

단백질 구조예측작업 에이전트의 프로토타입 설계

남덕우⁰ 진 훈 김인철

경기대학교 전자계산학과

경기도 수원시 팔달구 이의동 산 94-6, 446-760

Tel: +82-031-243-9670, Fax: +82-031-249-9673, E-mail: ndw76@kyonggi.ac.kr

요약

15년에 걸쳐 수행된 게놈프로젝트의 완성으로 인류는 본격적으로 프로테옴 시대로 접어들게 되었다. 90년대 중반 이후 전세계적으로 다량의 단백질 구조정보 및 예측을 위한 방법들이 소개되고 있지만 각 자원들마다 저장, 관리 형식이 다를 뿐만 아니라 이용하는 방법도 어렵다. 또한 결과적으로 컴퓨터기술을 이용한 단백질의 구조예측작업을 제대로 지원하기 어렵다. 본 논문에서는 개방형 다중 에이전트 시스템을 통해 이를 해결하고자 했으며 특히 단백질 자원 데이터베이스를 효과적으로 이용하기 위한 에이전트 설계방법에 대하여 기술하고자 한다. 단백질 구조 예측 지원을 위해 효과적인 에이전트 내부 구조를 설계함으로써 점차로 요구되는 온톨로지 기술이나, 자동 예측 기능과 같은 다양한 요구사항들을 충족시킬 수 있을 것이다.

Keyword.:

Protein Structure Prediction ; Multi Agent System ; APSS

1. 개 3

2003년 4월 12일에 최종 완성된 게놈프로젝트로 인해 인류는 드디어 본격적인 프로테옴 시대로 들어서게 되었다. 이제까지의 연구가 인간의 염기서열을 밝히는데 그 목적이 있었다면 앞으로는 밝혀진 염기서열, 즉 DNA서열을 기반으로 단백질의 기능을 밝히는 연구가 될 것이다. 단백질의 기능을 제대로 예측하기 위해서는 단백질의 3차 구조를 알아야 할 필요가 있다. 단백질의 3차 구조란 DNA가 전사된 단백질 서열들이 3차원 구조를 갖는 형태를 이루고 있는 것으로서 여러 종류의 단백질들이 3차원 내에서 서로 얽히는 관계를 이룰 때 비로서 다양한 인체 내의 각종 반응 및 분비물 등이 나타나는 것이다. 1970년대부터 단백질의

구조를 밝히려는 시도가 실험실 내 다양한 물리, 화학적 실험연구를 통해 이루어졌다. 당시에는 수년에 걸쳐 하나의 단백질 구조를 결정하는 것조차도 빠듯했던 것이 시간이 흐르면서 다양한 방법들이 소개되어 1년에도 몇 개의 단백질 구조를 밝히는 단계까지 와 있다. 하지만 휴먼게놈프로젝트의 여파로 컴퓨터를 이용하여 DNA서열을 밝히는 연구가 시작되면서 하루에도 여러 개의 유전정보가 밝혀지는 요즘 시점에 와서는 실험실 내 연구만을 통해서 단백질의 구조를 밝히고 기능을 예측하는 것은 불가능한 상황이 되고 말았다. 그래서 1990년대 중반 이후부터는 컴퓨터를 이용하여 단백질의 구조와 기능을 밝히려는 수많은 연구들이 진행되어 왔다. 현재 이와 관련하여 전 세계적으로 약 330여 종류의 단백질 관련 정보를 제공하는 데이터베이스들이 존재하고 있으며 그에 상응하는 다수의 분석용 소프트웨어 도구들이 개발되어 있다[3].

하지만 지금까지의 연구결과에서 컴퓨터기술을 이용한 단백질의 구조예측 방법들은 표준화되어 있지 못하다[5]. 그러므로 불과 얼마 전까지도 단백질의 구조와 관련된 데이터베이스 자원들은 구조 예측을 위한 전 과정을 다루고 있기보다는 주로 국부적인 세부 정보들만을 제공하였다. 최근 들어서는 이렇게 존재하는 여러 자원들을 서로 연결하여 정보를 공유하고자 하는 시도들이 진행되었는데 결과적으로 기존의 단백질 데이터베이스들을 서비스하는 유명 사이트들이 하이퍼링크를 통해 복합적인 형태의 정보를 제공하는 형태로 변모하게 되었다. 그럼에도 불구하고 이와 같은 자원들을 그대로 이용하여서는 날이 갈수록 새롭게 발견되고 분석 되어져야 하는 단백질들의 구조를 제대로 예측하기 어렵다. 이는 각 데이터베이스 자원들이 서로간의 데이터의 연결이나 통합을 전제로 개발된 것들이 아니기에 데이터 양식 및 용어체계가 다르고 이용방법 면에서도 차이가 있으며 각 자원들이 시간이 갈수록 질적, 양적으로 계속 성장함으로 인해 연결 작업

또는 연동작업을 위한 추가적인 작업이 계속적으로 요구되기 때문이다. 이러한 문제를 해결하기 위해 90년대 말부터 에이전트기술을 이용한 단백질 구조예측작업이 대두되었으며 이러한 연구결과로서 게놈 분석을 위해 다수의 데이터베이스 자원들을 에이전트화한 GeneWeaver, 의학 정보 검색을 위해 개발된 MeLiSA, 생물과학 연구를 지원하는 목적의 BioAgent, 단백질 구조예측을 위한 에이전트들 간의 행위를 주로 다룬 X-MACoP, 에이전트시티 네트워크에 기반한 다중에이전트를 이용하여 단백질학 연구를 지원하기 위한 APSS[1] 등이 있다.

본 논문에서는 APSS시스템을 위한 세부단계로서 시스템을 구성하는 단위 에이전트들에 대하여 기술하고자 한다. 각 에이전트들은 인터넷 상에 존재하는 각 단백질 데이터베이스 자원들에 대해 현재까지는 래퍼(wrapper)로서 작동하며 추가적인 설계 및 구현을 통해 에이전트들 간의 통신을 통해 원하는 정보를 검색한 후 온톨로지를 이용하여 정보의 저장 및 이용효율을 높이며 최종적으로 단백질 예측작업까지도 수행할 수 있는 기능을 갖도록 하고자 한다. 다만 현재까지는 생물정보학 데이터베이스 자원들의 특성을 고려한 에이전트 설계에 초점이 맞추어져 있다. 다음 장에서부터 관련연구로서 에이전트 기술 및 단백질 관련 자원들에 대해 소개하겠다.

2. 관련연구

2.1 생물정보학을 위한 에이전트 기술

에이전트 및 다중 에이전트 시스템에 관한 연구는 에이전트 기반 소프트웨어 시스템 개발에 도움이 되는 직접적인 연구 성과를 거두는 단계이며 에이전트 기법은 대규모 분산 소프트웨어 시스템 개발을 위한 차세대 개발 방법으로서 주목받고 있다[2]. 이에 생물정보학 분야에도 여러 곳에서 에이전트 기술을 활용한 연구가 발견되는데 그 중의 하나가 GeneWeaver시스템이다.

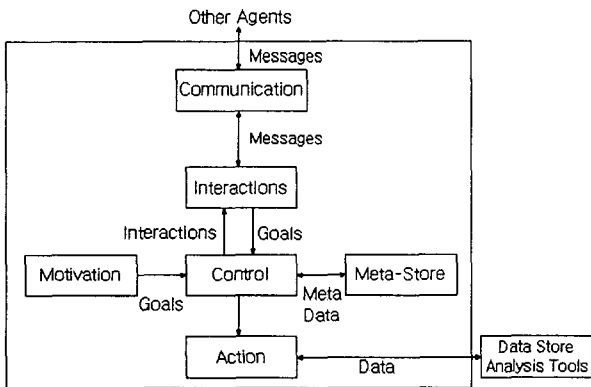


그림 1 GeneWeaver 에이전트 구조

이 시스템은 유전체 분석 목적 및 단백질 구조 예측을 위해 고안된 시스템이다. 특성을 보면 이질적인 생물학적 데이터베이스 자원들과 소프트웨어 도구들을 통합하여 발생하는 생물학 데이터를 분석하고자 하는 목적으로 개발되었다. GeneWeaver시스템에서 특이할 만한 점이 있는데 그것이 바로 분산되어 있고 동적으로 변화하는 환경에서 생물정보학 데이터를 분석하기 위해 다중 에이전트 개념을 도입하여 해결하고자 했다는 것이다. 다만 아쉬운 점은 시스템을 구성하는 모든 에이전트들이 에이전트로서의 기능에 치우쳐 공통된 구조를 가지도록 만들어졌다는 것과 래핑 작업을 수행하는 에이전트들 모두가 직접 저장장치나 분석 도구들과 통신함으로 인해 때로는 보다 나은 정보를 찾기 위해 에이전트끼리 빈번히 통신하며 발생하는 데이터들을 효율적으로 저장관리하기 어렵다는 것이다.

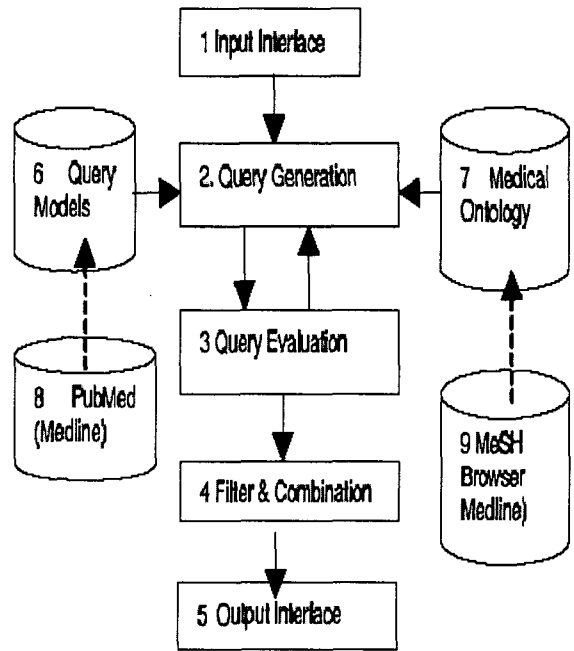


그림 2 BioAgent 에이전트 구조

MeLiSA시스템은 의학분야에서 정보검색을 위해 온톨로지 기반의 에이전트기술을 이용한 시스템이다. GeneWeaver와 달리 단일 에이전트 시스템으로서 [그림 2]와 같은 구조를 갖고 MeSH 단어목록에 기반하여 Medline데이터베이스를 검색하는 단순 검색용 에이전트이다. 특이할 점은 질의 생성을 위해 미리 MeSH단어목록을 기초로 한 의학 온톨로지 데이터베이스를 이용한다는 것이다.

이 외에도 유전체 데이터 분석연구 작업을 수행하는 생물학자들을 위한 이동 에이전트 시스템인 BioAgent, 현재 개발중인 시스템으로서 예측성능을 높이기 위해 에이전트들을 팀 단위의 협업작업을 지원하는 X-MACoP 시스템들이 있으나 에이전트의

내부 구조는 공개되어 있지 않다.

2.2 인터넷 상의 단백질 자원들

현재 인터넷 상에서 생물학 관련 자원들은 그 수와 양적 측면에서 급격하게 증가하고 있다. 특히나 게놈프로젝트의 말기부터는 단백질 관련 자원들이 놀라운 속도로 늘어나고 있는 추세이다. 그럼에도 불구하고 여전히 단백질 구조예측과 관련하여 많은 전문가들로부터 인정 받고 컴퓨터기술을 이용한 예측과정에서도 널리 쓰이는 자원들이 있는데, 바로 PDB, GenPept, 3D-SSSM, SwissProt 등이다. 또한 최근 들어서 급격하게 많이 연구되는 분야가 문헌정보를 이용한 단백질 기능예측인데 이와 관련하여 Medline 데이터베이스도 구조예측과 관련하여 중요하게 인정 받는 자원이다. 차례대로 소개하면 다음과 같다.

2.2.1 PDB

PDB(Protein Data Bank)는 미국 부록해본 연구소(Brookhaven National Laboratory)에서 설립하여 운영되고 있는 단백질의 3차구조 데이터베이스로서 현재 20800개의 구조정보를 저장하고 있다. PDB는 키워드 또는 PDB ID를 질의어로 사용하는 기본 검색방식과 단백질 서열 또는 부가정보를 질의어로 사용할 수 있는 맞춤검색 방식을 제공한다. 검색과정을 통해 제공되는 정보는 크게 나누어서 해당 단백질의 요약 정보, 구조파일 보기, 구조적 이웃 정보, 기하 정보, 관련 링크 정보, 구조 파일 다운로드 정보 등을 제공한다. 요약정보, 구조적 이웃 정보, 기하 정보, 링크 정보의 경우에는 대부분 텍스트로 이루어져 있고, 구조파일 보기는 다운로드 후 Add-on형태의 소프트웨어를 통해 작동가능하며, 다운로드의 경우는 pdb 파일 원본, 또는 압축파일 형태로서 http 또는 ftp를 이용하여 수행이 가능하다.

2.2.2 GenPept

GenPept는 NCBI 사이트 내에 존재하는 단백질 서열 데이터베이스로서 entrez를 통한 키워드 검색과 blast를 이용한 상동성 서열 검색기능을 수행할 수 있다. 최근 들어서는 자체적으로 MMDB라는 단백질 구조 데이터베이스를 따로 구축하여 GenPept 검색을 통해서도 구조정보를 얻을 수 있다. 하지만 PDB나 뒤에 설명할 SwissProt 데이터베이스에서 제공하는 정보에 비해서는 여러 가지 면에서 제한된 정보만을 제공한다. 이는 단백질 서열에 기반하여 blast검색 시에 점수와 E값을 기준으로 정렬되어 있음에도 불구하고 해당 지식이 없는 경우에 상동성을 갖는 단백질 정보를 구하기 어렵다는 단점을 가진다. 그럼에도 불구하고 GenPept는 단백질 구조예측을 위해 종종 이용되는데 이는 막대한 유전체정보를

바탕으로 이와 관련된 단백질 서열을 유도하는 핵산수준의 다양한 소스정보를 제공할 수 있기 때문이다.

2.2.3 3D-PSSM

3D-PSSM 서버는 PDB나 GenPept와는 달리 스테딩을 이용하여 예측기능을 수행하는 서버 소프트웨어이다. 하지만 사용자의 단백질 서열 질의만을 사용하여 밝혀진 상동 서열을 찾은 후 해당 서열이 갖는 3차 구조 정보와 가능한 기능을 예측하여 사용자에게 제공하는 기능을 가짐으로써 단백질 데이터베이스와 유사한 기능을 갖는 것으로 판단할 수 있다. 사용자는 질의 시에 예측된 결과를 수신할 수 있는 전자메일 주소와 질의 서열의 제목, 그리고 FASTA형식의 단백질 서열을 함께 입력한다. 수신된 결과는 얼라인먼트(alignment) 결과, 폴딩 라이브러리, 예측된 구조 정보, PSSM의 E값, SAWTED E값, biotext 정보 및 E값, SCOP 데이터베이스 검색을 통한 클래스, 폴딩 구조 정보, Family정보 등으로 구성된다. 예측된 구조 정보를 제외하고는 모두 텍스트 기반 하에 읽기가 가능한 정보들이다. 3D-PSSM의 경우 전자메일을 통해서 예측결과를 수신하거나 질의 시에 바로 리턴 되는 주소를 이용해서 예측결과를 얻는 특징을 갖는다.

2.2.4 SwissProt

SwissProt은 가장 잘 정의된 주석정보를 포함하고 구별되는 단백질 구조정보만을 저장한 데이터베이스로 인정 받고 있다. 유럽연합에서 운영하는 EBI 기관과 연관되어 있으므로 EBI에서 제공하는 다양한 생물정보학 자원들과 교차 참조 될 수 있는 장점을 제공함으로써 3차 구조 예측을 위해 PDB와 더불어 반드시 이용해야 하는 데이터베이스 자원이다. SRS 검색엔진을 통해 검색할 수 있으며 검색결과로서 풍부한 주석정보를 포함하여 해당 되는 단백질 서열의 도메인정보, Family정보, Motif정보, 2차 구조 정보, 3차 구조 정보 등을 제공한다.

2.2.5 Medline

Medline은 많은 출판물과, 생물, 생화학, 유전학 등의 학술 저널 등을 포함한다. Medline 사이트 접근은 유료 사용자에게만 허용되지만 NCBI내 PubMed 데이터베이스를 통한 접근통로는 열려 있어 이를 통해 이미 출판된 수많은 저널 들의 요약 정보를 이용할 수 있다. 또한 PubMed 데이터베이스는 GenBank, GenPept, MMDB 등 다양한 데이터베이스 결과들과 서로 링크되어 있어 해당 유전자 검색을 통해서도 관련 저널 등을 찾을 수 있는 기능을 제공한다.

3. 시스템 설계

본 논문에서는 APSS 시스템 구성을 위한 단위 에이전트를 설계하고자 하였다. APSS (Agent-based Proteomics Support System)는 다양한 단백질과 관련된 연구를 지원하기 위한 시스템으로서 개방형 네트워크를 지원하며 다중 에이전트를 기반으로 개발 중인 시스템이다[1].

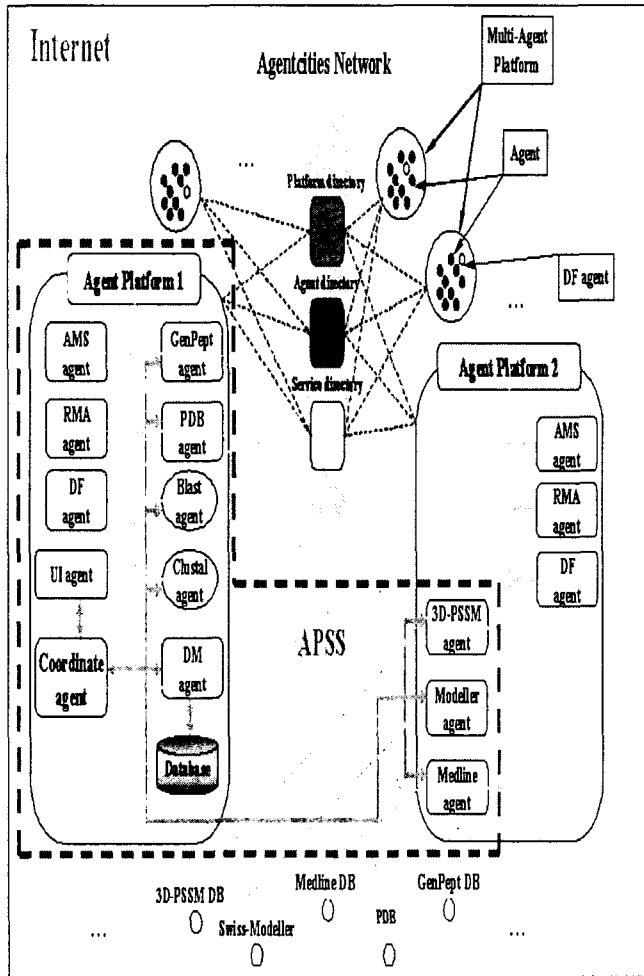


그림 3 전체 시스템 구조도

전체 시스템 구성을 위해 역할 및 수행 기능별로 에이전트들을 분류해 보면 크게 단일 자원 에이전트, 분석 에이전트, 관리 에이전트, 사용자 에이전트 그룹으로 나눌 수 있으며 나열하면 다음과 같다.

- 단일 자원 에이전트들 : PDB, GenPept, Medline, 3D-PSSM, SwissProt
- 분석 에이전트들 : BLAST, CLUSTALW, Modeller
- 관리 에이전트들 : 조정(Coordinate), DM
- 제어 에이전트들 : AMS, DF

● 사용자 에이전트 : UI

현재 개발 중에 있는 시스템은 우선적인 과제로서 단백질 구조예측을 지원하는 기능을 갖는 것으로서, 이를 위해서는 컴퓨터 기술을 이용하여 단백질 구조를 예측하는 프로세스를 지원하기 위해 위에서 나열한 에이전트들에 대한 설계 및 구현작업이 필수적이다.

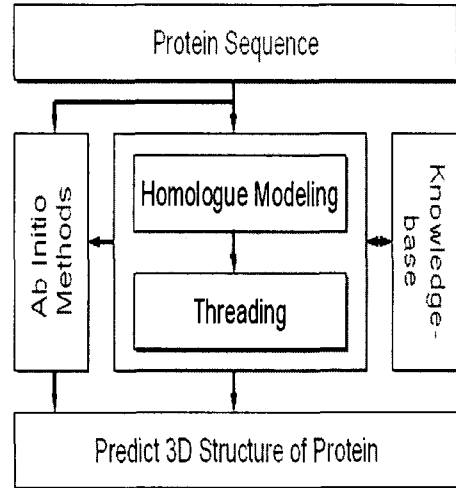


그림 4 컴퓨터 기술을 이용한 단백질 구조예측 방법

서론에서도 언급하였듯이 단백질의 3차 구조를 찾는 과정이 특별하게 정해져 있지 않은 상태에서 1차 구조만 밝혀진 상태에서도 3차원 구조가, 2차원 구조가 밝혀진 상태에서도 3차 구조가, 심지어는 1차 서열만 가지고서도 3차원 구조를 예측할 수 있는 다양성이 존재하기 때문에 이를 고려하여 세부적인 절차를 비교, 분석, 판단하고 최종적으로 예측과정을 설계해야 할 필요가 있다. 우선적으로 우리는 대표적인 단백질 자원 데이터베이스를 에이전트화하여 사용자의 수고를 덜 수 있도록 본 시스템을 사용함으로써 한 곳에서 쉽게 단백질 관련 정보를 찾고 통합할 수 있는 시스템을 고안하고자 한다. 이를 위해서는 단백질 서열에 맞는 1차 구조, 2차 구조 및 폴딩을 이루는 분자수준의 다양한 정보 등을 찾는 작업이 함께 병행되기도 하며, 때로는 분석용 소프트웨어 등을 이용해서 계산하는 작업도 필요하다. 이러한 방법은 새로운 알고리즘을 개발하는 과정이 아니며 이미 존재하는 자원들을 효율적으로 연결 및 연동시킴으로써 결과적으로 단백질의 구조예측 작업을 돕는 소프트웨어 도구를 개발하는 것이며 사용자에게 편리성을 제공하면서도 신뢰성 있는 결과를 제공할 목적을 갖는다. [그림 4]는 컴퓨터 기술을 이용하여 효과적으로 분석함으로써 단백질의 구조를 찾는 방법들을 도식화하여 나타낸 것이다.

3.1 시스템 구조

[그림 3]과 같이 에이전트 네트워크에 접속되는 플랫폼들은 최소한 AMS, RMA, DF 에이전트를 포함하며 UI 에이전트를 통해서 사용자는 예측 과정과 관련된 모든 작업을 수행할 수 있다. UI 에이전트를 통해 질의된 모든 작업은 조정 역할을 수행하는 조정에이전트가 담당한다. 우측의 플랫폼2는 연결된 다른 플랫폼을 예로 든 것이다. 자세히 보면 플랫폼2에는 2차 구조 예측 서버 및 3차원 모델 예측 서버를 위한 에이전트가 존재한다. 그리고 이들 에이전트들과 플랫폼1이 점선으로 연결된 것처럼 하나의 시스템을 구성하며 이는 전세계에 존재하는 수많은 생물정보 자원들을 대신하는 에이전트는 하나의 플랫폼 내에만 존재해야 할 필요가 없으며, 에이전트시티 네트워크 서비스에 등록만 되어있으면 다른 시스템에서 이미 개발되었거나, 앞으로 개발될 에이전트들까지 쉽게 연결하여 하나의 시스템으로 포함될 수 있음을 의미한다. 자원 에이전트는 실질적으로 단백질 구조예측을 위해 이용해야 할 단백질 구조 데이터베이스에 접속하여 질의를 하고 정보를 찾는 에이전트로서 일반 사용자들이 해당 데이터베이스에 웹 브라우저를 이용, 접속하여 사용하는 과정을 대신한다 분석 에이전트는 자원 에이전트로부터 얻어진 정보들을 계산하고 분석하는 과정을 수행한다 관리 에이전트는 에이전트 간 동작을 제어하고 조정하며 필요에 따라 저장기능까지도 수행하는 기능을 하는 에이전트이다. 관리 에이전트를 두는 이유는 시스템 내의 에이전트 각각의 특성에 맞는 상호 작용을 통해 사용자의 단백질 구조예측을 돕는 최종 목표를 수행하기 위해서이다. 제어 에이전트는 다중 에이전트 플랫폼 상에서 전체 에이전트들을 제어하고 감시하는 기능을 갖는 에이전트들이다. 그러나 기본적인 에이전트 속성이 단위 작업에 대해서 독립성을 추구하기 때문에 작업을 일관성 있게 감독하기 위한 관리 에이전트의 동작만으로 전체 작업에 대한 효과를 얻기란 힘들 것이다. 이와 함께 단위 에이전트간의 메시지 전달을 효율적으로 하는 일과 에이전트의 기능을 견고하게 하는 일, 본 논문에서는 기술하지 않지만 메시지 자체의 개념에 대한 장의를 통한 공유 개념을 만드는 일도 고려해야 하는 사항이다. 다음 절에서는 분류된 그룹에서 대표적인 몇 개의 단위 에이전트들의 수행 구조에 대하여 소개하였다.

3.2 에이전트 구조

프로토타입 수준의 구현을 위해 현재 개발중인 에이전트로는 PDB, GenPept, 3D-PSSM, Medline, Modeller 조정, DM, UI 에이전트들이 있다. 기능적인 면에서 아직까지 단순 질의만을 처리하는 수준인

것을 감안할 때 PDB 에이전트 구조가 단일 자원 에이전트들의 기능을, Modeller 에이전트가 분석 에이전트들, 그리고 조정 에이전트가 관리 에이전트들을 대표한다고 할 수 있다. 다음 절에서 이 3개의 에이전트의 수행구조에 대하여 소개하고 3.3절에서 단위 에이전트들의 상호작용에 대하여 설명하였다.

3.2.1 PDB 에이전트

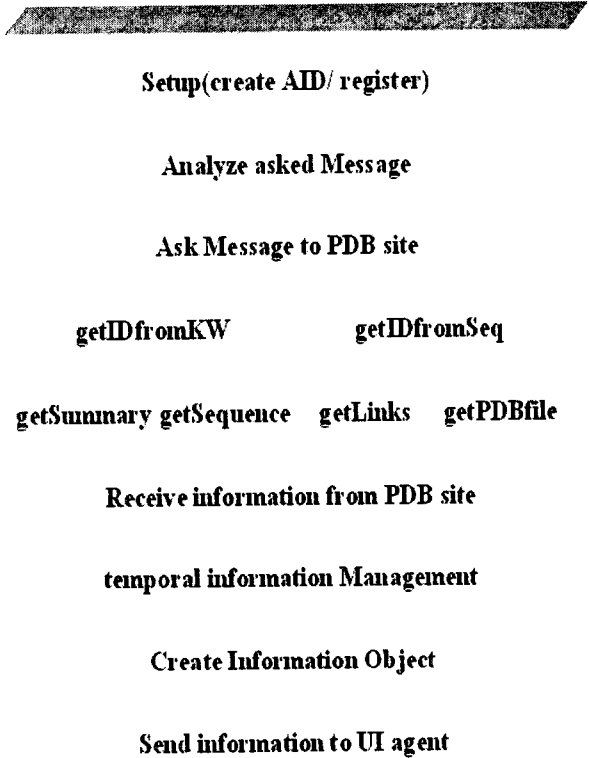


그림 5 PDB 에이전트의 수행 구조

먼저 모든 에이전트들은 기동과 동시에 에이전트 구별을 위해 에이전트 ID 생성 및 AMS, DF 에이전트에게 자신의 존재를 알리며 등록하는 셋업 과정을 수행한다. 그 후 UI 에이전트로부터 수신한 메시지를 분석하게 된다. 일반적으로 PDB 에이전트에게 전달되는 메시지는 사용자의 질의이며 내용으로는 일반적인 키워드, 또는 해당 PDB ID, 그리고 미지의 서열을 통한 질의로 나눌 수 있다. 이를 고려하여 현재까지 개발된 사항을 보면 ID와 서열로 질의할 수 있게 하였으며 검색대상 정보로는 해당 ID를 갖는 단백질의 요약정보, 단백질 서열 정보, 관련 링크 정보, 3차 구조 파일을 구할 수 있도록 하였다. PDB사이트로부터 수신된 모든 정보들은 객체화 되어 조정 에이전트를 거쳐 최종적으로 UI 에이전트에게 전달 되도록 하였다.

3.2.2 Modeller 에이전트

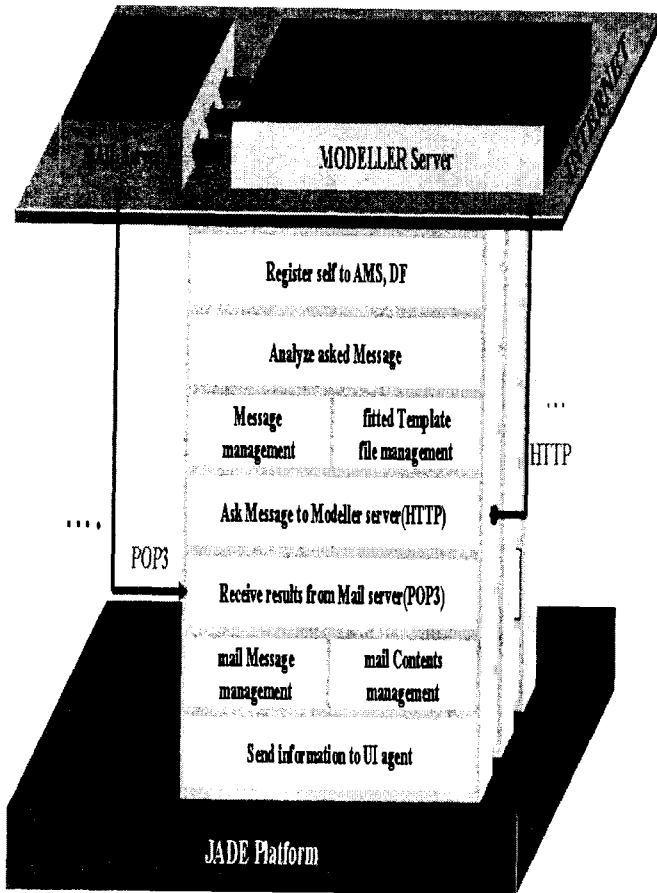


그림 6 MODELLER 에이전트 수행 구조

Modeller 에이전트 역시 셋업 과정을 거친 후 수신된 메시지를 분석하는데 이때 미리 Swiss-PDBViewer 소프트웨어를 통해 작성된 3차 구조를 갖는 주형정보 파일과 타이틀이 담긴 메시지를 수신하게 되므로 이의 관리가 필요하여 그 후 Modeller 서버에게 http 프로토콜을 이용하여 보내게 된다. Modeller 서버를 통해 계산되어진 최종 결과물인 예측된 3차 구조 파일은 pop3 프로토콜을 이용하여 수신하게 되며 역시 메시지와 파일을 관리해야 한다. 아무리 Modeller 서버를 통해 예측된 3차 구조 파일일지라도 사용자의 조정작업을 거쳐서 올바른 것인지 확인작업을 거치게 되는데 그렇지 않은 경우 예측과정을 살펴서 문제가 있다고 판단되는 과정에서부터 재차 예측과정을 따르게 된다. 그러므로 미지의 단백질 하나에 대하여 계속적으로 여러 개의 질의가 Modeller 에이전트를 통해 Modeller 서버에게 전달될 수 있고, 이를 위해서는 반드시 송신 전, 수신 후의 메시지들에 대한 관리작업이 필수적이다. 또한 조정 에이전트를 통해 동시에 여러 개의 질의를 수행할 필요가 발생할 것인데, 현재의 Modeller 에이전트는 자체 내의 다중화 작업을 지원하지 않고 대신 JADE 플랫폼 내에서 다중 인스턴스화 함으로써 작업을

수행할 수 있게 된다.

3.2.3 조정 에이전트

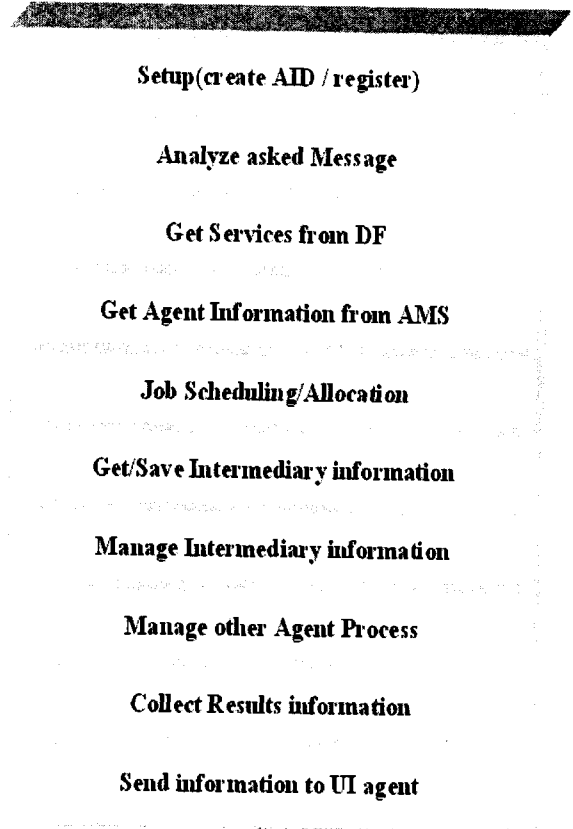


그림 7 조정 에이전트의 수행 구조

조정 에이전트는 본 시스템에서 가장 중요한 역할을 수행하는 에이전트로서 UI 에이전트로부터 받은 모든 질의에 대해 수행업무를 총괄하는 책임을 진다. 역시 셋업 과정을 수행 후 수신된 메시지를 분석한 후 수행 가능한 서비스들이 존재하는 지를 DF 에이전트를 통해 확인한 후 AMS 에이전트에게 해당 서비스를 수행하는 에이전트 정보를 구하게 된다. 메시지에 부응하는 서비스를 갖는 에이전트들이 구해지면 수행해야 할 작업들을 미리 정의된 내용에 따라 할당하고 작업수행을 하도록 한다. 실제 예측과정을 수행하는 에이전트들과 상호작용을 통해 발생하는 다양한 정보들은 DM에이전트를 통해 저장하는데 아직까지는 로컬 데이터베이스 구축이 안된 관계로 파일로만 저장하게 된다. 향후 온톨로지를 고려하여 다양한 자원들로부터 수신되는 정보들을 효과적으로 관리할 수 있도록 데이터베이스 스키마 형태를 갖는 DM 에이전트를 구현할 예정이다. 이러한 과정을 거쳐 얻어진 정보들은 현 단계에서는 모두 UI 에이전트에게 전달하여 보여지도록 하고 추가적인 과정을 거쳐서 선별적으로 UI 에이전트에게 전달되도록 할 예정이다.

3.3 에이전트 상호작용

위에서 기술한 단위 에이전트들이 시스템 내에서 어떻게 동작하는지 간략화 하여 기술하면 다음과 같다. 먼저 GenPept 에이전트는 NCBI의 GenPept 데이터베이스에 접속하여 질의서열과 blasting한 결과를 찾은 후 임계값에 따라 유전자의 Annotation 정보 또는 1차 서열을 구하는 작업을 수행한다. 이를 위해 Blast 에이전트가 이용되며 다중서열 정렬을 위해 Clustalw 에이전트가 이용된다. Medline 에이전트는 유전자와 관련된 문헌정보를 사용자에게 제공하여 구조예측을 돕는 작업을 수행한다. Modeller는 SWISS-Modeller 서버를 대신하는 에이전트이다. 미지의 서열 및 이미 얻어진 1차 서열들과 유사한 모형(template) 구조를 찾고 서열과 모형구조와의 맞춤작업(fitting) 후 이를 서버로 전달한다. 최종결과를 수신하면 사용자에게 전달한다. Coordinate 에이전트는 UI 에이전트로부터 질의된 내용을 처리하여 주 데이터베이스 에이전트들에 차례로 전달하고 필요에 따라 소프트웨어 에이전트를 호출하여 작업을 지시하기도 한다. 처리된 결과들은 다시 Coordinate 에이전트에게 전달되며 적절한 처리과정을 거쳐 UI 에이전트에게 전달된다. 때로는 복수 작업 또는 대량의 데이터 처리를 위해 DM 에이전트를 통해 로컬 데이터베이스에 저장하는 작업도 담당한다. DM 에이전트는 이외의 모든 에이전트들에서 수행 전후 과정을 통해 발생하는 데이터를 저장하고 관리하는 작업을 담당하며 이를 위해 데이터베이스를 내재한다. PDB 에이전트는 APSS를 이루는 주된 역할을 수행하는 에이전트이다. PDB 에이전트는 현재까지 밝혀진 모든 단백질 구조정보를 포함하는 PDB 데이터베이스 (<http://www.rcsb.org/pdb/>) 에 접속하여 질의 및 응답 과정을 수행한다. PDB 데이터베이스는 기본적으로 PDB ID 와 키워드 및 서열검색을 지원한다. 질의를 통한 응답결과로는 요약 정보, 구조 파일 보기, 구조 파일 받기, 유사 구조 파일 정보, 및 1차 서열 파일 등이다. PDB 에이전트는 UI 에이전트를 이용하여 사용자가 직접 PDB 데이터베이스 정보를 사용할 수 있도록 하기 위해 질의 종류와 요구 정보, 2가지로 구분하였다. 이를 통하여 사용자는 필요에 따라 원하는 정보를 얻기 위해 질의 형태를 다른 형태로 바꾸어 PDB 에이전트가 해야 하는 일들을 보다 세밀하게 명시할 수 있다. 각 에이전트들의 이러한 모든 행동 과정들은 에이전트간 메시지 전달을 통한 정보를 얻기 위한 목적을 달성하는 과정이 된다. 시스템에서 에이전트는 특징적인 역할을 가지고 존재하여 에이전트의 고유함을 지니고 있다. 역할을 수행 할 때는 에이전트로 하여금 정보를 얼마나 효율적으로 다루게 할 수 있는지가 단위 에이전트에 대한 기능을 견고하게 하여 결과적으로는 에이전트간 공동 수행능력을 향상시키게 된다.

4. 결론

지금까지 본 논문에서는 생물학 도메인에 사용되었던 에이전트들의 구조를 살펴보고 단백질 구조예측작업에 적절하게 이용될 수 있는 단위 에이전트 들을 설계하였다. APSS 시스템은 단백질학 연구지원을 위한 종합 시스템으로서의 목표를 갖고 개발중인 시스템으로서 시스템을 구성하는 단위 에이전트들은 아직까지는 에이전트 본래의 기능인 자율성 및 이기성 등과는 거리가 멀게 느껴질 정도로 단순한 기능을 내포한다. 하지만 기존 연구를 통해 밝혀진 에이전트들의 구조를 보면 에이전트로서의 기능 구현기능에 치우쳐 도메인의 특성을 제대로 반영하지 못한 채 구현된 것들이 있다. 이에 본 시스템에서는 현 단계에서 단백질 구조예측을 위한 데이터베이스 자원들의 특성을 고려하여 에이전트를 설계함으로써 독립성과 확장성 면에서 장점이 있다. 차후 연구 계획으로 에이전트 설계 상세화 및 협업을 통한 자동 예측 기능 추가, 그리고 온톨로지를 이용한 질의 확장 및 효율적인 정보 관리를 지원하는 기능을 제공하는 것이다.

참고문헌

- [1] 김현식 외. (2003). "단백질 구조예측을 위한 에이전트 시티 네트워크 기반의 다중 에이전트 시스템", 춘계정보과학회 Proceeding. Vol. 30, No. 1, pp.461-463
- [2] 한국정보과학회. (2000). "정보과학회지". Vol18, No5, ISSN 1229-6821.
- [3] Andreas D. Baxevanis*. (2002). "The Molecular Biology Database Collection: 2002 update", Nucleic Acids Research, Vol.30 No.1.
- [4] Bryson, K. and Luck, M. and Joy, M. and Jones, D. (2001). "Agent Interaction for Bioinformatics Data Management", Applied Artificial Intelligence, Vol.15, No.10, pp.917-947.
- [5] Cynthia Gibas, and Per Jambeck. (2001). *Developing Bioinformatics Computer Skills*, O'Reilly.
- [6] Rost, B. (1998). "Protein structure prediction in 1D, 2D, and 3D", *The Encyclopaedia of Computational Chemistry* (eds. PvR Schleyer, NL Allinger, T Clark, J Gasteiger, PA Kollman HF Schaefer III and PR Schreiner), No.3, pp.2242-2255.
- [7] Sycara and Katia. (1998). "Multiagent Systems", *AI Magazine* Vol.19, No.2, pp79-92.
- [8] Willmott, S.N. and Dale, J., Burg, B. and Charlton, C. and O'brien, P. (2001). "Agentcities: A Worldwide Open Agent Network", *Agentlink News*, Vol.8, pp.13-15.