

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第3607948号
(P3607948)

(45) 発行日 平成17年1月5日(2005.1.5)

(24) 登録日 平成16年10月22日(2004.10.22)

(51) Int. Cl.⁷

F I

G O 1 R 33/32

G O 1 N 24/02 5 3 O M

G O 6 T 17/40

G O 6 T 17/40 A

請求項の数 7 (全 18 頁)

(21) 出願番号	特願2001-333904 (P2001-333904)	(73) 特許権者	504173471
(22) 出願日	平成13年10月31日(2001.10.31)		国立大学法人 北海道大学
(65) 公開番号	特開2003-139832 (P2003-139832A)		北海道札幌市北区北8条西5丁目8番地
(43) 公開日	平成15年5月14日(2003.5.14)	(74) 代理人	100105371
審査請求日	平成13年10月31日(2001.10.31)		弁理士 加古 進
		(72) 発明者	稲垣 冬彦
			北海道札幌市北区北19条西三丁目21-136 サザンパレス北大402
		(72) 発明者	横地 政志
			北海道札幌市北区北21条西八丁目1-8 クィーンパレス21-302
		審査官	田中 洋介

最終頁に続く

(54) 【発明の名称】 NMRシグナル帰属支援システム

(57) 【特許請求の範囲】

【請求項1】

少なくとも、アミノ酸配列、2次元スペクトル、複数の3次元スペクトルを用いて行うNMRシグナル帰属支援システムであって、
前記アミノ酸配列を記録しているアミノ酸配列記録手段及びタンパク質のスペクトルを記録しているスペクトル記録手段とを有し、

前記アミノ酸配列記録手段から、アミノ酸配列の読み込みを指示するアミノ酸配列読み込み指示入力手段と、

前記スペクトル記録手段から前記2次元及び3次元スペクトルの読み込みを指示するスペクトル読み込み指示入力手段と、

前記2次元スペクトルを表示する2次元スペクトル表示手段と、

前記2次元スペクトル表示手段に表示されたピークが指定されることで、対応する前記複数の3次元スペクトルを表示する3次元スペクトル表示手段と、

表示された前記2次元スペクトル又は前記3次元スペクトルを用いて前記ピークが帰属するボンド・コンネクティビティを関連づけるボンド・コンネクティビティの指示を入力するボンド・コンネクティビティ指示入力手段と、

前記ボンド・コンネクティビティの原子型を指示する原子型指示入力手段と、

前記ボンド・コンネクティビティの一意に連続したものをまとめてクラスタとし、アミノ酸配列の候補と前記作成されたクラスタとを、両者の関連性の確率を計算し、計算された確率により対応させて表示するクラスタ表示手段と、

10

20

表示された前記アミノ酸配列の候補と、前記クラスタの両者を関連づけるクラスタ帰属指示を入力するクラスタ帰属指示入力手段と、
前記クラスタに含まれないピークをフラグメントとして検出し、前記アミノ酸配列の候補と対応させて表示するフラグメント表示手段と、

表示された前記アミノ酸配列の候補と前記フラグメントの両者を関連づけるフラグメントの帰属の指示を入力するフラグメント帰属指示入力手段と、

前記入力された指示をコマンドとして、コマンド記録手段に帰属過程を記録するログ手段とを備え、

帰属の支援を行うとともに帰属過程を記録することを特徴とするNMRシグナル帰属支援システム。

10

【請求項2】

請求項1に記載のNMRシグナル帰属支援システムにおいて、
さらに、前記コマンド記録手段に記録したコマンドを再度実行することにより、帰属過程を再現可能とするコマンド実行手段を備えることを特徴とするNMRシグナル帰属支援システム。

【請求項3】

請求項1又は2に記載のNMRシグナル帰属支援システムにおいて、
前記原子型指示入力手段は、予め設定したパターンを用いて自動的に原子型を帰属する自動帰属処理手段を含むことを特徴とするNMRシグナル帰属支援システム。

【請求項4】

請求項1～3のいずれかに記載のNMRシグナル帰属支援システムにおいて、前記クラスタ表示手段は、前記アミノ酸配列の候補と前記作成されたクラスタとの関連性の確率を、化学シフトに依存した項とクラスタ両末端の化学シフトの同一性に関わる項から計算して表示することを特徴とするNMRシグナル帰属支援システム。

20

【請求項5】

請求項4に記載のNMRシグナル帰属支援システムにおいて、前記クラスタ表示手段は、前記アミノ酸配列の候補と前記作成されたクラスタとの帰属指示が行われると、確率を計算し直して表示することを特徴とするNMRシグナル帰属支援システム。

【請求項6】

請求項1～5のいずれかに記載のNMRシグナル帰属支援システムの機能をコンピュータ・システムに実現させるプログラム。

30

【請求項7】

請求項1～5のいずれかに記載のNMRシグナル帰属支援システムの機能をコンピュータ・システムに実現させるプログラムを記録した記録媒体。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

本発明は、NMRによるタンパク質の立体構造の解析に関するものであり、特に、立体構造の解析に必要なNMRシグナルの帰属に対する支援システムに関するものである。

【0002】

【技術的背景】

核磁気共鳴法(NMR)は、溶液中におけるタンパク質の原子レベルにおける静的・動的立体構造を決定できる唯一の分光法である。いまや、NMRによって決定されたタンパク質の立体構造をもとに、さまざまな生命現象を分子認識の観点から説明することが可能になり、NMRは構造生物学を支える重要な柱のひとつになった。

40

【0003】

NMRを用いたタンパク質の立体構造解析の手順は以下のようになる。

(1) NMR測定

NMR測定には、1mM程度の水溶液(90% H₂O, 10% D₂O)を400μl用いる。測定対象によっては不規則な会合体を形成して、試料溶液が白濁することもある。こ

50

のような状態は、のちのNMR測定および解析に多大な障害をもたらすので、測定試料のpH、イオン強度などを変化させ至適測定条件を見つけだす。

(2) NMRシグナルの帰属

検出されたNMRシグナルがタンパク質中のどの原子核由来であるかを明らかにすることをNMRシグナルの帰属とよぶ。このステップはタンパク質の立体構造決定においてきわめて重要であり、慎重に行なうべきである。シグナル帰属法に関してはいくつかの2次元NMRをはじめとして3次元、4次元NMRスペクトルを使用する組織だった戦略が確立されている。

(3) 構造情報の収集

NMRスペクトルの中には、核オーバーハウザ - 効果 (NOE)、スピン結合などタンパク質の構造決定に有用な情報が多数含まれている。NMRスペクトル多次元表示することにより、たとえばプロトン核間の距離情報を反映したNOE交差ピークを一望に鳥瞰することができ、シグナル帰属に基づき溶液中のタンパク質のプロトン間距離を見積もることができる。

(4) 立体構造計算

3次元空間に存在する点の座標が与えられるならば、その座標からある2点間の距離を算出することは簡単なことである。ではこの逆、すなわち2点間の距離からそれぞれの点の3次元座標を求めることは可能であろうか？答えは'可'であり、ある立体を構成する点間の距離からそれぞれの点の座標を求める計算をディスタンス・ジオメトリー計算とよんでいる。したがって、NMRによるタンパク質の立体構造計算とは、ワークステーションを用いて、NOEから得られた距離情報などをもとにディスタンス・ジオメトリー法により立件構造を計算することである。X線結晶構造解析と異なり、NMRによるタンパク質の立体構造は近距離 (5 Å以内) 情報の積み重ねにより決定されている。実際の計算に際しては、共有結合周りの回転角度である2面角に関する情報など、できるかぎりの構造情報を利用してはいる。

【0004】

NMR法は、水溶液におけるタンパク質の立体構造を明らかにできる点で、構造生物学の重要な解析手法である。この解析手法では、上述したように、シグナル帰属を正確に行うことが必要である。シグナル帰属では、主鎖のシグナル帰属、側鎖のシグナル帰属、距離情報を得るためのNOEの帰属と、10 - 20種類以上におよぶスペクトルを組み合わせ

て解析し、必要な情報を取り出すことが必要である。図1にその例を示す。図1(a)に、66残基からなる2量体タンパク質の1次元スペクトルのアミドプロトン - 芳香環領域を示す。数十のアミドプロトン・シグナルが重なり合い、解析はほとんど不可能である。同じ領域を1H - 15N 2次元HSQC (図1(b)参照) で観測すると、アミドプロトン・シグナルはアミド窒素の化学シフト軸に展開され、シグナルの縮重が軽減されることがわかる。さらに、図1(c)に示すように、3次元HNCOスペクトルでは、アミドプロトン (H) 軸、アミド窒素 (N) 軸に加え、カルボニル炭素 (CO) 軸の導入によりシグナルの完全分離が達成される。

この帰属過程は複雑であり、スペクトルの組合せ等、研究者の経験や直感に依存する点が多い。従来の方法では、帰属のプロセスを振り返って変更したり、帰属に至る過程を他の研究者が再現することは困難である。

【0005】

【発明が解決しようとする課題】

本発明の目的は、NMRシグナルの正確な帰属を行うための支援システムを提供することである。

【0006】

【課題を解決するための手段】

上記の目的を達成するために、本発明は、少なくとも、アミノ酸配列、2次元スペクトル、複数の3次元スペクトルを用いて行うNMRシグナル帰属支援システムであって、前記アミノ酸配列を記録しているアミノ酸配列記録手段及びタンパク質のスペクトルを記録し

10

20

30

40

50

ているスペクトル記録手段とを有し、前記アミノ酸配列記録手段から、アミノ酸配列の読み込みを指示するアミノ酸配列読み込み指示入力手段と、前記スペクトル記録手段から前記2次元及び3次元スペクトルの読み込みを指示するスペクトル読み込み指示入力手段と、前記2次元スペクトルを表示する2次元スペクトル表示手段と、前記2次元スペクトル表示手段に表示されたピークが指定されることで、対応する前記複数の3次元スペクトルを表示する3次元スペクトル表示手段と、表示された前記2次元スペクトル又は前記3次元スペクトルを用いて前記ピークが帰属するボンド・コネクティビティを関連づけるボンド・コネクティビティの指示を入力するボンド・コネクティビティ指示入力手段と、前記ボンド・コネクティビティの原子型を指示する原子型指示入力手段と、前記ボンド・コネクティビティの一意に連続したものをまとめてクラスタとし、アミノ酸配列の候補と前記作成されたクラスタとを、両者の関連性の確率を計算し、計算された確率により対応させて表示するクラスタ表示手段と、表示された前記アミノ酸配列の候補と、前記クラスタの両者を関連づけるクラスタ帰属指示を入力するクラスタ帰属指示入力手段と、前記クラスタに含まれないピークをフラグメントとして検出し、前記アミノ酸配列の候補と対応させて表示するフラグメント表示手段と、表示された前記アミノ酸配列の候補と前記フラグメントの両者を関連づけるフラグメントの帰属の指示を入力するフラグメント帰属指示入力手段と、前記入力された指示をコマンドとして、コマンド記録手段に帰属過程を記録するログ手段とを備え、帰属の支援を行うとともに帰属過程を記録することを特徴とする。

10

【0007】

さらに、前記コマンド記録手段に記録したコマンドを再度実行することにより、帰属過程を再現可能とするコマンド実行手段を備えてもよい。

20

前記原子型指示入力手段は、予め設定したパターンを用いて自動的に原子型を帰属する自動帰属処理手段を含み、自動的に原子型を関連付けてもよい。

前記クラスタ表示手段は、前記アミノ酸配列の候補と前記作成されたクラスタとの関連性の確率を、化学シフトに依存した項とクラスタ両末端の化学シフトの同一性に関わる項から計算して表示してもよい。このとき、前記アミノ酸配列の候補と前記作成されたクラスタとの帰属指示が行われると、確率を計算し直して表示し、帰属が確定することに、判断が容易になるようにすることもできる。

また、上述のNMRシグナル帰属支援システムの機能をコンピュータ・システムに実現させるプログラムやプログラムを格納した記録媒体も本発明である。

30

【0008】

【発明の実施の形態】

以下、図面を参照して、本発明の実施形態であるNMRシグナル帰属支援システムを説明する。

さて、NMRシグナル帰属支援システムにおいて、帰属の操作を行うためには、

(1) 帰属の過程と帰属に必要な一連のスペクトルをリンクさせて帰属過程がスペクトル上で確認できること、

(2) 帰属の変更を必要とする場合は、スペクトル上で連動して変更できることが必要である。また、

(3) 帰属の確度を確かめるため、データベースを参照し、最も確度の高い帰属を確認しながら仕事を進めていくこと

40

も重要である。

【0009】

パソコン等のコンピュータ・システムに実装した実施形態のNMRシグナル帰属支援システムにおける処理の流れを、図2のフローチャートに基づき、詳しく説明する。

<帰属操作>

本支援システムを起動すると、図3のようなウインドウ200が表示される。本支援システムにおいては、化学シフトデータベースBioMagResBank (The BMRB Database) を利用して予測、診断を行っている。このため、このデータベースに行くためのボタン282がある。

50

ここで、画面280内の「OK」ボタン281を押してから、ウィンドウ200上部の「ファイル」210をクリックすることにより表示されるメニューを選択して、本支援システムで使用するファイルの読み込みを行う。

まず、試料のアミノ酸配列のファイルを読み込む(S102)。マルチチェーンの場合には鎖の数を入力して、アミノ酸配列ファイルが続けて読み込む。アミノ酸配列ファイルとして、一文字表記のアミノ酸一次配列のテキスト・ファイルを用いている。

このときに、読み込んだアミノ酸配列ファイルの名称(試料名)を基に、ログ・ファイルも自動的に作成される(S104)。このログファイルは、本支援システムにおける全ての帰属に関する操作が記録される。このログファイルを使って、前回の帰属状態を回復させることができる。このログファイルについては後で説明する。

10

【0010】

次に、スペクトル・ファイルとして、フーリエ変換ファイルおよびピークピック・ファイルを読み込む(S106)。フーリエ変換ファイルに対応するピークピック・ファイルを同時に指定して読み込ませる必要がある。3次元以上のスペクトル・ファイルで複数に分かれている場合も、どれか一つ指定すれだけでよい。フーリエ変換したスペクトルだけの読み込みは許されないので、事前にピークテーブル・ファイルを作っておく必要がある。本システムでは、スペクトル・ファイルとしては、NMRPipeでフーリエ変換処理したファイルおよびNMRDrawで作成したピークテーブル・ファイルを用いている。読みとったスペクトル・ファイルごとに、実験名を定義する(S108)。これにより、読みとったスペクトル・ファイルの表示名、軸名等を指定する。

20

なお、初期設定として、主鎖連鎖帰属のための実験である、HN-HSQC, HNCOC, HNCACB, HN(CO)CA, HNCACB, CBCA(CO)NH等がメニューに登録されており、これらに対するスペクトル・ファイルの読み込み、実験名の定義が容易にできる。

【0011】

NMRのスペクトルの帰属に必要なファイルを全て読み込む(S110でYES)と、帰属の操作を開始する。主鎖の帰属を行うため、まず、HSQCスペクトルのシグナルをピックアップする。そのために、図3において、「contour-2D」220をクリックして、実験名を指定し、読み込んだHN-HSQCの2次元ファイルを表示する(S112)。これを示したのが図4である。

30

図4に示した様に、本支援システムでは、HN-HSQCの2次元スペクトル300は、S/N分析、双3次補間、双曲線補間を用いて、常に適当な閾値が調整された等高線でスペクトルを表示している。各ピーク位置は、十字により表示されている。2次元スペクトル300には、実験名310, 軸名320と330も表示されている。また、各ピークに対しては、読み込んだピークピック・ファイルにおけるピーク番号(インデックス)が付与されている。

表示される2次元スペクトルは9つに分割して、ウィンドウ340で選択することで、分割されたそれぞれを拡大して表示できる。また、範囲を指定して、指定された範囲を拡大して表示することもできる。図4は拡大された2次元スペクトルである。

【0012】

40

重なり合ったり、肩になったピークについては、HN(CO)CA等の3次元スペクトルを投影して、ピーク位置を確定する。この操作は後で説明する。この時にも、インデックスが付与される。この様にして、HSQCスペクトル上に観測される全てのピークについてインデックス付けを行なう(例えば図4の360参照)。以降の帰属過程はこのピーク番号(インデックス)に基づいて行う。

ピーク情報には最大24文字までのメモを残すことができる。これは、ウィンドウ350内のテキスト・ボックス354に書き込むことで作成される。1つのピークには最大10個のメモを残すことができる。それらは日付管理され、通常最新のメモは強制的にスペクトル上に表示される。

ピークに与えられる帰属ラベルは、3次元実験との連結の確定、さらに帰属を確定するこ

50

とで、"アミノ酸一文字表記+残基番号"形式で表示されるようになる。リンクが確定しないうちは、何も表示されない。ラベル表示色については、以下の規則がある。

灰色：他の3次元実験と連結なし

黄色：3次元実験との連結があり、帰属は未確定

ピンク：帰属が確定

【0013】

<主鎖の帰属>

さて、主鎖連鎖帰属の作業手順は、

(a) ボンド・コネクティビティ (Bond Connectivity) の確立作業

(b) クラスタ・コネクティビティ (Cluster Connectivity) およ 10

びクラスタ・タグ (Cluster-Tag) の確立

(c) クラスタ・タグ (Cluster-Tag) の帰属

(d) フラグメント (Fragment) の帰属

(e) 帰属の評価

の5段階に分けられている。

ここで、ボンド・コネクティビティとは、参照する2次元実験のピークテーブルの化学シフト値で3次元実験の領域を切り出し、その面上にある3次元ピークと2次元ピークとを関連づける情報である。クラスタ・コネクティビティとは、2つのボンド・コネクティビティ間で、前後の化学シフトがユニークに一致した情報である。クラスタ・タグとは、複数のクラスタ・コネクティビティが一意に連続したものの集まりに、順番にタグ番号を与えて管理した情報である。クラスタ・タグは、インデックスのシーケンスとして表現される。フラグメント (Fragment) とは、クラスタ・タグに含まれなかった全てのボンド・コネクティビティのことである。 20

【0014】

(a) ボンド・コネクティビティ (原子型帰属) の確立

主鎖帰属でまず最初に行うことは、3次元実験を2次元実験の化学シフトの軸でスライスした面を切り出した上で、原子型帰属ボンド・コネクティビティ (原子型帰属) を確定することである。これを行うために、3次元のスペクトルを表示する必要がある。

このため、図4で、「Main Chain Assignment」230から、メニューを選択して、図5に示すウインドウ370を開き、インデックス入力欄372に、インデックスを入力することで、ピークを指定する (S114)。これにより、図6に示すように、指定したインデックスに対応する3次元スペクトル画面400が表示される (S116)。 30

画面400には、指定インデックスのシグナルに対応するストリップ・チャートである、HNCO410, HNCA420, HN(CO)CA430, CBCANH440, CBCA(CO)NH450が、各スペクトル・ファイルから自動的に切り出されて、表示されている。この表示されているストリップは、図5のウインドウ370内のボタン類376により、指定インデックスの化学シフトから、各軸に対して少しずつ切り出し面をずらして表示することができる。これらのボタン類376は、2次元実験で複数のシグナルが接近している場合に、3次元実験上の対応するシグナルを区別するために使用する。なお、切り出し面をずらすことは、各ストリップごとでもできるし、全ストリップ同時に行うこともできる。 40

【0015】

図6に示すように、表示された3次元のHNCAスペクトルでは、アミドプロトン、アミド窒素の化学シフト (HN, N) と同じ残基のカーボンCA(i)と前の残基のカーボンCA(i-1)との相関が表示される。同様に、HN(CO)CAでは (HN, N) とCA(i-1)、CBCANHでは (HN, N) とCA(i-1)、CB(i-1)、CA(i)、CB(i)の相関が、CACB(CO)NHでは (HN, N) とCA(i-1)、CB(i-1)との相関が表示される。

ここで、シグナルの示す強弱、正負の違いから、また、必要ならば隣接したスライスでシ 50

グナルの増減を確認するなどして、参照実験上で重なり合ったシグナルを完全に区別する。そして、インデックスと関連のある3次元ピークをクリックして、ウィンドウ470を表示し、「Create Bond Connectivity」ボタン472をクリックすることによって、ボンド・コネクティビティを確定する。この確定の後、炭素原子の型を、例えば、「Assign Atom Type」ボタン476をクリックして、メニュー480から、CO(-)/CO/CO(HB), CA(-)/CA, CB(-)/CBのように分類して帰属をする(S118)。いったん確定した関連性は、同じインデックスのストリップ面上でのみ、「Destroy Bond Connectivity」ボタン474をクリックすることで、解除することができる。なお、CO(HB)とは、水素結合のドナーのカルボニル炭素のことである。

10

各ストリップ上で全ての相関ピークが観測されるのではなく、重なったり、S/Nが悪いため、ピークが観測されないこともある。そこで、可能なパターンを各スペクトルについて生成し、全てのストリップ・チャートの組についてつじつまのあう帰属を求める。

【0016】

インデックス間で重なりがあったら、新たにピークを追加するなどして、2次元実験との関連性を一対一に保つよう帰属する。これを行うために、「Return Feedback to Ref - 2D」ボタン478をクリックする。これにより、新たなピークが形成されて、そのピークに対してインデックスが付与される。なお、付与されたピークには「Feedback from 3D」が表示される。また、2次元スペクトル上で少しでもピークの形状が歪んでいたら、インデックスの重なりがあるかどうか検討して、全ての曖昧さに決着をつける必要がある。

20

この様にして、2次元のHSQCスペクトル上のピークと、3次元の各スペクトル上でのピークとの対応と、それぞれのピークの帰属(CA(i-1), CB(i-1), CA(i), CB(i))が完了する(S120でYES)。

【0017】

(b) クラスタ・タグの作成

クラスタ・コネクティビティおよびクラスタ・タグは、メニュー「Main Chain Assignment」230(図6参照)をクリックすると現れる「Make Cluster Connectivities」ボタンをクリックすることで作成される(S122)。この処理では、複数のクラスタ・コネクティビティが一意に連続したもの集まりに、順番にタグ番号を与えて、クラスタ・タグを作成する。出来上がったクラスタ・タグは、「View Cluster Connectivities」メニューで、ウィンドウ490(図7参照)を表示することで確認できる。ここで、「Cluster-Tag Index」492として、#1が表示されている。クラスタ・インデックスは、「+」「-」ボタン494で増減でき、その番号のクラスタ・タグの内容は、「Show Cluster-Tags」ボタン496で、表示することができる。クラスタ・タグの内容は、「Cluster-Tag's Sequence」498に、インデックスの連続として表示される。

30

【0018】

(c) クラスタ・タグの帰属

クラスタ・タグの帰属では、アミノ酸一次配列と化学シフト・ライブラリーとを用いてアミノ酸予測を行い、クラスタ・タグとアミノ酸配列とを対応させる。この処理は、メニュー「Main Chain Assignment」230(図7参照)をクリックすると現れる「Assign Cluster-Tag Sequences」ボタンをクリックすることで行われ、図8に示す画面500が表示される。

40

図8において、画面500は、上下2段に別れ、各段の上部には、読み込んだアミノ酸の一次配列512, 514が表示されている。これに対応するクラスタ・タグ・シーケンスが、右側に表示されているクラスタ・タグ番号540の線上に、帰属可能な候補として黄緑のバーで明示されている。バーの横の数字は、アミノ酸配列番号と括弧内に表示されている予測による確率である。表示されている確率の計算については、後で説明する。この

50

バーをクリックすると、バーに対応したクラスタ・タグの内容がウインドウ 5 5 0 に表示される。

クラスタ・タグの帰属の処理は、当てはまり予測により、確かそうなクラスタ・タグから順番にアミノ酸配列に対して帰属させる。帰属操作は、ウインドウ 5 5 0 内の「Assign Cluster - Tag」ボタン 5 5 2 をクリックすることで行う (S 1 2 4)。一度タグによって帰属された配列は、次回のアミノ酸の型予測の対象から外されるので、帰属可能な候補の確率は再計算される。この帰属操作・再計算は繰り返される。従って、はじめは、帰属可能な配列の特定できなかったタグも他が帰属されるごとに、候補が絞り込まれていく。この帰属を繰り返すことによって、数分のうちに 7 割以上の帰属が終了する。

10

【0019】

(d) フラグメントの帰属

ある程度クラスタ・タグの帰属解析が進んでいくと、どうしても帰属ができない領域が生まれる。この原因は、クラスタ・コネクティビティが一意でなかったり、Cbの化学シフトが決まらなかったりして、生じたフラグメントによって説明される。それらの帰属は、アミノ酸の型予測とクラスタ・タグの端の化学シフトの一致、不一致の情報を頼りにして帰属を埋めていかなければならない。

これを行うために、メニュー「Main Chain Assignment」230 (図8参照) をクリックすると現れる「Assign Fragment Indices」によって行われ、図9に示す画面600が表示される。図9の画面600は、図8に示した画面500と同様の構成であり、上下2段に別れ、各段の上部には、読み込んだアミノ酸の一次配列612, 614が表示されている。フラグメントが検出され、アミノ酸配列612, 614に対応して、右側に表示されているフラグメント番号640の線上に、帰属可能な候補として黄緑の三角で明示されている。三角の横の括弧内の数字は、予測による確率である。この三角をクリックすると、三角に対応したフラグメントの情報がウインドウ650に表示される。

20

このウインドウ650に表示される情報を参考に、フラグメントの帰属を決定する。フラグメントを帰属させるためには、ウインドウ650内の「Assign Fragment」ボタン652をクリックする (S 1 2 6)。

【0020】

(e) 帰属の評価

決定された帰属の評価は、メニュー「Main Chain Assignment」230 (図9参照) をクリックすると現れる「Assess Main Chain Assignments」によって行われ、図10に示す画面700が表示される (S 1 2 8)。図10における画面700では、上下2段に分かれて、アミノ酸配列712, 714が表示されている。このアミノ酸配列に対応して、右側の表示に対するクラスタ間の化学シフトの不一致やアミノ酸型予測からの逸脱等が、色やバーの高さで表示される。緑<黄<赤の順、あるいはバーの高さが高いほどその帰属の信頼性が低いことを示している。それらの帰属を評価し直し、その原因を必ず追及することにより、フラグメントの帰属が解消される。帰属の結果は、メニュー「Main Chain Assignment」230 (図10参照) をクリックすると現れる「List Main Chain Assignments」ボタンによってテキスト保存される。また、アミノ酸の文字をクリックするとウインドウ540が現れ、帰属を外すことができる。

30

40

なお、上述では、3次元スペクトルを用いているが、4次元のスペクトルを用いることもできる。

【0021】

<側鎖の帰属>

側鎖についても、上述の主鎖における2次元実験とクラスタ・コネクティビティを確定し原子型を帰属する過程と同様の手順で、メニュー240を用いることにより帰属操作を行うことができる。

50

アミドプロトン - アミド窒素 (HN - N) を経由する脂肪族側鎖の帰属は、2次元実験として NH - HSQC を使用し、リンクを確定する実験として、3次元の HBHA (CO) NH, C (CO) NH, HC (CO) NH を使用する。

また、芳香族側鎖の帰属は、参照する2次元実験として、HN - HSQC, CT - HSQC の両方を用い、クラスタ形成に必要なスピン系を裏付ける実験として、2次元の COSY, TOCSY, NOESY も使われる。帰属を確定する実験としては、3次元の HCCH - COSY, HCCH - TOCSY, 15N edited NOESY, 13C edited NOESY あるいは4次元の HCCH - COSY, HCCH - TOCSY, C - C NOESY, C N NOESY などが使われる。

次に、例えば、芳香族側鎖の帰属の際には、主鎖帰属と同様に、化学シフトの一致と BMRB データベースとの比較、芳香環属プロトン間の COSY, TOCSY, NOESY シグナルの存在を調べた結果からクラスタ形成を行い、芳香環を一つのクラスタ・タグにまとめ、一括して帰属を行う。

< 構造計算段階の帰属 >

構造計算の段階では、帰属の完了した帰属テーブルを用いて、NOE シグナルの帰属を行う。NOE シグナルの帰属は各実験にリンクされており、帰属の修正を行う場合でも、修正に関連した全てのシグナルが同時に修正される。このため、構造計算と帰属過程とを同時に行うことができる。

以上、特定のシグナルとスペクトルをリンクすることにより、帰属過程を行うことが可能となり、初心者でも熟練者同様の確度でスペクトル解析を進めることができる。

【 0 0 2 2 】

< 操作のログ >

一連の帰属の過程で、HSQC スペクトル上で定義された特定のピーク番号 (インデックス) を持つシグナル (HN, N) は、他の3次元スペクトル上のシグナルとのリンクにより有機的に関連づけられている。また、特定のインデックスを持つシグナルは、アミノ酸配列ともリンクされる。これらのリンクを作成することにより帰属が行われている。この帰属の操作過程は、一連の操作コマンド列として、ログ・ファイルに書き込まれる。帰属を修正した場合でも、修正を含めて帰属過程の全てのログを残すことにより、帰属過程を再現することができる。これを図 2, 図 11, 図 12 を用いて説明する。

図 2 のフローチャートにおいて、上述したように、アミノ酸配列のファイルを読み込むことで、この帰属の処理を開始している (S102)。このときに、ログ・ファイルも自動的に作成されている (S104)。このログ・ファイルには、帰属のための操作を行うごとに、その操作により発行されるコマンドを格納している。

操作のときに発行されるコマンドの例を、図 11 に示す。例えば、アミノ酸配列ファイルの読み込み操作には、図 11 に示した「1... Read Amino Acid Sequence File」(S102 参照) を発行することで行っている。図 2 のフローチャートには、そのステップの操作により発行されるコマンドを例示している。

【 0 0 2 3 】

ログ・ファイルの内容の最初の部分を図 12 に示す。図 12 において、実行されない Remark (これは起動時の設定に自動的に書き込まれる) の次に、アミノ酸配列ファイルの読み込み (S102) と、一連のスペクトル・ファイルの読み込み (S104), 実験の定義 (S108) に対応するコマンドが読みとれる。また、その次には、ボンド・コネクティビティの確定 (S118) のためのコマンドが続いている。

この帰属のためのコマンドで構成されたログ・ファイルは、本支援システムにより、再実行できる。これにより、他の人が行った帰属処理を、ログ・ファイルを再実行することで、再現することができる。

本支援システムを用いることにより、膨大なスペクトル群と帰属過程を有機的に結合することにより、帰属過程の共有化を行うことが可能となる。また、ログファイルを用いることにより、他の研究者の行った帰属の過程を再現できる。

【 0 0 2 4 】

10

20

30

40

50

< 帰属の自動化 >

このシステムを用いることにより、原子型帰属の自動化を行うこともできる。この帰属の自動化により、初心者でも熟練した研究者同様に正確な帰属を進めることが可能となり、作業効率を大幅に向上することができる。

本支援システムでは、典型的な帰属を行うと、それらを学習して自動帰属を行う。学習を行うインデックスはプロファイル・インデックスと呼ばれ、自動帰属の見本となる。このため、このインデックスとしては、S/Nが良好で、一連の実験で見えるべきシグナルが揃っているインデックスを利用すべきである。このプロファイル・インデックスの帰属から、帰属、帰属シグナルの総数、シグナルの符号、シグナル間の強度比の情報から、帰属規則が分かる。

図13は、図6と同じ画面で、プロファイル・インデックスの帰属を決めて、自動帰属(Auto Assign)を行うところを示す図である。プロファイル・インデックスの帰属を決めてから、図5、図6でも表示されているウインドウ370内の「Auto Assign」ボタン374をクリックすると、図13に示した、自動帰属のためのウインドウ810、820、830が開く。ウインドウ810、820で、自動帰属のための各種パラメータを設定後、ウインドウ830の「Start」ボタン832をクリックすることで、自動帰属処理が開始する。

【0025】

自動帰属処理が開始されると、まず、どのシグナルにどの原子型が帰属するかを決定するパターンを作成する。このパターンは、どのシグナルにどの原子型帰属を当てはめるかの情報の集まりである。本支援システムでは、シグナルの総数によって、原子型帰属の種類になるパターンの最大数を変化させる。学習通りにシグナルが揃っている場合には、数十程度のパターンを生成することで、十分な精度が得られる。一つのパターンを生成する作業は以下のように行う。まずストリップ上のピークの数不足している場合に備えて、ピークピックを行う。ピークが巨大なシグナルのリッジピークであるかどうか調べ、そうであるならピークを参照する優先順位を下げる処置を行う。次にストリップ上のシグナルを正負に区別して大きさ順に並べ替え、ストリップ上のシグナルの総数に比例した偏差を与えたガウス分布に従う確率で、大きい方のシグナルをパターンに加える。ここで同じピークに異なる原子型帰属をしたり、また異なる実験の間で等しいはずの化学シフトが有意に異なるパターンは捨てる。同じパターンを避けて、それぞれに異なるパターンを次々生成し、設定した最大パターン数が集まるまで、あるいはパターン生成の試行数が最大パターンの二乗に達するまで続ける。

【0026】

生成した帰属パターンには、原子型帰属の規則の十分条件に違反するものも含まれる。例えば、当然加わるべきシグナルがパターンに含まれなかったことで、偶然に帰属規則の必要条件を満たしているように見せかけている偽のパターンが存在する。そこで原子型帰属パターンに対して以下の最適化を行う。

シグナルの補完、同一実験内で異なる原子型を与えるシグナルがその実験の分解能以上に近づいた場合にマイナーピークの排除、基準化学シフトからのずれが異常なピークの排除、同じ原子型帰属の化学シフトが実験の間で異なるものを排除、シグナルの欠落箇所ですpekトルの振幅の符号が違反したピークの排除。

(このとき、候補パターンを判定するのに必要なスペクトル上の強度は、一括してバッファにとっておき、随時参照するとよい。)

こうして、最適化された原子型帰属パターンの総数を単純に増やすことにより、帰属規則の必要十分条件を満たす範囲で原子型帰属の可能な組み合わせを調べ尽くすことができる。

【0027】

最後の評価ステージでは、評価関数を用いて原子型帰属パターンの優劣を判断する。このとき使用する評価関数は、それぞれ帰属段階の性質に合うように調整されている必要がある。評価項目として重要なのは、典型的なアミノ酸の化学シフト統計であるBMRBデー

10

20

30

40

50

データベースを利用する項目である。BMRB化学シフト統計は多くのソフトウェア(CSI, SHIFTY, TALOS, TATAPRO等)によってこれまで構造予測や自動帰属などに頻りに利用されてきた実績がある。

パターンの評価関数は他に、全シグナルの大きさの和、基準化学シフトからのずれ、異なる実験の間の化学シフトの同一性、シグナル強度比と学習プロファイルとの一致、最大帰属可能候補をどれだけ埋められたのか、原子型帰属とシグナルの大きさの順番に逸脱があるか等の複数の項目を考慮したものになる。最終的に全評価関数は、独立の事象に対する評価のように、各評価項目が乗算で寄与する複合関数となる。各評価項目の判定値にかかる指数因子によって、ユーザーが全評価関数を調整することができる。ある一定の閾値を越え、全評価関数値の最も優れた原子型帰属パターンを選択することにより、原子型帰属の自動化処理は完了する。処理結果は、ユーザーによる確認、修正を経た後で次のクラスターの帰属に進むことになる。

10

【0028】

帰属が一通り済んだら、3次元実験のピーク情報を2次元実験へ情報をフィードバックする。これは、2次元実験の重なり合ったピークやブロードな交換ピークを「Feedback」ボタン478(図6参照)で自動検出して、再び帰属をくり返すことで行う。これらの手順をまとめて自動的に行う便利なボタンを設けている。これが、「Main Chain Assignment」230内の「Auto Assign <-> Feedback Repeat」ボタンであり、これを利用すれば主鎖帰属が一発で完了する可能性がある。本支援システムでは、帰属とフィードバックとを3回繰り返している。

20

【0029】

< 確率の計算 >

(アミノ酸型分配確率)

まず、単純なアミノ酸型分配確率計算を示す。

化学シフトの中心値($x = \text{HN}, \text{N}, \text{C}^1, \text{Ca}, \text{Cb}, \dots$)が分かっているならば、あるアミノ酸 a に帰属する確率分布を以下のように計算できる。

【数1】

$$A(a) = \exp \left[- \sum_{\substack{x=\text{HN}, \text{N}, \text{C}^1, \text{Ca}, \text{Cb}, \dots \\ x \neq \phi}} \frac{(x - \bar{x}(a))^2}{2\sigma_x^2(a)} \right]$$

30

$$x_{\min}(a) - \sigma_x(a) \leq x \leq x_{\max}(a) + \sigma_x(a)$$

or

$$A(a) = 0 \quad x < x_{\min}(a) - \sigma_x(a) \text{ or } x > x_{\max}(a) + \sigma_x(a) \text{ or } \bar{x}(a) = \phi \cap x \neq \phi \text{ or } a = \phi$$

$$p(a) = \frac{A(a)}{\sum_{a'=\text{Ala}, \text{Cys}, \text{Asp}, \dots} A(a')}$$

40

ここで、 x は化学シフトの中心値、 $\bar{x}(a)$ 、 $\sigma_x(a)$ は BMRB 化学シフトの統計平均および標準偏差である。 $x_{\min}(a)$ 、 $x_{\max}(a)$ はその最小値および最大値である。

なお、第2式は統計上起こりえない事象を完全に除くためにある。

絶対的な帰属確率の値は必ずしも必要ではない。むしろ、確率分布量の最も大きなアミノ酸の型が何であるか知るだけでよい。

簡単のため、化学シフト値の広がり($d_{\text{HN}}, d_{\text{N}}, d_{\text{C}^1}, d_{\text{C}}, d_{\text{C}}, \dots$)を全て1とすると、ある化学シフトの中心値 x を与えたときにそのアミノ酸の型が a である分配確率は、次のように計算できる。

50

【数 2】

$$p(a) = \frac{A(a)}{\sum_{a'=Ala, Cys, Asp, \dots} A(a')}$$

【0030】

(クラスタの帰属予測)

クラスタの帰属予測では、上述の化学シフト中心値と化学シフト統計との間の対応を調べる他に、クラスタの両末端で既存の帰属の化学シフトが一致するかどうか考慮しなければならない。このため、主鎖帰属のクラスタの分配確率は、BMRB化学シフト統計情報に依存した前述の $p(a)$ 項とクラスタ両末端 (termini) の化学シフトの同一性 (前後の C^1 , Ca , Cb の化学シフトに関わる) に関わる項から計算される。

クラスタが一次配列 c の i 番目に帰属される分配確率は、

【数 3】

$$P_{c,i,n=1} = \frac{(1+t_{c,i-1})^{-1} p(a_{c,i-1}) p(a_{c,i}) (1+t_{c,i+1})^{-1}}{\sum_{c',i'} [(1+t_{c',i'-1})^{-1} p(a_{c',i'-1}) p(a_{c',i'}) (1+t_{c',i'+1})^{-1}]}$$

$$t_{c,i} = \frac{\sum_{\substack{x=C', Ca, C\beta \\ x_{c,i-1} \neq \phi \cap x_{c,i} \neq \phi}} \frac{(x_{c,i-1} - x_{c,i})^2}{\sigma_x^2}}{\sum_{\substack{x=C', Ca, C\beta \\ x_{c,i-1} \neq \phi \cap x_{c,i} \neq \phi}} 1}$$

ただし、分母が 0 の場合は、最大許容誤差を代入する。

【数 4】

$$t_{c,i} = err_{lmt}$$

ここで、 σ_x は化学シフトの標準偏差である。 1H , ^{13}C , ^{15}N には、それぞれ 0.01, 0.04, 0.1 ppm を与える。このとき、標準誤差、最大許容誤差程度の値を与える、 n 個のクラスタからなるクラスタの先頭が、一次配列 c の i 番目に帰属される分配確率は、

【数 5】

$$P_{c,i,n} = \frac{(1+t_{c,i-1})^{-1} \left\{ \prod_{i'=i}^{i+n} p(a_{c,i'-1}) \right\} (1+t_{c,i+n})^{-1}}{\sum_{c',i'} \left[(1+t_{c',i'-1})^{-1} \left\{ \prod_{i''=i'}^{i'+n} p(a_{c',i''-1}) \right\} (1+t_{c',i'+n})^{-1} \right]}$$

である。

【0031】

(分配確率計算に基づくクラスタ形成)

この分配確率を利用すると、タンパク質の一次配列と BMRB 化学シフト統計データと矛盾のないクラスタ形成が可能である。化学シフトの同一性だけを考慮したクラスタ形成には問題がある。不用意に長いクラスタが、実は別々の帰属に当てはまるべき、全く別の 2 つのクラスタでなければならないかもしれない。

そこで、分配確率の基本的な性質を理解する必要がある。

「クラスタ・サイズ n が十分大きい場合」、ある特定のサイトに対する帰属（正しい帰属と考えられる）への分配量が集まりやすい。クラスタ・サイズが1つ大きくなると、望ましい第一候補以外の対立候補であるアミノ酸の型が一致しなくなって分配量が減少する。すると、ますます第一候補への比重が増加する。その結果、分配確率の最大値は上昇する。よって、クラスタの形成が正しく行われているときには、分配確率の最大値は n が大きくなるに従い、分配確率の最大値は 1.0 に漸近しながら増え続ける。

問題は、アミノ酸の型が合っていないときにもクラスタが一定以上大きいと、やはり分配確率の最大値が増加する可能性があることである。偶然に化学シフトが一致して付加したクラスタは、正しい帰属サイト付近の絶対的な確率分配量は減少する。しかしながら、対立候補でも型が合わないことが起きているため、分配確率の最大値はこうしたエラーに鈍感でありうる。これは、できあがったクラスタを一つ一つ刈り込んで、同時に分配確率の最大値の変化率を監視することで問題は解決する。

10

【0032】

偶然、化学シフトが同一になった結果、誤ったクラスタが構成されている場合、最大分配確率の増加率は n の増加に従って、一時的に減少する点が存在するのである。そのような点を見いだしたときには、クラスタを分割しなくてはならない。また一方、「 n が3以下のような小さな場合」、 n を増やしていった場合に、最初のうち最大の分配確率を示す帰属サイトが正しいとは限らない。従って、ある程度、変化率の減少は起きていてもおかしくはない。結局、経験上、以下の判別式を用いることで目的が達成できる。

【数6】

20

$$(n+1) \cdot \text{maximum}(p_{c,i,n+1}) > n \cdot \text{maximum}(p_{c,i,n}) \text{ and } \text{maximum}(p_{c,i,n}) \geq \frac{n}{N_c}$$

この第一の判別式は、 n が小さい場合には、分配確率の最大値の変化率で減少する場合を許している。また n が大きいときには、その変化率が 1.0 に十分近づいて、分配確率の最大値は 1.0 が漸近的に近づくことを要求している。結局、この判別式は定性的に正しい振る舞いをしていることが理解できる。

また、第二の判別式はその長さのクラスタが最低限有意であることを保証する。ここで、 N_c は一次配列 c の長さである。つまり、クラスタ・サイズが配列に対して応分の長さがあるなら、その長さを占めるだけの責任に対して、 n/N_c は分配確率で負担しなければならないことを意味する。

30

実際のクラスタ作成は、まず、ブランチのないように化学シフトの同一性だけでクラスタ作成した後、それぞれのクラスタについて順次刈り込んでいく。そして、長さが1になるまで、上記の判別式を満たしているかどうかを確かめることを行えばよい。

【0033】

この支援システムを用いて、いくつかのタンパク質の帰属を進めた。従来の方法はスペクトルから帰属への流れで仕事は進み、流れを遡って帰属をチェックすることは困難である。本支援システムにより、遡って帰属のチェックを行うことが可能となり、帰属効率を大幅に向上することができた。また、構造計算の過程で明らかになった帰属の誤りをスペクトルに戻って見直すことができた。データベースを使用し、帰属の確度を統計的に処理するとともに、評価関数を設定することにより、確度の高い帰属を行うことができた。

40

ポストゲノム時代に入り、網羅的なタンパク質の構造解析が必要になっている。我が国も、3,000個のタンパク質の立体構造をNMRで決定することを国際的に約束している。また、製薬会社も、創薬のターゲットとしてタンパク質の立体構造決定が急務となっている。NMRによるハイスループットなタンパク質の立体構造決定は広く望まれている。本支援システムはこのような要請を満たすものであり、NMRを用いたハイスループットな構造解析を可能とする。

【0034】

【発明の効果】

50

上述するように、本発明の支援システムにより、従来では主鎖帰属に1ヵ月程度、構造決定までは半年以上を要したが、主鎖帰属は1 - 3日へ、構造決定も2ヵ月程度に短縮することができる。

本プログラムを使用することにより、初心者でも、熟練者同様に確度の高い帰属と精度の高い構造解析を行うことが可能となった。

【図面の簡単な説明】

【図1】NMRシグナルの帰属法について説明する図である。

【図2】本支援システムを用いて行う帰属操作を説明するフローチャートである。

【図3】本支援システムを起動すると現れる画面を示す図である。

【図4】2次元のスペクトルを表示した画面を示す図である。

【図5】2次元のピークと対応する3次元のスペクトルを表示するための操作を説明する図である。

【図6】3次元のスペクトルを表示した画面を示す図である。

【図7】クラスタ・タグ作成を示す図である。

【図8】クラスタ・タグの帰属を行う画面を示す図である。

【図9】フラグメントの帰属を行う画面を示す図である。

【図10】帰属の評価の画面を示す図である。

【図11】ログ・ファイルに記録される操作に対応したコマンドのリストである。

【図12】ログ・ファイルに記録された操作履歴の例を示す図である。

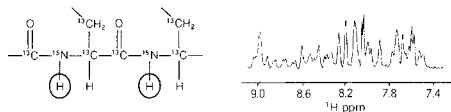
【図13】自動帰属を行う画面を示す図である。

10

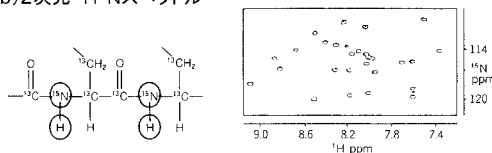
20

【図1】

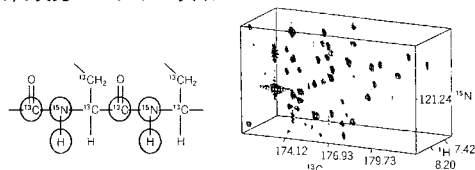
(a) 1次元 ^1H スペクトル



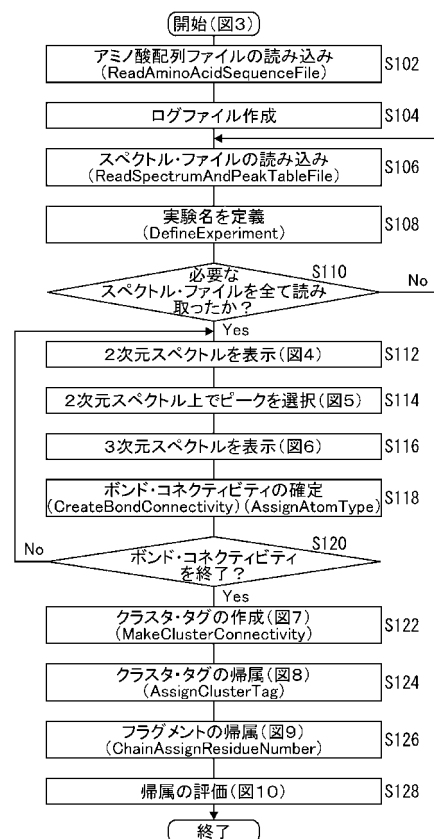
(b) 2次元 ^1H - ^{15}N スペクトル



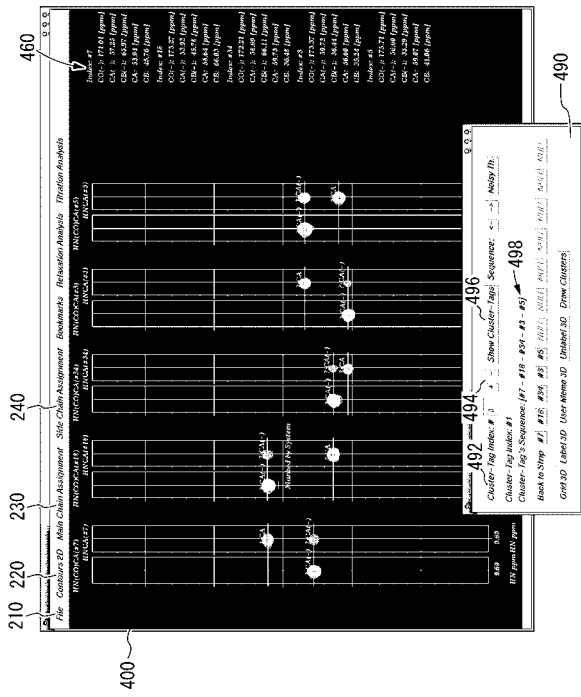
(c) 3次元 ^1H - ^{13}C - ^{15}N スペクトル



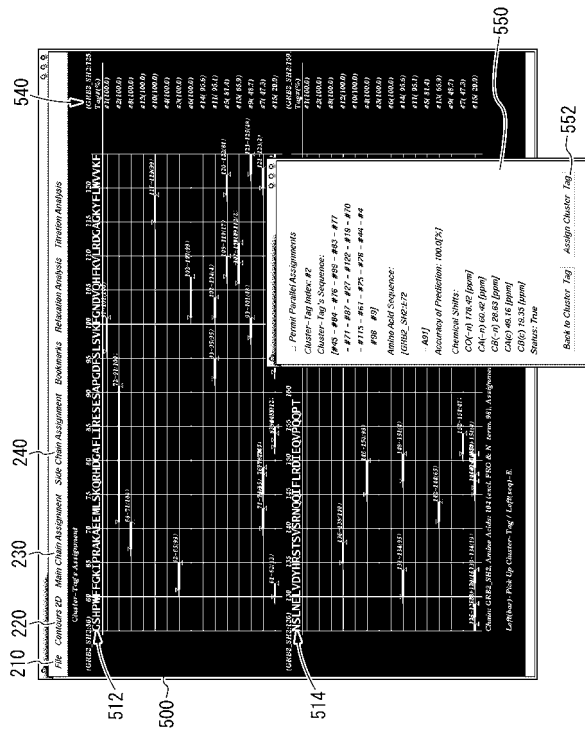
【図2】



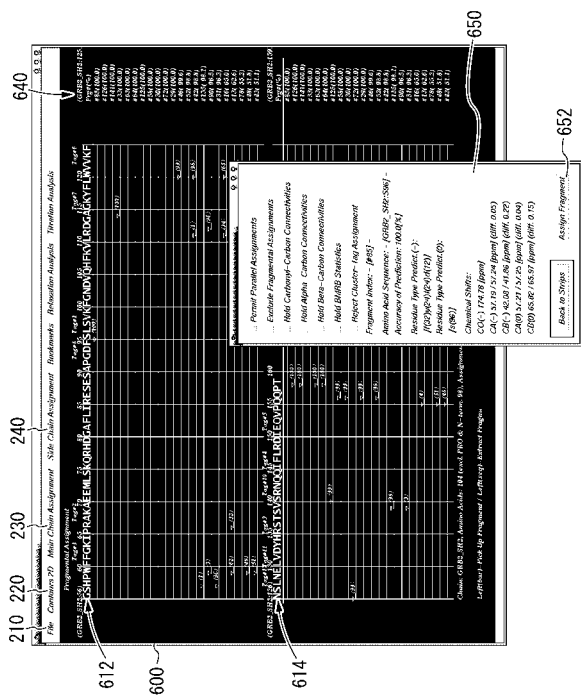
【 7 】



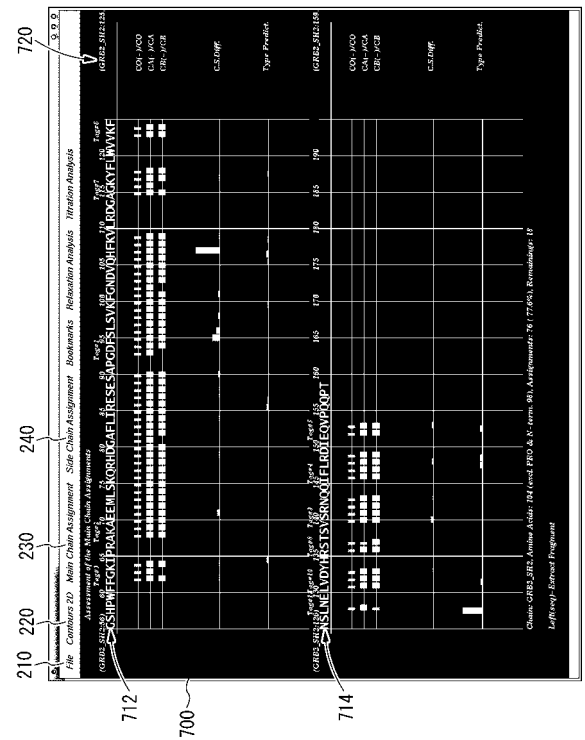
【 8 】



【 9 】



【 10 】



フロントページの続き

- (56)参考文献 特開平02-208579(JP,A)
特開平01-287452(JP,A)
特公平08-020387(JP,B2)
特許第2693252(JP,B2)
特開平10-281948(JP,A)
廣明秀一 他, 溶解度の限られている蛋白質の主鎖シグナルの効率的帰属, 第39回NMR討論会講演要旨集, 2000年11月, 第142-143頁
廣明秀一, 蛋白質のNMR シグナルの帰属法, 第2回 若手NMR研究会テキスト, 2001年 6月10日, 第1-18頁, インターネット<URL:http://kzmiu.hp.infoseek.co.jp/wakate/2nd_wakate_hp/>
稲垣冬彦, 先導的糖鎖構造解析技術の開発, 糖鎖の構造・機能解析のための共通基盤技術の開発に関する研究(第1期)成果報告書 平成3-5年度, 1995年, 第290-304頁
伊倉光彦, 多次元NMRと蛋白質の構造解析への利用, 蛋白質 核酸 酵素, 1992年, Vol.37 No.3, 第492-507頁
CHARLES D.EADS et al., Automated Location of Peak Patterns in Multidimensional NMR Spectra: A Tool for Resonance ..., JOURNAL OF MAGNETIC RESONANCE, 1993年, Series B 102, p.310-313
G.W.Vuister et al., Assignment Strategies in Homonuclear Three-Dimensional ¹H NMR Spectra of Proteins, Biochemistry, 1990年, Vol.29 No.7, 第1829-1839頁
稲垣冬彦, NMR法によるタンパク質の高次構造研究, 分光研究, 1993年, 第42巻 第5号, 第277-288頁

(58)調査した分野(Int.Cl.⁷, DB名)

G01N 24/00-24/14
G01R 33/20-33/64
JICSTファイル(JOIS)