

した。染色体バンドと言う光学顕微鏡レベルの現象が、遺伝子塩基配列と直接に関係を持つ可能性が示唆されたことになる。

3) マウス脳に存在する RNA の高分解能分離と遺伝子クローニング (池村・青田): 脳で発現する遺伝子群を総合的に解析することを目的に、マウス脳に存在する RNA 類を二次元ゲル電気泳動法で高分解能に分離することを試みた。本年度は約 300 ヌクレオチド長以下の比較的分子量 RNA の分離を行った所、数百個の分離したスポットを得た。それらの一部について塩基配列決定を行った。

上記の実験と並行して、先きにマウス脳の poly (A)⁺ RNA 画分を鋳型に作成した cDNA ライブラリーについて、数種類のクローンの塩基配列決定を行った。その内の 1 種類はレトロポゾン IAP であることが判明した。この pol 領域内の塩基配列は Martens らがマウス IgE-BF 遺伝子として報告した塩基配列の一部とほぼ同一な配列を持っていた。マウス IgE-BF 遺伝子が形成された過程における特異な機構の存在、ないしは彼等の塩基配列決定実験の問題点に関係すると思われる。

G-c. 遺伝情報分析研究室

1) DNA データベース導入 (丸山・五条堀): DNA データバンク構築のため、欧州及び米国からデータベースを導入し、日本国内の利用者へ配布している。現在、EMBL (欧州) の第 6 版, 約 4800 遺伝子, 457 bp, GenBank (米国) 38.0 版, 約 6000 遺伝子, 550 bp, 及び NBRF (米国) のタンパク質データベース, 3400 遺伝子, 78 アミノ酸残基が導入されている。これらの DNA 及びタンパクデータベースは、磁気テープあるいはパソコン用フロッピーディスクを媒体に用い、利用者が必要とする形式に改訂し、提供することが可能である。

2) DNA データベースの構築 (丸山・五条堀): DNA データバンクの任務の 1 つは、データベースの構築であり、昭和 59 年 6 月から、DNA 塩基配列を電算機に入力する作業を試験的に開始した。京都大学化学研究所の大井教授の形式を用いてデータベースを作成している。この形式は米国の GenBank の形式にほぼ準ずるものである。なお、論文の選出は、京大大井教授を介し、蛋白質研究奨励会に依頼している。過去 7 ヶ月の間に、仮入力ではあるが、約 800 遺伝子 60 万 bp のデータを入力した。

3) DNA データファイルを検索するプログラムの開発 (丸山・五条堀): DNA データバンクに納められている遺伝子数は 6000 以上であり、目的に沿ってこれらのファイルを検索し、データを処理することは大切な業務の 1 つである。EMBL (欧州), GenBank (米国), DDBJ (日本) のファイルを対象に、遺伝子名、著者名、キーワード、塩基配列、生物名などを指定し、該当するファイルを“AND”と“OR”の論理で検索するプログラムを作成した。このプログラムはパッチジョブまたは対話型で実行でき、目的とする遺伝子群のファイルのみを抽出して 1 つのファイルを作成するのに役立っている。

4) 解析プログラムの開発と導入 (五条堀・丸山): DNA 配列及びアミノ酸配列データベースを有効に利用するための解析プログラムが、約 60 種類以上開発された。例えば、

各データバンクにおいて、先に詳述された特定の配列や遺伝子座の検索プログラムをはじめ、相同性を持つ配列のサーチング、パリンδροーム構造の予測、制限酵素地図の作成、配列のアライメント、コドン頻度、塩基置換数の推定、系統樹の作成、制限酵素地図から DNA 多型の量的解析、等のプログラムがある。特に、相同性をもつ配列のサーチングプログラムは、DNA 配列あるいはその相補的配列が GenBank と EMBL の両方の DNA 配列データバンクで、またアミノ酸配列が NBRF のアミノ酸配列データベースで、比較的短時間にサーチングすることができる。さらに、結果もドットマトリックス方式による出力で視覚的に判断が付き易い。このため、未知の DNA 配列の同定や機能の予測に著しい実績をあげている。この他、スターデンによって開発された種々のプログラム（例えば、ショットガン方式によって得られた DNA フラグメントの整合検索プログラム等）、約 50 種類が、九州大学の久原博士の協力によって登録されている。また、ウィルバーとリップマンによって米国製コンピュータ VAX 用に開発された DNA 配列サーチング・プログラムとアライメント・プログラムが富士通社の協力により、日本で初めて本研究所の日本製コンピュータ F160M 機で動作するようになった。特にアライメント・プログラムは原則的 20,000 bp の長さをもつ 2 つの DNA 配列をギャップを入れて並べる能力をもつ。実際、それぞれ約 5,000 bp の長さの 2 つの DNA アライメントを約 10 分間で行なった実績をもち、今後の種々の研究に役立つことが期待される。

II. 遺伝情報の収集保存

DNA データバンクに集録されている配布可能な核酸およびタンパク質データベース.

Gen Bank (DNA データ) 29.0 版

約 4,700 遺伝子 約 400 万ベース

EMBL (DNA データ) 4 版

約 1,700 遺伝子 約 215 万ベース

DDBJ (DNA データ)

約 4,400 遺伝子 約 370 万ベース

NBRF (DNA データ) 23 版

約 1,500 遺伝子 約 260 万ベース

NBRF (タンパク質) 4.0 版

約 3,000 遺伝子 約 65 万残基