

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

1. 2002 年度の具体的な研究計画

- (1) 生物情報解析研究センター等で開発した分子動力学法のプログラム presto-X Basic 版 (Cosgene) を大阪大学グリッド環境上に実装し各機能単位の計算処理時間を測定する。
- (2) 測定結果に基づき、短距離力計算、長距離力計算やフーリエ変換等の計算負荷の大きい機能の並列アルゴリズムを検討し、並列版プログラムを開発する。
- (3) 入出力ファイルを介した並列版プログラムを機能単位に分離する(パーツ化)のために、インターフェース仕様とデータフォーマットを統一する XML 仕様を検討し、粗粒度レベルでパーツ化する。単一グリッド上で各パーツを分散実行する制御プログラムを開発し、グリッド環境上のフィジビリティを検証する。
- (4) 古典力学に基づく分子動力学計算と分子軌道法による電子状態計算の異なるレベルのシミュレーション計算を複合化するためのモデル化や連結方式を検討する。

2. 2002 年度の進捗状況と研究成果

- (1) BioPfuga (Bioinformatics Platform united for grid architecture) における連成プログラムのプロト版開発 (1.(3), (4) に対応)

1) QM/MM (Quantum Mechanics/Molecular Mechanics) の連結モデル

溶媒に囲まれた蛋白質と候補化合物からなる系全体のハミルトニアンを下記に示す。

$$H_{\text{Total}} = H_{\text{QM}}(X, X) + H_{\text{QM/MM}}(X, Y) + H_{\text{MM}}(Y, Y)$$

グリッド指向の連成プログラムの開発を目指し、QM/MM 連結モデルを下記(a)-(c)の近似法によって類型化し、将来的に拡張開発が容易なプロトタイプを開発する(図 2.1)。

- (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
- (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
- (c) 領域 X と Y の境界領域の設定と境界領域に存在する原子の末端処理

本年度に開発した連結モデルを下記に示す。

- (a) では、領域 X 内は分子軌道計算プログラム AMOSS を利用した HF 計算、領域 Y 内は Cosgene を利用した AMBER 力場を用いる。
- (b) では、 $H_{\text{MM}}(X, Y)$ と $H_{\text{QM}}(X, Y)$ から構成される。 $H_{\text{MM}}(X, Y)$ では、Cosgene を利用して、AMBER 力場の短距離結合(ボンド, アングル, 捻じれ結合)と VDW 結合を計算する。 $H_{\text{QM}}(X, Y)$ では、

AMOSS を利用して、領域 Y から領域 X への Coulomb 相互作用として領域 X 内の電子波動関数を用いたハートリー積分を計算する。

(c)では、領域 X と Y の境界領域の原子間が VDW 結合又は Coulomb 結合の場合を扱う。

$H_{QM}(X, X)$ 、 $H_{QM/MM}(X, Y)$ 、 $H_{MM}(Y, Y)$ に対して、その微分計算から得られる力を $F_{QM}(X, X)$ 、 $F_{QM/MM}(X, Y)$ 、 $F_{MM}(Y, Y)$ と示す。領域 X の原子に働く力 $F(X)$ 、並びに、領域 Y の原子に働く力 $F(Y)$ は、次式で与えられる。ここで、プロトタイプでは、連成プログラムの構築を最優先するため、 $F(Y)$ の Coulomb 相互作用として $F_{MM}^{coulomb}(X, Y)$ を用いた。この近似は、物理精度に劣るが、将来的に $F_{HF}^{coulomb}(X, Y)$ に拡張することは容易である。

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

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

$$F_{MM}(Y, Y) + F_{MM}^{coulomb}(X, Y) + F_{MM}^{VDW}(X, Y)$$

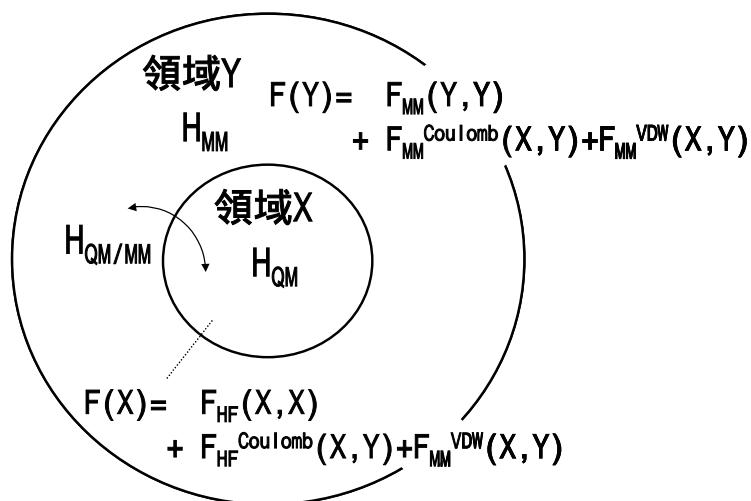


図 2.1 QM 計算と MM 計算の解析領域と力場モデル

2) 連成コントローラによるワークフローの制御

グリッド上で、連成シミュレーションを実行する連結方式として、各機能単位プログラムを制御するコントローラを開発した。連成コントローラの基本機能を下記に纏める。

コンポーネントプログラムの起動と実行終了メッセージの受信

XML ファイルの入力開始と出力完了メッセージの送受信

XML ファイル名の情報提供

本年度に開発したコントローラは、PC クラスタ等から構成された LAN 環境下で、コンポーネントの起動や XML ファイルを管理することができる。コンポーネントの起動とメッセージの送受信の機能は、MPI(Message Passing Interface)ベースで開発し、動的プロセス生成が可能な Indiana University の提供するフリーソフト LAM/MPI を用いた。又、XML ファイルは、共有ディスクに存在する場合に入出力アクセスが可能である。

次に、連成コントローラのもので、MD 計算プログラム Cosgene と MO 計算プログラム AMOSS を連結し、グリッド上で物理データを交換しながら同時実行する計算処理の手順を示す。

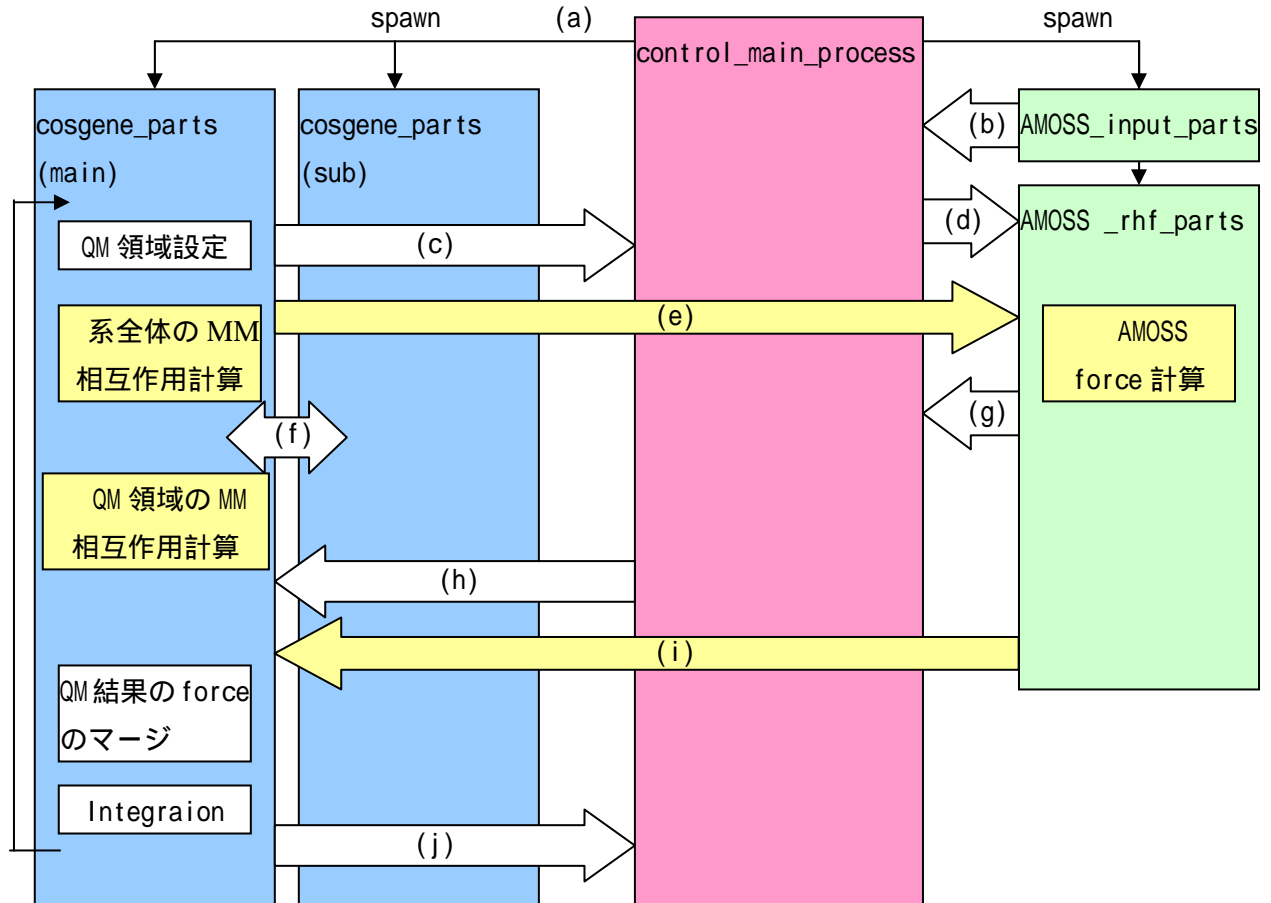


図 2.2 コントローラによる QM/MM 連成プログラムのワークフロー

- (a) コントローラからグリッド上に Cosgene と AMOSS_input を起動。
- (b) AMOSS_input コントローラへの実行終了メッセージ
- (c) Cosgene コントローラへの QM 領域と MM 領域ファイル出力完了メッセージ
- (d) コントローラ AMOSS_rhf の起動、QM 領域と MM 領域ファイル入力開始メッセージ
- (e) QM 領域と MM 領域の原子数、電荷、座標の put/get
- (f) Cosgene 内並列計算の broadcast, all reduce
- (g) AMOSS コントローラへの AMOSS の force ファイル出力完了メッセージ
- (h) コントローラ Cosgene の force ファイル入力開始メッセージ
- (i) QM 領域の原子数、force、ポテンシャルエネルギーの put/get
- (j) Cosgene コントローラの MD 実行終了メッセージ

物理データの交換には、XML(Extensible Markup Language)ファイルを用いて入出力を実行する XML ファイルインターフェース、並びに、連成コントローラを介してメッセージパッシングで物理データを送受信する MPI インターフェースを選択することができる。XML インターフェースでは、XML ファイルの入出力の同期を取るため、データの送り手はファイル出力完了時に、連成コントローラにファイル出力完了メッセージを送る。又、データの受け手は、連成コントローラからのファイル入力開始メッセージを受信した後にファイル入

力を実行する。MPI インターフェースでは、連成コントローラを介して、データの送り手は非同期送信を行い、データの受け手は同期受信を行う。

又、QM/MM の複合力 $F(X)$ と $F(Y)$ の計算は下記で与えられる。

$$\text{QM/MM の複合力} = \boxed{\begin{array}{c} \text{系全体の MM 相} \\ \text{相互作用計算} \end{array}} - \boxed{\begin{array}{c} \text{QM 領域の MM 相} \\ \text{相互作用計算} \end{array}} + \boxed{\begin{array}{c} \text{AMOSS} \\ \text{force 計算} \end{array}}$$

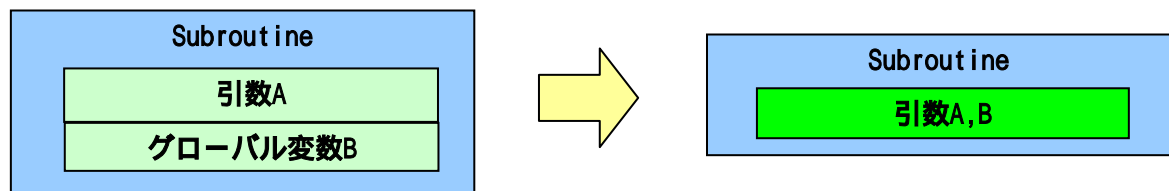
$F(X)$ では、 - より $F_{MM}^{VDW}(X,Y)$ 、 - より $F_{HF}(X,X) + F_{HF}^{coulomb}(X,Y)$ を求め、和をとる。

$F(Y)$ では、 - より、 $F_{MM}(Y,Y)$, $F_{MM}^{coulomb}(X,Y)$, $F_{MM}^{VDW}(X,Y)$ を求め、和をとる。

3) 機能単位プログラムのパーツ化とコンポーネント化

プログラムの拡張性や再利用性を高めるため、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した。UDSXML ライブラリは、大阪大学蛋白質研究所殿と日本電気殿が当 PJ で開発中のライブラリ群である。下記には、プログラムのパーツ化、並びに、コンポーネント化の手順を示す。

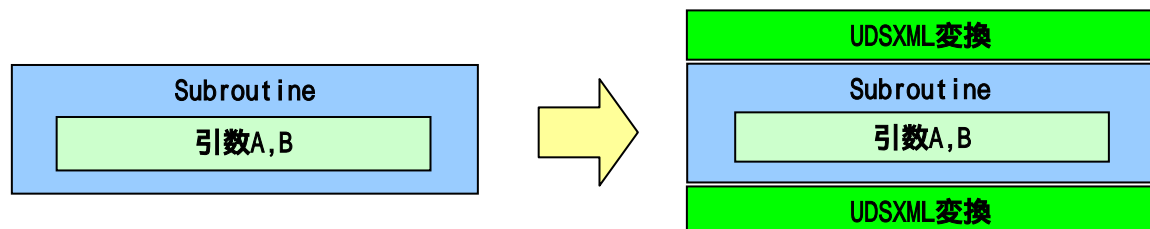
(a) プログラムを独立した疎粒度な機能単位に分離するため、サブルーティン内のグローバル変数を引数渡しに変更する。



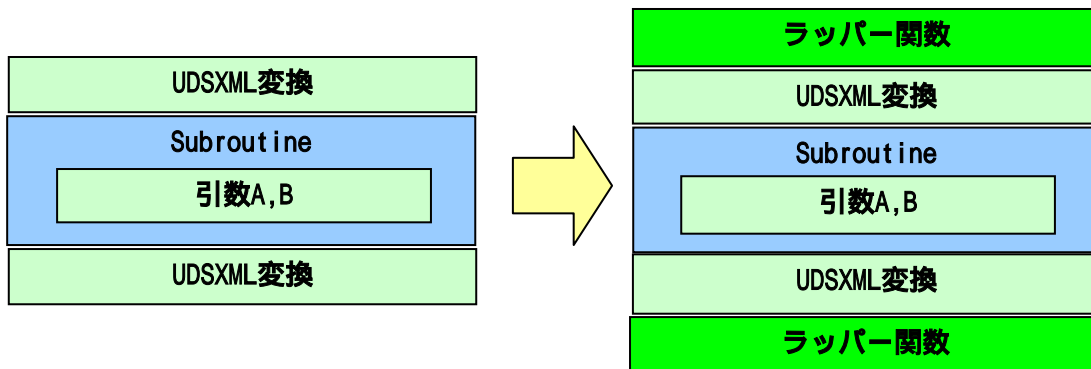
(b) 物理データを交換するためのファイルデータフォーマットを共通化するため、プログラム内で定義されたネイティブなデータフォーマットを XML ファイル形式に出力する UDSXML ライブラリを組み込む。UDSXML ライブラリは、シェアードライブラリ関数として動的にリンクを掛けて使用する。

(c) 物理データを交換するインターフェースとして、その中間データを活用する場合は UDSXML インターフェースを選択し、高速なデータ転送を重視する場合は MPI インターフェースを選択することができる。

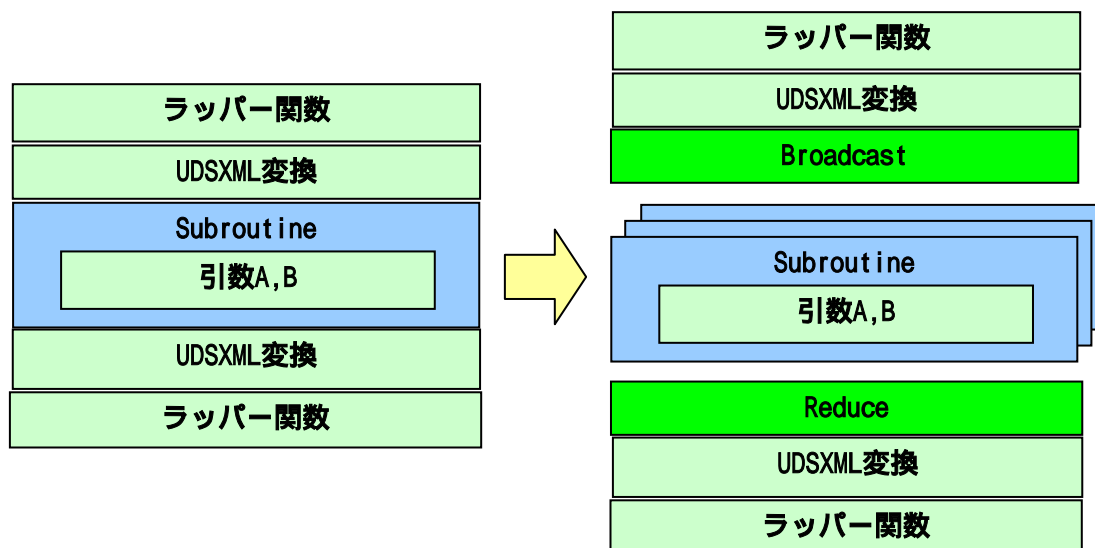
(d) 各機能単位プログラムを独立に実行可能なメイン(パーツ)プログラム化する。



(e) パーツプログラムに、連成コントローラとメッセージを送受信するラッパー関数を組み込み、連成シミュレーションの構成要素にコンポーネント化する。ラッパー関数は、連成コントローラから UDSXML インターフェースと MPI インターフェースで必要な情報、例えば、XML ファイル名、送受信先 ID、データ、個数、単位を受信する。



(f) コンポーネントプログラムの並列計算では、MPI ライブラリの Spawn 関数を用いて、連成コントローラから親コンポーネントを起動する。親コンポーネントは、ラッパー関数による実行終了メッセージを送信する。親子コンポーネント間のデータ交換は、MPI ライブラリの send 関数や receive 関数を用いて、子間のデータ交換は、Broadcast 関数や Allreduce 関数を用いる。



4) 連成プログラムの物理検証

QM/MM 連成プログラムの動作を確認するため、単一グリッドの PC クラスタ上で、エタノール C_2H_5OH の水和構造を解析し物理的な検証を行った。初めに、MD 計算として、エタノールの半径 13 Å に水分子 228 個を CAP 構造に配置した孤立系を設定する。エタノールに対しては、AMBER 力場を用いる。水分子に対しては、TIP4P モデルを用いた。一方、MO 領域内の水分子 2 個に対しては、TIP3P モデルを用いる。全原子数は 919 個になる。長距離 Coulomb 相互作用はノンカットオフモデルを使用する。系の粒子数、体積とエネルギーを一定にするカノニカルアンサンブルを用いた。平衡状態に至るプロトコルは、温度 3K 一定で 2ps (0.2fs、10000 ステップ)、3K から 283 K への温度の引き上げを 5ps (0.5fs、10000 ステップ)、カノニカルアンサンブルを元に温度 283 K 一定で 85.5ps (1.5fs、57000 ステップ) 計算した。次に、QM/MM 計算では、エタノールの近傍に位置した水分子 2 個を合わせて QM 領域として、他の水分子を MM 領域とする。QM 計算では、Huzinaga の MINI-4 を用いて軌道数は 35 基底である。シミュレーション時間 5ps (1fs、5100 ステップ) を計算し、50fs 毎の

計算結果(102 ステップ)を元に熱力学的アンサンブルを求めた。QM/MM と同様のプロトコルで、MM 計算による熱力学的なアンサンブルを求めた。尚、QM/MM 連成シミュレーションの実行は、大阪大学蛋白質研究所殿と日本電気殿と共同で行った。

エタノール C-C-O-H の捻じれ角分布を解析した。図 3 に示すように、QM/MM 計算の捻じれ角分布では、C-C-O から形成される面と C-O-H から形成される面の角度は、 π (または 0) にピークが現れる。これは、二つの面が平行であることを示し、Trans 構造と呼ばれている。又、 $\pi/2$ や $3\pi/2$ 近傍にもピークが見られるが、この場合は二つの面が垂直であることを示し、Gauche 構造と呼ばれている。ガス層では、Trans 構造は Gauche 構造よりも実験的に 0.12kcal/mol 低い測定結果が得られている。従って、室温 283 K において、Trans 構造に対する Gauche 構造の分布相対比は 0.82 程度と考えられる。捻じれ角分布の計算結果から、Trans 構造 $\pm \pi \pm 0.1$ のピーク分布数 40 に対して、Gauche 構造 $\pm \pi/2 \pm 0.1$ のピーク分布数 36 である。Trans 構造に対する Gauche 構造の分布相対比は 0.90 程度となり、測定結果の分布相対比と比較的良好一致が得られる。一方、MM 計算では、二つの面が平行になる Trans 構造分布のみにピークが現れており、Gauche 構造分布は存在しない。MM 計算では、捻じれ結合エネルギーは $V/2(1+\cos \theta)$ と表され、このエネルギー係数 $V=0.33\text{kcal/mol}$ の障壁を越えられず、ポテンシャルエネルギーの極小領域にトラップしていると考えられる。

AMBER 力場のボンド結合、アングル結合と VDW 結合は、単一アミノ基の固有分子振動数や等温・等圧条件下の体積密度の測定結果に一致するようにパラメータ値を決定する。一方、捻じれ結合は、QM 計算から得られるエネルギー値に合うようにパラメータ値を抽出している。捻じれ結合のエネルギー係数は数 kcal/mol オーダーと小さく、水分子や他分子との相互作用により変化を受けやすい。蛋白質と候補化合物間の相互作用は、捻じれを伴う構造緩和が生じるため、高精度な QM/MM 計算は極めて重要と考えられる。

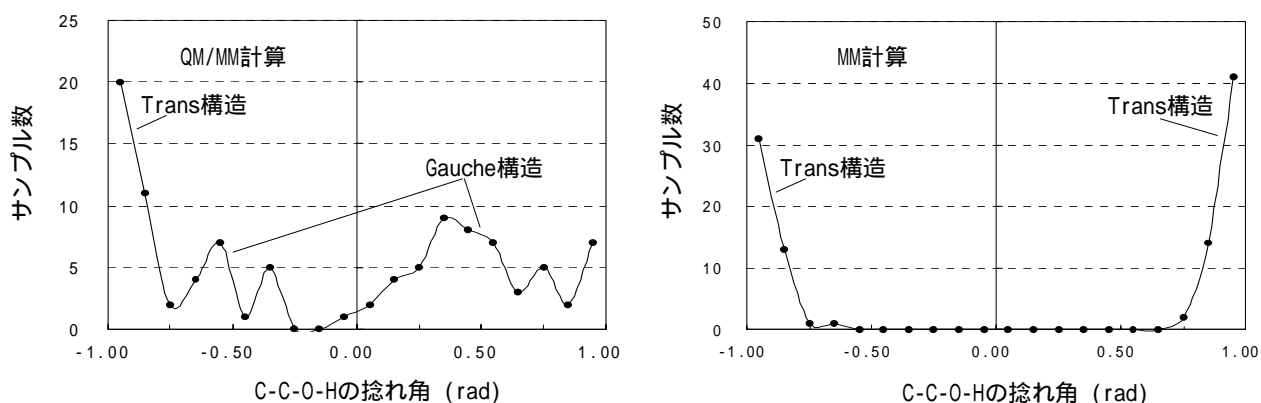


図 2.3 エタノール C-C-O-H の捻じれ角分布

5) 連成プログラムの実行性能の評価

Gbit イーサネットで繋がった Intel Pentium4 (2.8GHz)の PC クラスタからなる単一グリッド上で、QM/MM 連成プログラムを用いて、水溶液中のエタノールの水和構造を解析し、その実行性能を評価した。エタノールの周りに水分子 228 個を配置した孤立系、全水分子は TIP4P モデル、エタノールのみを QM 領域に設定する。全原子数は 921 個になる。長距離

Coulomb 相互作用はノンカットオフモデルを使用し、カノニカルアンサンブルを用いた。QM 計算では、RHF・ミニマム基底を用いて軌道数は 17 基底である。Cosgene と AMOSS を各々 1CPU 上で、1 タイムステップの実行性能を評価した。物理データの交換として、XML ファイルを用いた UDSXML インタフェース、並びに、Gigabit イーサネット上のメッセージパッシングを用いた MPI インタフェースの場合について計測した。

(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

表 2.1 UDSXML インタフェースを用いた QM/MM 連成シミュレーションの実行時間

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

表 2.2 MPI インタフェースを用いた QM/MM 連成シミュレーションの実行時間

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

Cosgene と AMOSS を各 1CPU 上で実行した場合は、QM 計算に殆どの計算時間を要する。従って、連成シミュレーションの負荷分散が今後の課題となる。又、UDSXML インターフェースでは、100KB の XML ファイルのデータ交換に 76ms を要する。一方、MPI インターフェースでは、メッセージパッシングで 30KB のデータ送受信に 3ms を要し、MD 計算の 1 ステップ 35ms の 1 割未満であり、実行負荷への影響は少ない。従って、連成シミュレーションで交換する物理データを活用しリポトリ化する場合は XML インターフェースが優れ、実行時間の高速性を追求する場合は MPI インターフェースが有効と考えられる。

(2) 分子動力学プログラム presto-X Basic 版(Cosgene)のグリッド上への実装と性能評価 (1.(1)に対応)

Gbit イーサネットで繋がった Intel Pentium4 (2.8GHz)の PC クラスタからなる単一グリ

ッド上に、生物情報解析研究センター等で開発した分子動力学法のプログラム presto-X Basic 版を実装し、その計算精度の検証と各機能単位の実行時間を測定した。

1) Cosgene の計算精度検証

水溶液中のエタノールの構造解析を対象に、Cosgene 開発環境の Compaq-ES40 と単一グリッド環境の Intel Pentium4 (2.8GHz) クラスタでの実行結果を比較する。エタノールの周りに水分子 4174 個を配置した周期系に対して、全水分子は TIP3P モデルを用いて、カノニカルアンサンブルの元で解析した。計算精度として 12 桁の精度を保証していることを検証した。

表 2.3 各マシン上での分子動力学プログラム Cosgene の計算精度比較

エネルギー	(1)Compaq(kcal/Mol)	(2)Intel(kcal/Mol)	(3)差分	(3)/(1)
BOND	854.526444202003	854.52644420199	1.2960E-11	1.5166E-14
ANGLE	0.263641725982449	0.263641725982449	0.0	0.0
TORSION	0.886602894631666	0.886602894631666	0.0	0.0
1-5 静電	-31067.061429103	-31067.0614290971	-5.9008E-09	1.8994E-13
1-5 VDW	-6091.97411294891	-6091.97411294843	-4.8021E-10	7.8827E-14

2) Cosgene の実行時間測定

単一グリッド上で、水溶液中のエタノールの構造解析を対象に実行時間を評価した。エタノールの周りに、(a)水分子 TIP4P モデルを 919 個配置した孤立系、(b)水分子 TIP3P モデルを 4174 個配置した孤立系、(c) 水分子 TIP3P モデルを 4174 個配置した周期系を評価した。周期系の計算には、Particle Mesh Ewald 法(PME 法)を用いた。カノニカルアンサンブルの元で、MD 計算の 100 ステップ当たりの実行時間を下記に示す。

全系は殆ど水分子が占有するため、MD 計算の実行時間は 1-5 相互作用計算に関連する処理で占められる。各ケースでの 1-5 相互作用計算に関連する処理の占有率は、(a)91%、(b)87%、(c) 96%である。今後、大規模な蛋白質分子を対象にする場合、その蛋白質を取り囲む水分子数も増加することから、Coulomb 相互作用を高精度に解析可能な PME 法を高速化することが不可欠になる。次章では、PME 法を並列化するために、3 次元フーリエ変換の並列プログラムの開発について進捗報告する。

表 2.4 各解析対象に対する Cosgene の実行時間比較

機能		(a)TIP4P	(b)TIP3P	(c)TIP3P+PME
相互作用計算	BOND	0.001s	0.720s	0.670s
	ANGLE	0.000s	0.000s	0.000s
	TORSION	0.002s	0.000s	0.000s
	1-5 相互作用	3.278s	38.790s	186.130s
	1-5 相互作用リスト作成	0.331s	37.320s	37.350s
拘束力計算	CAP	0.062s	-	-
積分器(原子速度と原子座標の更新)		0.313s	10.650s	9.270s
合計		3.987s	87.480s	233.420s

(3) Cosgene における三次元フーリエ変換プログラムの並列化 (1.(2)に対応)

前述したように、Cosgene の各機能の計算時間を比較すると、1-5 相互作用計算に関連する処理が最も大きい。この機能では、Particle Mesh Ewald 法(PME 法)を用いており、cutoff による実空間計算と 3 次元 FFT(Fast Fourier Transform)による逆空間計算から構成される。presto-X Basic 版は、逆空間計算は非並列で処理されているため、3 次元 FFT の並列処理機能プログラムを開発した。

x, y, z 方向の要素数を (N_x, N_y, N_z)、 f を各点でのフーリエ変換前の値、 F を各点でのフーリエ変換後の値とすると、3 次元 FFT は式(7.1)となる。

$$F(k_1, k_2, k_3) = \sum_{n_z=0}^{N_z-1} \left[\sum_{n_y=0}^{N_y-1} \left[\sum_{n_x=0}^{N_x-1} f(n_x, n_y, n_z) W_{N_x}^{n_x k_1} \right] W_{N_y}^{n_y k_2} \right] W_{N_z}^{n_z k_3} \quad (1.1)$$

$\underbrace{\hspace{15em}}_{x \text{ 方向 FFT 部分}}$
 $\underbrace{\hspace{25em}}_{y \text{ 方向 FFT 部分}}$
 $\underbrace{\hspace{35em}}_{z \text{ 方向 FFT 部分}}$

但し $W_N = \exp\left(-\frac{2\pi}{N} j\right), j^2 = -1$

3 次元 FFT の並列処理は次の手順で行う。

- 1) x 方向 FFT 計算
 - 2) y 方向 FFT 計算
 - 3) プロセッサ間の処理結果データ交換
 - 4) z 方向 FFT 計算
- } 同一プロセッサ内で計算する
— プロセッサを再分配して計算する

1), 2) では計算領域を z 方向に分割しプロセッサ毎に割り当てて処理する。また、4) では計算領域を x 方向に分割し、各プロセッサに配分する。1), 2) の分割方向と 4) の分割方向が異なるため、3) の通信処理が必要となる。本高速化手法においては、手順 2) の演算処理と 3) の通信処理をフレキシブルにオーバーラップさせることで高速な並列処理を可能にする 3 次元 FFT 並列処理機能プログラムを開発した。

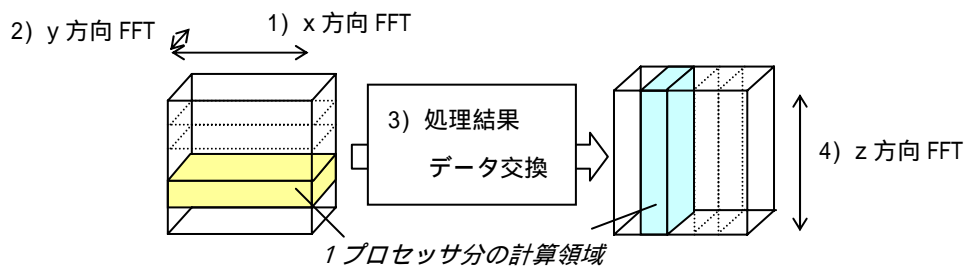


図 2.4 3 次元 FFT 並列処理方法

グリッド基盤システム 2 の PC クラスタ上で、プロセッサ数 1, 2, 4, 8, 64 での実行時間を計測した。解析対象として、x, y, z 方向の要素数 (N_x, N_y, N_z)=(64, 64, 64) を設定した。100 パターン分の入力データを用意し、全パターンに対し FFT 及び逆 FFT を実行した際の計算時間を 4 回測定し平均値を比較した。図 2.5 に、本方式、及び、2) の演算処理後に 3) の

通信処理を実行する非オーバーラップ方式の加速化率を示す。本方式では、64 台並列での CPU 時間の 1 台実行時に対する加速化率は約 50 倍であり、台数効果は更に増加傾向にある。一方、非オーバーラップ方式では加速化率は約 11 倍程度で飽和傾向を示す。来年度では、Cosgene に本年度開発した 3 次元 FFT 並列処理機能プログラムを組み込み、Cosgene の高速化、並びに、QM/MM 連成シミュレーションを高速化する予定である。

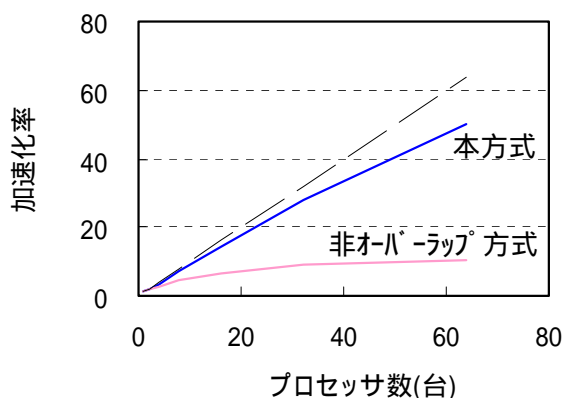


図 2.5 グリッド基盤システム 2 の PC クラスタ上での 3 次元 FFT の加速化率

(4) 疎粒度な機能単位プログラムへの Cosgene のパーツ化 (1 . (3) に対応)

presto-X Basic 版の Cosgene 単体プログラムを、グリッド対応に独立した機能単位プログラムに分割し、6 つのパーツプログラムから構成された Cosgene パーツ版を開発した。下記に Cosgene パーツ版のプログラムリストを示す。

- 1) shotrlength : 短距離相互作用計算 :
bond.f90, angle.f90, torsion.f90, torsional.f90
- 2) cutoff : カットオフありの 1-5 相互作用計算 :
Fast_Nonbonded.f90 の一部
- 3) nocut : カットオフなしの 1-5 相互作用計算 :
Fast_Nonbonded.f90 の一部
- 4) periodic : 周期境界の 1-5 相互作用の計算 :
Periodic_NonBonded.f90 の一部
- 5) pmeewald : PME/Ewald 法での実空間、逆空間の計算 :
Periodic_Nonbonded.f90 と Ewald_Method.f90
- 6) cosgene : MD 計算部分全体 (相互作用計算以外)

非パーツ化版の連成プログラムにて Cosgene を逐次実行した場合、パーツ化版の連成プログラムで各相互作用計算のパーツを並列実行した場合、両ケースの計算結果を比較し、パーツ化版が正常動作することを検証した。但し、Amoss の処理を実行せずにダミー化しており、実際に Amoss を連結したパーツ化版の連成プログラムは来年度に開発予定である。