

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第5055556号  
(P5055556)

(45) 発行日 平成24年10月24日(2012.10.24)

(24) 登録日 平成24年8月10日(2012.8.10)

(51) Int.Cl. F I  
 G O 6 N 3/00 (2006.01) G O 6 N 3/00 5 5 0 C  
 G O 1 N 23/207 (2006.01) G O 1 N 23/207

請求項の数 30 (全 73 頁)

(21) 出願番号	特願2007-549201 (P2007-549201)	(73) 特許権者	504139662 国立大学法人名古屋大学
(86) (22) 出願日	平成18年12月4日(2006.12.4)		愛知県名古屋市千種区不老町1番
(86) 国際出願番号	PCT/JP2006/324614	(74) 代理人	100081776 弁理士 大川 宏
(87) 国際公開番号	W02007/066787	(72) 発明者	坂田 誠 愛知県名古屋市千種区不老町1番 国立大 学法人名古屋大学内
(87) 国際公開日	平成19年6月14日(2007.6.14)	(72) 発明者	西堀 英治 愛知県名古屋市千種区不老町1番 国立大 学法人名古屋大学内
審査請求日	平成21年11月27日(2009.11.27)		
(31) 優先権主張番号	特願2005-350770 (P2005-350770)	審査官	新井 寛
(32) 優先日	平成17年12月5日(2005.12.5)		
(33) 優先権主張国	日本国(JP)		

最終頁に続く

(54) 【発明の名称】 ハイブリッドGAによる複数パラメータの最適化方法、パターンマッチングによるデータ解析方法、放射線回折データに基づく物質構造の推定方法、ならびに関連するプログラム、記録媒体お

(57) 【特許請求の範囲】

【請求項1】

各個体を構成する記号列のうち少なくとも一部は、所定の物理量に対応した複数のパラメータであり、これらのパラメータの適正な組合せを探索することを目的として、

これらのパラメータがそれぞれ適度に分布した複数の前記個体を初期世代のポピュレーションとし、

各世代において、当該世代のポピュレーションのうち少なくとも新たに生成された前記個体である新生個体の適応度を評価する評価過程と、このポピュレーションのうち複数の前記個体を親個体としてこれら親個体の間で行う交叉により新しい子個体を生成する交叉過程と、次世代に残すべき複数の前記個体を選択して他の前記個体を淘汰する淘汰過程とを繰り返すうちに、

これらの個体のうち一部に変更を加える突然変異過程を適宜交えながら、世代交代を重ねて前記適応度がより良好な前記個体の発生を待ち、もって前記複数のパラメータの最適な組合せを探索する、

GA（遺伝的アルゴリズム）による複数パラメータの最適化方法において、  
全世代のうち少なくとも一部にあたる複数の特定世代において、当該世代の前記ポピュレーションのうち一部にあたる特定個体を対象に、局所最適化操作が行われる局所最適化過程を有することと、

前記特定個体には、当該特定世代の前記ポピュレーションのうち前記適応度が比較的低い前記個体の中からも指定される可能性があり、前記適応度がより低い個体の方が、より

高い確率で前記特定個体に選定されることと、  
を特徴とする、

ハイブリッドGAによる複数パラメータの最適化方法。

【請求項2】

前記局所最適化過程は、前記全世代のうち一部にあたる前記複数の特定世代において、前記特定個体を対象に前記局所最適化操作を行う過程である、

請求項1に記載されたハイブリッドGAによる複数パラメータの最適化方法。

【請求項3】

前記局所最適化操作の対象たる前記特定個体は、当該世代の前記ポピュレーションのうち一つの前記個体である、

請求項1に記載されたハイブリッドGAによる複数パラメータの最適化方法。

【請求項4】

前記局所最適化過程が行われる前記特定世代は、前記全世代のうち少なくとも一部に散在している、

請求項1に記載されたハイブリッドGAによる複数パラメータの最適化方法。

【請求項5】

前記特定世代が各前記世代のうちに出現する頻度は、前記全世代のうち前記世代が下の方が高く設定されている、

請求項1に記載されたハイブリッドGAによる複数パラメータの最適化方法。

【請求項6】

前記特定世代は、前記全世代のうち少なくとも後半部で、所定の複数世代毎に周期的に出現する、

請求項1に記載されたハイブリッドGAによる複数パラメータの最適化方法。

【請求項7】

前記特定世代の前記ポピュレーションのうち、前記適応度が比較的低い複数個の前記個体からなるグループをもって、劣等グループと呼ぶことにし、

前記特定個体は、この劣等グループの中から指定される、

請求項1に記載されたハイブリッドGAによる複数パラメータの最適化方法。

【請求項8】

前記劣等グループは、前記ポピュレーションのうち前記適応度が低い方の所定数および所定割合のうち少なくとも一方に属する複数の前記個体からなるグループである、

請求項7に記載されたハイブリッドGAによる複数パラメータの最適化方法。

【請求項9】

前記特定個体は、当該世代の前記劣等グループの中からルーレット選択法およびランダム選択法のうち一方により選択される、

請求項7に記載されたハイブリッドGAによる複数パラメータの最適化方法。

【請求項10】

前記局所最適化過程は、前記特定世代の前記ポピュレーションの一部を前記特定個体として選び出して前記局所最適化操作を施す過程であり、

前記特定個体の選定に当たっては、前記適応度がより低い前記個体の方がより高い確率で選定されるという弱者優先の傾向を持つ、

請求項1に記載されたハイブリッドGAによる複数パラメータの最適化方法。

【請求項11】

前記淘汰過程では、当該世代のポピュレーションのうち前記適応度が比較的低い前記個体であっても次世代に生き残る可能性がある選択戦略が採られる、

請求項1に記載されたハイブリッドGAによる複数パラメータの最適化方法。

【請求項12】

前記淘汰過程で前記選択の際に採られるルールのことを選択戦略と呼ぶことにし、この選択戦略は、

この適応度とは関わりなく無作為に前記個体を選択されるランダム戦略と、

10

20

30

40

50

この適応度が比較的低い前記個体が優先的に選択される逆エリート戦略と、  
この適応度が比較的低い前記個体の方が選択される確率が高くなる逆ルーレット戦略と

、  
この適応度が比較的低い前記個体の方が優先的に選択されるようにした逆期待値方式、  
逆ランク方式および逆トーナメント方式とのうち何れかと、  
のうちいずれかである、

請求項 1 に記載されたハイブリッド G A による複数パラメータの最適化方法。

【請求項 1 3】

当該世代の前記ポピュレーションのうち前記適応度が最高である前記個体をトップ個体と呼ぶことにし、

当該世代からその次世代に世代交代する際に、このトップ個体は必ず生き残るというトップ保存戦略が併用される、

請求項 1 1 および請求項 1 2 のうち少なくとも一方に記載されたハイブリッド G A による複数パラメータの最適化方法。

【請求項 1 4】

各世代のポピュレーションを形成する個体のそれぞれは、所定の物理量に対応した複数のパラメータを含む記号列からなり、これらのパラメータがある程度の多様性をもって分布した初期世代から所定のルールで世代交代を重ねるうちに、より高い適応度をもつ前記個体が生成されることをもって、前記複数のパラメータの最適化を図る G A ( 遺伝的アルゴリズム ) による複数パラメータの最適化方法において、

全世代のうち少なくとも一部にあたる複数の特定世代において、当該世代の前記ポピュレーションのうち一部にあたる特定個体を対象に、局所最適化操作が行われることと、

前記特定個体には、当該特定世代の前記ポピュレーションのうち前記適応度が比較的低い前記個体の中からも指定される可能性があり、前記適応度がより低い個体の方が、より高い確率で前記特定個体に選定されることと、

を特徴とする、

ハイブリッド G A による複数パラメータの最適化方法。

【請求項 1 5】

請求項 1 ~ 請求項 1 4 のうちいずれかに記載された「ハイブリッド G A による複数パラメータの最適化方法」において、

各前記個体をもつ前記適応度は、当該個体の前記記号列に含まれる前記複数のパラメータに対応する前記物理量を仮定すると数値演算により決まる仮想データのパターンが、実測データのパターンとどの程度一致しているかを示す指標であり、

この適応度が高い前記個体を生成することによって、当該個体の前記仮想データと前記実測データとの間でパターンマッチングを行い、前記物理量を推定することを目的として前記実測データの解析を行うことを特徴とする、

パターンマッチングによるデータ解析方法。

【請求項 1 6】

請求項 1 ~ 請求項 1 4 のうちいずれかに記載された「ハイブリッド G A による複数パラメータの最適化方法」を、所定の試料物質に照射された放射線が形成する回折パターンに基づいてこの試料物質の構造を推定する物質構造の推定方法に応用し、

前記物理量のうち少なくとも一部は、この試料物質の物質構造を決定する構造パラメータであって、

各前記個体の前記適応度は、この試料物質が当該個体に相当するこれらの構造パラメータをもつと仮定すると生じる回折パターンが、実測された前記回折パターンに対してどの程度一致しているかを示す指標であることを特徴とする、

放射線回折データに基づく物質構造の推定方法。

【請求項 1 7】

前記構造パラメータとは、

前記試料物質の結晶構造を規定する格子定数と、

10

20

30

40

50

この試料物質を構成する分子内における三次元的な原子配列を規定する分子構造パラメータと、

この分子内での確率的な電子密度分布を規定する電子密度分布パラメータと、

この試料物質の結晶内での三次元的な原子配列を規定する結晶学的構造パラメータと、のうち少なくとも一つである、

請求項 1 6 に記載された放射線回折データに基づく物質構造の推定方法。

【請求項 1 8】

前記回折パターンは、その回折強度が同心円状に拡がって分布する回折環である、

請求項 1 6 に記載された放射線回折データに基づく物質構造の推定方法。

【請求項 1 9】

前記適応度は、前記回折環がもつ半径方向への強度分布の実測データと、仮定された前記構造パラメータに基づいて生成される仮定データとが、どの程度一致しているかを示す評価関数である、

請求項 1 8 に記載された放射線回折データに基づく物質構造の推定方法。

【請求項 2 0】

複数の結晶を含む粉末試料、多結晶集合体である多結晶試料、非晶質（アモルファス）試料および溶液試料のうち少なくとも一つである試料物質の構造を特定するために、

この試料物質に照射された放射線の回折によってもたらされる回折環の強度分布を示す回折データに基づき、この試料物質がもつ物質構造を示すいずれかの構造パラメータをデータマッチングにより推定するアルゴリズムとして、

前記構造パラメータがコーディングされた個体を、ある程度の多様性を持たせて複数個発生させ、初期世代のポピュレーションとする発生過程の後に、これらの個体の交叉および複製と、前記データマッチングの程度を示す適応度を評価してこれらの個体を淘汰する選択と、確率的に起こる突然変異とを繰り返す世代交代過程を行う G A（遺伝的アルゴリズム）を使用する、

放射線回折データに基づく物質構造の推定方法において、

前記 G A では、

前記世代交代過程の間に複数の世代において、当該ポピュレーションの一部である前記個体に対して局所最適化操作が行われることと、

この局所最適化操作の対象には、当該ポピュレーションのうち前記適応度が比較的低い前記個体もなる可能性があり、むしろ前記適応度がより低い個体の方が、より高い確率で当該操作対象に選定されることを特徴とする、

放射線回折データに基づく物質構造の推定方法。

【請求項 2 1】

放射線を試料物質に照射して得られた回折環の強度分布データに基づき、この強度分布データに適合する物質構造を特定するために、

この試料物質に含まれる結晶の格子定数を決定する格子定数決定プロセスと、

この強度分布データに含まれているバックグラウンド散乱の影響を除去ないし低減する背景散乱除去プロセスと、

この結晶が属する空間群の候補を絞り込む空間群絞り込みプロセスと、

この結晶の基本構造を初期的に決定して、この空間群を一つに特定する基本構造決定プロセスと、

この基本構造を精密化して、この結晶を構成する物質の構造をより精密に決定する構造精密化プロセスとのうち、

少なくとも、この格子定数決定プロセスと、この空間群絞り込みプロセスと、この基本構造決定プロセスとを有し、

この格子定数決定プロセスおよびこの基本構造決定プロセスのうち少なくとも一方で、G A（遺伝的アルゴリズム）に局所最適化操作を組み合わせたハイブリッド G A を利用した物質構造決定が行われる、

放射線回折データに基づく物質構造の推定方法であることを特徴とし、

10

20

30

40

50

前記ハイブリッドGAでは、世代交代の間に複数の特定世代で、当該世代のポピュレーションのうち一部の個体である特定個体を対象にして前記局所最適化操作が行われ、

この特定個体には、当該世代のポピュレーションのうちでは実測データとの一致度を示す適応度が比較的低い劣等グループからも選定される可能性があることを、さらに特徴とする、

放射線回折データに基づく物質構造の推定方法。

【請求項22】

所定の物理量に対応した複数のパラメータの適正な組合せを探索する機能を、所定のコンピュータに発揮させるべきコンピュータ・プログラムにおいて、

請求項1～請求項13のうちいずれかの請求項に記載された「ハイブリッドGAによる複数パラメータの最適化方法」を実行するために、

当該請求項に記載された前記初期世代の前記ポピュレーションの生成および読込の一方を行う初期世代設定ステップと、

当該請求項に記載された前記評価過程、前記淘汰過程、前記交叉過程、前記突然変異過程および前記局所最適化過程のそれぞれに対応する各演算ステップ（すなわち、評価ステップ、淘汰ステップ、交叉ステップ、突然変異ステップおよび局所最適化ステップ）とを有することを特徴とする、

コンピュータにより実行可能な最適パラメータ探索プログラム。

【請求項23】

所定の物理量に対応した複数のパラメータの適正な組合せを探索する機能を、所定のコンピュータに発揮させるべきコンピュータ・プログラムにおいて、

各世代のポピュレーションを形成する個体のそれぞれは、前記所定の物理量に対応した前記複数のパラメータを含む記号列からなり、これらのパラメータがある程度の多様性をもって分布した初期世代から所定のルールで世代交代を重ねるうちに、より高い適応度をもつ前記個体が生成されることをもって、前記複数のパラメータの最適化を図るGA（遺伝的アルゴリズム）の中に、

前記適応度に関する局所最適化操作を所定の前記個体に対して行う局所最適化ステップが組み込まれており、

この局所最適化ステップは、前記GAで生成される全世代のうち少なくとも一部にあたる複数の特定世代において、当該世代の前記ポピュレーションのうち一部にあたる特定個体を対象にする演算ステップであることと、

この局所最適化ステップにおいて、前記特定個体には、当該特定世代の前記ポピュレーションのうち前記適応度が比較的低い前記個体の中からも指定される可能性があり、前記適応度がより低い個体の方がより高い確率で前記特定個体に選定されることを特徴とする、

コンピュータにより実行可能な最適パラメータ探索プログラム。

【請求項24】

複数のパラメータの適正な組合せを探索する機能を、所定のコンピュータに発揮させるべきコンピュータ・プログラムにおいて、

各世代のポピュレーションを形成する個体のそれぞれは、前記複数のパラメータに対応した記号列を含んでおり、これらのパラメータがある程度の多様性をもって分布した初期世代のポピュレーションから始めて、所定のルールで世代交代を重ねるうちに、より高い適応度をもつ前記個体が生成されることによって前記複数のパラメータの最適化を図るGA（遺伝的アルゴリズム）の中に、

前記適応度に関する局所最適化操作を所定の前記個体に対して行う局所最適化ステップが組み込まれており、

この局所最適化ステップは、前記GAで生成される全世代のうち少なくとも一部にあたる複数の特定世代において、当該世代の前記ポピュレーションのうち一部にあたる特定個体を対象にする演算ステップであることと、

この局所最適化ステップにおいて、前記特定個体は、当該特定世代の前記ポピュレーシ

10

20

30

40

50

ヨンのうち前記適応度が比較的低い前記個体の中からも指定される可能性があり、前記適応度がより低い個体の方が、より高い確率で前記特定個体に選定されることとを特徴とする、

コンピュータにより実行可能な最適パラメータ探索プログラム。

【請求項 2 5】

所定の試料に照射された放射線が形成する回折パターンに基づいて当該試料を構成する物質の構造を規定する構造パラメータを推定する機能を、所定のコンピュータに発揮させるべきコンピュータ・プログラムにおいて、

前記回折パターンの実測データを数値的に読み込む実測データ読込ステップと、

請求項 2 2 に記載された「最適パラメータ探索プログラム」が有する前記初期世代設定ステップ、前記評価ステップ、前記淘汰ステップ、前記交叉ステップ、前記突然変異ステップおよび前記局所最適化ステップを有し、

前記適応度は、前記各個体の前記記号列に相当する仮定された前記構造パラメータから算出される前記回折パターンの予想値が、前記回折パターンの実測値に対してどの程度一致するかを示す指数であることを特徴とする、

コンピュータにより実行可能な放射線回折データに基づく物質構造推定プログラム。

【請求項 2 6】

所定の試料に照射された放射線が形成する回折パターンに基づいて、当該試料を構成している物質の構造を規定する構造パラメータを推定する機能を、所定のコンピュータに発揮させるべきコンピュータ・プログラムにおいて、

各世代のポピュレーションを形成する複数個の個体のそれぞれは、前記構造パラメータに対応する記号列を含んでおり、

これらの個体のそれぞれがもつ適応度は、当該個体に対応したこの構造パラメータから算出される回折パターンの予想値が、前記回折パターンの実測値とどの程度一致するかを示す指標であって、

この構造パラメータの各要素がある程度の多様性をもって分布するように、これらの個体を生成して初期世代ポピュレーションとする生成ステップをもち、

この初期世代ポピュレーションから淘汰ステップ、交叉ステップおよび突然変異ステップを含む所定の手順に従って世代交代を重ねるうちに、当該世代ポピュレーションの中にこの適応度がより高い前記個体が生成されることを待ち、もって前記構造パラメータの最適化を図る G A (遺伝的アルゴリズム) の中に、

前記適応度に関する局所最適化操作を所定の前記個体に対して行う局所最適化ステップが組み込まれており、

この局所最適化ステップは、前記 G A で生成される全世代のうち少なくとも一部にあたる複数の特定世代において、当該世代の前記ポピュレーションのうち一部にあたる特定個体を対象にする演算ステップであることと、

この局所最適化ステップにおいて、前記特定個体には、当該特定世代の前記ポピュレーションのうち前記適応度が比較的低い前記個体の中からも指定される可能性があり、前記適応度がより低い個体の方が、より高い確率で前記特定個体に選定されることとを特徴とする、

コンピュータにより実行可能な放射線回折データに基づく物質構造推定プログラム。

【請求項 2 7】

請求項 2 2 ~ 請求項 2 4 のうちいずれかに記載された「最適パラメータ探索プログラム」と、

請求項 2 5 および請求項 2 6 のうち一方に記載された「放射線回折データに基づく物質構造推定プログラム」とのうち、

少なくとも一方が記録されていることを特徴とする、

コンピュータ読み取り可能な記録媒体。

【請求項 2 8】

請求項 2 2 ~ 請求項 2 4 のうちいずれかに記載された「最適パラメータ探索プログラム

10

20

30

40

50

」と、

請求項 2 5 および請求項 2 6 のうち一方に記載された「放射線回折データに基づく物質構造推定プログラム」とのうち、

少なくとも一方がインストールされていることを特徴とする、  
ハイブリッド G A 搭載コンピュータ。

【請求項 2 9】

請求項 2 2 ~ 請求項 2 4 のうちいずれかに記載された「最適パラメータ探索プログラム」と、

請求項 2 5 および請求項 2 6 のうち一方に記載された「放射線回折データに基づく物質構造推定プログラム」とのうち、

少なくとも一方を実行可能なことを特徴とする、  
ハイブリッド G A を搭載した解析装置。

【請求項 3 0】

所定の試料物質に放射線を照射して、この試料物質から生じる回折パターンを測定する測定装置と、

この回折パターンから得られた実測データに基づいた解析を行うことにより、当該試料物質の構造を推定する請求項 2 8 記載のコンピュータおよび請求項 2 9 記載の解析装置のうち一方と、

を有することを特徴とする、

放射線回折パターンによる物質構造推定装置。

【発明の詳細な説明】

【技術分野】

【0 0 0 1】

( 1 ) 物質構造決定

本出願は、粉末試料に X 線を照射して得られた回折環がもつ強度分布のピーク位置データに基づいて、この試料を形成する物質の構造を同定しようという、発明者らの研究から生まれた一連の発明を特許出願するものである。

それゆえ本発明は、放射線回折により試料の物質構造を特定する「物質構造決定技術」の分野に属する。より詳しくは、例えば粉末試料に照射された X 線等の回折パターンを解析して同試料の物質構造を推定する「放射線回折データの解析技術」の分野に属する。

ふつう「物質構造」とは、その物質の結晶構造と、それに含まれる格子構造および分子構造との三者のうち、少なくともいずれかを指す言葉である。さらに物質の「結晶構造」とは、「格子構造」（当該結晶の格子定数）と、「分子構造」（同結晶内における分子の配列）との両者を含む技術用語である。ただし、「物質構造」とは、このような分子構造パラメータだけに限定されるものではなく、電子密度分布パラメータや結晶学的構造パラメータなどを指すこともある。また、ここで言う「構造決定」とは、このような物質構造のうちいずれかを可能な限り正しく精密に推定することである。

すなわち本発明は、単結晶試料ばかりではなく、多数の微細な結晶粒からなる粉末試料など、その放射線回折パターンが回折環の強度分布データとしてしか得られない試料物質についても、その物質の結晶構造や分子構造を高精度で決定することを可能にする技術である。それゆえ本発明によれば、ある程度以上の大きさを持つ単結晶を作ることが難しい物質の構造決定ができるようになるから、未知の分子構造をもつ医薬品や新素材などの構造決定に好適である。

( 2 ) ハイブリッド G A

本発明はまた、ハイブリッド G A によって多数のパラメータを最適化する「パラメータ最適化」ないし「最適パラメータ探索」の技術分野や「パターンマッチングによるデータ解析」の技術分野にも属する。

そして、これら一連の本願発明の中核となる発明概念は、遺伝的アルゴリズム（略称「G A」）に特定の方式で局所最適化操作を組み込んでハイブリッド化することにより、収束性を向上させたハイブリッド G A それ自体である。

10

20

30

40

50

なお、局所最適化操作が組み込まれたGAについては、いくつかのバリエーションがすでに公知であり、それらについては、以下の背景技術の項で公知文献を挙げて説明する。そして英語圏では、このようなアルゴリズムを、用不用説を唱えた先達の名を冠して「ラマルキアンGA」と呼ぶことが多い。一方、日本では同じものを「ハイブリッドGA」と呼ぶことが多いので、本明細書でもハイブリッドGAの名称を使うことにした。また、必要に応じてこれを「HGA」と略称してもよいものとする。

#### 【背景技術】

#### 【0002】

### 1. X線回折による物質構造決定技術の概況

#### (1) 直接法による構造決定

材料や医薬品などの研究開発において最も重要で基盤となる情報は、その材料の原子配列に関する情報である。すなわち、当該物質の分子内における立体的な原子配列がどうなっているかが分かれば、その物質の機能のある程度は予測することができるようになり、種々の応用研究を進めていくことができる。

このような状況は、その分野にかかわらず、分子性の物質の研究開発においてもほとんど変わるところがない。すなわち、分子を設計し、新たに創生した時、最初に求められるのは目的の分子ができたかどうかを判別する原子配列の決定である。ここで、原子配列が決定されてしまえば、目的としていた分子ができたかどうかを確認されるだけでなく、他にも重要な情報が得られる。たとえば、互いに隣り合う二つの原子を隔てる原子間距離から、両原子の結合形態（炭素原子ならその結合状態が2重結合か1重結合か）や、各原子のイオン価数など、分子の特性決定づける多くの情報を得ることができる。

さて、このような原子配列を決定する方法として、最も有力で直接的な証拠を与えられるのがX線回折法である。すなわち、他の構造測定手法には、たとえば核磁気共鳴(NMR)法などがあるが、これらの手法ではふつう状況証拠しか与えることができない。しかし、これに比べて、X線回折法によれば、当該分子中のすべての原子の立体的な配列を十分な精度(およそ0.01nm以下)で決定することができる。一方、無機材料に対しては原子を観測する手法として電子顕微鏡観察があるが、分子性の結晶は電子線に弱いので、このような材料に対しては、電子顕微鏡による原子配列の観測は不可能である。

現在、一般的なX線回折分野の常識として、分子性結晶の原子配列決定には、単結晶試料が必要である。なぜならば、単結晶試料を生成することさえできれば、タンパク質などの巨大分子を除いて、大半の材料についてほぼ自動的に原子配列(構造)を決定することができるからである。このように結晶試料のX線回折による構造決定方法は、1985年にノーベル賞を受賞したJ. Karle, H. A. Hauptmanによって開発された結晶構造を直接決定する方法であり、直接法と呼ばれている。

そして、X線装置技術の進歩と解析に用いるコンピュータの計算能力の進歩により、現在では回折装置とPC(パーソナルコンピュータ)1台があれば構造が決定できるシステムが構築され広く普及している。国内外の大学では理工系学科に少なくとも1台は、単結晶X線回折装置が設置されているのが普通である。また、このような回折装置は、医薬品関連企業や材料開発を行っている企業関係にも広く普及している。

#### (2) 単結晶の生成という難関

ところで、このような単結晶X線回折による原子配列決定方法を実行するうえで、避けて通れない最大の難関ないし隘路(ボトルネック)は、単結晶試料を得ることにある。なぜならば、単結晶試料として使用できるほどの大きさをもつ結晶の生成は、なかなか容易ではないことが多いからである。

すなわち、自然界に存在する物質のなかで単結晶で存在するものは、たいへんに稀であり、ほとんど存在しないと行って良い。仮に大きな単結晶が例外的に存在したとしても、それは自然界ではごく稀であるがゆえに高価な宝石であったりする。また、人工的に大型単結晶の作成が可能で実用化されている材料も、半導体材料(たとえばシリコン)くらいしかない。このように、単結晶の作成とは、通常では存在する状態ではない物質の状態を作り出すプロセスであるともいえる。一般に、単結晶を作成するためには、様々な結晶化

10

20

30

40

50

条件で試料物質の結晶化を試みる必要がある。そして、このプロセスにはかなりの年月と大量の試料とが必要とされ、膨大な労苦を費やすことになる。例えば、仮にある物質の構造決定に十年間を要したという場合に、そのうち八年間はX線回折に十分な大きさの単結晶を作ることには費やされていたということも珍しくなかった。

このような事情から、単結晶の生成に成功して実際に原子配列を決定することができるのは、創生される多数の物質のうちほんの数%程度にしか過ぎない。

こうして、単結晶試料が生成できないために原子配列が決定できないことは、分子性材料を用いた開発研究の大きな足かせとなっている。たとえば医薬品などでは、所望の分子ができていないかどうか分からないせいで、せっかく新物質を開発しても結局は薬として使うことができないことも少なくない。また、機能性材料の場合にも、電気伝導度などで興味深い特性を持った物質が創成できたとしても、その特性なり機能なりが、こういった物質構造に起因しているかを知ることができない。

10

以上のように、単結晶の試料物質を作れないのでX線回折パターンによる直接法で構造決定ができず、物質設計のための情報が不足してしまっていた。その結果、結局のところ場当たり的で試行錯誤が無駄に多い物質開発に終始せざるを得なかったというのが、今までの状況であった。

## 2. 代表分野での現状

以上のような理由で、粉末試料からその物質構造を決定する技術の開発は、分子性の材料を研究開発の対象とした極めて広い分野で切望されている。このことを具体的に理解していただけるように、以下では医薬品開発の分野とナノテクノロジー材料開発の分野とを例にとって、背景技術の現状を説明する。

20

### (1) 医薬品開発

一般的に言って、医薬品(ないし薬物)は粉末として用いられることが多いから、医薬品の研究開発において本来必要とされるのは、粉末状態での物質構造である。

ところが、一般に有機分子から構成される薬の場合、前述のように粉末では構造を決められないため、長期間をかけた苦労して単結晶を作成して構造を決定している。最近では、安全性などの観点から、物質構造を確認してからでなければ医薬品として認可されない場合が多い。すると、せっかく薬効のある物質が発見されても、その物質で大きな単結晶を作ることができないがために、その立体的な分子構造が解明できず、それが故に医薬品として用いられていないという状況に至ることも物質も少なくない。

30

そこで、もしも粉末試料からその薬品の分子構造が決定できるようになれば、新薬開発を飛躍的に進められることになる。そればかりではなく、医薬品開発の分野では次のような利点も生じる可能性がある。

すなわち、医薬品に用いられる有機分子は、たとえ分子式や構造式の表記は同じであっても、分子中の原子の立体配列が異なっていたり、結晶としては異なる構造で存在したりするものが多い。このように、構造式の表記などは同じであっても結晶構造が異なることがある分子を、医薬品の分野では「結晶多形」をもつ分子と呼んでいる。

そして、ある医薬品の分子が結晶多形である場合には、同じ構造式をもつ分子からなる粉末であっても、結晶構造が異なっているために、融点や密度、熱力学的安定性、溶媒に対する溶解度などが異なることがある。それゆえ、結晶多形のうち目的の相とは別の相を用いてしまうと、意に反して水に溶けず薬効を示さなかったりするなど、不都合を生じる可能性がある。それゆえ、この結晶多形という性質は、医薬品開発の分野では重大な問題である。

40

そればかりではなく、錠剤の作成など、医薬品を製品化するプロセスでも、粉末試料が圧縮されたり、熱処理されたりことがあるので、分子性の結晶は、これらのプロセスでも構造が変わる可能性がある。これらのプロセス中での構造の変化を追跡するには粉末試料からの構造決定が必要であるが、現状では粉末試料から構造を決定できる手法がないため行われていない。

以上のように、粉末試料から医薬品の構造を決定することができるようになれば、医薬品開発が飛躍的に加速され、最終的には世のすべて人々の健康の増進に繋がるのだが、そ

50

の手段として今までは実用に耐えるものがなかった。

## (2) ナノテク材料開発

有機伝導体やフラーレン関連物質など分子性材料は、十年先程度の近未来であってもナノテクノロジーの基幹物質として期待されており、国内外の多数の研究者によって研究が進められている。そして、これらの研究でもX線回折による構造情報は、材料開発の初期段階で最も重要である。

たとえば、現在ではボーリングのボール、テニスラケット、ゴルフクラブのヘッド、化粧品など10品目を超える製品に利用されるようになった炭素のサッカーボール型分子フラーレンも、発見当初の研究では、構造の決定が最も重要なテーマであった。すなわち、構造が決定された後に、ようやく超伝導体や磁性体など機能性材料の生成過程や様々な特性が、構造に基づき理解され、応用研究が進められていった。そして現在では、すでに年産数十トン程度の生産能力を持つフラーレン生成用の商用プラントが整備されており、さらに多くの製品への利用が始まっている。

さて、フラーレンに限らず、分子性の材料であれば結晶の構成単位が分子であるので、シリコンなどの無機材料と比較して生成可能な物質の種類は膨大な数に上る。それゆえ、研究者らによりおびただしい数の新物質が合成されているが、その構造決定には単結晶試料が必要とされていたがゆえに、未だに構造が明らかにされていない物質が多い。そして前述のように、構造が明らかにならないこと自体が、新物質生成の研究を推進するうえでの足かせになっている。

たとえば、新物質の物性測定などを先に進め、苦勞の末ようやく単結晶ができた時点でその物質構造が決定されてみれば、予測していた分子とは違っていたとか、別の方法で生成された分子と同じだったとかということは、よくあることである。しかし、そういうことがあると、それまでの研究に投下された資源は、ほとんどが無駄になってしまうわけである。

したがって、物質生成の初期であっても単結晶が作れない段階であっても、その粉末試料から構造決定ができる技術があれば、これらの研究を飛躍的に加速することができる。それゆえ、粉末試料からの構造決定技術は、分子材料を用いたナノテクノロジーの分野においても、その基盤技術となる可能性を持っているのだが、残念ながらその手段がなかったという実情が本発明の背景にある。

## 3. 粉末試料による構造決定が難しい理由

### (1) 概要

以上の説明では、もし粉末試料からの物質構造の決定技術があれば、医薬品や材料の開発に決定的な役割を果たすであろうことは明白でありながら、それが実現されていないということを述べてきた。では、粉末試料の構造決定がこれだけ切望されている技術でありながら今まで実用化されていなかったのは何故であるか、その理由の一端をここで説明しておく。

その理由を簡明に言うならば、あまりにも少ない情報量に基づいて、あまりにも多くの自由度を持つ物理量をしかも精密に求めることが要求されたので、今までの最適パラメータ探索技術の手に余ったからである。

### (2) 回折環の情報量

まず、粉末試料から得られたX線回折環に含まれる情報量は、極めてわずかではないことから説明する。

仮に、当該物質からある程度大きな(例えばサブミリメートル程度の)結晶を生成することができたものとし、この結晶を試料としてX線回折試験装置にかけた場合を考えてみよう。すると、図1(a)に示すように、その結晶に入射したX線は回折し、回折強度が格子状(ないし碁盤目状)に分布した回折パターンを生じる。

逆に、当該物質で回折試験に必要な大きさの結晶を作ることができず、仕方なしに当該物質の粉末試料をX線回折試験装置にかけた場合を考えてみよう。すると、図1(b)に示すように、その粉末試料に入射したX線は同軸の円錐状に回折し、回折強度が同心円状に分布した回折パターン(回折環)を生じる。

ここで、図1(a)の二次元的な回折パターンと図1(b)の回折環とを比較してみると、回折環の方は回折角の大きさに応じ半径方向に沿って回折強度が一次的に分布した強度分布データしか得られないことがわかる。それゆえ、粉末試料のX線回折である回折環からもたらされる情報量は、単結晶試料から得られた二次元的な回折パターンがもつ情報量に比べてずっと少ない。

そればかりではなく、試料結晶の格子面(hkl)の向きによってX線回折の方向および回折角が異なってくるから、粉末試料から得られた回折環では、その粉末試料に含まれている多数の微細な結晶が生じる回折ピークが互いに重複していることが多い。すると、互いに重なったピークを分離することができなくなり、回折環に含まれる情報はいよいよ少なくなってしまう。そして、粉末試料の回折強度分布では、あまりに広い角度範囲に少数のピークが分散しているので、多くの場合、回折データは結局のところ、ピーク位置とピーク強度との対を示す数値データに落とし込まれる。

また、定量的に説明すると、現行のX線回折装置ないしそのX線検出器から求めることができる精度は、回折角度 $2\theta$ にしておおよそ $0.1 \sim 0.01^\circ$ 程度の桁にある。一方、回折角度 $2\theta$ の範囲は $150^\circ$ くらいまでだから、その解像度ないし分解能は、階調数にして10の3~4乗程度でしかない。なお、回折環におけるX線強度の階調数(つまり明暗の段階数)は、10の3~6乗程度である。

### (3) 格子定数を決定したい場合

一方、試料物質の結晶が持つ格子定数を決定したい場合には、以上のように回折環に含まれる情報が極めて少ないのに対し、格子定数の決定に要求される精度は極めて高く、その解空間が極めて広いことが困難さの原因となっていた。

すなわち、粉末試料の回折環から1次元の強度分布データに基づいて、六つ一組の格子定数(a, b, c,  $\alpha$ ,  $\beta$ ,  $\gamma$ )と原点移動パラメータ(t)とを決定する際には、格子定数のそれぞれの決定に要求される精度が非常に高いことが問題になる。それに比べ、多くの場合、格子定数を決めるべき試料物質の標準的なa, b, cの値は10~30程度であるのに対し、これらの決定に要求される精度は $0.01 \sim 0.001$ 程度とかなり細かい。それゆえ、当該パラメータ(a, b, c)の探索空間での決定精度は、階調(値の取りうる幅に対する分解能)にして、各パラメータにつき10の5~6乗程度の桁で要求される。

したがって、決定すべき構造パラメータが格子定数の場合に、その要求精度を階調で表すと、a, b, cがそれぞれ10の5~6乗程度の桁に、 $\alpha$ ,  $\beta$ ,  $\gamma$ がそれぞれ10の4~5程度の桁になり、いずれにしる非常に大きいことになる。それに加えて、原点の位置ずれに関するパラメータtも、10の2~3乗程度の桁をもつ階調で決定しなければならない。

その結果、探索すべき複数パラメータが解空間で取りうる場合の数は10の29~36乗程度と、極めて大きいものになっているにもかかわらず、前述のように回折環のもつわずかな情報量からこれを決定しなければならない。格子定数の決定においては決定したいパラメータ数がわずか6個であっても、回折環に基づいてこれを決定することの難しさは実にここにある。

### (4) 分子の結晶構造を決定したい場合

この場合には、分子を構成する原子の数が多くなり、分子内での回転可能な結合部分が増えていくにつれて、分子内回転の自由度が増していく。このことから、複雑な分子であれば、その分子内で各原子がどのように三次元配置されているかを示す結晶構造を決定することは、多くの自由度をもつ多次元の解空間内で解を探索することになる。それゆえ、分子構造の自由度が増えるだけ(つまり決定すべきパラメータの数が増えるだけ)、指数関数的に構造決定が困難になっていくことは明らかである。

### (5) 遺伝的アルゴリズム(GA)の適用

先ほど、解空間の取りうる場合の数は極めて大きいことを説明したが、その数を控えめに見積もって10の20乗程度とし、使用可能なコンピュータで毎秒1万回の試行演算ができたと仮定してみよう。すると、絨毯爆撃方式の全探索を行うには3億年間程度かかる

10

20

30

40

50

ことになり、これは明らかに受け入れがたい計算時間である。

そこで、もう少し賢明に最適パラメータの探索を行い、物質構造の決定に要する計算時間を短縮することができるように、様々な試みがなされている。その中で、人工知能の分野で研究が盛んにされていたGA（遺伝的アルゴリズム）を用いる試みもなされており、本発明はこのような背景の中で行われた研究から生まれてきたものである。

次項には、発明者らが調査できた範囲で本発明に関連があると思われる公知文献を列挙し、ごく簡単に解説した。これらの中には、物質構造の決定方法に関する文献ばかりではなく、GAを改良したハイブリッドGAに関する文献もある。

#### 4．公知文献の簡単な説明

##### (1) 特許文献

###### ・特許文献1

この文献には、所定の空間群に属する斜方晶系のタンパク結晶の一種について、液滴蒸気拡散法によって所定の大きさの結晶を成長させ、単結晶試料のX線回折により複合同位体置換法を用いて三次元構造決定を行った旨が記載されている（第0010段落～第0011段落）。この技術では、X線回折試験に供することができるだけの大きさに単結晶試料を成長させており、回折環データの解析による物質構造決定は行われていない。

###### ・特許文献2

この文献には、X線などを試料に照射して、薄膜構造解析、単結晶構造解析や粉末構造解析などの物質構造解析を行う技術が開示されている。この技術の特徴は、構造パラメータの決定を高精度で短時間に行えるように、普通のGA（遺伝的アルゴリズム）またはその改良版である $\mu$ GAで最良解を探索した後、その最良解（適応度最高の個体に相当）に局所最適化を施していることである（同文献の図1参照）。なお、この技術で用いるGAや $\mu$ GAでは、適応度が高いエリート個体は次世代にそのまま残し、他の個体を親にして交叉を行っている（同文献の第0018段落および図2参照）。

この技術では、GAないし $\mu$ GAによる最適パラメータ探索が、その最良解を正解の近傍に収束させるまで待たなければならないが、そこまで待てば構造パラメータの決定を高精度化することができる。それゆえ、この技術によれば、単なるGAや $\mu$ GAに比べて構造パラメータの決定に要する計算時間を短縮することに、ある程度の効果を見込むことができる。

しかし、GA探索によってその最良解が正解の近傍に至るまでは局所最適化をかけても無駄であるから、最良解が正解の近傍に至ったと確信できるまでは待たなければならない、計算時間の短縮効果は限定的でしかない。

すなわち、基本的に言ってこの技術は、GAや $\mu$ GAで行う広域探索の最終段階でしか局所最適化をかけることができず、個体の進化に方向性を与えるものではない。そうではなく、この技術は、GAや $\mu$ GAによる確率的な広域探索では難しい最後の詰めを、確定的な操作である局所最適化によって加速するに過ぎない。そういう意味で、この技術は英語圏でいう Lamarckian GA（日本語表記では「ラマルキアンGA」）とは、いくぶん主旨が異なるようである。

###### ・特許文献3

この文献には、試料ホルダとその周辺部とに工夫を加えて、散乱X線による影響を低減したX線回折計測装置が開示されている。

###### ・特許文献4

この文献には、X線や放射光による粉末回折データの測定において空間分解能を上げるために、検出器をそのピクセル間隔よりも細かく移動させ、ピクセル間を補間するデータを測定する技術が開示されている。この技術は、一次元の検出器にも二次元の検出器にも適用可能である。

###### ・特許文献5

この文献には、X線回折試験装置において、粉末試料を収容したキャピラリーが細かったり、試料自体が細い線状であったりする場合にも、小さな試料を取り付ける際の作業が容易なる工夫がされた試料ホルダが開示されている。

10

20

30

40

50

## ・特許文献 6

この文献には、プロテインキナーゼの突然変異体を結晶化させてその構造を決定する方法とその結果とが開示されている。

## ・特許文献 7

この文献には、最適化問題に遺伝的アルゴリズムを適用するにあたって、ポピュレーションの多様性を維持する工夫が提案されている。その工夫の特徴は、ポピュレーションをいくつかの「島」と呼ばれる集団に分けて島の間で個体を交換したり、「タワー」と呼ばれる島の階層構造を導入したりすることのようである。

そして、同文献の「発明の名称」には、その冒頭に「ハイブリッド型遺伝的アルゴリズム」という文言があるが、そのハイブリッド型遺伝的アルゴリズムとはどのようなものであるのかは明記されていないようである。すなわち、「既存型アルゴリズムと組み合わせられた遺伝的アルゴリズム」（同文献の請求項 1 および第 0 0 6 3 段落）という程度の説明はあるが、これでは具体性を欠いている。さらに、「この最適化問題を解くために高速な既存型アルゴリズムがあると仮定する。ここで、既存型アルゴリズムとは、ヒューリスティックな解法に基づくアルゴリズムである。」（第 0 0 0 6 段落）とあっては、同文献の記載は仮定の話なのかと読者は思わずにいられない。そればかりではなく、同文献では具体的な数値例は一例も開示されておらず、「実施例」の項にある記載内容はむしろ「産業上の利用可能性」の項に記載すべき内容である。

それゆえ、同文献では所期の課題を解決することができるのかという疑問と、同文献は発明として完成されているのかという疑問とが払拭されない。また、単に適応度の高い解を探すことが目的である場合ではなく、X線回折による物質構造推定のように単一の正解が存在するデータ解析に適用することが目的である場合には、実用性があるのか否かも疑問である。

## ・特許文献 8

この文献には、遺伝的アルゴリズムを用いてパラメータの最適化ないし最適パラメータの探索を行うにあたり、遺伝的アルゴリズムの交叉過程にパラメータの連続性を考慮した工夫を取り入れた技術が開示されている。

すなわち、同文献では、その第 0 0 4 1 段落に記載されているように、交叉の際に両親として選択された二つの親個体について、まずそのパラメータ空間における中点を求め、次にその中点を中心とする正規分布などの連続的な確率分布に従って二つの子個体を生成する。この際、同文献の図 8 を参照して第 0 0 4 4 段落に記載があるように、正規分布の標準偏差は、両親となる親個体 1, 2 間の距離や、両親を結ぶ直線と第三の親個体 3 との距離に比例して設定される。その結果、パラメータ空間で互いに近傍にある両親の個体は遺伝子空間でも互いに近傍にあることが認識されて交叉されるから、要するに親個体に似た子個体が生まれるようになる。すると、探索所期にははつつ両親の間の距離が互いに遠いので比較的広い範囲で探索が行われ、探索の終盤では両親が互いに近接していることが多いので局所的な探索が行われる。

また、同文献の第 0 0 4 6 段落には、突然変異は子個体の遺伝子に正規乱数を加算する操作として、突然変異に局所的な探索を担当させることにより、探索終盤を効率よくすることができる旨の記載がある。さらに、同文献の第 0 0 5 8 段落には、探索の終盤で局所ピークに効率よく到達するために最急降下法などの局所的探索手法を組み合わせても良い旨の記載がある。

しかしながら、同文献ではその第 0 0 1 0 段落で初期収束を問題視していながら、十分に広域的な探索が行われうるのかは疑問である。なぜならば、交叉では親個体に似た子個体しか生まれず、突然変異で生成される新たな個体の範囲も正規分布で規制されるわけだから、初期世代において親個体の分布が偏っていれば、そのまま初期収束してしまうように推定されるからである。

また、前述のようにその第 0 0 5 8 段落には局所的探索手法を組み合わせる旨の記載があるが、その組合せ方は「遺伝的アルゴリズムによってパラメータを最適化した後に」ということなので、前述の特許文献 2 とおおむね同様であろうと想像される。これに関して

10

20

30

40

50

は、その「発明の効果」の項のうち第0143段落にも、「前記遺伝的アルゴリズムを用いて最適化したパラメータに対して、局所的探索手法による最適化を行うことにより」との記載がある。

なお、同文献では実施例1の変形例として「地域的な小集団」という概念を導入し、その小集団内で遺伝子操作を行う旨が、第0062段落および第0063段落に記載されている。しかし、これは一つの村の中で結婚を繰り返すようなものだから、遺伝子の多様性を求める方向とは全く逆であり、同文献で問題視していた初期収束を解決する作用があるものとは考え難い。

## (2) 非特許文献

非特許文献の紹介にあたっては、GAを用いて広域的探索を行う技術に関する文献から  
10  
先ず解説する。

### ・非特許文献1

この文献には、多峰性の著しい(つまり落とし穴の多い)大域的最適化の問題である板取り問題の解法として、GAに局所最適化を組み合わせたハイブリッドGAの一種が開示されている(1節末尾)。このハイブリッドGAは、準ニュートン法による局所最適化操作をGAの評価過程に組み込んだものである(同文献の4節冒頭および図4参照)。

そして、この組み込み方の特徴は、同文献の図9および4.8節を参照する限り、各個体に対して(あるいは交叉や突然変異で生じた新生個体の全てに対して)局所最適化操作を施すものようである。

なお、そのGAの選択過程(淘汰過程)の前処理として、  
20  
- 切り捨て(適応度による足切り)処理と、個体の多様性を維持するために適応度に対する線形スケール処理とが施されている。そして、選択過程では、適応度の高さに応じた数の個体が親個体を選択される期待値戦略と、最も高い適応度を持つ個体が生き残るエリート保存戦略とが採られている。

その結果、対象がかなり複雑な問題であるにも拘わらず、この解法でわずか30~40世代程度の間で収束をみるという成果を出している(同文献の図7参照)。しかしながら、その40世代にかかる計算時間は、28.5MIPS, 4.2Mflopsのワークステーションを用いても、約90時間すなわち4日間近くとなっている(同文献の5節末尾参照)。

### ・非特許文献2

この文献で開示されている技術内容は、一読したところ前述の非特許文献1とおおむね同一のようである。

### ・非特許文献3

この文献には、タンパク質立体構造予測問題を含む大規模な広域探索(ないし広域最適化)問題に適用するために、GAを用いた探索手法(広域最適化手法)を分散して並列処理する方法が開示されている。その方法としては、分散GA, 拡張分散GA, 分散確率モデルGA, 近傍培養型GA及び適応的重み付けGAと、これらを互いに組み合わせたハイブリッド手法とが紹介されている。ただし、ここでいうハイブリッド手法とは、上記の各種GAをどう組み合わせるかといったことであり、局所最適化操作をGAに組み込むこと  
40  
ではないようである。

### ・非特許文献4

この文献には、粉末試料のX線回折データを解析して同試料の物質構造を決定する方法について、GA適用の実際を含めて実に良くまとめられており、同文献を読めばその執筆  
50  
当時(1997年)までの背景技術がよく分かる。

この文献の導入部には、細かな粉末状の試料にX線を照射して得られた回折環データに基づいて、その資料物質の結晶構造を求めようとしても、強度ピークの重なり(ピーク・オーバーラップ)が大きいので、伝統的な方法によっては難しい旨が明記されている。そして同文献の主旨は、それがゆえに回折環のデータ解析にGAを用いることが研究されており、著者らはGAの適用にあたって何某かの工夫を加え、すでに構造が決定されている物質の粉末試料X線回折データに適用してみたということである。

しかし、同文献には局所最適化操作をGAに導入したハイブリッドGAについては特に言及はないので、結晶構造の特定にあたり決定すべきパラメータの自由度や精密さが増していくと、計算時間が爆発的に増大してしまう可能性が高い。

・非特許文献5

この文献では、粉末試料のX線回折データを解析して物質構造を決定する新手法としてラマルキアンGA（すなわちハイブリッドGA）が提案され、ごく少ない世代交代の間に構造決定がなされる演算例も提示されている。しかし、同文献で提案されている技術では交叉で生じた子個体と突然変異で生じた新個体との全てに局所最適化を施しているため、世代交代の数は少なくとも各世代での演算量が膨大なものとなる。その結果、同文献の技術を用いても、物質構造決定に要する計算時間の短縮効果は限定されたものにしかならないであろう。

10

・非特許文献6

この文献では、粉末試料のX線回折データから、前述のように物質構造（すなわち分子内の原子配置）を決定するのではなく、試料物質の結晶格子定数を求める手法にGAを用いることが提案されており、実際の回折データ解析例が開示されている。同文献には格子定数決定の困難さとその原因とが整理されて説明されているが、広域的なパラメータ探索に使用されているGAはおおむね公知のものである。

・非特許文献7

この文献は、GAの交叉におけるBLX-法を説明するための資料として参照した。すなわち、GAにおける交叉過程では、両親となる二つの個体がもつ遺伝子パラメータに対して最適化すべきパラメータ空間中で適度な範囲で拡がりを持たせる手法がある。この手法は、ブレンド・クロスオーバー `blend crossover` と呼ばれ、適正な拡がりを設定するために新たなパラメータを導入するので「BLX-」と簡素に記されることが多い。なお、パラメータの値が例えば0.5に設定されている場合には、BLX-0.5と表現される。

20

・非特許文献8～非特許文献10

結晶構造を直接的に決定する技術に関する研究の歴史は古く、このような方法に関する文献としては、代表的なところで非特許文献8～非特許文献10などを挙げる事ができる。すなわち、非特許文献8には原理的な数式の導出が開示されており、非特許文献9には数式を解く際の有力な関係式が開示されている。また、非特許文献10には、数式の結晶学的取り扱いがまとめられている。

30

(3) 直近の従来技術

何らかの形で局所最適化をGAに組み込んだハイブリッドGAに関する記載があるものは、以上の公知文献のうち特許文献2および非特許文献1, 2, 5くらいのものである。それゆえ、これら4本の文献が本発明に対して直近の従来技術に当たるものと、発明者らは認識している。

そこで、本明細書中では、GAに局所最適化を組み合わせた広域最適化に関する従来技術を、特許文献2と非特許文献1, 2, 5との二つの系統に分類して論ずることとする。すなわち、特許文献2に開示されているように、GAによるパラメータ空間の広域的探索を行った後、最終世代に属するポピュレーションのうち適応度最高の個体にだけ局所最適化を施す広域最適化技術をもって、「従来技術1」（図2(a)参照）と称することとする。一方、非特許文献1, 2, 5に開示されているように、GAで生じた新生個体の全てに局所最適化をかけて広域探索を行おうとする所謂ラマルキアンGAのことを、「従来技術2」（図2(b)参照）と称することとする。

40

(4) 公知文献リスト

【特許文献1】特開2000-86696号公報 梶昭、アンダース・リルジャス 「タンパクの結晶及び該結晶から得られる三次元構造情報に基づくラショナルドラゲデザインへの応用」

【特許文献2】特開2001-83106号公報 アレキサンダー・ウリヤネンコフ、原田仁平 「物質構造解析方法およびその装置」

50

【特許文献3】特開2002-71589号公報 西堀英治、高田昌樹 「X線回折計測装置」

【特許文献4】特開2003-149180号公報 高田昌樹、西堀英治、坂田誠、原田仁平 「・・・粉末X線回折データ測定方法」

【特許文献5】特開2004-37345号公報 原田仁平、西堀英治、涌井義一、熊澤克芳、鷺見高雄、小塚基樹、高田昌樹、坂田誠 「試料ホルダ及びX線装置」

【特許文献6】特表2005-500844号公報 ノバルティス・・・バイオメディカル・リサーチ社(スイス)出願 「キナーゼ結晶構造ならびにキナーゼ活性化のための・・・」

【特許文献7】特開2001-117901号公報 笹川文義、品川明雄(出願人:富士通株式会社)「ハイブリッド型遺伝的アルゴリズムの島モデルによる最適化問題処理方法およびその処理プログラム」

10

【特許文献8】特開2003-288579号公報 村川正宏、樋口哲也(出願人:産業技術総合研究所)「最適フィッティングパラメータ決定方法および・・・」

【非特許文献1】藤田喜久雄、赤木新介、廣川敬康 「遺伝的アルゴリズムと極小探索アルゴリズムとのハイブリッド化による板取り問題の一解法」 Approach for Optimal Nesting Using Generic Algorithm and Local Minimization Algorithm 日本機械学会論文集(C編)第59巻第564号(1993年)2576-2583頁 論文No.92-1875

20

【非特許文献2】Kikuo Fujita, Shinsuke Akagi and Noriyasu Hirokawa Hybrid Approach for Optimal Nesting Using a Generic Algorithm and a Local Minimization Algorithm, Proceedings the 1993 ASME Design Automation Conference, DE-Vol.65-1, (Sept.1993), pp.477-484, Albuquerque, New Mexico, U.S.A.

【非特許文献3】廣安知之「進化的手法による知的最適化とその並列化」 Development and Parallelization of Intelligent Optimization by Evolutionary Computation 同志社大学 学術フロンティア 「知能科学とその応用」に関するシンポジウム-AFIIIS SYMPOSIUM 2001(2002年)

30

【非特許文献4】K.HARRIS, R.JHONSTON and B.KARIUKI: Acta Cryst. (1998). A54, 632-645, The Generic Algorithm: Foundation and Applications in Structure Solution from Powder Diffraction Data

【非特許文献5】G.Turner, E.Tedesco, K.Harris, R.Jhonston and B.Kariuki: Chemical Physics Letters 321(2000)183-190, Implementation of Lamarckian concepts in a Genetic Algorithm for structure solution from powder diffraction data

40

【非特許文献6】B.Kuriuki, S.Belmonte, M.McMahon, R.Jhonston, K.Harris and R.Nelmes: Journal of Synchrotron Radiation (1999)6, 87-92, A new approach for indexing powder diffraction data based on whole-profile fitting and global optimization using a generic algorithm

50

【非特許文献7】L. J. Eshelman and J. D. Schaffer: Real-Coded Genetic Algorithm and Interval Schemata, Foundation of Generic Algorithm 2, pp187-202 (1993)

【非特許文献8】D. Harker and J. S. Kasper: Acta Cryst. 1 (1948) 70., Phases of Fourier Coefficients directly from Cristal Diffraction Data

【非特許文献9】D. Sayre: Acta Cryst. 5 (1952) 60., The Squaring Method: a New Method for Phase Determination

10

【非特許文献10】W. COCHRAN: Acta. Cryst. 8 (1955) 473., Raration between the Phases of Structures

【発明の開示】

【発明が解決しようとする課題】

【0003】

(背景技術の不都合)

前述のように、従来のX線回折による直接法では、ある程度の大きさ(例えば直径0.数mm程度)にまで単結晶を成長させることができない限り、実質的に言って物質構造を決定することができなかつた。

20

しかしながら、単結晶が得られない段階の新物質であっても、粉末なら得られるという場合は非常に多い。表現を変えて言うと、物質の創生間もない初期段階では、当該物質の粉末しか得られない場合がほとんどである。それゆえ、もしもこの時点で物質の構造を決定することができれば、目的とした物質ができていないか否かが迅速に判断できるようになる。そして、物質構造が明らかになればその機能特性の理解も進むため、構造に基づく物質設計が推し進められることとなる。

(本発明の課題)

そこで本発明は、単結晶ではなく粉末状等の試料物質に放射線を照射して得た回折環の実測データからでも、現実的な範囲の計算負荷で当該試料の物質構造を高精度で決定できるようにする手段を提供することを解決すべき課題とする。すなわち本発明の課題は、多くの新物質のように、ある程度大きな単結晶を作ることが難しい物質であっても、例えばその粉末試料から得たX線回折環の強度分布データから、その結晶構造や分子配置などの構造パラメータを決定することができるようにすることである。

30

このような課題の解決手段として、より具体的には、方法発明としての「放射線回折データに基づく物質構造の推定方法」と、そのソフトウェア関連発明としての「物質構造推定プログラム」及びその「記録媒体」と、それらに関連した各種の装置発明とを、本発明は提供するものである。

そしてこの中核技術として、GAの探索速度を加速して収束時間を短縮し、従来よりも多数のパラメータを精密かつ効率良く最適化または特定ないし推定することを、現実的なコスト範囲で可能にする手段を提供することをも、本発明は解決すべき課題としている。そして、より高速なパターンマッチングを行い、情報量が少ない実測データに基づきながら、従来よりも多数のパラメータを精密かつ効率良く最適化ないし推定することができる手段を提供することをも、本発明の解決すべき課題である。

40

より具体的には、方法発明としての「ハイブリッドGAによる複数パラメータの最適化方法」及び「パターンマッチングによるデータ解析方法」を本発明は提供する。併せて、そのソフトウェア関連発明である「最適パラメータ探索プログラム」やその「記録媒体」をも提供する。

【課題を解決するための手段】

【0004】

50

[ 発明の経緯 ]

本願発明の発明者らは、シンクロトロン放射光設備「S P r i n g - 8」(固有名称)などのX線回折測定装置を用いて、X線を始めとする各種放射線を試料物質に照射してその回折パターンを測定し、当該物質の構造解析を行ってきた(例:特許文献3, 4, 5)。そして、すでに背景技術の項や課題の項でも述べたように、粉末試料の放射線回折データ(回折環の実測データ)に基づく物質構造決定の必要性を痛感していた。

そこで、結晶の格子定数や分子の立体構造などの構造パラメータの決定に適した強力な広域最適化方法(大域的最適化方法ともいう)と、それを実現するソフトウェアとを開発し、回折環から物質構造を決定できるようにすべく努力を続けてきた。その結果、僥倖にも恵まれて、以下のような発明に一応の完成をみたので、ここにその一部を出願するに至った。

10

なお、本発明の適用は、S P r i n g - 8(固有名称)など巨大設備での測定データに限られるものではなく、企業や大学の研究室にあるような小規模なX線回折測定装置等で得られた測定データに対しても可能である。

[ 出願の単一性 ]

本願発明の中核となる部分は、技術分野の項で触れたように、G A(遺伝的アルゴリズム)に特定の方式で局所最適化操作を組み込んだ「ハイブリッドG A」それ自体である。すなわち、本願発明は、このハイブリッドG Aを実行して複数パラメータを最適化する方法とそのプログラムとを提供し、これを利用して構造パラメータを多峰性の多次元パラメータ空間中で広域的に十分に最適化し、もって回折環などの放射線回折データから物質構造を十分な確度で推定(すなわち決定)する技術を提供する。

20

したがって、本項で解決手段として列挙される本願の各発明は、以下の名称で呼ばれるが、いずれもこの中核的な単一の発明概念(すなわち「ハイブリッドG A」の利用)をもって貫かれており、同一の特別な技術的特徴を共通に有している。

(1) 方法発明

- ・「ハイブリッドG Aによる複数パラメータの最適化方法」
- ・「パターンマッチングによるデータ解析方法」
- ・「放射線回折データに基づく物質構造の推定方法」

(2) ソフトウェア関連発明

- ・「最適パラメータ探索プログラム」
- ・「物質構造推定プログラム」
- ・「記録媒体」

30

(3) 装置発明

- ・「ハイブリッドG A搭載コンピュータ」
- ・「ハイブリッドG Aを搭載した解析装置」
- ・「放射線回折パターンによる物質構造推定装置」

それゆえ、本願発明たる以下の各手段の間では、単一性の要件が満たされているものと考え、一つの出願にまとめた。なお、この項に記載された各解決手段の番号は、本願の出願時点では特許請求の範囲に記載された各請求項の番号に対応している。また本項では、各手段の構成上の特徴に併せて、その構成がもたらす作用効果についても簡潔に説明する。

40

[ 方法発明のカテゴリー ]

1. ハイブリッドG Aによる複数パラメータの最適化方法

(1) 第1手段

(第1手段の枠組み)

本発明の第1手段は、以下に示す「ハイブリッドG Aによる複数パラメータの最適化方法」である。なお、局所最適化過程が組み込まれた遺伝的アルゴリズム Genetic Algorithm(略してG A)のことを、ここでは「ハイブリッドG A」、略記すれば「H G A」(英語表記では Hybrid G A または H G A)と呼ぶことにする。

50

なお、「ハイブリッドGA」なる名称は、前述のように日本での用語を踏襲したものであり、対応する用語として英語圏では「ラマルキアンGA」( Lamarckian GA )の名称が広く使われている。ただし、日本での「ハイブリッドGA」という用語と、英語圏での「ラマルキアンGA」という用語とが、互いに厳密に等価であるとまでは断言することができない。なぜならば、特許文献2の解説で触れたように、日本でいう「ハイブリッドGA」が最後の詰めを局所最適化で加速しようとするものが多いのに対し、英語圏でいう「ラマルキアンGA」は全ての段階で新生個体の全てを局所最適化して進化を加速しようとする傾向があるからである。

すなわち本手段は、各個体を構成する記号列のうち少なくとも一部は所定の物理量に対応した複数のパラメータであり、これらのパラメータの適正な組合せを探索することを目的としたGAによって、複数パラメータを広域最適化する方法の一種である。

(第1手段の定義)

本発明の第1手段は、各個体を構成する記号列のうち少なくとも一部は、所定の物理量に対応した複数のパラメータであり、これらのパラメータの適正な組合せを探索することを目的として、これらのパラメータがそれぞれ適度に分布した複数の前記個体を初期世代のポピュレーションとし、各世代において、当該世代のポピュレーションのうち少なくとも新たに生成された前記個体である新生個体の適応度を評価する評価過程と、このポピュレーションのうち複数の前記個体を親個体としてこれら親個体の間で行う交叉により新しい子個体を生成する交叉過程と、次世代に残すべき複数の前記個体を選択して他の前記個体を淘汰する淘汰過程とを繰り返すうちに、これらの個体のうち一部に変更を加える突然変異過程を適宜交えながら、世代交代を重ねて前記適応度がより良好な前記個体の発生を待ち、もって前記複数のパラメータの最適な組合せを探索するGA(遺伝的アルゴリズム)による複数パラメータの最適化方法において、全世代のうち一部にあたる複数の特定世代において、当該世代の前記ポピュレーションのうち少なくとも一部にあたる特定個体を対象に、局所最適化操作が行われることと、全世代のうち少なくとも一部にあたる複数の特定世代において、当該世代の前記ポピュレーションのうち一部にあたる特定個体を対象に、局所最適化操作が行われることとのうち、少なくとも一方である局所最適化過程を有することを特徴とするハイブリッドGAによる複数パラメータの最適化方法である。

なお、第1手段を定義する記載がこのままでは一読して理解することは難しいので、読者に理解が容易になるよう、本手段の構成を前提部分と特徴部分とに分けて以下で説明する。

(第1手段の前提)

本手段の前提構成は、簡単に言ってしまうとGAを有することである。

すなわち本手段は、たとえば、これらのパラメータがそれぞれ適度に分布した複数の前記個体を初期世代のポピュレーションとし、各世代において、当該世代のポピュレーションのうち少なくとも新たに生成された前記個体である新生個体の適応度を評価する評価過程と、このポピュレーションのうち複数の前記個体を親個体としてこれら親個体の間で行う交叉により新しい子個体を生成する交叉過程と、次世代に残すべき複数の前記個体を選択して他の前記個体を淘汰する淘汰過程とを繰り返すうちに、これらの個体のうち一部に変更を加える突然変異過程を適宜交えながら、世代交代を重ねて前記適応度がより良好な前記個体の発生を待ち、もって前記複数のパラメータの最適な組合せを探索するGAによる複数パラメータの最適化方法であることを、構成上の前提とする。

言い換えると、本手段が構成上の前提としてもつGAは、これらのパラメータがそれぞれ適度に分布した複数の前記個体を初期世代のポピュレーションとし、世代交代を重ねて前記適応度がより良好な前記個体の発生を待つアルゴリズムである。このGAは、各世代において評価過程、交叉過程および淘汰過程とを繰り返すうちに、突然変異過程を適宜交えながら世代交代を進めていく。ここで、評価過程は、当該世代のポピュレーションのうち少なくとも新たに生成された前記個体である新生個体の適応度を評価する過程である。また、交叉過程は、このポピュレーションのうち複数の前記個体を親個体として、これら親個体の間で行う交叉により新しい子個体を生成する過程である。さらに、淘汰過程は、

次世代に残すべき複数の前記個体を選択して他の前記個体を淘汰する過程である。そして突然変異過程は、これらの個体のうち一部に変更を加える過程である。本手段は、このように世代交代を重ねるGAにより、前記複数のパラメータの最適な組合せを探索することを目的とした複数パラメータの最適化方法である。

ここで、「世代交代を重ねて前記適応度がより良好な前記個体の発生を待ち」とあるうちの「前記個体」は、ふつう局所最適化操作または交叉ないし突然変異の結果、新たに生じた個体のうち適応度が最良のものといった意味合いである。ここに「前記」とあるは、既に記載された要件を満たしているといった程度の意味であり、けっして以前の世代から生き残っている個体であるという限定を意味するわけではない。

なお、GAの構成については種々のヴァリエーションがあるので、少なくとも交叉過程と突然変異過程とを有し、世代交代を重ねて適応度の高い個体の発生を待つGAでさえあれば、前提構成が異なっても構わない。

すなわち本手段は、上記の前提構成のうち各処理過程を列挙している部分が、たとえば「各世代において、当該世代のポピュレーションのうち少なくとも一部の前記個体について当該個体の適応度を評価する評価過程と、このポピュレーションから複数の前記個体を選択して親個体とする淘汰過程と、これらの親個体の中で交叉を行って新しい子個体を生成する交叉過程とを繰り返すうちに、…」などとなっても構わない。

#### (第1手段の特徴)

本手段の特徴は、前述のGAにおいて、全世代のうち一部にあたる複数の特定世代において、当該世代の前記ポピュレーションのうち少なくとも一部にあたる特定個体に対して局所最適化操作が行われることと、全世代のうち少なくとも一部にあたる複数の特定世代において、当該世代の前記ポピュレーションのうち一部にあたる特定個体に対して局所最適化操作が行われることとのうち、少なくとも一方である局所最適化過程を有するハイブリッドGAによる複数パラメータの最適化方法であることである。

これは要するに、局所最適化操作の対象は、全ての世代の全ての個体ではなく、全世代のうち一部の複数世代であったり、逆に全世代であっても一部の個体であったり、あるいは一部の複数世代のポピュレーションのうち一部個体であったりということである。換言すれば、局所最適化操作の対象は、全世代のうち一部の複数世代での一部個体であるか、全世代での一部個体であるか、一部の複数世代での全個体であるかのうちいずれかであるということである。

なお、特定世代の分布には色々な場合がありうる。すなわち、特定世代が飛び飛びに散在する場合にも、必ずしも所定世代毎に周期的に特定世代が出現しなければいけないわけではなく、特定世代の出現頻度ないし出現周期が世代交代の間に変動しても良い。また、特定世代の出現が初期世代にもあるようにしても良いし、逆に初期世代は特定世代にはなり得ないようにしておいても良い。

また、各特定世代で選定される特定個体の数も、全ての特定世代の間で一定していても良いし、あるいは世代交代の間が変わっていても良い。ここで、特定個体がポピュレーションのうち一部である場合には、特定個体の数を絶対的な個数で決めても良いし、ポピュレーションのうち所定パーセンテージ(例えば1%とか0.1%とか)が特定個体に選定されるようにしても良い。両者はポピュレーションが一定であれば等価であるが、世代交代によってポピュレーションが変動する場合には両者の間に違いが出る。あるいは、特定個体の数は、交叉や突然変異によって生じた新生個体の数に対して所定パーセンテージを占めるように設定されていても良い。もちろん、このようなパーセンテージも一定である必要はなく、世代が代われれば変動するようにしても良い。

例えば、GAにより1000個の個体からなるポピュレーションが世代交代を繰り返す場合を想定してみよう。

まず、一方の極端な実施形態では、世代交代毎に各世代のポピュレーションのうち一つが特定個体に選ばれて局所最適化操作を施される。そして、これが最も好ましい実施形態であることも少なくない。

なぜならば、仮に各世代のポピュレーション(個体数1000個)のうち半分(500

10

20

30

40

50

個)程度が新生個体であるとすれば、各世代での局所最適化操作の回数が、ラマルキアンGAに比べてこの実施形態では約500分の1に減っているからである。これは劇的な演算量の低減を意味しているとともに、ラマルキアンGAとは違って当該世代のポピュレーションのうちほとんどの個体は局所最適化操作を受けないから、個体の多様性はほとんど失われずに済む。それでありながら、初期世代を入れるか入れないかにかかわらず、世代交代毎に全ての世代で進化に幾分か方向性が与えられて、継続的に進化が加速され続けるから、広域的な最適値にごく小さな演算量で到達する可能性が高まる。

次に、中間的な実施形態では、例えば二三世代毎ないし十世代毎に特定世代があり、当該世代のポピュレーションのうち一つないし数個程度の特定個体が最適化操作される。

そして、他方の極端な実施形態では、例えば百世代毎に一度、当該世代のポピュレーションを形成する個体の全てに対して局所最適化操作が施される。

ところで、複数の個体に局所最適化を施した結果、同一世代のポピュレーションに属する個体のうち幾つかが、パラメータ空間中で同じ局所最適値に相当する位置にまとまってしまうことがありうる。そこで、実質的に同一点または等価と言えるほど近傍に集合した複数の個体は、一つの個体と見なしてしまい、こうして個体数が減った分については、交叉や突然変異で新生個体を補充するようにしても良い。

また、世代交代を重ねるうちには、以前の世代ですでに局所最適化されている個体が再び特定個体に選ばれ、局所最適化操作をかけられる可能性が生じる。そこで、この可能性を封じるために局所最適化済みかそうでないかを示す印(フラッグなど)を個体に付けておくようにしても良い。ただし、このようなことをしなくても、局所最適化操作が施される特定個体の数がポピュレーションに対してごく僅かなものであれば、同一の個体が重複して特定個体として選ばれることはごく稀であるから、実質上問題になることはない。そして、仮に同一個体に対して重ねて局所最適化操作が行われた場合にも、二回目以降の同操作では即座に完了してしまうので、ほとんど演算処理の負担が生じることはない。

さて話を本筋に戻すと、このように、本手段においては、特定世代の分布と当該世代のポピュレーションから選出される特定個体の個数とに、かなり広いヴァリエーションがある。

ただし、前述の公知文献のうち特許文献2に開示されているように、個体に局所最適化が施されることがある特定世代は、最終世代にあたる一世代だけである場合は、本手段に含まれていない。また、非特許文献1, 2, 5に開示されている Lamarckian GAのように、全ての世代で全ての新生個体に局所最適化を施す(すなわち全世代が特定世代でその全個体が特定個体である)場合も、本手段には含まれない。

そして、本手段では、前提となるGA自体の構成にも、そのGAに組み合わせられる局所最適化操作の手法自体にも、特段の限定はない。すなわち、局所最適化操作をGAの中にもどのように組合せてハイブリッド化するかという点に、本手段の特徴がある。

それゆえ、本手段が前提として有するGAの構成は、単に「各世代で評価過程、淘汰過程(選択過程)および交叉過程を繰り返しつつ、突然変異過程が適宜交えられて世代交代を重ねるうちに、適応度が良好な個体の生成を待つGA」としても良い。あるいは、評価過程を選択過程に含めてしまい、もっと簡単に「選択過程(淘汰過程)、交叉過程、突然変異過程をもつGA」としても良い。

ここで、一部の用語に定義を加えておく。まず、放射線としては、X線が代表的であるが、中性子線や電子線などの粒子線も含まれる。X線の中には、放射光(シンクロトロン軌道放射)のような高エネルギーX線や、X線領域でのコヒーレント光なども、もちろん含まれるものとする。次に、回折とは、透過回折が主流ではあるが、反射現象によるものも含まれるものとする。そして、各個体の適応度が探索範囲の全域で連続量であることは、適応度に関する局所最適化操作を個体に施すための十分条件である。

(第1手段の作用効果)

本手段では、従来技術1(図2(a)参照)のように、方向性が与えられず確率的に進化する単純なGAによって広域的な探索を行い、そのGAによって最適解近傍への収束が確信されるに至るのを待つというわけではない。あるいは、従来技術2(図2(b)参照)

10

20

30

40

50

のように、各世代の全ての個体（または新生個体の全て）に対して局所最適化操作が行われることはない。

そうではなく、本手段の好ましい実施形態を挙げれば、本手段での局所最適化操作の対象となる「特定個体」は、全世代のうち一部にあたる分散した複数世代（「特定世代」）に散在しており、当該世代のポピュレーションに属するごく一部の個体である。ただし極端な例としては、全世代において各世代のポピュレーションのうちごく一部の個体であるとか、逆に、全世代のうち分散したごく一部の複数世代の全個体であるかのうちいずれかであるとかいう場合も、本手段の実施形態に含まれる。

それゆえ本手段では、特定世代の出現頻度の設定如何とその世代での特定個体の選定方法の如何とによって、局所最適化操作の回数を従来技術 2 よりも劇的に減らすことが可能になる。

10

ここで留意すべきことは、局所最適化を行うには、特定個体の近傍において適応度の勾配を算出する微分処理か、あるいはそれに類する演算処理が繰り返し必要であり、なかなか大きな演算処理が必要となることである。そして、最適化すべきパラメータの個数ないし自由度が増え、探索すべきパラメータ空間の次元数が増えるにつれて、このような微分操作に要する演算負荷も大きくなるから、局所最適化操作の回数を減らすことができれば、それがもたらす演算負荷の低減効果は大きくなる。

したがって、本手段の実施によって局所最適化操作の回数を従来技術 2 よりも劇的に減らすことができれば、最適化すべきパラメータの数が多い場合には演算負荷を大幅に減らすことができる。それゆえ本手段によれば、最適化すべきパラメータの数が多ければ多いほど、従来技術 2 よりも計算時間を大幅に短縮することができるという効果が見込める。その結果、計算時間が長くなるのがネックになって今まで実用にならなかったようなパラメータの数が多い問題であっても、本手段によってその問題を解くことができるようになる可能性がある。

20

そればかりではなく、本手段では、従来技術 2 のように各世代のポピュレーションがもつ多様性が急速に失われるような不都合は生じることがない。それゆえ、本手段は漏れの少ない広域的な探索能力を強力に発揮することができ、結果的に短い演算時間のうちに広域的な最適値に達することができる。

一方、本手段では、従来技術 1 とは違って、GA の進化に方向性をなんら与えることなく、GA の最良解が正解の近傍にまで収束したことが確信できるまで待っていることは、必要ではない。より正確に言うと、本手段では、世代交代の間に適宜その進化に方向性が与えられるので世代交代の回数が少なく済むうえに、広域的な最適値に至る最後の詰めも局所最適化でなされることがほとんどである。

30

すなわち本手段では、世代交代の途中であっても、幾つもの世代で何れかの個体に局所最適化操作を施すことにより進化に方向性を与えることができるから、広域的な最適値に収束するまでの世代交代の回数自体をも劇的に減らすことができる。それでいながら、交叉や突然変異で生じる全ての新生個体の数に比べて局所最適化操作の回数がずっと少ないようにしておけば、前提構成として持つ GA に比べて、世代交代毎の平均的な演算負荷はそれほど増えないで済む。

したがって本手段によれば、世代交代の間に進化の方向性が適度に与えられるようになるにもかかわらず、単純な GA で広域探索を行う従来技術 1 に比べても世代交代毎の平均的な演算負荷があまり増えないで済む。すなわち本手段では、世代交代の間に進化の方向性が適度に与えられながら、単純な GA に比べてもあまり演算負荷が増えることがないので、広域的な最適値に達するまでの世代交代の回数が、従来技術 1 に比べて劇的に低減される。特に、広域的な最適値（正解）が形成している峰の一部（たとえ裾野であっても）に個体が達し、その個体に局所最適化がかけられたならば、一気にその正解に達しうる。それゆえに本手段では、従来技術 2 ばかりではなく従来技術 1 に比べても、広域的な最適値に達するまでの演算時間を短縮することができる。

40

すなわち本手段によれば、従来技術 1 および従来技術 2 のいずれに比べても、広域的な探索能力が劇的に向上しており、多数のパラメータの広域最適化にかかる計算時間を大幅

50

に短縮することができるという効果がある。その結果、広域的に探索すべき多峰性空間があまりに複雑で、従来技術1でも従来技術2でも現実的には今まで解けなかった複数パラメータの最適化問題であっても、本手段によってその問題を解くことができるようになる可能性がある。

(第1手段に関する考察)

・ポイント

さて、本手段のように、局所最適化を適正に組み合わせることによってGAの収束特性が改善され、多数のパラメータを広域的に最適化するのに要する時間が大幅に短縮する理由は、局所最適化とGAとが互いにその欠点を補完し合うためであると考えられる。

すなわち、GAによる複数パラメータの広域探索というのは、確率的(stochastic)な探索方法である。それゆえGAは、多峰性の(すなわち落とし穴の多い)多次元パラメータ空間を、たとえ粗くであっても素早く(つまり小さな演算負荷で)広域的に探索し、そこそこ良い適応度をもつパラメータの組合せを見つけ出すことを得意としている。逆に、広域的な最適点(つまり求める「正解」)の近傍まで肉薄していても、そのことを認識することはできないので、ピンポイントでその最適点に達することは不得手である。つまり、局所最適化すれば最適点にたどり着けるような最適点の近傍に個体が生じても、その個体は適応度がそれほど良くはないという理由などで淘汰されてしまい、次世代に生き残れないと行ったことが、世代交代の過程で少なからず起きているものと考えられる。

一方、局所最適化操作による探索は、確定的(deterministic)な探索方法である。それゆえ局所最適化は、多次元パラメータ空間の中であっても、多峰性のない領域に限ってしまえば、速やかかつ容易に最適点に到達することが得意であるが、逆に、強い多峰性を持つ多次元空間中を広域的に探索することは不得手である。

そこで、GAと局所最適化とを上手に組み合わせれば、互いの長所が互いの短所を補って補完し合い、多峰性の多次元空間中を広域的に探索して回り、ピンポイントで最適点に達するという困難な課題であっても、達成することができるようになる可能性が生じる。本手段は、このような考えに基づいて開発された技術であるから、本手段でこのような効果がどの程度大きく得られるかは、主として局所最適化とGAとを如何に巧妙に組み合わせるかという点に大きく依存する。

ここで、局所最適化とGAとの組み合わせる上での重要なポイントには、次の二つがある。第一のポイントは、世代交代の間における特定世代の出現頻度と、当該世代のポピュレーションに対する特定個体の個数ないしパーセンテージとである。第二のポイントは、特定世代のポピュレーションの中から所定数の特定個体を選び出す際に、どのような選定基準を採用するかである。なお、局所最適化とGAとの組合せ効果を十全に発揮するためには、淘汰過程で生き残り個体を選択するルールや、交叉過程で交叉させる親個体の選定や交叉の仕方などのルールなど、GA自体の設計にも留意した方がよい。

したがって、本手段に従属している解決手段には、GAと局所最適化との組合せ方に関する限定か、あるいは組み合わせる際にGA自体をどう設計するかといったことに関する限定を含むものが多い。このことには、以下の記載を読み進める上で留意されたい。

・両従来技術の組合せに対する阻害要因

前述のように、局所最適化操作をGAに組み合わせたハイブリッドGAには、従来から大きく分けて二通りがある。この二通りとは、前述のように、ずっとGAで世代交代を繰り返して広域を探索したうえで最後の詰めだけを局所最適化操作で行う従来技術1と、GAの世代交代毎に全ての新生個体を局所最適化する従来技術2とである。

すなわち従来技術1では、単純なGAによってダーウィンのような進化を行う世代交代が繰り返され、確率的に広域的な探索が行われた後に、最終的に残った世代のポピュレーションのうち適応度が最高の個体にだけ局所最適化が施される。この技術では、淘汰過程での選択戦略や交叉過程での親個体の選択はあるものの、最終世代に至るまでは通常のGAと変わるところは基本的にない。それゆえ、最終世代に至るまでGAに進化の方向性が与えられることは全くないから、GAによる進化が加速されることはない。そして、たまたま

10

20

30

40

50

最終世代の最優秀個体が広域的な最適解の近傍にあればよいが、そうでなければ広域的な最適解には到達できないままに探索が終了する。

いっぽう従来技術2では、GAの世代交代毎に全ての新生個体が局所最適化されるので、進化の方向付けが全ての世代で強烈に行われ、ごく初期の数世代程度では確かに急速に適応度が向上し、特に各世代のポピュレーションの平均適応度の向上は著しい。しかしながら、その間にポピュレーションを形成する個体から多様性が失われてしまうらしく、その後は適応度があまり向上しないところをみると、むしろ狙いとは逆に、多峰性空間における広域的な探索能力が損なわれてしまっている可能性さえある。しかも、計算量が大変に多い局所最適化操作を全ての新生個体に施すので、一回の世代交代にかかる演算量は膨大となり、多数のパラメータを広域的に最適化するのにかかる計算時間を短縮する効果は小さいものと考えられる。

10

端的に言ってしまうと、従来技術1は、世代交代の途中で進化に方向性を与えず、徹底してダーウィンの適者生存の進化を繰り返した後、最後の詰めで最優秀個体にだけ方向性をあたえる広域探索方法である。逆に従来技術2は、全ての新生個体を局所最適化してその進化に方向性を強要するあまり、そのポピュレーションから多様性を奪ってしまい、多峰性の強い多次元パラメータ空間では簡単に落とし穴に陥って捉えられてしまう傾向がある。

さて、それでは本発明の第1手段のように局所最適化をGAの一部にだけに組み込むというコンセプトを発想することが、なぜ専門家にとっても難しかったか、何がその阻害要因になっていたかを考察してみよう。

20

それは、一言で言うと、世代交代を重ねてもポピュレーションから多様性が失われないようにすることと、進化に方向性を付与することとは、両立することが今まではできなかったのではないと思われる。そもそも、多様性の確保と進化の方向付けとの両者をバランスをとって両立させることの必要性は、認識されていなかったようであり、その必要性が認識されていたという証拠を発明者らは寡聞にして知らない。

すなわち、従来技術1では、GAで行われる進化に方向性を与えようという試みが、最終世代に至るまで全く行われていない。一方、従来技術2では、広域的な探索を効率よく行うために必要な個体の多様性が確保されておらず、最初の数世代を経るだけで速やかにポピュレーションから多様性が失われてしまっている。本願発明者らの見るところ、本発明以前には、このような両従来技術の欠点がしっかりと認識されていたとは考え難い。ましてや、世代を重ねてもポピュレーションに属する個体の多様性が失われないようにしながら、進化の過程で適度な方向性が与えられるようにしたGAを開発することが、当該技術分野で課題として認識されていたとは、いっそう考え難い。

30

そして、ダーウィンの適者生存の進化論は、不作為な確率重視であるから、各個体で行った努力が子孫にまで継承されるとするラマルクの要不要論とは、思想的に相容れない傾向が強い。この傾向は、従来技術1と従来技術2とを組み合わせることに対する阻害要因であると考えべきである。したがって、基本的な発想の過程から考えても、従来技術1と従来技術2とを組み合わせれば容易に本願発明に達することができるというのは、妥当ではない。

それ故に、遺伝的アルゴリズムやその改良としてのハイブリッドGAやラマルキアンGAの分野の専門家が本願発明のような技術を完成させるには、小さからぬ阻害要因があったものとするのが自然である。

40

#### ・まとめ

なお、第1手段を発明するに至るためには、世代交代を繰り返すうちに散在する飛び飛びの世代で、ごく少数の個体に局所最適化操作を施すだけでも、進化に方向性を与えることができるということに気づくことが必要である。そして、本手段での局所最適化とGAとの組合せ方の良いところは、このように進化に方向性を与えることができながら、世代が下っていても広域的な探索能力に必要な多様性がポピュレーションから失われないで済むことである。

その結果、本手段によれば、多峰性が強い多次元パラメータ空間であっても落とし穴に

50

囚われることなく、両従来技術のいずれにも勝って強力な広域的探索能力を発揮することができるという効果がある。しかも、世代交代の間に広域的な最適値の近傍に達した個体があり、たとえ偶然であってもその個体に局所最適化が施されたならば、それで複数パラメータの広域的最適化という目的は即時に達成される。そして、このような偶然は少なくともほとんどの場合に於いて両従来技術のいずれよりも少ない計算時間で起こるから、結果的に本手段によればずっと短い計算時間のうちに広域的な最適値（ないし「正解」）に達しうる。

#### (2) 局所最適化操作の頻度に関する限定

以下では先ず、局所最適化操作の頻度に関する限定を、前述の第1手段に対して加えた解決手段から説明する。

ここで、先ずこのような限定を加えるのは何故かということ、局所最適化操作を行うには多大な演算量を要するので、局所最適化操作を行う頻度を減らしたいからである。そしてその一方で、GAの進化に方向性を持たせる局所最適化の作用が失われないようにすることが、多峰性パラメータ空間における広域的な最適値の探索過程全体の演算負荷を減らす上で大切であるからである。

#### (第2手段)

本発明の第2手段は、前述の第1手段の「ハイブリッドGAによる複数パラメータの最適化方法」において、前記局所最適化過程は、前記全世代のうち一部にあたる前記複数の特定世代において、当該世代の前記ポピュレーションのうち一部にあたる前記特定個体を対象に、前記局所最適化操作を行う過程であることを特徴とする。すなわち、本手段では、局所最適化操作の対象は、全世代のうち一部の複数世代での一部個体である。

ここで、特定世代の出現頻度と特定個体の個数との組合せを具体的に例示しておこう。例えば、特定世代の出現頻度は二～三世代毎ないし数世代毎に一回で、その世代における特定個体の数は一つだけという場合があり得る。また、特定世代の出現頻度は十世代毎ないし数十世代毎に一回で、その世代における特定個体の数は一つないし十個程度とかでもよい。さらに、特定世代は、世代交代の度に連続した世代で出現するが、初期の幾つかの世代では出現しなかったり、または初期の幾つかの世代でだけ飛び飛びであったり、あるいは飛び飛びに出現しない世代があったりしてもよい。

本手段では、前述の第1手段の範囲では局所最適化操作の頻度が比較的低い方に属し、設定の仕方如何では、進化の方向性が適度に与えられながら、局所最適化操作の頻度を大きく減らすことができる。すなわち、GAの作用により広域的な探索を確率的に行いながら、たまたま広域的な最適値の周辺にある個体が局所最適化されることがあれば、一気に求める最適値に達することが可能である。

したがって本手段によれば、前述の第1手段の中でも広域的な最適値に達するまでの演算負荷が低減される効果が大きくなり、特に最適化すべきパラメータの数が多い場合には計算時間の短縮効果が著しい。

#### (第3手段)

本発明の第3手段は、前述の第1手段の「ハイブリッドGAによる複数パラメータの最適化方法」において、前記局所最適化操作の対象たる前記特定個体は、当該世代の前記ポピュレーションのうち一つの前記個体であることを特徴とする。

本手段では、局所最適化操作が行われる特定世代（全ての世代でも良い）に属するポピュレーションのうち、たった一つの個体に対してのみ局所最適化操作が行われるので、前述の第1手段の範囲では局所最適化操作の頻度が特に低い方に属する。それゆえ、本手段によれば、計算時間の短縮効果が大きくなるという効果があり、この効果は特に最適化すべきパラメータの数が多い場合に顕著になる。

#### (第4手段)

本発明の第4手段は、前述の第1手段の「ハイブリッドGAによる複数パラメータの最適化方法」において、前記局所最適化過程が行われる前記特定世代は、前記全世代のうち少なくとも一部に散在していることを特徴とする。

本手段では、局所最適化操作が行われる世代が全世代に全般的に散在している必要はな

10

20

30

40

50

く、全世代のうち一部区間では全く局所最適化操作が行われなかったことがあっても構わない。例えば、世代交代の浅いうちには局所最適化が全く行われなかった世代があったり、局所最適化が行われなかった世代がしばらく続いたりしても構わない。

したがって本手段によれば、全世代における特定世代の分布を設定する自由度が増しているため、その設定次第ではより効率的に広域的な最適化を進めることができる可能性があるという効果がある。

実は、実際に回折環に基づいて物質構造決定を行ったケースのうち幾つかで、世代交代のうち始めの幾つかの世代までは局所最適化を行わなかった方が、ハイブリッドGAの収束が速く、計算時間が短縮されたということがあった。そこで、その特徴のうち一部を本手段で定義しておいた。

なお、次の第5手段および第6手段も、その特徴を部分的に表現を変えて記載したものである。

#### (第5手段)

本発明の第5手段は、前述の第1手段の「ハイブリッドGAによる複数パラメータの最適化方法」において、前記特定世代が各前記世代のうちに出現する頻度は、前記全世代のうち前記世代が下の方が高く設定されていることを特徴とする。

本手段では、その世代のポピュレーションのうちどれかの特定個体に局所最適化操作が行われる特定世代は、その出現頻度を世代交代の途中で変え、しかも増えるという設定がなされている。ただし、特定世代の出現頻度の増え方に限定はないから、世代交代の途中から一度だけステップ状に出現頻度が増えても良いし、世代交代の一部区間または全区間で徐々に出現頻度が増えても良い。

本手段によっても、ハイブリッドGAの収束がもっと速くなり、計算時間がさらに短縮されるという効果が発揮される場合がある。

#### (第6手段)

本発明の第6手段は、前述の第1手段の「ハイブリッドGAによる複数パラメータの最適化方法」において、前記特定世代は、前記全世代のうち少なくとも後半部で、所定の複数世代毎に周期的に出現することを特徴とする。

実際に回折環に基づいて物質構造決定を行った際に、初期世代から所定世代までは局所最適化を行わないでおき、所定世代以降ではどれかの個体に局所最適化を施す特定世代を周期的に出現させるようにしたところ、ハイブリッドGAの収束が速くなり、計算時間が短縮されたということがあった。そこで、本手段でその特徴を定式化しておくことにした。

本手段によっても、ハイブリッドGAの収束がもっと速くなり、計算時間がさらに短縮されるという効果が発揮される場合がある。

### (3) 局所最適化対象の指定方法に関する限定

#### (当該限定の意義)

- ・「正解」の必要性

次に、局所最適化の行われる特定世代のポピュレーションから、その対象となる特定個体を選び出す方法について限定を加えた解決手段を開示する。

ここでのポイントは、意外なことに、適応度の高い個体に局所最適化操作を施した場合よりも、むしろ逆に適応度の低い個体に局所最適化操作を施した方が、その局所最適化の結果、正解にたどり着く確率が高くなるということである。発明者らがこのことを発見したのは、プログラムミスが切っ掛けになって、不思議に収束が速いケースがあったという僥倖(セレンディピティー)にも恵まれたせいもある。

さて、先ほど「正解」という言葉を使ったが、これは発明者らが放射線回折による物質構造の決定という技術分野における研究者であるからである。

すなわち、この技術分野においては、最適設計などの分野とは違って、単にデータマッチングの適応度が比較的高いというだけでは不足であり、広域的に最適な(つまり適応度が最高の)正解であることが必要である。なぜならば、物質構造を決定するうえで、実測データとよく似た回折パターンを形成するのでデータマッチングの適応度は高いのだけ

10

20

30

40

50

ど、実は正解とは異なる構造を持つ物質でしたというのでは、ほとんど全く意味がないからである。

・思考実験

そして、適応度の低い個体に局所最適化操作を施した方が正解にたどり着く確率が高くなるということは、例えば世界地図の中から世界最高峰を探索する問題を想定してみると、理解しやすい。これは一種の思考実験である。

すなわち、世界地図は緯度経度という二次元のパラメータ空間であり、その陸地の部分は所定の拘束条件を満たす範囲であって、各個体が位置する地点の標高は当該個体の適応度に相当する。そして、陸地のうちなるべく広範な範囲に多数の個体を散在させ、本発明のハイブリッドGAによって、世界最高峰の山頂を極めることを目的とした探索を行うことを想像してみよう。

10

すると、適応度が高い個体は、すでに標高が高い山岳部なり山麓なり高地なりにいるわけだが、それがアルプス山脈やウラル山脈、ロッキー山脈あるいはアンデス山脈に連なる部分にいては、局所最適化をかけて上り坂を登らせ山頂に至ったところで、世界最高峰には達し得ない。そうではなく、局所最適化をかければ世界最高峰のエベレスト（チョモランマまたはサガルマータともいう）山頂に至る可能性があるのは、むしろガンジス川北岸の低地に佇む個体である。

このように考えれば、比較的適応度が低い個体も局所最適化操作の対象に選ぶことができる方が、より速く正解にたどり着くことになるということは、思考実験によって妥当な説明が付くことではある。

20

しかしながら、局所最適化操作する個体を選ぶうえでこのような方法を採用することは、適応度の高い個体を優遇するエリート戦略やルーレット戦略などを重宝してきたGAの世界では、全く常識の逆を行く発想であった。なぜならば、このような概念は、適者生存として自然淘汰が行われるダーウィンの進化論を模して誕生したGAとは相容れない性格を持つものだからである。

以下の各手段は、一連の本願発明のうち、このような発想に基づいて設けられた限定を色々に規定したものである。

（第7手段）

本発明の第7手段は、前述の第1手段の「ハイブリッドGAによる複数パラメータの最適化方法」において、前記特定個体は、当該特定世代の前記ポピュレーションのうち前記適応度が比較的低い前記個体の中からも指定される可能性があることを特徴とする。

30

本手段では、前述のような作用が得られる。その結果、本手段によれば、ふつうに適応度が高い個体の中ばかりから局所最適化操作の対象を選ぶようにした場合よりも、正解への収束が速くなり、計算時間がさらに短縮されるという効果が発揮される。

（第8手段）

本発明の第8手段は、前述の第7手段の「ハイブリッドGAによる複数パラメータの最適化方法」において、前記特定世代の前記ポピュレーションのうち、前記適応度が比較的低い複数個の前記個体からなるグループをもって、劣等グループと呼ぶことにし、前記特定個体は、この劣等グループの中から指定されることを特徴とする。

本手段では、前述の第7手段よりも限定が狭くなされているので、前述の作用がもっと強くなる。その結果、本手段によれば、ハイブリッドGAがもっと速く正解へ収束する可能性が高まり、計算時間がさらに短縮されるという効果がある。

40

（第9手段）

本発明の第9手段は、前述の第8手段の「ハイブリッドGAによる複数パラメータの最適化方法」において、前記劣等グループは、前記ポピュレーションのうち前記適応度が低い方の所定数および所定割合のうち少なくとも一方に属する複数の前記個体からなるグループであることを特徴とする。

本手段では、前述の第8手段よりもさらに狭く限定されているので、前述の作用がもっと強くなる。その結果、本手段によれば、ハイブリッドGAによる探索がもっと速やかになり、正解に至るまでの計算時間がよりいっそう短縮されるという効果がある。

50

## (第10手段)

本発明の第10手段は、前述の第8手段の「ハイブリッドGAによる複数パラメータの最適化方法」において、前記特定個体は、当該世代の前記劣等グループの中からルーレット選択法およびランダム選択法のうち一方により選択されることを特徴とする。

本手段でも、前述の第8手段よりもさらに狭く限定されているので、前述の作用がよりいっそう強くなる。その結果、本手段によっても、ハイブリッドGAがいっそう速く正解へ収束し、計算時間がよりいっそう短縮されるという効果がある。

## (第11手段)

本発明の第11手段は、前述の第1手段の「ハイブリッドGAによる複数パラメータの最適化方法」において、前記局所最適化過程は、前記特定世代の前記ポピュレーションの一部を前記特定個体として選び出して前記局所最適化操作を施す過程であり、前記特定個体の選定に当たっては、前記適応度がより低い前記個体の方がより高い確率で選定されるという弱者優先の傾向を持つことを特徴とする。

ただし本手段では、「弱者優先」の傾向を持つとはいえ、前述の第8手段および第9手段とは異なり、劣等グループには属さないエリート集団の中からも特定個体が選ばれる可能性が残っている。

本手段でも、前述のように適応度が低い個体の中にこそ局所最適化操作の結果、正解に至るものが多いという考えから、適応度がより低い個体の方がより高い確率で選定されるようになっている。その結果、本手段によっても、ハイブリッドGAがいっそう速く正解へ収束し、計算時間がいっそう短縮されるという効果が得られる。

## (4) 淘汰過程の選択戦略に関する限定

ここでは、次世代に生き残る個体を選択する選択戦略において、適者生存という自然淘汰の思想(ダーウィニズム)とは逆行し、言うなれば「不適者生存」ともいべき逆転の発想を採り入れた解決手段を幾つか開示する。

したがって、以下の幾つかの解決手段が持つ特徴は、淘汰過程で生き残る個体の選択に関するものであり、局所最適化操作の対象となる特定個体の選択に関するものではないことに留意されたい。

## (第12手段)

本発明の第12手段は、前述の第1手段の「ハイブリッドGAによる複数パラメータの最適化方法」において、前記淘汰過程では、当該世代のポピュレーションのうち前記適応度が比較的低い前記個体であっても次世代に生き残る可能性がある選択戦略が採られることを特徴とする。

前述の第1手段では、交叉なり突然変異なりで新たに生成された個体の全てに局所最適化操作を施すわけではないので、局所最適化が施されれば正解に至る個体が、局所最適化を施されず適応度が低いままに放置されている可能性がある。そこで本手段では、このような個体が世代交代の折に全滅してしまうことを、なるべく回避することができるようにした。

すなわち、本手段の淘汰過程では、適応度の高い個体だけが生き残るエリート戦略を採らず、適応度が比較的低い個体であっても次世代に生き残る可能性がある選択戦略を取っている。その結果、局所最適化が施されれば正解に至る個体は、その世代では局所最適化が施されずに放置されていても、次世代に生き残る可能性があるから、世代を下ったところで局所最適化されて正解に至る可能性を持つ。

したがって本手段によれば、このような可能性がある分だけ、ハイブリッドGAが速く正解へ収束する可能性が高まり、計算時間が短縮される可能性が高まるという効果が得られる。

## (第13手段)

本発明の第13手段は、前述の第1手段の「ハイブリッドGAによる複数パラメータの最適化方法」において、前記淘汰過程で前記選択の際に採られるルールのことを選択戦略と呼ぶことにし、この選択戦略は、この適応度とは関わりなく無作為に前記個体を選択されるランダム戦略と、この適応度が比較的低い前記個体が優先的に選択される逆エリート

戦略と、この適応度が比較的低い前記個体の方が選択される確率が高くなる逆ルーレット戦略と、この適応度が比較的低い前記個体の方が優先的に選択されるようにした逆期待値方式、逆ランク方式および逆トーナメント方式とのうち何れかとのうちいずれかであることを特徴とする。

前述の第12手段では、適応度が比較的低い個体であっても次世代に生き残る可能性が否定されなかつただけであるが、本手段では、適応度とは関わりなく無作為に生き残り個体を選択されるか、むしろ積極的に適応度が低い個体の生存が優先される。すなわち、本手段は、不適者の生存を優先する選択戦略を提案するものである。

その結果、仮に局所最適化が施されたとすれば正解に至るであろう個体が、局所最適化さらないうままで放置されていても、この個体は、次世代に生き残る可能性が高まるので、世代を下ったところで局所最適化されて正解に至る可能性を持つ。したがって本手段によれば、このような可能性が高まる分だけ、ハイブリッドGAが速く正解へ収束する可能性が高まり、計算時間が短縮される可能性が高まるという効果が得られる。

(第14手段)

本発明の第14手段は、前述の第12手段または第13手段の「ハイブリッドGAによる複数パラメータの最適化方法」において、当該世代の前記ポピュレーションのうち前記適応度が最高である前記個体をトップ個体と呼ぶことにし、当該世代からその次世代に世代交代する際にこのトップ個体は必ず生き残るというトップ保存戦略が併用されることを特徴とする。

本手段では、トップ個体の生存が保証されているので、仮に探索が正解に至っていた場合には、世代交代の際にその正解に至った個体が淘汰されてしまうような不都合は防止される。その結果、いったん正解に収束してしまえば、そこから外れてしまうことがなくなり、しかも世代交代を重ねてもトップ個体が交代することはなくなるから、収束判定が容易になる。

したがって本手段によれば、正解に収束したらもはや動かず、収束判定が容易になるという効果がある。

(5) 代替手段(第1手段の別表現)

(第15手段)

本発明の第15手段は、前述の第1手段とおおむね同様の「ハイブリッドGAによる複数パラメータの最適化方法」である。

本手段は、各世代のポピュレーションを形成する個体のそれぞれは、所定の物理量に対応した複数のパラメータを含む記号列からなり、これらのパラメータがある程度の多様性をもって分布した初期世代から所定のルールで世代交代を重ねるうちに、より高い適応度をもつ前記個体が生成されることをもって、前記複数のパラメータの最適化を図るGAによる複数パラメータの最適化方法であることを前提とする。本手段の特徴は、全世代のうち一部にあたる複数の特定世代において、当該世代の前記ポピュレーションのうち少なくとも一部にあたる特定個体を対象に、局所最適化操作が行われることと、全世代のうち少なくとも一部にあたる複数の特定世代において、当該世代の前記ポピュレーションのうち一部にあたる特定個体を対象に、局所最適化操作が行われることとのうち、少なくとも一方であることである。

本手段の作用効果は、第1手段の作用効果とおおむね同様である。

なお、前述の第1手段に従属した上記の各解決手段に相当する限定を設けたヴァリエーションを、本手段に対しても同様に従属して設定することができ、それぞれ対応する上記解決手段とおおむね同様の作用効果が得られる。

2. パターンマッチングによるデータ解析方法

(第16手段)

本発明の第16手段は、前述の第1手段～第15手段のうちいずれかに記載された「ハイブリッドGAによる複数パラメータの最適化方法」において、各前記個体をもつ前記適応度は、当該個体の前記記号列に含まれる前記複数のパラメータに対応する前記物理量を仮定すると数値演算により決まる仮想データのパターンが、実測データのパターンとどの

10

20

30

40

50

程度一致しているかを示す指標であり、この適応度が高い前記個体を生成することによって、当該個体の前記仮想データと前記実測データとの間でパターンマッチングを行い、前記物理量を推定することを目的として前記実測データの解析を行うことを特徴とする「パターンマッチングによるデータ解析方法」である。

ここでいう適応度は、普通はパターンマッチングないしデータマッチングの程度を示す指標であるから、逆に二乗誤差などマッチング不足の程度（ずれの大きさ）を示す指標を利用して良い。

本手段では、その中核となるハイブリッドGAが、前述の第1手段～第15手段のうちいずれかに記載された特徴を有しているため、それぞれに特有の作用効果が得られる。その結果、本手段によれば、パターンマッチングの正解に対して、ハイブリッドGAがより速やかに収束するようになり、データ解析方法に要する計算時間が短縮されるという効果がある。

10

### 3. HGAの応用たる物質構造の推定方法

#### (第17手段)

本発明の第17手段は、以下の特徴を持つ「放射線回折データに基づく物質構造の推定方法」である。ここでは控えめに「推定」という言葉を用いたが、本手段は、かなりの（少なくとも両従来技術を大幅に凌駕する）確度をもって推定することができ、しかもその推定精度もかなり高いので、実質的には物質構造の「決定」を行う方法である。

本手段の特徴は、前述の第1手段～第15手段のうちいずれかに記載された「ハイブリッドGAによる複数パラメータの最適化方法」を、所定の試料物質に照射された放射線が形成する回折パターンに基づいてこの試料物質の構造を推定する物質構造の推定方法（ないし決定方法）に応用したことである。そして、本手段では、前記物理量のうち少なくとも一部は、この試料物質の物質構造を決定する構造パラメータである。また、各前記個体の前記適応度は、この試料物質が当該個体に相当するこれらの構造パラメータをもつと仮定すると生じる回折パターンが、実測された前記回折パターンに対してどの程度一致しているかを示す指標である。

20

本手段では、その中核となるハイブリッドGAが、前述の第1手段～第15手段のうちいずれかであり、当該手段の項に記載された特徴を有しているため、その特徴に起因してそれぞれに特有の作用効果が得られる。すなわち、放射線回折パターンのパターンマッチングが行われる際に、本手段が用いるハイブリッドGAの特徴によって構造パラメータの正解に対する収束が、従来技術1や従来技術2に比べて格段に速やかになる。

30

したがって本手段によれば、より多数のパラメータ空間を広域に探索してより精密に最適化することができるようになり、物質構造の推定ないし決定に要する計算時間が著しく短縮されるという効果がある。その結果、両従来技術によっては実際的に構造決定ができなかった物質まで、費用および時間の許容範囲内でその構造決定をすることが可能になるという効果がある。

#### (第18手段)

本発明の第18手段の「放射線回折データに基づく物質構造の推定方法」は、前述の第17手段において、前記構造パラメータとは、前記試料物質の結晶構造を規定する格子定数と、この試料物質を構成する分子内における三次元的な原子配列を規定する分子構造パラメータと、この分子内での確率的な電子密度分布を規定する電子密度分布パラメータと、この試料物質の結晶内での三次元的な原子配列を規定する結晶学的構造パラメータとのうち、少なくとも一つであることを特徴とする。

40

すなわち本手段は、前述の第18手段における「構造パラメータ」という用語の内容を具体的に限定するものである。

なお、ここでいう「試料物質の結晶内での三次元的な原子配列を規定する結晶学的構造パラメータとは、結晶の各軸方向の長さ $a$ 、 $b$ 、 $c$ をそれぞれ単位長さ1として規格化した分極座標 $(x, y, z)$ である。また、結晶内での分子全体の向き（姿勢）を定義する角度パラメータも含まれる。ここで、 $x, y, z$ はそれぞれ0～1の範囲の値で表すのが普通である。

50

## (第19手段)

本発明の第19手段は、前述の第17手段の「放射線回折データに基づく物質構造の推定方法」において、前記回折パターンは、その回折強度が同心円状に拡がって分布する回折環であることを特徴とする。

本手段によれば、試料物質から形成される放射線の回折パターンが回折環であっても、その試料物質の構造を決める構造パラメータを決定することができるようになるという効果がある。ここで、回折環ではピーク重複が大きく、回折データに含まれる情報量は単結晶が作る格子状の回折パターンに比べて格段に少なくなってしまうから、ある程度複雑な物質構造を決定することは困難であったことを付言しておく。

換言すれば、構造を決定したい物質で試料に適当な大きさの結晶を作ることが未だできなくても、その物質が形成している微細な結晶の格子定数や当該物質を構成している分子中の原子の立体配置構造などを決定することが、本手段によりできるようになる。その結果、結晶を成長させる技術が未だ確立されないほど生成間もない新物質の構造決定が可能になるから、医薬品や新材料の開発や物質設計の分野に大きな貢献をすることが、本手段には期待できる。

それは、本手段で使用されるハイブリッドGAが、極めて強力な広域での最適値探索能力と精密な局所最適化との両方を併せ持っているからである。すなわち、前述のように本手段の中核技術は本発明の「ハイブリッドGA」であり、それによって多数のパラメータを広域で探索して精密に最適化するのに要する計算時間が大幅に短縮されていることが、本手段の実用化の最大要因である。

## (第20手段)

本発明の第20手段は、前述の第19手段において、前記適応度は、前記回折環がもつ半径方向への強度分布の実測データと、仮定された前記構造パラメータに基づいて生成される仮定データとが、どの程度一致しているかを示す評価関数であることを特徴とする、「放射線回折データに基づく物質構造の推定方法」である。

本手段は、前述の第19手段において、その限定をより具体的にしたものである。本手段では、適応度として、例えば実測データと仮定データとの間にある誤差の二乗和などの評価関数を用いることができ、適応度の算出が容易になる。そうすれば、局所最適化操作には、最小二乗誤差法による多種多様な局所最小化手法の中から選択することができるようになる。

したがって本手段によれば、適応度の算出に関する演算負荷が低減されるので、計算時間がいっそう短縮されるという効果がある。

## (第21手段)

本発明の第21手段は、前述の第17手段などによく似た「放射線回折データに基づく物質構造の推定方法」である。

本手段では、複数の結晶を含む粉末試料、多結晶集合体である多結晶試料、非晶質（アモルファス）試料および溶液試料のうち少なくとも一つである試料物質の構造を特定するために、この試料物質に照射された放射線の回折によってもたらされる回折環の強度分布を示す回折データを解析する。すなわち、この回折データに基づき、この試料物質がもつ物質構造を示すいずれかの構造パラメータを、データマッチングにより推定する。

その推定アルゴリズムとしては、遺伝的アルゴリズム（GA）に局所最適化操作を組み込んだハイブリッドGA（HGA）を使用し、このHGAの構成に本手段の特徴がある。

ここで、前提となるGAとは、初期世代のポピュレーションを生成する発生過程と、交叉、複製、選択を繰り返して突然変異を交える世代交代過程とをもつ。すなわちGAでは、先ず発生過程で、前記構造パラメータがコーディングされた個体を、ある程度の多様性を持たせて複数個発生させ、初期世代のポピュレーションとする。この発生過程の後に、これらの個体の交叉および複製と、前記データマッチングの程度を示す適応度を評価してこれらの個体を淘汰する選択と、確率的に起こる突然変異とを繰り返す世代交代過程を行う。

そして、本手段の特徴には二つあり、第一の特徴は、このGAが、世代交代過程の間に

10

20

30

40

50

複数の世代において、当該ポピュレーションの一部個体に対して局所最適化操作を行うHGAであることである。第二の特徴は、この局所最適化操作の対象には、当該ポピュレーションのうち適応度が比較的低い個体もなる可能性があることである。

本手段では、GAの世代交代のうち、複数世代で当該ポピュレーションのうち適応度が比較的低い個体も局所最適化操作の対象になるから、第7手段とほぼ同様の作用効果が得られる。その結果、本手段によれば、推定すべき構造パラメータの正解に収束するまでの計算時間が大幅に短縮され、今までは計算時間が膨大になるなどして実現できなかった物質構造の決定が可能になるという効果がある。

(第22手段)

本発明の第22手段は、以下のような解析手順を構成要素とした「放射線回折データに基づく物質構造の推定方法」である。

すなわち本手段は、放射線を試料物質に照射して得られた回折環の強度分布データに基づき、この強度分布データに適合する物質構造を特定するために、格子定数決定プロセスと背景散乱除去プロセスと空間群絞り込みプロセスと基本構造決定プロセスと構造精密化プロセスとのうち、少なくとも格子定数決定プロセスと空間群絞り込みプロセスと基本構造決定プロセスとを有する。そして本手段では、格子定数決定プロセスおよび基本構造決定プロセスのうち少なくとも一方で、GA(遺伝的アルゴリズム)に局所最適化操作を組み合わせたハイブリッドGAを利用した物質構造決定が行われることを特徴とする。

第一に、格子定数決定プロセスとは、この試料物質に含まれる結晶の格子定数を決定するプロセスである。すなわち、このプロセスでは、当該結晶の単位格子を定義する格子定数、例えば格子定数(a, b, c, , )などが決定される。

第二に、背景散乱除去プロセスとは、この強度分布データに含まれているバックグラウンド散乱の影響を、除去ないし低減するプロセスである。ここで、バックグラウンド散乱の原因には種々のものがあるが、いずれも有用な情報ないし信号を埋もれさせるノイズとして作用するので、その影響をできるだけ減らすことが必要である。

第三に、空間群絞り込みプロセスとは、この結晶が属する空間群の候補を絞り込むプロセスである。このプロセスでは、回折パターンのピークないし反射の出現位置に基づき、結晶が取りうる230種類の空間群のうちから可能性の低いものを除き、可能性が高い空間群を数種類程度に絞り込むことが行われる。

第四に、基本構造決定プロセスとは、この結晶の基本構造を初期的に決定して、この空間群を一つに特定するプロセスである。すなわち、このプロセスでは、前述の空間群絞り込みプロセスで絞り込まれた空間群の候補の中から、最も確からしい空間群が決定され、単位格子内での分子の位置及び姿勢と三次元的な原子配列とが、やや粗めの精度で決定される。

第五に、構造精密化プロセスとは、この基本構造を精密化して、この結晶を構成する物質の構造をより精密に決定するプロセスである。さらに、この構造精密化プロセスでは、分子内での電子密度の分布をも推定ないし決定することがある。

本手段の最大の特徴は、前記ハイブリッドGAでは、世代交代の間に複数の特定世代で、当該世代のポピュレーションのうち一部の個体である特定個体を対象にして前記局所最適化操作が行われることと、この特定個体には、当該世代のポピュレーションのうちでは実測データとの一致度を示す適応度が比較的低い劣等グループからも選定される可能性があることである。

本手段では、放射線回折データに基づいて物質構造を段階的に推定ないし決定していく手順が明らかにされており、さらにその手順の中でハイブリッドGAを用いるべきプロセスはどれであるかも明らかにされている。そればかりではなく、そのハイブリッドGAがもつ特徴も限定されており、これによって構造パラメータが正解に速やかに収束し、計算時間が大幅に短縮される。

したがって本手段によれば、粉末試料が形成する回折環に基づく物質構造の決定など、今までは計算時間が膨大になって実現できなかった新物質の構造決定も、所定の手順を踏んで実現することが可能になるという効果がある。

10

20

30

40

50

## [ 物の発明のカテゴリー ]

## 1. ソフトウェア関連発明

## (1) 最適パラメータ探索プログラム

## (第23手段)

本発明の第23手段は、コンピュータに実行可能な「最適パラメータ探索プログラム」であり、所定の物理量に対応した複数のパラメータの適正な組合せを探索する機能を、所定のコンピュータに発揮させるべきコンピュータ・プログラムであることを前提にする。本手段の特徴は、第1手段～第14手段のうちいずれかに記載された「ハイブリッドGAによる複数パラメータの最適化方法」を実行するために、前記初期世代の前記ポピュレーションの生成および読込のうち一方を行う初期世代設定ステップと、前記評価過程、前記淘汰過程、前記交叉過程、前記突然変異過程および前記局所最適化過程のそれぞれに対応する各演算ステップ（すなわち、評価ステップ、淘汰ステップ、交叉ステップ、突然変異ステップおよび局所最適化ステップ）とを有することである。

10

換言するならば、本手段の「最適パラメータ探索プログラム」を所定のコンピュータに実行させることにより、前述の第1手段～第14手段の「ハイブリッドGAによる複数パラメータの最適化方法」は実現される。

## (第24手段)

本発明の第24手段は、コンピュータにより実行可能な「最適パラメータ探索プログラム」であり、所定の物理量に対応した複数のパラメータの適正な組合せを探索する機能を、所定のコンピュータに発揮させるべきコンピュータ・プログラムであることを前提とする。

20

本手段の特徴は、まず、各世代のポピュレーションを形成する個体のそれぞれは、前記所定の物理量に対応した前記複数のパラメータを含む記号列からなり、これらのパラメータがある程度の多様性をもって分布した初期世代から所定のルールで世代交代を重ねるうちに、より高い適応度をもつ前記個体が生成されることをもって、前記複数のパラメータの最適化を図るGA（遺伝的アルゴリズム）の中に、前記適応度に関する局所最適化操作を所定の前記個体に対して行う局所最適化ステップが組み込まれていることである。

本手段はさらに、この局所最適化ステップが、前記GAで生成される全世代のうち一部にあたる複数の特定世代において、当該世代の前記ポピュレーションのうち少なくとも一部にあたる特定個体を対象にする演算ステップであることと、前記GAで生成される全世代のうち少なくとも一部にあたる複数の特定世代において、当該世代の前記ポピュレーションのうち一部にあたる特定個体を対象にする演算ステップであることとのうち、少なくとも一方であることを特徴としている。

30

本手段では、GAが世代交代を重ねるうちに、複数の特定世代で、適応度に関する局所最適化操作を特定個体に対して行う局所最適化ステップが組み込まれているから、最適パラメータ探索の収束が速くなり、計算時間が大幅に短縮される。その結果、本手段によれば、従来は計算時間が膨大にかかって実際上不可能であった複雑な複数パラメータの広域探索問題が、実用的な計算時間内に解けるようになるという効果がある。

なお、本手段についても、前述の第1手段に従属した第2手段～第14手段のうちいずれかに相当する限定を付けた各種ヴァリエーションが実施可能であり、それぞれに相当する作用効果が得られる。

40

## (第25手段)

本発明の第25手段は、コンピュータにより実行可能な「最適パラメータ探索プログラム」であり、複数のパラメータの適正な組合せを探索する機能を、所定のコンピュータに発揮させるべきコンピュータ・プログラムであることを前提にしている。

本手段の特徴は、各世代のポピュレーションを形成する個体のそれぞれは、前記複数のパラメータに対応した記号列を含んでおり、これらのパラメータがある程度の多様性をもって分布した初期世代のポピュレーションから始めて、所定のルールで世代交代を重ねるうちに、より高い適応度をもつ前記個体が生成されることによって前記複数のパラメータの最適化を図るGA（遺伝的アルゴリズム）の中に、前記適応度に関する局所最適化操作

50

を所定の前記個体に対して行う局所最適化ステップが組み込まれていることである。

本手段はさらに、この局所最適化ステップが、前記GAで生成される全世代のうち一部にあたる複数の特定世代において、当該世代の前記ポピュレーションのうち少なくとも一部にあたる特定個体を対象にする演算ステップであることと、前記GAで生成される全世代のうち少なくとも一部にあたる複数の特定世代において、当該世代の前記ポピュレーションのうち一部にあたる特定個体を対象にする演算ステップであることとのうち、少なくとも一方であることをも特徴としている。

すなわち本手段は、前述の第24手段から、最適な組合せを探索すべき複数のパラメータが物理量に関するものであるとの限定を外したものであり、その結果、物理量以外を探索対象とする多数のパラメータの最適な組合せを探索する問題にも適用することができるようになる。したがって本手段によれば、前述の第23手段や第24手段とほぼ同様の作用効果が得られながら、その適用範囲が、政治決定や経営問題などの社会的な問題やテーブルゲームなどを含むより広い分野にまで広がるという効果がある。

なお、本手段についても、前述の第1手段に従属した第2手段～第14手段のうちいずれかに相当する限定を付けた各種ヴァリエーションが実施可能であり、それぞれに相当する作用効果が得られる。

## (2) 物質構造推定プログラム

### (第26手段)

本発明の第26手段は、コンピュータにより実行可能な「放射線回折データに基づく物質構造推定プログラム」であり、所定の試料に照射された放射線が形成する回折パターンに基づいて当該試料を構成する物質の構造を推定する機能を、所定のコンピュータに発揮させるべきコンピュータ・プログラムであることを前提にしている。

本手段の特徴は、前記回折パターンの実測データを数値的に読み込む実測データ読込ステップと、第23手段記載の「最適パラメータ探索プログラム」が有する前記初期世代設定ステップ、前記評価ステップ、前記淘汰ステップ、前記交叉ステップ、前記突然変異ステップおよび前記局所最適化ステップとを有し、前記適応度は、前記各個体の前記記号列に相当する仮定された前記構造パラメータから算出される前記回折パターンの予想値が、前記回折パターンの実測値に対してどの程度一致するかを示す指数であることである。

本手段では、放射線回折データに基づく物質構造の推定が、前述の第23手段に相当する強力な探索能力を持つプログラムによって行われる。その結果、本手段によれば、従来は計算時間が膨大にかかって實際上不可能であった複雑な複数パラメータの広域探索問題が、実用的な計算時間内に解けるようになるという効果がある。

### (第27手段)

本発明の第27手段は、コンピュータにより実行可能な「放射線回折データに基づく物質構造推定プログラム」である。本手段は、所定の試料に照射された放射線が形成する回折パターンに基づいて、当該試料を構成している物質の構造を規定する構造パラメータを推定する機能を、所定のコンピュータに発揮させるべきコンピュータ・プログラムであることを前提にしている。

本手段の特徴は、各世代のポピュレーションを形成する複数個の個体のそれぞれは、前記構造パラメータに対応する記号列を含んでおり、これらの個体のそれぞれがもつ適応度は、当該個体に対応したこの構造パラメータから算出される回折パターンの予想値が、前記回折パターンの実測値とどの程度一致するかを示す指標であって、この構造パラメータの各要素がある程度の多様性をもって分布するように、これらの個体を生成して初期世代ポピュレーションとする生成ステップをもち、この初期世代ポピュレーションから淘汰ステップ、交叉ステップおよび突然変異ステップを含む所定の手順に従って世代交代を重ねるうちに、当該世代ポピュレーションの中にこの適応度がより高い前記個体が生成されることを待ち、もって前記構造パラメータの最適化を図るGA(遺伝的アルゴリズム)の中に、前記適応度に関する局所最適化操作を所定の前記個体に対して行う局所最適化ステップが組み込まれていることである。本手段はさらに、この局所最適化ステップが、前記GAで生成される全世代のうち一部にあたる複数の特定世代において、当該世代の前記ポピ

10

20

30

40

50

ュレーションのうち少なくとも一部にあたる特定個体を対象にする演算ステップであることと、前記GAで生成される全世代のうち少なくとも一部にあたる複数の特定世代において、当該世代の前記ポピュレーションのうち一部にあたる特定個体を対象にする演算ステップであることとのうち、少なくとも一方であることを特徴とする。

本手段では、放射線回折パターンに基づく物質構造の推定が強力な探索能力を持つハイブリッドGAを利用して行われる。その結果、本手段によれば、従来は計算時間が膨大にかかって実際上不可能であった複雑な複数パラメータの広域探索問題が、実用的な計算時間内に解けるようになるという効果がある。

なお、本手段についても、前述の第1手段に従属した第2手段～第14手段のうちいずれかに相当する限定を付けた各種ヴァリエーションが実施可能であり、それぞれに相当する作用効果が得られる。

10

### (3) 記録媒体

#### (第28手段)

本発明の第28手段は、コンピュータ読み取り可能な記録媒体である。そして本手段の特徴は、前述の第23手段～第25手段のうちいずれかである「最適パラメータ探索プログラム」と、前述の第26手段～第27手段のうち一方である「放射線回折データに基づく物質構造推定プログラム」とのうち、少なくとも一方が記録されていることである。

本手段によれば、これらのプログラムの保存および流通が容易になる。

## 2. 装置発明

#### (第29手段)

本発明の第29手段は、「ハイブリッドGA搭載コンピュータ」である。本手段の特徴は、前述の第23手段～第25手段のうちいずれかである「最適パラメータ探索プログラム」と、前述の第26手段～第27手段のうち一方である「放射線回折データに基づく物質構造推定プログラム」とのうち、少なくとも一方がインストールされたコンピュータであることである。

20

本手段のコンピュータにより、これらのプログラムの実行が可能になり、その結果、多次元の広域で最適パラメータを精密に探索することや、今まではできなかった物質構造の決定ができるようになる。

#### (第30手段)

本発明の第30手段は、「ハイブリッドGAを搭載した解析装置」であり、前述の第23手段～第25手段のうちいずれかである「最適パラメータ探索プログラム」と、前述の第26手段～第27手段のうち一方である「放射線回折データに基づく物質構造推定プログラム」とのうち、少なくとも一方を実行可能な解析装置であることを特徴とする。

30

本手段の解析装置により、当該装置がコンピュータに属するか否かの判断に関わりなく、これらのプログラムの実行が可能になる。

#### (第31手段)

本発明の第31手段は、「放射線回折パターンによる物質構造推定装置」であり、所定の試料物質に放射線を照射してこの試料物質から生じる回折パターンを測定する測定装置と、この回折パターンから得られた実測データに基づいた解析を行うことにより当該試料物質の構造を推定する第29手段のコンピュータおよび第30手段の解析装置のうち一方とを有することを特徴とする。

40

ここで、上記測定装置と上記コンピュータまたは解析装置とが同じ敷地にあるか否かとか、両者の間で行われる回折データの移送がオンラインかオフラインかとか、両者の所有権ないし使用権の帰属先が同一か否かとかは、本質的な問題とはならない。

本手段によれば、試料物質に放射線を照射して回折パターンを測定し、この回折パターンに基づいて当該試料の物質構造を決定するまでの一連の操作手順を一貫して行うことができる。しかも、前述のように、従来は現実的なコストや期間の範囲内ではできなかった物質の構造決定までが可能になるという効果がある。

### 【図面の簡単な説明】

【0005】

50

図1は、単結晶試料と粉末試料とでX線回折パターンを比較して示す組図である。同図のうち、図1(a)は、単結晶試料で生じるX線回折とその回折パターンとを示す模式図であり、図1(b)粉末試料で生じるX線回折とその回折環とを示す模式図である。

図2は、本発明の中核をなすHGAの第一特徴を従来技術と比較して示す組図である。同図のうち、図2(a)は従来技術1としてのHGAの概要構成を示す流れ図であり、図2(b)は従来技術2としてのHGAの概要構成を示す流れ図であり、図2(c)は本発明の実施形態でのHGAの概要構成を示す流れ図である。

図3は、本発明の中核をなすHGAの第二特徴とその作用とを示す概念図である。

図4は、X線回折の比較試験に供された各種粉体試料の物質構造を示す組図である。同図のうち、図4(a)は $Pd(tmdt)_2$ の、図4(b)は $Au(tmdt)_2$ の、図4(c)はシチジンの、図4(d)はタウリンの、図4(e)は $1-FcAqH(ClO_4)$ の、図4(f)はコハク酸ブレドニゾロンの、それぞれ立体分子構造および結晶構造を示す斜視図である。

10

図5は、本発明の一実施例でのHGA収束速度を両従来技術と比較して示すグラフである。

図6は、本発明によって決定された $1-FcAqH(ClO_4)$ の物質構造を示す斜視図である。

図7は、本発明の実施形態としての物質構造の決定方法の手順を示す流れ図である。

図8は、粉末試料から得た回折環の強度分布を例示する実測データのグラフである。

図9は、分子中の三次元的な原子配列と結晶構造とを模式的に示す斜視図である。

20

図10は、単位格子とその格子定数とを例示する模式図である。

図11は、本実施形態がもつ格子定数決定プロセスの手順概要を示す流れ図である。

図12は、回折環のX線強度分布データにおけるピーク位置を例示するグラフである。

図13は、強度分布の実測データに含まれるバックグラウンド散乱を例示するグラフである。

図14は、本実施形態がもつ背景散乱除去プロセスの作用を示す組図である。同図のうち、図14(a)は粉末試料 $CsPbCl_3$ からのX線強度分布実測データのグラフであり、図14(b)はガラスキャピラリーから生じるハローパターンを示すグラフである。また、図14(c)は空気散乱の強度分布を示すグラフであり、図14(d)は粉末試料からの温度散漫散乱が示す強度分布のグラフであり、図14(e)実測データに各背景散乱をフィッティングさせた結果を示すグラフである。

30

図15は、基本構造決定プロセスで決定される構造パラメータの種類を示す組図である。同図のうち、図15(a)は構造パラメータのうち分子の位置(x, y, z)を示す斜視図であり、図15(b)は構造パラメータのうち分子の姿勢( , , )を示す斜視図である。また、図15(c)は構造パラメータのうち複数の分子内回転角  $\theta_i$ を示す斜視図である。

図16は、構造精密化プロセスにより調整されるサイズパラメータを示す模式図である。

図17は、構造精密化プロセスによるフィッティング結果例1を示す強度分布グラフである。

40

図18は、構造精密化プロセスによるフィッティング結果例2を示す強度分布グラフである。

図19は、本実施形態で使用するハイブリッドGAの処理手順を示す流れ図である。

図20は、格子定数決定プロセスでHGAが行う格子定数のグループ化を示す模式図である。

図21は、パラメータのグループ化をしないと生じる不都合を例示する組図である。同図のうち、図21(a)は親個体の記号列に相当する格子定数をもつ単位格子を示す模式図であり、図21(b)は子個体の記号列に相当する格子定数をもつ単位格子を示す模式図である。

図22は、格子定数決定プロセスのHGAでの三斜晶系の交叉パターンを示す組図であ

50

る。同図のうち、図 2 2 ( a ) は交叉パターン 1 の、図 2 2 ( b ) は交叉パターン 2 の、図 2 2 ( c ) は交叉パターン 3 の、それぞれ親個体と子個体とを比較して示す模式図である。

図 2 3 は、格子定数決定プロセスの H G A での単斜晶系の交叉パターンを示す組図である。同図のうち、図 2 3 ( a ) は交叉パターン 1 の、図 2 3 ( b ) は交叉パターン 2 の、図 2 3 ( c ) は交叉パターン 3 の、それぞれ親個体と子個体とを比較して示す模式図である。

なお、上記各図中で使用されている符号のうち主たるものは、以下のようなものである。

- P 1 : 格子定数決定プロセス (ハイブリッド G A で最適パラメータ探索) 10
- P 2 : 背景散乱除去プロセス
- P 3 : 空間群絞り込みプロセス
- P 4 : 基本構造決定プロセス (ハイブリッド G A で最適パラメータ探索)
- P 5 : 構造精密化プロセス
- D : 読込ステップ (回折環の X 線強度分布データの読み込み)
- S 0 : 前処理ステップ (回折角 2 軸上で強度ピーク位置を抽出)
- S 1 : ハイブリッド G A ステップ
- S 1 1 : 初期世代設定ステップ (第 1 世代ポピュレーションを生成)
- S 1 0 2 : 記録ステップ (当該世代ポピュレーションの各個体を記録)
- S 1 0 3 : 交叉ステップ (親個体を選択して交叉させ新生個体を生成) 20
- S 1 0 4 : 複製ステップ (生き残る個体を選定し複製して次世代に保存)
- S 1 0 5 : 淘汰ステップ (新生個体および複製個体から選択して淘汰)
- S 1 0 6 : 判定ステップ (突然変異を起こすか確率的に判定)
- S 1 0 7 : 突然変異ステップ (特定個体の記号列の一部を強制的に変更)
- S 1 0 8 : 判定ステップ (特定世代の判定および特定個体の選定)
- S 1 0 9 : 局所最適化ステップ (特定個体に対し局所最適化操作)
- S 1 1 0 : 収束判定ステップ (解の収束判定等による終了判定)
- S 1 1 1 : インクリメントステップ (世代番号のインクリメント)
- S 1 2 : 結果出力ステップ (決定された構造パラメータ等を入力)
- L : 世代交代ループ 30
- LC : 格子定数
  - ・三斜晶系の格子定数 ( a , b , c , , )
  - ・単斜晶系の格子定数 ( a , b , c , )
- d : 結晶中の格子面間隔 (数 2 参照)
- t : X 線検出器原点の位置ずれ (探索すべきパラメータの一つ)
- R : 一致度 (格子定数から算出されたピーク位置とその実測データとの差異、数 3 参照)
- R w p : 信頼度因子 (数 1 で定義される数値であり、低いほど適応度としての信頼度が高い。)
- 【発明を実施するための最良の形態】 40
- 【 0 0 0 6 】
- 本項は記載内容が大量でありその構成も複雑であるから、読者が本項の全体像を把握しやすいように、先ず目次 (項目のみ) を設けておく。
- この目次をご覧になってお判りのように、本実施形態の説明の順番としては、先ず物質構造の決定技術の分野において本発明が発揮する効果を理解していただくために、本発明の実施によって得られた研究成果を先に提示する。しかる後、そのような研究成果を得るには、具体的にどのような構成の実施形態で本発明が実施されていたのかを説明する。
- [ 目次 ]
- 1 . 本実施形態の意義
- ( 1 ) 巨大設備での先端的研究
- ( 2 ) 実験室レベルへの簡易化 50

## 2. 本発明の実施例

- (1) 本発明の実施形態がもつ二大特徴
- (2) 本発明の実施例1, 2と比較例との対比
- (3) 実施例で実施された解決手段

## 3. データ解析方法の全体構成

- (0) 解析手順の概要
- (1) 格子定数決定プロセス
- (2) 背景散乱除去プロセス
- (3) 空間群絞り込みプロセス
- (4) 基本構造決定プロセス
- (5) 格子定数決定プロセス

10

## 4. 格子定数決定プロセスでのHGA

- (1) HGAの概要
- (2) HGAの詳細

## 5. 基本構造決定プロセスでのHGA

- (1) HGAの概要
- (2) HGAの詳細

## 6. その他の実施例

### 1. 本実施形態の意義

この項では、本発明の「放射線回折データに基づく物質構造の推定方法」及びその関連発明の実施形態と、これらの中核技術たるハイブリッドGAの実施形態とについて、この分野の他の研究者等にも実施可能な理解が得られるように、明確かつ十分に説明する。ここに記載されていることは、おおよそ発明者らが実際にX線回折環の測定試験およびそのデータ解析を行って好結果を得たという実績に基づいている。

20

#### (1) 巨大設備での先端的研究

前述のように、発明者らはSpring-8(大規模設備の固有名詞:本註以下略)で生成される世界最高クラスのエネルギー強度ないし輝度を持つ放射光X線を用い、各種の試料物質を対象にその構造解析ないし構造決定に取り組んできた。そしてSpring-8によれば、前述のようにそのX線の強度ないし輝度が抜群に高いので、ほんの数mg程度の粉末試料からでも、他の設備では得られないほど極端に鋭く立ち上がったピークを多

30

数含む世界最高精度の回折環データが得られる。ところが、これだけ高い精度で回折環の実測データが得られるようになると、逆にこの計測精度に対し、有効にデータ解析を行って物質構造を決定することができるほどの強力なデータ解析方法がなかった。それ故、今まではSpring-8で回折環が測定されても、そのデータに基づいて構造決定を行うことができず、せっかくのSpring-8の測定能力を生かし切れていないという皮肉な状態が続いていた。

そこで発明者らは、前述の発明を完成させ、以下に記載する形態で本願発明を実施することにより、Spring-8で粉末試料に高エネルギーX線を照射して得た回折環の実測データに基づいて、各種の物質の構造を決定することができるようにした。

その結果、従来は構造決定の信頼性も精度も低く、間違いも多かったために、医薬品業界では構造決定とは呼ばず「構造予測」と位置づけていたものを、十分な確度と精度とを持って「構造決定」できるようになった。すなわち、医薬品に使われる有機低分子の構造決定が、極めて高い実測データとの一致度(例えば98%,信頼度因子Rwp=2%)をもってできるようになっており、その際の誤差は、単位格子の各辺でわずか0.0001nm程度、原子の配列でも同じく0.0001nm程度と推定される。

40

このようにして世界で初めて決定した物質構造の数々については、これまでも発明者らが学会の講演会や論文で発表してきた。しかし、本願発明の物質構造の決定技術それ自体と、その中核技術たるハイブリッドGAとについては、本願の出願までに具体的に公開したことがない。

#### (2) 実験室レベルへの簡易化

50

本願発明者らは、前述のように世界最高クラスの大規模設備による先端的な研究を進める一方で、このような物質構造の決定方法を大学や企業に普及している実験室レベルの普及型設備でも実施することを試みた。

その結果、大学や企業の実験室に普及しているようなX線回折装置で測定した回折環のデータを解析しても、前述のSpring-8での測定データを用いて行った物質構造の決定に準ずる成果を上げられるようになった。しかも、そのデータ解析に使用したのは、上級機とはいえ市販のPCであった。

ここで、Spring-8のような大規模設備と実験室レベルの普及型設備との間に、X線回折装置としてどのような違いがあるかということ、本質的にはX線強度とエネルギー分解能、そしてX線の平行度において大規模設備の方が優れているということである。すなわち、X線強度の違いは、短時間測定が可能か否かの違いにつながり、ひいてはごく微量の試料による測定の可否につながる。また、エネルギー分解能の違いとX線の平行度の違いとは、そのままS/N比の違いとなって現れる。要するに、X線回折環データではピーク重複などによってただでさえ情報量が少ないというのに、大規模設備に比べて普及型設備ではいっそう少ない情報量しか与えられないというわけである。それゆえに、普及型設備で測定された回折データに基づいての構造決定は、大規模設備に比べていっそう難しくなるわけである。

また、このように世界最高レベルの大規模設備から実験室レベルの普及型設備に簡易化して本願発明を実施できるようにするにあたり、データ解析手法には特筆すべき変更点はない。すなわち、本願発明の中核となるハイブリッドGAについては基本的に同一のものを同一手法で使用しており、データ解析プログラムも基本的に同一なら、それを走らせるPCも同一である。

したがって、本実施形態のように、本発明は、どこの大学や企業であっても自前で用意できる程度の普及型設備でも、物質構造の決定を進められるという広範な普及性をもつ。その結果、本発明によれば、大規模設備がなくてもそれに準ずる結晶構造ないし物質構造の決定ができるようになるから、構造決定を要する研究の費用対効果およびペースが劇的に向上することが期待できる。

## 2. 本発明の実施例

### (1) 本発明の実施形態がもつ二大特徴

本発明の実施形態たる「放射線回折データに基づく物質構造の推定方法」では、格子定数決定プロセスと基本構造決定プロセスとで、本願発明の中核となるハイブリッドGAが使用されている。物質構造の決定方法としての本実施形態の詳細については後に譲り、ここでは、このハイブリッドGA（「本HGA」と呼ぶ）がもつ特徴点のうち二つだけについて、概念的に説明するものとする。

#### ・第一特徴：局所最適化操作が散在

本HGAがもつ第一の特徴は、図2(c)に示すように、GAの世代交代を重ねるうちに分布した複数の世代（特定世代）において局所最適化操作が行われ、しかもその対象となる個体（特定個体）は例えば一つというようにごく少数であることである。

ここで、特定世代は「分布した複数の世代」である旨の表現は、全世代のうちに特定世代が飛び飛びに広く散在する場合だけではなく、世代交代する毎に特定世代があり特定世代が連続する場合をも含む。それゆえ、本HGAを示す図2(c)では、ラマルキアンGAを示す図2(b)との差異をより際立たせるために、隔世で（つまり飛び飛びに）しか局所最適化が行われなないように描いたが、世代交代毎に局所最適化が行われるように描いても良い。

すなわち、本実施形態で使用されているハイブリッドGAでは、局所最適化操作が行われるのは時折でしかなく、その対象となる個体はポピュレーションのうちごく少数でしかない。具体的には、特定世代は世代交代のうち所定周期（例えば毎世代とか、あるいは二三世代毎ないし十世代毎とか）で現れる。そして、当該世代のポピュレーション（例えば個体数で千個とか五百個とか、あるいは絞り込んでもっと少数個とか）のうち特定個体の数は、例えば一個だけとかせいぜい二三個ないし数個とかの程度である。そのうち本実施

10

20

30

40

50

形態で最も多用した設定は、特定世代は毎世代（つまり全世代）であり、各世代のポピュレーションのうち特定個体は一個だけであった。それゆえ、本実施形態では、演算量の大きい局所最適化操作が行われる回数が新生個体の発生個数に比べて極めて少なく抑制されている。

ここで、特定世代の分布を、飛び飛びにはあっても全世代のうちほぼ全域に散在するように設定しているのは、単に局所最適化操作の回数を減らして演算負荷を軽減したいというだけではない。それだけではなく、進化に対する方向性が与えられる程度が、世代交代の全域でほぼ均等になるようにという意図もあってのことである。

また、特定個体の数が当該ポピュレーションのうち一つとかというように極めて少ない数に限定されているのは、局所最適化操作の回数を減らして演算負荷を軽減したいというだけではない。それだけではなく、なるべく特定世代の数を増やして頻繁に進化の方向性を与える操作をしたいという意図もあってのことである。

それゆえ、世代交代する毎に全ての新しい世代が特定世代となり、その特定世代での特定個体の数は極めて少数（例えば一つだけ）であるという実施形態が、おそらく最良の実施形態である。なぜならば、局所最適化操作の行われる回数が新生個体数に対して極めて小さいので、ラマルキアンGAに比べて演算負荷が劇的に低減されておりながら、継続的に方向性が適度に与えられて進化が加速されるからである。しかも、ラマルキアンGAとは違い、局所最適化される個体の数に比べて新生個体の数の方が圧倒的に多いので、ポピュレーションを構成する個体の多様性が損なわれることなく世代交代が進められる。

さて、このように所定世代あたりの局所最適化操作の回数が、図2(b)に示す従来技術2に比べてごく少数であるから、本HGAでは遺伝的アルゴリズムが本来持っている高速性が損なわれることがない。それでいながら本HGAでは、時折行われる局所最適化操作によってGAに進化の方向性ないし指向性が導入されているので、進化に方向性が与えられていない通常のGAに比べ、広域的な最適解に対する収束性が格段に向上している。すなわち、図2(a)に示す従来技術1のように、確率的な探索しかできないGAが最適値の近傍に収束するまで、かなり長い時間を待たなければならないということはない。

その結果、多峰性のきつい多次元パラメータ空間を探索する場合であっても、従来技術1および従来技術2のいずれに比べても広域的な最適解に達するまでの演算量が桁違いに小さくなっている。すなわち、探索すべきパラメータ空間の次元数（つまり自由度）が高くその多峰性も強い場合であっても、局所最適の陥穽に捉えられることが避けられ、極めて効率よく広域的な探索を進めることが可能になる。

したがって、本HGAによれば、パラメータの数が多すぎたりパラメータ空間の多峰性がきつかったりして従来は解けなかったような最適化問題を、実用的な演算時間で精密に解くことができるようになる。つまり、本HGAによれば、演算負荷が小さく高速であるばかりでなく、極めて効率的に広域探索を進められるから、いずれの従来技術でも実際上できなかった多数パラメータの広域最適化ができるようになる。

#### ・第二特徴：局所最適化操作の対象選択

さて、ポピュレーションの多様性を確保する上でも、演算負荷の都合上でも、新生個体全てを局所最適化できず、上記のように局所最適化操作が稀にしか行われたいとするならば、次にはその個体をどう選ぶかを工夫しなければならない。

そこで本HGAは、その第二の特徴として、特定世代のポピュレーションのうちから局所最適化の対象個体（特定個体）を選ぶにあたり、GAの淘汰過程というエリート戦略のように適応度の高い個体を選ばず、むしろ適応度の低い個体を選んでいく。

本HGAでは何故このような選択志向を取っているかということ、図3に示すように、当該ポピュレーションのうちで適応度の低い個体の中にこそ、局所最適化を施せば広域的な最適解（「正解」と呼ぶ）に至る個体が紛れている可能性が高いからである。

さて本HGAでは、当該世代のポピュレーションの中で適応度が比較的低い劣等グループから特定個体を選定するにあたって、具体的には「逆ルーレット選択」（より詳しくは「値の制限付き逆ルーレット選択」）と名付けた選定方法を採用している。

10

20

30

40

50

この方法では、先ず当該世代（つまり特定世代）のポピュレーションに属する個体のうち、予め設定した割合に相当する適応度の高い個体（例えば上位10%）をエリートグループとして除いておく。次に、残りの（この例では下位90%の）個体に適応度で序列を付け、適応度が最悪の個体（0に該当）から最良の個体（1に該当）までを実数0～1の範囲に等間隔で割り振る。しかる後、一様乱数を発生させ、0～1の範囲に収まる個体で「劣等グループ」を構成する。そして、この劣等グループに属する個体を改めて適応度の序列に従って実数0～1の範囲に割り振ったうえで、一様乱数を発生させ、この範囲に相当する個体をこの劣等グループから選択する。こうして選択された個体を特定個体とし、当該ポピュレーションのうちこの個体にだけ局所最適化を施す。ここで、特定個体の数が複数に設定されている場合には、この処理を複数回繰り返すようにするか、あるいは一様乱数をその回数だけ発生させて劣等グループから指定個数の特定個体を選び出すようにする。

10

この「逆ルーレット選択」を本HGAに導入して特定個体の選択を行うと、適応度の高い個体を優先的に特定個体を選ぶ通常のルーレット選択に比べて、どの程度の効果があるかを次の表1に示す。

【表1】

選択方法	適応度	効果(計算時間)
ルーレット選択	平均0.75付近	自由度12～14では約1週間を要す。 (100～200時間程度)
逆ルーレット選択	平均0.25付近	自由度12～14では約1時間で済む。 (自由度20でも回折環から構造決定可能)

20

すなわち本HGAで、特定個体の平均的な適応度が0.25付近になるように設定した「逆ルーレット選択」をもって局所最適化操作の対象個体を選択するようにした場合には、同0.25付近のルーレット選択を採用した場合に比べ、広域的探索能力が飛躍的に向上している。同表にあるように、本願発明者らが行った数値解析例では、自由度12～14の多峰性パラメータ空間を広域的に探索する時間が、おおむね1/100～1/200程度に短縮されている。

30

それゆえ、特定個体の選定方式に逆ルーレット選択を導入すれば、通常のルーレット選択方式に比べて二桁以上も本HGAの探索能力が向上することが分かる。その結果、本HGAによって自由度の数が20を超える物質構造モデルについても、回折環データを解析して構造決定ができるようになっている。

余談ながら、発明者らは仲間内で、ある世代のポピュレーションのうち適応度の低い個体のことを「オチコボレ」（劣等生や落ち穂の意）と呼んでいる。また、このような志向で特定個体の選択をすることを「オチコボレ拾い」（落ち穂拾い *gleaning* に日本語では発音も意味も似た言葉）と呼び習わしている。

40

表1の基になった複数の実施例では、分子が百個程度の原子からなり決定すべき自由度が12～14程度の物質構造を決定するために、本発明のプログラムを1,000ポピュレーションで走らせて回折環データを解析している。この際の特定世代の出現頻度は毎世代毎にであり、各世代での特定個体の個数は一つである。また、表1の各実施例で物質構造決定の対象とした粉体試料の物質は、後述の実施例1や実施例2で構造決定されたものと同じである。

そして、まさにこれらの実施例で計算時間が異様に短くなる現象が発現し、その原因を追及したところ、発明者らは「逆ルーレット選択」に該当するプログラムミスを発見した。すなわち、異様に収束が速かった実施例では、プログラム中のバグにより意図せずに「

50

逆ルーレット選択」方式で特定個体が選ばれていることが分かった。その僥倖の結果、適応度の低い個体を優先的に選択して局所最適化した方が、最適パラメータの広域探索に要する計算時間が劇的に短縮されることを発明者らは発見することができた。

ただし、表1のデータはこのような発見をしたときのものであり、HGAを用いた本発明のプログラムでは、ポピュレーションの個体数など設定パラメータに未だ調整の余地を残していた。それゆえ、本プログラムの設定パラメータを適正にチューニングすれば、表1中に記載された計算時間をさらに短縮できる可能性がある。

・まとめ

さて、以上を簡単にまとめると、GAと局所最適化とが互いに補完し合って多峰性のパラメータ空間であっても効率的に広域探索を行うことができるように、本HGAは第一特徴および第二特徴の二つの特徴を有する。すなわち、第一特徴は、GAと局所最適化とをどのように組み合わせるかを工夫して、局所最適化操作の頻度を減らすことである。そして第二特徴は、ポピュレーションから局所最適化する個体をどう選ぶかを工夫して、その個体が局所最適化によって正解（すなわち広域的最適解）に至るものである確率をなるべく増すことである。

なんとすれば、ふつつ局所最適化をGAに導入する狙いは、GAのもつ広域的な探索を高速で（つまり小さい演算量で）行う能力を、局所最適化のもつ局所的な精密さで補うことにあるからである。

すなわちGAには、広域的な探索を高速で行うことができるという長所がある一方、遺漏が生じやすく、しかも詰めが甘いという短所がある。逆に局所最適化には、広域的な最適値の近傍からスタートすれば、遺漏なく精密にその最適値に到達することができるという長所がある一方、広域的な探索には向かないという短所がある。そこで本HGAでは、前述の第一特徴および第二特徴を併せ持つことにより、GAと局所最適化とが極めて巧みに組み合わせられて、それぞれの長所が最大限に生かされるようになった。

(2) 本発明の実施例1, 2と比較例との対比

本実施形態の効果は、次の表2を参照すれば一目瞭然である。

【表2】

ケース名→			比較例		実施例1		実施例2	
試験装置→			SPring8				実験室レベル	
符号	試料の化合物名	自由度	解析時間	一致度	解析時間	一致度	解析時間	一致度
(a)	Pd(tmdt) <sub>2</sub>	6	67.6秒	8.53%	1.1秒	8.00%	0.6秒	10.50%
(b)	Au(tmdt) <sub>2</sub>	6	36.7秒	5.30%	0.2秒	5.11%	0.2秒	3.40%
(c)	シチジン	6	33.3秒	2.10%	0.8秒	2.10%	0.8秒	8.60%
(d)	タウリン	8	構造決定不可		(未実施)		0.4秒	5%
(e)	1-FcAqH(C <sub>6</sub> H <sub>4</sub> O)	12	構造決定不可		7秒	9.00%	29秒	9%
(f)	コハク酸プレドニゾン	26	構造決定不可		約8時間	3.00%	(未実施)	

この表は、本実施形態の「放射線回折データに基づく物質構造の推定方法」によって決定された構造が決定された6種類の物質(a~f)を、その立体分子モデルの自由度と併せて一覧表にして示したものである。この表には、二種類の実施例(実施例1および実施例2)と比較例とで、その物質の構造決定にかかった解析時間とその結果の一致度とが、互いに比較できるように提示されている。

(実施例の概要)

ここで、実施例1というのは、前述のSPring-8で測定した世界最高レベルの精度をもつ回折環データに基づき、本実施形態の方法により粉末試料の物質構造を決定した本発明の実施例である。

一方、実施例 2 では、実施例 1 と同じ 6 種類の粉末試料について X 線回折環データを測定する装置として、実施例 1 の S P r i n g - 8 とは異なり、発明者らの属する大学の実験室に設置した X 線回折測定装置を使用している。この測定装置は、発明者らの出した要求仕様に基づき、株式会社リガク（固有組織名）によって制作されたものである。そして、この測定装置の取得費用は五千万円程度（おおむね四十万米ドルに相当）で済んでおり、この額は S P r i n g - 8 の建設に要した設備投資に比べてわずか千分の一ほどである。なお、同装置のスペックは、おおむね以下の通りである。

- ・ X 線の波長：0.7 [ ] ( S P r i n g - 8 と同程度 )
- ・ X 線のエネルギー密度：正確には分からないが、おそらく  $10^7 \sim 10^8$  [ 光子数 / 秒  $\text{mm}^2$  (  $\text{mrad}$  )<sup>2</sup> 0.1% バンド幅 ] 程度 ( S P r i n g - 8 の千分の一程度 )
- ・ センサの解像度：50 [  $\mu\text{m}$  ]  $\times$  50 [  $\mu\text{m}$  ] ( S P r i n g - 8 と同程度 )

10

そして比較例というのは、実施例 1 と同じ S P r i n g - 8 で測定したデータを解析するのに、従来技術 1 の手法を採用してみたものである。ここで、比較例に従来技術 1 を採用して従来技術 2 の手法による記載を省いた理由は、従来技術 2 ではあまりに計算時間が長くなってしまふからである。たとえば試料 ( e ) の場合でも従来技術 2 の手法を適用すると、計算時間に数日間以上を要してしまふ。

( 試料物質 )

また、実施例 1 および実施例 2 と比較例とは、前述のようにそれぞれ 6 種類の物質の構造決定を試みており、実施例 1 で決定された物質構造は、図 4 ( a ) ~ ( f ) に示されている。したがって実施例 1 とは、より正確には実施例 1 ( a ) ~ 実施例 1 ( f ) からなる一連の実施例シリーズであり、実施例 2 および比較例についても同様である。

20

ここで、図 4 ( a ) は、 $\text{Pd}(\text{tmdt})_2$  の分子内での原子の立体配置とその分子が作る結晶構造とを示す斜視図である。同様に、図 4 ( b ) は  $\text{Au}(\text{tmdt})_2$  の、図 4 ( c ) はシチジンの、図 4 ( d ) はタウリンの、図 4 ( e ) は 1 - F c A q H (  $\text{C}1\text{O}_4$  ) の、図 4 ( f ) はコハク酸プレドニゾロンの物質構造をそれぞれ示す斜視図である。

参考のため、これら 6 種類の試料物質を次に列挙する。

- ( a )  $\text{Pd}(\text{tmdt})_2$  : P d - t r i m e t h y l e n e t e t r a t h i a f u l v a l e n e d i t h i o l a t e
- ( b )  $\text{Au}(\text{tmdt})_2$  : A u - t r i m e t h y l e n e t e t r a t h i a f u l v a l e n e d i t h i o l a t e
- ( c ) シチジン：ピリミジンヌクレオシドの一種
- ( d ) タウリン：アミノエチルスルホン酸
- ( e ) 1 - F c A q (  $\text{C}1\text{O}_4$  ) : 1 - f e r r o c e n y l e t h y n y l - a n t h r a q u i n o n e ,  $\text{C}1\text{O}_4$
- ( f ) コハク酸プレドニゾロン：合成副腎皮質ホルモン剤の一種

30

( 比較検討 )

先ず、自由度 6 の試料物質 ( a ) ~ ( c ) については、決定すべき自由度の数が少ないので、比較例であってもデータ解析に数十秒程度の演算時間をかければ概ね構造決定をすることができ、同じデータを解析した実施例 1 に準ずる一致度を得ることができる。

しかし、同じデータを本実施形態の方法により解析した本発明の実施例 1 によれば、解析時間は数十分の一 ~ 百数十分の一で済み、試料物質三つのうち二つまでは一致度（すなわち適応度）も向上している。ここで、一致度は二乗誤差に基づく実測データと仮想データとのミスマッチの程度であるから、一致度がゼロに近づくほど評価は高い（つまり適応度が高い）。

40

したがって、計算時間と一致度との両方から見て、S P r i n g - 8 での高精度測定データに基づいて広域的最適値を探索する構造決定能力は、本発明の実施例 1 の方が従来技術による比較例よりも二桁程度は向上しているものと推定される。

一方、実施例 2 では、S P r i n g - 8 からではなく実験室レベルの X 線回折測定装置で測定をしているので、実測データの精度が低くなっている。それにもかかわらず、実施例 2 でもおおむね実施例 1 と同等の解析時間で構造決定をするに至っており、求められた

50

最適解も実施例1のそれに近い。このようにして、本実施形態によれば、実験室レベルの測定装置での測定データからであっても、世界最高レベルの精度をもつ測定データからと比肩しうる内容で物質構造決定ができるという効果が確認された。なお、一致度に関する比較は、実施例1と実施例2とでは互いに異なる測定データに基づいて物質構造の決定がなされているから、あまり意味をなさない。

次に、決定すべき自由度がそれぞれ8, 12, 26である試料物質(d)~(f)については、比較例ではSpring-8での測定データに基づいても構造決定をすることができなかった。しかし、本発明の実施例1によれば、自由度26の試料物質(f)に至るまで、構造決定することができる。

そして、本発明の実施例2(d)では、このタウリンの物質構造が、大学の実験室で測定したデータを解析して僅か0.4秒で構造決定するに至っている。

また、自由度12の試料物質(e)についての構造決定は、実施例1(e)で7秒程度、実施例2(e)では29秒程度でなされている。同物質の構造決定にかかった演算時間は、基になる測定データの精度が違うので、実施例2(e)では実施例1(e)の数倍になっているが、その絶対値は問題にならない程度の短さである。

ここで、現況では実施例1(d)が未実施となっている理由は、ひとえにタウリンの粉末試料についてSpring-8の利用時間(マシンタイム)を確保することができなかったせいである。しかし、タウリンの自由度は8に過ぎず、しかも前述のように、実施例2(d)では実験室レベルの装置による測定データによっても構造決定ができています。それゆえ、Spring-8の利用時間を割いてタウリンの測定データさえ取ることができれば、実施例1(d)でも実施例2(d)とおおむね同等の解析結果が得られるものと推察される。

また、実施例2(f)が未実施になっている理由も、本発明によるデータ解析に供するための実験データを用意することができなかったからである。つまり、実験室レベルのX線回折装置では、測定に一週間から十日間程度の時間が必要であるが、この実施例については、本出願までに同装置の利用時間の配分を受けることができなかった。

(表1との差異)

なお、実施例1(e)では12自由度でありながら7秒間で最適解に達しており、前述の表1中に記載がある逆ルーレット選択では自由度12~14で約1時間を要するという実施例よりも、二~三桁ほど解析時間が短くて済んでいる。これは何故かということ、その理由は大きく分けて次の二つである。

第一に、表1中の実施例では解析対象物質の一分子中の原子数が百個程度であったのに対し、実施例1(e)ではその数が半数以下に減っていて物質構造が簡素であるからである。すなわち、各個体に相当して仮定された物質構造が簡素であれば、その粉末試料が形成するであろうX線回折像を算出する演算負荷も小さくなるからである。

第二に、表1中の実施例では、本発明のプログラムを実行するに当たりHGAのポピュレーションの個体数を1000個としていたのに対し、実施例1(e)では、ポピュレーションの個体数も最適化されていて、わずか10~50個体程度に抑えているからである。すなわち実施例1(e)は、本発明のプログラムを実行するに当たって、ポピュレーション中の個体数など幾つかの設定パラメータを適正に設定した最速値だからである。

さらに、自由度26の試料物質(f)については、比較例では全く構造決定をすることができず、本発明の実施例1によってのみ約8時間の解析時間をかけて物質構造の決定を行うことができた。この解析時間は、コンピュータを一晩だけ放置すれば済む程度のものであるから、十分に実用に耐えるものである。

ここで、前述のように、Spring-8によらず実験室レベルの測定データに基づいてであっても、本発明の実施例2では自由度12の試料物質(e)までは短時間のうちに構造決定を行うことができ、本発明の有効性を実証することができた。

以上のように、本実施形態によれば、Spring-8による高精度の回折環データに基づいて試料物質の構造決定を行う場合、その決定すべき自由度が6程度までであれば、従来技術よりもずっと短時間のうちにデータ解析を終えることができた。また、試料物質

10

20

30

40

50

(d)以降のように自由度が8を超えてしまうと、従来技術では構造決定が困難になるが、本実施形態では可能である。そして、例えば試料物質(f)のように自由度が26であっても、Spring-8の測定データを数時間で解析してその物質構造を決定することができることを実証した。

なお、試料物質のうち(f)にあたるコハク酸ブレドロニゾンの構造は、本発明の実施例1によって初めて決定されたものである。

そして、本願発明者らを含む研究グループは、本発明の本実施形態によって決定された物質構造をすでに幾つかの学会に報告しており、参考までに学会発表および論文のうち主要なものを次に列挙する。

[学会発表]

- ・「中性単一分子金属Ni(tmdt)<sub>2</sub>とAu(tmdt)<sub>2</sub>のMEM電子密度分布」藤城雄一<sup>1</sup>、西堀英治<sup>1</sup>、高田昌樹<sup>1</sup>、坂田誠<sup>1</sup>、鈴木和佳子<sup>2</sup>、藤原絵美子<sup>2</sup>、小林昭子<sup>2</sup>、藤原秀紀<sup>3</sup>、小林速男<sup>3,4</sup>(名大院工<sup>1</sup>、東大院理<sup>2</sup>、分子研<sup>3</sup>、CREST<sup>4</sup>)日本結晶学会 平成15年年会 平成15年12月1~2日 熊本産業文化会館
- ・「ベンゼンを加えた大型フェロセン-アントラキノン共役結合錯体の結晶構造」吉田雅則、青柳忍、西堀英治、坂田誠、近藤美欧、村田昌樹、西原寛(名大院工、東大院理)日本結晶学会 2005年 平成17年12月6日、7日 イーグレひめじ

[論文]

- ・「放射光粉末X線回折法による精密結晶構造解析法の開発と展開」西堀英治, 日本結晶学会誌, Volume 48(3), 2006. p216-223
- ・Recent progress in development of single-component molecular metals. , Kobayashi A, Fujiwara E, Suzuki W, Sasa M, Fujishiro Y, Eiji Nishibori, Masaki Takata, Makoto Sakata, Okano Y, Fujiwara H, Kobayashi H., Journal De Physique IV 114(2004)419-424
- ・Structures of a single-component palladium complex with extended TTF-type dithiolate ligands, bis-(tetrathiafulvalenedithiolato)palladium determined by powder X-ray diffraction. , Suzuki, W., Fujiwara, E., Kobayashi, A., Fujishiro, Y., Eiji Nishibori., Masaki Takata., Makoto Sakata., Okano, Y., Kobayashi, H., Chemistry Letters 32(2003)1106-1107.

(従来技術1,2との比較)

図5のグラフに、本実施形態と従来技術1,2とで一致度Rの時間変化を比較して示す。すなわち同図は、回折環データに基づいて物質構造を決定するにあたり、各世代のポピュレーションのうち適応度が最も良い最優秀個体の一致度Rが収束していく様子を、前述の実施例2(e)に相当する本実施形態と従来技術1,2とで比較して示したグラフである。ここで、本実施形態と従来技術1,2とで共通して解析に供された回折環データは、1-FcAqH(Cl<sub>4</sub>O)の粉末試料を実験室レベルのX線回折装置にかけて測定したデータであり、前述の実施例2で使用したものである。

同図に明らかなように、従来技術1では、進化の方向性がGAに与えられていないので、最優秀個体の一致度Rがなかなか収束しようとはしない。逆に従来技術2では、全ての新生個体が局所最適化されるので、最優秀個体の一致度Rは演算の初期段階では急速に向上する。しかし、その際に個体の多様性も急速に失われてしまうらしく、ほんの数秒間から先では一致度Rの収束は遅々として進まなくなってしまう。

一方、本実施形態での一実施例(実施例2(e)に相当)は、進化の方向性が適宜与えられていながら、ポピュレーション中の個体の多様性も失われなくなっている。そ

10

20

30

40

50

れゆえ本実施例では、最優秀個体の一致度 R がほんの百秒間ほどで従来技術 2 のそれを追い越し、その後も収束を続ける。そして、千秒間（十数分間）程度が経過した頃には、一致度 R がおおむね収束しきっており、その段階での一致度は従来技術 1 や従来技術 2 よりもずっと改善されている。

### (3) 実施例で実施された解決手段

以上の実施例 1 および実施例 2 で実施された本願発明の解決手段を、なるべく前述の解決手段の項の記載に沿って表現すれば、以下のようなものである。

#### [ 方法発明 ]

##### ・パラメータ最適化方法

実施例 1 および実施例 2 は、いずれも本発明のうち「ハイブリッド GA による複数パラメータの最適化方法」の好ましい実施形態である。そう解釈すると、実施例 1 ( a ) ~ ( f ) および実施例 2 ( a ) ~ ( f ) のうち未実施の 2 ケースを除く 10 ケースの大半は、本発明の複数パラメータの最適化方法にあたる第 1 手段 ~ 第 15 手段のうち、第 2 手段および第 4 手段 ~ 第 6 手段を除くいずれにも該当する。

ここで、これら大半のケースが第 2 手段に該当しない理由は、これらのケースでは、HGA の設定が、世代交代の毎に各世代で一つの個体だけを局所最適化するようになってからである。すなわち、初期世代を除く全ての世代がいずれも特定世代であるようでは、第 2 手段のように特定世代は全世代のうち一部であるとは、言い難いからである。

さて、両実施例の最適化方法は、各個体を構成する記号列のうち少なくとも一部は、所定の物理量に対応した複数のパラメータであり、これらのパラメータの適正な組合せを探索することを目的としている。この方法は、前提として、評価過程と交叉過程と淘汰過程とを繰り返すうちに突然変異過程を適宜交えながら世代交代を重ねる GA ( 遺伝的アルゴリズム ) を使用する。

そして、この方法にはいくつかの特徴がある。

まず、第一の特徴は、初期世代を除く全世代で、各世代のポピュレーションのうちそれぞれ一個体ずつに局所最適化が施されることである。

すなわち、両実施例では、原則として特定世代の出現頻度は世代交代のある毎に ( つまり連続した各世代毎に ) であり、当該ポピュレーションのうち特定個体の数は一個だけということである。それゆえ両実施例の全ケースは、いずれも第 3 手段に該当している。ただし、実施例 1 の系統と実施例 2 の系統とでは、特定世代の分布に若干の違いがある。すなわち、実施例 1 の各ケースでは、初めから特定世代が連続しているが、一方、実施例 2 の各ケースでは、第 10 世代までは特定世代が出現せず、第 11 世代から特定世代が連続して出現している。それゆえ実施例 2 の各ケースは、第 2 手段および第 5 手段にも該当している。

このような第一の特徴は、特定世代の出現頻度と当該世代のポピュレーションでの特定個体の個数とに関する特徴であり、両者の積によって、一世代あたりに局所最適化操作を行う平均的な回数がおおむね決まる。そして、両者のバランスを適正に取りながら、両者の積を適度に低減することができれば、計算負荷の大きな局所最適化操作の回数を減らすことにより、高速で世代交代を重ねながら、多様性を保ちつつ進化に適度の方向性を与えることができる。

次に、この方法の第二の特徴は、最適化過程で局所最適化操作の対象となる特定個体を選ぶ際に、適応度が比較的低い個体の中からも特定個体を選ばれる可能性があるということである。

すなわち、両実施例のいずれのケースでも、特定世代のポピュレーションのうち適応度が低い方の所定割合 ( この割合もランダムに設定される ) に属する個体からなる劣等グループの中から、ランダム選択法によって特定個体を選ばれるようになっている。特定個体の選定に際してこのような方法を採用すると、実質的には適応度の低い個体の方が適応度が高い個体よりも高い確率で選ばれるような弱者優先の傾向を持つルーレット選択法によって特定個体を選ばれる。そして、このような特定個体の選択方法のことを、発明者らは前述のように「逆ルーレット選択」と呼んでいる。それゆえ、両実施例の全ケースは第 7 手段

10

20

30

40

50

～第11手段のいずれにも該当している。

そして、この方法の第三の特徴は、HGAの淘汰過程では、当該世代のポピュレーションのうち適応度が比較的低い個体の方が選択される確率が高くなる逆ルーレット戦略を取っていることである。

その結果、適応度が比較的低い個体であっても（むしろその方が）、次世代に生き残る可能性が高くなっている。ただし、せっかく広域的な最適値にまで達した個体が生き残れないのでは困るから、当該世代の前記ポピュレーションのうち適応度が最高のトップ個体だけは世代交代の際に必ず生き残るというトップ保存戦略も併用されている。それゆえ、両実施例の全ケースは第12手段～第14手段のいずれにも該当している。

・データ解析方法

次に両実施例は、本発明の「パターンマッチングによるデータ解析方法」の実施形態でもあり、前述の解決手段のうち第16手段に該当する。

すなわち両実施例ではさらに、前述の「ハイブリッドGAによる複数パラメータの最適化方法」において、次のような特徴を持つ。まず、各前記個体もつ前記適応度は、当該個体の前記記号列に含まれる前記複数のパラメータに対応する前記物理量を仮定すると数値演算により決まる仮想データのパターンが、実測データのパターンとどの程度一致しているかを示す指標である。そして、この適応度が高い前記個体を生成することによって、当該個体の前記仮想データと前記実測データとの間でパターンマッチングを行い、前記物理量を推定することを目的として前記実測データの解析を行う。

・構造決定方法

両実施例はまた、本発明の「放射線回折データに基づく物質構造の推定方法」の実施形態でもあり、前述の解決手段のうち第17手段～第20手段ならびに第21手段および第22手段に該当する。

すなわち、両実施例では前述の「ハイブリッドGAによる複数パラメータの最適化方法」を、所定の試料物質に照射された放射線が形成する回折パターンに基づいてこの試料物質の構造を推定する物質構造の推定方法に応用している。そして、前述の物理量は主に、この試料物質の物質構造を決定する構造パラメータである。また、各前記個体の前記適応度は、この試料物質が当該個体に相当するこれらの構造パラメータをもつと仮定すると生じる回折パターンが、実測された回折パターンに対してどの程度一致しているかを示す指標である。

ここで、両実施例で決定している構造パラメータは、試料物質の結晶構造を規定する格子定数と、この試料物質を構成する分子内における三次元的な原子配列を規定する分子構造パラメータと、この試料物質の結晶内での三次元的な原子配列を規定する結晶学的構造パラメータとである。また、その回折パターンとは、その回折強度が同心円状に拡がって分布するX線回折環である。さらに、適応度とは、この回折環がもつ半径方向への強度分布の実測データと、仮定された構造パラメータに基づいて生成される仮想データとが、どの程度一致しているかを示す評価関数である。

この物質構造の推定方法を、第21手段の記載に準拠して表現すると次のようになる。

すなわち、この方法は、多数の細かい結晶を含む粉末試料の物質構造を特定するために、この試料物質に照射されたX線回折環がもつ強度分布の回折データに基づいて、その構造パラメータをデータマッチングにより推定する方法である。その際に用いられるアルゴリズムは、前記構造パラメータがコーディングされた個体を、ある程度の多様性を持たせて複数個発生させ、初期世代のポピュレーションとする発生過程の後に、これらの個体の交叉および複製と、前記データマッチングの程度を示す適応度を評価してこれらの個体を淘汰する選択と、確率的に起こる突然変異とを繰り返す世代交代過程を行うGA（遺伝的アルゴリズム）を基礎としている。そして、前記GAでは、前記世代交代過程の間に複数の世代において、当該ポピュレーションの一部である前記個体に対して局所最適化操作が行われ、この局所最適化操作の対象には、当該ポピュレーションのうち前記適応度が比較的低い前記個体もなる可能性があることを特徴としている。

あるいは、この物質構造の推定方法を、第22手段の記載に準拠してプロセスに分けて

10

20

30

40

50

表現すれば次のようになる。

すなわち、両実施例の「放射線回折データに基づく物質構造の推定方法」は、放射線としてのX線を試料物質に照射して得られた回折環の強度分布データに基づき、この強度分布データに適合する物質構造を特定することを目的とする。そしてこの目的を果たすために、背景散乱除去プロセスと、基本構造決定プロセスと、空間群絞り込みプロセスと、基本構造決定プロセスと、構造精密化プロセスとが、順に行われる。

まず、格子定数決定プロセスでは、この試料物質に含まれる結晶の格子定数が決定され、次に背景散乱除去プロセスでは、この強度分布データに含まれているバックグラウンド散乱の影響が除去ないし低減される。そして、空間群絞り込みプロセスでは、この結晶が属する空間群の候補が絞り込まれ、そのうえで基本構造決定プロセスでは、この結晶の基本構造が初期的に決定されてこの空間群が一つに特定される。最後に、構造精密化プロセスでは、この基本構造が精密化されてこの結晶を構成する物質の構造がより精密に決定される。

この格子定数決定プロセスおよびこの基本構造決定プロセスでは、それぞれGA（遺伝的アルゴリズム）に局所最適化操作を組み合わせたハイブリッドGAを利用した物質構造決定が行われる。このハイブリッドGAでは、世代交代の間に複数の特定世代で、当該世代のポピュレーションのうち一部の個体である特定個体を対象にして前記局所最適化操作が行われ、この特定個体には、当該世代のポピュレーションのうちでは実測データとの一致度を示す適応度が比較的低い劣等グループからも選定される可能性がある。

[ソフトウェア関連発明]

・最適パラメータ探索プログラム

両実施例では、「最適パラメータ探索プログラム」が使用されており、このプログラムは本発明の第23手段～第25手段の一実施形態である。すなわち両実施例では、X線回折環のデータ解析に同プログラムが実施されており、同プログラムの広域探索機能により求められた最適パラメータとして、物質構造が決定されている。

・物質構造推定プログラム

同様に両実施例では、「物質構造推定プログラム」が使用されているとも捉えることができ、そうすると同プログラムは第26手段および第27手段の一実施形態にも該当する。

・記録媒体

両実施例で使用されている前述の「最適パラメータ探索プログラム」および「放射線回折データに基づく物質構造推定プログラム」のうち少なくとも一方をコンピュータ読み取り可能な記録媒体に記録したとしよう。すると、その記録媒体は、本発明の第28手段の一実施形態に該当する。なお、具体的な記録媒体の種類としては、発明者らは主として広く市販されているCD-ROMやUSBメモリを使用している。

[装置発明]

・HGA搭載のコンピュータ

両実施例のデータ解析にあたっては、前述のように市販PCの上級機ないしワークステーションで上記プログラムを実行している。このPCは、3[GFLOPS]程度の演算能力をもつもので、より具体的には、HPCシステムズ株式会社（固有組織名：旧株式会社HITが他社と合併して社名変更した）製の型式名Silent-SCC（商品名）である。そしてその構成は、PCクラスタ（20CPUおよび8CPU）と呼ばれるものである。なお、このようなPCクラスタを単にPCやワークステーションと呼んでしまうことに抵抗のある向きもあるが、その議論は本発明の本質とは関わりがないので、単に強力な市販PCということでお許しをいただきたい。

そして、このPCには前述の「最適パラメータ探索プログラム」がインストールされており、このプログラムは「放射線回折データに基づく物質構造推定プログラム」でもある。それゆえ、このPCは本発明の第29手段「ハイブリッドGA搭載コンピュータ」の一実施形態に該当する。

・HGA搭載の解析装置

さらにこのPCを解析装置として捉えれば、このPCはやはり前述の「最適パラメータ探索プログラム」ないし「放射線回折データに基づく物質構造推定プログラム」を実行可能な解析装置である。それゆえこのPCは、第30手段たる「ハイブリッドGAを搭載した解析装置」の一実施形態にも該当する。

・物質構造推定装置

両実施例の物質構造決定結果を得るに当たって用いた設備全体は、当該試料物質に放射線を照射して生じる回折パターンを測定する測定装置と、この回折パターンから得られた実測データに基づいた解析を行うことにより当該試料物質の構造を推定する解析装置とを有する。ここで、この測定装置として使用されているのは、実施例1ではSpring-8であり、実施例2では前述の実験室レベルのX線回折装置である。一方、解析装置としては、両実施例で共通して前述のPCを使用している。

10

それゆえ、両実施例のうちいずれで使用した「放射線回折パターンによる物質構造推定装置」も、本発明の第31手段の一実施形態に該当する。

3. 物質構造決定の手順

(0) 全体像

本発明の実施形態たる「放射線回折データに基づく物質構造の推定方法」は、図7に示す解析手順(プロセスP1~P5)を踏んで、粉末試料の物質構造を十分な確度および精度をもって推定(すなわち決定)する方法である。本実施形態は、格子定数決定プロセスP1と、背景散乱除去プロセスP2と、空間群絞り込みプロセスP3と、基本構造決定プロセスP4と、構造精密化プロセスP5とを有する。

20

第一に、格子定数決定プロセスP1は、この粉末試料に含まれる多数の微細な結晶の格子定数を決定するプロセスである。すなわち、このプロセスでは、当該結晶の単位格子を定義する格子定数、例えば三斜晶系の結晶と特定されればその格子定数( $a$ ,  $b$ ,  $c$ ,  $\alpha$ ,  $\beta$ ,  $\gamma$ )が決定される。

第二に、背景散乱除去プロセスP2は、本実施形態の入力データINたるX線回折の強度分布データに含まれているバックグラウンド散乱の影響を、除去ないし低減するプロセスである。

第三に、空間群絞り込みプロセスP3は、この粉末試料に含まれる微細な結晶が属する可能性がある空間群を、ある程度少ない候補に絞り込むプロセスである。このプロセスでは、入力データINたる回折強度のピークないし反射の出現位置に基づき、結晶が取りうる230種類の空間群のうちから可能性の低いものを除き、可能性が高い空間群を数種類程度に絞り込むことが行われる。

30

第四に、基本構造決定プロセスP4は、この結晶の基本構造を初期的に決定して、この空間群を一つに特定するプロセスである。すなわち、このプロセスでは、前述の空間群絞り込みプロセスで絞り込まれた空間群の候補の中から、最も確からしい空間群が決定され、単位格子内での分子の位置及び姿勢と三次元的な原子配列とが、やや粗めの精度で決定される。

第五に、構造精密化プロセスP5は、この基本構造を精密化して、この結晶を構成する物質の構造をより精密に決定するプロセスである。さらに、この構造精密化プロセスでは、分子内での電子密度の分布をも推定ないし決定することがある。

40

さて本実施形態は、粉末試料をガラスキャピラリーに封入し、Spring-8(固有名詞)の高エネルギー放射光X線をこの試料物質に照射して実測された回折環の強度分布データを入力データINとする。入力データINは、再び図1(b)に示すように、粉末試料で回折したX線が形成する回折環がもつ強度分布の実測データであるから、図8に示すように、回折角 $2\theta$ に沿って一次元的に展開した回折X線の強度分布データである。そして、本実施形態では所定のプロセスP1~P5を順に経て、入力データINに適合する物質構造が特定され、出力結果OUTとして出力される。出力結果OUTは、図9に示すように、粉末試料に含まれる多数の微細な結晶の格子定数と、その微細な結晶を構成する物質の分子が当該結晶内にもつ三次元的な原子配列とである。

そして本実施形態では、格子定数決定プロセスP1と基本構造決定プロセスP4との両

50

方で、GA（遺伝的アルゴリズム）に局所最適化操作を組み合わせたハイブリッドGAを利用した試料物質の構造決定が行われている。

このハイブリッドGAでは、GAの世代交代過程の間に、複数の特定世代において、当該世代のポピュレーションのうち一部の個体である特定個体を対象にして、前記局所最適化操作が行われている。また、この特定個体には、当該世代のポピュレーションの中で、実測データとの一致度を示す適応度が比較的低い劣等グループからも選定される可能性がある。

前述のように、本実施形態の「放射線回折データに基づく物質構造の推定方法」では、格子定数決定プロセスP1と基本構造決定プロセスP4との両方で、GAに局所最適化操作を組み合わせたハイブリッドGAを利用した試料物質の構造決定が行われている。そして、これらの両プロセスP1, P5自体も、本願発明の中核をなすハイブリッドGAを用いて処理されていることを条件に、本発明の「放射線回折データに基づく物質構造の推定方法」の一実施形態である。

10

以上で、本実施形態のうち「放射線回折データに基づく物質構造の推定方法」についてその手順構成の概要を明らかにすることができたので、次に各プロセスP1～P5について具体的に説明する。

#### (1) 格子定数決定プロセス

格子定数決定プロセスP1は、前述のように、格子定数を決定するプロセスであり、例えば三斜晶系の結晶であれば、図10に示すように、単位格子UCの各辺長 $a$ ,  $b$ ,  $c$ と各辺がなす角度 $\alpha$ ,  $\beta$ ,  $\gamma$ との六つのパラメータからなる格子定数LCが決定される。

20

その手順は、図11に示すように、先ず読込ステップDで、Spring-8（固有名詞）で粉末試料から得たX線回折環の実測データをコンピュータ（図略）に読み込んでから、前処理ステップS0でピーク位置を抽出して回折ピーク位置表を作成する。ただし、実施例2(a)～(f)では、前述のように実験室レベルのX線回折装置で測定した回折環データに基づいたピーク位置表を用いている。

ここで、回折ピーク位置表とは、回折環の強度分布データが図12に示すようなものである場合には、その各ピークに回折角 $2\theta$ の小さい方から順に番号を振ったものである。すなわち、次の表3に示すように各ピークの回折角 $2\theta$ を列挙したピークデータの一覧表が、回折ピーク位置表である。

【表 3】

番号	$2\theta_i [^\circ]$
1	5.352514
2	5.788506
3	6.141745
4	7.544511
5	8.8211
6	9.05712
7	9.324903
8	9.452379
9	9.982389
10	10.17055
11	10.54705
12	10.68341
13	11.6108
14	12.30847
15	13.30555
16	13.40638
17	14.08942
18	14.17574
19	14.60032
20	14.82903
21	15.13145
22	15.39832
23	16.99598
24	17.33527

10

20

そして、再び図 1 1 に示すように、ハイブリッド G A ステップ S 1 で、そのピークデータに対するデータマッチングを行って最適な（つまり正解の）格子定数を探索し、格子定数（ $a, b, c, \dots$ ）を決定するに至る。なお、ハイブリッド G A ステップ S 1 で行われる格子定数決定の具体的な処理内容については、後ほど「ハイブリッド G A」の節を立てて詳しく説明する。

30

なお、このプロセスで用いるピークデータに含まれているのは、ピーク位置に関する情報だけであって、ピーク強度に関する情報は含まれていない。

なぜならば、回折線の各ピークの位置だけの情報から、X 線の回折が起こる条件式（ $2d \cdot \sin \theta = \lambda$ ）に基づいて、各ピークの面間隔  $d$  の大きさを求めることができるからである。このとき、各ピークの面間隔  $d$  は、面を示す単位格子の軸の整数分の 1 のベクトルの組み合わせであらわされるミラー指数（ $hkl$ ）と格子定数（ $a, b, c, \dots$ ）とを用いて、所定の関係式から算出される。すなわち、実測データに含まれる全ての（場合によってはほとんどの）ピーク位置に対応する  $d$  を実現できる格子定数（ $a, b, c, \dots$ ）の組合せを探り出すことができれば、それをもって正しい格子定数であると決定することができる。このプロセスでは、この格子定数の探索にハイブリッド G A を使用している。

40

## （2）背景散乱除去プロセス

背景散乱除去プロセス P 2 は、前述のように、X 線回折の強度分布データに含まれている背景（バックグラウンド）散乱の影響を除去ないし低減するプロセスである。

すなわち、図 1 3 に模式的に示すように、回折環の X 線強度の実測データには、本当に観測したい粉末試料自体からの回折 X 線の他に、様々な原因で生じるバックグラウンド散乱による X 線が混じっている。そして、構造決定を行う試料からの回折強度を正確に見積

50

もるためにはバックグラウンドの強度を取り除く必要がある。最近の粉末X線回折においては、必要とされる粉末試料がごく微量であることと、実験の簡便さから頻繁にデバイシエラー法によるデータ測定が行われることから、粉末試料をガラス製のキャピラリーに封入して実験を行うため、ガラスのハローパターンのためにバックグラウンドが複雑な形状になりがちである。

さて、このようなバックグラウンドを構成する散乱の因子のうち、主たるものは次の四つである。

- ・ X線の発生源から粉末試料を経て検出器に至るまでに存在する空気による散乱
- ・ ガラスキャピラリーおよび粉末試料から発生する蛍光X線の等方的な散乱
- ・ ガラスキャピラリーからのハローパターンを作る散乱
- ・ 粉末試料からの温度散漫散乱

10

たとえば、X線回折試料がCsPbCl<sub>3</sub>の粉末からなる場合には、図14(a)に示すようなX線強度の実測データが得られるが、この実測データにもこれら四つの因子による散乱が含まれている。すなわち、ガラスのハローパターンは図14(b)に示すように、空気散乱は図14(c)に示すように、温度散漫散乱は図14(d)に示すように、それぞれ分布している。一方、蛍光X線は、等方的に散乱することが予め分かっているから、その分布パターンも明らかである。そこで、図14(e)に示すように、これら四つの因子によるバックグラウンド散乱を取り除けば、ほぼ純粋に粉末試料からの回折X線だけの強度分布を割り出すことができる。

### (3) 空間群絞り込みプロセス

20

空間群絞り込みプロセスP3は、前述のように、X線回折環の実測データに基づいて、結晶が取りうる230種類の空間群のうちから可能性の低いものを除き、可能性が高い空間群の候補を絞り込むプロセスである。

このプロセスでは、先ほど構造精密化プロセスP5で求められている格子定数(a, b, c, , )から、粉末試料に含まれる結晶が七つの結晶系(三斜晶、単斜晶、斜方晶、正方晶、三方晶、六方晶、立方晶)のうちどれに属するかを決定する。なぜならば、X線回折により構造を決定するためには、結晶系から空間群を決定する必要があるからである。

ここで、空間群とは、結晶の原子配列の対称性を記述した群であり、全ての結晶は230個の空間群のどれかに属している。ちなみに各結晶系に含まれる空間群の数は、三斜晶で2個、単斜晶で13個、斜方晶で59個、正方晶で68個、三方晶で25個、六方晶で27個、立方晶で36個である。これらのうち同一の結晶系に属する空間群の違いは、X線回折データで見ると、消滅する反射を決定する消滅則にあり、消滅則とは、結晶系からは出現するはずの反射が、空間群の対称操作で周期的に消滅する規則を示す。例えばb軸方向に軸の1/2の大きさの並進操作がある場合には、ミラー指数(hkl)が0k0で表される反射のうち、k=2n+1(奇数)の反射は消滅する。

30

発明者らは、空間群を粉末回折データから、結晶系に含まれる全ての空間群に対してLe Bail解析(A. Le Bail et al, Mat. Res. Bull. 23, 447, (1988))を適用し、Le Bail解析の信頼度因子の値R<sub>wp</sub>から可能性のある空間群を絞り込む方法を開発した。ここで、信頼度因子R<sub>wp</sub>は次の数1で定義される数値である。

40

【数1】

$$R_{wp} = \left[ \frac{\sum_i^N w_i [Y(2\theta_i)_{obs} - Y(2\theta_i)_{calc}]^2}{\sum_i^N w_i Y(2\theta_i)_{obs}^2} \right]^{1/2}$$

50

この数 1 中で、 $N$  は観測点の数、 $w_i$  は第  $i$  番目の観測点に対する荷重であり、観測強度の逆数  $w_i = Y(2\theta_i)_{obs}^{-1}$  を用いる。また、 $Y(2\theta_i)_{calc}$  は計算プロファイル強度、 $Y(2\theta_i)_{obs}$  は観測プロファイル強度である。この方法では、最初に各結晶系で消滅側を持たない空間群により Le Bail 解析を行う。この解析から回折反射の形状を決める回折プロファイルパラメータを求める。その後、このプロファイルパラメータを用いて、可能性のある空間群全てに対する Le Bail 解析を自動で行う。

そして、Le Bail 解析の結果として得られた信頼度因子  $Rwp$  の数値が低い空間群が、すなわち現実の状態に該当する可能性の高い空間群であって、正しい空間群の候補となる。(そういう意味では、 $Rwp$  には「信頼度因子」という呼称よりも、むしろ逆に「不信度因子」ないし「過誤確率指数」といった呼称の方が相応しい。)

10

ここで、空間群を決定するうえで従来取られていた方法では、複数の空間群に対して一つずつ消滅側を確認しようとする、多大な手間と時間を必要としていた。なぜならば、実測データと反射位置とから消滅側を決定し、空間群を絞り込んでいこうとすると、粉末回折データは単結晶回折データと比較して回折強度が弱い、目視できない反射が実際に消滅しているか判明することが難しかったからである。その結果、従来法では、強度データに反射が見られなければ、正解であっても候補から外れてしまうことがあり、正解の空間群が絞り込んだ空間群から漏れる可能性があった。

一方、発明者らが開発した方法では、前述のように、Le Bail 解析空間群を変化させながら自動で繰り返すので、結晶系から対象となる全ての空間群を網羅的に探索することができる。その結果、可能性のある全ての空間群を、正解の候補として漏れなく抽出できるようになった。

20

その実例として、1 - FcAqH(ClO<sub>4</sub>) つまり「1 - フェロセンアントラキノン接合錯体プロトン付加体(ClO<sub>4</sub>)」の解析例を次の表 4 に示す。

【表 4】

	Rwp[%]		Rwp[%]
P222	○ 3.93	Pnnn(1)	23.37
P222 <sub>1</sub>	○ 3.89	Pnnn(2)	23.35
P2 <sub>1</sub> 2 <sub>1</sub> 2	○ 3.97	Pccm	32.24
P2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	○ 3.90	Pban(1)	24.69
C222 <sub>1</sub>	34.72	Pban(2)	24.68
C222	34.73	Pmma	9.12
F222	40.27	Pnna	21.37
I222	23.84	Pmna	21.21
I2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	23.84	Pcca	32.23
Pmm2	○ 4.26	Pbam	23.55
Pmc2 <sub>1</sub>	7.90	Pccn	33.41
Pcc2	32.23	Pbcm	32.15
Pma2	○ 4.42	Pnnm	8.34
Pca2 <sub>1</sub>	23.65	Pmmn(1)	8.48
Pnc2	8.39	Pmmn(2)	8.43
Pmn2 <sub>1</sub>	7.86	Pbcn	31.79
Pba2	23.59	Pbca	31.96
Pna2 <sub>1</sub>	○ 4.82	Pnma	9.51
Pnn2	8.34	Cmcm	40.31
Cmm2	34.76	Cmca	41.55
Cmc2 <sub>1</sub>	40.31	Cmmm	34.78
Ccc2	39.62	Cccm	39.63
Amm2	25.06	Cmma	36.20
Abm2	41.55	Ccca(1)	40.89
Ama2	25.08	Ccca(2)	40.90
Aba2	36.45	Fmmm	40.28
Fmm2	40.36	Fddd(1)	38.44
Fdd2	38.43	Fddd(2)	38.43
Imm2	23.85	Immm	23.86
Iba2	31.61	Ibam	31.61
Ima2	23.56	Ibca	33.04
Pmmm	○ 4.12	Imma	25.84

10

20

30

40

この表に示すように、単斜晶の空間群に対して軸の及び原点の取替えを含めて行った 384 例に試したところ、可能性のある空間群をわずか数分（実際には 1 分間程度）で 8 種類にまで絞り込むことができた。ここで、可能性のある空間群の条件としては、信頼度因子  $Rwp$  が 5% 未満のものであることを採用し、同表ではそのパーセンテージの前に印を付けて示してある。なお、この際に X 線回折環のデータ解析に用いたコンピュータは、前述の市販 PC である。

50

#### (4) 基本構造決定プロセス

基本構造決定プロセスP4は、前述のように、絞り込まれている空間群の候補の中から最も確からしい空間群を正解として決定し、単位格子内での分子の位置及び姿勢と三次元的な原子配列とを大まかに決定するプロセスである。すなわち、結晶系、空間群が決定されたうえで、単位格子内の大まかな原子または分子の配列を決定することを基本構造の決定という。X線回折データの回折ピークの強度は、結晶内の原子の配列を反映しているため、回折データの強度関係を満足する大まかな原子配列を決定することで、基本構造の決定となるわけである。

さて、一般に単結晶X線回折の場合には、各回折ピークの強度を独立に求めることが容易であるため、この強度を用いて1985年にノーベル賞に輝いた直接的に結晶構造を決定する方法により基本構造が決定できる。しかし、粉末試料の回折データの場合には、前述のように回折環の強度ピークが重なり合ってしまうので、各回折ピークの強度を求めることができない。このため、従来の直接的に結晶構造を決定する方法では、本実施形態のように粉末試料から得られたX線回折環データに基づいて構造決定を行うことはできない。

10

そこで、本実施形態の基本構造決定プロセスP4では、単位格子の実空間中における分子の位置姿勢およびその形状を求める実空間法にハイブリッドGAを適用して、粉末回折データから基本構造を決定することができるようにした。

ここで用いた「実空間法」とは、原子もしくは幾つかの原子を集めた剛体（つまり分子）が単位格子中で取る位置および向き（つまり姿勢）を決定し、その剛体（分子）の内部にある回転自由度について解を与えて剛体の形状をも決定する方法である。すなわち、図15(a)~(c)に示すように、単位格子中の原子の配列を原子、剛体（分子）の位置 $(x, y, z)$ および向きないし姿勢 $(\theta_1, \theta_2, \theta_3, \dots)$ と、内部回転 $(\phi_1, \phi_2, \phi_3, \dots)$ のパラメータで表すものとする。そして、基本構造決定プロセスP4では、これらのパラメータをハイブリッドGAによる実空間法で決定する。ふつう、決定すべきパラメータの個数は、分子が単位格子内で取る位置姿勢の6自由度分と内部回転の箇所数との和ということになる。

20

基本構造決定プロセスP4で使用するハイブリッドGAは、前述の構造精密化プロセスP5で使用するものと基本的には同じものであり、最適化すべきパラメータと適応度の定義と各種の設定などが異なるだけである。発明者らが開発したハイブリッドGAの詳細については、後に節を立てて説明する。

30

#### (5) 構造精密化プロセス

前述のように、本実施形態での構造精密化プロセスP5は、前述の基本構造決定プロセスP4で決定された基本構造の精度を増して精密化するプロセスである。このプロセスでは、さらに分子内での電子密度の分布をも決定することも、やろうと思えばできる。

ここで、構造精密化とは、基本構造決定で求められた構造をもとに個々の原子の位置を最小自乗最適化により求めることである。粉末回折の構造最適化の手法としては、Rietveld法(H. M. Rietveld, J. Appl. Cryst. 2 (1969), 65)が用いられており、ふつう原子の位置が0.001nm程度の精度で求められる。発明者らはさらに、基本構造決定とRietveld法の精密化の流れを円滑にし、最終の構造精密化には、Rietveld法だけでなく、マキシマムエントロピー法(M. Sakata et al, Acta Cryst. A 46, 263, (1990))を組み合わせ構造の精度を高める方法を開発した。その結果、粉末回折データから決定した物質構造と回折環データとの一致度は98%程度にも達し、原子配列の誤差も0.0001nm程度と、極めて精密な構造決定が可能になった。そして、これだけの精度をもって物質構造を決定することができれば、新物質の機能予測や医薬品の開発にも十分に利用可能である。

40

すなわち、構造精密化プロセスP5では、前述の基本構造決定プロセスと同様に、分子を剛体として取り扱い、基本構造決定と同様の位置・向き・内部回転のパラメータを精密化する。また、図16に示すように、剛体全体の大きさを示す三軸方向のサイズパラメー

50

タ ( $x$ -size,  $y$ -size,  $z$ -size) をも最適化すべきパラメータに採り入れて、最小二乗最適化により精密化する。そしてさらに、有限温度における結晶中の原子は熱振動を起こしており、このような熱振動の影響は回折パターンに比較的大きく寄与するから、剛体全体の熱振動をも精密化した。その結果、分子を剛体として取り扱ったままで、回折パターンとの一致度を基本構造決定の場合よりも高めることができる。

構造精密化プロセス P5 の結果として、Pd ( $dt$ )<sub>2</sub> の決定された構造による X 線回折の実測データとの一致度を示すフィッティンググラフを図 17 に示し、同様に Au ( $tmdt$ )<sub>2</sub> の一致度を示すフィッティンググラフを図 18 に示す。ちなみに、実測データとの一致度としての信頼度因子 Rwp の値は、Pd ( $dt$ )<sub>2</sub> では 3.6% であり、Au ( $tmdt$ )<sub>2</sub> では 1.1% であり、どちらも原子間の距離、角度まで含めた高精度な解析がなされていることが分かる。

さらに、Pd ( $dt$ )<sub>2</sub> の構造からバンド計算により電子状態を計算したところ、この物質が半導体であることが導き出されたので、その予測を確認すべく実際にその電気伝導度を測定したところ、やはり半導体的であった。このことから、このようにして精密化した構造は、物質の機能を予測するのに十分な精度を持つことが分かった。

なお、Pd ( $dt$ )<sub>2</sub> の正式名称は Pd-tetrathiafulvalenedithiolato であり、本実施例によるこの物質の構造決定の結果については、前述の文献のうち後者で報告されている。読者の便宜のために、同文献を重ねて次に記載しておく。 Structures of a single-component palladium complex with extended TTF-type dithiolate ligands, bis-(tetrathiafulvalenedithiolato)palladium determined by powder X-ray diffraction. , Suzuki, W., Fujiwara, E., Kobayashi, A., Fujishiro, Y., Eiji Nishibori, Masaki Takata, Makoto Sakata, Okano, Y., Kobayashi, H., Chemistry Letters 32 (2003) 1106 - 1107.

#### 4. 格子定数決定プロセスでの HGA

##### (1) HGA の概要

いよいよ、以上の各プロセス P1 ~ P5 のうち、格子定数決定プロセス P1 と基本構造決定プロセス P4 とで使用されるハイブリッド G A (HGA と略記することもある) の構成について説明する。なお、両プロセス P1, P4 は、本願発明のうち「ハイブリッド G A による複数パラメータの最適化方法」及び「パターンマッチングによるデータ解析方法」並びに「最適パラメータ探索プログラム」及び「物質構造推定プログラム」の一実施形態でもある。

先ずこの節では、格子定数決定プロセス P1 で用いられる HGA について、具体的に説明する。

すなわち、本実施形態の「ハイブリッド G A による複数パラメータの最適化方法」及び「最適パラメータ探索プログラム」等の中核をなすハイブリッド G A は、図 19 に示すように、初期世代設定ステップ S11 の後、交叉ステップ S103, 淘汰ステップ S105, 突然変異ステップ S107, 局所最適化ステップ 109 などを繰り返す世代交代ルーブルとを有する。なお、各処理手順については、本願発明を方法発明として捉えた場合には「過程」と呼び、プログラムとして捉えた場合には「ステップ」と呼んでいるが、処理する内容は実質的に同じであるから、以後、過程とステップとの呼称の区別には拘らないことにする。

この HGA では、各個体を構成する記号列は一組の物理量パラメータである格子定数 [例えば ( $a, b, c, \dots$ )] であり、格子定数の適正な組合せを広域的に探索することが目的である。

先ず、初期世代設定ステップ S11 では、格子定数がそれぞれ適度に分布した所定数の個体からなる初期世代 (第 1 世代ともいう) のポピュレーションが生成される。そして、

各世代において、当該世代のポピュレーションを記録しておく記録ステップS 1 0 2と、評価ステップ（図略）と交叉ステップS 1 0 3と淘汰ステップS 1 0 5と、確率的に起こる突然変異ステップS 1 0 7及び局所最適化ステップS 1 0 9とを含む世代交代ループLが構成されている。

ここで、評価ステップ（図略）は、当該世代のポピュレーションのうち新たに生成された前記個体である新生個体の適応度を評価する過程である。評価ステップでは、初期世代設定ステップS 1 1と交叉ステップS 1 0 3と突然変異ステップS 1 0 7とで生成された新生個体の全てについて、各々の適応度が算出される。また、交叉ステップS 1 0 3は、このポピュレーションのうち複数の個体を親個体として選び、これら親個体の間で所定の仕方で交叉を行って、新しい子個体を生成する過程である。そして、淘汰ステップS 1 0 5は、交叉ステップS 1 0 3による新生個体と、複製ステップS 1 0 4で第N世代のポピュレーションから複製された個体との中から、次世代に残すべき個体を選択して他の個体を淘汰する過程である。

10

そして、突然変異ステップS 1 0 7は、それに先立つ判定ステップS 1 0 6により所定の確率で実行され、やはり確率的に選ばれた個体を構成する記号列に対し、強制的な変更を加える過程である。

さらに、局所最適化ステップS 1 0 9は、やはりこれに先立つ判定ステップS 1 0 8によって、確率的なルールに則って選ばれた一部の世代（特定世代）において、やはり所定の確率的なルールで選ばれた個体（特定個体）に対してだけ、その適応度に関して格子定数のパラメータを局所的に最適化する過程である。判定ステップS 1 0 6の判定ロジックは調整可能であるが、特定世代の出現頻度は例えば10世代毎に一回程度であり、特定個体の指定は、その特定世代のポピュレーションを構成する所定数の個体のうちで一つか、せいぜい数個程度に抑えている。

20

最後に、収束判定ステップS 1 10で解の収束判定をした結果、収束していなければ指定の制限回数の範囲で、インクリメントステップS 1 11で世代番号を改めた後、ステップS 1 0 2へ戻り、第N世代のポピュレーションとして、再び世代交代ループLが繰り返される。こうして世代交代ループLを幾度も繰り返しつつ、世代交代を重ねるうちにポピュレーションの中に適応度が十分に良好な個体が発生することを待ち、それが正解ないし最良解に収束する至った判定をもって格子定数の最適な組合せが探索される。

そして、収束判定ステップS 1 10で最良の適応度をもつ個体が生成されたことが判定されれば（あるいは世代交代が制限値に達したら）、世代交代ループLを終え、結果出力ステップS 1 2で、最適の個体をもつ決定された構造パラメータ式とその適応度等とが出力される。併せて、最終世代のポピュレーションを構成する個体全部の情報も、記録ステップS 1 0 2と同様に記録される。

30

## （2）HGAの詳細

本実施形態のHGA、すなわち格子定数決定プロセスP1でのハイブリッドGAは、以上のようにして格子定数（ $a$ 、 $b$ 、 $c$ 、 $\dots$ ）の組み合わせを決定するが、その詳細を以下に記す。

### （位置ずれパラメータ $t$ ）

本実施形態のHGA（図19参照）において、初期世代設定ステップS 1 1、交叉ステップS 1 0 3、突然変異ステップS 1 0 7及び局所最適化ステップS 1 0 9には、新たに生成された個体の適応度を算出する評価ステップ（図略）が内包されている。この評価ステップで、適応度算出のために基準として参照される実測データは、前述のように、測定された粉末回折データからピーク位置を回折角 $2\theta$ で求めたピークデータである（図11、図12、表3参照）。

40

ここで、元になる粉末X線回折データは、回折角 $2\theta$ について等間隔もしくは非等間隔ステップを持った離散点での強度データとして得られる。それゆえ、その回折角 $2\theta$ の測定値には、原点（回折環の中心点）付近でX線検出器がもつ空間分解能の範囲で、微細な原点の位置ずれ（ $t$ ）が生じており、この原点ずれ（ $t$ ）の値は、X線検出器からの実測データからだけでは不明である。

50

そして、本実施形態のHGAで最適解を求めて探索する複数パラメータには、格子定数 ( a , b , c , , , ) の他にも、この位置ずれ t が加えられており、探索パラメータは ( a , b , c , , , , t ) の七つである。すなわち、位置ずれ t を含んだこれら七つのパラメータを探索し、位置ずれ t をも正しく決定してこそ、初めてピークデータとの正確なデータマッチングが行えるようになる。

(初期世代の発生過程)

先ず初期世代設定ステップS11(図19参照)では、初期世代のポピュレーションを形成する第一世代個体は、最初に使用者が設定した個体数Np組の格子定数 ( a , b , c , , , ) と原点ずれ ( t ) とが、適正な乱数で発生させられる。

その際、格子定数 ( a , b , c , , , ) には、必ず平行六面体を形成しなければならないことと、その単位格子の軸 a , b , c の大きさの最小値は、原子の大きさから 0.25 nm よりも小さいことなどの制約条件が課せられている。ここで、平行六面体の条件は、  $\alpha + \beta < 360^\circ$  ,  $\alpha + \gamma < 360^\circ$  ,  $\beta + \gamma < 360^\circ$  である。この条件を満たさない場合、例えば  $\alpha + \beta = 360^\circ$  といった場合には、単一の平面状に軸 a , b , c の方向が入ってしまい、単位格子が平面に収まってしまって平行六面体ではなくなってしまう。また、格子は原子のサイズより大きいという制限から、 a , b , c > 0.25 nm の範囲で発生させる。これらの制限により、乱数により必ず平行六面体を現す格子定数 ( a , b , c , , , ) の組を発生させることができる。

一方、原点ずれ ( t ) に関しては、データ点のステップ幅(つまり空間解像度)に相当するずれの範囲を設定し、その範囲でパラメータ t をランダムに発生させる。典型的な原点ずれ ( t ) の値は 0.001° のオーダーである。さらに、 ( a , b , c , , , ) に関しては、発生させるパラメータに、以下に示す多数の制約条件を付加することができる。

さらに、探索範囲を制限して効率的に解を探索するために、パラメータ ( a , b , c , , , ) の各値に上限、下限を設定することができる。また、必要に応じて、 a > b > c , a > c , b > c , c > a 等の制約も与えることができる。そして、構成原子や分子からおよその単位格子の体積が予測できる場合には、体積の上限値 ( Vmax ) と下限値 ( Vmin ) を入力することにより、上限 ( Vmax ) と下限 ( Vmin ) の範囲内の体積 ( V ) をもつ平行六面体を形成する格子定数 ( a , b , c , , , ) の組を発生させることができる。

(評価過程)

各個体には、このような格子定数 ( a , b , c , , , ) と原点ずれ ( t ) とがその記号列にコーディングされており、各個体は、評価ステップ(図略)でその格子定数が実験で得られた回折ピークを満足するかどうかで評価される。

すなわち、適応度の評価は、その個体が表す格子定数の平行六面体から次の数2により得られる d を回折条件式 (  $2d \cdot \sin \theta = \lambda$  ) に代入して計算される回折ピーク位置の仮定データ (  $2\theta_{calc}$  ) と、実験データから抽出した回折ピーク位置の実測データ (  $2\theta_{obs}$  ) との一致度を用いて行う。

【数2】

$$\frac{1}{d_{hkl}^2} = \frac{\frac{h^2}{a^2} \sin^2 \alpha + \frac{k^2}{b^2} \sin^2 \beta + \frac{l^2}{c^2} \sin^2 \gamma + \frac{2hk}{ab} (\cos \alpha \cos \beta - \cos \gamma) + \frac{2kl}{bc} (\cos \beta \cos \gamma - \cos \alpha) + \frac{2lh}{ca} (\cos \gamma \cos \alpha - \cos \beta)}{1 - \cos^2 \alpha - \cos^2 \beta - \cos^2 \gamma + 2 \cos \alpha \cos \beta \cos \gamma}$$

なお、一致度を示す評価関数は、次の数3から算出される。

10

20

30

40

【数3】

$$R = \frac{\sum_i^N \frac{|2\theta_{obs} - 2\theta_{ca1}|}{2\theta_{obs}}}{N}$$

ここで、Nは入力した反射の本数である。そして、 $2\theta_{obs}$ を中心とした $\pm \Delta$ 値の範囲を設定し、その範囲内に含まれる $2\theta_{ca1}$ との間で一致度Rを計算する。 $\Delta$ の値は、おおむね回折ピークの半値幅の半分から半値幅の全体まで程度の値とする。この数式を用いるにあたり、もしも $2\theta_{obs} \pm \Delta$ の範囲に $2\theta_{ca1}$ が存在しない場合には、上記数3の分子( $|2\theta_{obs} - 2\theta_{ca1}| / 2\theta_{obs}$ )の値を1とする。逆に範囲内に存在する場合には、( $|2\theta_{obs} - 2\theta_{ca1}| / 2\theta_{obs}$ )の値は0.005前後の値のため、1となるピークが存在した場合には、一致度Rの値は著しく大きくなる。

10

さて、説明が前後するが、前述の初期世代設定ステップS11では、生成される個体の表すパラメータの範囲に、一致度Rの値によって制限を加えることもできる。すなわち、一致度Rが設定した値より大きいパラメータに相当する個体が生成された場合には、その個体を破棄して、改めて別のパラメータに相当する個体を発生させる。そして、使用者が設定した各種の制約条件を満足するNp個の個体が得られたら、第一世代人口の発生は終了する。

20

(交叉過程)

交叉とは、生物の交配を模した操作である。交叉ステップS103で行われる交叉の操作では、以下の数4から求められたi番目の個体の適応度を用いて、交叉を行う親個体(格子定数パラメータ及び原点ずれの組)を選択する。

【数4】

$$\rho(i) = 1 - \frac{R - R_{min}}{R_{max} - R_{min}}$$

30

ここで、 $R_{max}$ 、 $R_{min}$ は、それぞれ第N世代のポピュレーションに属する個体の中で、前述の数3により算出される評価関数Rの最大値と最小値とである。

そして、交叉ステップS103では、親個体の選択にはルーレット選択を用いる。ルーレット選択とは、特定世代のポピュレーションからランダムに選ばれた個体がもつ適応度が、0~1の範囲で生成される一様乱数の値より大きければ、親個体として採用するという選択法である。逆に、各個体の適応度の値が上記一様乱数よりも小さければ、他の個体を選んで同じ操作を繰り返し、選択された親個体が所定数になるまでこのような操作が続けられる。

40

さて、交叉ステップS103で行う交叉方法には幾つかの種類があり、交叉のために選択される親個体の数は交叉方法によって異なる。すなわち、パラメータを交換するタイプの1点交叉タイプ、2点交叉タイプ及び一様交叉タイプの三つのうちいずれかと、実数値のGAで使用されるブレンドタイプの交叉とのうち、どれかを行うことができる。この際、親個体の選択数は、1点・2点・一様交叉タイプの方式で交叉を行う場合には、2個体1組なので、使用者が設定した $2Nm$ 個でNm組となが、一方、ブレンド交叉タイプの場合には、2個~Nb個となる。

交叉ステップS103では、有機物の分子性結晶の良く見られる単斜晶系と三斜晶系とに対するパラメータ探索効率を改善するために、特殊な工夫がなされている。すなわち、格子定数パラメータ(a, b, c, , )をそのまま交叉させるのではなく、図20に示すように、パラメータの性質に応じた遺伝を考慮して、パラメータのグループ化を

50

行っている。

たとえば、三斜晶系のグループ化では、平行六面体の1つの面を保存する。すなわち、単位格子のうち互いに並行ではない三つの面は、それぞれ2つの軸と一つの角度であらわすことができるから、この3つのパラメータを ( a , b , ) , ( b , c , ) , ( a , c , ) というふうに分ける。あるいは、単斜晶系であれば、四つ一組の格子定数 ( a , b , c , ) を、 ( b , c ) , ( a , b ) , ( a , c , ) というふうに分ける。

このようなグループ化の結果、親個体の平行六面体の3種類の面のうち1種類は子に遺伝することになる。

逆に、もしもこのようなグループ化を行わないと、どうなるかを思考実験してみれば、このグループ化のもつ意義が理解できる。たとえば、粉末試料に含まれる物質が三斜晶系であれば、( a , b , c , , ) の6パラメータのうち、図21(a)~(b)に示すように、, , を抽出してこれらの値を変更してしまうと、元の親個体とは全く形状の異なった平行六面体に相当する子個体となってしまう。そして、この平行六面体から計算される回折ピーク位置は、親個体のそれとは全く異なってしまい、たいへん不都合である。

10

なぜならば、一般的に言ってGA(遺伝的アルゴリズム)では、親個体の遺伝子を子個体に伝えることによって解の探索性能を向上させているからである。したがって、この例のように、形状も回折ピーク位置も、親個体とは全く異なる平行六面体に相当する子個体が生成されてしまえば、GAがもつ長所である探索効率の良さがすっかり損なわれてしまう。格子定数決定プロセスP1では、このような理由で交叉過程ではパラメータのグループ化が行われている。

20

したがって、交叉方式が1点・2点・一様タイプのうちいずれかである場合には、交叉ではグループ化されたパラメータのうち一つのグループは保存され、それ以外グループに属するパラメータだけが、両親(2つの親個体)の間で交換される。このようなグループ化が交叉過程で行われた場合には、三斜晶系の場合に可能な交叉の仕方は、図22(a)~(c)に示す三つのパターンに限られ、単斜晶系の場合にも同様に、図23(a)~(c)に示す三つのパターンに限られる。このような交叉が行われた結果、Nm組の親個体から2Nm個の子個体が生成される。

30

一方、交叉の仕方にブレンド交叉タイプを選んだ場合には、交叉はグループ化したパラメータを保存し、それ以外のパラメータを2~Nb個の個体の平均値周りに一様乱数/正規乱数で発生させることにより行う。すなわち、子個体のパラメータPcは、次の数5に基づいて、親個体のパラメータPpと乱数とから算出される。

【数5】

$$P_c = \frac{\sum_{i=1}^N P_{p_i}}{N_b} + \frac{\left| \frac{\sum_{i=1}^N P_{p_i}}{N_b} - P_{p_j} \right|}{N_b} \times (1 - \alpha) \times (\text{乱数})$$

40

さて、ブレンド方式の交叉過程では、上記数5の操作が2Nm回繰り返され、2Nm個の子個体が生成される。

そして、交叉の方式に拘わらず、交叉過程で生成された全ての個体は、前述の一致度を示す評価関数R(数3参照)により評価される(評価ステップが行われる)。

(淘汰過程)

突然変異ステップS107では、当該第N世代のNp個のパラメータと交叉過程により新たに生成された2Nm個の子個体とを合計したNp+2Nm個の個体から、次世代に残すNp個の個体を選択する。その際の選択戦略としては、大きく分けて以下の2種類があ

50

る。

先ず一つめの選択戦略は、各個体の適応度を計算し、適応度の高い順に  $N_p$  個を選択するエリート戦略である。

一方、もう一つの選択戦略は、第  $N$  世代の  $N_p$  個のうち  $(N_p - 2)$  個の個体を残し、残りの 2 個を、交叉で生成された  $2N_m$  個と交叉の親として選択した個体のうちの 2 個体とを合わせた  $2N_m + 2$  個体の中から選択する。その際、1 個体を  $(2N_m + 2)$  個の個体の中で適応度が最も高い個体（最良個体）とし、もう 1 個体は最良個体を除いた  $2N_m + 1$  個体からルーレット選択により選択するという選択戦略である。

以上の二つの選択法のうちどちらを使用するかについては、本実施形態の「最適パラメータ探索プログラム」を使用する使用者が選択する。

（突然変異過程）

突然変異過程とは、生物の進化の過程で起こる突然変異を模した操作を GA に取り入れたものである。

本実施形態における格子定数決定プロセス P1 では、 $N_p$  個の個体の中から、一様乱数により  $N_x$  個の個体を変位対象として選択し、それらの個体に対して突然変異の操作を行う。選択された個体の格子定数と原点のずれパラメータを乱数で変化させる。

その際に、格子定数決定プロセス P1 で前述したように、パラメータの遺伝を考慮して、三斜晶系では  $(a, b, c, \quad, \quad)$  のうち 1 つのパラメータ、単斜晶系では  $(a, b, c, \quad)$  のうち 1 つのパラメータを選択し、その値を乱数と置き換える。また、原点のずれパラメータ  $(t)$  は、常に乱数と置き換える。

そして、このような操作を  $N_x$  個の個体について繰り返し、突然変異によって作成された個体は、当該世代のポピュレーションのうち変異前の個体と置換される。したがって、この操作では世代の個体数  $N_p$  は変化しない。

なお、突然変異の対象個体の個数  $N_x$  の設定は、使用者が予め行っておく。

（複製過程）

複製ステップ S104 では、第  $N$  世代のポピュレーションについて、その  $N_p$  個の個体もつ全てのパラメータと評価関数値  $R$  とを保存し、次の淘汰ステップ S105（つまり淘汰過程ないし選択過程）の操作に送る。

（局所最適化過程）

局所最適化ステップ S109 は、7 つのパラメータ  $(a, b, c, \quad, \quad, t)$  の値を非線形制約条件付きの最小自乗法により局所的に最適化する処理ステップである。すなわち、 $N_p$  個体のうち  $N_n$  個の個体を選択し、適応度について局所最適化を行うのだが、これに伴う制約条件は、前述の第一世代発生時の制約に従う。

この際、局所最適化操作の対象とする  $N_n$  個の個体を選択する方法が特徴的であって、ルーレット選択により、常識とは逆に適応度の低い個体から選択する。その理由には二つあって、うち一つは、前述のように、既に適応度が高い個体は多峰性の多次元パラメータ空間中の落とし穴に陥っている可能性が小さくないことであり、今ひとつは適応度の低い個体の方が、以前に局所最適化されていない可能性が高く、局所最適化の効果が期待できることである。

もう一つ、局所最適化を行う個数  $N_n$  には限定があって、ポピュレーションに含まれる個体数の  $1/100$  以下とする。なぜならば、一回の局所最適化操作にかかる計算時間は、1 回の評価関数計算と比較して、ふつう  $10 \sim 100$  倍程度かかるので、それ以上に局所最適化の対象とする個体の数  $N_n$  を増やしてしまうと、演算負荷が軽いという GA 本来の良さが損なわれてしまうからである。

局所最適化操作によって新たに作成された個体は、元の個体と入れ替られる。したがって、この操作でも当該世代のポピュレーションに含まれる個体数  $N_p$  は変化しない。

なお、局所最適化操作を行う特定世代の頻度と、局所最適化操作の対象とする特定個体の個数  $N_n$  とは、使用者が予め設定しておく。

（収束判定）

ハイブリッド GA の収束判定は、適応度としての評価関数の値が、使用者により入力さ

10

20

30

40

50

れた値を下回った場合に収束したものの判定とする。ただし、もし何時までも収束に至らなかった場合には、使用者が予め設定しておいた世代交代のサイクル回数で計算は終了し、その旨のメッセージが出力される。

## 5. 基本構造決定プロセスでのHGA

### (1) HGAの概要

次にこの節では、基本構造決定プロセスP4で用いられるHGAについて、具体的に説明する。

すなわち、本実施形態の「ハイブリッドGAによる複数パラメータの最適化方法」は、再び図19に示すように、大筋においては前述の格子定数決定プロセスP1で使用されるHGAのプログラムを利用するものである。

ただし、大きく違う点としては、正解となる組合せを探索すべき複数パラメータは、前述のように(図15(a)~(c)参照)、分子の位置(x, y, z)および姿勢( , , )ならびに当該分子内部の回転角(  $\theta_i$ )である。また、各個体の適応度  $F$  には、前述の数4中のRをRwpに置き換えた式をもって充てる。一方、細部の違いについては、次の項で明らかにする。

### (2) HGAの詳細

本項では、再び図19を参照しつつ、基本構造決定プロセスP4で用いられるHGAについて、具体的に説明する。

#### (初期世代生成過程)

先ず初期世代設定ステップS11では、初期世代のポピュレーションが、予め使用者が設定した個体数Np個について、その記号列に相当する構造パラメータ式を乱数を用いて発生させる。そして、発生された構造パラメータから個体数Np個相当の構造モデルが作成されると、作成された構造モデルと、格子定数、空間群及び先に求められたプロファイルパラメータとを用いて、計算回折パターンが計算される。この計算回折パターンは、実験で得られた回折パターンとの信頼度因子Rwp(数1参照)により評価される。すなわち、初期世代設定ステップS11では、生成された個体数Np個分の構造パラメータとそれらに対する信頼度因子Rwpとが計算される。

#### (交叉過程および評価過程)

交叉ステップS103では、前述のように、数4の右辺のうち一致度Rを信頼度因子Rwpに置き換えた式から求められた適応度  $F$  を用いて、交叉を行う親個体を選択する。ここで、Rmax, Rminはそれぞれ、N世代の個体の中で評価関数の最大と最小の値である。

親個体の選択戦略には、前述の格子定数決定プロセスP1と同様のルーレット選択法を用いている。一方、交叉の方式は、前述の1点交叉と一様交叉と、実数値のGAで使用されるブレンドタイプの交叉とのうちから、使用者によって予め指定されている。これらの交叉方式のそれぞれについては、格子定数決定プロセスP1の項で説明したものと同様である。

#### (淘汰過程、突然変異過程、局所最適化過程および収束判定)

淘汰ステップS105、淘汰ステップS105、局所最適化ステップS109および収束判定ステップS110は、それぞれ前述の格子定数決定プロセスP1の各ステップと同様のやり方で処理を行うステップである。

## 6. その他の実施例

本発明のその他の実施例として、本実施形態の「放射線回折データに基づく物質構造の推定方法」の格子定数決定プロセスP1によって、Pd(tmdt)<sub>2</sub>の粉末試料からその格子定数(a, b, c,  $\beta$ ,  $\gamma$ ,  $\delta$ )を決定した。その結果、次の表5に例示するように、局所最適化がない普通のGAによる格子定数決定方法に比べて、本発明の中核となるHGAを用いた格子定数決定プロセスP1では、格子定数決定の成功確率とそれに要した平均サイクル数(世代交代数)とにおいて、顕著な効果が確認された。

【表 5】

試料名	Nn=100(10%)の場合		Nn=0場合 (最小自乗なし)		DICVOL 成否	晶系	A(A)	B(A)	C(A)	$\alpha$ (degree)	$\beta$ (degree)	$\gamma$ (degree)
	成功確率 (%)	平均収束 cycle	成功確率 (%)	平均収束 cycle								
Pd(tmdt) <sub>2</sub>	100	732.8	100	1799.5	否	三斜晶 三斜晶	12.08904	8.53774	6.37215	122.64774	83.73503	95.20243

10

20

30

40

この表のうち左端近くの項目に「Nn = 100 (10%) の場合」とあるは、本実施例による解析結果を示す。ただし、本実施例では前述の逆ルーレット選択はしておらず、特定世代のポピュレーションから特定個体を選択するには、ランダム方式で選択している。

本実施例での解析は、HGAでのサイクル数の制限値 = 5000回、ポピュレーションの個体数  $N_p = 1000$ 、交叉の数  $N_m = 500$ 、突然変異の数  $N_x = 500$  (つまり十

50

世代毎に一回の突然変異)、特定世代の出現頻度は世代交代毎に各世代で、局所最適化数(つまり特定個体の数) $N_n = 100$ (すなわちポピュレーションのうち10%に局所最適化をかける)の条件下で行った。このようなHGAによる格子定数の決定用プログラムにおいては、その成否が初期乱数、DELTAの値に依存するために、その値を変化させてそれぞれの実測データに対して10例の解析を行った。

その右隣に「 $N_n = 0$  場合(最小二乗なし)」とあるは、上記設定のHGAから局所最適化操作をなくしてしまった単純なGAではどうなるか計算してみた比較例である。GAによるこの比較例では、前述のHGAによる実施例よりも二倍半ほど多いサイクル数(世代交代回数)で最適値への収束を見た。

さらにその右隣に「DICVOL成否」とあるは、物質構造決定用の定番とも言えるフリープログラムによる解析結果を、先ほどとは別の比較例として提示したものである。すなわち、三斜晶でも適用可能なLouer & Louer(固有人名)により開発されて公開されているプログラムDICVOL04(固有のプログラム名)を用いて同じデータを解析した結果、格子定数を決定することはできなかった。なお、このフリープログラムによる回折データの解析方法は、いわゆる一定体積内での絨毯爆撃方式である。

その結果、本実施形態での格子定数決定プロセスP1によれば、全ての場合で解析に成功しているだけでなく、 $Pd(tmdt)_2$ の例のように、既存のDICVOLプログラムでは得られない場合にも解が得られている。また、この二つの例では、パラメータのグループ化を行わない場合には、HGAでも解が得られなかった。さらに、局所最適化操作を行わなかった場合には、成功確率が落ちていたり収束に要したサイクル数(つまり世代交代回数)が倍以上に増えていたりする。

このような比較から、本実施例の「物質構造推定プログラム」によれば、格子定数決定プロセスP1においても、只のGAや定番のフリープログラムに比べて明瞭に良好な計算結果が得られていることが分かる。

#### 【発明の効果】

本発明の「放射線回折データに基づく物質構造の推定方法」や「物質構造推定プログラム」等によれば、単結晶を作ることが難しい物質であっても、その粉末試料にX線を照射して得た回折環の強度分布データから、その物質構造を決定することができるという効果がある。その結果、この方法によって医薬や素材などの新物質の構造が速やかに特定できるようになり、新物質の開発が飛躍的に加速されるので、各種産業や医療に関する研究開発に大きな福音がもたらされるものと期待される。

また、本発明の「ハイブリッドGAによる複数パラメータの最適化方法」や「パターンマッチングによるデータ解析方法」及び「最適パラメータ探索プログラム」等によれば、多峰性の多次元空間に於いて極めて強力な広域探索能力と精密な最適化能力とが発揮される。すなわち、本発明の各種プログラムやその搭載コンピュータ等によれば、実測データに対するパターンマッチングによって、従来よりも多数のパラメータをより精密かつ効率良く最適化または特定ないし推定することができるようになる。その結果、実測データに含まれる情報量が少なすぎたり、推定すべきパラメータの数が多すぎたり、複数パラメータの最適化に要求される精度が高すぎたりして、従来は実際上できなかった最適値決定やデータ解析に基づく判定までが可能になるという効果がある。

#### 【産業上の利用可能性】

#### 【0007】

##### (新物質の研究開発)

本発明の「放射線回折データに基づく物質構造の推定方法」等は、前述のように放射線回折環から物質構造を決定する能力が大幅に向上しているため、物質設計、医薬品開発、ナノテクノロジー、液晶技術などの多くの産業分野で利用可能である。しかしながら、本願発明の中核的な技術的特徴をもつ「ハイブリッドGA」は、以下に例示するように、さらに広い産業分野で利用可能である。

##### (データ解析)

すなわち、本発明の「ハイブリッドGAによる複数パラメータの最適化方法」や「最適

10

20

30

40

50

パラメータ探索プログラム」および「パターンマッチングによるデータ解析方法」等は、前述のように実測データを解析する技術だけではなく、逆に最適設計を行う設計技術にも利用が可能である。

先ず、データマッチングによる実測データの解析技術としては、前述の実施形態のように、多数の物理量を精密に推定ないし決定することが代表的である。そして、その応用分野は、ソナーの音響解析、レーダーの反射像解析、地震波の伝搬による地殻構造解析や鉱脈探査など、極めて広いものと考えられる。

(最適設計)

逆に、これらの本発明を逆問題を解くレゾルバとして利用し、極めて多数の設計パラメータについて最適ないし適正な組合せを探索する設計技術(設計パラメータの最適化)に適用することも可能である。このような最適設計の技術分野としては、例えば航空機の翼型設計および主翼を含む機体の構造設計や、多段ロケットの最適設計問題、通信ネットワークの最適設計、打上ロケットや人工衛星などの宇宙機の軌道設計、空港騒音問題に考慮した航空機の最適経路問題などがある。

ここで、逆問題やそのレゾルバとはどういうものであるか具体的なイメージが湧くように、翼型設計を例にとって説明しよう。翼型設計における逆問題とは、所与の形状の翼型が二次元流中に置かれた場合に周囲の空気流を求めてその翼型の特性を求めるという順方向の問題ではなく、逆に希望する空力特性を発揮するためには翼型の形状をどう設計すれば良いかという問題である。これを解くためには、先ず適当な形状の翼型を初期形状として与えておき、その翼型の周囲の流れ場を算出したうえで、望みの空力特性に近づくように翼型の形状を変えていくという試行錯誤を繰り返して、徐々に翼型の形状を修正していく。そして、このような演算を行って解を求める演算手段のことを、逆問題のレゾルバという。

こうして、最終的に最も望ましい形状の翼型が得られるまで、何度でも翼型の形状を修正しては改めてその空力特性を求めするので、その演算量は膨大になる。それゆえ、このような逆問題においては、問題の設定が適当でなかったりそのレゾルバの能力が十分でないと、要求を満たす解としての翼型は現実的なコストや時間では求まらない。そこで、本発明の「最適パラメータ探索プログラム」をこのような逆問題のレゾルバに利用すれば、レゾルバの大幅な能力向上が見込めるので、従来は現実的な範囲で解けなかった逆問題を解くことができるようになる可能性がある。その結果、本発明の技術によって、最適設計の技術分野にも発展をもたらすことができるであろう。

(最適経営問題など)

さらに、本発明の中核をなすハイブリッドGAは、物理量に対応するか否かに拘わらず多数のパラメータを最適化に調整するうえで、計算負荷を劇的に低減することができるという効果がある。

それゆえ、現象のモデル化や適応度設定など克服すべき難しさは多いが、調整すべきパラメータが極めて多い最適経営問題にも、本発明の「最適パラメータ探索プログラム」などを適用することができる可能性がある。なお、ここでいう最適経営問題とは、たとえば会社が持つ各種の経営資源を如何に配分すると良いかとか、あるいは国家の各種資源をどう配分したら良いかといった一種の最適設計問題である。

(人工知能)

また、本発明の技術によれば、同様の理由で人工知能(AI)の一分野としてのGAが発揮する処理速度を加速し、その収束時間ないし学習時間を短縮することができるから、人工知能としてのGAにも新たな実用範囲が開ける可能性が生じる。

たとえば、幾つかの惑星をかすめて探査機を飛行(フライバイまたはスウィングバイ)させ、その惑星の重力を利用して軌道エネルギーを稼ぎ、併せて軌道要素を変えながら、次々に幾つもの惑星やその衛星の観測するという宇宙探査ミッションがある。しかし、その軌道設計には極めて多くの試行錯誤を要し、研究者に非人間的な努力を強いる場合が少なくない。そこで、GAを使った人工知能によって最適パラメータ探索を行い、自動的に軌道設計させようというアイデアは既にあるが、本発明によればその探索時間を節約し

10

20

30

40

50

てミッション設計を容易にすることができる可能性がある。特に、惑星の衛星をも次々に観測していくミッションでは、衛星重力の影響をも受けるうえに、さらに制約条件が多くなって軌道設計の困難が予想されるので、本発明を利用して初めて設計可能になるミッションもあるものと想像される。

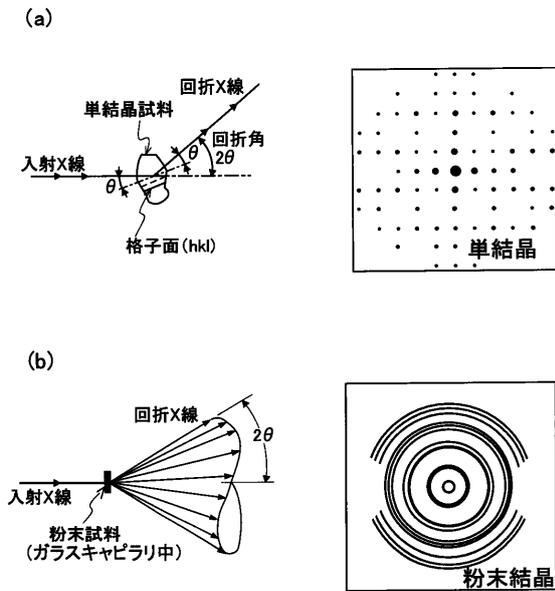
あるいは、理論的にはGAで解くことが可能でありながら、事象を表現するモデルが複雑になりパラメータが増えるなどして、その計算時間の制約から実用にならなかった問題に対しては、本発明のHGAが有効な解決手段になる可能性がある。このような利用分野としては、例えば天気予報や株価予測などがある。すなわち、一週間後のことを予測するのに一ヶ月間の計算時間がかかっている用をなさないが、1時間で解けるようになってしまえば有用だという場合があるからである。

(結論)

本発明のハイブリッドGA技術を応用すれば、多峰性の強い広大なパラメータ空間内でも強力な広域探索能力を発揮することができるので、従来技術1や従来技術2などでは解けなかった問題をも解くことができるようになる。その結果、本発明は前述の実施形態で詳述した粉末試料のX線回折による物質構造の決定ばかりではなく、極めて広い産業分野でも利用可能であり、様々な分野で強力な最適値探索ツールとなりうる。

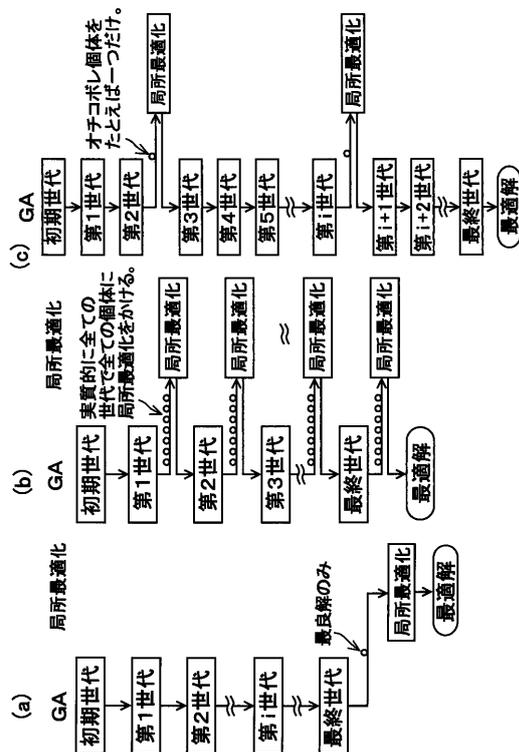
【図1】

図1

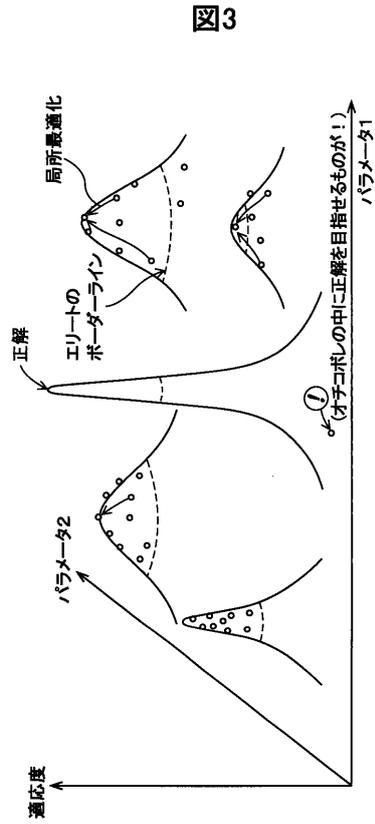


【図2】

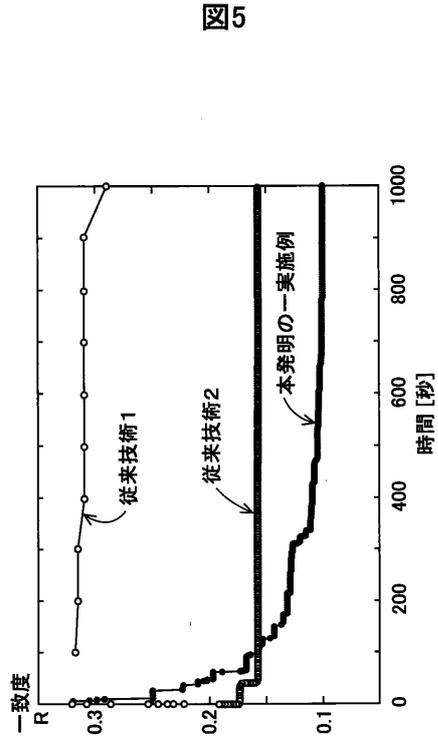
図2



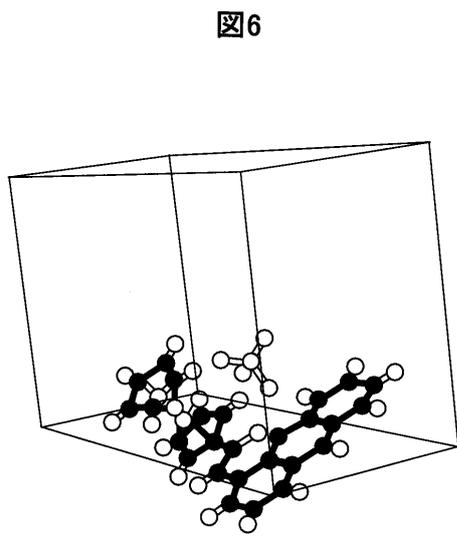
【 図 3 】



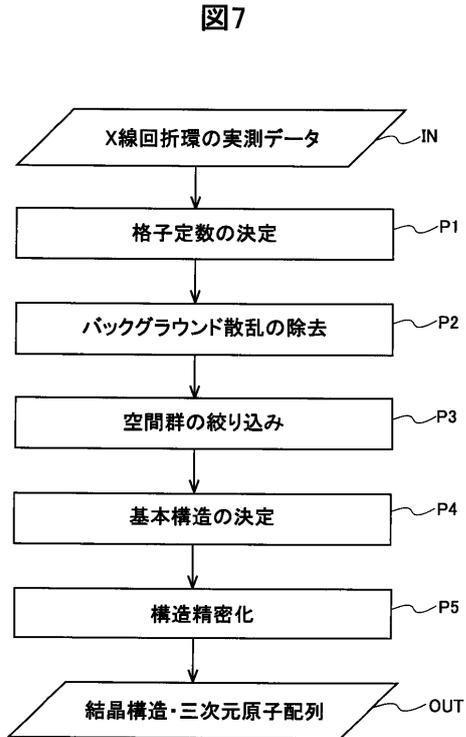
【 図 5 】



【 図 6 】

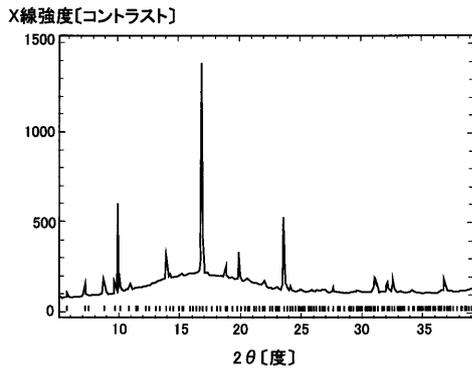


【 図 7 】



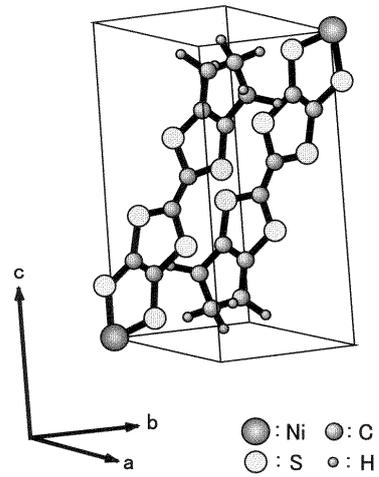
【 図 8 】

図8



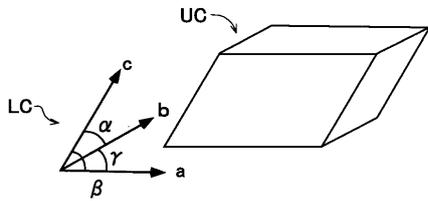
【 図 9 】

図9



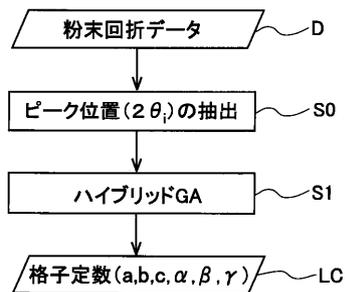
【 図 1 0 】

図10



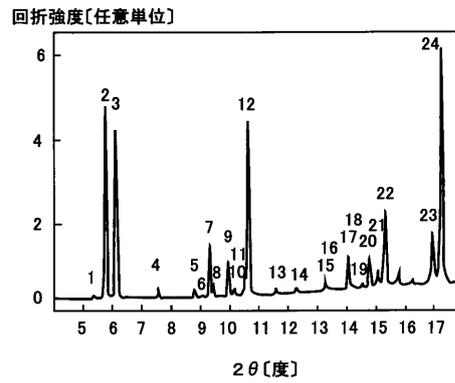
【 図 1 1 】

図11



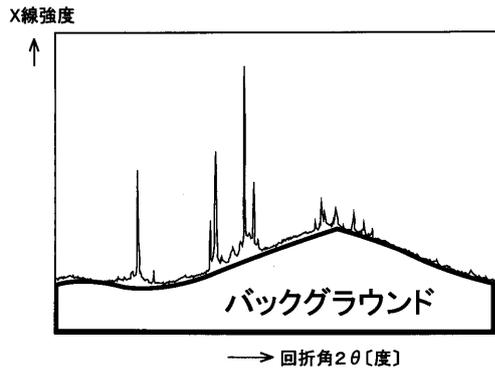
【 図 1 2 】

図12



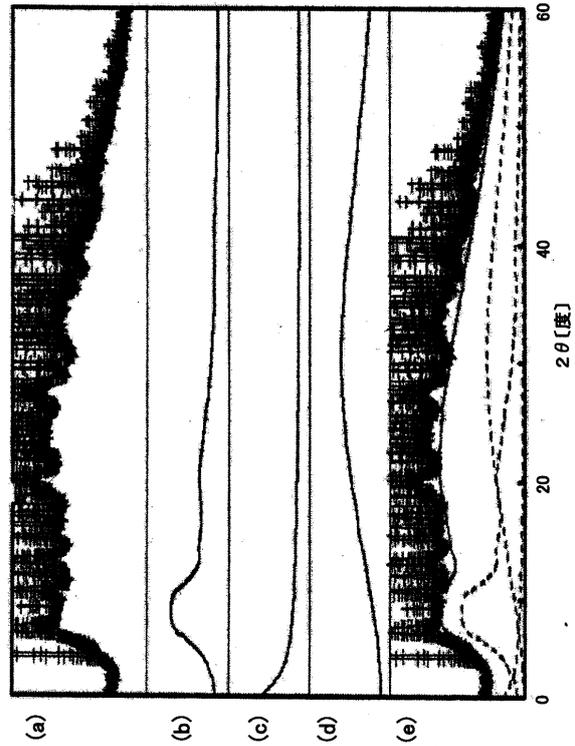
【 図 1 3 】

図13



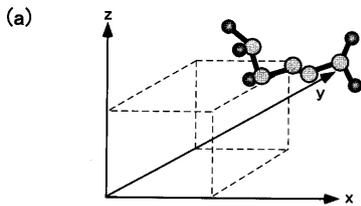
【 図 1 4 】

図14



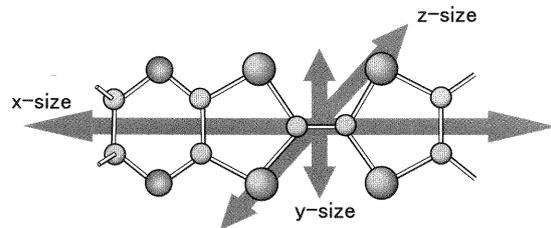
【 図 1 5 】

図15

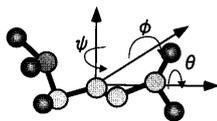


【 図 1 6 】

図16



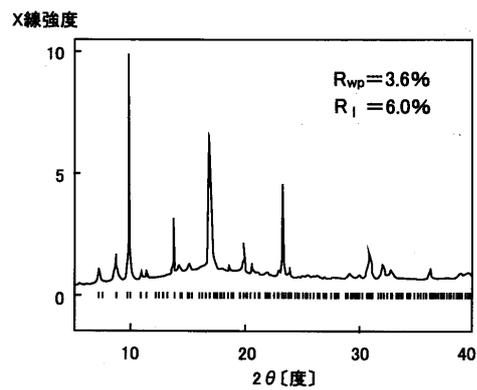
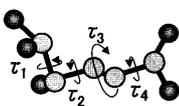
(b)



【 図 1 7 】

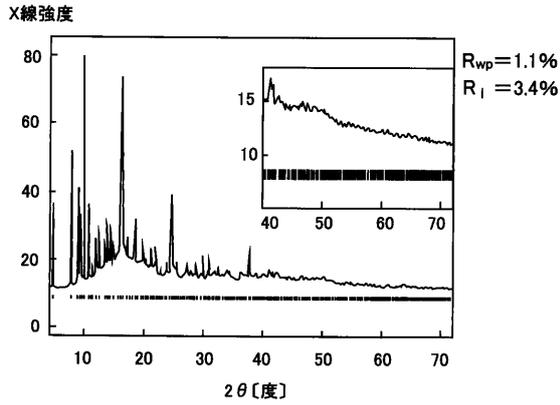
図17

(c)



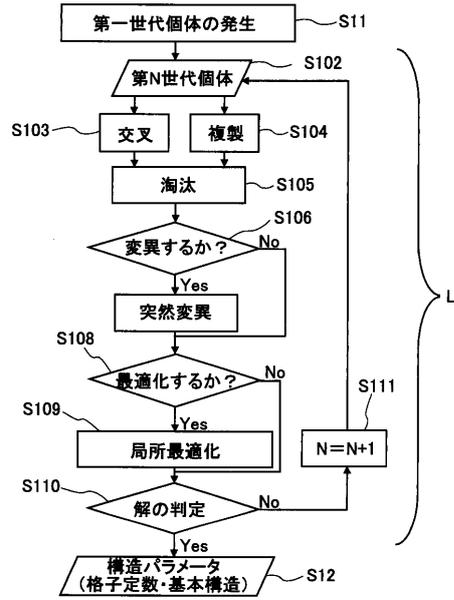
【 図 1 8 】

図18



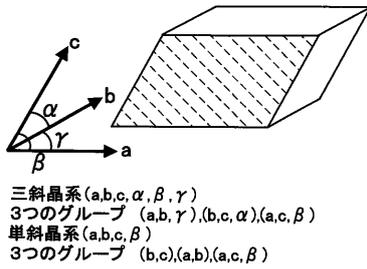
【 図 1 9 】

図19



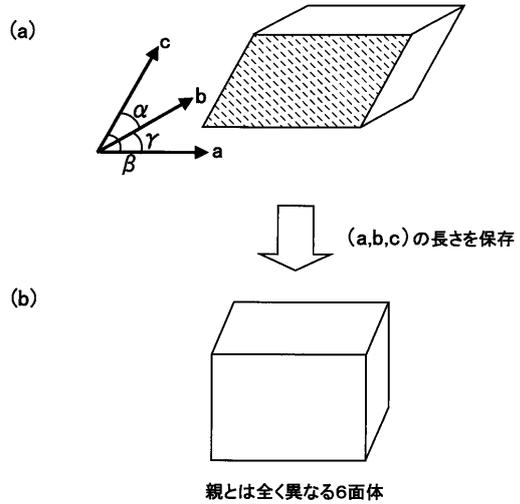
【 図 2 0 】

図20



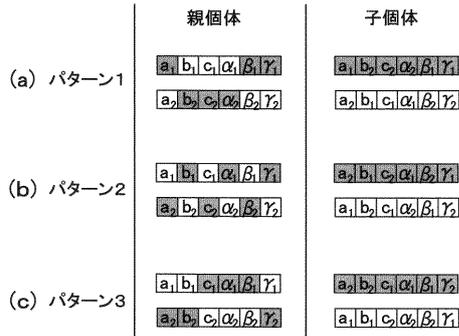
【 図 2 1 】

図21



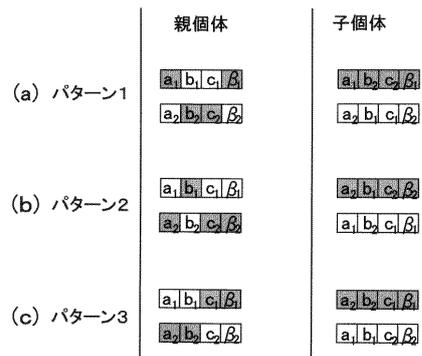
【 図 2 2 】

図22



【 図 2 3 】

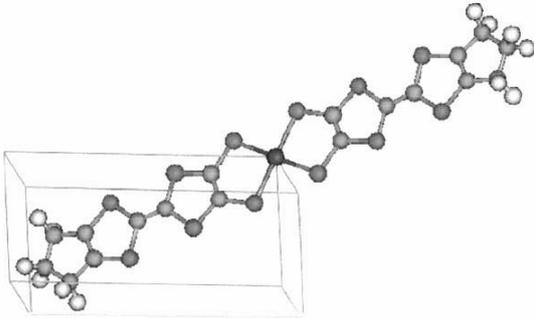
図23



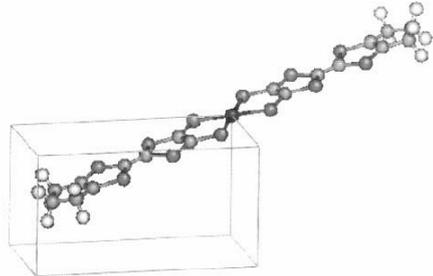
【 図 4 】

図4

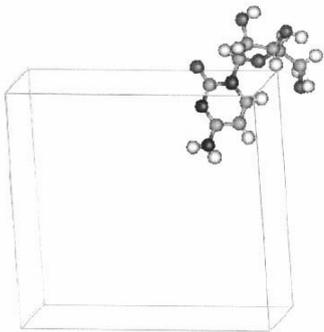
(a)



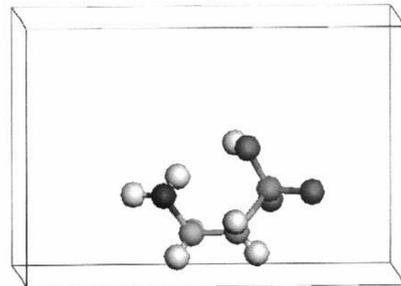
(b)



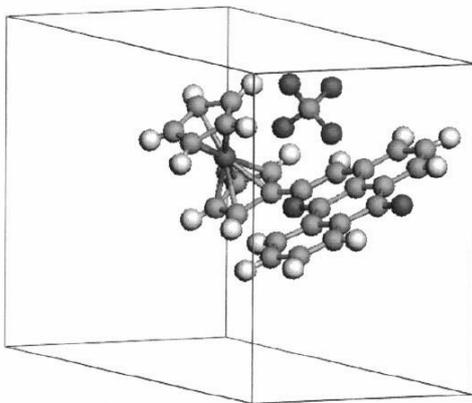
(c)



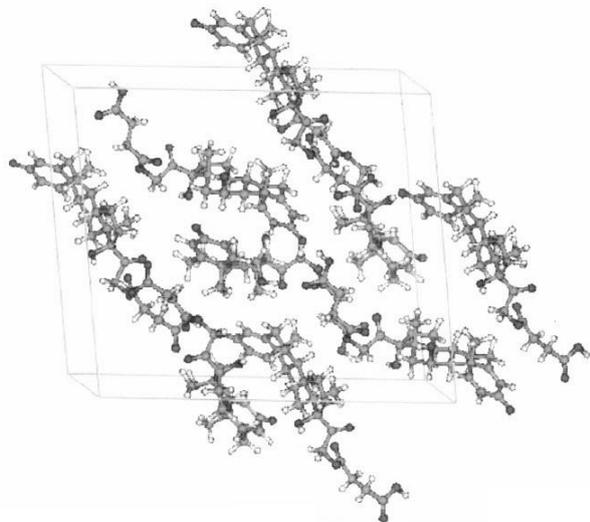
(d)



(e)



(f)



---

フロントページの続き

- (56)参考文献 特開2005-190372(JP,A)  
特開2003-219558(JP,A)  
特開2001-083106(JP,A)  
Giles W. Turner, et al., Implementation of Lamarckian concepts in a Genetic Algorithm for structure solution from powder diffraction data, Chemical Physics Letters, 2000年 4月28日, vol.321, p.183-190

- (58)調査した分野(Int.Cl., DB名)  
G06N 3/00-3/12  
G01N 23/20-23/207

- (54)【発明の名称】ハイブリッドGAによる複数パラメータの最適化方法、パターンマッチングによるデータ解析方法、放射線回折データに基づく物質構造の推定方法、ならびに関連するプログラム、記録媒体および各種装置