

Development and Parallelization of Intelligent Optimization by Evolutionary Computation

Tomoyuki HIROYASU*

(Received March 18, 2002)

進化的手法による知的最適化とその並列化

同志社大学 工学部 廣安知之

1. はじめに

筆者はこれまでにシミュレーションベースによる問題解決手法の検討を行っている。その中で最も重要なツールの一つが最適化手法である。最適化とはある制約条件の下で目的関数の値を最小もしくは最大とするような設計変数の値を決定することである。この方法において、まず問題はモデル化され、モデル化に基づき解析器を構築する。われわれは最適器を用意することで問題を解決することが可能である。これらを図にまとめると図1となる。

図1においてはなんらかの形で問題の解析器の構築ができれば、最適器とのやりとりの中で問題が解決できることを示している。

このシミュレーションベースにて問題解決することにより、これまで以上に高度な問題の解決への可能性があり、意思決定においてそれに必要なコストを低下させることが可能である。また、同時に、問題解決が可能となるだけでなく、その過程において目的、制約、設計パラメータなどを明確化することにより問題の把握、理解ができるという側面がある。

最適化によって解決可能な問題のレベルは種々存在する。目的関数が静的な設計変数によって陽に表せる問題から、設計変数が動的な場合、目的関数が複数のサブ目的から構成される場合、サブ目的が最適化問題となっているような場合など、簡単な最適化問題から複雑な問題まで存在する。これらの最適化問題のレベルはそれぞれであるが、いずれもパラメータの最適化

問題に帰結可能である。

しかしながら、大規模もしくは複雑な問題を対象とした場合に、パラメータ最適化問題は以下の2点の問題を有する。すなわち、

1) 局所解が複数存在するため大局的な解の探索が困難

2) 最適解を探索するためには膨大な計算コストが必要となる。

点である。

大局的な解の探索のためには、なんらかの形で知的な最適化手法の開発は必須である。次に、膨大な計算コストに対応するためには、より良い最適化手法の開発、適切な解析器の選択、高速な計算処理可能なアーキテクチャの採用などが考えられる。

そこで、本研究では、これまでとは違った生物の仕組みを模倣した最適化手法の検討を行うことを目的とする。生物は長い間を通じて、それらが持つシステム自体が環境に適応すべき知的なシステムを有していると考えられ、そこから抽出可能なシステムは複雑な問題にも対処可能であると考えられる。また、同時に、生物は複数(多点)を特徴の一つとするため、並列処理との親和性は極めて高い。そのためシステムを並列処理することにより、大規模問題にも適応可能であると考えられる。

生物の遺伝や進化といった生物の仕組みを模倣した最適化手法の開発を行う。ここでは、生物の仕組みだ

* Department of Knowledge Engineering and Computer Sciences, Doshisha University, Kyoto
Telephone: +81-774-65-6932, Fax: +81-774-65-6780, E-mail: tomo@is.doshisha.ac.jp

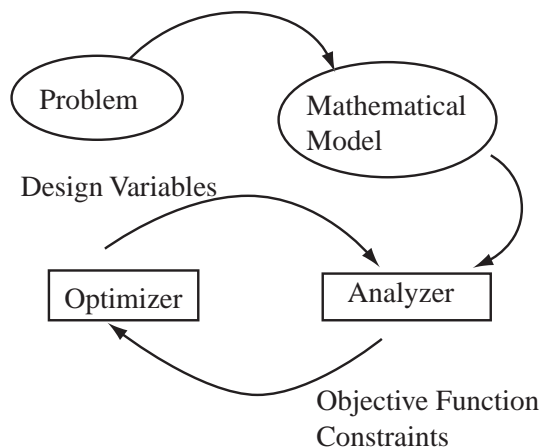


Fig. 1. Problem Solving using Optimization.

けでなく、生物の持つ、知的な手法というものも同時に抽出する。それにより複雑な問題も解決することのできる手法の開発が行えるものと考えられる。また、進化的計算手法による知的最適化手法の特徴である多点探索に基づき、それに適した並列処理のシステムを考案する。

本年度は以下の研究を中心に行った。

- PC クラスタの性能評価: 本プロジェクトでは 2 種類の超並列 PC クラスタを導入している。Linpack ベンチマークを使用してその性能を評価した。
- 分散遺伝的アルゴリズムの開発: 遺伝的アルゴリズム (Genetic Algorithm) は進化的計算手法の代表的な手法である。その並列モデルの一つである分散遺伝的アルゴリズムの改良を行った。
- GRID 環境を考慮した進化計算システム: GRID とはインターネット接続された複数のコンピュータを一つのシステムとして利用するインフラストラクチャである。GRID 環境において進化的計算を行うシステムの構築を開始した。
- 進化的計算手法による多目的最適化: 実問題における最適化問題においては、多くの場合、一つの目的関数ではなく、複数の目的関数、すなわち、多目的最適化問題となる。進化的計算手法による多目的最適化アルゴリズムの開発・改良を行った。
- タンパク質の 3 次元構造解析: タンパク質の持つエネルギーを最小化することにより 3 次元構造を同定することが可能である。本年はさらに自由度の大きなタンパク質の構造同定に成功した。

- ディーゼルエンジン噴射タイミングの最適化: 燃料噴射の量とタイミングを最適化することにより出力が一定で、すすや NO_x などの量を軽減できるディーゼルエンジンを設計することが可能である。

本稿ではこれらの中からタンパク質構造解析の結果について報告を行う。

2. 遺伝的交叉を用いた並列シミュレーテッドアニーリングによるタンパク質立体構造予測

本プロジェクトでは遺伝的アルゴリズムとシミュレーテッドアニーリングのハイブリッド手法である PSA/GAc を提案している。本年度は PSA/GAc により 2 種のタンパク質の立体構造予測を行った。以下にアミノ酸 13 残基の C-peptide, アミノ酸 34 残基のヒト副甲状腺ホルモンのフラグメントの構造予測結果を示す。

2.1 C-peptide の立体構造予測

本実験で構造予測の対象とした C-peptide は、13 個のアミノ酸残基 Ly+, Gl-, Thr, Ala, Ala, Ala, Ly+, Phe, Glu, Ar+, Gln, Hi+, Met からなるものである。岡本らの実験では、エネルギー関数 ECEPP/2 に基づいた気相中において、7 つのアミノ酸残基が α -helix となるとときに最小エネルギー構造をとることが明らかになっている。このときのエネルギー値は約 $-42kcal/mol$ である。 α -helix は、3 個以上連続したアミノ酸残基の二面角 (ϕ_i, ψ_i) が、 $(-60 \pm 45^\circ, -50 \pm 45^\circ)$ の値をとるときに形成されるものとする。なお、設計変数は C-peptide の主鎖における 26 個の二面角と、側鎖における 38 個の二面角である。二面角のとり得る値は $[-180^\circ, 180^\circ]$ の範囲で表現した。つまりこのタンパク質は 64 個の設計変数を持っており、1MCsweep によって 64 回の Metropolis 判定が課されるものとした。

PSA/GAc の Master-slave モデルによる C-peptide の立体構造予測では 41646MCsweep、24 個体の計算を 7 試行行った。初期構造は乱数を用いて各二面角をランダムに生成した。実験には、600MHz の PentiumIII を搭載した PC クラスタシステムの 6 ノードを用いた。計算時間は試行ごとにおよそ 35 時間を要した。

最も低いエネルギー値の構造を得たときの主鎖の二面角は Table 1 に示す値となった。このときのエネルギー値は $-57.8kcal/mol$ であり、9 アミノ酸残基の α -helix を持つ構造をとっている。得られた構造のエネルギー値は、岡本らの結果よりも低い値であり、また、

安定構造である α -helix となる残基数も多いことから、PSA/GAc の Master-slave モデルによる C-peptide の立体構造予測は成功したといえる。得られた構造を Fig. 2 に示す。

Table 1. Main-chain structure of the lowest-energy conformation of C-peptide gained by master-slave model of PSA/GAc.

Residue	ϕ	ψ	ω
1	18	-55	180
2	-161	162	180
3	-79	76	180
4	*	-72	180
5	*	-68	180
6	*	-63	180
7	*	-65	180
8	*	-63	180
9	*	-68	180
10	*	-64	180
11	*	-71	180
12	*	-65	180
13	-78	96	180

$E[\text{kcal/mol}] = -57.8$

2.2 ヒト副甲状腺ホルモンのフラグメントの立体構造予測

ヒト副甲状腺ホルモン (Parathyroid Hormone : PTH) のフラグメント (1-34) は、NMR 実験によって 2 つの α -helix(3-9, 17-28) の存在が示唆されている^{?)}。しかし、結晶化による解析と X 線実験ではその存在が明らかになってはいない。エネルギー関数 ECEPP/2 に基づいた気相中における岡本らの実験では、得られた最小エネルギー構造に 2 つの α -helix(2-7, 18-22) が出現し、そのときのエネルギー値は -210kcal/mol であることが示されている。なお、二面角 (ϕ_i, ψ_i) が $(-60 \pm 45^\circ, -50 \pm 45^\circ)$ の値をとるときに α -helix が形成され、 $(-105 \pm 50^\circ, 135 \pm 45^\circ)$ の値をとるときに β -sheet が形成される。PTH(1-34) は主鎖に 68 個、側鎖に 110 個の二面角を持ち、これらを設計変数としたため、1MCsweep によって 178 回の Metropolis 判定が課される。二面角のとり得る値は $[-180^\circ, 180^\circ]$ の範囲で表現した。

PSA/GAc の Master-slave モデルによる PTH(1-34) の立体構造予測では 4165MCsweep, 24 個体の計算を

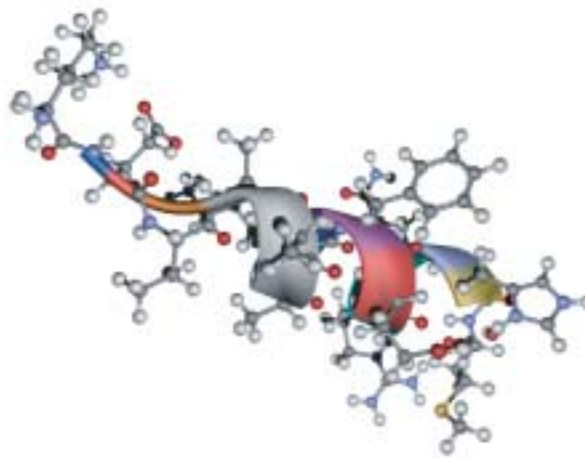


Fig. 2. Lowest-energy conformation of C-peptide gained by master-slave model of PSA/GAc.

2 試行を行った。初期構造は乱数を用いて各二面角をランダムに生成した。なお、実験には 1GHz の PentiumIII を搭載した PC クラスタシステムの 6 ノードを用いた。試行ごとにおよそ 40 時間の計算時間を要した。

2 試行中、得られた最も低いエネルギー値は -240.0kcal/mol であった。このときの構造は Table 2 に示す二面角を持つ。二面角 (ϕ_i, ψ_i) が $(-60 \pm 45^\circ, -50 \pm 45^\circ)$ の値をとるときに α -helix が形成され、 $(-105 \pm 50^\circ, 135 \pm 45^\circ)$ の値をとるときに β -sheet が形成されるため、得られた構造は 2 つの α -helix(2-8, 18-20) と 1 つの β -sheet(27-30) を持つ。得られた構造を Fig. 3 に示す。

岡本らの実験で得られた最小エネルギー構造には、アミノ酸残基 2-7 と 18-22 がそれぞれ α -helix の構造となった。先に述べたように、NMR 実験によっても 2 つの α -helix が確認されている。PSA/GAc による立体構造予測でも、アミノ酸残基 2-8, 18-20 に α -helix が出現したため、PTH(1-34) は 2 つの α -helix を持つと考えられる。また、岡本らの実験で得られた構造には存在しなかったが、PSA/GAc によって得られた構造ではアミノ酸残基 27-30 に β -sheet が出現した。 α -helix や β -sheet は安定構造であるが、誤った部分に出現するとエネルギー値が高くなる。PSA/GAc によって得られた構造のエネルギー値は、岡本らの結果よりも低い値であるため、PTH(1-34) の立体構造は 2 つの α -helix と 1 つの β -sheet を持つ可能性が高いと考えられる。この結果から、PSA/GAc の Master-slave モデルによる PTH(1-34) の立体構造予測は成功した

Table 2. Main-chain structure of the lowest-energy conformation of PTH(1-34) gained by master-slave model of PSA/GAc.

Sequence	5					10							
	Ser	Val	Ser	Glu	Ile	Gln	Leu	Met	His	Asn	Leu	Gly	Lys
ϕ	140	-70	-64	-68	-69	-66	-68	-56	-89	-170	-126	-149	-67
ψ	154	-26	-35	-38	-40	-39	-45	-49	60	-52	35	-90	-41
	-	a	a	a	a	a	a	a	-	-	-	-	a
	15					20					25		
	His	Leu	Asn	Ser	Met	Glu	Ar+	Val	Glu	Trp	Leu	Ar+	Lys
	-82	61	-73	-122	-72	-80	-90	-94	-71	-95	-90	-103	-56
	-60	91	84	145	-22	-26	-39	88	147	78	138	-47	-64
	a	-	-	b	a	a	a	-	b	-	b	-	a
	30					34							
	Lys	Leu	Gln	Asp	Val	His	Asn	Phe					
	-80	-68	-67	-124	-75	-108	-62	-110					
	160	118	145	132	-36	67	137	-46					
	b	b	b	b	a	-	-	-					

$$E[kcal/mol] = -240.0$$

といえる。

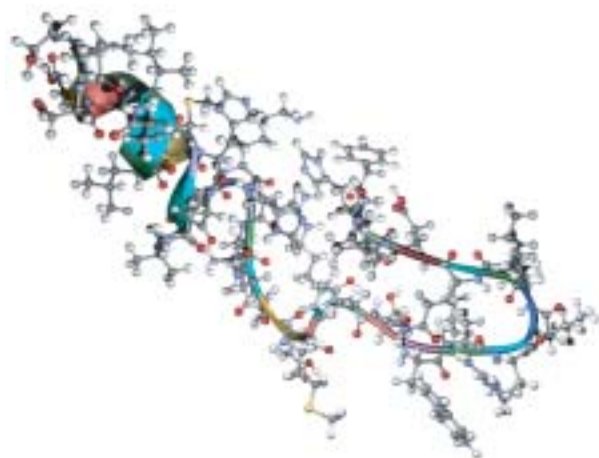


Fig. 3. Lowest-energy conformation of PTH(1-34) gained by master-slave model of PSA/GAc.

3. 結論

本研究では、提案手法である遺伝的交叉を用いた並列シミュレーテッドアニーリング (PSA/GAc) によるタンパク質立体構造予測を行った。

提案手法は、並列に複数実行しているシミュレーテッ

ドアニーリング (SA) の解の伝達時に、遺伝的アルゴリズム (GA) のオペレータである遺伝的交叉を用いたハイブリッドアルゴリズムである。アミノ酸数 17 の C-peptide とアミノ酸数 34 のヒト副甲状腺ホルモンについて立体構造予測を行った結果、従来用いられていた逐次 SA よりもよりエネルギー値の低い立体構造を予測することに成功した。

タンパク質のエネルギー関数は連続空間に存在し、局所的に無数の極小値、大域的にもいくつかの極小値を持つ。並列 SA は局所的な探索が得意でかつ計算の高速化を図ることができ、また GA は大域的な探索が得意でかつ部分解の組み合わせで最適解が得られる問題に有効である。そのため、並列 SA に GA オペレータを取り入れた本手法は、逐次 SA よりもタンパク質立体構造予測において有効であったと考えられる。

これらの結果から、PSA/GAc はタンパク質立体構造予測に有効な手法であることが示された。