

‘03.08.19 グリッド協議会 第3回ワークショップ

グリッド指向のQM/MM連成 シミュレーション技術(*BioPfuga*)

(株)日立製作所 基礎研究所
何 希倫

目次

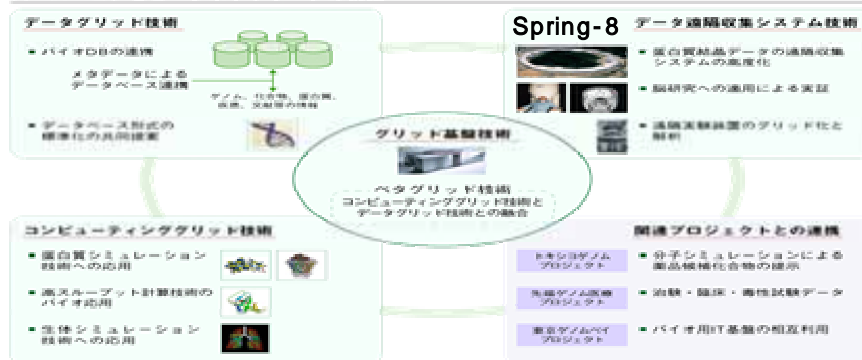
- 1 . バイオグリッドプロジェクト(biogrid-PJ)のご紹介
- 2 . 創薬プロセスにおけるQM(Quantum Mechanics)/
MM(Molecular Mechanics)技術の位置づけ
- 3 . QM/MM連成シミュレーション
- 4 . *BioPfuga*の技術的特徴
- 5 . プロト版の動作検証と実行性能の評価

文部科学省ITプログラム スーパーコンピュータネットワークの構築



スパコンネット上に分散した観測装置や、巨大で多様なデータベースを統合的かつ安全に連携処理するデータグリッド技術、そして、それぞれのデータベースの有機的な連携利用や多大な計算資源を必要とするデータ処理の橋渡しを実現するコンピューティンググリッド技術を開発する。

<http://www.biogrid.jp/>より転載

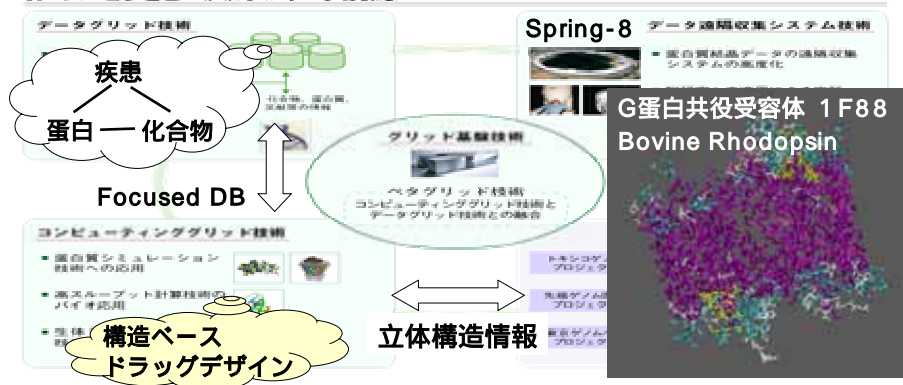


文部科学省ITプログラム スーパーコンピュータネットワークの構築



スパコンネット上に分散した観測装置や、巨大で多様なデータベースを統合的かつ安全に連携処理するデータグリッド技術、そして、それぞれのデータベースの有機的な連携利用や多大な計算資源を必要とするデータ処理の橋渡しを実現するコンピューティンググリッド技術を開発する。

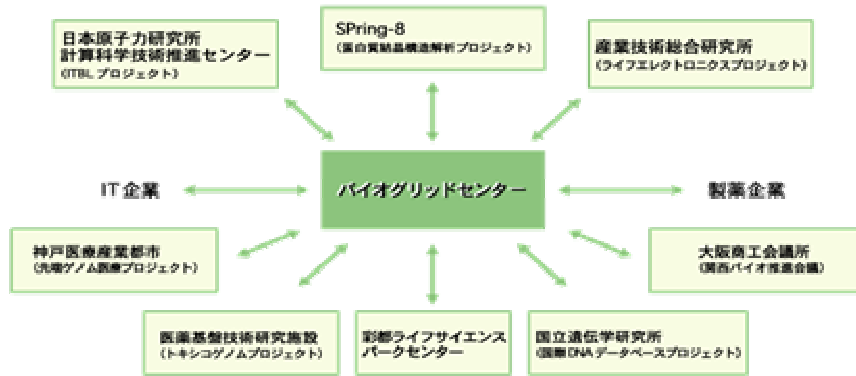
<http://www.biogrid.jp/>より転載



バイオ関連研究機関との連携協力体制



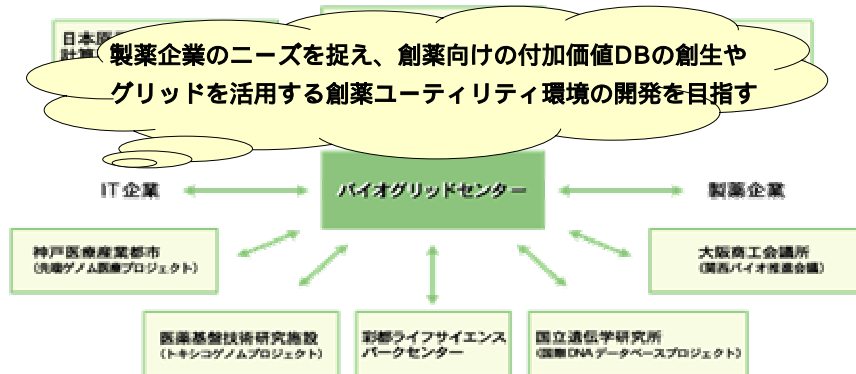
- ・技術開発に結びつく研究シーズよりの部分で情報関連企業の研究者が携り、最先端の技術動向を反映した研究開発を推進。
- ・製薬企業7社からなる創薬研究会を組織し、創薬の現場からの実践的評価を実施し、シーズとニーズのバランスを念頭においた研究開発を推進。



バイオ関連研究機関との連携協力体制

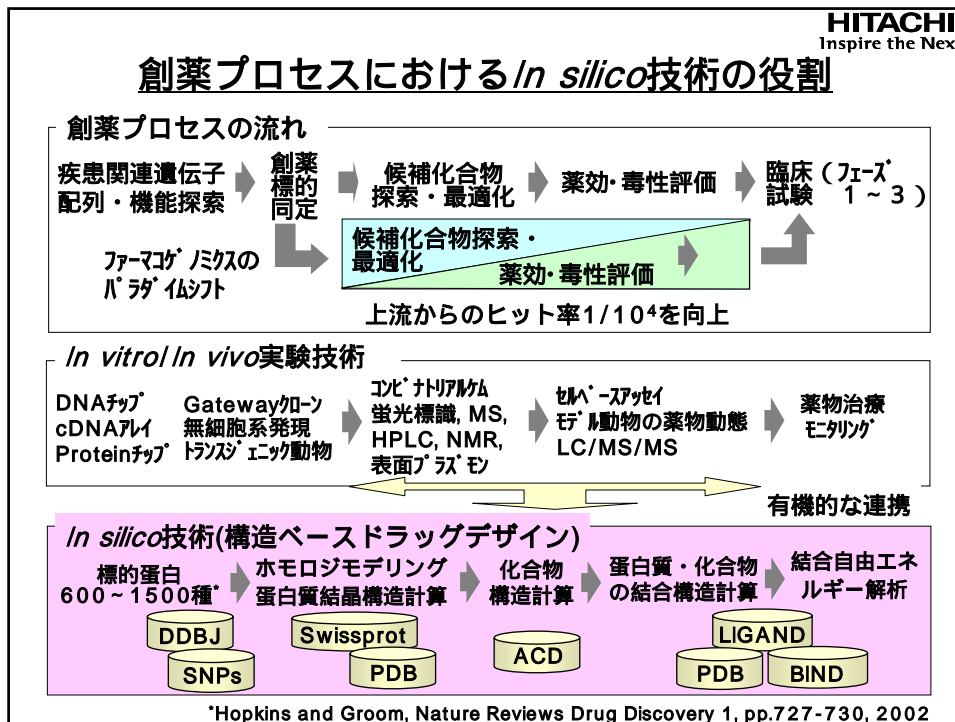


- ・技術開発に結びつく研究シーズよりの部分で情報関連企業の研究者が携り、最先端の技術動向を反映した研究開発を推進。
- ・製薬企業7社からなる創薬研究会を組織し、創薬の現場からの実践的評価を実施し、シーズとニーズのバランスを念頭においた研究開発を推進。



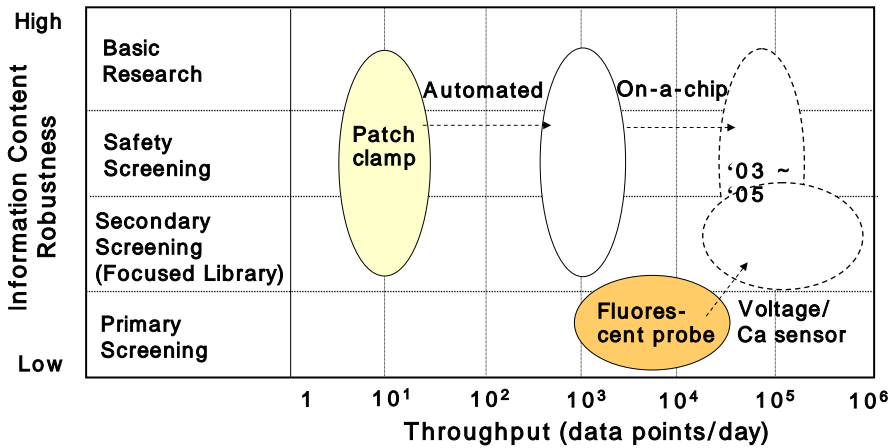
目次

- 1 . バイオグリッドプロジェクト(biogrid-PJ)のご紹介
- 2 . 創薬プロセスにおけるQM/MM技術の位置づけ
- 3 . QM/MM連成シミュレーション
- 4 . *BioPfuga*の技術的特徴
- 5 . プロト版の動作検証と実行性能の評価



In vitroアッセイ技術とIn silico技術の動向

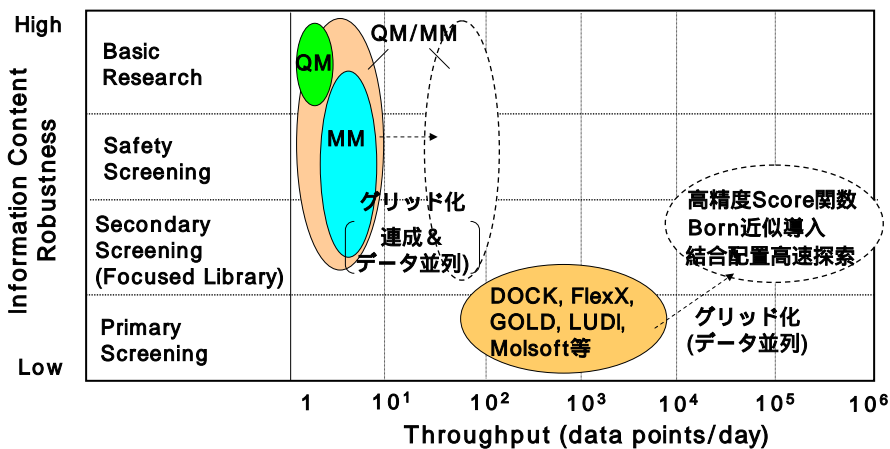
- HCS化(High Content Screening)とHTS化(High Throughput Screening)の潮流
- 高度情報を測定するPatch clampは自動化やオンチップ化によりHTS化



Jia Xu et al., Drug Discovery Today, Vol. 6 No. 24, pp.1278-1287, 2001

In vitroアッセイ技術とIn silico技術の動向

- HCS化(High Content Screening)とHTS化(High Throughput Screening)の潮流
- QMとMMを有機的に連結した連成シミュレーションを構築しHCSを追究

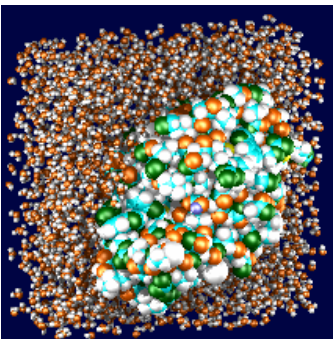
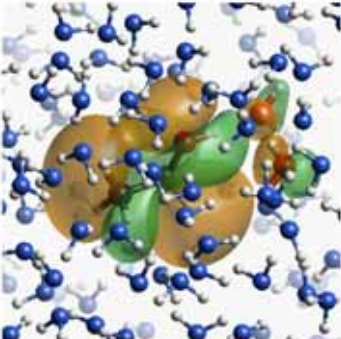


□ : 10TF(10GF × 1000PE)

目次

- 1 . バイオグリッドプロジェクト(biogrid-PJ)のご紹介
- 2 . 創薬プロセスにおけるQM/MM技術の位置づけ
- 3 . QM/MM連成シミュレーション
- 4 . *BioPfuga*の技術的特徴
- 5 . プロト版の動作検証と実行性能の評価

MMとQM手法の比較

項目	MM (Molecular Mechanics)	QM (Quantum Mechanics)
特徴	原子の相互作用パラメータを設定 熱力学的な水和構造アサブル	原子位置から電子状態を決定 活性解析可, 水和構造解析難
物理モデル 蛋白質 化合物		
水溶液効果		
計算オーダー		
構造最適化		
代表プログラム	AMBER, CHARMm, Presto-X(cosgene)他	Gaussian, GAMESS, AMOSS, FMO, Protein-DF他

MMとQM手法の比較

項目	MM (Molecular Mechanics)	QM (Quantum Mechanics)
特徴	原子の相互作用パラメータを設定 熱力学的な水和構造アングル	原子位置から電子状態を決定 活性解析可, 水和構造解析難
物理モデル 蛋白質 化合物	Bond, Angle, vdwパラメータ (実験データにフィット) Coulomb, Torsionパラメータ (アミノ基単位のQM計算)	原子の基底関数 STO-3G, 3-21G, 6-31G*, ---
水溶液効果	溶媒水分子の粗視化 PME, BH, FMM, カット 時間スケールのマルチステップ化	電子相関(積分計算) HF, B3LYP, MP2, DFT, QCISD, FMO, ---
計算オーダー	$O(N) \sim O(N \log N)$ (原子数: N)	$O(N^2) \sim O(N^7)$ (基底数: N)
構造最適化	最急勾配, 共役勾配,	Truncated ニュートン
代表プログラム	AMBER, CHARMM, Presto-X(cosgene)他	Gaussian, GAMESS, AMOSS, FMO, Protein-DF他

目次

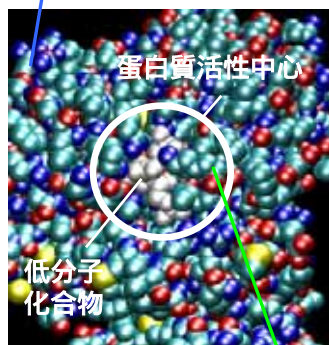
- 1 . バイオグリッドプロジェクト(biogrid-PJ)のご紹介
- 2 . 創薬プロセスにおけるQM/MM技術の位置づけ
- 3 . QM/MM連成シミュレーション
- 4 . *BioPfuga*の技術的特徴
- 5 . プロト版の動作検証と実行性能の評価

QM/MM連成シミュレーション技術 *BioPfuga* (*Biosimulation PlatForm United on Grid Architecture*)



分子動力学と分子軌道法の各プログラムを機能単位に分割、グリッド上で同時実行しXMLファイルを通じてデータ交換する連成プログラムを開発。

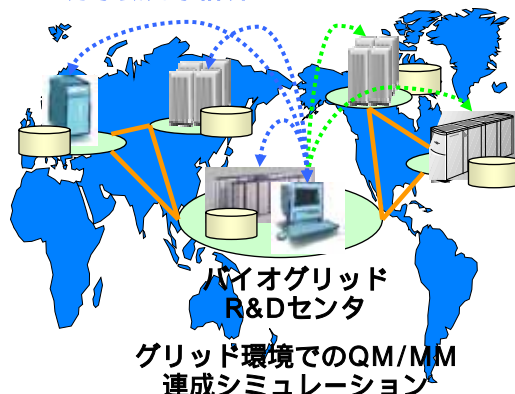
分子動力学の解析領域



分子軌道法の解析領域

分子動力学計算

分子軌道計算



QM/MMの物理モデル



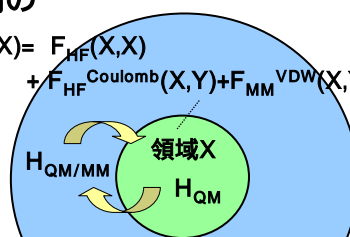
$$H_{Total} = H_{QM}(x,x) + H_{QM/MM}(x,y) + H_{MM}(y,y)$$

(a) 領域XまたはY内の原子・電子間の相互作用エネルギー計算

QM:HF, MM:Amber力場

$$F(X) = F_{HF}(X,X) + F_{HF}^{Coulomb}(X,Y) + F_{MM}^{VDW}(X,Y)$$

(b) 領域XとY間に跨る原子・電子間の相互作用エネルギー計算



$$F(Y) = F_{MM}(Y,Y) + F_{MM}^{Coulomb}(X,Y) + F_{MM}^{VDW}(X,Y)$$

(c) 領域XとYの境界領域の設定
境界領域に存在する原子の末端処理

クーロン結合, VDW結合から構成された境界面

QM並びにMMシミュレーションプログラム



蛋白質複合体シミュレーションプログラム

prestoX-basic(Cosgene)

In silico drug screeningを目的に、産業技術総合研究所JBIRC、大阪大学蛋白質研究所、横浜市立大学、日立製作所、富士通で共同開発中の古典力学に基づく分子動力学計算プログラム。

電子状態シミュレーションプログラム

AMOSS (Ab initio Molecular Orbital Simulation on Supercomputer)

NEC基礎研究所が開発した非経験的な分子軌道法計算のプログラム。巨大な系の並列下における高速計算、高精度計算が可能。CI(Configuration interaction:配置間相互作用)計算も開発中。

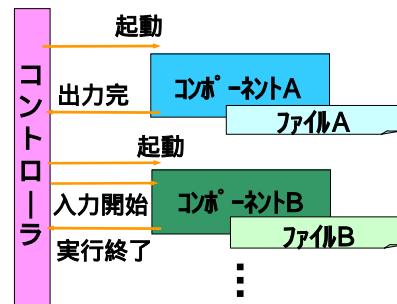
連成コントローラ



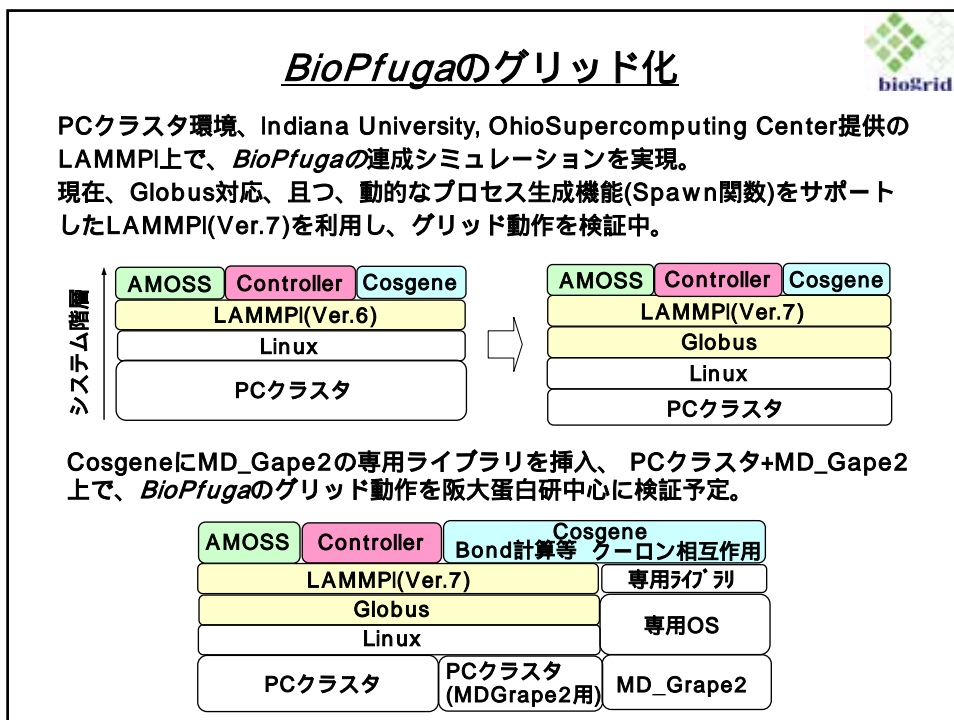
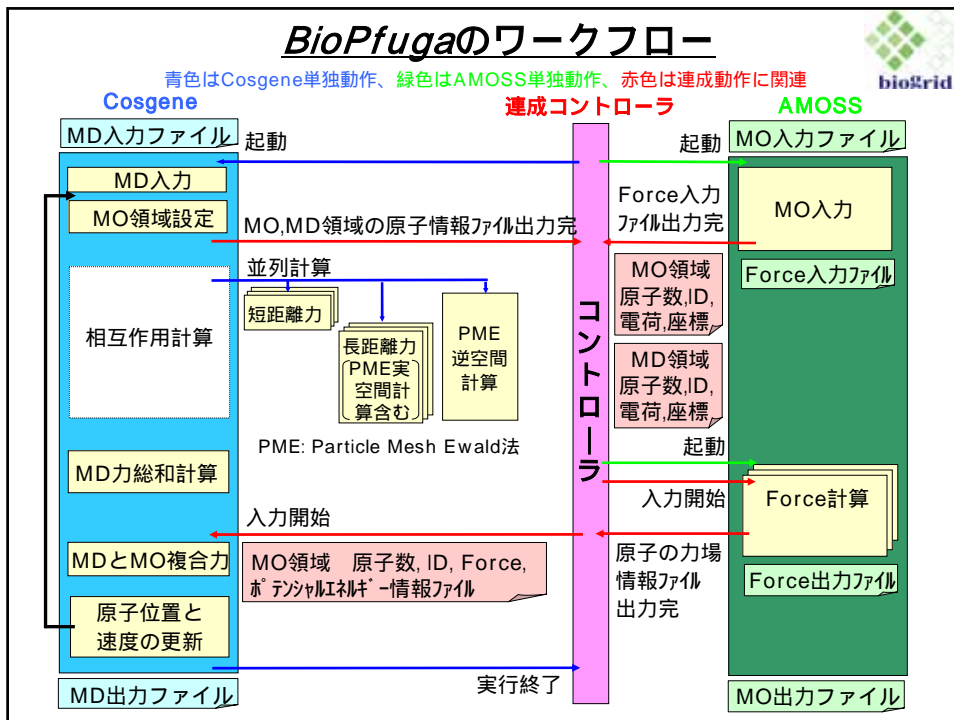
グリッド上で、複数のプログラムを連結し同時実行する方式として、各機能単位プログラムの起動やXMLファイルの入出力を管理するコントローラを開発。

基本機能

- (a)コンポーネントの起動と
実行終了メッセージの受信
- (b)XMLファイルの入力開始と
出力完了メッセージの送受信
- (c)XMLファイル名の情報提供



MPI(LAM/MPI, Indiana University 提供のフリーソフト)ベースで開発。XMLファイルは、共有ディスク上において入出力可。

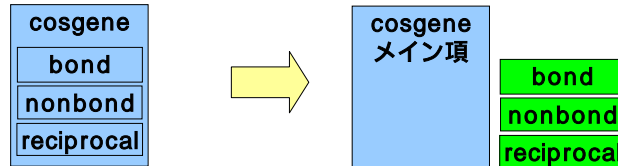


プログラムのコンポーネント化 (1)

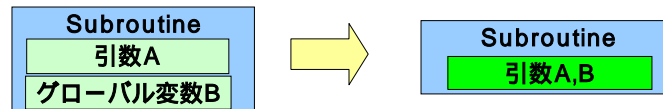


プログラムの拡張性や再利用性を高めるため、機能単位プログラムをメイン (パーツ) 化し、UDSXMLライブラリやMPIライブラリに対応するインターフェースとして、ラッパー関数を開発。

Cosgeneでは、Cosgeneメイン項、短距離力計算項、長距離力計算項とPME法の逆格子空間計算項に分離したパーツプログラムを開発。



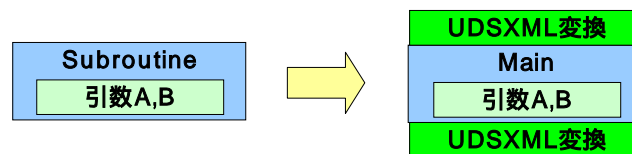
(a) 各機能単位のプログラムにおいて、変数の独立性を高めるため、グローバル変数を引数渡しに変更する。



プログラムのコンポーネント化 (2)

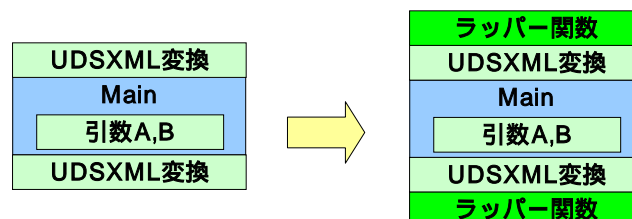


(b) UDSXMLライブラリを用いて入出力データフォーマットを共通化。プログラムを独立に実行可能なメイン(パーツ)プログラム化する。



(c) 連成コントローラとパーツプログラム間のメッセージを送受するラッパー関数を組み込み、コンポーネント化。

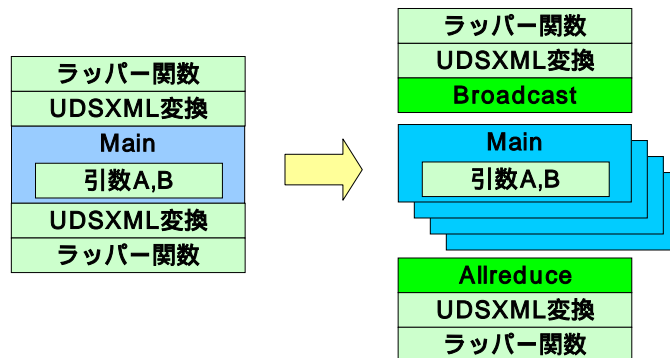
メッセージ: パーツの実行完了、XMLファイル入力開始と出力完了、並びに、XMLファイル名の情報取得等である。



プログラムのコンポーネント化 (3)



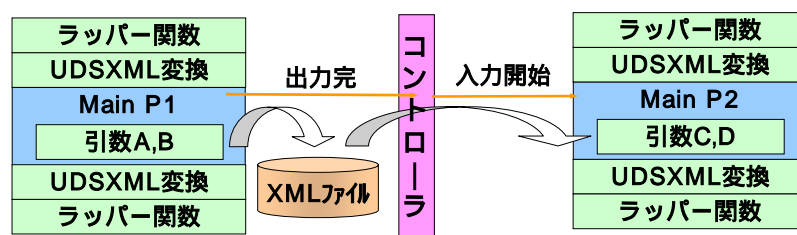
- (d) コンポーネントプログラムの並列計算では、MPIの動的なプロセス生成機能(Spawn関数)を用いて、親コンポーネントを起動。親子間のデータ交換は、Broadcast関数やAllreduce関数を用いて複数プロセスを分散実行する。



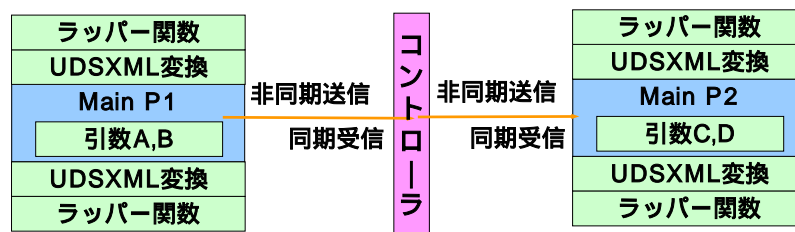
ラッパー関数



- (a)UDSXMLインターフェース(計算データの再利用や高付加価値化)



- (b)MPIインターフェース(計算処理高速化、データ交換時間の短縮化)



UDSXMLインターフェース (Universal Data Set-eXtensible Markup Language)



```
call udsxml_put 又は int udsxml_get
(const char *url, /*書き込み又は読み取りXMLファイルのURL*/
 void *data, /*物理データのバッファ*/
 const int size, /*1つの物理データのバイト数*/
 int count /*物理データの数*/
 int *p_form, /*物理データの記録形式*/
 char *p_unit, /*単位の格納先指定*/
 const int unitsz, /*単位の格納先サイズ*/
 char *p_cost, /*内容説明の格納先指定*/
 const int contsz, /*内容説明の格納先サイズ*/
 char *p_cmnt, /*コメントの格納先指定*/
 const int cmntsz, /*コメントの格納先サイズ*/
 );
```

UDSXML入出力ファイル例



UDS-XML (form = Base64) (x4/3)

```
<?xml version="1.0" encoding="UTF-8"?>
<uds_data size="8" count="200" unit="AU" form="b64">
  <uds_content> sample_hxd.xml </uds_content>
  <uds_comment> Sample of UDS-XML (hexdec) </uds_comment>
  <uds_array_count>4</uds_array_count>
  <uds_array seq="1" element="character" length="76">
pLWkq6TipMik0qS1pLekzrrupMOKv6XQpaSlyqXqpcvhKW/pPKlqKXzpbOhvKXJp
LekxqS9pM6l<uds_array>
  <uds_array seq="2" element="character" length="76">
x6G8pb+k8qPYo82jzLfBvLCkx6XVpaGlpKXrpMvK3cK4pLmk66XXpe2lsKXppeCkz
qXGpbmlyKTH</uds_array>
  <uds_array seq="3" element="character" length="76">p
Lmho6XXpe2lsKXppeCkrMC1vu+ky8awpKSkxqSkpOu+7LnnpM+ks6TOyrikrMC1p
Lekr8bJpOGk<uds_array>
  <uds_array seq="4" element="character" length="40">xqSkpOukz6S6pMekuaG
jpMmkpqTHpLek56Smoak=</uds_array>
</uds_data>
```

目次

- 1 . バイオグリッドプロジェクト(biogrid-PJ)のご紹介
- 2 . 創薬プロセスにおけるQM/MM技術の位置づけ
- 3 . QM/MM連成シミュレーション
- 4 . *BioPfuga*の技術的特徴
- 5 . プロト版の動作検証と実行性能の評価

*BioPfuga*プロト版の動作検証



*BioPfuga*プロト版を開発、水溶液中のエタノールの構造解析を実施。PCクラスタ Pentium4(2.8GHz) + Gbitインターネット。

(a)MM計算

$C_2H_5OH + H_2O$ 228分子
(孤立系:全原子数919個)

水分子モデル: TIP4P

C_2H_5OH 近傍の H_2O 2分子:TIP3P

カノニカルアンサンブル, 300K

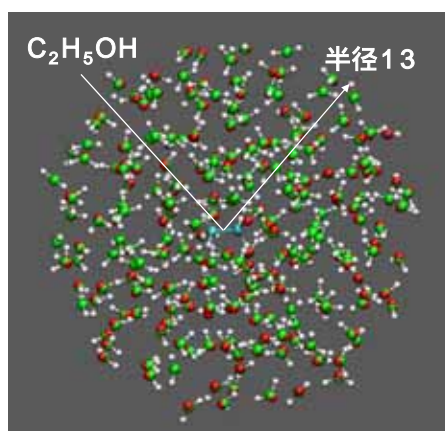
85.5ps(MD)+5ps(QM/MM)

クーロン相互作用はノンカットフ

(b) QM計算

$C_2H_5OH + H_2O$ 2分子

HuzinagaMINI-4(軌道数35基底)

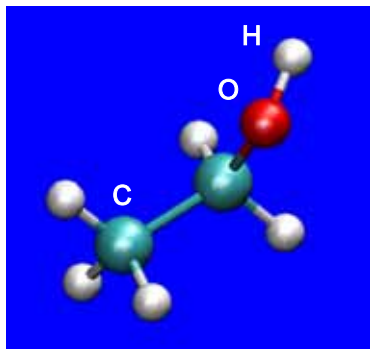


エタノールC-C-O-Hの捻じれ角分布(1)



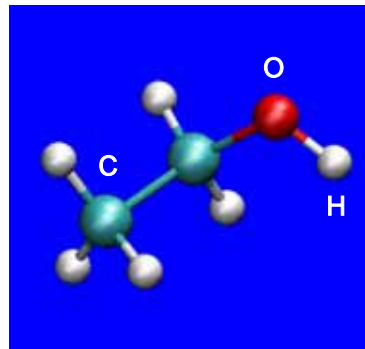
Gas層では、Trans構造はGauche構造に比べてエネルギー的に12kcal/mol低い。Trans構造とGauche構造の分布比は1:0.82。

Trans構造(二面角度0rad)



MD計算(85.5ps後)

Gauche構造(二面角度 $\pi/2$ rad)



QM/MM計算(MD85.5ps+2ps)

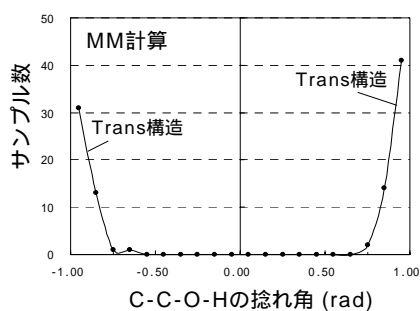
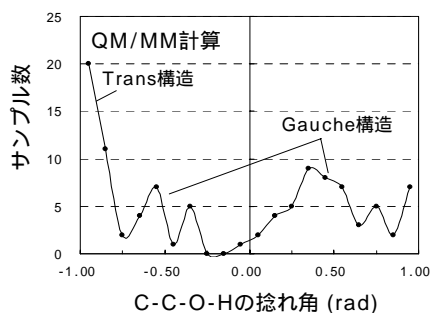
エタノールC-C-O-Hの捻じれ角分布(2)



Gas層では、Trans構造はGauche構造に比べてエネルギー的に12kcal/mol低い。Trans構造とGauche構造の分布比は1:0.82。

QM/MM計算では、Trans構造($\pm 1 \pm 0.1$) のピーク分布に対して、Gauche構造($\pm 1/2 \pm 0.1$) のピーク分布の相対比は0.9程度。

MM計算では、Trans構造のみに分布(Torsionパラメータの精度低)



実行性能評価



(a)UDSXMLインタフェース

連成1回のデータ転送量: Cosgene AMOSS: 約100KB
 AMOSS Cosgene: 約800B

Cosgene機能	計測時間(s)	AMOSS機能	計測時間(s)
MD計算	0.035	QM計算	10.171
データ送信	0.020	データ受信	0.076
データ受信	0.002	データ送信	0.003
QM/MM複合力計算	0.000	-	
合計	0.057	合計	10.250

(b)MPIインタフェース

連成1回のデータ転送量: Cosgene AMOSS: 約30KB
 AMOSS Cosgene: 約240B

Cosgene機能	計測時間(s)	AMOSS機能	計測時間(s)
MD計算	0.035	QM計算	10.158
データ送信	0.000	データ受信	0.003
データ受信	0.002	データ送信	0.000
QM/MM複合力計算	0.000	-	
合計	0.037	合計	10.161

BioPfuga連結方式のまとめ



項目		機能仕様
単体アプリ	プログラミング言語	Fortran, C
	プログラミングスタイル	オブジェクト指向コード (パーツ化)
	並列プログラミング方式	動的プロセス生成
連成アプリ	連成の制御プログラム	コントローラ(RPC)
	コンポーネント化	ファイル制御用ラッパー関数
	データ交換形式	UDSXMLインタフェース, (MPI)
	通信パターン	一方向, 双方向
	インターコネクション	LAMMPI, MPI2
通信プロトコル		TCP/IP
実装マシン, OS		Linux, 専用OS PCクラスタ, MD_Grape2



Acknowledgements

Haruki Nakamura, Yasushige Yonezawa
(*Institute for Protein Research, Osaka University*)

Toshikazu Takada, Toshihiro Sakuma, Kazuto Nakata (*NEC*)

Shinji Shimojyo, Hideo Matsuda, Tsuneaki Sakata,
Kizashi Yamaguchi (*Osaka University*)

Yoshifumi Fukunishi (*BIRC-AIST*)

Nobuyuki Nakajima (*IPR, Osaka University*)

Narutoshi Kamiya (*BERI*)

Takashi Kurosawa, Yuuki Imai (*Hitachi*)



謝辞

文科省 ITプログラム
スーパーコンピュータネットワークの構築(biogrid-PJ)
の一環として実施された研究成果の一部である。
ここに、感謝の意を表します。