

# 일본 JST의 고기능 기반 DB 구축 사업 소개

박영서(E-mail:ygs@skkku.ac.kr)  
산업기술정보원 산업정보분석실장

- I. 서 론
- II. 고기능 기반 DB의 개발
- III. JST의 고기능 기반 DB와 복합 슈퍼컴퓨터 시스템

## I. 서 론

결국 더 다양화·고도화되고 있는 정보화에 대한 니즈와 부단히 진보하고 있는 정보기술에 대응하기 위해 일본 과학기술 정보 분야의 중추적 기관인 일본과학기술진흥공사(Japan Science and Technology Corporation, JST)은 계산과학이나 컴퓨터 그래픽(CG)의 발달·정보 유통 네트워크의 고도화를 배경으로 해서 물질·재료 및 생명과학의 분야에서 연구 지원형의 새로운 DB인 「고기능 기반 DB」를 개발하고 있다.

또한 슈퍼컴퓨터를 도입해서 계산재료과학 분야의 계산 물질 데이터 창제·생명과학분야의 대규모 염기 서열(sequencing)의 자원이나 재질 해석 등을 하는 환경구축을 계획하고 있다.

본 고에서는 이와 같이 슈퍼컴퓨터의 도입으로 새로운 전기를 맞고 있는 JST의 「고기능 기반 DB」 구축 사업을 소개한다.<sup>1)</sup>

## II. 고기능 기반 DB의 개발

JST는 「연구 정보 DB 개발사업」으로서 물질·재료분야 및 생명과학분야의 고기능 기반 DB를 개발하고 있다.

고기능 기반 DB는 신규 데이터를 추가하거나 세어 데이터를 보완하는 기능, 연구자가 직접 데이터를 등록할 수 있는 기능, 검색이나 출력용 코드로 편리하게 제공하는 기능 등을 주목적으로 하고 있다.

### 1. 고기능 기반 물질 DB의 개발 (물질·재료 분야)

JST의 고기능 기반 물질 DB의 개발은 합금 분야와 고분자 분야를 대상으로 해서 추진되고 있다. 물질·재료 분야를 생각할 때 원자·분자 레벨의 마이크로한 레벨로부터 매크로사의 마이크로한 현상까지가 대상이다.

예를 들어 물질의 구조라고 하는 기본 데이터로부터 그 물질의 성질을 예측하거나, 반대로 키지(觀測) 물질의 데이터를 기초로 유용한 기능을 가진 물질을 계산기상에서 상제하는 기능, 거시 데이터를 기초로 결실된 데이터를 추산·보완하는 기능, CG 기술을 이용해서 알기 쉬운 출력 결과를 얻는 기능, 그리고 네트워크를 통해서 많은 연구자가 간단하게 데이터를 제공하거나 이용할 수 있는 기능을 중심으로 개발하고 있다.

이러한 기능을 가진 것으로서, 결정 구조·교육 특성, 모선 회귀, 상태도의 각 데이터로 구성된 고기능

합금 DB 개발을 국제협력으로 추진하고 있다. 또한 고분자에 대해서도 물질적 뿐만 아니라 구조·형상의 상호·묘환 기능이나 해석·추산 기능을 가진 DB를 개발하고 있다.

합금 DB는 마이크로한 계산 물질 DB, 마이크로한 엔지니어링 DB, 그 중간에 있는 키즈 DB의 3가지 축으로 개발되고 있으며 금속계로 강도 DB, 압력용기 재료 DB, 3차원 상태도 표시 시스템, 계산 물질 DB, 기초적 DB 등이 있다.<sup>2)</sup>

고분자 DB PdyInfo는 고분자를 화학구조어 근거해서 식별·동정(同定)하는 시스템인 고분자 사관을 개발하고 있다.<sup>3)</sup>

## 2 고기능 기반 생체 DB의 개발 (생명과학 분야)

생명과학 부문에서는 JICST 시절인 1995년부터 코도르 복잡한 생체인 인간을 타겟으로, 가장 기초적인 생체 정보인 인간 게놈 DNA의 염기 서열 데이터를 기본으로 해서 펩 아미노산 배열, 단백질 구조 등 유전자 관련 데이터의 종합 검색이 가능한 고기능·기반·생체 DB (Advanced Lifesence Information System, ALIS)를 개발하고 있다.

관련 분야의 연구자가 인간 게놈 DNA 해독 데이터를 활용할 수 있는 시스템을 목표로, ① 대규모 인간 게놈 서열의 해독과 DB화, ② GDB 등 관련된 공공정보와의 통합, ③ 게놈 데이터 코드 이용을 위한 정보처리 기술을 개발하고 있다.

JST는 ALIS 프로젝트의 일환으로 인간 게놈 연구를 목적으로 하고 있는 외부 연구기관에 대해서 인간 게놈 특정 영역의 DNA 염기 서열을 위탁하는 게놈 프로젝트를 추진하고 있다.<sup>4)</sup>

외부 위탁기관은 담당 영역을 염기 서열화 함으로써 해당 영역의 DNA 염기 서열을 획득한다.

여기서 획득된 염기 서열 데이터는 위탁기관 내에서의 게놈 연구게코가 될과 동시에 염기 서열 데이터

상술에 관련된 마크 데이터가 JST에 제출된다. 이와 같이 해서 획득한 데이터를 체계적으로 정비·DB화 해서 네트워크 연구 정보 자원으로서 그 일부를 전세계에 공개하고 있다.

그리고 유전자 관련 데이터를 연구자가 유효하게 이용하기 위한 해석 기능 해석·결과의 표시 기능도 개발하고 있다. 해석 기능으로서의 DNA 염기 서열의 결합 연결 보경 기능·유전자 영역의 추정 기능 등이 있다.

또한, 네트워크를 이용해서 검색·해석·결과를 연구자가 간단한 조작으로 이용할 수 있는 환경구축도 개발하고 있다.

## 가. 고기능 기반 생체 DB 개발의 목표

고기능 기반 생체 DB 개발은 인간 게놈 해석·계획에 의해 얻어진 광대한 DNA 염기 서열·데이터를 DB에 수록해서 공개함과 동시에 그로써

데이터 해석·예측 등을 실시하는 응용소프트웨어를 갖춘 DB 시스템을 제공함으로써 염기 서열·데이터 중에서 유전자 등의 의미 있는 정보를 추정, 해석, 검색할 수 있는, 연구활동 지원 등의 개발을 목표로 하고 있다.

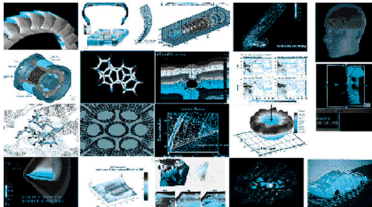
## 나. 고기능 기반 생체 DB의 개발 단계

고기능 기반 생체 DB는 다음과 같이 크게 4단계로 나누어 개발되고 있다.

### 1) 대규모 염기 서열 결정

고기능 기반 생체 DB의 키즈 데이터가 되는 100만 염기의 연속한 DNA 염기 서열 데이터를 해독하는 염기 서열 결정(sequencing) 작업을 한다. 염기 서열 결정 작업은 염기 서열 결정에 필요한 인간 게놈의 맵 및 기술을 보유한 연구자가 소속되어 있는

“ 슈퍼컴퓨터를 도입해서 · 계산재료과학 분야의 · 계산 물질 데이터 창제 · 생명과학분야의 대규모 염기 · 서열의 지원이나 배열 해석 · 등을 하는 환경정비를 계획 · 하고 있다. 이와 같이 · 슈퍼컴퓨터의 도입으로 · 새로운 전기를 맞고 있는...”



대학이나 연구기관에 위탁한다.

1995년부터 모델 데이터를 획득할 목적으로 대량의 게놈 서열 결정 콤팩트 프로그래밍을 연속 100만 염기 이상의 인간 게놈 서열 결정 작업을 4기관에 위탁하고 있다.<sup>2)</sup>

### 2) 기반 DB의 개발

DB 개발의 제 1단계는 염기 서열로 얻어진 DNA 염기 서열 데이터를 받아 평가, 공기하는 3개의 서브시스템(평가 서브시스템, 평가 서브시스템, 공기 서브시스템)으로 구성된 「기반 DB」를 개발하는 것이다.

### 3) 인간 게놈 통합 DB의 개발

고기능 기반 생체 DB의 중심 DB는 「인간 게놈 통합 DB」이다. 이 DB에는 2)에서 상술한 DB 외에 지도 정보, 유전자 배열 정보, 유전자 발현 정보 등 인간 게놈 해석 프로그래밍의 진전에 따라 축적되어 온 다양한 정보를 수집·평가·재 구축한 DB(기반 DB)가 포함된다.

### 4) 데이터 해석 소프트웨어의 개발

기반 DB-인간 게놈 통합 DB의 개발과 병행해서

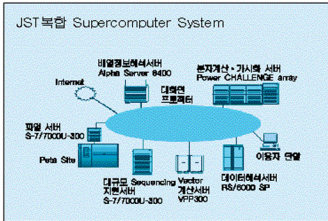
유전자 발현의 예측이나 유전자 제어 발현의 예측과 같은 고도의 게놈 정보를 해석하고, JST 독자적인 응용소프트웨어를 개발할 계획이다.

또한, 염기 배열·정보로부터 단백질 체인의 해석(기능 예측, 2차 구조 예측, 입체구조예측 등)을 수행할 수 있는 응용소프트웨어의 개발도 추진할 계획이다.

## III. JST의 고기능 기반 DB와 복합 슈퍼컴퓨터 시스템

JST에서는 [연구 기반 DB 개발 사업]으로서 물리·재료분야 및 생명과학분야의 고기능 기반 DB를 개발하고 있다. 이 고기능 기반 DB는 연구 정보 네트워크의 보급을 배경으로 한 연구개발 활동의 고도 정보화에 대응하여, 다양한 과학기술 분야의 연구자를 대상으로 해서 그 특화된 연구 활동이 활성화되도록 정보나 기능을 제공함을 목적으로 하고 있다.

이러한 사업을 위해 JST가 도입하여 JST 동경 본부의 슈퍼컴퓨터실에 설치한 복합 슈퍼컴퓨터 시스템은, 계산 목적에 따른 5종류의 고속 계산 서버 및 기타 워크스테이션 등이 상호 고속 네트워크 연결되어 있으며, 서브간 효율적 연대를 가능케 하고 있



■ 그림 1 JST 복합슈퍼컴퓨터 시스템

다. 이 시스템은 고기능 기반 DB가 갖고 있는 다양한 기능의 개발에 필요한 고속 계산 처리 알고리즘의 연구·개발과 데이터 해석 등의 실제의 계산 처리 시스템의 개발에 위력을 발휘하여 기타 JST의 기초 연구·개발의 지원과 공동연구에도 이용되고 있다.

본 시스템은 1997년 8월 25일부터 본격적으로 가동되었다.

#### 1. 도입 경위

JST에서는 생명과학분야 및 물질·재료분야의 고기능 기반 DB를 개발하고 있다. 이 고기능 DB는 신규 데이터의 투산이나 여러 데이터를 보관하는 [데이터 창체기능], 연구자와 DB가 대화하면서 데이터 평가나 데이터 등록을 하는 [인터랙티브 기능], 또한 검색 및 출력시 고도의 편리성을 제공하는 [이용자 친화 기능] 등을 주축으로 하고 있다.

#### 2. 시스템 소개

본 시스템은 기본적으로 5종류의 고속계산서버, 과일사계, 메인프레임 라이브러리, 네트워크 장치 등으로 구성되어 있다.

가. 시스템의 특징

1) 기능·성능에 대응할 수 있는 다양한 계산기 환경  
고기능 DB를 개발하는 2분야에서는, 계산에 이용되는 알고리즘이 다르다. 모든 계산을 대형 계산기 1대로 하는 것이 아니라 각각의 계산에 맞는 최적 계산기를 도입했다.

2) 세계 최초의 **데이터센터 프로세서**

Supercomputing과 Visualization의 유기적 결합을 목표로 동경 본부의 지하·출처 복합 슈퍼컴퓨터 시스템과 직접 연결된 120인치 대화면 프로젝터를 설치했다.

역경 서버 방식의 인경을 활용함으로써 카시화(可視化)·서버로 작성된 CG나 인더터 상에 공개되고 있는 VRML 데이터를 일체 화상으로 볼 수 있다.

### 3) 대용량 기억장치

DB를 개발하기 위해서는 대량의 데이터 처리가 필요하다. 때문에 시스템 권제한 97GB, 디스크 23TB, 테이프 기록장치 6TB라고 하는 대용량 기억장치를 갖추었다.

### 나. 시스템의 개요

#### 1) 하드 계산 서버

액터화 및 액터화 병렬로 대규모 수치 계산을 한다.

#### 2) 분자 계산·가시화 서버

계산과 가시화를 한다. 기능시켜 결정구조의 변화 등 계산 과정을 그래픽으로 표시한다. 또한 그 결과를 대형 프록터로 투영(投影)시키거나 비디오로 녹화할 수 있다.

#### 3) 대규모 염기 서열 결정 서버

다른 생명과학 관계 기관과의 선대를 중시한 다양한 소프트웨어 이용이나 교류를 하기 위한 플랫폼으로 한다. 고속(高速) 염색이나 DNA로 합 등에서 이용된다.

#### 4) 서열 정보 해석 서버

신규 알고리즘의 검토나 연구개발을 위해 전자(電子)·CPU 성능이 높은 계산기이다. DNA 서열의 주석(注釋) 등에도 이용된다.

#### 5) 데이터 해석 서버

데이터 해석을 위한 병렬 처리시키는 방법의 연구개발 등에 이용된다.

네트워크부는 기능 분산화를 도모하기 위해 HIPPI와 FDDI의 2종류 네트워크를 중심으로 한 LAN으로 구성되어 있다.

### 3. 시스템의 활용 예

생명과학 분야에서 생각하고 있는 장면의 제마는 고속 동족·검색 유전자 구조 예측, 인간 DNA 서열 중의 시그널 패턴 추출이나 유전자 양의 추정·생체 고분자의 구조 해석 등이다.

기타 JST의 기초 연구·개발의 지원과 공동연구에의 이용이 기대되고 있다.

### ■ 주

- 1) 일본 JST의 홈페이지: <http://www.jst.go.jp>
- 2) 新藤洋行, 『情報管理』, JST, 제41권 제10호 1998. 1, pp.234-245
- 3) 高杉弘 編, 『情報管理』, JST, 제42권 제3호 1998. 5, pp.230-237
- 4) 今井建雄, 『情報管理』, JST, 제41권 제12호 1998. 12, p.201
- 5) 平川英典, 『情報管理』, JST, 제41권 제1호, 1998. 7, pp.254-268.