

DNAシーケンス入力解析システム“DNASIS”

DNA Sequence Input and Analysis System

遺伝子の本体であるDNA(デオキシリボ核酸)はアデニン、グアニン、シトシン、チミンの4種類の塩基で構成されている。最近、この4種類の塩基配列決定方法が急速に進歩し、DNAから得られる遺伝子情報は急激に増加している。また、求めたい塩基配列をランダムに切断した後に全領域を重複させながら、その配列を決定するショットガン法が導入され、研究に拍車がかかりつつある。しかし、その解析は研究者自らの手と目で実施されているため、非常に煩雑である。そこで、今回開発したDNASISとショットガン法を組み合わせる使用することにより、これを解決した。その結果、例えば5,000塩基を決定するのに11箇月かかったのが、13日でできるようになった。

鈴木 理* *Saloru Suzuki*奈須永典* *Hisanori Nasu*

1 緒 言

マキサム・ギルバート法¹⁾とサンガーらの方法はDNA*塩基配列決定法(DNAシーケンス法)を代表する2大技術である。前者は二重鎖DNAを直接シーケンスが決定できるのに対し、後者は単鎖DNAに限定され、取り扱いにくい面があった。しかし、M13ファージベクター系でのクローニング法²⁾、すなわちショットガンDNAシーケンス法の開発により、この障害は取り除かれた。本方法は塩基配列を決めたい二重鎖DNAをランダムに切断し、切断した各々のDNA断片を電気泳動によりDNAシーケンスゲル化するもので、従来の手法ではDNAの末

端からせいぜい350塩基を読むのが限度であったが、DNAをランダムに切断し、500塩基以上を無作為にシーケンスが決定できるものである。一方、従来の手法も含め、電気泳動によるDNAシーケンスゲルの読取り及び解析は、研究者自らの目と手で実施され、非常に効率が悪いものであった。これらの問題点を解決するために、安価なマイクロコンピュータを利用して、解析作業を省力化するDNASIS(DNAシーケンス入力解析システム)を開発した。



図1 DNASIS(DNAシーケンス入力解析システム)の外観 B-16EXIIを中心にソニック デジタルイザ、スピーチ シンセサイザ及びCD-ROM(コンパクトディスク-Read Only Memory)を示す。

※) DNA: デオキシリボ核酸(Deoxyribonucleic acid)の略称である。遺伝子の本体とされている物質で、共通の軸の周りを糖とリン酸が結合した2本の鎖で取り巻く二重らせん構造に

なっている。DNA分子の幅は2 nm、長さは生物の種によって異なるが極めて長い。DNAの持つ遺伝子情報は、その塩基の配列によって決まる。

* 日立ソフトウェアエンジニアリング株式会社

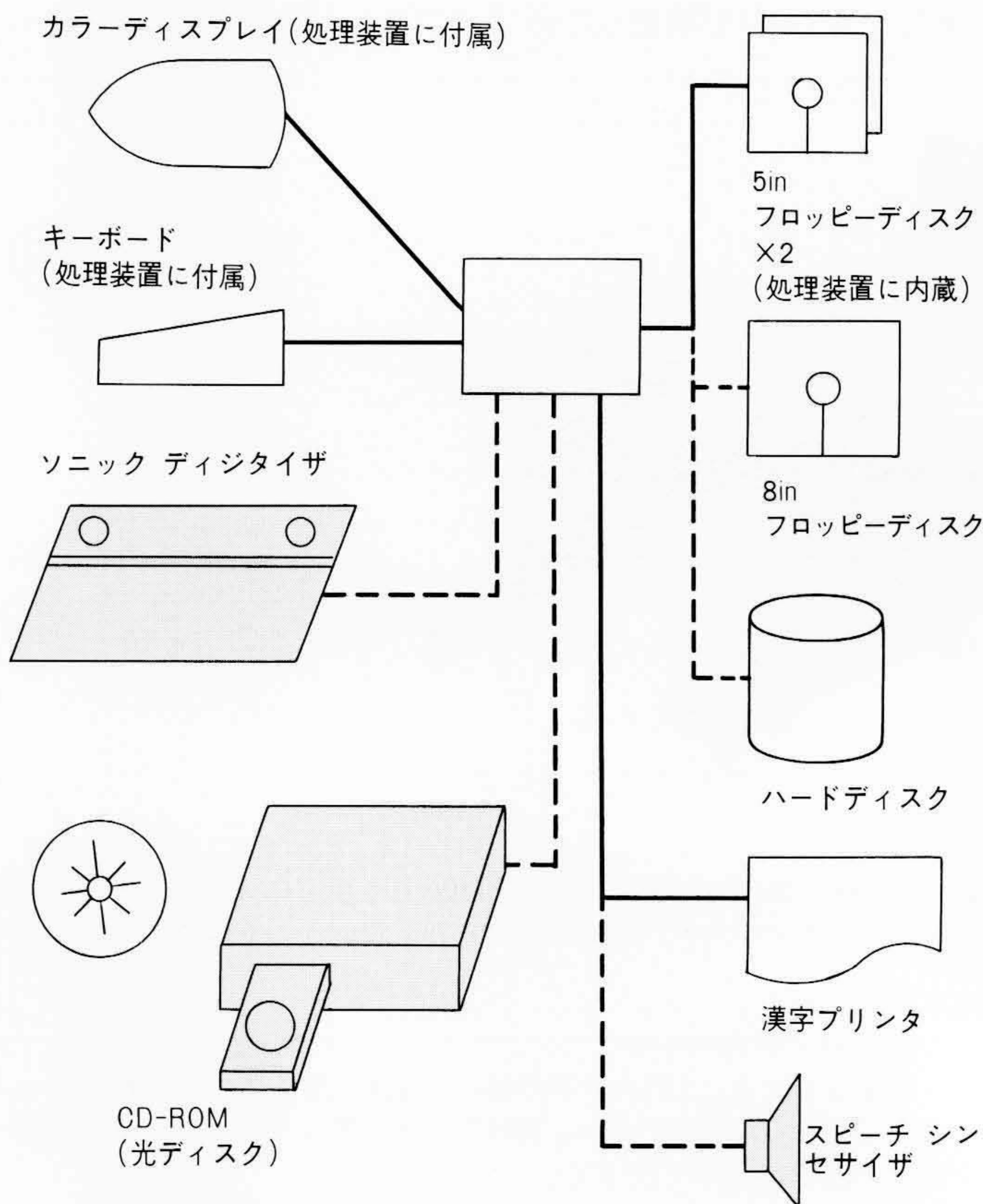
2 DNASISの概要

DNASISは薬品業界、化学業界を代表とする新薬の開発などを行うために必要な遺伝子の塩基配列の決定、及び解析を行うシステムである。コンピュータを使用したことから、人手では考えられない大容量の情報を迅速に、しかも正確に処理することができる。その処理の概要を以下に述べる。

塩基配列の決定を行うに当たって、まず目的の遺伝子を取り出す。その遺伝子を細かく裁断し、その各々について塩基の配列を決定する。そして、これらの断片を結合することによって1本の遺伝子の塩基配列を決定する。これら断片の塩基配列を速やかにコンピュータに入力し、最終的には1本の遺伝子の塩基配列を決定するところから、DNASISの適用は有効となる。更に、コンピュータに入力された塩基配列からアミノ酸の翻訳、二次構造の予測などの情報処理を行うことができる。その結果は単なる数値情報の羅列だけにとどまらず、グラフなどにまとめることができるので、研究者により速く適切な判断を促すことができる。DNASISの外観を図1に示す。

2.1 ハードウェア構成

DNASISは日立製作所製B-16シリーズを中心に入力装置としてソニック デジタイザ、音声による確認装置としてスピーチ シンセサイザ及び遺伝子情報データベース用にCD-ROM(コンパクトディスク-Read Only Memory)から構成さ



注：-----線は、オプションを示す。

図2 DNASISハードウェア構成 点線部分の装置はオプションになっており、ユーザーは予算に合わせて購入が可能となっている。

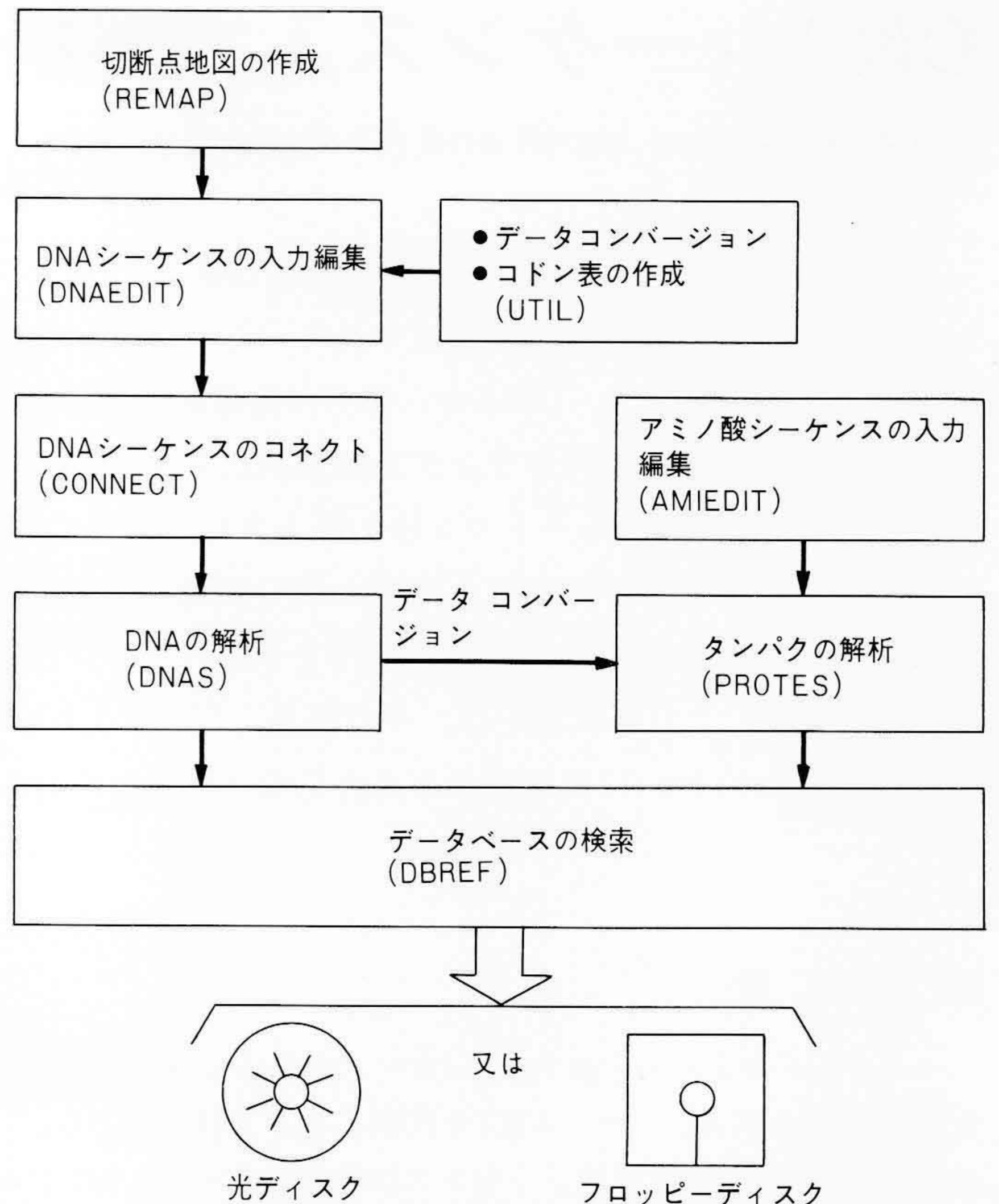


図3 DNASISソフトウェア構成 DNASISのソフトウェアは、八つのサブシステムから成っている。

れ、研究所に導入しやすいようにコンパクトなシステムとなっている(図2)。

2.2 ソフトウェア構成

DNASISのソフトウェアはユーザーフレンドリーを第一に設計され、DNAシーケンスの解析からタンパク質の解析まで豊富に用意されている。これらソフトウェアの構成を図3に、プログラム一覧を表1に示す。

3 DNASISの処理方式

3.1 設計方針

DNASISは従来、研究者自らの手と目で処理していた手法をコンピュータで情報処理するもので、研究者とのマンマシンインタフェースを中心に次のような設計方針を設定した。

(1) 誤りのない、簡便なデータ入力及び確認方式

現在のゲルフィルムは濃淡で判別すること、かなり湾曲があることなどから、人間の目で読むには非常に煩雑な作業となる。更に、研究者が求めたDNAシーケンスと文献などから得られるシーケンスの確認も大変煩雑な作業になっている。したがって、塩基データの入力及び確認方法に関して、誤りのない、簡便方法を実現する必要がある。

(2) 既存遺伝子情報との比較確認方式

研究者は自分の発見したDNAシーケンス又は遺伝子組換え後のDNAシーケンスと、既に公表されている他のDNAシーケンスなどの類似性を比較する方式を実現する必要がある。

表1 DNASISプログラム一覧 DNASISのプログラムは、その解析種別ごとに合計46本から成っている。

項番	サブシステム名	プログラム名	
1	DNAEDIT	DNAシーケンスの入力編集	
2		DNAシーケンスのコンペアエディット	
3		DNAシーケンスファイルの削除, 名変更, リバース	
4		DNAシーケンスのプリント	
5	DNAS	オープンフレームの表示	
6		アミノ酸の翻訳	
7		コドン使用頻度の集計	
8		制限酵素認識部位の探索	
9		スタッキングサイトの探索	
10		ハープロットの表示	
11		G, C含有量の表示	
12		タンデムリピートの探索	
13		コード使用頻度の集計	
14		ヘアピン構造の探索	
15		パリンδροームの探索	
16		高ホモロジー領域の探索	
17		スプライサ	
18		マキシマム マッチング	
19		DNAキーワードの探索	
20		二次構造予測	Zuker-Stiegler
21		プロテイン領域の予測	Fickett
22		t-RNA領域の予測	Staden
23	z-DNA領域の予測	Vass-Wilson	
24	融解温度, 浮遊密度の予測	Poland	
25	CONNECT	手動連結	
26		自動連結	
27	REMAP	切断点地図作成	
28		切断点地図更新	
29		切断点地図削除	
30		切断点地図画面出力	
31		切断点地図リスト出力	
32	AMIEDIT	アミノ酸シーケンスの入力編集	
33		アミノ酸シーケンスのコンペアエディット	
34		アミノ酸シーケンスファイルの削除, 名変更	
35		アミノ酸シーケンスのプリント	
36	PROTES	DNAアミノ酸データ変換	
37		タンパク質の疎水性, 親水性	
38		タンパク質ホモロジープロット	
39		プロテイン高ホモロジー領域の予測	
40		プロテインキーワードの探索	
41		二次構造予測	Chou-Fasman
42		二次構造予測	Robson
43		アミノ酸組成の計算	
44		比容の計算	
45		露出領域の予測	Kyte-Doolittle
46		細胞外分泌の予測	Klein-Kanehisa
47		分子量の計算	
48	DBREF	DNAシーケンスデータベース	
49		PROTEINデータベース	
50	UTIL	DNAシーケンスファイル変換	
51		コドン表作成	

注：略語説明 G(グアニン), C(シトシン)

(3) ショットガン法によるDNAシーケンスの構築方式

3~4年前に出現したショットガン法によって得られる遺伝子情報は急激に増加しており、人間の目で処理するには限界にきている。したがって、ショットガン法によるDNAシーケンスの構築方式を実現する必要がある。

これらシステムに対する課題と実現方式概要を図4に示す。

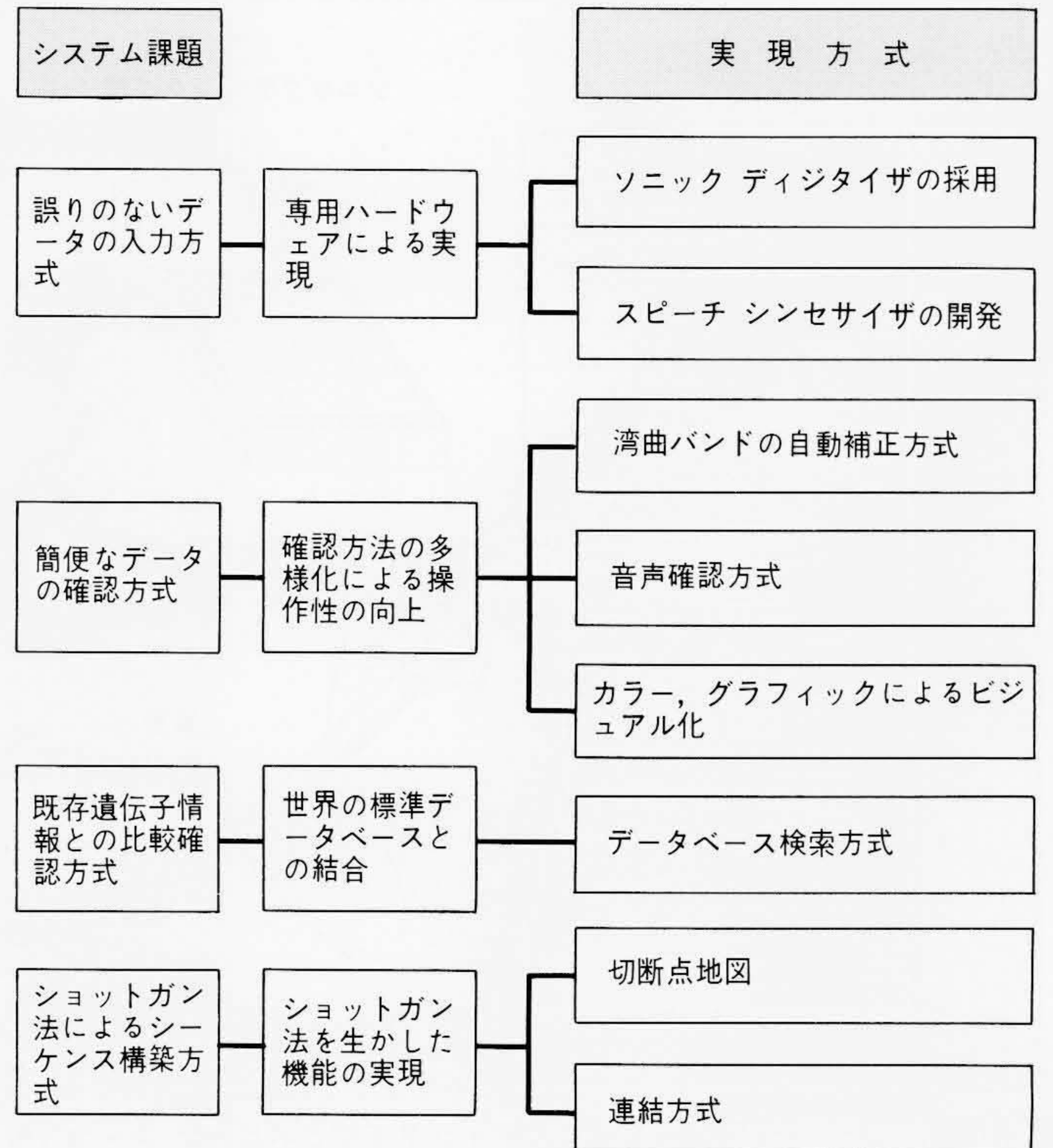


図4 システム課題と実現方式 DNASIS設計時に課題となった項目と、その実現方式の概要を示す。

3.2 実現方式

3.2.1 専用ハードウェアによる実現

(1) ソニック デジタルイザの採用

ゲルフィルムの入力として市販されているライトボックス上でデジタルイザできるソニックデジタルイザを採用した。その特徴は、ライトボックスによりフィルムが見やすいこと、入力有効範囲が45cm×60cmと広いこと、フィルムを分割することなく、かつ濃いバンドでも鮮明に入力できることである。

(2) スピーチ シンセサイザの開発

人間の肉声をROMに焼き付け、コンピュータからの指示により任意の音声を出力する装置の開発によって確認装置としての効力を発揮する。

これら二つの手法は、煩雑なDNAシーケンス読取りや、確認作業を大幅に軽減するものである。

3.2.2 確認方法の多様化による操作性の向上

(1) 湾曲バンドの自動補正方式

フィルム上のバンドが曲線及び末広がり状態でも、バンド位置を再定義、分割入力することなく、一気に連続して入力可能なように、自動補正機能を組み込んである。

(2) 音声確認方式

デジタルイザからフィルムのバンドを押下するごとに、スピーチシンセサイザでエー(A), ジー(G), シー(C), ティー(T)など肉声で返答する。更に、文献などから入力したデータをアデニン, グアニン, シトシン, チミンなど肉声で読み上げさせる。これによって、フィルム又は文献上から視点をずらさずにその確認ができ、従来二人以上で実施していたものを一人で可能とした。本方式の概要を図5に示す。

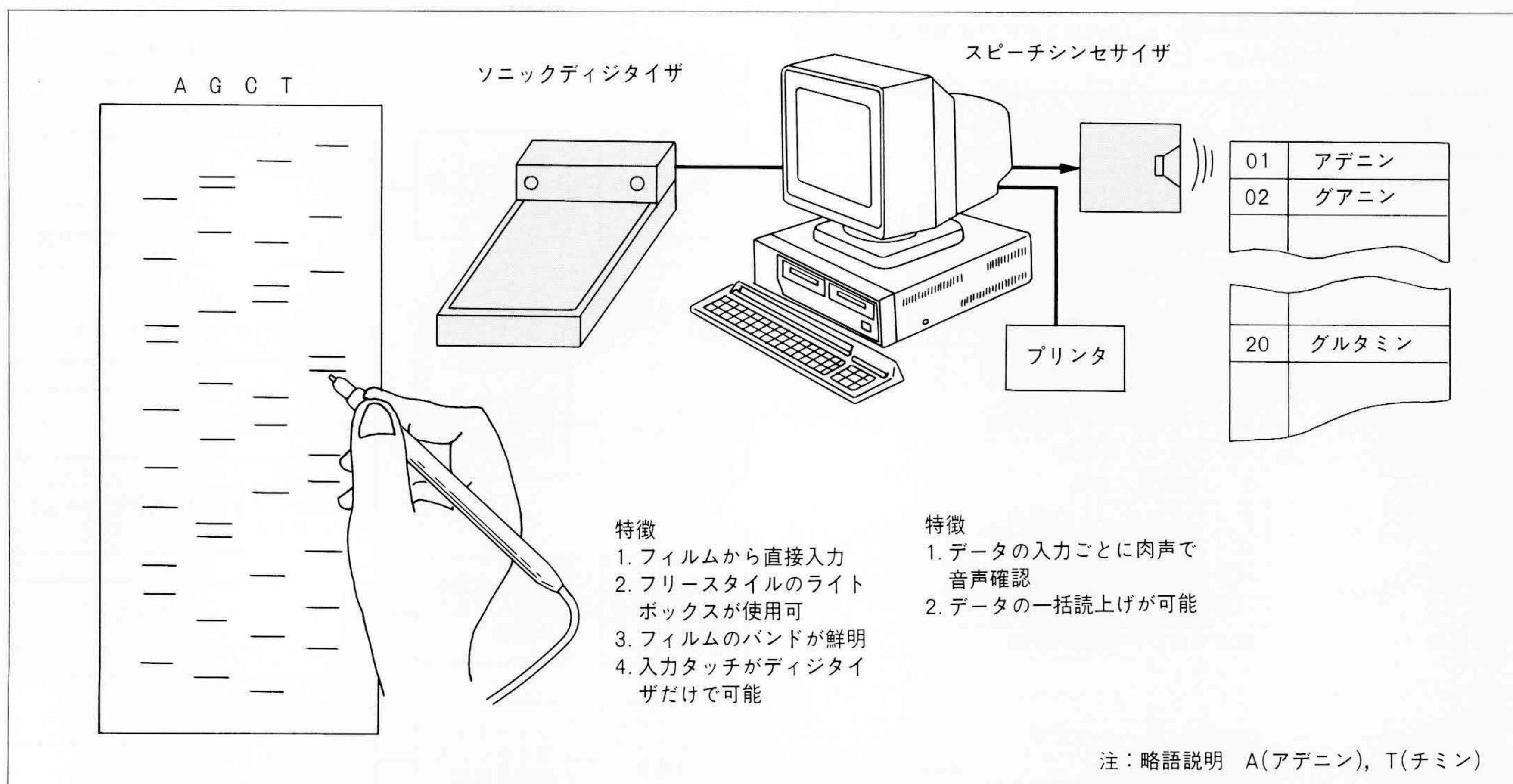


図5 音声確認方式 ソニック デジタイザとスピーチ シンセサイザによる音声確認方式を示す。

(3) カラー、グラフィックによるビジュアル化

ゲルフィルムから読み取った断片のDNAシーケンスは連結され、目的のシーケンス、遺伝子情報であるかどうかを表1に示した機能により解析される。その解析結果は、最先端のコンピュータのグラフィック技術により処理され、研究者にとって見やすく、かつワードプロセッサと組み合わせて論文作成にも使用可能とした。

3.2.3 CD-ROMによるデータベース検索方式

世界で標準的に使用されているDNAシーケンスとタンパク質シーケンスはGen Bank, EMBL(European Molecular Biology Laboratory), NBRF (National Biomedical Research Fundation)としてデータベースセンタから磁気テープ若しくはフロッピーの媒体で世界の研究者に供給されているが、その取扱いは煩雑である。これを世界で初めて1枚の光ディスクに格納した。これにより、媒体の交換作業なしで自由に各種の検索を行うことができる。その光ディスクを図6に示す。

3.2.4 ショットガン法によるシーケンス構築方式

(1) 切断点地図

DNAシーケンスを制限酵素で切断した結果のゲルフィルム上のマーカサイズと各酵素の断片の長さを指定することにより、DNAシーケンスの全体像、すなわち切断点地図を出力する。その出力例を図7に示す。

(2) 連結方式

各々の断片のDNAシーケンスを入力することにより、自動的又は会話形式で連結することができる。自動連結では入力時のミス補って、インサクション及びデリーションを自動的に行い連結する。また連結結果は、その状態が一目で分かりやすい状態にして出力する。その出力例を図8, 9に示す。

4 システムの導入効果

従来のマキサム・ギルバート法²⁾ではゲルフィルムの作成に関し1日の読取り塩基数は平均75であるために、5,000塩基の場合は11箇月を要する。また、ショットガン法では1日の読取りの平均が3,000塩基となるので、本手法の場合は重複部分が存在することを考慮しても、5,000塩基の場合は3日で可能とする。更に、DNAシーケンスの構築には従来の手法の場合



図6 DNASIS CD(コンパクトディスク) 遺伝子情報の光ディスクの例を示す。

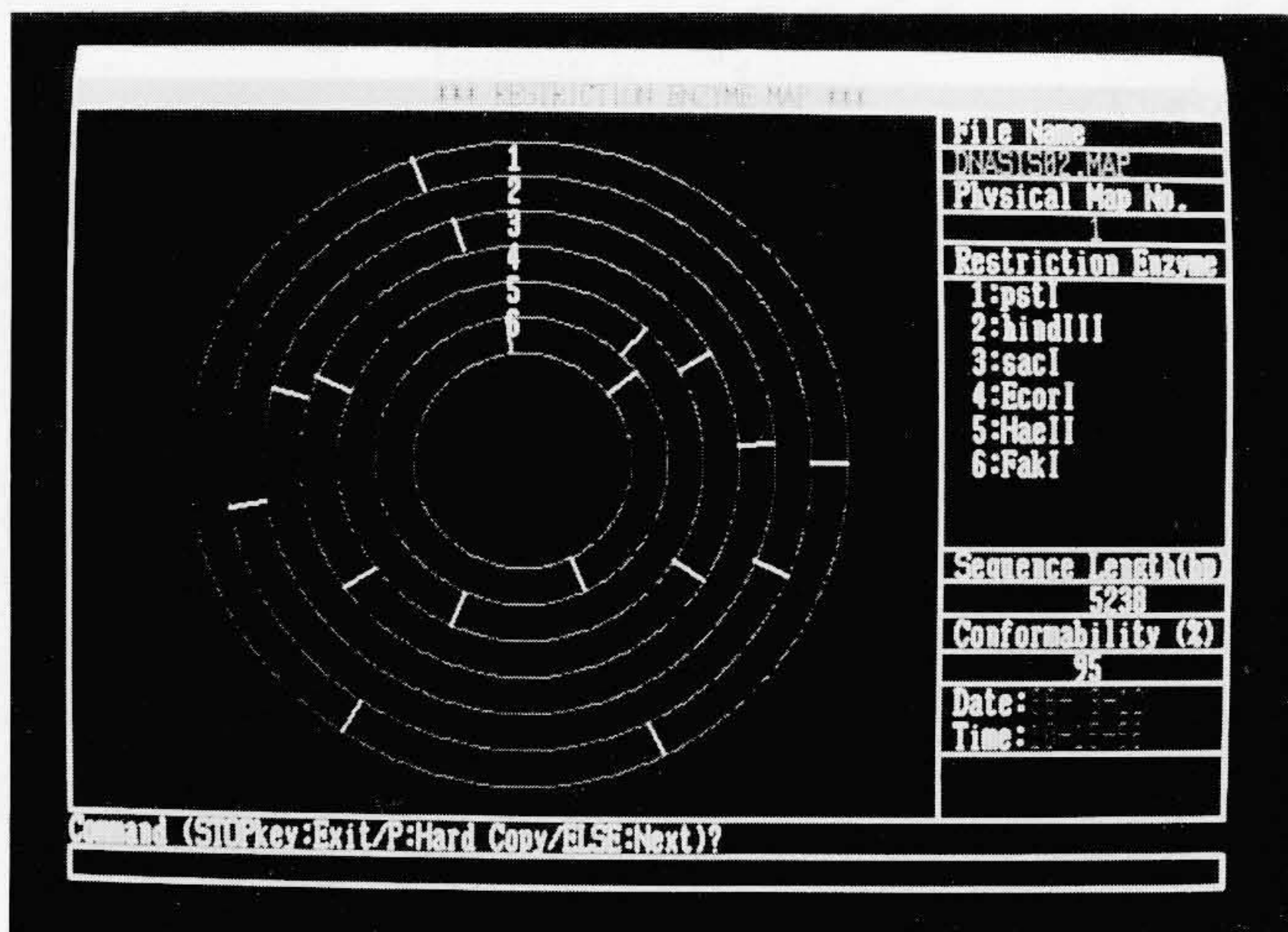


図7 DNAシーケンス切断点地図 切断地図の出力例を示す。

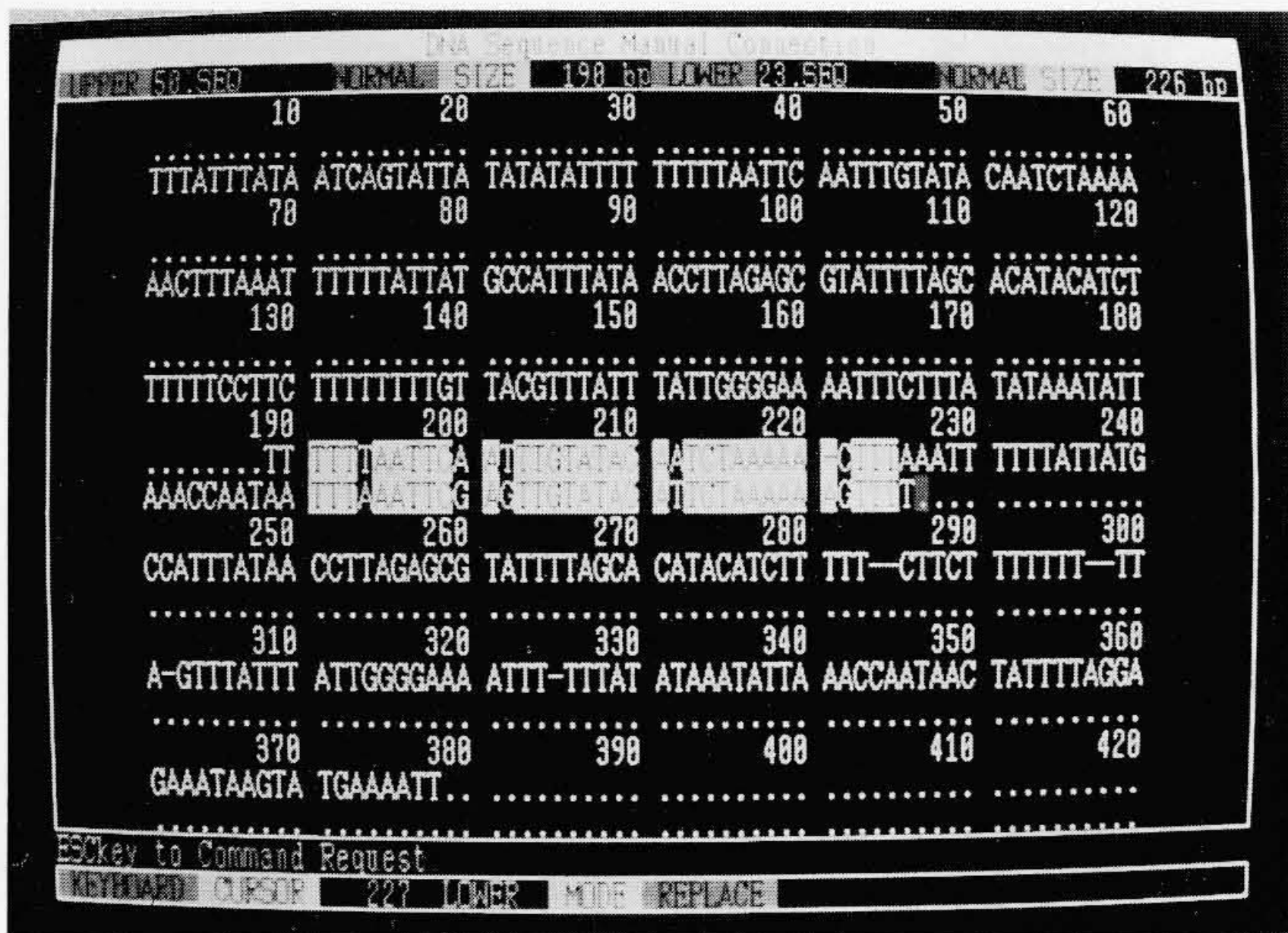


図8 DNAシーケンスの結合 ショットガン法による断片DNAシーケンスの重複部分の結合状態出力例を示す。

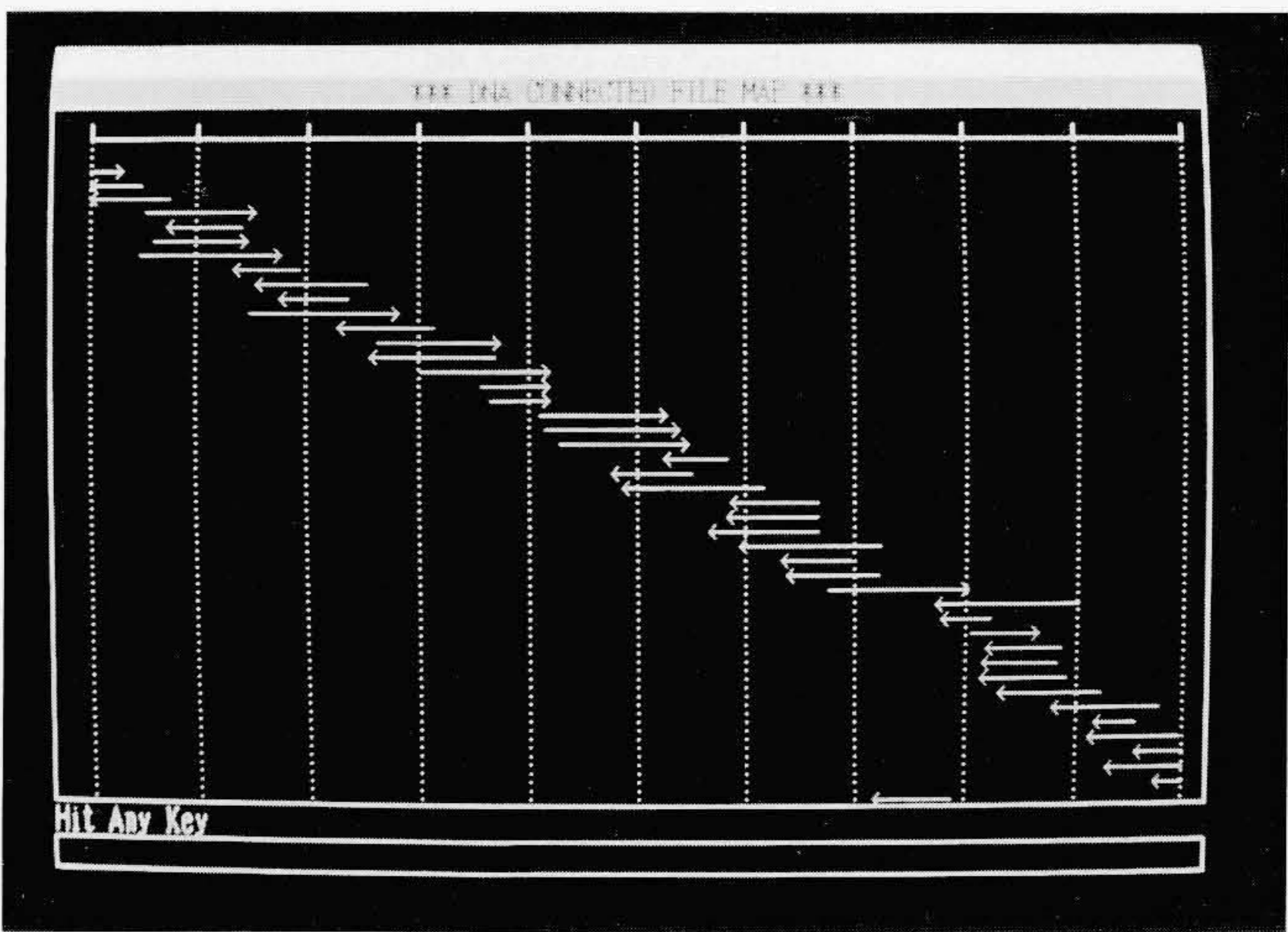


図9 DNAシーケンス連結マップ ショットガン法による各々の断片DNAシーケンスの全体の連結マップ出力例を示す。

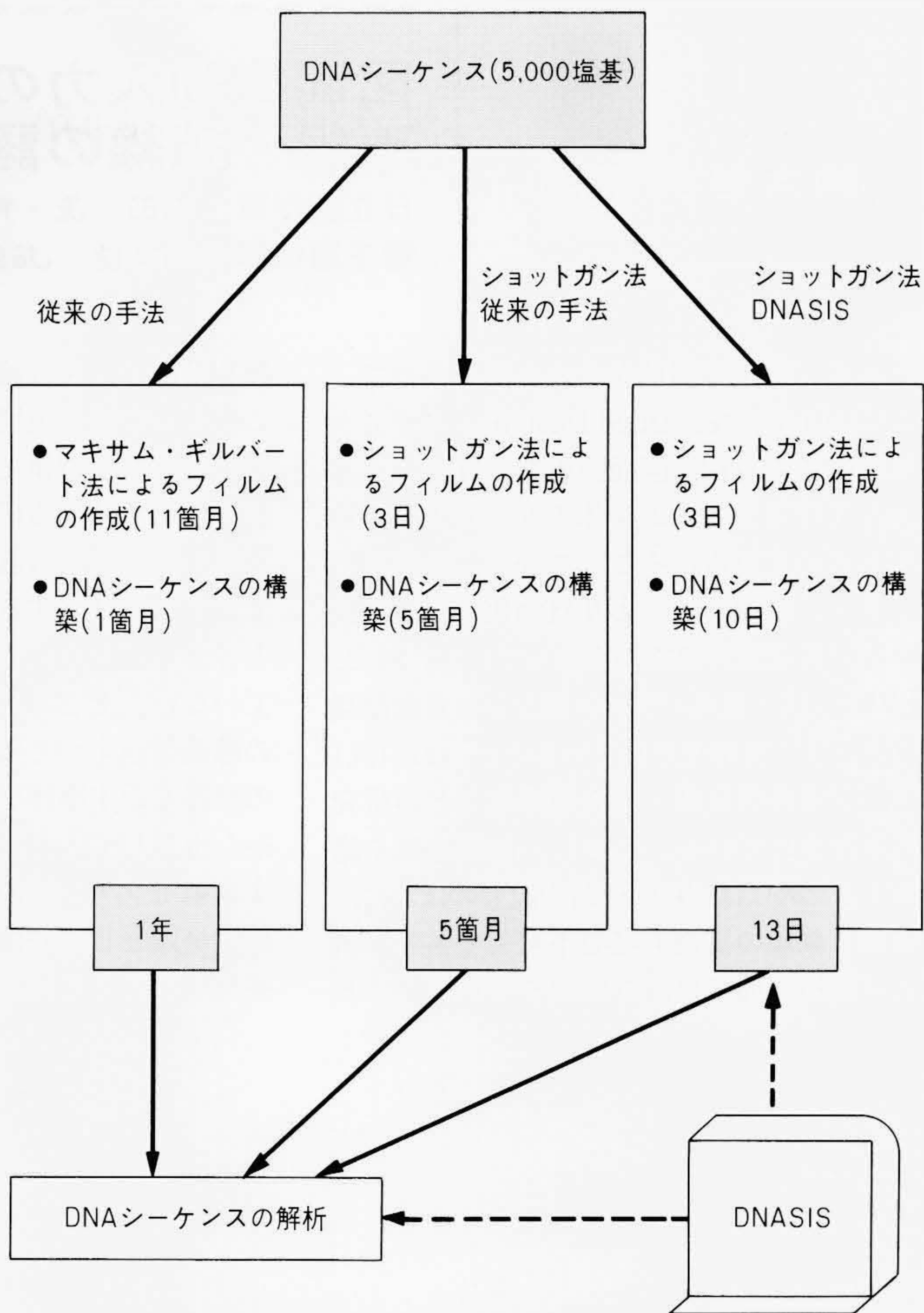


図10 システム導入効果比較 DNASIS導入効果を、従来の手法と比較して示す。

1箇月を、またショットガン法では5箇月を要する。そこで本システムを導入した場合、本構築は10日でできる。これら従来の手法との効果比較を図10に示す。

5 結 言

バイオテクノロジー分野の発展への期待は非常に大きいものがある。DNASISは、本分野の研究を支援するシステムとして安価なマイクロコンピュータで実現したもので、研究者はパーソナルユースが可能となり、その研究作業の雑務から解放された。今後共、日立ソフトウェアエンジニアリング株式会社は情報処理の観点から本分野の研究動向にフィットしたものを開発する考えである。

参考文献

- 1) 南雲：生体における情報処理，岩波講座，情報科学(1982年10月)
- 2) 添田：組換DNAの形質転換と生体機能，文部省特定研究(I) (1984年2月)



図面自動入力のための 破線・鎖線の認識アルゴリズム

日立製作所 嶋田 茂・角本 繁・他1名

電子通信学会論文誌 J69-D, 5, 759~770 (昭61-5)

現在、電子・機械などの各種の分野で、CAD(Computer Aided Design)システムが稼動中であるが、図面情報を計算機に入力する工程はシステム全体のなかで大きな比率を占めている。そこでこれを、計算機のパターン認識機能により自動化する、いわゆる図面認識の要求が高まっている。既に、LSIセル図・液晶パターン図などが自動認識技術により実用化されているほか、はん(汎)用的な回路図認識・等高線認識などの研究も報告されている。しかし、図面全体を認識するためには、実線経路や記号のほかに、破線・鎖線の認識も必要不可欠である。例えば機械図面では、外形線を実線、かくれ

線を破線で、中心線を鎖線で書いており、このような線の種別は、線の意味を与えるための重要な情報となっている。従来の破線や鎖線に関する研究は、一直線状に並んだ短い線列から破線を構成する線セグメントをグループ化するだけの実験的な段階にとどまっており、まだ実用の段階には達していない。そこで本論文では、既に図面内の線画像がすべて実線として認識され、線の経路を示す座標データになっている段階から出発して、これから破線や鎖線を構成する線セグメントの群を抽出し、破線・鎖線の経路として認識するアルゴリズムを提案する。ここでは、線セグメントの隣接関

係に着目したローカル(局所的)な認識処理系と、LSIセル図・地形図など各図面固有のシンタックスに基づき、図面全体が矛盾しないような認識を行うグローバル(大局的)な認識処理系とで構成する。特に後者の認識では、図面シンタックスに基づく処理をルール形式で記述することにより、図面固有の特殊性を吸収し、処理手続きがはん用化する点に特長がある。そして最後に、このアルゴリズムに基づいたプログラムを実際のLSIセル図と地形図の破線・鎖線の認識に適用した結果を示す。

エレクトロニクス産業から見た ナノメートル技術への期待

日立製作所 森山茂夫

計測と制御 25-5, 443 (昭61-5)

最近の半導体素子、情報関連機器を中心とするエレクトロニクス製品の著しい性能向上の一部は、半導体・電子・光学部品など構成要素の精密・微細化により達成されたものであり、その点から、よりいっそうの精密・微細化を進めるための加工・計測技術の開発が強く望まれている。

本稿では、エレクトロニクス産業界から早急に実用化が期待されている、ナノメートル領域の計測技術の幾つかについて紹介する。

(1) 微細表面形状計測技術

最近の超LSI製造工程では、膜厚が厚いレジストを加工したり、あるいは基板に深溝を掘ったりする3次元的微細加工技術が用いられ、その加工精度を評価するための微細表面形状計測技術が重要になっている。

これまでは電子顕微鏡を用い、所望位置

で切断された断面を観察していたが、これに代わる新しい非破壊計測技術の出現が望まれている。現在、電子ビーム技術及びSTM(Scanning Tunnel Microscope)技術などによるアプローチが研究されており、その将来が期待される。

(2) 高精度3次元形状計測技術

シンクロトロン放射光応用装置や半導体露光装置では、外径300mm程度の球面あるいは非球面の反射鏡が用いられ、その形状精度として数十ナノメートルのオーダーが必要とされている。

球面や球面からの偏差が小さい形状に対しては、干渉法など光学的計測技術が適用可能であるが、それ以外の任意形状に対しては、今のところ3次元座標測定による方法が唯一となっている。しかし、従来の3次元測定器では機械精度の制約から、サブ

ミクロン程度の測定精度が限界となっており、今後の飛躍的な精度向上が望まれている。

(3) 超精密位置検出技術

半導体素子や磁気ディスク・光ディスク装置などの製造工程では、超精密位置決め技術が重要な役割を果たしている。位置決め精度の向上には位置検出器の精度向上が不可欠であるが、現在、最も精度が高いとされるレーザ干渉測長器の場合でも、大気屈折率変動などの影響のために、実的に得られる精度は100mm当たり数十ナノメートルが限界となっている。環境変化に強いレーザ干渉測長技術とともに、レーザ干渉測長器に代わる新しい小形・低価格な超精密位置検出器の出現が期待される。