

(19) 日本国特許庁(JP)

(12) 公開特許公報(A)

(11) 特許出願公開番号

特開2006-318048
(P2006-318048A)

(43) 公開日 平成18年11月24日(2006.11.24)

(51) Int. Cl.	F I	テーマコード (参考)
G06F 17/30 (2006.01)	G06F 17/30 210D	5B075
	G06F 17/30 170F	
	G06F 17/30 380E	

審査請求 有 請求項の数 12 O L (全 33 頁)

(21) 出願番号	特願2005-137690 (P2005-137690)	(71) 出願人	504132272 国立大学法人京都大学 京都府京都市左京区吉田本町36番地1
(22) 出願日	平成17年5月10日(2005.5.10)	(74) 代理人	110000338 特許業務法人原謙三国際特許事務所
		(72) 発明者	山下 富義 京都府京都市左京区吉田下阿達町46-2 9 国立大学法人京都大学大学院薬学研 究科内
		(72) 発明者	伊藤 貴之 京都府京都市左京区吉田本町36番地1 国立大学法人京都大学大学院情報学研 究科内
		Fターム(参考)	5B075 ND35 NR12 PQ02 PQ13 PQ46 UU18

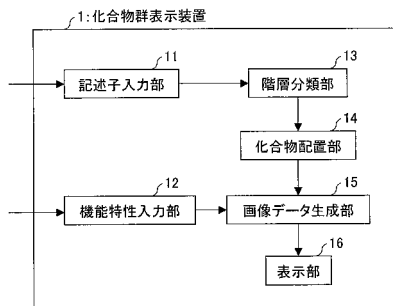
(54) 【発明の名称】 化合物群表示装置、化合物群表示方法、プログラム、及びコンピュータ読み取り可能な記録媒体

(57) 【要約】

【課題】 複数の化合物について第1の特性と第2の特性とを関連付けて効率的に表示できる化合物群表示装置であって、第1の特性として4つ以上の特性をも同時に取り扱うことのできる化合物群表示装置を提供する。

【解決手段】 化合物群表示装置1は、化合物を第1の特性情報に基づいて複数の階層的なクラスターに分類する階層分類部13と、これによって分類された階層的なクラスターに基づいた再帰的な入れ子構造となるように、化合物の2次元平面内における配置を決定する化合物配置部14と、これによって決定された配置で、第2の特性情報に基づいた色によって化合物が描画された画像データを生成する画像データ生成部15と、これによって生成された画像データに基づいた画像を表示する表示部16とを備えている。

【選択図】 図1



【特許請求の範囲】**【請求項 1】**

複数の化合物について、当該各化合物の第 1 の特性を表す第 1 の特性情報と第 2 の特性を表す第 2 の特性情報とに基づいて、第 1 の特性と第 2 の特性との関係を表示する化合物群表示装置であって、

上記化合物を、第 1 の特性情報に基づいて複数の階層的なクラスターに分類する階層分類手段と、

上記階層分類手段によって分類された階層的なクラスターに基づいた再帰的な入れ子構造となるように、上記化合物のアイコンの 2 次元平面内における配置を決定する化合物配置手段と、

上記化合物配置手段によって決定された配置で、第 2 の特性情報に基づいた色、模様、及び / 又は形状によって上記化合物のアイコンが描画された画像データを生成する画像データ生成手段と、

上記画像データ生成手段によって生成された画像データに基づいた画像を表示する表示部とを備えていることを特徴とする化合物群表示装置。

10

【請求項 2】

第 1 の特性情報が、 n 個（ただし、 n は自然数）の量的変数からなるものであり、

上記階層分類手段が、上記化合物を、各量的変数を座標軸とする n 次元空間における点として扱い、各化合物間又は各クラスター間の距離に基づいて、上記化合物を分類することを特徴とする請求項 1 に記載の化合物群表示装置。

20

【請求項 3】

上記階層分類手段が、第 1 の特性情報を説明変数、第 2 の特性情報を目的変数として、決定木を作成することによって上記化合物を分類することを特徴とする請求項 1 に記載の化合物群表示装置。

【請求項 4】

第 2 の特性情報に基づいて上記化合物を非階層的なクラスターに分類する非階層分類手段をさらに備え、

上記階層分類手段が、第 1 の特性情報を説明変数、上記非階層分類手段によって形成されたクラスターに対応するカテゴリ変数を目的変数として、決定木を作成することによって上記化合物を分類することを特徴とする請求項 1 に記載の化合物群表示装置。

30

【請求項 5】

第 2 の特性情報に基づいて上記化合物を非階層的なクラスターに分類する非階層分類手段をさらに備え、

上記画像データ生成手段が、上記非階層分類手段によって形成された各クラスターに対応する色、模様、及び / 又は形状によって上記化合物のアイコンが描画された画像データを生成することを特徴とする請求項 1 に記載の化合物群表示装置。

【請求項 6】

第 2 の特性情報が複数の量的変数からなるものであり、

上記非階層分類手段が、上記第 2 の特性情報をパターンベクトルとした自己組織化マップ法によって、上記化合物を非階層的なクラスターに分類することを特徴とする請求項 4 又は 5 に記載の化合物群表示装置。

40

【請求項 7】

第 1 の特性情報が、

(a) 化合物の分子構造の構造フラグメントもしくは構造トポロジーを表す 1 つ以上の分子記述子；

(b) 計算もしくは実験によって求められた物理化学的性質に対応する 1 つ以上の分子記述子、

の少なくとも何れか一方を含んでいることを特徴とする請求項 1 から 6 の何れか 1 項に記載の化合物群表示装置。

【請求項 8】

50

第2の特性情報が、上記化合物の生物活性、毒性、物理化学的性質、薬物動態学的性質の少なくとも何れかのパラメータであることを特徴とする請求項1から7の何れか1項に記載の化合物群表示装置。

【請求項9】

第2の特性情報がカテゴリ変数からなることを特徴とする請求項1から8の何れか1項に記載の化合物群表示装置。

【請求項10】

請求項1から9の何れか1項に記載の各手段として、コンピュータを動作させるためのプログラム。

【請求項11】

請求項10に記載のプログラムが記録されたコンピュータ読み取り可能な記録媒体。

【請求項12】

複数の化合物について、当該各化合物の第1の特性を表す第1の特性情報と第2の特性を表す第2の特性情報とに基づいて、第1の特性と第2の特性との関係を表示する化合物群表示装置における化合物群表示方法であって、

上記化合物群表示装置が、階層分類手段、化合物配置手段、画像データ生成手段、及び表示部を備え、

上記階層分類手段が第1の特性情報に基づいて複数の階層的なクラスターに分類する階層分類工程と、

上記化合物配置手段が上記階層分類工程において分類された階層的なクラスターに基づいた再帰的な入れ子構造となるように、上記化合物のアイコンの2次元平面内における配置を決定する化合物配置工程と、

上記画像データ生成手段が上記化合物配置工程において決定された配置で、第2の特性情報に基づいた色、模様、及び/又は形状によって上記化合物のアイコンが描画された画像データを生成する画像データ生成工程と、

上記表示部が上記画像データ生成工程において生成された画像データに基づいた画像を表示する表示工程とを備えていることを特徴とする化合物群表示方法。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、複数の化合物を分類して表示する化合物群表示装置及び化合物群表示方法と、この装置を実現するためのプログラム、さらには、このプログラムを記録したコンピュータ読み取り可能な記録媒体に関するものである。

【背景技術】

【0002】

医薬品、農薬、化粧品、及びその他の化学工業の分野では、活性や物性等の機能特性が優れた化合物を探索する研究が常に行われている。最近では、例えば、コンビナトリアルケミストリー技術とハイスループットスクリーニング法との組み合わせが頻繁に利用されている。これにより、多数の化合物を同時にかつ高速に合成し、評価することが可能となり、化合物の機能特性に関する大規模なスクリーニングデータが蓄積されるようになった。

【0003】

このようにして蓄積されたスクリーニングデータを基に、さらに機能特性の優れた新規化合物を合成展開するためには、得られたスクリーニングデータを化合物の構造と対応付けて整理することが不可欠である。

【0004】

化合物の構造とスクリーニングデータとの関係を整理するためには、まず、化合物の構造を数量的に表現する手法が必要である。この手法として、これまでに数多くの分子記述子が考案されている。その代表的なものとしては、構造キー (Daylight Chemical Information Systems社、MDL Information Systems社)、Molconn-Zディスクリプタ (Hall Assoc

10

20

30

40

50

iate Consulting社)、ClogP(Biobyte社)、ACD/LogD(ACD Labs社)等が挙げられる。

【0005】

一方、これらの分子記述子を用いてスクリーニングデータを管理する方法についても、これまで多くの方法が提案されている。その方法は以下の二つに大別される。

【0006】

一つは、スクリーニングデータを目的変数、分子記述子を説明変数として多変量解析を行い、それらの関係を定量的に表現するモデル式として管理する方法である(例えば、非特許文献1)。これにより、未知化合物についても、化学構造から分子記述子を計算し、これをモデル式に適用することによって機能特性を予測することが可能となる。

【0007】

もう一つの方法は、各化合物の分子記述子とスクリーニングデータ(機能特性)とを対応付けたデータベースを構築する方法である(例えば、非特許文献2)。構造的に類似すれば機能特性も類似することが予想されるため、未知化合物と構造的に類似する化合物をデータベースから検索すれば、検索した化合物のスクリーニングデータに基づいて未知化合物の機能特性を予測することが可能となる。この場合、類似性/非類似性の評価は、分子記述子に基づいて行われる。

【0008】

上記の各方法は、何れも統計学又は推計学的なアプローチに基づくものであり、化合物における分子記述子と機能特性との関係を経験的に予測するものである。従って、これらの方法では、機構論に基づく理論的な予測に比して計算結果の妥当性が問われることが多く、妥当性を客観的に評価しうる表現方法が望まれている。これを解決する極めて有効な方法がデータの可視化である。データを可視化すれば、人間が経験によって培った高度な分析能力を活用して適切な判定を行うことができる。また、可視化技術により、大規模な情報リソースを共有することが容易になるため、プロジェクト研究において意思の統一を図り、プロジェクトを効率よく推進することも可能になる。

【0009】

従来、化合物の機能特性を構造情報と関連付けて可視化する方法として、(a)単一の分子記述子、又は(b)複数の分子記述子に基づく総合特性値、を軸とする3次元以下の特性空間に化合物をマッピングする方法が採用されてきた(例えば、非特許文献3、特許文献1)。これにより、データベースに収載される化合物群について、機能特性と構造情報とを対応付けて表示することができる。そして、特性の評価を行いたい未知化合物があれば、構造情報に基づいて未知化合物を特性空間上にプロットすることによって、データベースに蓄積された化合物の機能特性から未知化合物の特性を視覚的に評価することができる。

【特許文献1】特公2002-531894(平成14年(2002)9月24日公開)

【非特許文献1】C. Hansch et al., Chem-bioinformatics: comparative QSAR at the interface between chemistry and biology. Chem. Rev., 102: 783-812, 2002

【非特許文献2】R.P. Sheridan and S. K. Kearsley, Why do we need so many chemical similarity search methods? Drug Discovery Today, 7: 903-911, 2002

【非特許文献3】Y. Takahashi et al., MolSpace: a computer desktop tool for visualization of massive molecular data. J. Mol. Graph Model., 21: 333-339, 2003

【発明の開示】

【発明が解決しようとする課題】

【0010】

しかしながら、上記従来の方法には以下の問題がある。

【0011】

まず、従来の方法では、特性空間における各座標軸が間隔尺度を表すものであるため、プロットされた化合物の配置が離散的になったり、偏在したりする。これにより、ディスプレイの表示領域の一部に化合物の存在しない空白部分ができたり、それとは逆に、一部に化合物が密集して重複部分ができたりするおそれがある。このように、従来の方法では

10

20

30

40

50

、化合物を特性空間内にプロットする際に、ディスプレイの表示領域を有効に利用できない。

【0012】

また、ディスプレイの表示領域では、3次元の特性空間を表示するのが限界であり、4次元以上の特性空間を表示することはできない。これは、化合物の構造情報として、4つ以上の構造情報を同時に取り扱えないことを意味する。よって、構造情報を何らかの形で3つ以下に減少させる必要があり、このときに、化合物の構造情報の一部が失われてしまう。このように、従来の方法では、化合物について、多数の構造情報からなる構造特性と機能特性との関係を包括的に評価することができない。

【0013】

本発明は上記課題に鑑みてなされたものであり、その目的は、複数の化合物について、例えば構造特性及び機能特性のような、第1の特性と第2の特性とを関連付けて効率的に表示できる化合物群表示装置であって、第1の特性として多数の特性情報をも同時に取り扱うことのできる化合物群表示装置を提供することにある。

【課題を解決するための手段】

【0014】

上記課題を解決するために、本発明に係る化合物群表示装置は、複数の化合物について、当該各化合物の第1の特性を表す第1の特性情報と第2の特性を表す第2の特性情報とに基づいて、第1の特性と第2の特性との関係を表示する化合物群表示装置であって、上記化合物を、第1の特性情報に基づいて複数の階層的なクラスターに分類する階層分類手段と、上記階層分類手段によって分類された階層的なクラスターに基づいた再帰的な入れ子構造となるように、上記化合物のアイコンの2次元平面内における配置を決定する化合物配置手段と、上記化合物配置手段によって決定された配置で、第2の特性情報に基づいた色、模様、及び/又は形状によって上記化合物のアイコンが描画された画像データを生成する画像データ生成手段と、上記画像データ生成手段によって生成された画像データに基づいた画像を表示する表示部とを備えていることを特徴とする。

【0015】

なお、第1の特性としては、例えば、化合物の構造特性等が挙げられ、第2の特性としては、例えば、化合物の機能特性等が挙げられる。

【0016】

上記構成によれば、階層分類手段が化合物を第1の特性情報に基づいて階層的なクラスターに分類する。そして、化合物配置手段及び画像データ生成手段が、分類したクラスターに基づいた入れ子構造となるように化合物を2次元平面内に表示する。これにより、各化合物は、アイコンの含まれる入れ子構造によって第1の特性が表現され、アイコンの色、模様、及び/又は形状によって第2の特性が表現される。以上のように、本発明の化合物群表示装置は、化合物群を、第1の特性と第2の特性とを関連付けて表示することができる。なお、上記の色には、モノクロ及びグレースケールも含まれるものとする。

【0017】

ここで、階層的なクラスターに分類するにあたって、第1の特性情報に含まれる変数の個数は複数であってもよい。すなわち、第1の特性情報が多次元ベクトルからなるものであっても、化合物を階層的なクラスターに分類することができる。よって、化合物群表示装置は、第1の特性として多数の特性情報をも同時に取り扱うことができる。

【0018】

また、化合物は、第1の特性情報を座標軸とする特性空間における点によって表現されるのではなく、第1の特性情報に基づいた再帰的な入れ子構造におけるアイコンによって表現される。ここで、各アイコンの位置は、第1の特性情報の数値に基づいて一点に決定されるわけではなく、類似する化合物のアイコン同士は近隣に存在する、という風に相対的な位置関係が大まかに決定される。よって、アイコンの配置が離散的になったり、偏在したりしないように、なるべく一様な間隔になるように配置することができる。このように、ディスプレイの表示領域を有効に利用できる。

10

20

30

40

50

【0019】

また、第1の特性情報は、 n 個（ただし、 n は自然数）の量的変数からなるものであり、上記階層分類手段が、上記化合物を、各量的変数を座標軸とする n 次元空間における点として扱い、各化合物間又は各クラスター間の距離に基づいて、上記化合物を分類してもよい。

【0020】

上記構成によれば、化合物の座標やクラスターの重心などの座標から算出した距離を非類似度の尺度として、化合物を分類することができる。

【0021】

また、上記階層分類手段は、第1の特性情報を説明変数、第2の特性情報を目的変数として、決定木を作成することによって上記化合物を分類してもよい。

10

【0022】

上記構成によれば、第1の特性情報に基づいて決定木を作成することによって分類されたクラスター内の化合物同士は、第2の特性情報が一樣になる傾向にある。よって、第1の特性と第2の特性との対応関係が明確になり、未知の化合物がどのような第2の特性を有するかを、その化合物の第1の特性情報に基づいて容易に推定できるようになる。なお、決定木には回帰木も含まれるものとする。

【0023】

また、上記化合物群表示装置は、第2の特性情報に基づいて上記化合物を非階層的なクラスターに分類する非階層分類手段をさらに備え、上記階層分類手段が、第1の特性情報を説明変数、上記非階層分類手段によって形成されたクラスターに対応するカテゴリ変数を目的変数として、決定木を作成することによって上記化合物を分類してもよい。

20

【0024】

上記構成によれば、第2の特性情報が複数の変数からなるものであっても、非階層分類手段が化合物を第2の特性情報に基づいて非階層的なクラスターに分類するので、化合物が有する第2の特性情報を、クラスター数と同数の値を取り得るカテゴリ変数として扱うことができるようになる。これにより、決定木の各クラスターにおける化合物の第2の特性を明確にすることが可能になる。

【0025】

また、上記化合物群表示装置は、第2の特性情報に基づいて上記化合物を非階層的なクラスターに分類する非階層分類手段をさらに備え、上記画像データ生成手段が、上記非階層分類手段によって形成された各クラスターに対応する色、模様、及び/又は形状によって上記化合物のアイコンが描画された画像データを生成するものであってもよい。

30

【0026】

上記構成によれば、第2の特性情報が複数の変数からなるものであっても、非階層分類手段が化合物を第2の特性情報に基づいて非階層的なクラスターに分類するので、各化合物が有する第2の特性を、分類したクラスターに基づいた色、模様、及び/又は形状によって表現することができる。

【0027】

また、第2の特性情報が複数の量的変数からなるものであり、上記非階層分類手段が、上記第2の特性情報をパターンベクトルとした自己組織化マップ法によって、上記化合物を非階層的なクラスターに分類してもよい。

40

【0028】

自己組織化マップ法によれば、多次元のベクトルにより表されたデータを、その特徴を残し他のデータとの相互関係を保ったまま、2次元のマップに写像することが出来る。この2次元のマップ上では、類似する多次元データを有するノードが近接して配置され、データの関係が視覚的に容易に理解できる。

【0029】

また、第1の特性情報が、(a)化合物の分子構造の構造フラグメントもしくは構造トポロジーを表す1つ以上の分子記述子；(b)計算もしくは実験によって求められた物理化学

50

的性質に対応する1つ以上の分子記述子、の少なくとも何れか一方を含んでいてもよい。

【0030】

また、第2の特性情報が、上記化合物の生物活性、毒性、物理化学的性質、薬物動態学的性質の少なくとも何れかのパラメータであってもよい。

【0031】

また、第2の特性情報がカテゴリ変数からなるものであってもよい。

【0032】

ところで、上記化合物群表示装置における各手段は、ハードウェアで実現してもよいし、プログラムをコンピュータに実行させることによって実現してもよい。具体的には、本発明に係るプログラムは、上記各手段としてコンピュータを動作させるプログラムであり、本発明に係る記録媒体は、当該プログラムが記録されたものである。

10

【0033】

これらのプログラムがコンピュータによって実行されると、当該コンピュータは、上記化合物群表示装置の各手段として動作する。したがって、上記化合物群表示装置と同様の効果が得られる

上記課題を解決するために、本発明に係る化合物群表示方法は、複数の化合物について、当該各化合物の第1の特性を表す第1の特性情報と第2の特性を表す第2の特性情報とに基づいて、第1の特性と第2の特性との関係を表示する化合物群表示装置における化合物群表示方法であって、上記化合物群表示装置が、階層分類手段、化合物配置手段、画像データ生成手段、及び表示部を備え、上記階層分類手段が第1の特性情報に基づいて複数の階層的なクラスターに分類する階層分類工程と、上記化合物配置手段が上記階層分類工程において分類された階層的なクラスターに基づいた再帰的な入れ子構造となるように、上記化合物のアイコンの2次元平面内における配置を決定する化合物配置工程と、上記画像データ生成手段が上記化合物配置工程において決定された配置で、第2の特性情報に基づいた色、模様、及び/又は形状によって上記化合物のアイコンが描画された画像データを生成する画像データ生成工程と、上記表示部が上記画像データ生成工程において生成された画像データに基づいた画像を表示する表示工程とを備えていることを特徴とする。

20

【0034】

なお、第1の特性としては、例えば、化合物の構造特性等が挙げられ、第2の特性としては、例えば、化合物の機能特性等が挙げられる。

30

【0035】

上記構成によれば、階層分類工程において化合物を第1の特性情報に基づいて階層的なクラスターに分類する。そして、化合物配置工程及び画像データ生成工程において、分類したクラスターに基づいた入れ子構造となるように化合物を2次元平面内に表示する。これにより、各化合物は、アイコンの含まれる入れ子構造によって第1の特性が表現され、アイコンの色、模様、及び/又は形状によって第2の特性が表現される。以上のように、本発明の化合物群表示方法は、化合物群を、第1の特性と第2の特性とを関連付けて表示することができる。なお、上記の色には、モノクロ及びグレースケールも含まれるものとする。

【0036】

ここで、階層的なクラスターに分類するにあたって、第1の特性情報に含まれる変数の個数は複数であってもよい。すなわち、第1の特性情報が多次元ベクトルからなるものであっても、化合物を階層的なクラスターに分類することができる。よって、本方法は、第1の特性として4つ以上の特性をも同時に取り扱うことができる。

40

【0037】

また、化合物は、第1の特性情報を座標軸とする特性空間における点によって表現されるのではなく、第1の特性情報に基づいた再帰的な入れ子構造におけるアイコンによって表現される。ここで、各アイコンの位置は、第1の特性情報の数値に基づいて一点に決定されるわけではなく、類似する化合物のアイコン同士は近隣に存在する、という風に相対的な位置関係が大まかに決定される。よって、アイコンの配置が離散的になったり、偏在

50

したりしないように、なるべく一様な間隔になるように配置することができる。このように、ディスプレイの表示領域を有効に利用できる。

【発明の効果】

【0038】

本発明に係る化合物群表示装置は、以上のように、化合物を第1の特性情報に基づいて複数の階層的なクラスターに分類する階層分類手段と、階層分類手段によって分類された階層的なクラスターに基づいた再帰的な入れ子構造となるように、化合物のアイコンの2次元平面内における配置を決定する化合物配置手段と、化合物配置手段によって決定された配置で、第2の特性情報に基づいた色、模様、及び/又は形状によって化合物のアイコンが描画された画像データを生成する画像データ生成手段と、画像データ生成手段によ

10

【0039】

また、本発明に係る化合物群表示方法は、階層分類手段が第1の特性情報に基づいて複数の階層的なクラスターに分類する階層分類工程と、化合物配置手段が階層分類工程において分類された階層的なクラスターに基づいた再帰的な入れ子構造となるように、化合物のアイコンの2次元平面内における配置を決定する化合物配置工程と、画像データ生成手段が化合物配置工程において決定された配置で、第2の特性情報に基づいた色、模様、及び/又は形状によって上記化合物のアイコンが描画された画像データを生成する画像データ生成工程と、表示部が画像データ生成工程において生成された画像データに基づいた画像を表示する表示工程とを備えた構成となっている。

20

【0040】

よって、上述したように、複数の化合物について第1の特性と第2の特性とを関連付けて効率的に表示でき、さらに、第1の特性として多数の特性情報をも同時に取り扱うこともできるという効果を奏する。

【発明を実施するための最良の形態】

【0041】

〔実施形態1〕

本発明の一実施形態について図1から図5に基づいて説明すると以下の通りである。本実施形態の化合物群表示装置1は、複数の化合物について、第1の特性と第2の特性とを関連付けて表示するものである。第1の特性の種類としては特に限定されるものではないが、本実施形態では、一例として構造特性が用いられ、これを表す構造情報として分子記述子が用いられている。なお、構造情報は、単一の分子記述子であってもよいし、n個の分子記述子、すなわち、各分子記述子の値を成分とするn次元ベクトルとして表現されていてもよい。また、分子記述子としては、化合物の分子構造の構造フラグメントもしくは構造トポロジーを表すものであってもよいし、計算もしくは実験によって求められた物理化学的性質に対応するものであってもよいし、これらの組み合わせであってもよい。

30

【0042】

また、第2の特性についても、第1の特性情報と同様に特に限定されるものではなく、例えば、化合物の生物活性、毒性、物理化学的性質、又は薬物動態学的性質等の機能特性等を用いることができ、これを表す機能特性情報として各種パラメータを用いることができる。また、これらのパラメータを適宜組み合わせてもよい。本明細書では、これらのパラメータについて以下「機能特性情報」という。

40

【0043】

なお、本実施形態において、機能特性情報は、1次元ベクトルで表される量的変数、カテゴリ変数の何れであってもよい。量的変数とは、例えば、化合物の活性値等の間隔尺度で測られる変数であり、離散変数などであってもよく、また、カテゴリ変数とは、例えば、化合物の活性の種類(A活性、B活性、又はC活性)等の複数の値を取りうる変数をいう。

【0044】

このような構成により、本実施形態の化合物群表示装置は、化合物の構造と機能特性と

50

を関連付けて表示することができる。よって、ユーザは、未知の化合物の機能特性を知りたい場合、構造情報から未知の化合物がどこに表示されるかを知ることによって、構造が類似する化合物の機能特性情報から未知の化合物の機能特性を推測することができる。また、表示されている既知の化合物の分布から、構造と機能特性とがどのように関係しているかを、視覚的に判断することもできる。

【0045】

図1は、本実施形態の化合物群表示装置1の機能ブロック図である。化合物群表示装置1は、図1に示すように、記述子入力部11、機能特性入力部12、階層分類部(階層分類手段)13、化合物配置部(化合物配置手段)14、画像データ生成部(画像データ生成手段)15、及び表示部16を備えている。

10

【0046】

記述子入力部(第1の入力部)11は、ユーザ又は他の装置等から、表示する化合物の分子記述子(第1の特性情報)を取得するためのものである。記述子入力部11のハードウェア構成としては、例えばキーボード、マウス、タッチパネル、又はタブレット等のようにユーザからの入力を受け付けるものであってもよいし、各種入出力インターフェース等のように外部記憶装置や他のコンピュータからの入力を受け付けるものであってもよい。

【0047】

また、機能特性入力部(第2の入力部)12は、ユーザ又は他の装置等から、表示する化合物の機能特性情報(第2の特性情報)を取得するためのものである。機能特性入力部12のハードウェア構成については、記述子入力部11と同様である。

20

【0048】

階層分類部13は、化合物を分子記述子(第1の特性情報)に基づいて複数の階層的なクラスターに分類するためのものである。より具体的には、分子記述子に基づいて、各種アルゴリズムによって化合物を階層的に分類する。

【0049】

化合物配置部14は、階層分類部13によって分類された階層的なクラスターに基づいた再帰的な入れ子構造となるように、化合物の2次元平面内における配置を決定する。より具体的には、化合物配置部14は、階層的に分類された化合物を葉ノードとして、子ノードが親ノードに順次含まれた入れ子構造となるように、各化合物を2次元平面内に配置する。なお、このように配置するためのアルゴリズムとしては、例えば、平安京ビュー(「平安京ビュー ~ 階層型データを碁盤上に配置する視覚化手法」、伊藤貴之、小山田耕二、可視化情報学会第9回ビジュアライゼーションカンファレンス抄録、2003年発行)を用いることができる。平安京ビューでは、まず階層型データ中の葉ノード群を格子状に配列する。続いてこれらの葉ノードの親にあたる枝ノードを画面上に効率よく配置することで、限られた画面空間上に大量の情報を表現する。

30

【0050】

画像データ生成部15は、化合物配置部14によって決定された配置で、第2の特性情報に基づいた色によって化合物のアイコンが描画された画像データを生成する。より具体的には、画像データ生成部15は、化合物配置部14によって決定された位置に化合物が配置された画像データを生成する。このとき、葉ノードに相当する各化合物は、例えば、円形、正方形、又は点等の簡単な図形で表現してもよいし、化合物の種類が分かるようなシンボルによって表現してもよい。以下、葉ノードを表現する図形又はシンボルをアイコンと称する。また、このとき、各化合物(葉ノード)に相当するアイコンを、第2の特性情報に基づいて色分けする。さらに、画像データ生成部15は、葉ノードである各化合物が属する親ノードを枠で囲み、枠で囲んだ各ノードが属するさらなる親ノードを枠で囲み...、といった風に、順次、ノードを枠で囲んでいく。これにより、葉ノードである各化合物が、どのような階層構造のどのノードに属しているかが表現される。

40

【0051】

なお、上記の色分けに用いられる色は必ずしもカラーである必要はなく、モノトーンの

50

濃淡（グレースケール）等でもよい。また、化合物の機能特性を色によって表現する代わりに、網掛け等の模様や、アイコンの形状によって表現してもよい。もちろん、これらを適宜組み合わせ用いてもよい。

【0052】

表示部16は、画像データ生成部15によって生成された画像データに基づいた画像を表示するためのものである。表示部16のハードウェア構成としては、画像をカラー表示できるものであれば特に限定されるものではなく、例えば、カラー表示可能な液晶ディスプレイやCRT（Cathode Ray Tube）ディスプレイ等の各種ディスプレイを備えていてもよいし、カラープリンタを備えていてもよい。これにより、表示部16は、画像データ生成部15によって生成された画像データによって表現された画像を表示することができる。

10

【0053】

なお、上記の階層分類部13、化合物配置部14、及び画像データ生成部15は、CPU等の演算装置がROMやRAM等の記憶装置に格納されたプログラムコードを実行することで実現される機能ブロックである。

【0054】

次に、本実施形態の化合物群表示装置の動作について説明する。図2は、本実施形態の化合物群表示方法の工程図である。

【0055】

まず、ステップS10において、機能特性入力部12が機能特性情報を取得する。取得する機能特性情報は、本実施形態では、1次元ベクトルで表される量的変数、カテゴリ変数の何れであってもよい。表1は、取得した機能特性情報の例を示すテーブルである。

20

【0056】

【表1】

	活性値
化合物1	2.5
化合物2	1.0
化合物3	0.3
化合物4	1.5

30

【0057】

取得した機能特性情報は、画像データ生成部15に送信される。

【0058】

次に、ステップS11において、記述子入力部11が分子記述子を取得する。取得する分子記述子は一種類のみでもよいが、本実施形態ではn種類とする。これらの分子記述子は、化合物と対応付けられている。表2は、取得した分子記述子の例を示すテーブルである。

【0059】

【表2】

	分子記述子1	分子記述子2	分子記述子3
化合物1	0.5	2	0
化合物2	0.9	4	1
化合物3	0.1	1	0
化合物4	0.3	2	0

40

【0060】

なお、表2では、各化合物に対応付けられた分子記述子の種類は3つであるが、本発明はこれに限定されるものではなく、任意の個数の分子記述子を取り扱うことができる。

50

【0061】

そして、取得した分子記述子は階層分類部13に送信される。

【0062】

次に、ステップS12において、階層分類部13が、受信した分子記述子に基づいて化合物を階層的なクラスターに分類する。なお、本ステップにおいて、特に記述しない場合は、階層分類部13が各処理を行っているものとする。

【0063】

説明変数が上述の分子記述子のような量的変数である場合、階層分類部13は、各化合物をn次元空間内の点として取り扱う。すなわち、表1の場合、分子記述子の種類が3つであるので、化合物1は、3次元空間内の点(0.5, 2, 0)となる。また同様に、化合物2は、点(0.9, 4, 1)となる。なお、このとき、分子記述子に対して標準化処理を行ってもよい。具体的には、例えば、各分子記述子の化合物間における平均値が0となり、標準偏差値が1となるように標準化すること等が挙げられる。

10

【0064】

また、構造情報が複数の分子記述子を要素とする多次元ベクトルからなる場合、構造情報について主成分分析を行い、構造情報の次元(変数の数)を縮約してもよい。

【0065】

そして、各化合物がこのように配置される空間において、化合物間の距離から化合物間の非類似度を算出する。すなわち、化合物間の距離の遠いものほど化合物間の類似度が小さいものとして扱い、同一クラスターに分類されないようにする。なお、非類似度を算出する際の距離としては、例えば、ユークリッド距離、マンハッタン距離、及びべき乗距離等のうち好適な距離を利用することができる。

20

【0066】

なお、変形例としては、距離に基づく非類似度の代わりに類似度を算出してもよい。類似度を算出する方法としては、例えば、ピアソン相関係数、Tanimoto係数、コサイン関数等を利用することができる。

【0067】

また、距離に基づいて階層的なクラスターに分類するアルゴリズムとしては、例えば、最近隣法、群平均法、又はWard法等の公知のものを用いることができる。

【0068】

以上の処理により、例えば図3のようなデンドログラムを表現できる各クラスター間の非類似度(距離)及び階層構造の情報が得られる。なお、図3における階層構造及び距離は正確なものではない。得られた情報のうち、階層構造の情報が、化合物配置部14に送信される。表3は、階層構造を示すデータ(階層構造テーブル)の一例である。

30

【0069】

【表3】

枝ノードID	子枝ノードID	葉ノード(化合物)ID
1(根ノード)	2, 3	null
2	null	1, 2
3	null	3, 4

40

【0070】

表3に示すように、階層構造テーブルには、枝ノードのIDと、その枝ノードに含まれる子ノード及び葉ノードのIDが対応付けられて格納されている。

【0071】

次に、ステップS13において、化合物配置部14が、上記階層構造テーブルによって表現される階層構造に基づいて、2次元平面内に各化合物(葉ノード)を配置する。このとき、同一の枝ノードに含まれる葉ノード同士が近接するように、かつ、隣接する葉ノード同士の間隔が葉ノード間の距離にかかわらず略一定の間隔になるように、各葉ノードを

50

配置する。

【0072】

図4は、2次元平面内における化合物の配置の一例を示す図である。なお、図中、各葉ノードは、黒塗りの正方形で表現されている。また、各葉ノードを囲む枠は、階層構造を視覚的に認識できるように付記したものであり、化合物配置部14によって作成される配置データには含まれていない。すなわち、化合物配置部14によって生成される配置データには、化合物(葉ノード)とその化合物の表示位置(例えば、図中の正方形における左上の座標)の情報とが対応付けて含まれている。生成された配置データは、画像データ生成部15に送信される。

【0073】

次に、ステップS14において、画像データ生成部15が、受信した配置データに基づいて、画像データを生成する。なお、生成された画像データでは、各化合物(葉ノード)が枝ノードを示す枠によって囲まれ、枠に囲まれた枝ノードがさらにその親ノードを示す枠によって囲まれ、再帰的な入れ子構造となっている。また、各化合物(葉ノード)のアイコンは、ステップS10で取得した機能特性情報に基づいて、色分けされる。

【0074】

例えば、入力された機能特性情報がカテゴリ変数である場合、画像データ生成部15は、各カテゴリ変数と色とが対応付けられたテーブル等に基づいて、化合物の色を決定する。すなわち、画像データ生成部15は、同じカテゴリ変数を有する化合物が同じ色となるように、各化合物の色を決定する。図5は、化合物1と化合物2、化合物3と化合物4が、それぞれ同じカテゴリ変数からなる機能特性情報を有している場合の色分けの一例である。

【0075】

また、入力された機能特性情報が量的変数である場合、画像データ生成部15は、各機能特性情報の値を、変換式によって色情報(例えば、R、G、Bの何れか1つの色における明度等)に変換して、アイコンの色が機能特性情報の値に応じて連続的に変化するように、化合物の色を決定してもよい。

【0076】

このようにして生成された画像データは、表示部16に送信される。なお、ステップS14において、画像データ生成部15は、平面内に配置された各化合物(葉ノード)を示すアイコンに対して、高さを与えてもよい。例えば、各化合物が1次元ベクトルで表現される第3の特性情報を有している場合、第3の特性情報の値を高さに変換し、各化合物のアイコンが高さを有するように立体的に描画された画像データを生成してもよい。これにより、化合物群表示装置1は、第1から第3の特性情報を関連付けて表示することができる。

【0077】

最後に、ステップS15において、表示部16が、受信した画像データに対応する画像を表示する。図5は表示部16による表示の一例である。なお、化合物群表示装置1では、表示部16によって表示された画像における各化合物をポインタで指定すると、化合物の分子記述子や機能特性情報が表示されることが好ましい。

【0078】

〔実施形態2〕

本発明の別の実施形態について図6及び図7に基づいて説明すると以下の通りである。なお、上述した実施形態1と同一の機能を有する部材については、実施形態1と同一の番号を付記し、説明を省略する。

【0079】

本実施形態の化合物群表示装置2は、複数の化合物について、第1の特性情報と第2の特性情報とを関連付けて表示するものである。

【0080】

第1の特性情報の種類としては特に限定されるものではなく、量的変数、カテゴリ変数

10

20

30

40

50

の何れであってもよく、また、それらの変数の数は、単一であっても複数であってもよい。なお、本実施形態では、実施形態 1 と同様に複数の分子記述子が用いられるものとする。

【0081】

また、第 2 の特性情報についても、第 1 の特性情報と同様に特に限定されるものではなく、例えば、化合物の生物活性、毒性、物理化学的性質、又は薬物動態学的性質等のパラメータを用いることができる。また、これらのパラメータを適宜組み合わせてもよい。本明細書では、これらのパラメータについて以下「機能特性情報」という。

【0082】

なお、本実施形態の化合物群表示装置 2 は、化合物を階層的なクラスターに分類する際に、分子記述子を説明変数、機能特性情報を目的変数として、決定木を作成する点が実施形態 1 と異なっている。これにより、階層的に分類された化合物が、クラスター内において一様な機能特性を示すようになる。よって、構造と機能特性との関係が明確になり、未知の化合物がどのような機能特性を有するかを、その分子記述子に基づいて容易に推定できるようになる。

10

【0083】

また、機能特性情報が、1次元ベクトルで表される量的変数、カテゴリ変数に加えて、 n 次元ベクトル($n \geq 2$)で表される量的変数であってもよい点も実施形態 1 と異なっている。すなわち、本実施形態の化合物群表示装置 2 は、機能特性情報が n 次元で表されるような化合物についても、化合物の構造と機能特性とを関連付けて表示することができる。

20

【0084】

図 6 は、本実施形態の化合物群表示装置 2 の機能ブロック図である。化合物群表示装置 2 は、図 6 に示すように、記述子入力部 11、機能特性入力部 12、階層分類部(階層分類手段) 23、非階層分類部(非階層分類手段) 27、化合物配置部(化合物配置手段) 14、画像データ生成部(画像データ生成手段) 15、及び表示部 16 を備えている。

【0085】

機能特性入力部 12 は、実施形態 1 と基本的に同様であるが、本実施形態では、 n 個の特性値(量的変数)を取得するものとする。

【0086】

非階層分類部 27 は、機能特性入力部 12 に入力された n 個の特性値(量的変数)に基づいて化合物を非階層的なクラスターに分類するためのものである。

30

【0087】

階層分類部 23 は、化合物を分子記述子(第 1 の特性情報)に基づいて複数の階層的なクラスターに分類するためのものである。より具体的には、分子記述子を説明変数、非階層分類部 27 によって形成されたクラスターをカテゴリとして、決定木を作成する。

【0088】

なお、第 2 の特性情報が単一の量的変数からなる場合は、階層分類部 23 は、分子記述子を説明変数、第 2 の特性情報を目的変数として決定木(回帰木)を作成することによって、化合物を階層的に分類してもよい。

40

【0089】

画像データ生成部 25 は、基本的には実施形態 1 と同様であるが、色分けを行う際に、非階層分類部 27 によって形成されたクラスターに基づいた色によって化合物のアイコンが描画された画像データを生成する点が、実施形態 1 と異なっている。

【0090】

記述子入力部 11、化合物配置部 14、及び表示部 16 については、実施形態 1 と同様である。

【0091】

なお、上記の非階層分類部 27、階層分類部 23、化合物配置部 14、及び画像データ生成部 25 は、CPU 等の演算装置が ROM や RAM 等の記憶装置に格納されたプログラ

50

ムコードを実行することで実現される機能ブロックである。

【0092】

次に、本実施形態の化合物群表示装置の動作について説明する。図7は、本実施形態の化合物群表示方法の工程図である。

【0093】

まず、ステップS20において、機能特性入力部12が機能特性情報を取得する。取得する機能特性情報は、本実施形態では、一例としてn次元ベクトルで表される量的変数であるものとする。表4は、取得した機能特性情報の一例である。

【0094】

【表4】

10

	A活性の値	B活性の値	C活性の値
化合物1	0	1	0
化合物2	2	0	0
化合物3	1	2	2
化合物4	1	0	2

【0095】

なお、表4では、各化合物に対応付けられた機能特性情報（活性値）の種類は3つであるが、本発明はこれに限定されるものではなく、任意の個数の機能特性情報を取り扱うことができる。

20

【0096】

そして、取得した機能特性情報は非階層分類部27に送信される。

【0097】

次に、ステップS21において、記述子入力部11が分子記述子を取得する。このステップについては、実施形態1のステップS11と同様である。取得した分子記述子は階層分類部23に送信される。

【0098】

次に、ステップS22において、非階層分類部27が、受信した機能特性情報に基づいて化合物を非階層的なクラスターに分類する。以下、機能特性情報に基づいて分類したクラスターのことを機能クラスターともいう。ここで、非階層的なクラスターに分類する方法としては、公知のものを利用することができる。一般的なものとしては、k-means法等が挙げられる。また、ニューラルネットワークモデル、例えば、コホネン（Kohonen）の自己組織化マップ法を用いてもよい。さらに、複数の方法を組み合わせてもよい。

30

【0099】

これにより、各化合物の所属する機能クラスターと、各機能クラスターの平均的な機能特性とが得られる。換言すれば、各化合物が、どのような機能特性を有するどのクラスターに所属するのかが決定される。なお、各クラスターの機能特性は、例えばk-means法の場合は、クラスターの重心位置の座標として得られ、自己組織化マップ法の場合は、ノードのパターンベクトルとして得られる。

40

【0100】

表5は、分類によって得られた、化合物の所属する機能クラスターが示されたテーブルの一例である。

【0101】

【表 5】

	機能クラスターID
化合物1	1
化合物2	1
化合物3	2
化合物4	1

【0102】

表 5 に示すように、上記テーブルでは、化合物と機能クラスターの ID とが対応付けられている。 10

【0103】

分類によって得られた、化合物の所属する機能クラスターの情報は、階層分類部 2 3 及び画像データ生成部 2 5 に送信される。

【0104】

次に、ステップ S 2 3 において、階層分類部 2 3 が、記述子入力部 1 1 から受信した分子記述子に基づいて、化合物を階層的なクラスターに分類する。ここで、本実施形態では、クラスターに分類する際に、各クラスター内の化合物の有する機能特性がなるべく一樣になるように分類する。具体的には、化合物の分子記述子を説明変数、非階層分類部 2 7 から受信した化合物の所属する機能クラスターを目的変数として、決定木を作成する。このときの分岐ルールとしては、公知のものを用いることができ、例えば、クインラン (Quinlan) による C 5 . 0、C & R T、Q U E S T、C H A I D 又は実施例に示すミッチェルらによる情報利得 G を用いる手法等が挙げられる。 20

【0105】

これにより、実施形態 1 の表 3 のような階層構造の情報が得られる。得られた階層構造の情報は、化合物配置部 1 4 に送信される。

【0106】

次に、ステップ S 2 4 において、化合物配置部 1 4 が、階層分類部 2 3 によって形成された階層構造の情報に基づいて、2 次元平面内に各化合物 (葉ノード) を配置する。このステップは、実施形態 1 のステップ S 1 3 と同様である。 30

【0107】

次に、ステップ S 2 5 において、画像データ生成部 2 5 が、受信した配置データに基づいて、画像データを生成する。このステップは、各化合物 (葉ノード) がステップ S 2 3 で決定された機能クラスターに基づいて色分けされる点を除いて、実施形態 1 のステップ S 1 4 と同様である。

【0108】

最後に、ステップ S 2 6 において、表示部 1 6 が、受信した画像データに対応する画像を表示する。このステップは、実施形態 1 のステップ S 1 5 と同様である。

【0109】

以上のように、本実施形態の化合物群表示装置 2 は、化合物を階層的なクラスターに分類する際に、分子記述子を説明変数、機能特性情報を目的変数として、決定木を作成する。これにより、階層的に分類された化合物が、クラスター内において一樣な機能特性を示すようになる。よって、構造と機能特性との関係が明確になり、未知の化合物がどのような機能特性を有するかを、その分子記述子に基づいて容易に推定できるようになる。 40

【0110】

また、本実施形態の化合物群表示装置 2 は、各化合物に関する機能特性情報が複数の特性値からなるような場合であっても、化合物を機能特性情報に基づいて非階層的なクラスター (機能クラスター) に分類し、所属する機能クラスターをその化合物の機能特性情報として扱う。そして、画像データ生成部 2 5 がアイコンの色分けをする際に、この機能クラスターに応じて色分けすることによって、化合物の機能特性を表現している。 50

【0111】

なお、上記各実施形態では、化合物群表示装置を構成する階層分類部、非階層分類部、化合物配置部、及び画像データ生成部が、「CPUなどの演算装置がROMやRAMなどの記憶装置に格納されたプログラムコードを実行することで実現される機能ブロックである」場合を例にして説明したが、同様の処理を行うハードウェアで実現してもよい。また、処理の一部を行うハードウェアと、当該ハードウェアの制御や残余の処理を行うプログラムコードを実行する上記演算手段とを組み合わせても実現することもできる。さらに、上記各部材のうち、ハードウェアとして説明した部材であっても、処理の一部を行うハードウェアと、当該ハードウェアの制御や残余の処理を行うプログラムコードを実行する上記演算手段とを組み合わせても実現することもできる。なお、上記演算手段は、単体であ

10

【0112】

上記演算手段によって直接実行可能なプログラムコード自体、または、後述する解凍などの処理によってプログラムコードを生成可能なデータとしてのプログラムは、当該プログラム（プログラムコードまたは上記データ）を記録媒体に格納し、当該記録媒体を配付したり、あるいは、上記プログラムを、有線または無線の通信路を介して伝送するための通信手段で送信したりして配付され、上記演算手段で実行される。

【0113】

なお、通信路を介して伝送する場合、通信路を構成する各伝送媒体が、プログラムを示す信号列を伝搬し合うことによって、当該通信路を介して、上記プログラムが伝送される。また、信号列を伝送する際、送信装置が、プログラムを示す信号列により搬送波を変調することによって、上記信号列を搬送波に重畳してもよい。この場合、受信装置が搬送波を復調することによって信号列が復元される。一方、上記信号列を伝送する際、送信装置が、デジタルデータ列としての信号列をパケット分割して伝送してもよい。この場合、受信装置は、受信したパケット群を連結して、上記信号列を復元する。また、送信装置が、信号列を送信する際、時分割/周波数分割/符号分割などの方法で、信号列を他の信号列と多重化して伝送してもよい。この場合、受信装置は、多重化された信号列から、個々の信号列を抽出して復元する。いずれの場合であっても、通信路を介してプログラムを伝送できれば、同様の効果が得られる。

20

30

【0114】

ここで、プログラムを配付する際の記録媒体は、取外し可能である方が好ましいが、プログラムを配付した後の記録媒体は、取外し可能か否かを問わない。また、上記記録媒体は、プログラムが記憶されていれば、書換え（書き込み）可能か否か、揮発性が否か、記録方法および形状を問わない。記録媒体の一例として、磁気テープやカセットテープなどのテープ、あるいは、フロッピー（登録商標）ディスクやハードディスクなどの磁気ディスク、または、CD-ROMや光磁気ディスク（MO）、ミニディスク（MD）やデジタルビデオディスク（DVD）などのディスクが挙げられる。また、記録媒体は、ICカードや光カードのようなカード、あるいは、マスクROMやEPROM、EEPROMまたはフラッシュROMなどのような半導体メモリであってもよい。あるいは、CPUなどの

40

【0115】

なお、上記プログラムコードは、上記各処理の全手順を上記演算手段へ指示するコードであってもよいし、所定の手順で呼び出すことで、上記各処理の一部または全部を実行可能な基本プログラム（例えば、オペレーティングシステムやライブラリなど）が既に存在していれば、当該基本プログラムの呼び出しを上記演算手段へ指示するコードやポインタなどで、上記全手順の一部または全部を置き換えてもよい。

【0116】

また、上記記録媒体にプログラムを格納する際の形式は、例えば、実メモリに配置した状態のように、演算手段がアクセスして実行可能な格納形式であってもよいし、実メモリ

50

に配置する前で、演算手段が常時アクセス可能なローカルな記録媒体（例えば、実メモリやハードディスクなど）にインストールした後の格納形式、あるいは、ネットワークや搬送可能な記録媒体などから上記ローカルな記録媒体にインストールする前の格納形式などであってもよい。また、プログラムは、コンパイル後のオブジェクトコードに限るものではなく、ソースコードや、インタプリタまたはコンパイルの途中で生成される中間コードとして格納されていてもよい。いずれの場合であっても、圧縮された情報の解凍、符号化された情報の復号、インタプリタ、コンパイル、リンク、または、実メモリへの配置などの処理、あるいは、各処理の組み合わせによって、上記演算手段が実行可能な形式に変換可能であれば、プログラムを記録媒体に格納する際の形式に拘わらず、同様の効果を得ることができる。

10

【実施例】**【0117】****〔実施例1〕**

本発明の一実施例について以下に説明する。なお、本発明はこれに限定されるものではない。本実施例は、化合物の水に対する溶解度の測定結果の表示に関するものである。具体的には、化合物の構造と溶解度とを関係付けて表示した。

【0118】

本実施例では、第2の特性情報（機能特性情報）として、文献（Y. Ran, et al., Prediction of aqueous solubility of organic compounds by the general solubility equation (GSE). J. Chem. Inf. Comput. Sci., 41: 1208-1217, 2001; G. Yan, et al., Prediction of the aqueous solubility: comparison of the general solubility equation and the method using an amended solvation energy relationship. J. Pharm. Sci., 91: 517-533, 2002）より収集した化合物の溶解度データを用いた。収集した溶解度データの総数、すなわち、本実施例で扱った化合物の総数は908個であった。なお、溶解度のデータは、化合物と対応付けられたテーブルとして取得した。

20

【0119】

また、第1の特性情報（構造情報）としては、Lipinskiら（C.A. Lipinski, et al., Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Adv. Drug Deliv. Rev. 23: 3-25, 1997）が薬剤様化合物を識別するために提示している4つの分子記述子、具体的にはClogP、分子量、水素結合供与基数、及び水素結合受容基数を用いた。これらの分子記述子のデータは、化合物と対応付けられたテーブルとして取得した。この対応テーブルの一部を表6に示す。

30

【0120】

【表 6】

ID	化合物名	ClogP	分子量	水素結合 供与基数	水素結合 受容基数
1	p-chloronitrobenzene	2.6	157.556	0	4
2	4-nitroaniline	1.26	138.126	1	4
3	p-nitrophenol	1.85	139.11	1	4
4	p-nitroanisole	2.1	153.137	0	4
5	1,4-dinitrobenzene	1.63	168.109	0	6
6	ethylbenzene	3.17	106.167	0	0
7	styrene	2.87	104.151	0	0
8	benzyl chloride	2.7	126.585	0	1
9	benzylamine	1.09	107.155	1	1
10	benzonitrile	1.57	103.123	0	1
11	phenylmethanol	1.1	108.14	1	1
12	benzaldehyde	1.5	107.132	0	1
:	:	:	:	:	:

10

【0121】

そして、これらの各分子記述子について、化合物間で平均が0、かつ、標準偏差が1となるような標準化を行った。標準化を行った後の各分子記述子の対応テーブルを表7に示す。

20

【0122】

【表 7】

ID	化合物名	ClogP	分子量	水素結合 供与基数	水素結合 受容基数
1	p-chloronitrobenzene	0.034	-0.159	-0.675	0.745
2	4-nitroaniline	-0.686	-0.384	0.398	0.745
3	p-nitrophenol	-0.369	-0.373	0.398	0.745
4	p-nitroanisole	-0.234	-0.210	-0.675	0.745
5	1,4-dinitrobenzene	-0.487	-0.037	-0.675	1.641
6	ethylbenzene	0.340	-0.755	-0.675	-1.047
7	styrene	0.179	-0.778	-0.675	-1.047
8	benzyl chloride	0.088	-0.518	-0.675	-0.599
9	benzylamine	-0.777	-0.743	0.398	-0.599
10	benzonitrile	-0.519	-0.790	-0.675	-0.599
11	phenylmethanol	-0.772	-0.732	0.398	-0.599
12	benzaldehyde	-0.557	-0.744	-0.675	-0.599
:	:	:	:	:	:

30

40

【0123】

次に、標準化後の分子記述子に基づいて、908個の化合物を階層的なクラスターに分類した。分類するアルゴリズムには、Ward法 (J.H. Ward, Hierarchical Grouping to Optimize an Objective Function. J. Am. Stat. Assoc., 58: 236-244, 1963) を用いた。Ward法は、初期状態 (未結合状態) における情報の欠損を0とし、クラスター合併における各結合のステップで情報欠損 $S_{p,q}$ の増加を最小にするようなクラスターの組を選ぶ方法であり、情報欠損 $S_{p,q}$ は次式で与えられる。

【0124】

50

【数 1】

$$\Delta S_{pq} = \frac{n_p n_q}{n_p + n_q} \sum_{m=1}^L (\bar{x}_{pm} - \bar{x}_{qm})^2$$

【0125】

ただし、バーのついた x_{pm} 、 x_{qm} は、それぞれ結合前のクラスター p、クラスター q の重心における m 番目の分子記述子の値であり、 n_p 、 n_q はクラスター p、q に属するクラスター又は化合物の数である。また、L は、本実施例では分子記述子の種類数、すなわち 4 である。

10

【0126】

この方法に基づき、クラスター数が 1 になるまで、クラスターの合併を繰り返した。そして、情報欠損 S_{pq} の閾値を 1.0 に設定し、それ以下のものは同一階層とみなして 908 個の化合物を分類した。

【0127】

次に、階層的に分類した化合物について、階層的なクラスターに基づいた再帰的な入れ子構造となるように、2次元平面内における配置を決定した。この配置を決定するアルゴリズムには、平安京ビュー（「平安京ビュー ~ 階層型データを基盤上に配置する視覚化手法」、伊藤貴之、小山田耕二、可視化情報学会第9回ビジュアルイゼーションカンファレンス抄録、2003年発行）を用いた。

20

【0128】

そして、908 個の化合物について、決定された配置で、溶解度の対数値を色相とした色でアイコンを描画した画像データを作成した。この画像データに基づく画像を図 8 に示す。なお、図中、化合物を示すアイコンはカラー表示されている。

【0129】

図 8 に示すように、比較的溶解度の高いもののみ、もしくは、低いもののみが集まっているクラスターが認められ、化学構造が類似する化合物は比較的良く似た溶解度を有することが示された。

【0130】

また、各化合物を示すアイコンに高さを与えたものを図 9 (a)、図 9 (b) に示す。図 9 (a) は、各化合物の ClogP をバーの高さによって表したものである。一方、図 9 (b) は、ClogP の代わりに分子量をバーの高さによって表したものである。このように、908 個もの化合物に関する非常に大規模なデータであるにもかかわらず、ClogP の値が小さいほど、また分子量が小さいほど水に対して溶けやすいことを視覚的に容易に捉えることができた。

30

【0131】

〔実施例 2〕

本発明の一実施例について以下に説明する。本実施例は、薬物代謝酵素チトクロム P 450 (CYP) による化合物 (薬物) の代謝パターンの分類に関するものであり、化合物の構造と代謝パターンとを関連付けて表示した。

40

【0132】

なお、CYP は、生体の物質代謝や異物解毒に関わる重要な薬物代謝酵素群の一つであり、ヒトでは約 20 種類の分子種が同定されている。生体にとって異物である薬物の多くは CYP による代謝を受けて消失するため、CYP による代謝は薬物の有効性や安全性と深く関わっているといえる。CYP スーパーファミリーの中でも、CYP1A2、CYP2C9、CYP2C19、CYP2D6、CYP3A4 の 5 つは薬物代謝に関わる主要な分子種であり、薬物代謝の約 90% がこれらによって説明できると言われている。したがって、これら 5 つの CYP による薬物の代謝を評価・予測することは創薬研究において非常に重要な課題となっている。

【0133】

本実施例では、第 2 の特性情報 (機能特性情報) として、ボナブリーらが調査した、1

50

61個の薬物についての、上記の5つのCYPによる代謝感受性のデータ(P. Bonnabry, et al., Quantitative drug interactions prediction system (Q-DIPS): a dynamic computer-based method to assist in the choice of clinically relevant in vivo studies. Clin. Pharmacokinet., 40: 631-640, 2001)を用いた。なお、CYPによる代謝感受性のデータは、3段階で評価されている。代謝感受性のデータは、化合物と対応付けられたテーブルとして取得した。実施例1と同様に標準化した後の代謝感受性のデータの一部を表8に示す。

【0134】

【表8】

ID	薬物名	代謝感受性				
		CYP1A2	CYP2C9	CYP2C19	CYP2D6	CYP3A4
1	aceclofenac	-0.354	2.299	-0.322	-0.587	-1.454
2	acenocoumarol	-0.354	2.299	-0.322	-0.587	-0.306
3	alfentanil	-0.354	-0.463	-0.322	-0.587	0.841
4	alprazolam	-0.354	-0.463	-0.322	-0.587	0.841
5	amiodarone	-0.354	-0.463	-0.322	-0.587	0.841
6	amitriptyline	-0.354	-0.463	1.754	1.868	0.841
7	amprenavir	-0.354	-0.463	-0.322	-0.587	0.841
8	astemizole	-0.354	-0.463	-0.322	-0.587	0.841
9	bromocriptine	-0.354	-0.463	-0.322	-0.587	0.841
10	buprenorphine	-0.354	-0.463	-0.322	-0.587	0.841
11	buspirone	-0.354	-0.463	-0.322	-0.587	0.841
12	caffeine	3.443	-0.463	-0.322	-0.587	-0.306
:	:	:	:	:	:	:

10

20

【0135】

また、第1の特性情報(構造情報)として、Molconn-Zによって計算される220個の分子記述子(Molconn-Zディスクリプタ)のうち、パス長が7以上のもの、及びデータセット内での分散が0となる分子記述子を除いた計115個の分子記述子を用いた。なお、各薬物のMolconn-Zディスクリプタは、Molconn-Z(商標、Hall Associate Consulting社製)を用いて化学構造から算出した。各化合物のMolconn-Zディスクリプタの一部を表9に示す。

30

【0136】

【表 9】

ID	薬物名	Molconn-Zディスクリプタ				
		nrings	Mw	x0	x1	...
1	aceclofenac	2	354.189	16.8196	10.9692	...
2	acenocoumarol	3	353.331	18.8446	12.3799	...
3	alfentanil	1	416.522	21.5623	14.5412	...
4	alprazolam	0	308.769	15.1041	10.7372	...
5	amiodarone	3	645.319	22.3801	14.9939	...
6	amitriptyline	4	277.408	14.6565	10.254	...
7	amprenavir	3	505.628	25.4241	16.6459	...
8	astemizole	3	458.577	23.3299	16.6742	...
9	bromocriptine	3	654.603	30.5847	20.3686	...
10	buprenorphine	5	467.647	23.8885	16.0166	...
11	buspirone	7	385.508	19.3991	13.6514	...
12	caffeine	7	194.193	10.4557	6.5366	...
:	:	:	:	:	:	:

10

【0137】

20

次に、本実施例では、各化合物の機能特性情報（代謝パターン）が5つの量的変数（5次元ベクトル）からなるため、化合物を機能特性情報に基づいて非階層的な機能クラスターに分類した。機能クラスターに分類するアルゴリズムには、コホネンの自己組織化マップ法（T. Kohonen, "Self-Organizing Maps", Springer, Berlin, Heidelberg, 1995）とk-means法とを組み合わせたものを用いた。これらについて以下に詳細に説明する。

【0138】

まず、各CYPによる代謝強度を属性とする5次元のパターンベクトルを考え、取得したCYPによる代謝感受性のデータを用いて自己組織化マップ（ $7 \times 9 = 63$ ノード）の学習を行った。学習によって得られた各ノードのパターンベクトルを示すテーブルの一部を表10に示す。

30

【0139】

【表 1 0】

ノード ID	ノードの x座標	ノードの y座標	代謝感受性				
			CYP1A2	CYP2C9	CYP2C19	CYP2D6	CYP3A4
1	1	1	-0.27654	1.7609	0.092761	0.018831	0.81583
2	1	2	-0.21332	1.199	1.0854	-0.24721	0.58301
3	1	3	-0.30101	0.61868	2.4121	-0.33353	0.16956
4	1	4	-0.27004	-0.05826	3.364	-0.43803	-0.28109
5	1	5	-0.05443	0.68401	2.5887	-0.5074	-0.53191
6	1	6	-0.03098	0.99163	1.4581	-0.54118	-0.8285
7	1	7	-0.13131	1.7935	0.26393	-0.57342	-1.0613
8	1	8	-0.24256	1.9349	-0.22058	-0.54454	-1.171
9	1	9	-0.24982	2.2411	-0.31235	-0.56888	-1.1455
10	2	1	-0.03221	1.2976	0.23421	-0.27541	0.78261
11	2	2	0.075406	0.54769	1.532	-0.42896	0.43891
12	2	3	-0.24352	0.049709	2.9809	-0.41496	-0.01958
:	:	:	:	:	:	:	:

10

【0 1 4 0】

20

また、上記の表 1 0 を可視化したものを図 1 0 に示す。図 1 0 では、自己組織化マップの各ノードが有するパターンベクトルの各成分を色相によって表示している。図 1 0 において、CYP3A4 に対応する要素については広範囲のノードが高い値を示すことから、多くの薬物が CYP3A4 によって代謝されることが視覚的に捉えられた。また、CYP1A2 と CYP2C9 あるいは CYP2C19 との間ではほとんどオーバーラップがないことから、CYP1A2 は CYP2C9 や CYP2C19 とは大きく異なった基質選択性を示すことが明らかとなった。

【0 1 4 1】

30

次に、自己組織化マップにおけるノード数は 6 3 個であり、機能クラスターとしてそのまま用いるには数が多いため、ノードを k - m e a n s 法によって最適なクラスター数にさらに分類することにした。ここで、最適なクラスター数を求める指標として、次式で表される Davies-Bouldin Index (D B) を用いた。

【0 1 4 2】

【数 2】

$$DB = \frac{1}{n} \sum_{i=1}^n \max_{i \neq j} \left\{ \frac{S_n(Q_i) + S_n(Q_j)}{S(Q_i, Q_j)} \right\}$$

【0 1 4 3】

40

ただし、 $S_n(Q_i)$ はクラスター Q_i におけるクラスター中心と各クラスター要素とのユークリッド距離の平均値を表し、 $S(Q_i, Q_j)$ はクラスター Q_i と Q_j の中心間のユークリッド距離を表す。

【0 1 4 4】

D B により、最適なクラスター数は 6 であることが求められた。

【0 1 4 5】

次に、自己組織化マップにおける 6 3 個のノードを k - m e a n s 法によって 6 つのクラスターに分類し、これを機能クラスターとした。これにより、自己組織化マップにおける各ノードが 6 つのうちどのクラスターに属するかを示すテーブルが得られた。そのテーブルの一部を表 1 1 に示す。

【0 1 4 6】

50

【表 1 1】

ノードID	クラスターID
1	2
2	2
3	4
4	4
5	4
6	2
7	6
8	6
9	6
10	2
11	2
12	4
:	:

10

【0 1 4 7】

また、表 1 1 を可視化したものを図 1 1 に示す。図 1 1 において、各ノードが所属する

20

クラスターは、色相によって表現されている。

【0 1 4 8】

次に、各薬物の代謝のパターンベクトルと自己組織化マップの各ノードのパターンベクトルとの距離から薬物の帰属するノード（ベストマッチノード）を決定し、そのノードが上記の 6 つのクラスターの何れに帰属するかに応じて、薬物を機能クラスターに分類した。表 1 2 は各機能クラスターに分類された薬物の CYP による代謝感受性の平均スコアを示したものである。

【0 1 4 9】

【表 1 2】

機能クラスターID	薬物数	CYP代謝感受性の平均スコア				
		CYP1A2	CYP2C9	CYP2C19	CYP2D6	CYP3A4
1	11	0.273	1.27	0.545	0.364	1.64
2	8	0	0	2	0.125	1
3	26	0.038	1.54	0	0.077	0.192
4	76	0	0	0	0.066	1.96
5	26	0.077	0	0.038	2	0.462
6	14	1.71	0	0.143	0.929	0.857

30

【0 1 5 0】

表 1 2 から、機能クラスター 1 は CYP3A4 及び CYP2C9 による代謝が、機能クラスター 2 は CYP2C19 による代謝が有意に高く、機能クラスター 3 は CYP2C9 でのみ、機能クラスター 4 は CYP3A4 でのみ、機能クラスター 5 は CYP2D6 によって、機能クラスター 6 は CYP1A2 によって代謝される薬物の集合であることがわかる。

40

【0 1 5 1】

次に、構造情報に基づいて化合物を階層的に分類するために、二分木を作成した。二分木を作成するにあたって、説明変数として Molconn-Z ディスクリプタ、目的変数として機能クラスターを用い、分岐ルールには次に示す情報利得 G (T.M. Mitchell, "Machine Learning", McGraw-Hill, Singapore, 1997) を用いた。

【0 1 5 2】

50

【数 3】

$$G = -\sum_{i=1}^N \frac{c_i}{s} \log_2 \frac{c_i}{s} - \left(-\frac{s_1}{s} \sum_{i=1}^N \frac{c_{1,i}}{s_1} \log_2 \frac{c_{1,i}}{s_1} - \frac{s_2}{s} \sum_{i=1}^N \frac{c_{2,i}}{s_2} \log_2 \frac{c_{2,i}}{s_2} \right)$$

【0153】

ただし、Nは薬物を機能クラスターに分類した際のクラスター数 (= 6) を示し、sは分割前の集合における薬物の総数を、 c_i は機能クラスター*i*に属する薬物数を示し、 s_1 及び s_2 は二分割されたそれぞれの集合における薬物の総数を、 $c_{1,i}$ 及び $c_{2,i}$ は二分割されたそれぞれの集合における、機能クラスター*i*に属する薬物数を示す。

10

【0154】

ここで、二分木の各分割ステップにおいて、最大の情報利得Gが得られる条件(分割の判定に用いるMolconn-Zディスクリプタの種類及び閾値)を決定し、得られる最大の情報利得Gの値が0.25以下になるまで再帰的に分割を繰り返した。二分木の作成によって得られたデータの一部を表13に示す。

【0155】

【表13】

ノードID	分割に用いた Molconn-Z ディスクリプタ	分割に用い た閾値	第1子ノード ID	第2子ノード ID	薬物ID
1 (ルート)	SHCsats	4.80369	3	2	1,2,3,...,161
2	SHHBA	31.98136	31	30	1,2,4,6,...,161
3	SHother	16.7977	5	4	3,5,...,160
4	Gmin	-0.8893	15	14	5,11,...,152
:	:	:	:	:	:
30	SHHBA	19.61258	55	54	4,6,...,160
:	:	:	:	:	:
54	totop	137.0427	69	68	6,13,...,150
:	:	:	:	:	:
68	xp6	3.638	77	76	16,31,...,134
:	:	:	:	:	:
77	null	null	null	null	16,31,117
:	:	:	:	:	:

20

30

【0156】

表13に示すように、ルートノード(ノード1)では、Molconn-ZディスクリプタのSHCsatsが4.80369未満か否かで薬物を2つの子ノード(ノード2及びノード3)に分割している。そして、同様に、ノード2では、Molconn-ZディスクリプタのSHHBAが31.98136未満か否かで薬物を2つの子ノード(ノード30及びノード31)に分割している。

40

【0157】

次に、階層的に分類した薬物について、階層的なクラスターに基づいた再帰的な入れ子構造となるように、2次元平面内における配置を決定した。この配置を決定するアルゴリズムには、平安京ビューを用いた。

【0158】

そして、161個の薬物について、決定された配置で、機能クラスターの種類を色相とした色でアイコンを描画した画像データを作成した。この画像データに基づく画像を図12に示す。なお、画像中において、薬物を示すアイコンはカラー表示されており、アイコ

50

ンの色は、代謝パターンに基づいて分類した際に帰属した機能クラスターに応じた色となっている。

【0159】

図12に示すように、表示された画像では、共通の構造的特徴を有している薬物が同じ代謝パターンを示すように分類されている。このように、161個もの大規模な数の薬物について、構造と代謝パターンとの関係を効果的に視覚化できることが示された。

【0160】

また、図13は、チトクロムP450の各分子種CYP1A2、CYP2C9、CYP2C19、CYP2D6、CYP3A4による代謝感受性を個別の画像で示したものである。画像中において、代謝感受性（3段階）はアイコンの色で表現されている。このように、表示された画像を一望するだけで各薬物代謝酵素の感受性を包括的に理解することができ、本発明の有用性が示された。

10

【0161】

なお、本発明は上述した実施形態及び実施例のみに限定されるものではなく、請求項に示した範囲で種々の変更が可能であり、異なる実施形態及び実施例にそれぞれ開示された技術的手段を適宜組み合わせ得られる実施形態についても本発明の技術的範囲に含まれる。

【産業上の利用可能性】

【0162】

本発明は、スクリーニングデータの全体像を効果的に視覚化できるので、候補化合物の選択を支援するとともに、機能的にさらに優れた化合物の合成展開を行う方向性を示すことができる。したがって、医薬品、農薬に限らず、その他の機能性化学物質の探索研究を行う化学工業分野一般において利用可能であり、非常に有用性が高いと考えられる。

20

【図面の簡単な説明】

【0163】

【図1】本発明の一実施形態を示すものであり、化合物群表示装置の機能ブロック図である。

【図2】本発明の一実施形態を示すものであり、化合物群表示装置の処理工程を示す工程図である。

【図3】階層的な分類によって得られたデンドログラムの一例を示す図である。

30

【図4】化合物配置部による化合物の配置の一例を示す図である。

【図5】画像データ生成部によって生成された画像データの一例を示す図である。

【図6】本発明の別の実施形態を示すものであり、化合物群表示装置の機能ブロック図である。

【図7】本発明の別の実施形態を示すものであり、化合物群表示装置の処理工程を示す工程図である。

【図8】本発明の一実施例を示すものであり、化合物群表示装置によって表示された画像を示す図である。

【図9】本発明の一実施例を示すものであり、化合物群表示装置によって表示された画像を示す図である。

40

【図10】本発明の別の実施例を示すものであり、自己組織化マップにおける各ノードのパターンベクトルの各成分を色相で示した図である。

【図11】本発明の別の実施例を示すものであり、自己組織化マップにおける各ノードが帰属する機能クラスターを色相で示した図である。

【図12】本発明の別の実施例を示すものであり、化合物群表示装置によって表示された画像を示す図である。

【図13】本発明の別の実施例を示すものであり、化合物群表示装置によって表示された画像を示す図である。

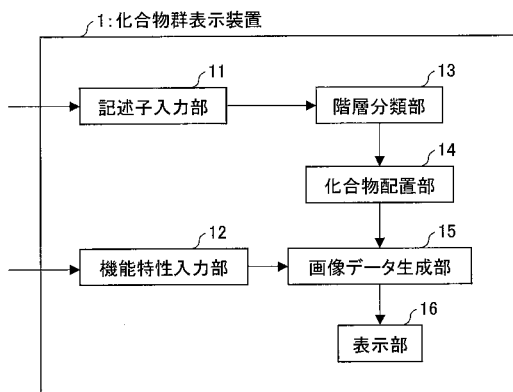
【符号の説明】

【0164】

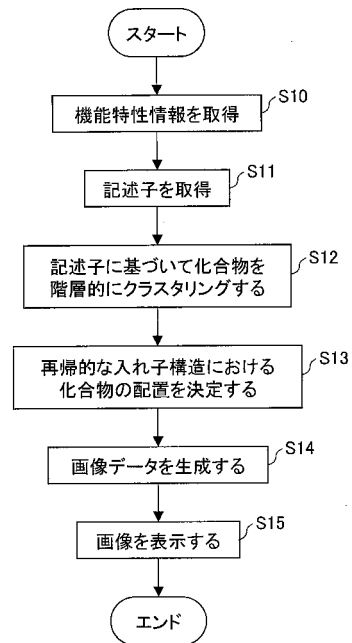
50

- 1, 2 化合物群表示装置
- 13, 23 階層分類部 (階層分類手段)
- 14 化合物配置部 (化合物配置手段)
- 15, 25 画像データ生成部 (画像データ生成手段)
- 16 表示部

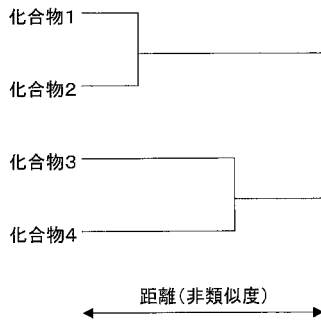
【図1】



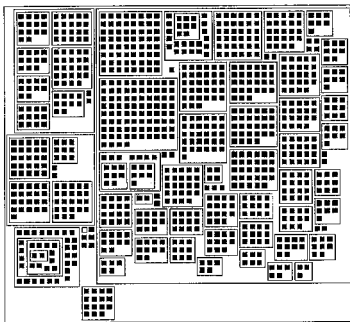
【図2】



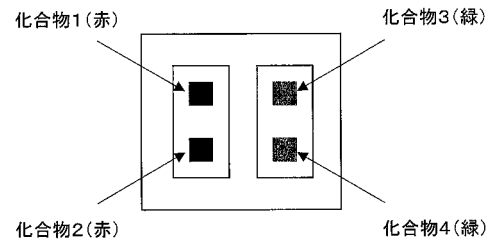
【 図 3 】



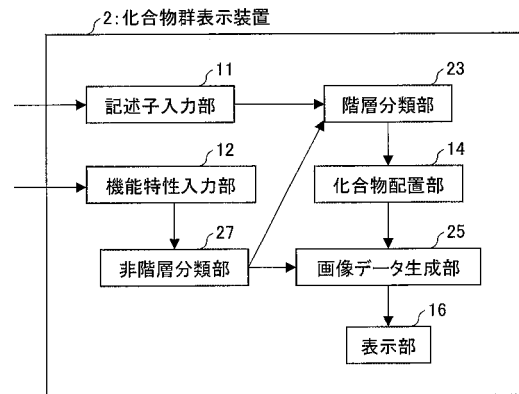
【 図 4 】



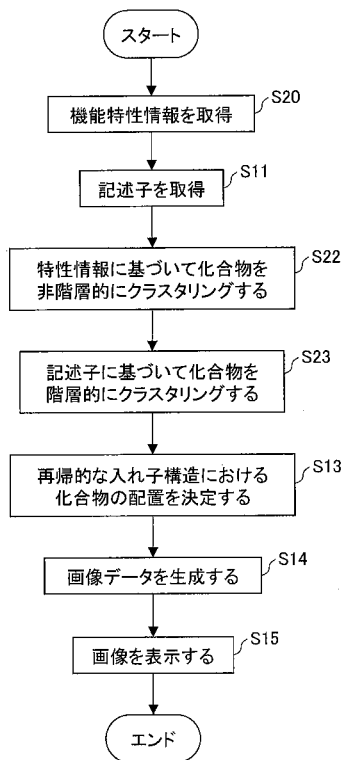
【 図 5 】



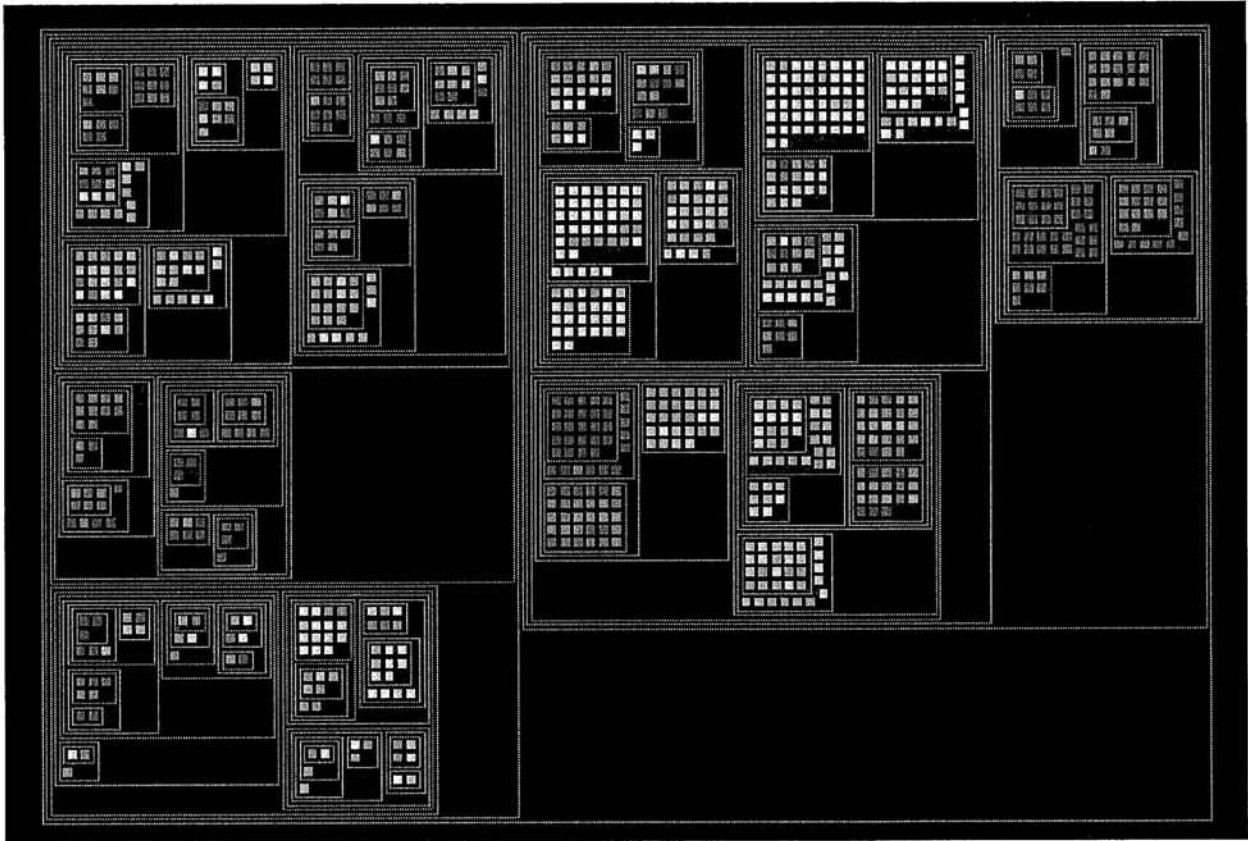
【 図 6 】



【 図 7 】

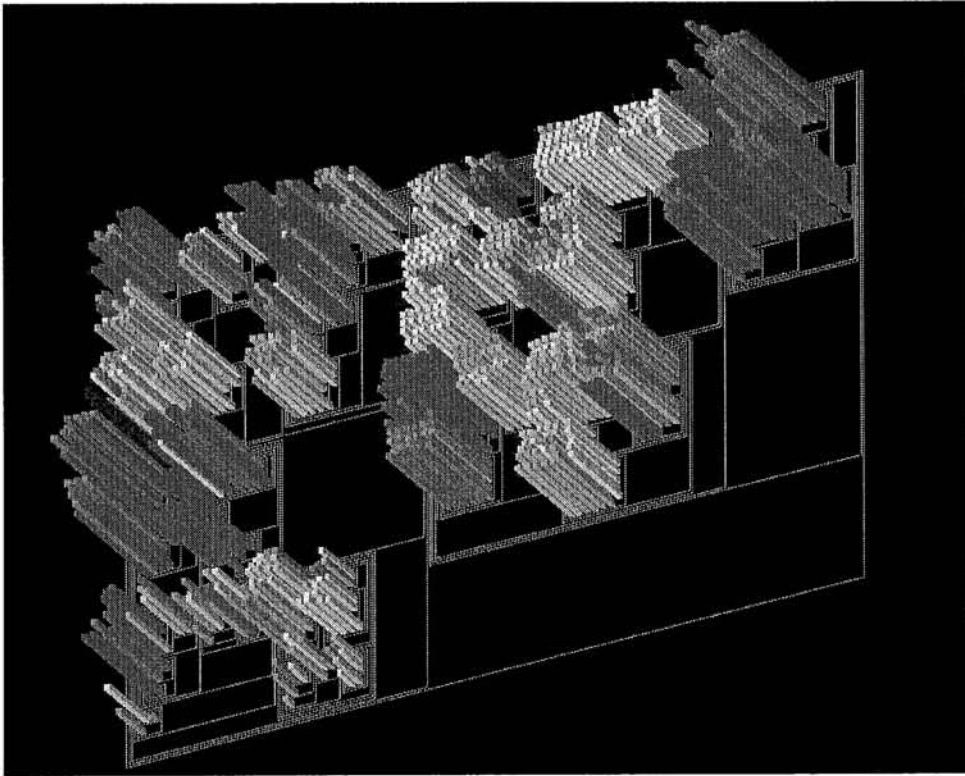


【 図 8 】

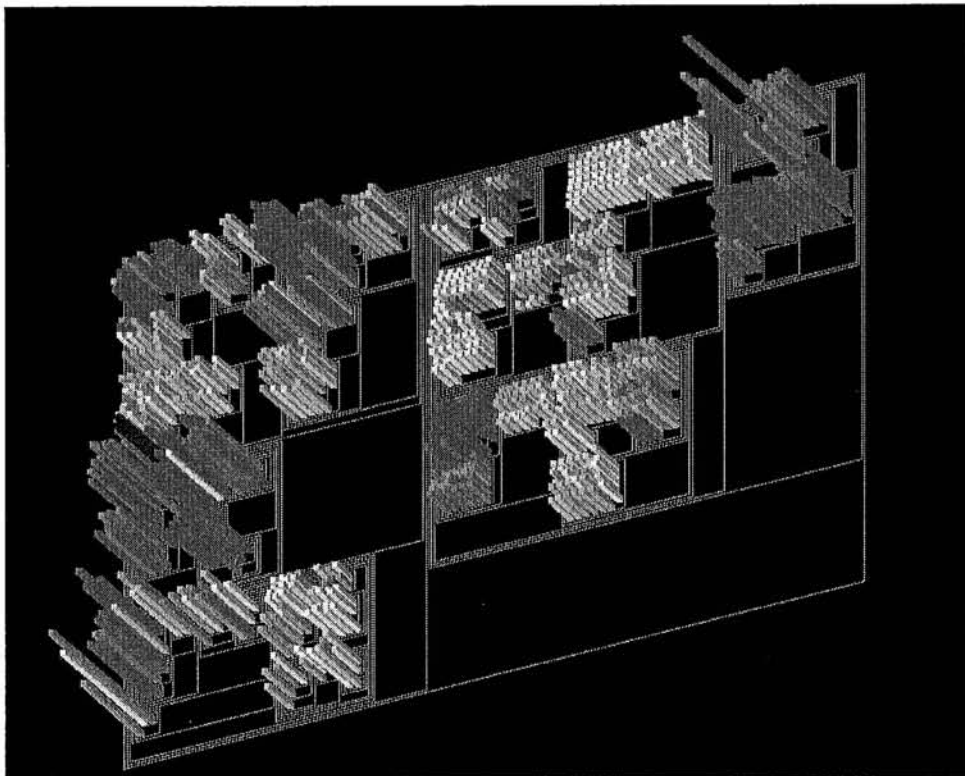


【 図 9 】

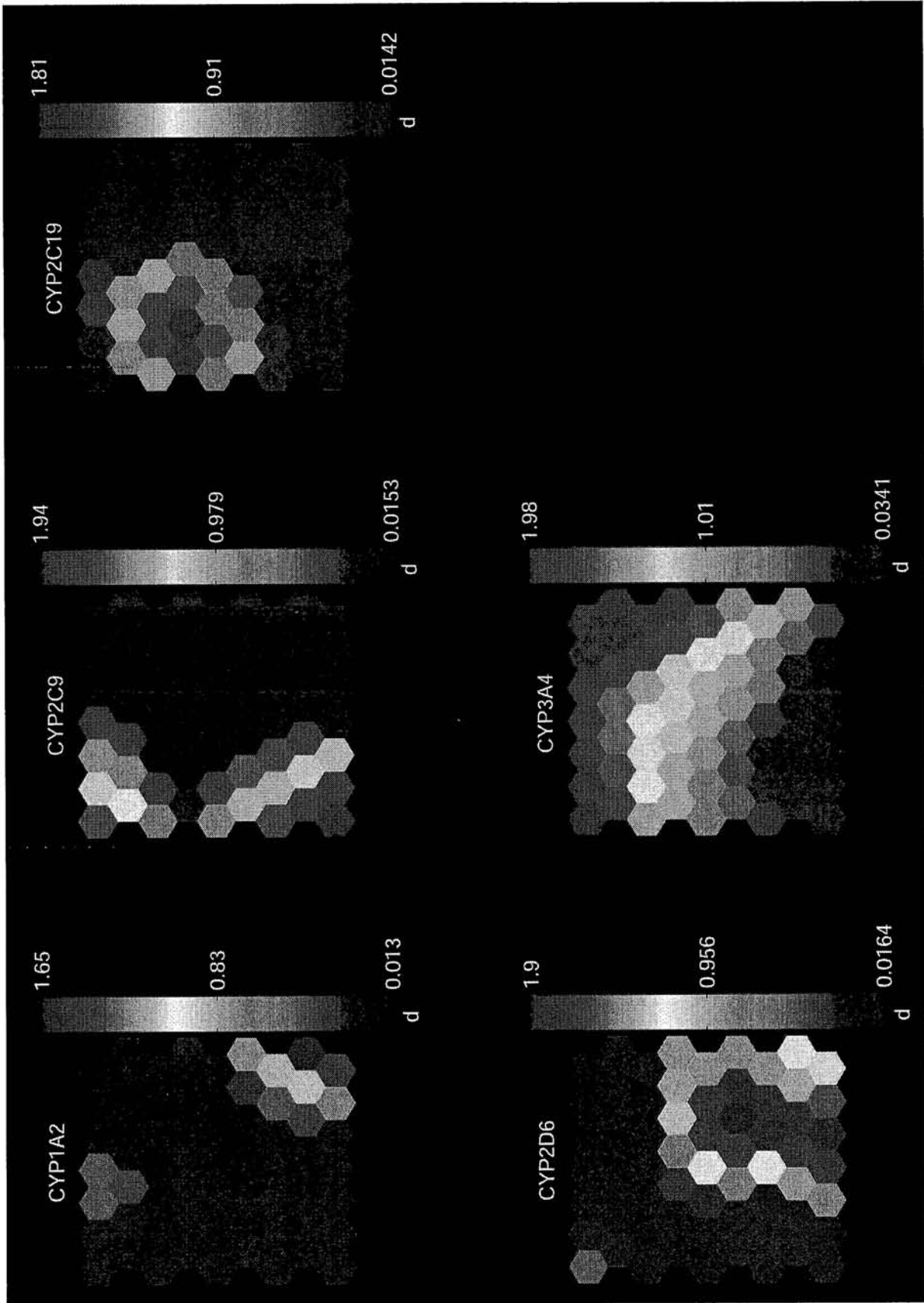
(a)



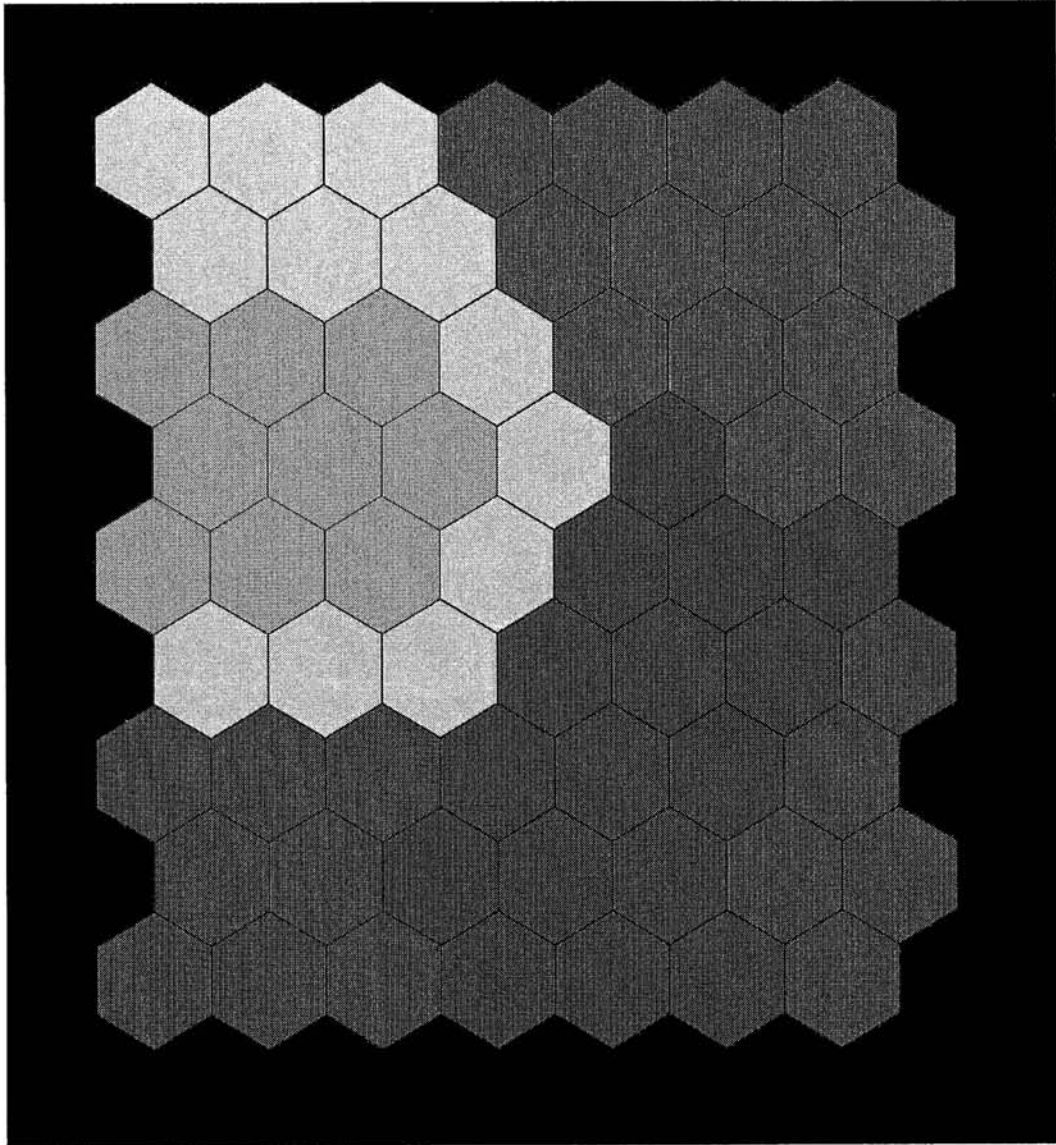
(b)



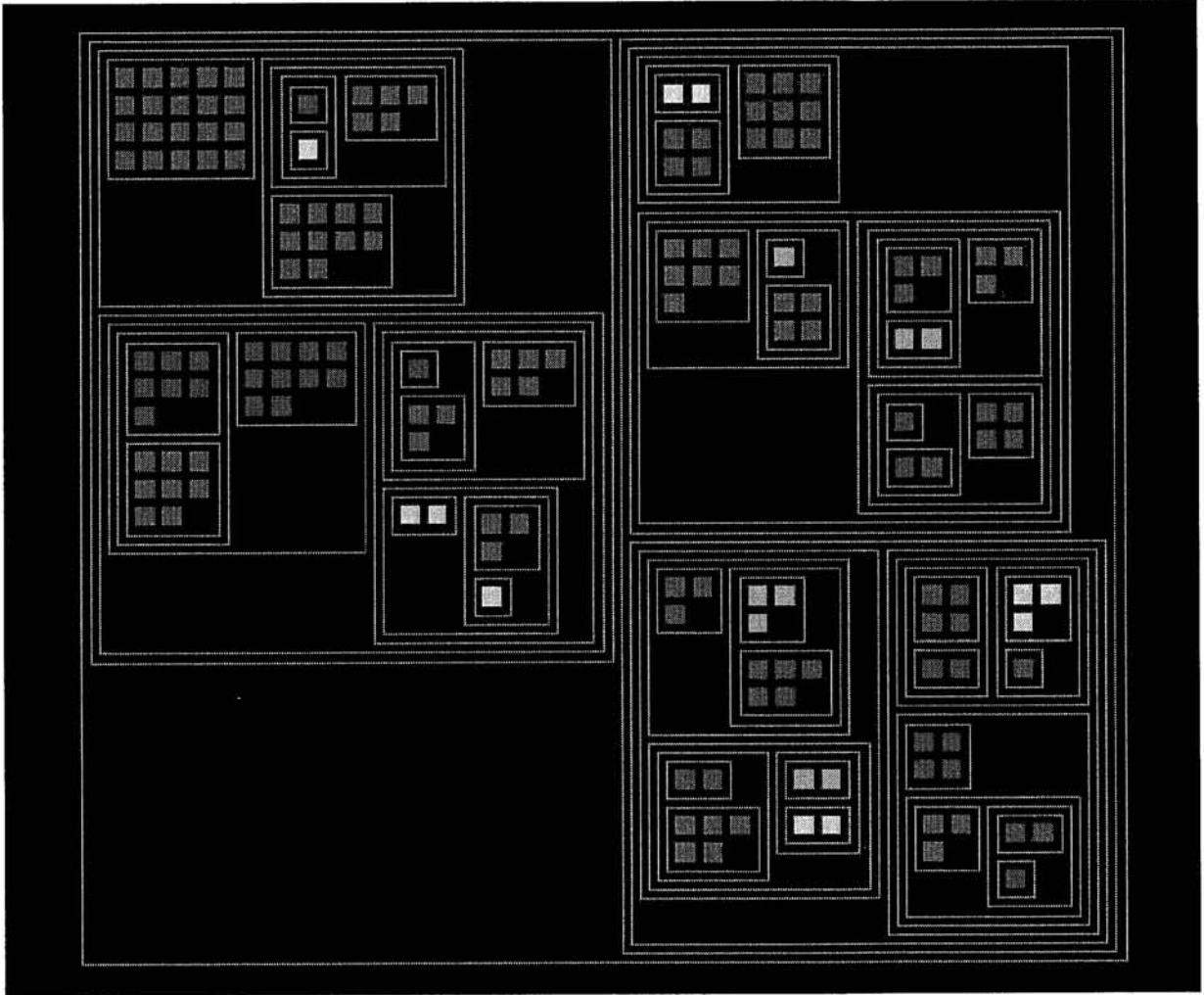
【 図 1 0 】



【 図 1 1 】



【 図 1 2 】



【 図 1 3 】

