

遺伝的交叉を用いた並列 SA による タンパク質立体構造のエネルギー最小化

廣安知之[†] 三木光範[†] 小椋信弥^{††}
青井桂子^{††} 吉田武史^{††} 岡本祐幸^{†††}

これまでの研究で、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) が小規模なタンパク質である Met-enkephalin の立体構造エネルギー最小化において高い解探索能力を示すことが明らかになっている。PSA/GAc は、並列に実行しているシミュレーテッドアニーリング間で、一定間隔のアニーリングごとに遺伝的アルゴリズムのオペレータである遺伝的交叉を用いて情報交換を行う最適化手法である。本論文では、PSA/GAc を Met-enkephalin よりも大規模なタンパク質である C-peptide と PTH (1-34) の立体構造エネルギー最小化に適用し、その探索性能について検討する。また、PSA/GAc の持ついくつかのパラメータのうち、特に探索性能に影響を与えていると考えられる交叉間隔、個体数、および総ステップ数について、いくつかの値を適用した数値実験を行い、その結果を検討し、考察を行った。その結果、独立型並列 SA (PSA) と比較して PSA/GAc は対象タンパク質に対して高い探索性能を示した。また、本稿において用いたパラメータでは、交叉を行う頻度が高いほど、総ステップ数は長いほど、より高い精度の立体構造が得られた。このことより、PSA/GAc は大規模なタンパク質のエネルギー最小化にも有効であることが明らかとなった。

Energy Minimization of Protein Tertiary Structure by Parallel Simulated Annealing using Genetic Crossover

TOMOYUKI HIROYASU,[†] MITSUNORI MIKI,[†] SHINYA OGURA,^{††}
KEIKO AOI,^{††} TAKESHI YOSHIDA^{††} and YUKO OKAMOTO^{†††}

From our recent research, it has been clarified that Parallel Simulated Annealing using Genetic Crossover (PSA/GAc) has a high searching ability on a minimization of an energy of small protein called Met-enkephalin. PSA/GAc performs genetic crossover, one of the operations of Genetic Algorithm (GA), among the Parallel SAs (PSAs) to exchange their informations. In this paper, PSA/GAc is applied to the minimization of the energy of C-peptide and Parathyroid Hormone Fragment(1-34). Also, among the parameters of PSA/GAc, crossover interval and total number of searching steps, which are supposed to have the high influences on the searching ability, are modified to some values in order to examine and study their influences. The result shows that PSA/GAc provides lower energy of the target proteins than Parallel SA. Furthermore, among the parameters used in this article, higher frequency of crossover and longer searching steps are confirmed to derive the lower energies. From the results we conclude that PSA/GAc is also effective on the energy minimization of larger proteins.

1. はじめに

タンパク質は生命現象に直接関わる重要な物質であるため、構造を解明することは生命現象の仕組みを説明することにもつながる。タンパク質の立体構造はエ

ネルギーの最小状態に対応しており、構造予測の一手法としてタンパク質の持つエネルギーを最小化する手法が挙げられる。これまでエネルギー最小化によるタンパク質の立体構造予測においてはシミュレーテッドアニーリング (Simulated Annealing: SA) が使用されてきた¹⁾。

しかし、タンパク質立体構造のエネルギー関数は非常に複雑で、大域的にいくつかの、局所的には無数の極小値を持つと考えられる。そのため、高速に大域的な最適解を発見するためには従来の逐次 SA (Sequential SA: SSA) に加えて、解探索性能を向上させる何

[†] 同志社大学工学部

Department of Engineering, Doshisha University

^{††} 同志社大学大学院

Graduate School of Engineering, Doshisha University

^{†††} 岡崎国立共同研究機関 分子科学研究所

Department of Theoretical Studies, Institute for Molecular Science

らかのメカニズムを導入したハイブリッド手法が望まれる．そこで筆者らは、局所的な探索が得意な SA に、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遺伝的アルゴリズム (Genetic Algorithm: GA) のオペレータを取り入れた手法として遺伝的交叉を用いた並列 SA (Parallel Simulated Annealing using Genetic Crossover: PSA/GAc) を提案した²⁾．

これまでに、PSA/GAc が小規模なタンパク質である Met-enkephalin のエネルギー最小化において、SSA よりも高い解探索能力を示すことが明らかとなった²⁾．そこで本研究では、PSA/GAc を Met-enkephalin よりも大規模なタンパク質である C-peptide およびヒト副甲状腺ホルモンのフラグメント (PTH (1-34)) のエネルギー最小化に適用し、その探索能力を確認する．また、PSA/GAc の持ついくつかのパラメータのうち、特に探索性能に影響を与えていると考えられる交叉間隔および探索ステップ数について、いくつかの値に変化させることによってその結果を検討する．これらの PSA/GAc の結果は独立型並列 SA (Parallel SA: PSA) の結果と比較している．なお本稿においては、PSA/GAc の対象問題における解探索性能を中心に検討し、並列性能などの検討は今後の課題とする．

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

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 (Parallel Simulated Annealing using Genetic Crossover: PSA/GAc) は、並列に実行している各 SA の解の伝達時に、GA のオペレータである交叉を用いた最適化手法である³⁾．本研究では、GA のオペレータを用いた SA であるため、探索点を個体と呼び、SA の探索点の総数 (並列数) を個体数と呼ぶこととする³⁾．

PSA/GAc での探索手順を以下に示す．

- step1 初期解を生成し、複数ある探索点が並列に SA の処理である生成処理、受理判定、クーリングを一定ステップ行う⁴⁾．
- step2 アニーリングが一定期間 d (交叉間隔) に達すると、並列に実行している SA の解からランダムに 2 つずつ解を選びペアを生成する．このときすべての個体がペアを組むため、個体数の半数のペアが生成される．
- step3 ペアを組む 2 つの個体を親として遺伝的交叉を行い、2 個体の子を生成する．この交叉法については、後で詳細を説明する．

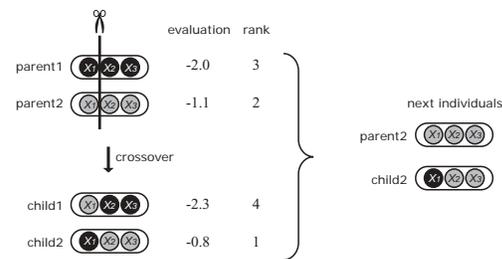


図 1 PSA/GAc における交叉と選択

Fig. 1 Crossover and selection in PSA/GAc

- step4 もとの親と生成した子との 4 個体のうち評価値の高い 2 個体を選択する．
- step5 選択された 2 個体から一定期間 d のアニーリングを行う．
- step6 すべてのペアにおいて step3 ~ step5 の処理を行う．
- step7 終了条件を満たすまで step2 ~ step6 の処理を繰り返す．

Step 3 における遺伝的交叉を簡単な例を用いて説明する．図 1 には、3 設計変数 (x_1, x_2, x_3) の場合が示されている．parent1 と parent2 が交叉の対象として選択された個体である．それぞれの個体は評価値を持っており、この場合 parent1 の評価値は -2.0, parent2 の評価値は -1.1 である．これらの個体に対して交叉が行われるが、これは通常のビット表現における一点交叉や 2 点交叉などとは異なり、交叉点は常に設計変数の境界に存在し、設計変数の値そのものは変化しない．同図では、設計変数 x_1 と x_2 の境界が交叉点としてランダムに選択され、交叉点以降の設計変数を個体間で入れ替える．この設計変数間交叉によって新しい 2 個体が生成される．生成された child1 の評価値は -2.3, child2 の評価値は -0.8 である．parent1, 2, child1, 2 の 4 個体それぞれの評価後、評価値の高かった parent2 および child2 が次のステップの探索点に選択されている．

3. PSA/GAc によるタンパク質のエネルギー最小化

本研究では、C-peptide およびヒト副甲状腺ホルモン (Parathyroid Hormone: PTH) のフラグメント (1-34) のエネルギー最小化に PSA/GAc を適用し、その有効性を確認し、いくつかのパラメータについても検討を行う．

3.1 実験条件

本研究においては、それぞれのタンパク質の主鎖お

よび側鎖の二面角を設計変数とし、エネルギー関数 $ECEPP/2^{5)~7)}$ に基づいた気相中のエネルギー最小化を行う。それら二面角の取りうる値は $[-180^\circ, 180^\circ]$ とした。各二面角において順に SA の生成・受理判定を行ってから 1 回のクーリングを行うこととし、これらの処理を 1 Monte Carlo sweep (MCsweep) と呼ぶこととする。SA の受理判定には、Metropolis 基準⁸⁾を用いた。Metropolis 基準では、次の状態 x' のエネルギー $E' = f(x')$ と現在の状態 x のエネルギー $E = f(x)$ との差分 $\Delta E (= E' - E)$ 、および温度パラメータ T から、次の状態への推移を受理するかどうかの判定を行う。Metropolis 基準は、式 1 で表される。

$$P_{ACCEPT} = \begin{cases} 1 & \text{if } \Delta E \leq 0 \\ \exp(-\frac{\Delta E}{T}) & \text{otherwise} \end{cases} \quad (1)$$

本研究で行ったタンパク質のエネルギー最小化において、初期構造はいずれも乱数を用いて各二面角をランダムに生成した。生成処理において、次の状態は近傍内に一様分布を用いて確率的に生成した。近傍の範囲 $[max, min]$ は式 2 で与えた。

$$\begin{cases} max = 180^\circ - \frac{180^\circ \times 0.7 \times \# \text{ sweep}}{\text{Total } \# \text{ sweeps}} \\ min = -max \end{cases} \quad (2)$$

SA における温度 T は、エネルギーの改悪方向への推移確率に重大な影響を与えるパラメータである。温度が高い場合、この推移確率が高く、温度が低い場合は低い値となる。最適解への漸近収束性を保証するためには、第 k ステップの温度を T_k としたとき、温度 T_{k+1} は $T_{k+1} = T_k / \log k$ 以上に急速に冷やしてはならない。しかしこのクーリングスケジュールではあまりにも解への収束が遅いため、本研究では最適解への収束性を犠牲にしたクーリング手法である、指数型クーリングを用いている。本研究で用いたクーリングスケジュールは、式 3 で表すことができる。すなわち、現在の温度 T_k で 1MCsweep の探索を行った後、次の温度 T_{k+1} は T_k にクーリング率 γ を乗じることにより決定される。

$$T_{k+1} = \gamma T_k \quad (0.8 \leq \gamma < 1) \quad (3)$$

本研究では、1MCsweep ごとにクーリングを行うため、全探索中のクーリングステップ数は、各個体における MCsweep 数と等しくなる。パラメータには表 1 に示したものをを用いた。表 1 において、最高温度および最低温度は、それぞれ岡本らと同じ値を用いている⁹⁾。本研究では、PSA/GAc の性能を岡本らの実験結果と比較することによって検証するため、実験パラメータを岡本らのものと等しく設定している。な

お、クーリングには指数型クーリングを用いており、1MCsweep ごとに一定のクーリング率を現在の温度に乗じるものとした。

表 1 PSA/GAc のパラメータ
Table 1 Parameters of PSA/GAc

Parameter	Value
Initial Temperature	2.0 (1000K)
Last Temperature	0.1 (50K)
Crossover Interval	8, 16, 32, 64
Cooling Rate	0.999281

3.2 対象とするタンパク質

本研究で対象とするタンパク質は、C-peptide および PTH (1-34) である。C-peptide は、13 個のアミノ残基からなり、主鎖および側鎖にそれぞれ 26 個と 38 個の二面角を持つ。PTH (1-34) については、34 個のアミノ残基からなり、主鎖および側鎖にそれぞれ 68 個と 110 個の二面角を持つ。つまり、C-peptide においては 1MCsweep によって 64 回の Metropolis 判定が課され、PTH (1-34) については 1MCsweep によって 178 回の Metropolis 判定が課される。

C-peptide について、岡本らの数値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、8 つのアミノ残基 (4-11) が α ヘリックスとなるときに最小エネルギー構造をとることが明らかであり、このときのエネルギー値は約 $-42kcal/mol$ である¹⁰⁾。また、岡本らの実験で得られた構造は、X 線結晶構造解析法などの実験的なタンパク質立体構造解析法によって得られた構造とよく一致している¹¹⁾。したがって本研究では、PSA/GAc を用いて岡本らの実験と同様の立体構造と $-42kcal/mol$ 以下のエネルギー値を得ることが目標となる。

PTH (1-34) は、NMR 実験によって 2 つの α ヘリックスの存在が示唆されている¹²⁾。岡本らの実験においては、得られた最小エネルギー構造に 2 つの α ヘリックスが形成され、そのときのエネルギー値は $-210kcal/mol$ である¹³⁾。したがって本研究では、PSA/GAc を用いて、2 つの α ヘリックスを持ち、かつエネルギー値が $-210kcal/mol$ 以下となる PTH (1-34) の立体構造を得ることが目標となる。

3.3 C-peptide のエネルギー最小化

C-peptide のエネルギー最小化において、岡本らは 10,000MCsweep \times 20 試行を用いて実験を行っている¹⁰⁾。総 MCsweep 数を岡本らと同等にするため、PSA/GAc による C-peptide のエネルギー最小化では、24 個体 \times 4,165MCsweep の計算を 2 試行行った。

交叉間隔を 8,16,32,64 としたときのそれぞれの試行で得られたエネルギー値を表 2 に示す.

表 2 C-peptide のエネルギー値
Table 2 Energies of C-peptide

Interval	Best	Worst
8	-43.7	-43.3
16	-40.9	-40.6
32	-53.9	-41.9
64	-44.9	-37.6

実験の結果, 交叉間隔が 32 のときに最も低いエネルギー値を得た. そのときの二面角は表 3 に示す値となり, エネルギー値は -53.9kcal/mol であった. α ヘリックスは, 3 個以上連なったアミノ残基の二面角 (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¹⁰⁾ ため, 本実験で得られた C-peptide は 9 アミノ残基が α ヘリックスを持つ構造をとっている. 表 3 において, a と記述したアミノ残基が α ヘリックスを形成したアミノ残基である. 得られた立体構造を図 2 に示す. 本実験で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 ヘリックスとなる残基数も多いことから, PSA/GAc は C-peptide のエネルギー最小化において高い解探索能力を持っているといえる.

表 3 PSA/GAc で得られた C-peptide の主鎖における二面角値
Table 3 Dihedral angles of C-peptide obtained by PSA/GAc

Sequence	Ly+	Gl-	Thr	Ala	Ala
	23	-79	-76	-67	-66
	-66	102	86	-27	-32
	-	-	-	a	a
	Ala	Ly+	Phe	Glu	Ar+
	-79	-62	-65	-65	-63
	-40	-43	-41	-41	-41
	a	a	a	a	a
	Gln	Hi+	Met		
	-72	-69	-82		
	-30	-40	105		
	a	a	-		

$Energy = -53.9\text{kcal/mol}$

3.4 PTH (1-34) のエネルギー最小化

PTH (1-34) のエネルギー最小化において, 岡本らは $10,000\text{MCSweep} \times 20$ 試行を用いて実験を行っている¹³⁾. 総 MCSweep 数を岡本らと同等にするため, PSA/GAc による PTH (1-34) のエネルギー最小化

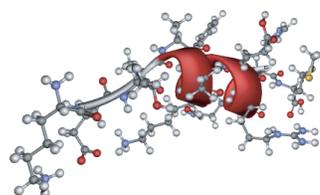


図 2 PSA/GAc で得られた C-peptide の立体構造
Fig. 2 Lowest energy conformation of C-peptide obtained by PSA/GAc

では, 24 個体 \times $4,165\text{MCSweep}$ の計算を 2 試行行った. 交叉間隔を 8,16,32,64 としたときのそれぞれの試行で得られたエネルギー値を表 4 に示す.

表 4 PTH (1-34) のエネルギー値
Table 4 Energies of PTH(1-34)

Interval	Best	Worst
8	-246.0	-242.2
16	-239.2	-222.4
32	-231.7	-231.1
64	-225.6	-222.5

実験の結果, 交叉間隔が 8 のときに最も低いエネルギー値である -246.0kcal/mol が得られた. このときの構造は表 5 に示す二面角を持つ. また, 得られた構造は 2 つの α ヘリックス (2-8, 15-18) を持つ. PSA/GAc で得られた構造を図 3 に示す.

3.2 節で述べたように, PTH (1-34) は NMR 実験および岡本らの実験により, α ヘリックスを 2 つ持つことが確認されている. PSA/GAc による実験においても, PTH (1-34) の立体構造には 2 つの α ヘリックスが出現することが確認された. また, このときのエネルギー値は, 岡本らの実験で得られたエネルギー値よりも低い値であった. この結果より, PSA/GAc は PTH (1-34) のエネルギー最小化において高い解探索能力を持っているといえる.

3.5 MCSweep 数と交叉の検討

3.5.1 実験概要

PSA/GAc では, 並列に実行している各 SA の探索途中の解の伝達に GA のオペレータである遺伝的交叉を用いている. タンパク質を持つエネルギー値を最小とするような最適解の一部の設計変数の値が, すでに探索点において発見されている場合, 設計変数間の交叉操作により, その設計変数の値を他の SA 探索に伝達することができるため, アニーリングの収束を早めることができると考えられる. したがって, 遺伝的交叉を行う頻度が解探索能力に大きな影響を与える

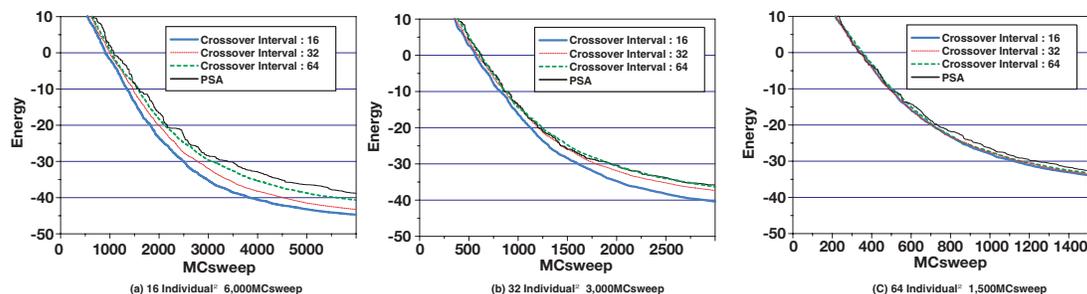


図 4 C-peptide のエネルギー履歴
Fig. 4 Energy transition of C-peptide

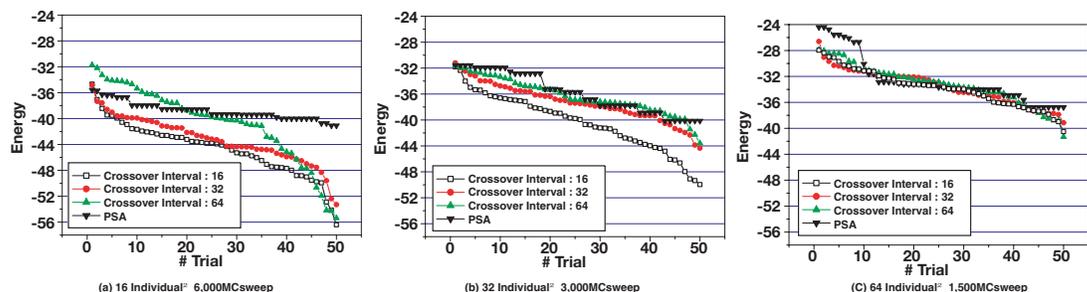


図 5 探索終了時のエネルギー値 (C-peptide)
Fig. 5 Lowest energies of C-peptide

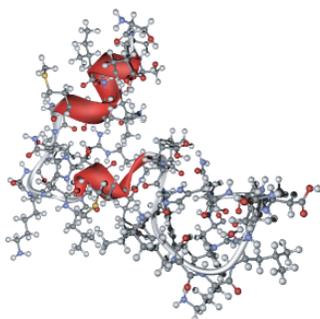


図 3 PSA/GAc で得られた PTH (1-34) の立体構造
Fig. 3 Lowest energy conformation of PTH (1-34)
obtained by PSA/GAc

と考えられる．そこで本節では，MCsweep 数と交叉を行う頻度が解探索に及ぼす影響を検査するための実験を行う．本実験で用いた PSA/GAc は，16 個体 \times 6,000MCsweep，32 個体 \times 3,000MCsweep，64 個体 \times 1,500MCsweep であり，それぞれの PSA/GAc においてさらに交叉間隔を 16MCsweep，32MCsweep，64MCsweep と設定した．本研究で対象としているタンパク質のエネルギー最小化問題においては，非常に繰り返し数が多く，計算コストの高い問題であり，本

研究で用いた PC クラスタシステム においても，16 個体 \times 6,000MCsweep，交叉間隔 32 の PSA/GAc でそれぞれのタンパク質のエネルギー最小化を行うと 1 試行に表 6 に示した時間を要する．なお表 6 には，比較のためにアミノ酸残基数 5，二面角数 19 からなる小規模なタンパク質である Met-enkephalin の結果についても示した．また，一回のタンパク質のエネルギー計算は本クラスタシステムにおいて C-peptide で 1.98×10^{-3} [sec]，PTH (1-34) で 1.83×10^{-2} [sec] であるため，効率良く計算させるためには，できるだけ交叉間隔を長く設定したい．そのために，これらの交叉間隔での実験が適当であると判断した．また，本実験ではこれらの PSA/GAc と共に，交叉を行わない独立型並列 SA (Parallel Simulated Annealing : PSA) を用いた．これは逐次 SA を並行数回試行し，その中から最良解を取り出す手法である．

なお，すべての場合での終了時の総計算回数は大よそ統一している．

3.5.2 C-peptide の実験結果

図 4 はそれぞれ 16 個体 \times 6,000MCsweep，32 個体 \times 3,000MCsweep，64 個体 \times 1,500MCsweep の

CPU : Intel Xeon 2.4GHz \times 128 2CPUs/node
Memory : 1GB/node
Network : Myrinet2000

表 5 PSA/GAc で得られた PTH (1-34) の主鎖における二面角値

Table 5 Dihedral angles of PTH (1-34) obtained by PSA/GAc

					5
Sequence	Ser	Val	Ser	Glu	Ile
	93	-65	-68	-76	-71
	157	-26	-36	-37	-35
	-	a	a	a	a
					10
	Gln	Leu	Met	His	Asn
	-64	-70	-55	-64	-112
	-41	-44	-43	-96	145
	a	a	a	-	-
					15
	Leu	Gly	Lys	His	Leu
	-105	76	-95	-90	-73
	108	-90	165	-1	-22
	-	-	-	-	a
					20
	Asn	Ser	Met	Glu	Ar+
	-85	-85	-74	-88	-142
	-20	-37	-30	64	-60
	a	a	a	-	-
					25
	Val	Glu	Trp	Leu	Ar+
	-134	-85	-64	-65	-155
	-58	73	131	108	112
	-	-	-	-	-
					30
	Lys	Lys	Leu	Gln	Asp
	-113	-105	-148	-64	-65
	63	-32	148	-55	145
	-	-	-	a	-
					34
	Val	His	Asn	Phe	
	48	-55	-60	-84	
	71	-46	153	118	
	-	-	-	-	

$$Energy = -246.0kcal/mol$$

表 6 各タンパク質のエネルギー最小化に要する時間
Table 6 Required time for the energy minimization of each protein

Protein	Time[sec]
Met-enkephalin	41
C-peptide	762
PTH(1-34)	17,306

PSA/GAc および PSA を用いて, C-peptide のエネルギー最小化を 50 試行行ったときのエネルギー履歴の平均値である. またそれぞれのグラフには, PSA/GAc における交叉間隔を 16, 32, 64 としたときと PSA の結果を示した. 横軸は MCsweep 数を, 縦軸はタンパク質のエネルギー値 [kcal/mol] を示す.

それぞれのパラメータを用いた実験において, 探索

終了時に得られたエネルギー値を降順に並べたものを図 5 に示した. 横軸が試行回数を, 縦軸がエネルギー値 [kcal/mol] をあらわす.

3.5.3 PTH (1-34) の実験結果

図 6 はそれぞれ 16 個体 \times 6,000MCsweep, 32 個体 \times 3,000MCsweep, 64 個体 \times 1,500MCsweep の PSA/GAc および PSA を用いて, PTH (1-34) のエネルギー最小化を 20 試行行ったときのそれぞれのエネルギー履歴の平均値を示したものである. またそれぞれのグラフには, PSA/GAc における交叉間隔を 16, 32, 64 としたとき, および PSA の結果を示した. 横軸が MCsweep 数を, 縦軸がタンパク質のエネルギー値 [kcal/mol] である.

それぞれのパラメータを用いた実験において, 探索終了時に得られたエネルギー値を降順に並べたものを図 7 に示した. 横軸が試行回数を, 縦軸がエネルギー値 [kcal/mol] をあらわす.

3.5.4 各タンパク質における交叉間隔とエネルギー値

それぞれのタンパク質において, 用いた交叉間隔と得られたエネルギー値の関係を図 8 に示す. 同図において, 横軸が交叉間隔を, 縦軸が得られたエネルギーの平均値を示したものである. なお, 比較のために図 8(b) の C-peptide と図 8(c) の PTH(1-34) に加え, Met-enkephalin のエネルギー最小化の結果を図 8(a) に示した. また図 8(a), (b), (c) のそれぞれには, 16 個体 \times 6,000MCsweep, 32 個体 \times 3,000MCsweep, および 64 個体 \times 1,500MCsweep の結果を示した.

3.5.5 結果の検討

C-peptide および PTH (1-34) の両方の結果において以下のことが確認される.

まず, 各個体における MCsweep 数が長い場合と短い場合の比較を行う. 図 4 は C-peptide のエネルギー履歴を, 図 6 は PTH (1-34) のエネルギー履歴を示している. C-peptide については, 図 4(c) と図 4(a) を比較すると, 平均して 10kcal/mol のエネルギー値の改善が見られる. 同様に, PTH (1-34) については, 図 6(c) と図 6(a) より平均して 30kcal/mol のエネルギー値の改善が見られる. また, 図 8(b), (c) に示した交叉間隔ごとの平均エネルギー値からもこの傾向を見て取ることができる. すなわち同じ交叉間隔でも, 各個体における MCsweep 数が長いほど, 得られるエネルギー値が低くなっていることが分かる. これらの結果より, いずれの対象問題においても個体数を増加させるよりも MCsweep 数を長く取る場合が探索に有効であるといえる. これは MCsweep 数は SA の温度

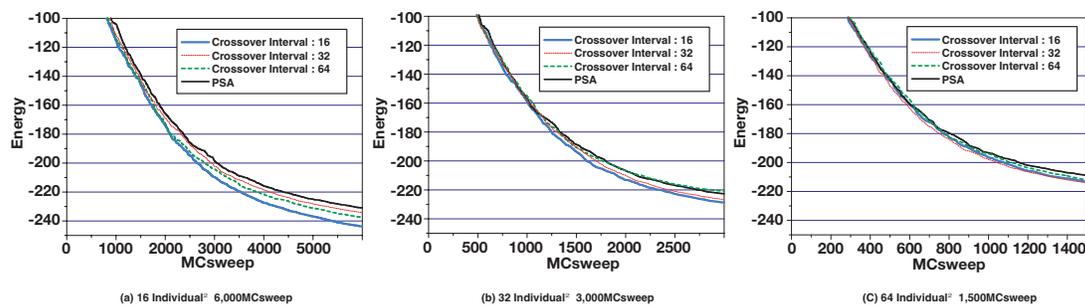


図 6 PTH (1-34) のエネルギー履歴
Fig. 6 Energy transition of PTH(1-34)

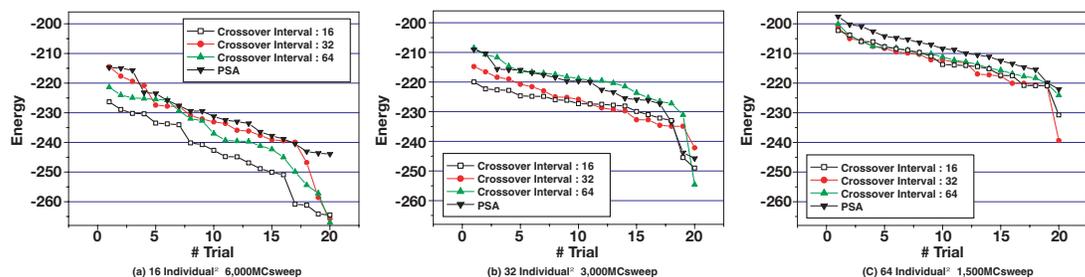


図 7 探索終了時のエネルギー値 (PTH (1-34))
Fig. 7 Lowest energies of PTH(1-34)

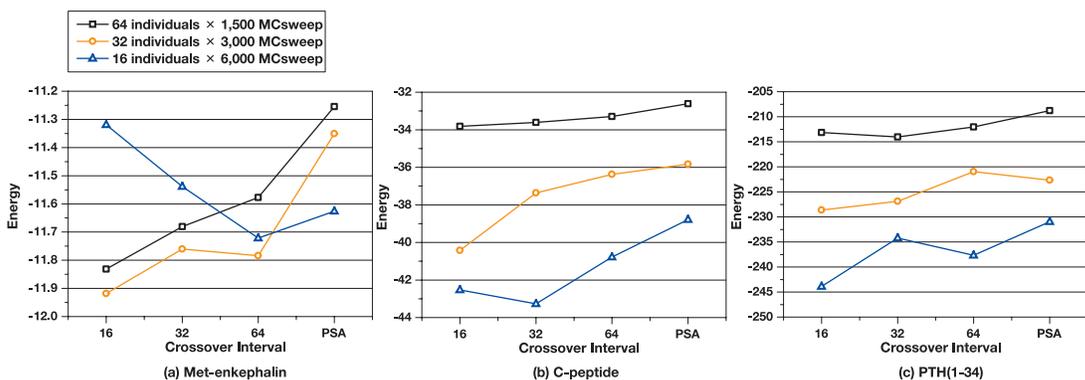


図 8 各タンパク質における交叉間隔とエネルギー値
Fig. 8 Crossover intervals and energies of each protein

変化と連動しているため、MCsweep 数が長い方が大局的および局所的な探索が十分行えたと考えられるからである。一方、図 8(a) に示した Met-enkephalin についての実験結果からは、32 個体 \times 3,000MCsweep、交叉間隔 16 の PSA/GAc が最も良い性能を示しており、次いで 64 個体 \times 1,500MCsweep、交叉間隔 16 の PSA/GAc が高い性能を示していることが確認できる。Met-enkephalin は二面角数が 19 の小規模なタンパク質であることから、最小化問題としては簡単な問題であると予想される。そのため、本問題においては、3,000MCsweep あるいは 1,500MCsweep が探索

には十分であり、その分探索点数である個体数を増やすことによって有効な探索が行えたと考えられることができる。

一方、C-peptide および PTH (1-34) は大規模なタンパク質であるため、MCsweep 数を長く設定する必要がある。

次に交叉の影響について検討を行う。図 5(c) および図 7(c) は、64 個体 \times 1,500MCsweep の PSA/GAc を両タンパク質のエネルギー最小化に適用した結果である。同図の結果からは、PSA と PSA/GAc との性能に大差が見られないことが分かる。次に、図 5(b)

および図 7(b) に示した 32 個体 \times 3,000MCsweep の PSA/GAc および PSA の結果を見ると、交叉間隔が 16 の PSA/GAc では、多くの試行において PSA よりも低いエネルギー値を得ていることが分かる。このことは、図 8(b), (c) から明らかである。すなわち、図 8(b), (c) において 32 個体 \times 3,000MCsweep の PSA/GAc を両タンパク質に適用した結果、交叉間隔が 16 の時にいずれにおいても PSA よりも低いエネルギー値を得ている。さらに、図 5(a) の 16 個体 \times 6,000MCsweep の PSA/GAc を C-peptide に適用した結果からは、交叉間隔 16, 32 の時にいずれの試行においても PSA よりも低いエネルギー値を得ていることが分かる。また図 7(a) に示した 16 個体 \times 6,000MCsweep の PSA/GAc を PTH (1-34) に適用した結果からは、交叉間隔が 16 の PSA/GAc がいずれの試行においても PSA より低いエネルギー値を得ていることが分かる。また、この時のエネルギー値の差は図 7(b) よりも顕著に現れている。十分に長い MCsweep 数を用いて探索を行うとき、交叉間隔が短くなるとより低いエネルギー値を得るという傾向は、図 4(a) および図 6(a) から見て取ることができる。同図は、16 個体 \times 6,000MCsweep の PSA/GAc および PSA の結果を表したものであるが、交叉間隔が短くなるごとに探索終了時のエネルギー値が $3 \sim 6 \text{ kcal/mol}$ だけ低くなっていることが分かる。これらのことより、十分に長い MCsweep 数を用いて探索を行うとき、交叉を 16MCsweep 程度で行うことは C-peptide および PTH (1-34) のエネルギー最小化に有効であると考えることができる。

小規模なタンパク質である Met-enkephalin のエネルギー最小化においては、十分に長い MCsweep 数、すなわち 16 個体 \times 6,000MCsweep で探索を行うと、図 8(a) より、交叉間隔 64 の PSA/GAc は PSA よりも性能が高まるものの、交叉間隔を 32, 16 とさらに短くすることによって性能が下がっていることが分かる。一方、64 個体 \times 1,500MCsweep あるいは 32 個体 \times 3,000MCsweep の PSA/GAc においては、交叉間隔を短くすることによっておおむね性能が向上していることが分かる。このことから、十分に長い MCsweep 数を用いて探索を行う場合、Met-enkephalin においては交叉を頻繁にするほど性能が低下するといえる。すなわち小規模なタンパク質については、不必要に長い MCsweep 数を用いて探索を行う場合、交叉間隔を短く設定すると個体の多様性が失われるため、性能が低下する。これらのことより、PSA/GAc および交叉操作は、特に長い MCsweep が必要であると考えられ

る大規模なタンパク質に対して有効であると考えられる。

4. 結 論

本研究では、アミノ酸 5 残基からなる Met-enkephalin のエネルギー最小化において有効性が確認されていた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) を、より大規模なタンパク質である C-peptide とヒト副甲状腺ホルモンのフラグメント PTH (1-34) の立体構造エネルギー最小化に適用した。

まず、PSA/GAc と岡本らの実験で用いられている逐次 SA (Sequential SA) の性能を比較するために、PSA/GAc を C-peptide および PTH (1-34) のエネルギー最小化に適用した。その結果、いずれのタンパク質においても PSA/GAc が SSA よりも高い解探索能力を示すことが明らかとなった。

次に、PSA/GAc と独立型並列 SA (PSA) を上記のタンパク質のエネルギー最小化に適用して性能の比較を行った。このとき、MCsweep 数と交叉が解探索に及ぼす影響を検討するために、個体数 \times MCsweep 数を一定 (評価計算回数がほぼ同等) に設定した。実験の結果、C-peptide および PTH (1-34) については、十分に長い MCsweep 数を用いて探索を行う場合、PSA/GAc は PSA よりも低いエネルギー値を得た。また、交叉間隔が 16MCsweep 程度に設定した場合、多くの試行において低いエネルギー値を得ることが確認された。

以上の結果より、PSA/GAc における交叉オペレータがタンパク質のエネルギー最小化において有効に働いていることが確認された。また、PSA/GAc がより大規模なタンパク質のエネルギー最小化計算に対しても有効であることが示された。

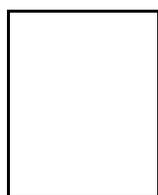
参 考 文 献

- 1) Hikaru Kawai, Takeshi Kikuchi, and Yuko Okamoto. A prediction of tertiary structures of peptide by the Monte Carlo simulated annealing method. *Protein Engineering*, Vol. 3, No. 2, pp. 85-94, 1989.
- 2) 廣安知之, 三木光範, 小掠真貴, 岡本祐幸. 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の検討. *情報処理学会論文誌*, Vol. 43, No. SIG7(TOM6), pp. 70-79, 2002.
- 3) 廣安知之, 三木光範, 小掠真貴. 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. 第 44 回 システム制御情報学会 研究発表講演会講演論文集, pp. 113-114, 2000.

- 4) S. Kirkpatrick, C. D. Gelatt, Jr., M. P. Vecchi. Optimization by Simulated Annealing. *Science*, Vol. 220, No. 4598, pp. 671–680, 1983.
- 5) F.A. Momany, R.F. McGuire, A.W. Burgess, and H.A. Scheraga. *J. Phys. Chem.*, Vol. 79, pp. 2361–2381, 1975.
- 6) G. Nemethy, M.S. Pottle, and H.A. Scheraga. *J. Phys. Chem.*, Vol. 87, pp. 1883–1887, 1983.
- 7) M.J. Sippl, G. Nemethy, and H.A. Scheraga. *J. Phys. Chem.*, Vol. 88, pp. 6231–6233, 1984.
- 8) Bruce E. Rosen, 中野良平. シミュレーテッドアニーリング -基礎と最新技術-. 人工知能学会誌, Vol. 9, No. 3, 1994.
- 9) Yuko Okamoto, Takeshi Kikuchi, and Hikaru Kawai. Prediction of Low-Energy Structures of Met-Enkephalin by Monte Carlo Simulated Annealing. *CHEMISTRY LETTERS*, pp. 1275–1278, 1992.
- 10) Yuko Okamoto, Masataka Fukugita, Takashi Nakazawa, and Hikaru Kawai. α -Helix folding by Monte Carlo simulated annealing in isolated C-peptide of ribonuclease A. *Protein Engineering*, Vol. 4, No. 6, pp. 639–647, 1991.
- 11) Ulrich H. E. Hansmann and Yuko Okamoto. Tertiary Structure Prediction of C-Peptide of Ribonuclease A by Multicanonical Algorithm. *J. Phys. Chem. B*, Vol. 102, No. 4, pp. 653–656, 1998.
- 12) W. Klaus, T. Dieckmann, V. Wray, D. Schomburg, E. Wingender, and H. Mayer. *Biochemistry*, Vol. 30, pp. 6936–6942, 1991.
- 13) Yuko Okamoto, Takeshi Kikuchi, Takashi Nakazawa, and Hikaru Kawai. α -Helix structure of parathyroid hormone fragment (1-34) predicted by Monte Carlo simulated annealing. *INTERNATIONAL JOURNAL OF PEPTIDE & PROTEIN RESEARCH*, Vol. 42, pp. 300–303, 1993.

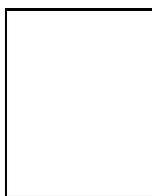
(平成?年?月?日受付)

(平成?年?月?日採録)



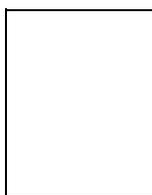
廣安 知之 (正会員)

1966年生. 1997年早稲田大学理工学研究科後期博士課程修了. 同年早稲田大学理工学部助手. 1998年より同志社大学工学部助手. 創発的計算, 進化的計算, 最適設計, 並列処理などの研究に従事. IEEE, 電気情報通信学会, 計測自動車制御学会, 日本機械学会, 超並列計算研究会, 日本計算工学会各会員.



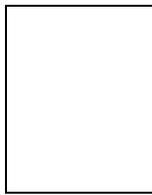
三木 光範 (正会員)

1950年生. 1978年大阪市立大学大学院工学研究科博士課程修了, 工学博士. 大阪市立工業研究所研究員, 金沢工業大学助教授を経て1987年大阪府立大学工学部航空宇宙工学科助教授, 1994年同志社大学工学部教授. 進化的計算手法とその並列化, および知的なシステム的设计に関する研究に従事. 著書は「工学問題を解決する適応化・知能化・最適化法」(技法堂出版)等多数. IEEE, 米国航空宇宙学会, 人工知能学会, システム制御情報学会, 日本機械学会, 計算工学会日本航空宇宙学会等会員. 超並列計算研究会代表. 通産省産業技術審議会委員.



小椋 信弥 (学生会員)

1979年生. 2002年同志社大学工学部知識工学科卒業. 同年, 同志社大学大学院工学研究科知識工学専攻入学. 並列処理, 最適化計算, バイオインフォマティクス等に興味を持つ.



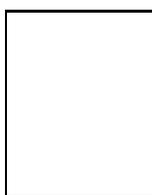
青井 桂子

1979年生. 2002年同志社大学工学部知識工学科卒業. 同年, 同志社大学大学院工学研究科知識工学専攻入学. 並列処理, 最適化計算, グリッドコンピューティング等に興味を持つ.



吉田 武史

1978年生. 2001年同志社大学工学部知識工学科卒業. 2003年同志社大学大学院工学研究科知識工学専攻修了. 同年, IBM ビジネスコンサルティングサービス(株)入社. 並列処理, 最適化計算等に興味を持つ.



岡本 祐幸

岡崎国立共同研究機構分子科学研究所助教授, 総合研究大学院大学数物科学研究科助教授(併任). 1984年コーネル大学大学院理学研究科博士課程修了, Ph.D. 1986年奈良女子大学理学部助手等を経て1995年より現職.