresentada nas palestras iniciais desta etapa, através do módulo JPD[1] (*Java Protein Dossier*) do SMS (*Sting Millennium Suite*). Todas as atividades desenvolvidas estão descritas detalhadamente em item posterior.

---

[1] JPD – *Java Protein Dossier* ferramenta computacional interativa do programa SMS que apresenta em forma gráfica as principais características físico-químicas de estruturas macromoleculares descritas em arquivos PDB.

2) Modelagem e análise de estruturas protéicas: quando a seqüência de uma determinada proteína tem alto grau de identidade (~50%) com outra, cuja estrutura já foi determinada, é possível construir um modelo tridimensional da proteína genômica cuja estrutura é desconhecida. A modelagem de proteínas utiliza ferramentas computacionais para construção de modelos da estrutura protéica. A partir da busca de outras proteínas com estrutura conhecida e seqüência similar de aminoácidos, faz-se um alinhamento múltiplo das seqüências, escolhe-se os melhores homólogos potenciais cujas estruturas tridimensionais serão usadas como base para construção da proteína a ser modelada. Pode-se fazer também a construção de estruturas de mutantes da proteína através de alterações de sua seqüência. A modelagem, quando possível, reduz o tempo necessário para determinação de uma estrutura além de ter um custo econômico mais baixo. Existem vários programas para modelagem de estruturas protéicas, sendo que neste curso usou-se o programa Modeller (Marti-Renom et al., 2000).

3) Impressão digital de proteínas: é o registro de contornos e propriedades eletrostáticas da superfície da proteína, útil para análise das possibilidades de ligação da proteína com outras moléculas. É na interação que ocorre entre as superfícies proteína-proteína, proteína-ligante que acontecem as reações mais importantes das macromoléculas. A complementariedade destas superfícies é fundamental para que essas interações aconteçam. Existe um método que calcula uma impressão digital bidimensional das proteínas e ligantes, possibilitando encontrar quais impressões digitais se complementam, reduzindo assim o tempo necessário para se encontrar um inibidor adequado para determinada proteína. Este método, desenvolvido pelo colaborador Bernard Maigret, foi apresentado no curso.

A metodologia para a realização do curso foi separar os participantes em nove duplas, permitindo que apenas dois participantes ocupassem uma máquina, nas quais foram abertas contas específicas para uso durante o curso. Com a disponibilidade de uma máquina para cada duas pessoas, foi possível praticar todos os exercícios. As etapas são descritas no decorrer do documento.

Cronograma do Curso

29/10/2002

Manhã	Atividade
08:30 - 09:40	1. Palestra descrevendo os objetivos do curso. (Goran Neshich)
	2. Apresentação introdutória descrevendo a dinâmica do curso. (Goran Neshich)
09:40 - 10:00	*Intervalo*
10:00 - 12:00	3. Exploração completa do SMS e seu componente *Java Protein Dossier* (JPD)4. *Exercício Prático I*: estudando propriedades estruturais de proteínas. Uso de ferramentas do SMS e JPD para estudar estruturas protéicas.
Tarde	Atividade
13:30 - 15:00	*5. Exercício Prático II*: Modelando e Analisando Estruturas de Proteínas
15:00 - 15:20	*Intervalo*
15:20 - 17:00	*Exercício Prático II*: Modelando e Analisando Estruturas de Proteínas (continuação)

30/10/2002

Manhã	Atividade
08:30 - 09:40	6. Seminário do prof. Dr. Bernard Maigret: "Criando impressões digitais"
09:40 - 10:00	*Intervalo*
10:00 - 12:00	7. *Exercício Prático III*: Impressões digitais de proteínas
Tarde	Atividade
13:30 - 15:00	8. Entendendo as harmônicas esféricas.
15:00 - 15:20	*Intervalo*
15:20 - 17:00	Entendendo as harmônicas esféricas

Parte 1

Para dar início ao curso, o Dr. Goran Neshich, líder do Núcleo de Bioinformática Estrutural da Embrapa, apresentou os objetivos e a forma como seriam ministradas as palestras e atividades práticas. Em seguida ministrou uma palestra explorando de forma abrangente o software *STING Millennium Suite* (SMS) e seu componente *Java Protein Dossier* (JPD), dando início à primeira etapa do curso.

O *STING Millennium Suite* (SMS) (Neshich et al., 2003) é uma suite de programas para análise de estruturas de proteínas e a relação com a sua função. O *Java Protein Dossier* é uma apresentação das propriedades da proteína (um *cartoon*). Sua forma de apresentação tem a vantagem de permitir que diversas propriedades sejam observadas simultaneamente, facilitando a análise das propriedades da proteína.

Após a palestra inaugural, os participantes do curso tiveram a oportunidade de executar uma atividade prática utilizando o módulo descrito.

Foram apresentadas três serino proteases complexadas a respectivos inibidores para que os participantes pudessem analisar três complexos de proteína/inibidor no módulo JPD:

1SLU - enzima tripsina complexada com ecotina.

1MTU - Inibidor do fator Xa em complexo com tripsina bovina

1PPF - elastase complexada com inibidor de ovomucoide

Estes três complexos foram pré-selecionados pela equipe de organização, sendo considerados bons modelos para finalidade didática, facilitando a compreensão dos participantes e contribuindo para um aprendizado efetivo.

Os recursos do módulo JPD foram amplamente demonstrados aos participantes, explicando como esta ferramenta pode facilitar o estudo de estruturas macromoleculares. A Fig. 1 mostra o JPD em uso com o complexo 1SLU, um mutante N143H, E151H de tripsina aniônica de rato complexada com o inibidor ecotina A86H.

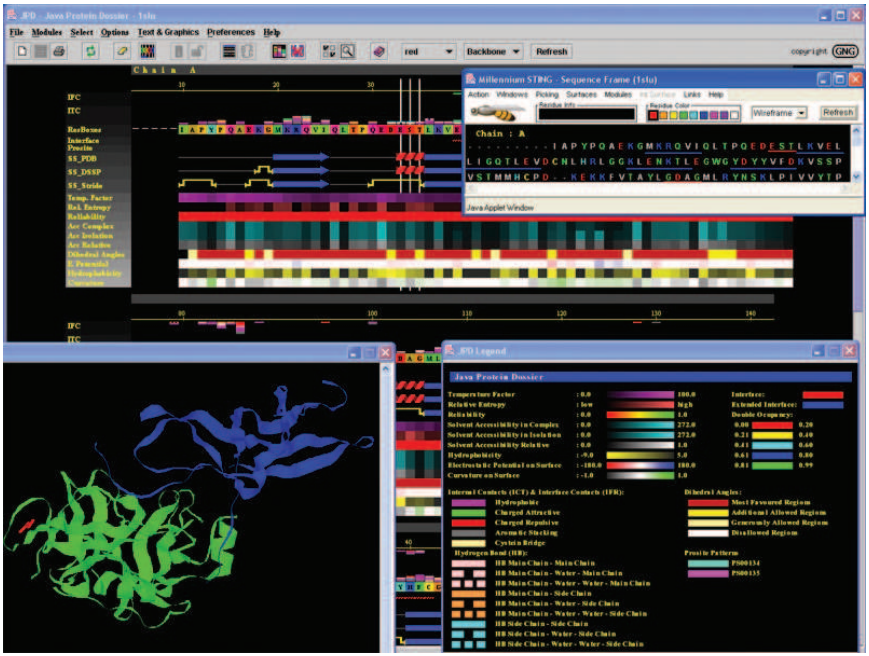

Fig. 1. Exemplo do tipo de figura que foi utilizada no curso para ser analisada e estudada pelos particpantes. Dossier de várias características estruturais do complexo 1SLU criado pelo programa JPD.

Parte 2

Na segunda etapa do curso, foram seguidos passos para se entender como funciona a modelagem molecular. As aulas teórica e prática foram ministradas pela Dr. Paula Kuser Falcão e pelo colaborador Dr. Jorge Hernandez Fernandez. O fluxograma a seguir descreve as etapas que foram seguidas para explicar como funciona a modelagem de macromoléculas, desde a busca por seqüências homólogas até a obtenção de um modelo da proteína em questão. O fluxograma que se segue foi apresentado durante o curso. Os participantes acompanharam a atividade prática de modelagem seguindo os passos do fluxograma.

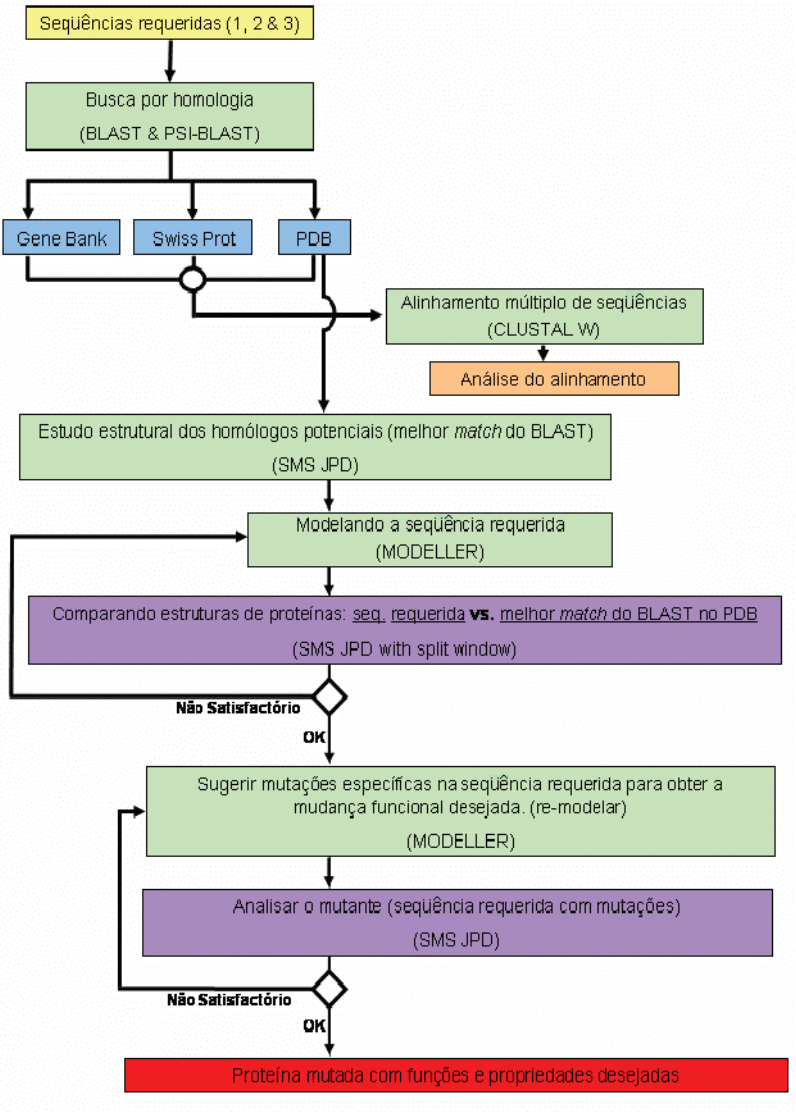

Fig. 2. Fluxograma da segunda etapa do curso: modelagem e análise estrutural de proteínas.

Como descrito no fluxograma (Fig. 2), os participantes receberam uma de três possíveis seqüências de aminoácidos, simulando uma proteína de interesse sem estrutura resolvida. Os partipantes deveriam fazer busca por seqüências potencialmente homólogas em diferentes bancos de dados, incluindo o *Protein Data Bank* (PDB) (Berman et al., 2000). A busca por seqüências homólogas foram executadas com o auxílio do software Blast (Altschul et al., 1990), instalado localmente. Uma vez encontradas essas seqüências foi feito o alinhamento múltiplo das seqüências obtidas na busca com o programa ClustalW (Thompson et al., 1994). Após obter o alinhamento, o módulo JPD do SMS, foi utilizado para escolher uma estrutura de uma proteína semelhante à proteína de interesse, para ser utilizada na modelagem molecular. Eram considerados o maior número de aminoácidos idênticos, ou conservados, as estruturas que tivessem sido resolvidas a uma melhor resolução, ou seja, com mais detalhes, a melhor estereoquímica, entre outros aspectos. Após a escolha do modelo inicial, foi feita a modelagem com o programa Modeller (Marti-Renom et al., 2000).

O uso do programa Modeller exigiu um acompanhamento passo a passo, conforme descrito no parágrafo a seguir.

Roteiro para modelagem recebido e acompanhado pelos participantes do curso:

1) Logar na máquina como usuário tempus e senha senhatemp;

2) Abrir a pasta SMScg no canto superior direito da tela.

3) Abrir o Internet Explorer na página do curso.

4) Abrir em Program Details o II Practical Exercise e ler o fluxograma.

Busca por homologia nos bancos de dados (Sequence Databases Search):

5) Abrir o arquivo texto com a seqüência inicial em formato fasta.

6) Abrir em Softwares o programa Blast, e fazer uma busca por homologia da seqüência inicial contra o banco de nucleotídeos - nr(nucleotide) - usando o tblastn.

7) Observar a quantidade de seqüências nucleotídicas que a busca por homologia retornou nessa base de dados.

8) Voltar ao browser com a página do curso e abrir novo programa Blast. Desta vez fazer uma busca por homologia da seqüência inicial contra um banco de proteínas - nr (protein) - usando o blastp.

9) Observar a quantidade de hits retornados nessa busca e comparar com a busca anterior.

10) Executar mais uma vez o programa Blast, mas dessa vez contra um banco de estruturas - pdb - usando o blastp.

11) Abrir em Softwares o programa PSI-Blast, e fazer a mesma busca anterior.

12) Comparar o resultado da última busca com os resultados das buscas anteriores.

Alinhamento (Sequence Alignment):

13) Abrir na página do curso, no campo "Data Bases" o "PDB" e conferir se a base de dados PDB está selecionada.

14) Selecionar e copiar ("Ctrl C") os 10 melhores hits e colar ("Ctrl V") no campo de "Identifiers". Clicar em "Submit".

15) Copiar ("Ctrl C") as seqüências do PDB requeridas.

16) Abrir o arquivo da seqüência inicial na pasta SMScg. Colar ("Ctrl V") as seqüências de melhores hits do Blast, e em seguida salvar esse conjunto de seqüências como "alig_strc.txt" na máquina gaviao.

17) Executar o programa X-Server (o link está na pasta do curso - SMScg).

18) Abrir um SSH, logando na máquina "gaviao" como usuário "guest0?" e senha smscg.

19) Digitar "tcsh" para mudar o shell de bash para tcsh. Em seguida digitar "setenv DISPLAY nome_da_máquina_local:0.0".

20) Executar o programa ClustalX digitando""clustalx".

O resultado do alinhamento das seqüências deve ser avaliado manualmente.

Verificando a Estrutura (Looking at structure):

21) Launch Java Protein Dossier from:

http://bsgi.cbi.cnptia.embrapa.br/SMS/

Colocar código do PDB para proteína que foi escolhida como modelo no Blast Sting it!

Construindo a estrutura (Homology Modeling):

21) Modeling com Modeller :

A partir do folder SMScg, abrir duas shells SshClient .

Conectar na máquina "ganso" nas duas shells: Quick Connect :

Host name: ganso

Username: guest0?

Password: smscg

21.1) Shell 1:

Entrar no nodo específico

$rsh node? (ex. guest01 ' node1)

$tcsh

21.2) Shell1

Copiar pdb que será usado como modelo:

$cd modeller

$cp /db/dpb/xxx.pdb .

21.3) Shell2 : Editar arquivo ALIGN2D.top

$cd modeller

$ pico ALIGN2D.top - script para fazer alinhamento entre modelo e target

Mudar:

file.pdb : pdb que sera usado como modelo

pdb name: código do pdb

seq.txt : arquivo onde será colocado a seqüência a ser modelado (target)

seq : código da seqüência

file.ali : nome do arquivo onde será colocado o alinhamento entre modelo e target

21.4) Shell2: Editar arquivo SEQ.txt e mudar nome

$ pico SEQ.txt

Colocar seqüência target no lugar de XXXX (manter "" no final da seqüência).*

Adicionar o sinal > antes de P1.

Mudar "name"

$ mv SEQ.txt xxx.txt

21.5) Shell1 : Rodar programa

$ mod6v2 ALIGN2D.top

Quando o programa termina de rodar imprime a flag STOP na tela e gera arquivos .log e .ali

21.6) Shell2 : Editar arquivo MODEL.top

$pico MODEL.top

Mudar: "file.ali", "pdb name", "seq"

21.7) Shell1 : Executar programa

$ mod6v2 MODEL.top

Avaliação da qualidade do modelo construído (Evaluating the quality of the model):

22) Checar a qualidade da estrutura com Procheck(Laskowski,) e SMS

22.1) Procheck

Colocar o pdb modelado no diretório procheck

$ procheck xxx.pdb cadeia(se existir) resolução

Procheck cria vários arquivos .ps

Transferir arquivos .ps para PC .

Abrir programa gsview que está na pasta SMScg e visualizar arquivos .ps

23) Comparar estruturas com SMS Java Protein Dossier (JPD)

23.1) Copiar os arquivos pdb da ganso para a gaviao: Clicar no ícone "New File Transfer Window", nas duas máquinas e transferir o arquivo, arrastando de uma máquina para outra.

23.2)Abrir SMS Java Protein Dossier (JPD) do endereço: http://bsgi.cbi.cnptia.embrapa.br/SMS/

Escrever nome do arquivo pdb e clicar em Sting it.

23.3) Colocar o pdb modelo para comparar

Clicar no ícone "New PDB" (folha em branco), ou no comando File, New PDB.

Escolher Split window e escrever o nome do PDB.

Após a obtenção do modelo, novamente o módulo JPD foi utilizado para comparar o modelo construído com a estrutura molde. Foram então introduzidas mudanças nos aminoácidos que eram diferentes no modelo inicial e na proteína modelo, com a finalidade de obtenção de propriedades desejadas para o modelo construído. Todas as possibilidades de modelagem foram discutidas entre os participantes e os organizadores.

Parte 3

Na terceira etapa, tivemos um seminário do colaborador prof. Dr. Bernard Maigret (CNRS – Nancy, France) sobre esféricas harmônicas e como elas podem representar a impressão digital (fingerprints) de estruturas macromoleculares. As harmônicas esféricas são funções matemáticas que podem ser usadas para decompor outras funções usando-as como base. Foi demonstrado como pode-se futuramente obter o pareamento (*matching*) de algumas moléculas, como enzimas, com inibidores específicos para um determinado sítio, utilizando-se os mapas de contorno, que são representações gráficas das esferas harmônicas das estruturas moleculares, ou seja, as próprias impressões digitais.

O processo de *docking* consiste na utilização de ferramentas computacionais para prever complexos entre estruturas de proteínas e ligantes, ou complexos proteína-proteína, proteína-DNA. A maioria dos processos biológicos ocorrem devido a estas interações. O conhecimento dos complexos fornece informações sobre o funcionamento das proteínas. Os programas de *docking* tentam encontrar o encaixe perfeito para duas estruturas tentando várias orientações, criam bilhões de possíveis complexos, avaliam estes complexos e prevêem qual deve ser o complexo correto. O programa mais utilizado atualmente para este tipo de cálculo é o AutoDock (Morris et al. 1998, 1996; Goodsell & Olson, 1990).

Fig. 3. *Cartoon* explicativo descrevendo de forma visual como será feito o casamento (*docking*) de pares enzima/inibidor através dos mapas de contorno (impresões digitais).

O programa desenvolvido por Cai et al. (2001) do laboratório de um dos colaboradores do grupo, Dr. Maigret, foi instalado nas máquinas do NIB especialmente para o curso. A Dra. Paula Kuser desenvolveu um tutorial para operar este programa. Os participantes do curso, seguiram o tutorial e utilizaram o *script* que já havia sido desenvolvido para criar "fingerprints" de três enzimas e de três ligantes que haviam sido previamente selecionados, e misturados. Com um programa desenvolvido pelo Dr. Michel Yamagishi, foi possível encontrar qual enzima seria complementar a qual ligante, fazendo então os pareamentos enzima/ligante. Esta etapa é de grande importância para os pesquisadores que estão interessados em encontrar inibidores para os sítio-ativos de suas proteínas de estudo, uma vez que reduz muito o processo de busca.

O tutorial para este exercício está descrito a seguir:

- *Abra SSH Secure Shell Client, e log in na máquina "gaviao" com o user name" "guest0?" (password smscg).*
- *Digite $ cd FINGERPRINT para entrar no diretório.*
- *Digite $ pico Input para editar o arquivo FINGERPRINT.*
- *Selecione "LIG" (para calcular o superfície do ligante)) ou" "CAV" (para calcular a superfície da cavidade). Esta escolha deve ser feita na linha 27. Para ver o número das linhas digite "Ctrl C". Quando o arquivo estiver editado, digite"Ctrl X" para fechá-lo, "Y" para confirmar e "Enter" e.*
- *Digite $ ll PDB/ para verificar se o diretório PDB contém o arquivo para calcular PDB os mapas. Caso o arquivo não estiver lá copie-o digitando type $ cp /pdb/filename.pdb ./PDB*
- *Após digite $ pico pdb_list e escreva " "filename.pdb" .*
- *Quando todos os arquivos estiverem corretos, digite $ fingerprint para rodar o algoritmo.*
- *Após o final do cálculo, se você selecionou LIG, digite $ mv PDB/filename.kin PDB/filename_lig.kin; se você selecionou CAV, type $ mv PDB/filename.kin PDB/filename_cav.kin*

Visualização dos mapas de contorno

- *Para ver os mapas de contorno gerados com FINGERPRINT, digite $ cd PDB, e "mage".*

- *Uma janela gráfica será aberta na tela.*
- *Clicar no ícone "Proceed".*
- *Clicar em "File", "Open New Kin File".*
- *Selecionar seu filename.kin e clicar "OK*

Considerações Finais e Avaliação

O curso mostrou-se bastante produtivo, tendo obtido excelente avaliação feita pelos próprios participantes. A seguir a Tabela 1 apresenta a opinião dos participantes. O curso deve ser oferecido novamente no ano de 2003.

Tabela 1. Avaliação dos participantes do curso.

0 - péssimo	1-ruim	2 - razoável	3 - bom	4 - muito bom	5 - excelente

Questões	0	1	2	3	4	5	média
Conforto das instalações	0	0	0	4	5	8	**4,2**
Tempo de duração do curso	0	0	2	3	6	6	**3,9**
Objetividade e clareza	0	0	1	2	5	9	**4,3**
Didática na transmissão das informações	0	0	0	3	6	8	**4,3**
Conhecimento sobre os assuntos abordados	0	0	1	1	3	12	**4,5**
Nível de aproveitamento do conteúdo abordado em minhas atividades profissionais	0	0	0	7	3	6	**3,9**
Avaliação geral do curso	0	0	0	0	7	10	**4,6**

Também foram feitos comentários aos organizadores do curso, transcritos a seguir:

> *"O curso foi excelente. Sugiro somente que vocês dêem uma noção de UNIX para todos no início. Acredito que assim vocês economizariam tempo. Parabéns, foi ótimo!"*
>
> *"Apesar de ser um assunto" 'não-habitual' no meu trabalho, foi uma introdução muito bem feita ao assunto, despertando mais minha atenção para o assunto. Como agora possuo um conhecimento*

mínimo, posso agora pensar e ter dúvidas sobre o assunto. Devido a dificuldade de alguns temas, não entendi tudo mas acredito que foi muito bem levado. Parabéns pelo curso!"

"Sugiro a criação de cursos de 1 dia sobre um dos vários programas existentes (treinamento). Colocam tutoriais na rede, caso ainda não existam, com exemplos simples e complicados."

"Precisa-se de mais tempo pela complexidade do tema."

"Na minha opinião o curso poderia ter uma duração de 3 ou 4 dias."

"Achei o curso bastante amplo. Gostaria que outros cursos mais específicos fossem oferecidos. Gostaria também de parabenizar os organizadores. No geral foi excelente. Uma sugestão é fazer cursos menores (menos pessoas) e pessoas com conhecimentos anteriores parecidos."

"Mais tempo para exercícios práticos - curso on-line."

Referências Bibliográficas

ALTSCHUL, S. F.; GISH, W.; MILLER, W.; MYERS, E. W.; LIPMAN, D. J. Basic local alignment search tool. **J. Mol. Biol.**, v. 215, p. 403-410, 1990.

BERMAN, H. M.; WESTBROOK, J.; FENG, Z.; GILLILAND, G.; BHAT, T. N.; WEISSIG, H.; SHINDYALOV, I. N.; BOURNE, P. E. The Protein Data Bank. **Nucleic Acids Research**, v. 28, p. 235-242, 2000.

CAI, W.; SHAO, X.; MAIGRET, B. Protein-ligand recognition using spherical harmonic molecular surfaces: twards a fast and efficient filter for a large virtual throughput screening. **Journal of Molecular Graphics and Modelling**, v. 5301, p. 1-16, 2001.

GOODSELL, D. S.; OLSON, A. J. Automated docking of substrates to proteins by simulated annealing.'**Proteins**: Struc. Func. and Genet., v. 8, p. 195-202, 1990.

MARTI-RENOM, M. A.; STUART, A.; FISER, A.; SÁNCHEZ, R.; MELO, F.; SALI, A. Comparative protein structure modeling of genes and genomes. **Annu. Rev. Biophys. Biomol. Struct.**, v. 29, p. 291-325, 2000.

MORRIS, G. M.; GOODSELL, D. S.; HALLIDAY, R. S.; HUEY, R.; HART, W. E.; BELEW, R. K.; OLSON, A. J. Automated docking using a Lamarckian genetic algorithm and empirical binding free energy function."**J. Computational Chemistry**, v. 19, p. 1639-1662, 1998.

MORRIS, G. M.; GOODSELL, D. S.; HUEY, R.; OLSON, A. J. Distributed automated docking of flexible ligands to proteins: parallel applications of AutoDock 2.4. **J. Computer-Aided Molecular Design**, v. 10, p. 293-304, 1996.

NESHICH, G.; TOGAWA, R. C.; MANCINI, A. L.; KUSER, P. R.; YAMAGISHI, M. E. B.; PAPPAS JUNIOR, G.; TORRES, W. V.; CAMPOS, T. F. e; FERREIRA, L. L.; LUNA, F. M.; OLIVEIRA, A. G.; MIURA, R. T.; INOUE, M. K.; HORITA, L. G.; SOUZA, D. F. de; DOMINIQUINI, F.; ÁLVARO, A.; LIMA, C. S.; OGAWA, F. O.; GOMES, G. B.; PALANDRANI, J. F.; SANTOS, G. F. dos; FREITAS, E. M. de; MATTIUZ, A. R..; COSTA, I. C.; ALMEIDA, C. L. de; SOUZA, S.; BAUDET, C.; HIGA, R. H. STING Millennium: a Web based suite of programs for comprehensive and simultaneous analysis of protein structure and sequence. **Nucleic Acids Research**, v. 31, n. 13, p. 3386-3392, 2003.

THOMPSON, J. D.; HIGGINS, D. G.; GIBSON, T .J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. **Nucleic Acids Research**, v. 22, p. 4673-4680, 1994.


Embrapa
Informática Agropecuária