

G-d. 遺伝情報分析研究室

本研究室では、DNA 配列データを主とする遺伝情報を対象として、コンピュータや理論的手法を用いたデータ解析を行なっている。また、研究事業として、DNA 配列データベースの構築を、日本 DNA データバンク (DNA Data Bank of Japan: DDBJ) として、アメリカ合衆国の GenBank とヨーロッパ共同体の EMBL データ・ライブラリーと共同して行なっている。

本年は、宮沢三造助教授と林田秀宜助手が本研究室の教官として活動したが、7月16日付で五條堀孝が教授として進化遺伝研究部門から本研究室に移った。また、11月1日付で、鶴川義弘が理化学研究所研究員から本研究室助手として赴任した。また、遺伝実験生物保存研究センター遺伝資源研究室の館野義男助教授がデータバンク研究事業 (DDBJ) に意欲的に協力した。本研究室の研究及び DDBJ 活動の補佐を、堀江元乃、鈴木成子、岩瀬正子、佐藤由美子、下山メアリー、野口由利子、山崎由紀子、鈴木利紀子が行なった。

日本 DNA データバンク (DDBJ) 活動

1990年2月5日-10日、GenBank (米国、ロスアラモス) でデータバンク (DDBJ/EMBL/GenBank) の実務協議が開かれ、DDBJ を代表して宮澤が参加した。議題は現行データベースから関係データベースへの移行及び新注釈テーブルへの移行のスケジュール、関係データベース移行後における EMBL/GenBank/DDBJ 間でのデータベース交換のための手順の決定、及び DDBJ の念願であるデータ収集における学術雑誌分担方式から地域分担方式への移行のためのスケジュールが議論された。また1990年3月15日-17日には、三島でデータバンクのための第三回国際諮問委員会が開かれ、EMBL/GenBank/DDBJ より関係データベースへの移行とデータ収集における地域分担方式への移行が提案され賛意を得た。

(a) ニュースレターの発行 (宮澤・林田): DNA データバンク活動の報告のため、1990年5月ニュースレター No. 9 を発行した。

(b) DNA データベースの導入 (宮澤・林田): 米国から GenBank データベース、欧州から EMBL, SwissProt データベースを磁気テープで取り寄せ、希望者に配布した。

(c) DNA データベースの構築 (宮澤・林田): 1990年1月に6版 (496 エントリー、841,236 塩基)、1990年7月に7版 (681 エントリー、1,154,211 塩基) をリリースした。1990年1月から7月までの期間に処理したデータは全世界で処理されたデータ量の約11%である。この処理量はほぼ日本で解析された DNA 配列データに相当する。その内約44%は7版でリリースされ、約21%は入力処理されたものの論文が出版されるまでリリースが延期されているもの、残りの約37%は EMBL Data Library との約束で EMBL に転送されたデータである。今年度における処理件数の増加は、データ収集の地域分担方式に向け、GenBank 担当の論文雑誌に発表されるデータも DDBJ に提供されたものに関しては1989年末より入力処理を開始した結果である。

(d) DNA データ入力、管理システムとデータ検索システムの構築、ネットワークデー

データベース機能と電子メールによるデータベースの自動更新機能の追加 (宮澤): データベース検索ソフトウェア FLAT にコマンドを電子メールで送付することによりデータベースを検索することが可能なネットワークデータベースサーバー機能と電子メールによるデータベースの自動更新機能を追加した, Flat Database and Sequence Analysis for DNA and Proteins, version 1.3 をリリースした. このソフトウェアを使用して DDBJ は勿論のこと EMBL/GenBank が入力した最新のデータを DDBJ 計算機の利用者及び電子メールを用い全世界に提供している.

(1) アミノ酸置換による蛋白質構造の安定性変化の予測 (宮沢・R. L. Jernigan (NIH)): 単一アミノ酸置換による蛋白質の安定性の変化を予測する簡単な経験的方法を提唱する. この方法は蛋白質構造は変化しないと仮定した時, 置換されたアミノ酸と周囲のアミノ酸間の相互作用エネルギーがどの程度変化するかを 20 種のアミノ酸間での接触エネルギーの見積もりから予測する. Yutani *et al.* (1987) 及び Matsumura *et al.* (1988) 等は, 各々 tryptophan synthase a subunit と bacteriophage T4 lysozyme において単一アミノ酸置換による unfolding Gibbs free energies の変化を測定し疎水性残基への置換においてはその疎水性の程度とよく相関することを示し, 安定性の変化は疎水相互作用の変化に起因すると説明した. ここで提唱する方法では, 疎水性残基への置換だけでなく 20 種のアミノ酸すべてについて予測値は実験値とよく相関することがわかった. また tryptophan synthase a subunit の場合, Yutani *et al.* (1987) による疎水性の変化の見積もりは unfolding Gibbs free energies の変化に比べ小さすぎるが, この方法はより妥当な値を生むことが示された.

(2) ホモロジーサーチのための scoring matrix (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 Scoring 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. Scoring 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, グローバルホモロジーサーチ及びローカルホモロジーサーチに使用した結果, 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.

G-e. 遺伝子ライブラリー研究室

本研究室では, 遺伝子ライブラリーの構築, 管理, 配布という業務と, このための新しい方法論の開発を行い, 並行して, 遺伝子ライブラリーを活用して, 動物発生過程の遺伝子発現ネットワークの解明をめざす研究を始めている.

研究室構成としては, 2 月に小原雄治助教授が英国への長期出張から帰国し, 研究を開始した. 前年にひきつづき, 永田妙子が研究補助業務を行ったが, フランス留学のため, 7 月 31 日付で辞職した. 後任として, 10 月 11 日より西垣明子と西村かよ子が研究補助

で標識されている): 473 株

- a) 遺伝的背景の異なる株 (1983~1987 年の Journals に掲載された株のコレクション): 203 株
- b) 遺伝的背景が野生型の株 (Singer *et al.* 1989. Microbiol, Rev. 53: 1-24 に掲載された kit): 190 株
- c) Hfr 株の kit: 80 株
- (3) クラーク・カーボンの pLC コレクション* (合成 Co1E1 ハイブリド・プラスミド 2,000 種を含む大腸菌のジーン・バンク. Clarke & Carbon. 1976. Cell 9: 91-99): 2,000 株

* Nishimura, A. et. al.: Correlation of PLC-plasmids to the physical map of *Escherichia coli* K12. in preparation.

- (4) 広田の大腸菌温度感受性変異株のコレクション: 約 5,000 株
- | | |
|--------------|-----------|
| DNA 複製欠損変異株 | 115 株 |
| RNA 合成欠損変異株 | 100 株 |
| ムレイン生合成欠損変異株 | 55 株 |
| 細胞分裂欠損変異株** | 353 株 |
| 染色体分配欠損変異株** | 45 株 |
| 膜蛋白欠損変異株 | 22 株 |
| リボソーム蛋白変異株 | 79 株 |
| 未同定欠損変異株 | 約 3,800 株 |

** Nishimura, A. et. al.: Mapping of a whole set of cell division genes in *Escherichia coli* K-12. In "Control of cell growth and division". H. Yoshikawa and A. Ishihama eds. Japan Sci. Soc. Press, Tokyo/Springer-Verlag, Berlin (1991) (pp. 205-223).

- (5) *Escherichia* のファージ: T₂, T₃, T₄, T₄GT7, T₅, T₆, T₇, Pl·kc, Pl·vir, Mu, λpapa, λvir, λgt·λC, λcb2, λcI₈₅₇·S7, λTn5, λTn10, φX174wild, φX174am3, f1, Qβ, その他

その他

Bacillus subtilis (枯草菌): 200 株

Salmonella typhimurium (ネズミチフス菌): 1,370 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ようである。

DNA 塩基配列データ:

| | | | |
|---------|----------------|--------------|---------------|
| DDBJ | 8 版 (01/91) | 879 エントリー | 1,573,442 塩基 |
| EMBL | 25 版 (11/90) | 41,580 エントリー | 52,900,354 塩基 |
| GenBank | 67.0 版 (03/91) | 43,903 遺伝子 | 55,169,276 塩基 |

| | | | |
|--------|----------------|-------------|--------------|
| NBRF | 36.0 版 (03/90) | 3,355 エントリー | 8,128,496 塩基 |
| HIV-N | 1988 年版 | | |
| KABAT | 1987 年版 | | |
| Miyata | 1988 年 3 月版 | | |

蛋白質アミノ酸配列データ:

| | | | |
|-----------|----------------|------------|--------------|
| DDBJ | 8 版 (01/91) | | |
| PIR | 27.0 版 (12/91) | 26,798 蛋白質 | 7,620,668 残基 |
| SWISSPROT | 16 版 (10/90) | 18,364 蛋白質 | 5,986,949 残基 |
| KABAT | 1987 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):

2.0 版 (08/90)

以下は各データベースの簡単な収集内容である。

1. GenBank Release 67.0

| グループ | | | エントリー数 | 塩基数 |
|------|---|-----|--------|-----------|
| 霊 | 長 | 類 | 8,206 | 9,814,969 |
| ゲ | ッ | 歯類 | 8,400 | 8,546,574 |
| 哺 | 乳 | 類 | 1,638 | 2,077,398 |
| 脊 | 椎 | 動物 | 1,965 | 2,249,342 |
| 無 | 脊 | 椎動物 | 3,383 | 4,307,294 |
| 植 | | 物 | 3,187 | 5,069,146 |
| オ | ル | ガネラ | 1,402 | 2,021,305 |
| バ | ク | テリア | 4,616 | 7,572,518 |
| R | | N A | 1,735 | 518,967 |
| ウ | イ | ル | 4,032 | 7,019,564 |
| フ | ァ | ー | 602 | 705,269 |
| 人 | 工 | 合成 | 1,053 | 550,086 |
| 無 | | 注 | 3,684 | 4,716,844 |

2. EMBL Release 25

| グループ | | | エントリー数 | 塩基数 |
|------|----|-----|--------|-----------|
| バク | テリ | オフ | 654 | 786,251 |
| 真 | 菌 | 類 | 1,568 | 2,455,385 |
| 無 | 脊 | 椎動物 | 3,229 | 3,988,907 |
| オ | ル | ガネラ | 1,921 | 2,689,062 |
| 哺 | 乳 | 類 | 1,683 | 2,144,554 |
| 脊 | 椎 | 動物 | 2,012 | 2,336,080 |
| 植 | | 物 | 2,256 | 3,241,216 |

| | | | | |
|---|---|-----|-------|-----------|
| 霊 | 長 | 類 | 7,781 | 9,431,997 |
| 原 | 核 | 生 物 | 5,505 | 8,148,434 |
| ゲ | ッ | 菌 類 | 7,229 | 7,737,436 |
| 人 | 工 | 合 成 | 957 | 528,212 |
| 無 | 注 | 釈 | 3,011 | 2,565,069 |
| ウ | イ | ル ス | 3,774 | 6,847,751 |

3. PIR Release 27.0

| グループ | | エントリー数 | 残 基 数 |
|------|---------|--------|---------|
| 真 | 核 生 物 | 3,827 | 907,467 |
| | 哺 乳 動 物 | 2,071 | 518,330 |
| | 植 物 | 548 | 119,334 |
| | 真 菌 類 | 235 | 84,065 |
| 原 | 核 生 物 | 1,547 | 423,198 |
| 動 物 | ウ イ ル ス | 1,808 | 901,924 |
| 植 物 | ウ イ ル ス | 121 | 54,007 |
| バクテリ | オファー | 447 | 101,953 |