情報学基礎 23－5
（1991. 9. 10）

マルチアライメントについて

大矢 雅則　宮崎 智　緒方 浩二

東京理科大学 理工学部 情報科学科

この報告では，シュミレーティドアニーリングを用いたマルチアライメントのアルゴリズムについて説明する．シュミレーティドアニーリングは最小値探索問題を解くために開発されたアルゴリズムのひとつで，実際にセールスマン問題等に適用されよい実績をあげている．ここでは，まずアニーリングの数学的裏付けについて簡単に触れた後，アニーリングをアライメントに適用する上で必要となる摂動行列，受理行列や目的関数を実際に構成し，比較的短い６本の配列に対して整列化を実行した結果について考察する．我々が開発したマルチアライメントは実行時間において多少の課題を残しているものの，より正確なアライメント結果を得るという点で適しており，得られた結果について報告をする．

ON MULTIPLE ALIGNMENT

Masanori Ohya　Satoru Miyazaki　Kouji Ogata

Depertment of Information Sciences
Faculty of Science & Technology
Science University of Tokyo

In this paper, we introduce multiple alignment as an application of "Simulated Annealing" method. Simulated Annealing has been applied to some combinational optimization problems such as traveling salesman problem. After giving short mathematically explanation of this method, we construct some matrices and a genetic distance called the object function in annealing theory for the multiple alignment. Our method is shown to be better in the sense that we obtain a result having a smaller value for the genetic distance. We discuss further development along this new method.

1．まえがき

近年，遺伝子配列データの蓄積により，これらのデータを用いて分子進化を説明したり遺伝子配列の構造を数理的に解明しようとする試みがなされている．より正確な分析を行うためにはより多くのデータを一度に比較しやすい形に整列化できるた方が都合がよい．そこで，本報告では，最小値探索問題に対する解法の一つであるシュミレーティドアニーリング法を応用し文字道理ｎ本の遺伝子配列を扱うことができるマルチアライメントのアルゴリズムについて説明する．これまでに実用されているマルチアライメントは本質的に２本の遺伝子配列に対するアライメント法を用いているようである．そのためギャップの入り方が不自然になったり結果の復元性がない等の欠点を克服するには限界があると思われる．しかし，ここで提案するアライメント法はこうした問題を理論的には十分カバーしているものである．ここでは，まず，シュミレーティドアニーリングの概要について触れ，次にこれをアライメントに応用した一例を紹介し，具体例をもとに，他のアライメント法と比較した場合の優位性について報告する．

2．シュミレーティドアニーリングの基礎[3]

この節では，シュミレーティドアニーリング（以下，単にアニーリングと呼ぶ）の一般的な基本概念を簡単に復習しておく．アニーリングは次に述べるような数学的裏付けに基づいて考案された最小値探索法の一つである．いま，Ｘを空でない有限集合とし，Ｘの要素ｘを基本状態と呼ぶ．また，ｘから実数への関数ｆ：ｘ→ℝを考えこれを目的関数と呼ぶことにする．このとき，我々は集合Ｘの要素の中でｆの値を最小にするものを見つける問題を最小値探索問題と呼んでいる．この問題を解く最も単純なやり方は，Ｘの全ての要素に対してｆ(ｘ)の値を計算し比較することである．しかし，Ｘの要素の数があまりにも大きいと計算時間が膨大になり事実上解を求めることは不可能である．しかし，近年になって全ての要素を考えることなしに最適解を求める手法のひとつとしてアニーリングと呼ばれる方法が考案された．それではアニーリング法を説明するために必要ないくつかの概念　をあげておく．まず，正のパラメータβと目的関数ｆから次のような確率分布q_βを考えておく．

$$q_\beta(x) \equiv \frac{1}{Z}\exp\{-f(x)\beta\} \quad (x \in X)$$

ここで，

$$Z \equiv \sum_{x \in X}\exp\{-f(x)\beta\}$$

である．また，Ｘ上でｆの最小値をf_{min}とし，f_{min}をあたえるｘの要素の集合をX_{min}とする．

$$X_{min} \equiv \{x; x \in X, f(x) = f_{min}\}$$

このとき確率分布q_0を

$$q_0(x) \equiv \begin{cases} 1/|X_{min}| & (x \in X) \\ 0 & (x \notin X) \end{cases}$$

で与えておく．ここで|・|は集合の要素の個数である．

このq_βとq_0に対して，次の定理が知られている．

<定理 2.1>　$\beta \to \infty$のときq_βはq_0に収束する．

$$\lim_{\beta \to \infty} q_\beta = q_0$$

<定理 2.1>はq_βに従いＸに値を取る乱数が発生できれば，十分に大きなβをとればq_0に従いＸに値を取る乱数が得られることを示している．q_0はｆで最小値をとるｘ∈Ｘの分布になっていたから，

q_0に従う乱数の発生⇐⇒

ｆで最小値をとるｘ∈Ｘの決定

という図式が成り立つことになる．こうなると，ｆで最小値をとるｘを見つけるためにはq_βを求めればよい．しかしq_βもｆによって定義されているからq_βを求めることも不可能であるようにみえる．しかし，次に述べる摂動行列，受理行列，推移行列というも概念を導入することによりq_βを近似することが理論的に可能となる．

[定義 2.2]　$|X|\times|X|$の行列$P=(P(x,y))$で次の性質を満たすものを摂動行列という.

(1) $\forall x, y \in X : P(x,y) \geqq 0$

(2) $\forall x \in X : P(x,x) = 0$

(3) $\forall x \in X : \sum_{x \in X} P(x,y) = 1$

(4) $\forall x, y \in X : P(x,y) = P(y,x)$

(5) $\forall x, y \in X : \exists t : P_t(x,y) > 0$

ここで, $P_t(x,y)$ は行列 P の t 個の積の x 行 y 列の成分を示す.

[定義 2.3]　次に示す$|X|\times|X|$の行列A_βを受理行列と呼ぶ.

$$A_\beta(x,y) \equiv \min\{1,\ q_\beta(y)/q_\beta(x)\}$$
$$= \min\{1,\ \exp(f(y)-f(x))\beta\}$$

さらに, 摂動行列と受理行列を用いて推移行列が次のように定義されている.

[定義 2.4]　次式で定義される行列$\Lambda_\beta = (\Lambda_\beta(x,y))$を推移行列と呼ぶ.

$$\Lambda_\beta(x,y) \equiv \begin{cases} P(x,y)A_\beta(x,y) & (x=y\text{のとき}) \\ 1-\sum_{x \in X} P(x,y)A_\beta(x,y) & (x\neq y\text{のとき}) \end{cases}$$

この推移行列はマルコフ連鎖の性質を有し, さらに規約性, 非周期性の性質を満たすことがわかっている. また, q_βはΛ_βの平衡分布となっていることも知られている. すると, マルコフ連鎖のエルゴード定理によって, 任意の確率分布 q に対して

$$\lim_{t\to\infty} q\Lambda_\beta^t = q_\beta$$

が成り立つ. これと<定理 2.1>より

$$\lim_{\beta\to\infty}\lim_{t\to\infty} q\Lambda_\beta^t = q_0 \qquad (2.1)$$

となることがわかる. つまり, 理論的には推移行列Λ_βから定まるマルコフ連鎖に従うX上の乱数を発生させていけば, 十分に大きいβとtにおいてはそれらの乱数は高い精度でq_0に従うことになる. すなわち, f で最小値をとる x を見つけることができることになる. このようにΛ_βに従いβやtが十分に大きくなるまでXに値を取る乱数を発生させることをアニーリングと呼ぶ. しかし, 実際には, βもtも無限大になるまで計算を続けることはできないのでいかにうまくこれらの量を制御するかが問題となる. また, 最小値探索問題をアニーリング法によって解く場合に実際にここで述べた各種の行列が必要となるわけではない. これらの概念はむしろ後からアニーリングの正当性を裏付けるために導入されたように思われる. 実際にはつぎのようなアルゴリズムに沿ってXの要素 x を次々にシュミレートするばよいのである.

<アニーリングのアルゴリズム>

(1) βの初期値を決定する. t=0 とする.

(2) $x \in X$を一つ決める.

(3) βが十分大きくなるまで以下の(4)～(9)を繰り返す.

(4) t が十分大きくなるまで以下の(5)～(9)を繰り返す.

(5) 摂動行列Pを用いて x を摂動させ次の候補 y を選ぶ.

(6) $\Lambda_\beta(x,y)$を計算する.

(7) [0, 1]の一様乱数を発生させる. これをrandomとする.

(8) $\Lambda_\beta(x,y) >$ randomならば y を新たに x として記憶しておく.

(9) t:=t+1 .

(10) 得れられた x が f の最小値を与える基本状態である.

3. アニーリング法を用いたマルチアライメント

この節では, 前節で導入したアニーリング法をn本の遺伝子配列を整列化するアルゴリズムに応用する. そのためには基本状態, 目的関数, 摂動行列を設定すればよい. そうすれば, 前節の最後に示したアルゴリズムを用いてアニーリングを行うことができる. さて, これらをうまく設定する

ために２本の遺伝子配列を整列化するアルゴリズムを見直してみることにする．今，次に示す２本の遺伝子配列$\mathcal{A}$，$\mathcal{B}$考える．

$\mathcal{A}$：MNPQY

$\mathcal{B}$：MPQR

ここで，現在よく使われている２本の遺伝子配列を整列化するアルゴリズムを解析してみると，"２本の遺伝子配列を整列化すること"は，いわば，$\mathcal{A}$，$\mathcal{B}$を図 3.1 のような格子点グラフ上の２つの座標軸上に配置したときに，ちょうど"P_0(原点)からP_N(終点)にいたるパスの中で最適なものを見つけること"であることがわかる．例えば，図 3.1 のパス上の格子点を$\mathcal{A}$，$\mathcal{B}$の要素に対応させると

$\mathcal{A}$：MNPQY

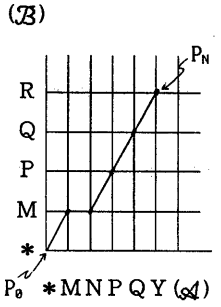
$\mathcal{B}$：MMPQR

が得られ，列$\mathcal{B}$でMが２つ現れるのはおかしいからどちらかを＊に置き換えると，

$\mathcal{A}$：MNPQY

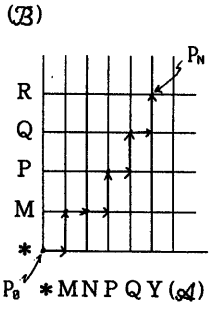
$\mathcal{B}$：M＊PQR

となる．こうして，パスと整列結果がよく対応していることがわかる．

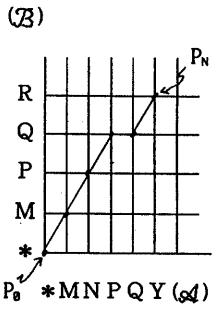


(図 3.1)

ところで，P_0からP_Nにいたるパスの数は，２本の配列を比較する場合(以降これを２次元と呼ぶ)でも，かなりの大きさになる．しかし，２次元の場合に限っては，動的計画法やそれに類似の手法を使って，これらのパスの中から最適なもの(２本の遺伝子配列の距離を最小とするもの)をうまく選択しているのである．このことをヒントに考えると，ｎ本の遺伝子配列を整列化することも原理的には，２本の場合と同様，ｎ次元格子点グラフのｎ本の座標軸上に，比較したいｎ本の遺伝子配列を配置し，原点から終点にいたるパスを考えれば良いように思える．そこで，アニーリング法を用いたマルチアライメントの一例として基本状態にパスをとるものを考えることにする．



(図 3.2)



(図 3.3)

図 3.1 のパスを単位ベクトルに分解したところ

x'から得られるパス

パスを基本状態にとるといっても，それがうまく表現できなければ結局アニーリングを実行することは不可能である．そこで，次にパスの表現方法を論じておく．具体的にもう一度２次元の場合，先に挙げた$\mathcal{A}$，$\mathcal{B}$を例にとって説明を進めていこう．パスはそれが通る格子点を順に起点，終点にするベクトルのつながりとみなすことができる．また，格子の一辺の長さを一単位とした各軸方向の単位ベクトルを考えると，各々のベクトルは単位ベクトルに分解することができるから，結局ｎ次元格子点グラフ上のパスは，ｎ種類の単位ベクトルをつらねた列になる(図 3.2)．この単位ベクトルの列は，それぞれの単位ベクトルをそれぞれある記号に対応させれば，記号列として扱うことができる．また，ある整列化において対象となるパスは全て同数の単位ベクトルで表すことができるから，記号列の長さが一定になり，パスの摂動を考えるときにも都合がよい．すなわち，一般にｎ本の配列の整列結果を示すパスはｎ種類の記号を用い，ｎ本の配列の長さを全て加えた長さの記号列で表現できる．例えば，図 3.1 の２次元格子

点グラフ上に現れるパスはx軸方向の単位ベクトルを"1"，y軸方向のそれを"2"と書くことにすると，全て5個の"1"と4個の"2"，からなる記号列に置き換えることができる．このとき，図 3.1 に描かれたパスは，

$$x = 121121212$$

表すことができる．また，記号列を 図 3.1 のようなパスのイメージに直すときには，記号を頭から読んでいき，いままでに読んだ記号の中に同じ記号が出てきたらその記号の前までの部分列がもともとある同一のベクトルを分解して得られた単位ベクトルの列であると解釈すればよい．例えば，$x' = 111212212$ は(1) (1) (12) (12) (21) (2)と考えるのである((·)が同一のベクトルを構成)．すると 図 3.3 のようなパスが得られることになる．さらにパスをこのように表現すると，摂動行列はつぎのように考えることができる．まず，記号(シンボル)列内のどこか異なった場所にある，異なった種類のシンボルを入れ換えてxを摂動させることにする．例えば，図3.1 上のパスについて$x = 12$ 1121212の4番目の1と7番目の2 を入れ換えれば $x' = 121221112$を得る．このxが示す整列化結果は，

$$\mathcal{A}：MNPQY$$
$$\mathcal{B}：M*PQR$$

であり，x'が示す整列結果は，

$$\mathcal{A}：MNPQY$$
$$\mathcal{B}：MPQ*R$$

である．さて，あるxに対して，上記のような操作で作ることができるパスの集合をS_x とする．このとき，次に推移するパスの候補をS_x の中から等確率で選ぶことにする．つまり，摂動行列のx行y成分を

$$P(x,y) \equiv \begin{cases} 1/|S_x| & (y \in S_x \text{のとき}) \\ 0 & (y \notin S_x \text{のとき}) \end{cases}$$

で与えるのである．このP(x,y)が[定義 2.2]の条件を満たしていることは容易に証明できる．

次に，目的関数について考えよう．今，あるパスxが表す遺伝子配列が，

$$\mathcal{A}_1 : a_1^1\, a_2^1 \cdots \cdots a_N^1$$
$$\mathcal{A}_2 : a_1^2\, a_2^2 \cdots \cdots a_N^2$$
$$\cdot$$
$$\cdot$$
$$\cdot$$
$$\mathcal{A}_m : a_1^m\, a_2^m \cdots \cdots a_N^m$$

であったとする．このとき，目的関数f(x)は次のように定めることができる．

$$f(x) = \sum_{n=1}^{N} \{ (\sum_{i=1}^{m-1} \sum_{j=i+1}^{m} d(a_n^i, a_n^j)) / m(m-1)/2 \}$$

ここで，

$$d(a_n^i, a_n^j) = \begin{cases} 0 & (a_n^i = a_n^j \text{のとき}) \\ 1 & (a_n^i \neq a_n^j \text{かつ } a_n^i \neq *, a_n^j \neq *) \\ 2 & (a_n^i \neq a_n^j \text{かつ } a_n^i = * \text{あるいは } a_n^j = *) \end{cases}$$

このfは$\mathcal{A}_i$，$\mathcal{A}_j$ (i≠j)間の平均距離をあらわしている．アライメントは比較したい遺伝子配列の違いが最も小さくなるようにするための操作であるから，目的関数として遺伝子配列の間で定義された距離をとることは妥当であると思われる．以上のように基本状態，摂動行列，目的関数を設定したものを，パス表現方式と呼ぶことにしておく．ここで設定した基本状態等を使い2節の終わりに示したアルゴリズムを用いればアニーリングによってアライメントを行うことが可能となる．これで，アニーリングによるマルチアライメントの一例が完成した．

4．計算機による実験結果と考察

ここでは実際に計算機を用いて我々が行った実験結果について述べておく．3節で導入した基本状態，目的関数，摂動行列を用いて実際にマルチアライメントを行った．整列化がより自然な形になっているかあるいはほんとうに目的関数の最小解になっているかを判断しやいように，実験には比較的短いつぎの6列を用いた．

$\mathcal{A}_1$: GGIPQGDVEKGKTIFKQRCAQCHTV

$\mathcal{A}_2$: GVPAGDVEKGKKLFVQRCAQCHTV

$\mathcal{A}_3$: GGIPQGFGGIPQGFVEKKTFIKQRC

$\mathcal{A}_4$: GGIPQGFVEKKTFIKQRCGGIPQGF

$\mathcal{A}_5$: GGIPQGFGGIPQGFVEKKTFI

$\mathcal{A}_6$: GGIPQGFVEKKTFIKQRCD

また，初期状態としては，

$\mathcal{A}_1$: GGIPQGDVEKGKTIFKQRCAQCHTV

$\mathcal{A}_2$: GVPAGDVEKGKKLFVQRCAQCHTV*

$\mathcal{A}_3$: GGIPQGFGGIPQGFVEKKTFIKQRC

$\mathcal{A}_4$: GGIPQGFVEKKTFIKQRCGGIPQGF

$\mathcal{A}_5$: GGIPQGFGGIPQGFVEKKTFI****

$\mathcal{A}_6$: GGIPQGFVEKKTFIKQRCD******

をとり，βの初期値は1.00，t は9591にしてアニーリングを実行した．これは，パスの表現に必要な記号列の長さが139であることに注意して，

$$_{139}C_2 = 9591$$

から算出した数である．また，βは1.5倍ずつ増やすことし，一つのβ中に一度もパスが変移しなくなったとき計算を中止することにした．このとき次のような結果を得ることができた(表 4.1)．

$\mathcal{A}_1$: GGIPQGDVEKGKTIFKQRCA*QCHTV

$\mathcal{A}_2$: GV*PAGDVEKGKKLFVQRCA*QCHTV

$\mathcal{A}_3$: GGIPQGFGGIPQGFVEKKTF*IKQRC

$\mathcal{A}_4$: GGIPQGFVEK*KTFIKQRCGGIPQGF

$\mathcal{A}_5$: GGIPQGFGGIPQGFVEKKTF*I****

$\mathcal{A}_6$: GGIPQGFVEK*KTFIKQRCD******

(表 4.1)

このとき，

$$f(x) = 17.9333\cdots$$

であった．この結果を評価する一つの指標として同じ6列を用いグループアライメント[7]によって得られた結果と比べてみた．グループアライメントもn本の遺伝子配列を同時に整列化するために開発されたアルゴリズムであるが，これは基本的に2本の遺伝子配列を整列化するアルゴリズムを用いているため理論的な段階で"結果が近似的な域を越えられない"という欠点がわかっているものである．しかし，ある程度自然な結果を短時間で得られるという利点があり，現在使われているn次元アライメントの代表的なアルゴリズムであると思われるものである．先に示した6列についてグループアライメントから得られた結果は次のようになった(表 4.2)．

$\mathcal{A}_1$: GGIPQGDVEKGKTIFKQRCA*QCHTV

$\mathcal{A}_2$: *GVPAGDVEKGKKLFVQRCAQC*HTV

$\mathcal{A}_3$: GGIPQGFGGI*PQGFVEKKTFIKQRC

$\mathcal{A}_4$: GGIPQGFVEK*KTFIKQRCGGIPQGF

$\mathcal{A}_5$: GGIPQGFGGI*PQGFVEKKTFI****

$\mathcal{A}_6$: GGIPQGFVEK*KTFIKQRCD******

(表 4.2)

また，このとき f(x)は，

$$f(x) = 18.1333\cdots$$

であった．表 4.1 と 表 4.2 のどちらがより自然な結果であるかを判断することはむずかしいが，f(x) の値がより小さいという点においてはアニーリングを用いたアライメントの方がよい結果になっていると言える．ただし，十分に時間をかけないと，表 4.1 のようなよい結果を得ることができない．すなわち，結果を得るまでの時間は，グループアライメントに分があるのが現状である．これは，式(2.1) から判るように，そもそもアニーリングは t，β を無限大にして初めて完全な解を与えることが保証されているので，短い実行時間で最適解を得ることはあまり期待できないのである．また，3節で導入したパス表現方式にも問題がなかったわけではない．パスを基本状態として考えたことは，2本の整列化アルゴリズムを念頭におけば，n次元への拡張としてはかなり自然な成り行きであったと思われる．これを活かすためにはどうしてもある種のコーディングを用いて

パスを表現しなければいけない．しかし，目的関数の値は遺伝子配列からでないと計算できないので，次の候補を捜すたびに，記号列と遺伝子配列の変換が必要となった．この変換が，計算時間にかなりの影響を与えていたようである．ところで，n本の遺伝子配列を整列化するということは，単に，n本の遺伝子配列を並列にし，その各配列にいくつかの＊を挿入し長さのそろったn本の配列をつくることであると考えることもできる．そう考えれば，基本状態として，n本の遺伝子配列を並列にしたもの(ちょうど表 4.1 や 表 4.2に示したような整列結果)を直接とることが可能となる．すなわち，アライメントしたい m 本のアミノ酸配列を，

$$\mathcal{A}_1 : a_1^1\, a_2^1 \cdots\cdots a_{N(1)}^1$$
$$\mathcal{A}_2 : a_1^2\, a_2^2 \cdots a_{N(2)}^2$$
$$\cdot$$
$$\cdot$$
$$\mathcal{A}_m : a_1^m\, a_2^m \cdots\cdot a_{N(N)}^m$$

(ここで，a_i^jはアミノ酸を示す．)

とするとき，集合Xとして

$$X = \left\{ x \mid x = \begin{matrix} a_1^1\, a_2^1 \cdots \cdots a_N^1 \\ a_1^2\, a_2^2 \cdots \cdots a_N^2 \\ \cdot \\ \cdot \\ a_1^m\, a_2^m \cdots \cdots a_N^m \end{matrix}, \right.$$

$$\left. a_i^j \text{はアミノ酸または＊，} N \geqq \max N(i) \in \mathbb{N} \right\}$$

をとるのである．このとき，摂動は，例えば次のように考えればよい．今，xが

$$x = \begin{matrix} a_1^1\, a_2^1 \cdots a_N^1 \\ a_1^2\, a_2^2 \cdots a_N^2 \\ \cdot \\ \cdot \\ a_1^m\, a_2^m \cdots a_N^m \end{matrix}$$

であるとして，

(1) m行の中から任意に1行選ぶ．

(2) その行の中から任意に2つの列を選ぶ．

(3) 選ばれた要素の種類によって以下の操作をつぎの3つにわける．

(イ) 一方が＊の場合はアミノ酸の場合

アミノ酸の前に＊を挿入し＊であった要素を削除する．

(ロ) 両方とも＊の場合

どちらか一方の＊の前に＊を挿入しもう一方の＊を削除する．

(ハ) 両方ともアミノ酸の場合

各行において任意に1列選びその列の前に＊を挿入する．

(4) こうして，得られたxにおいて全ての要素が＊となる列が発生した場合にはその列を削除する．

さて，上記の基本状態と摂動を用い，3節の目的関数で同様の実験をおこなったところ，実行時間は約1／3に減少した．しかし，グループアライメントよりも小さなfを与えるxを見つけようとすると，まだかなりの実行時間が必要であった．

5．まとめ

本報告では，最小値探索問題の解法の一つであるシュミレーティドアニーリングを用いることによるマルチアライメントのアルゴリズムを紹介しその具体例を示した．そして，比較的短い6本の列を用いて実際に整列化を行ってみた．その結果，アニーリングによるマルチアライメントの整列結果から計算されるf(x)の方が，既存の代表的なマルチアライメントアルゴリズムを用いた整列結果のそれよりも小さな値をとることができるということがわかった．このことから直ちにアニーリングによるマルチアライメントがより自然な整列結果を与えるとは言えないが，アライメントの精

度は遺伝子配列間に定義された距離（類似度）のとりかたに依存するものであるから，遺伝子配列の関係を正確に示す目的関数が見つかれば，我々の方法はより有効的になるであろう．実行時間について言えば，現状では，既存のマルチアライメントの方がはるかに速い．しかし，xの初期値の選び方やt，βの制御の仕方をもっと工夫すればかなりのグループアライメントの実行時間に近くなることが期待できる．一般の最小値探索では，最適解がどこにあるをかまったく予想できないのが普通である．ところが，例えば，こうしたマルチアライメントは相同タンパク質を構成するアミノ酸配列や似通ったDNA配列に対して用いられることが多い．すると，整列結果は比較したい遺伝子配列を単に並行にならべただけのものに近い様相になる場合が多いと考えられることになり，その結果，最適解の位置をだいたい想像することができる．すなわち，3節で述べたパス表現法の言葉をかりれば，結果となるパスはn次元格子グラフの対角線付近に集中すると思われるのである．したがって，xの初期値をほぼ対角線にとり，摂動も対角線からのある幅に絞ってアニーリングを実行すれば，より短い実行時間でより正確な結果を得ることが期待できるのである．また，本報告ではまったく触れなかったが，最小値探索問題の解法の一つでアニーリングよりも収束が速いニューラルネットの考え方を合わせることにより，より高速な方法も考案中である．

参考文献

[1] M.O.Dayhoff: "Atlas of Protein Sequence and Structure", N.B.R.F..

[2] S.B.Needleman and C.D.Wunsch: A General Method Applicable to Search for Similarities in the Amino Acid Sequence of Two Proteins, J.Mol.Biol.48 443-453 (1970).

[3] Emile Aarts and Jan Korst: "Simulated Annealing and Boltzmann Machines", John Wiley & Sons Ltd. (1989).

[4] 大矢，梅垣："量子論的エントロピー"，共立出版(1984).

[5] 大矢，宮崎，大島：アミノ酸配列の整列化法，Viva.Origino, 17, 139-151 (1989).

[6] M.Ohya: Information theortical treatments of genes, The Transctions of IEICE, Vol.E72, No.5,556-560 (1989).

[7] M.Ohya and Y.Uesaka: Amino Acid Sequences and DP Matching, to appear in International Journal of Information Science.

[8] P.H.Sellers: On The Theory and Computation of Evorutionary Distances, SIAMJ.APPL.MATH. Vol.26 No.4 (1974).

[9] T.F.Smith, M.S.Waterman, and W.M.Fitch: "Comparative Biosequence Metrics", J. Mol.Evol 18. 38-46 (1980).