



バイオグリッドプロジェクト

文部科学省RR2002 「スパコンネットの構築」

大阪大学サイバーメディアセンター
下條真司

<http://www.ais.cmc.osaka-u.ac.jp/~shimojo>

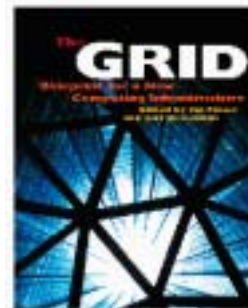
www.biogrid.jp



グローバルコンピューティング

The Grid

- 広域分散計算 (High-Performance Distributed Computing)
- 仮想的な大規模並列計算機
 - Metacomputing [Smarr87]
 - “The GRID” [Foster98]
- さまざまなツールキットも登場
 - Globus, Ninf
 - Web computingへの傾倒 – OGSA
- 最近は
 - 科学とITのコラボレーション
 - Peer-to-peer Grid
 - Grobal Grid Forumによる標準化
 - IBM, HP, SUNなどの企業の参加





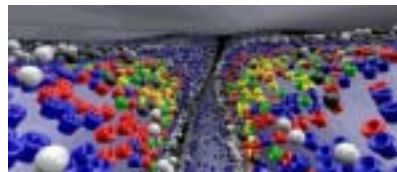
なぜ、今GRIDなのか

- スーパーコンへの反省
- 超高速ネットワークアプリケーションへの期待
- ユビキタスコンピューティングへの予感
- 米国、欧州のScienceにおけるIT技術の利用の進展



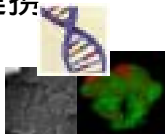
Technical challenge of Grid

- 大規模計算
 - 無限のPC
- 大規模データ管理・処理



HPC/Distributed Computing
Meets Internet

– データベース連携





世界の現状: 各種Gridプロジェクト

- 80年代 - 分散コンピューティング
- 90年代初期 - ギガビットテストベッド
 - 主にネットワーキングの研究
- I-Way 1995
 - アプリケーションの feasibilityが主
- Alliance (NCSA) Virtual Machine Room
- PACIs (NCSA/SDSC NSF National Technology Grid) 1998 ~
- NASA Information Power Grid 1999~
- ASCI DISCOM 1999~
- GriPhyN (Grid Physics Network), PPDG, 2000~
- eGrid (European Grid), (EU/CERN) DataGrid, 2000~
- ApGrid (Asia-Pacific Grid) 2000~
- NCSA-SDSC Distributed Terascale System 2001~, IDVGL (Distributed Virtual Grid Lab) 2001~

松岡



世界中で計画される ビッグプロジェクト

- US Blue ribbon panel on Cyberinfrastructure
 - ネットワークインフラおよびセンターインフラに4.3億ドルで、ソフトウェアおよびアプリサポートで全体で毎年6.5億ドル以上(750億円程度)投資する。(2年後に開始)
- EUの第六世代(6th Initiative)
 - インフラで3億Euro, その上のソフトウェアR&Dに3億Euroを4年間(つまり、650億円/4年)
- 日本 NaReGi (National Research Grid Initiative)
 - 5年間。全国にナノとグリッド基盤、二ヶ所のグリッド拠点を置き、100TFlopsを目指す。

Thanks to Dr. Matsuoka



日本、アジアの最近の動き (2002年)

- [ApGrid](#)
 - 5月14日–17日(台北)
- グリッド協議会
- [バイオグリッド始動](#)
- [PRAGMA](#)(Pacific Rim and Grid Middleware Assembly)
 - 7月11, 12日(韓国)
 - 1月23,24日(福岡) with APAN, 特定(C)



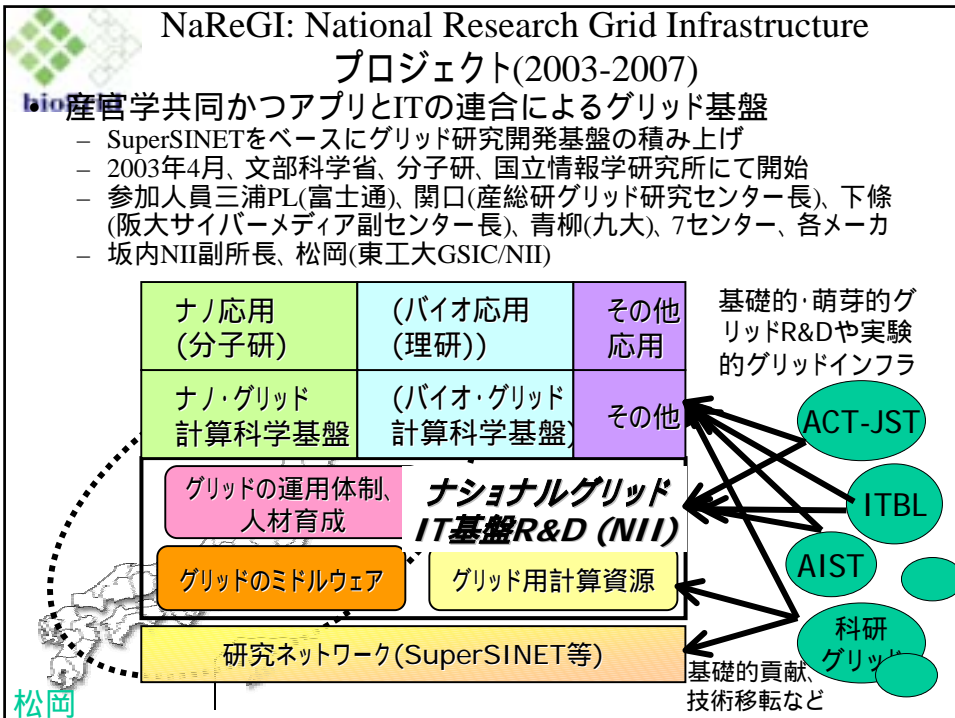
日本、アジアの最近の動き (2003年)

- バイオグリッドシンポジウム(東京) 3月3日
- [PRAGMA](#)(Pacific Rim and Grid Middleware Assembly)
 - 1月23,24日(福岡) with APAN, 特定(C)
- 特定(C)成果報告会
 - 1月22日(NII) with Fran Berman
- GGF7 at Shinjuku (3/4/2003-3/6/2003)
 - Life Science mini workshop (3/5, 6)
- NaReGi始動
- ccGrid (5/2003)



わが国での動き

- 文科省 特定領域研究(C)「ITの深化の基盤を拓く情報学研究」[2001/10-2005/3]
 - A05:最先端の情報通信システムを活用した新しい研究手法
松岡聡、下條真司
- NII スーパーSINET [2001/10-2005/3]
 - 7つのスーパーコンピュータセンターをGRIDで接続
 - NECはSX5にGLOBUSを移植
- 日本原研他 ITBL (IT based Laboratory)
 - ITによって科学研究を進化させる。
- NaReGi (National Research Grid Initiative)
- ビジネスグリッド





グリッド基盤ソフト&ネットワーク研究

グリッド基盤ソフト

グリッド大規模計算プログラミング

Grid RPC, Grid MPI

産総研

グリッド環境における資源管理

スーパースケジューラ、ブローカー、クラスタスケジューラ
情報モニタリング、資源管理ツール

東工大, F, H, N

グリッドソフトの統合・運用技術

ミドルウェアの検証・統合・配布、
運用管理者ツール、セキュリティ、認証

NII, F, H, N

グリッドアプリケーション開発用ソフトと環境

ワークフローツール&言語、可視化、PSE

NII, F, H, N

ネットワーク

ネットワーク通信基盤

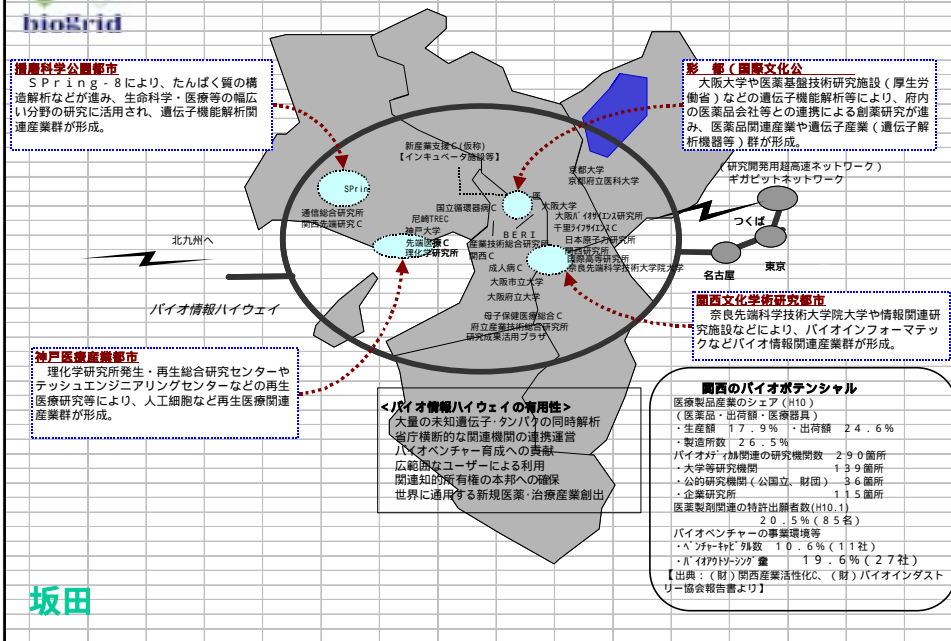
広域計測フレームワーク/ツール、マルチクラス/マルチパス
経路制御、GridFTP向け通信プロトコル、性能評価

阪大, 九工大

三浦



バイオ情報ハイウェイによる産業クラスターについて (イメージ)



坂田

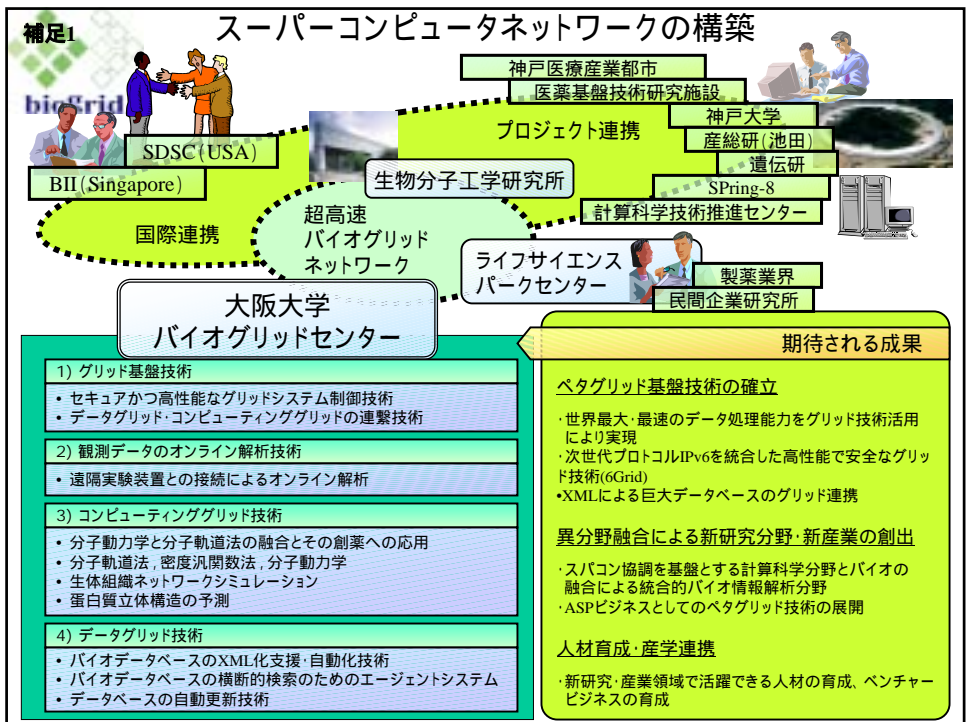
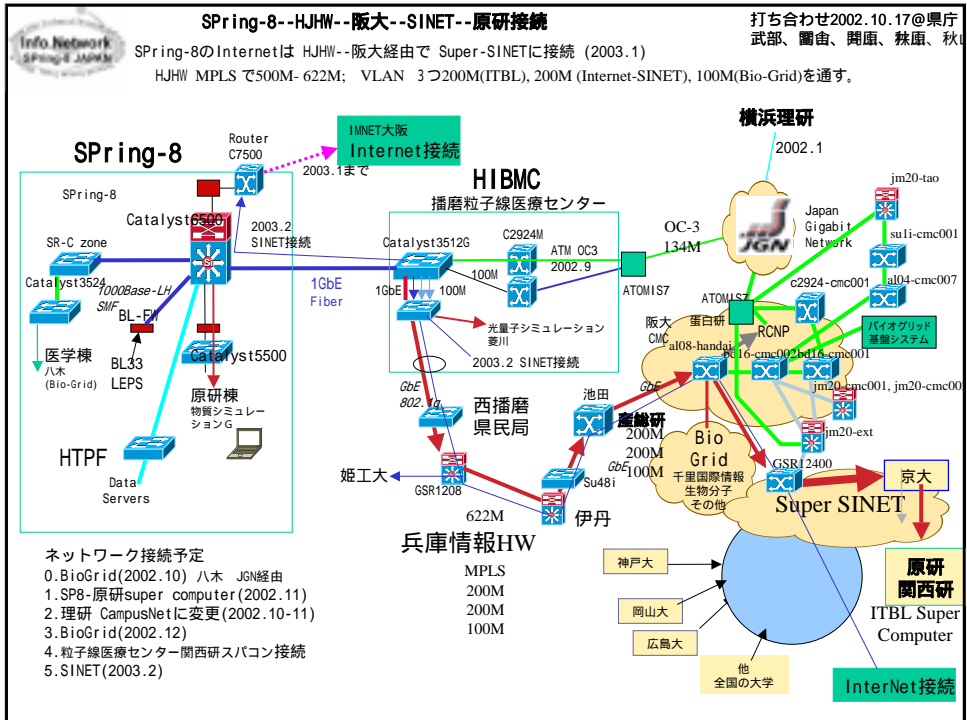


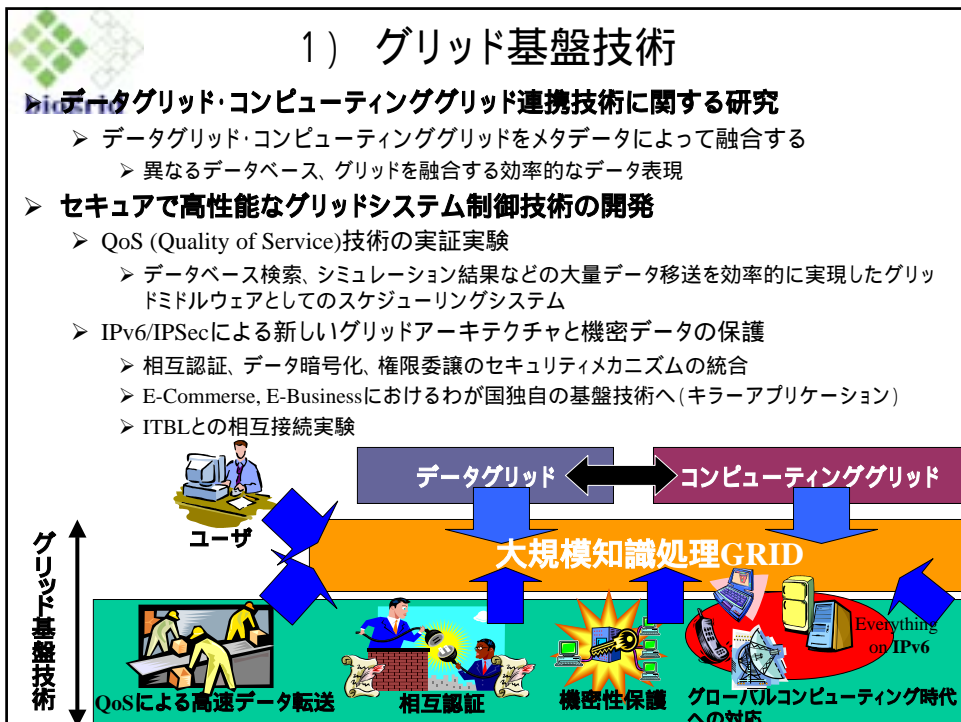
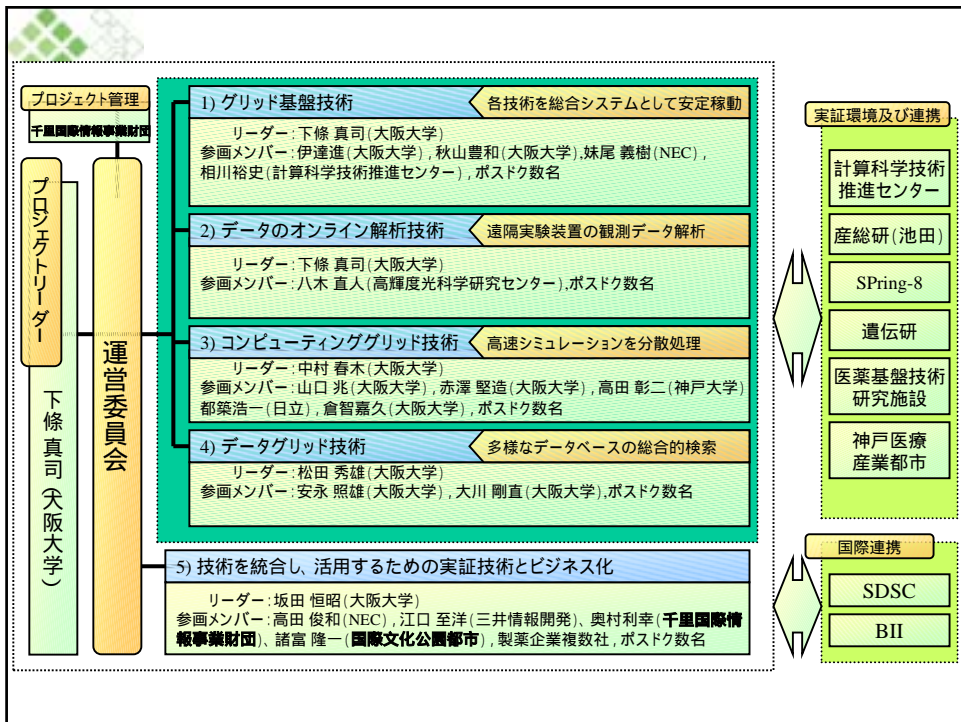
従来の衛星回線を用いた 遠隔観察システム



バイオグリッド・センター構想

- 超高速ネットワーク、超並列計算機、大規模データベースを核とするリエゾンセンター
- 情報通信技術(IT)の蓄積である大型計算機センターを利用
- 情報通信技術の本格的な取り込みと人材の育成, 関連機関のネットワーク化、ビジネス化を行う
- Gridによる国際協調





2) コンピューティング・グリッド技術 高速シミュレーションの異種大規模計算機群による分散処理とデータグリッド、Intelligent Gridとの融合

オープンソース開発環境の構築と並列化技法の開発

・分子軌道法、密度半閉数法、分子動力学法、ネットワークシミュレーションなどの計算ソフトウェア(大阪大学等で開発)を**オブジェクト指向のオープンソース・プログラム**として整備。

・グリッド上での**並列化の推進**。
・**コンピューティング・グリッドを活用する新たなソフトウェアの開発とそのデータグリッドとの融合**

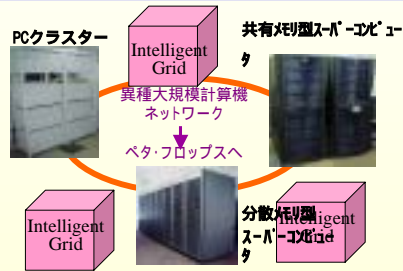
- ・in silico **薬物スクリーニング**
- ・薬物・受容体の電子状態および反応経路解析
- ・生体分子立体構造予測
- ・細胞内・生体組織のネットワークシミュレーション

データグリッド・コンピューティンググリッドの協調

・データグリッドとの協調によるバイオデータ解析・**データマイニング** (バイオ情報の新たな属性や関連性の探索)

コンピューティンググリッド・ユーザーのためのプラットフォーム構築

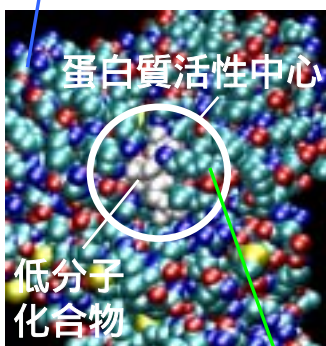
・具体的な**新産業創生**の場をベンチャー企業等へ提供



グリッドによる分子動力学と分子軌道法の融合と その創薬への応用に関する研究

候補新薬と蛋白質受容体との相互作用エネルギーを高速且つ高精度に解析するため、異なるレベルのシミュレーション計算を複合化した**In silico**創薬ソフトウェアの開発と実証

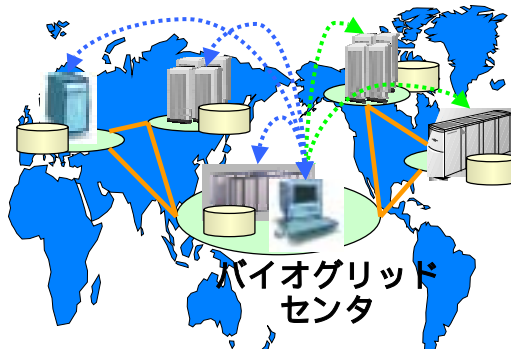
分子動力学の解析領域



分子軌道法の解析領域

分子動力学計算

分子軌道計算



グリッド環境での連成解析

何(日立)

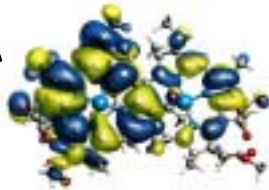


電子状態の計算:

Quantum Mechanics Calculations

by AMOSS (NEC基礎研が開発した非経験的分子軌道計算プログラム: 巨大な系の高精度計算が可能)

光合成活性中心
(1,100軌道)の
HOMO



生体高分子の計算:

Molecular Mechanics Calculations

by presto (阪大蛋白研が開発した分子動力学計算プログラム: 効率的な構造探索が可能 → 分子ドッキングへ)



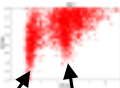
K⁺-イオンチャネルと
阻害剤のドッキング

蛋白質形状の予測:

(神戸大理が開発した粗視化モデルによる予測: 国際コンテストで高い評価)



Protein G

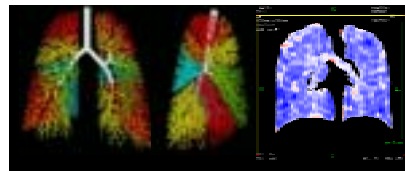


予測結果
RMSD = 3 Å

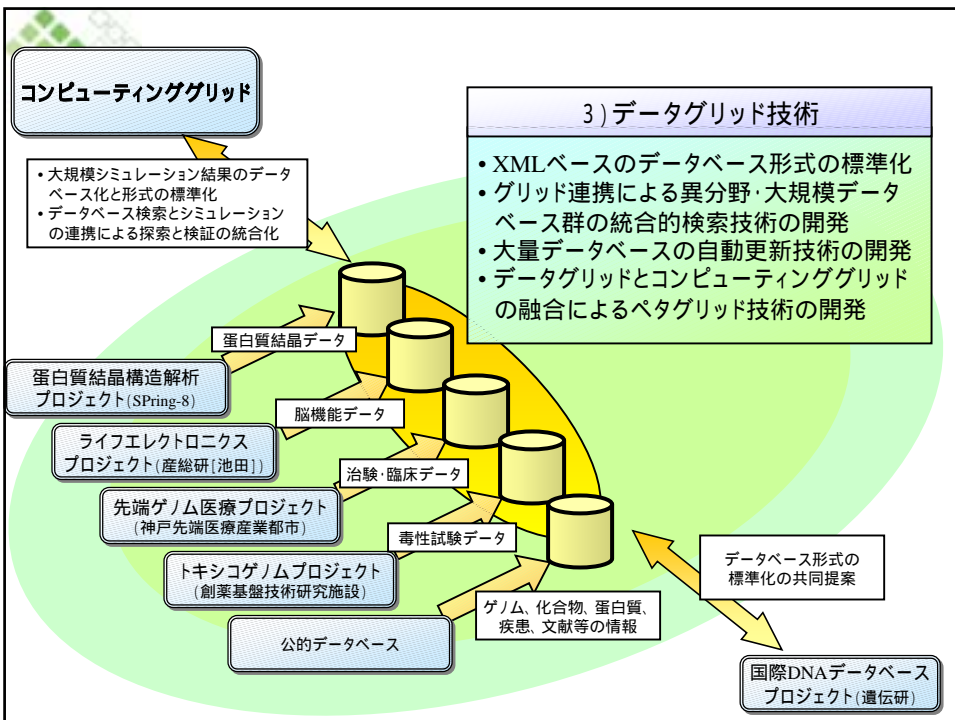


細胞・生体シミュレーション:

(阪大情報・医が開発中のシミュレーションプログラム: 臨床応用を目指す)



樹本-北岡法による換気分布のシミュレーション





「統合データベース」と「データベース連携」

統合データベース

- 検索対象となる複数のデータベースを静的に結合して、物理的に単一のデータベースを構成。

データベース連携

- 検索対象のデータベースはそのまま保持し、検索時に各データベースに問合せを出して、その結果を動的に結合する。

	システム構築コスト	検索コスト	DB更新コスト
統合DB	× (統合DB構築に膨大なコスト)	(検索は通常のDBと同じ)	× (検索対象DB更新のたびに大きな更新コストが発生)
DB連携	(検索対象DBをそのまま保持)	(検索時に各DBの検索結果の結合が必要)	(検索対象DBの更新をそのまま反映させるだけ)

データベース書式の異種性

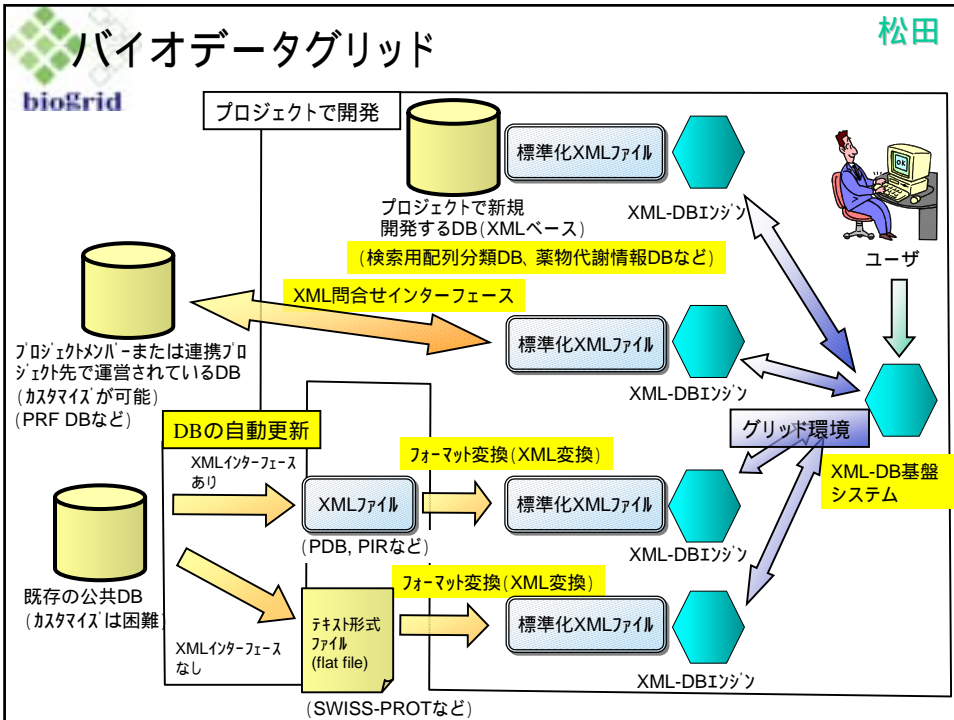


1D DNA塩基配列 (DDBJ)

タンパク質アミノ酸配列 (SWISS-PROT)

タンパク質立体構造 (PDB)



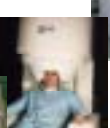
<p>LOCUS ECORBS 6197 bp DNA 30-OCT-1994</p> <p>DEFINITION E.coli ... rbsB ... transport system, ...</p> <p>ACCESSION M13169 M13517</p> <p>NID g147511</p> <p>KEYWORDS high affinity ribosome</p> <p>SOURCE E.coli K12 DNA.</p> <p>ORGANISM Escherichia coli</p> <p>REFERENCE 1 (sites)</p> <p>AUTHORS ...</p> <p>TITLE ...</p> <p>JOURNAL ...</p> <p>MEDLINE 84032513FEATURES</p> <p>gene 122..127 /gene="rbsB" /translation="MKKGTVLNS...</p> <p>ORIGIN 94 bp upstream of BclI 1 ctcagggttcg aaatctaac</p>	<p>ID RBSB_ECOLI ...</p> <p>AC P02925;</p> <p>DT 21-JUL-1986 (... CREATED)</p> <p>DT 21-JUL-1986 (... UPDATE)</p> <p>DT 01-NOV-1995 (... UPDATE)</p> <p>DE D-RIBOSE-BINDING ...</p> <p>GN RBSB OR RBSP OR PRLB.</p> <p>OS ESCHERICHIA COLI.</p> <p>OC PROKARYOTA; GRACILICUTES; ENTEROBACTERIACEAE.</p> <p>RN [1]RP SEQUENCE FROM N.A. ...</p> <p>RX MEDLINE; 84032513.</p> <p>RA GROARKE J.M., MAHONEY W.C.; ZALKIN H., HERMODSON M.A.; J. BIOL. CHEM. 258:...</p> <p>RL J. BIOL. CHEM. 258:...</p> <p>CC -I- FUNCTION: INVOLVED IN ...</p> <p>KW TRANSPORT; SUGAR TRANSPORT; 3D-STRUCTURE.</p> <p>KW SEQUENCE 296 AA; ...</p> <p>SQ MNMKKLATLV SAVALSATVS ANA. ...</p>	<p>HEADER SUGAR TRANSPORT 23-SEP-9</p> <p>COMPND D-RIBOSE-BINDING PROTEIN</p> <p>COMPND 2 (G134R) COMPLEXED WITH</p> <p>SOURCE (ESCHERICHIA COLI)...</p> <p>SOURCE 2 EXPRESSION PLASMID)</p> <p>AUTHOR S.L.MOWBRAY,A.J.BJORKMAN</p> <p>REVDAT 1 26-JAN-95 1DRJ 0</p> <p>JRNL AUTH A.J.BJORKMAN</p> <p>JRNL TITL PROBING PR...</p> <p>JRNL TITL 2 RIBOSE-BINDING</p> <p>JRNL REF TO BE PUBLISHED</p> <p>JRNL REFN ASTM</p> <p>SEQRES 1 271 LYS ASP THR ...</p> <p>SEQRES 2 271 PRO PHE PHE ...</p> <p>...</p> <p>HELIX 1 A PRO 14 LEU ...</p> <p>HELIX 2 B PRO 43 LEU ...</p> <p>...</p> <p>ATOM 1 N LYS 1 x1 y1 z1</p> <p>ATOM 2 CA LYS 1 x2 y2 z2</p> <p>...</p>
--	---	---



4) データ遠隔収集技術

世界最高性能を誇る観測機器を超高速ネットワークで共有

- 超高压電子顕微鏡 (UHVEM)
 - 大阪大学超高压電子顕微鏡センター (大阪府茨木市)
- 脳磁計 (MEG)
 - 産業技術総合研究所ライフエレクトロニクスラボ (大阪府池田市)
- Spring-8
 - 高輝度光科学研究センター (西播磨)

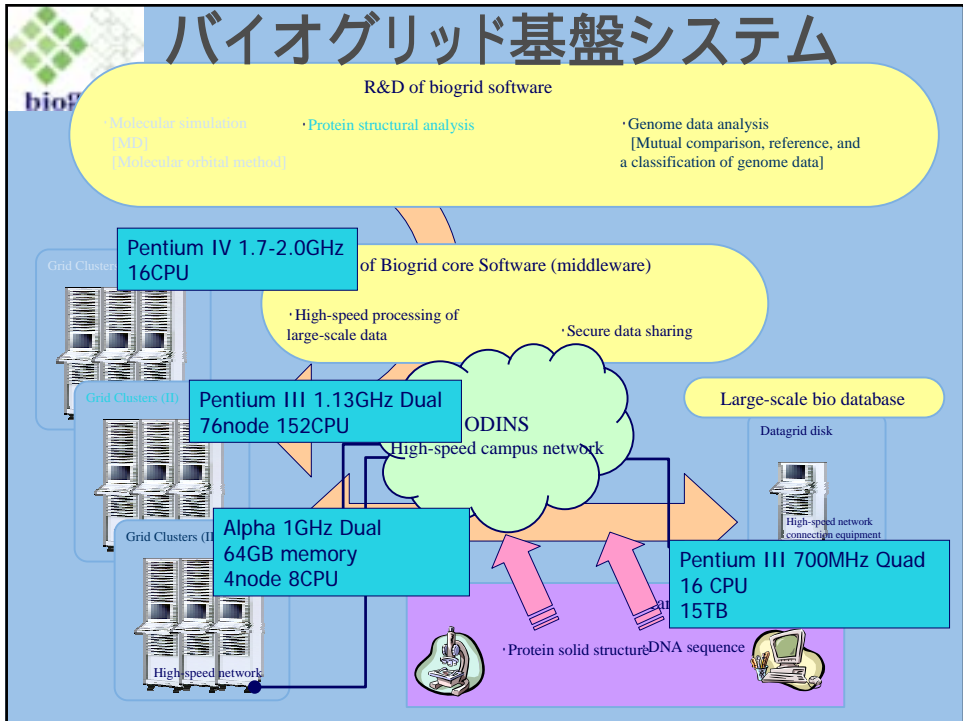
解析・シミュレーション

データ観測

データグリッド、Intelligent Gridへの融合



- 仮想研究空間の実現 (E-Science)
 - データ解析手法、シミュレーション手法の共有 (知識の共有)
 - シームレスなオンライン解析の実現
 - 観測機器の遠隔操作
 - 観測データ、可視化データ等の可視化実現へ





最新の国際連携

- Pragma 3rd (福岡) 共催
 - アジア太平洋地域でのアプリケーションでのグリッド連携を推進する国際連携
 - 日本のグリッドアクティビティの一つとして紹介
- GGF7でのLife Science Workshop

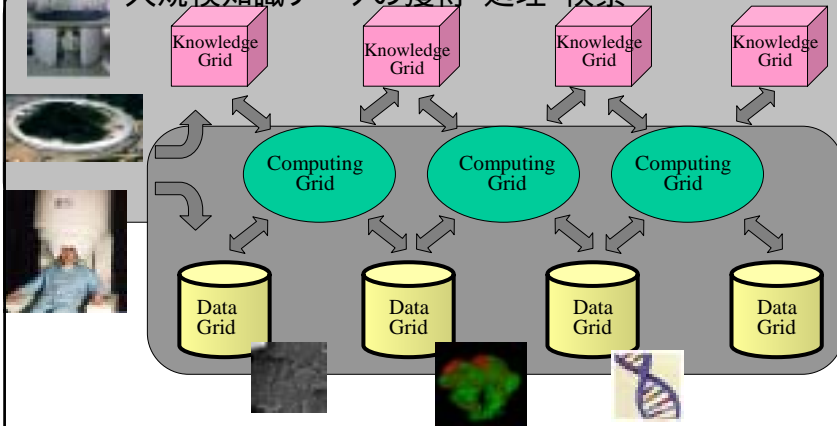





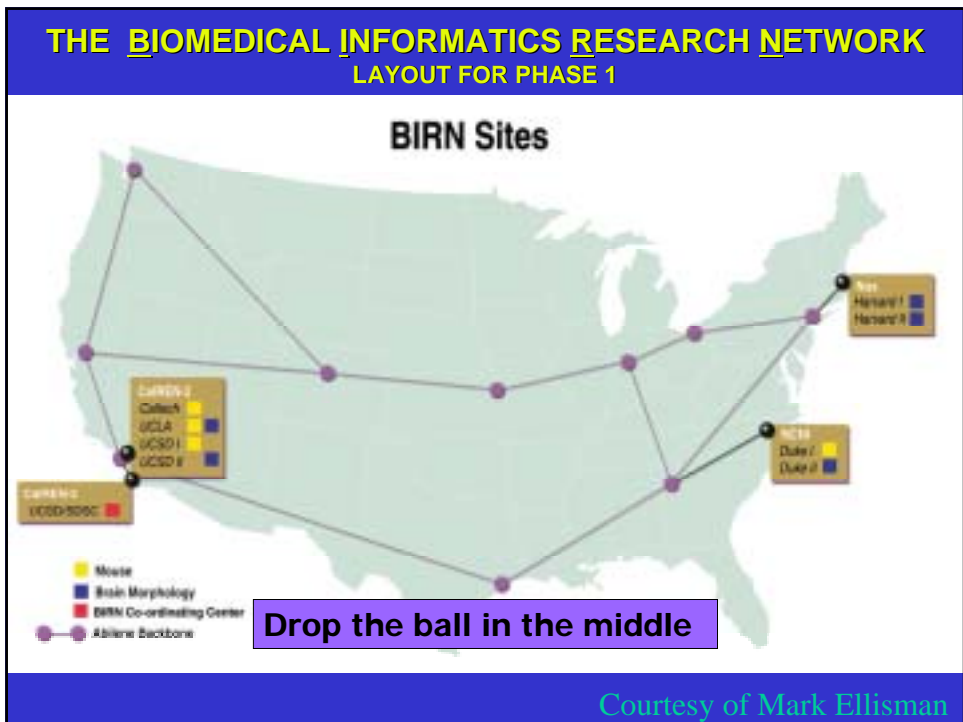
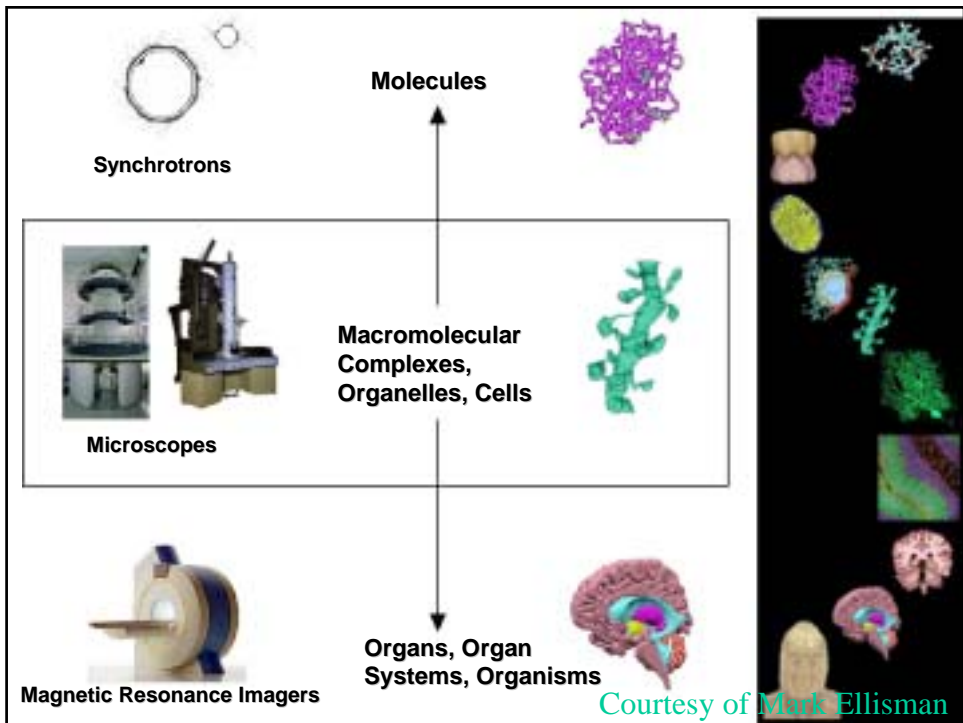
Technical Challenge of Grid

Coordinated Computing

大規模知識データの獲得・処理・検索

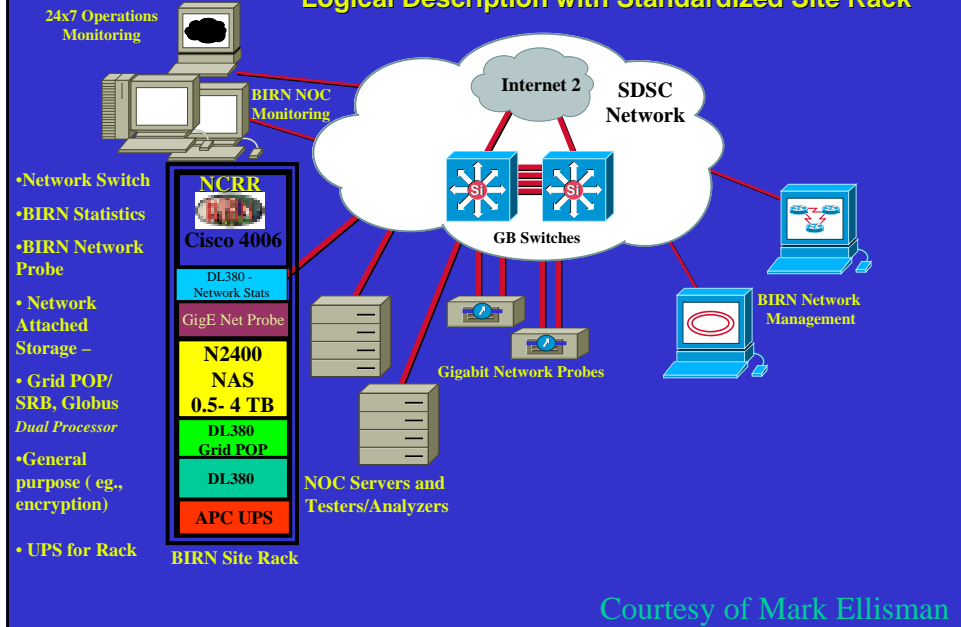


<http://birn.ncrr.nih.gov/>



BIRN Network Operations Center (NOC)

Logical Description with Standardized Site Rack



Gridのもうひとつの挑戦

- 科学とITのコラボレーション
- 異文化の融合
 - 言語
 - 評価システム
- 相互理解が重要

箱物や技術開発だけではうまくいかない



新しいナショナルプロジェクトのあり方

- 大学、国研と企業ががっぷり四つ
- 大学側の非常勤研究者(ポスドク)、雇用DRが中心。これに企業からの出向者を客員で向かえる
 - ベンチャー、新技術の温床とする
- 予算の制約をできるだけ緩和
- 開発した技術は戻さない
 - 開発の途中から、共同でベンチャーにしていく。



バイオグリッドの現状

- バイオグリッド研究会
- 文科省との契約。プロジェクトがスタート(6月1日)
- 創薬WGがスタート(7月27日)
- バイオグリッドビジネスセンター開設(8月1日)
- ネットワーク開通(8月20日)
 - BERI, BGBC, 阪大、産総研

プロジェクトの一年

- 基盤G
 - 安全で効率のよいグリッド基盤ソフトの開発 (6 Grid, GSI-SFS)
 - 応用システムのグリッド統合(MEGGrid)
 - 基盤システムの構築<hard+soft>
- コンピューティングG
 - 各種シミュレーションソフトのグリッド化とシームレスな結合
- データグリッドG
 - データベースの共通形式の開発 (XML標準形式)、プロトタイプの開発
- データのオンライン解析
 - 電子顕微鏡の遠隔操作、初のHDTV伝送
- World-wide Collaborative Research Activity
 - SC2002, PRAGMA
 - GGF Life-Science workshop

本プロジェクトの目標

