

DNA トポイソメラーゼ II (Hirose, S. *et al.* (1988) *J. Biol. Chem.* **263**, 3805-3810) であった。一方、素通り画分に含まれる成分は DNA 超らせん化因子と名付けられ、均一な状態にまで精製された。この因子は分子量 50 kd の酸性タンパクで、DNA より少ない分子数で超らせん化反応を起すので、触媒的に働くと考えられる。

(2) 真核生物の遺伝子発現制御 (廣瀬・田淵・水谷): カイコ後部絹糸腺抽出液によるフィブリン遺伝子の転写は Sarkosyl に対する感受性に基づき、次の 3 段階に分けられる: ① 0.025% Sarkosyl に感受性な転写開始複合体形成, ② 0.05% Sarkosyl に感受性な開始複合体から RNA 鎖伸長複合体への変換, ③ その後の RNA 鎖伸長。②と③の反応は速く、鋳型 DNA のトポロジーによって影響を受けなかったのに対し、①の段階は転写の律速段階であり、鋳型 DNA の超らせん化により著しく促進されることが判明した (Tabuchi, H. and Hirose, S. (1988) *J. Biol. Chem.* **263**, 15282-15287)。開始複合体は、TFIIA, TFIIIB, TFIID, TFIIE と RNA ポリメラーゼ II から成ることが知られている。そこで、これらのタンパクを HeLa 細胞核抽出液から部分精製して再構成転写系を組み、それに完全に精製した DNA トポイソメラーゼ II と超らせん化因子を加えて解析した。その結果、TFIID (TATA 結合タンパク) のプロモーターへの結合が開始複合体形成の律速段階であり、鋳型 DNA の超らせん化によりこの段階が加速されることが分った。

(3) カイコの *fushi-tarazu* 遺伝子に関する研究 (上田・孫・廣瀬): 胚発生における形態形成の機構を分子レベルで解明する端緒として、カイコの *fushi-tarazu* 遺伝子のクローニングを試み、候補となるクローンを得たので現在解析を進めている。また、ショウジョウバエ *fushi-tarazu* 遺伝子の発現制限領域に結合する因子 NF ftz と同一の塩基配列を認識するタンパクが、カイコ後部絹糸腺抽出液中にも存在することが分り、その精製を開始した。

(4) マウスホメオティック遺伝子の発現とクロマチン構造 (浦・北川・廣瀬): 胚発生における形態形成の機構を探る別のアプローチとして、マウスのホメオボックス遺伝子 *Hox 2.1* の発現について調べた。F9 細胞をレチノイン酸処理して分化を誘導すると、誘導前には検出されなかった *Hox 2.1* mRNA が 24 時間をピークとして出現し、5 日目には消失することが分った。トポイソメラーゼ II の特異的阻害剤 VP-16 により、この発現は抑えられ、*Hox 2.1* 遺伝子の発現に DNA の高次構造が関与していることが示唆された。

(5) アデノウイルス初期遺伝子の転写調節因子の解析 (半田・広瀬): アデノウイルス E4 遺伝子の上流に存在するエンハンサーは、カイコ後部絹糸腺抽出液を用いて DNA を超らせん化させると *in vitro* でも転写促進機能を果す。エンハンサー領域内のいろいろな部分を欠失した DNA の転写活性を測定した結果、鋳型の超らせん化による転写促進にはサイクリック AMP 応答領域が関与していることが示唆された。現在、部分精製した転写因子を用いて転写系を再構成し、より詳細な機構を解析している。

G-d. 遺伝情報分析研究室

当研究室 (宮沢・林田) は、遺伝情報の研究及び DNA データバンクの運営 (データベ-

ス構築, 検索, 解析プログラムの開発, データベースの配布と計算機のオンライン利用のサポート, 講習会, ニュースレター(の出版等)に従事した。

1988年2月15日~16日 NIH (米国, ベセスダ) で第一回データバンクのための国際諮問委員会が開かれ, DDBJ を代表して宮沢が参加した。また宮沢は7月4日~8日 EMBL (ハイデルベルグ) で開かれたデータバンクの会合に出席し, DNA 配列の注釈に関する書式の共同作成に参加した。さらに, 宮沢, 林田は9月5日~15日 EMBL (ハイデルベルグ) で開かれたデータバンクの年会に出席し関係データベースの共同構築に関する議論に参加した。議題は関係データベース設計, データ交換のための方式, 研究者自身によるデータ入力のための支援ソフトウェア, データベースの同一性をどのようにして保つか, 研究者自身によるデータ入力のための支援ソフトウェア, CD-ROM のためのデータフォーマット等である。

宮沢は7月11日~9月2日米国 NIH に滞在し蛋白質及び DNA のコンフォメーションの研究に従事した。また12月12日~16日米国サンタフェで開かれた“核酸塩基配列解析と計算科学との間の接点”ワークショップに出席した。

海外からは, NIH の GenBank 予算担当官 J. Cassatt 博士が10月19日, EMBL Biocomputing Programme 所属の C. Sander 博士が10月24日に研究所を訪問され, データバンクに関して意見を交換した。

(1) 日本 DNA データバンク (DDBJ) 活動

(i) ニュースレターの発行 (丸山・宮沢・林田): DNA データバンク活動の報告のため, 2月ニュースレター No. 7 を発行した。今年度 493 部配布した。

(ii) DDBJ 利用者講習会 (宮沢・林田): 計算機初進者のための講習会を6月17日~18日に開催した。分子遺伝研究部門の藤田氏の協力を得た。受け入れ可能な人数 (20人) の2倍程の応募があった。

(iii) DDBJ 計算機利用デモンストレーション: 9月20日~22日癌学会 (進化遺伝学部門五條堀他), 10月5日~6日生化学会 (宮沢), 12月20日~23日分子遺伝学会 (林田) で計算機利用によるデータ提出及び解析プログラムの使用に関してデモンストレーションを行った。その後, 研究者自身によるデータの提出が増加の傾向にある。

(iv) DNA データベースの導入 (宮沢・林田): 米国から GenBank, NBRF データベース, 欧州から EMBL データベースを磁気テープで取り寄せ, 希望者に配布している。配布媒体は GenBank の場合は磁気テープとフロッピーディスク, その他は磁気テープのみである。磁気テープの配布総数は 612 本, フロッピーディスクの配布枚数は 572 枚である。端末エミュレータープログラム (Kermit) を 19 件, フロッピーにして 57 枚希望者に配布した。

(v) DNA データベースの構築 (宮沢・林田): 1月に2版 (142 エントリー, 199, 392 塩基), 7月に3版 (230 エントリー, 345, 850 塩基) をリリースした。研究者自身によるデータ作成に向けて, 幾つかの学術雑誌と論文受理の際研究者にデータ提出を呼びかける協定を結んだ。また日本で生産される DNA 配列データの入力に向けて EMBL, Gen-

Bank 担当の論文雑誌であっても収集する試みを開始した。

(vi) “The DNA/EMBL/GENBANK Feature Table: Definition, version 1” (EMBL Data Library, GenBank, DDBJ (宮沢・林田)): EMBL Data Library, GenBank と共同で DNA データベースの注釈に関する定義マニュアルを完成させた。新注釈フォーマットは GenBank と EMBL フォーマット間の共通項目を増やすだけでなく、従来のフォーマットにおける多くの問題点を解決し、また近年明らかになった新知識を注釈として計算機可読な形で表現することを可能にする。新注釈フォーマットの採用に向けて、各種マニュアルの作成、旧注釈フォーマットを新フォーマットに変換するためのソフトウェア、新フォーマットで入力するための支援ソフトウェア等をデータバンクで共同して開発する計画である。

(vii) DNA Data Bank of Japan 共同利用電子計算機利用の手引 (宮沢): DDBJ 計算機システムのオンライン利用支援のため online information retrieval system (“getinfo”) を開発した。これによりデータバンク活動に関する広報及び計算機システム利用に関する情報のオンラインによる提供が可能となった。また特別なアカウント (ddbnews) で DDBJ 計算機システムをアクセスし “getinfo” の指示に従うことにより、誰でも容易に解析した DNA データをデータバンクへサブミッションすることが可能である。“getinfo” のハードコピー版である利用の手引は DDBJ により希望者に配布されている。

“getinfo” は VAX/VMS システムの help コマンドに一見似ているが、help とは異なり個々の情報は単一のファイルに納められ、必要な場合 symbolic link または pseud symbolic link を用い tree-like なファイル構造として管理される。利用者は得たい情報をメニューの中から選択する。getinfo は指定された情報を画面に表示し、また次のレベルの情報が利用可能な場合はそのタイトルを表示する。ファイルは UNIX 付属ツールである pager (pg または less) を使用して表示するので、ページごとに表示されると同時にもし必要ならファイルとして得ることも可能である。

(viii) The Manual of the Flat Database and Sequence Analysis System for DNA and Proteins, version 1.0 (beta) (宮沢): UNIX システムの上で稼働する DNA 及び蛋白質配列データベースのための検索システムを開発した。このシステムは簡単な機能を果たす様々なツールからなる。基本ツールの例は、

- 指定されたタイプのレコードを出力する。
- 指定された文字列を含むエントリーのエントリー名を出力する。
- エントリー名からなるセットに関する and, or, xor.
- 指定されたエントリーを出力する。

等である。ほとんどのツールは UNIX におけるフィルターとして働く。このようなツールを UNIX のパイプで組み合わせることにより、著者名、論文名、生物種、遺伝子名、キーワード等による検索が可能である。文字列は通常 UNIX の正規表現で指定する。よってあいまいな文字列による検索が可能である。また特異な塩基配列をもつ遺伝子の検索においても塩基配列を正規表現で表現できるプログラムを作成した。ファイルシステムとし

ては保守の簡単なフラットファイルを用いた。このシステムは、大型、小型、パーソナルコンピューターを問わず UNIX システムなら移植可能であるという利点を持つ。

またホモロジー検出プログラムパッケージ Fasta (Pearson & Lipman 作成) を移植した。

(ix) The Manual of the Qanalys Sequence Analysis System for Molecular Evolution, version 1.0 (beta) (林田): DNA 及びアミノ酸配列の分子進化学的解析のためのソフトウェアパッケージ。残基含量等の計算の他、配列間の相違度の計算、Alignment用のソフトウェア及び計算結果の表示用のソフトウェアを含む。系統樹作用のソフトウェアは現在移植中。

(2) Score Matrix の評価 (宮沢・R. L. Jernigan (NIH)): ホモロジー検出等で使用されるアミノ酸置換に関する Score Matrix をアミノ酸置換によるアミノ酸間相互作用エネルギーの損失を平均値として見積ることにより評価した。アミノ酸間相互作用エネルギーの残基依存性は、立体構造が既知の蛋白質において観測されるアミノ酸残基間接触数から評価したアミノ酸間接触エネルギーとして近似した。核酸塩基置換の塩基タイプ依存性は無視し、塩基置換は平衡状態にあると仮定した。Score Matrix は 250 PAM に対応するアミノ酸置換行列の評価から計算し、ホモロジー検出に使用した結果、Dayhoff 等により評価されたものとはほぼ同じ程度の検出力を示した。これは 7~8 月に米国 NIH でした研究である。

(3) 局所探査法の開発 (林田): 分子進化学の知見を利用し進化的に不変なアミノ酸配列上の座位に着目することによって、従来の相同性探査法より高性能の相同性探査法を開発した。この方法により、新しい Ig 超遺伝子族に属する遺伝子を 2 種発見した。

H. 放射線アイソトープセンター

当センターは昭和 63 年 4 月に発足した。前身は、およそ 30 年前に設立された放射線実験室である。センター長には変異遺伝部門から移った定家義人助教授が就任し、原登美雄技官、4 月に新規採用された谷田勝技官と共に次の研究を行った。また米国カンサス大学 D. M. Shankel 教授は 12 月に大仁で開かれた第 2 回国際突然変異、発ガン抑制機構会議の準備のため滞在し、当研究室で以下の共同研究を行った。さらに相川練二 (東洋醸造)、岡部正実 (協和醸酵) は受託研究員として、可徳小四郎 (株式会社九動) は研修生として当研究室で研究、研修を行った。オランダ、ライデン大学 Sankaranarayanan 博士とは放射線遺伝学に関する研究討議を行った。当センターは上記国際会議の庶務を行った。

本年度の当センターにおける研究活動は以下の通りである。

(1) 線虫 *Caenorhabditis elegans* 生殖細胞における DNA 修復機構に関する研究 (定家): 生殖細胞における DNA 修復機構の解明には、分子遺伝解析可能な変異株があり、多数扱え、充分な数の子を作り、生殖細胞の起源がはっきりしている材料が望ましい。線

Q β

Bacillus のファージ

PBS 1, SP 10, SPO 1, SPO 2 など

K. 培養細胞

1. 低増殖性2倍体細胞

ヒト胎児肺細胞 15 株

2. 高増殖性正常細胞

ヒト胎児肺細胞 8 株

チャイニーズ・ハムスター肺細胞 Don-6, クローン株 10 株

チャイニーズ・ハムスター肺細胞 V 79, クローン株 4 株

マウス繊維芽細胞 5 株

3. 高増殖性癌細胞

ヒト子宮癌細胞 HeLa S 3, クローン株 3 株

ラット肝癌細胞 10 株

4. 薬剤抵抗性および修復欠損変異細胞

ヒト・レッシ・ナイハン症由来細胞 (8-アザグアニン抵抗性) 3 株

ヒト・色素性乾皮症由来細胞 (修復欠損) 5 株

シリアン・ハムスター・5-ヨードウリジン抵抗性細胞 5 株

チャイニーズ・ハムスター・8-アザグアニン抵抗性細胞 10 株

チャイニーズ・ハムスター・6-チオグアニン抵抗性細胞 12 株

チャイニーズ・ハムスター・5-ヨードウリジン抵抗性細胞 25 株

チャイニーズ・ハムスター・金属塩抵抗性細胞 10 株

II. 遺伝情報の収集保存

現在 DDBJ で利用可能な核酸および蛋白質データベースは以下のものである。今年より組み換え研究室青田、池村氏作成のコードン使用頻度データベースと、分子生物学関係のデータベースをリストしたデータベース (LiMB) が利用可能となった。

DNA 塩基配列データ:

DDBJ	3 版 (07/88)	230 エントリー	345,850 塩基
EMBL	17 版 (11/88)	20,695 エントリー	24,211,054 塩基
GenBank	58.0 版 (12/88)	21,248 遺伝子	24,690,876 塩基
NBRF	34.0 版 (11/88)	3,069 エントリー	6,735,091 塩基
HIV-N	1988 年版		
KABAT	1983 年版		
Miyata	1988 年 3 月版		

蛋白質アミノ酸配列データ:

DDBJ	3 版 (07/88)		
PIR	18.0 版 (09/88)	5,556 蛋白質	1,510,026 残基
SWISSPROT	7 版 (04/88)	6,821 蛋白質	1,885,771 残基
HIV-N	1987 年版		
KARAT	1983 年版		

コドン使用頻度データベース: (GenBank 50.0 版に対応)

LiMB (Listing of Molecular Biology Database):

1 版 (02/88)

以下は各データベースの簡単な収集内容である。

1. GenBank Release 58.0

グループ	エントリー数	塩基数
霊 長 類	3,255	4,001,807
ゲ ッ 歯 類	3,660	3,626,245
哺 乳 類	727	840,709
脊 椎 動 物	947	956,226
無 脊 椎 動 物	1,503	1,652,160
植 物	1,487	2,079,610
オ ル ガ ネ ラ	685	1,185,381
バ ク テ リ ア	2,022	3,010,657
R N A	1,033	141,936
ウ イ ル ス	1,957	3,139,106
フ ァ ー ジ	400	494,730
人 工 合 成	574	210,001
無 注 釈	2,998	3,352,308

2. EMBL Release 17

グループ	エントリー数	塩基数
人 工 合 成	453	166,731
ク ロ プ ラ ス ト	269	614,737
Genetic elements	83	80,672
ミトコンドリア遺伝子群	542	596,193
原 核 生 物	2,255	2,862,227
ウ イ ル ス / フ ァ ー ジ	2,245	3,433,080
真 核 生 物	11,328	12,604,460
そ の 他	43	60,854
無 注 釈	3,477	3,792,100

3. NBRF Release 34.0

グループ	エントリー数	塩基数
真核生物	1,669	3,150,511
哺乳動物	974	1,840,248
植物と真菌類	310	736,948
真核生物ウイルス	575	2,092,124
原核生物	721	1,192,250
バクテリオファージ	104	300,206
動物ウイルス	519	1,933,693
植物ウイルス	56	158,431
大腸菌	457	839,054
真菌類	207	323,525
人類	380	925,184
ミトコンドリア	67	212,009
クロロプラスト	32	326,191

4. PIR Release 18.0

グループ	エントリー数	塩基数
真核生物	3,209	703,134
哺乳動物	1,768	411,551
植物	384	77,506
真菌類	192	64,402
原核生物	1,048	266,685
動物ウイルス	931	441,239
植物ウイルス	69	30,212
バクテリオファージ	301	69,994