



## 신약개발을 위한 바이오정보기술

아이디알 김승목

### 1. 서론

2000년 이후 많이 논의되고 있는 Post Genome 시대의 큰 특징은 여러가지가 있을 수 있지만 가장 눈에 띄며 가장 실용적인 측면에서 보자면 질병 치료의 가능성이 커졌다는 점을 들 수 있다. 흔히 Genome Project의 목적에 대해 단순히 생명체의 신비를 밝힌다든가, 복제인간의 탄생을 가능케 한다든가 하는 조금은 추상적이고 공상 과학적인 측면들이 강조되어 왔으나, 이는 아마도 생명 과학을, 또는 Genome Project를 일반인들에게 홍보하고 정책적인 지원을 얻어내기 위한 방편이 아닌가 생각된다. 실제로 미국이 1990년대 초반 Genome Project를 추진할 때, 가장 중요한 목표로 여겼던 점은 다양한 신약타겟의 발굴과 이를 통한 신속한 신약개발과 다양한 생명체의 Genome 해석을 통해 대체 자원(주로 에너지 자원)의 개발이었다. 인간에 대한 Genome 연구의 초기 단계가 완료된 지금, 실제로 많은 수의 질병타겟이 새로이 발굴되었으나 이것이 오히려 신약개발에 대한 개발에 대한 압박으로 작용하고 있다. 즉 통계에 의하면 지난 10년 간 발굴된 질병타겟에 대해 단 25%만이 유용한 선도물질이 발굴되어 있다고 한다[1]. 이러한 원인은 21세기 생명 과학 기술 또는 신약개발 기술이 아직은 전통적인 방법을 답습하고 있는 반면 Genome Project의 연구 결과는 너무나 방대하고 복잡하여 전통적인 연구 방법으로는 Genome 연구 결과를 따라갈 수 없다는 말로 설명할 수 있다. 결국 Genome Project의 전개와 예상 밖의 많은 결과들은 관련 분야에 패러다임의 변화(Paradigm Shift)를 유도하게 되었고 신약개발 방법도 빠른 변화를 보이고 있는 실정이다.

### 2. 신약개발 방법

#### 2.1 신약개발의 과정

신약개발은 일반적으로 질병타겟(Disease Target) 발굴, 신약선도물질(Hit Compound) 발굴, 신약후보(Lead Compound)물질 발굴과 최적화를 거친 물질을 전임상과 3단계의 임상 시험을 거친 후 상품화까지 이르게 되어 있다. 일반적으로 신약후보물질 발굴 단계까지를 신약발굴 단계(Drug Discovery)라 하고 전임상부터 임상 3상까지를 신약개발 단계(Drug Development)라 칭하며 한 개의 신약이 발굴부터 시판까지는 보통 12~15년이 소요되며 3억에서 6억 달러의 비용이 소요된다고 알려져 있다 [그림 1]. 더구나 2000년 통계에 의하면 시판되는 약 10개 중에 단 3개만이 개발비를 회수할 수 있다고 한다.

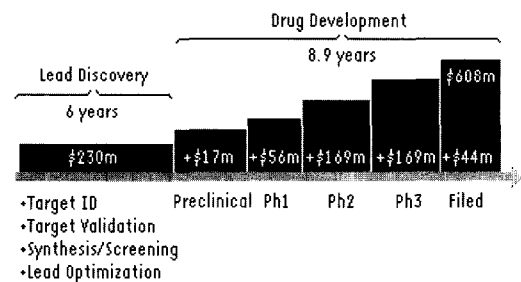


그림 1 신약개발 과정 및 비용

이처럼 신약개발은 성공에 이르기까지 막대한 연구 개발비의 투자를 감수해야 하는 과정을 거쳐야 하므로 전세계적으로 신약개발에 도전할 수 있는 제약

회사는 극히 제한되어 있다고 볼 수 있으며, 처음부터 독자적으로 신약을 개발할 경우 많은 시행착오와 출시 지연, 낮은 성공률에 따른 위험부담을 가지게 된다. 그러나 Genome Project에 의한 생명 과학의 발전은 신약개발 전략에 있어 많은 변화를 주고 있다. 가장 큰 특징은 전통적인 방법에 비해 많은 선도 물질을 최단 시간 내에 찾아내고 이를 유용한 후보물질로 발전시키는 시간을 단축시켜 가고 있다. 그림 1에서 알 수 있는 것처럼 신약개발은 동물과 사람에 대한 실제적인 약효실험(in vivo experiment)을 바탕으로 하고 있기 때문에 아직은 시간과 비용을 절약할 수 있는 실제적인 방법이 개발되어 있지 않으나 신약 후보물질 발굴은 비 생체실험(in vitro experiment)과 화학합성에 기반을 두고 있기 때문에 시간과 비용을 절약할 수 있는 실질적인 방법이 충분히 개발되어 있는 상태이다. 지금까지 전통적으로 사용해 오던 방법은 질병타겟이 발굴되고 이에 대한 스크리닝 시스템이 갖추어지면 수만에서 수십만 개의 화합물의 약효를 자동화된 실험장치에 의해 수일만에 검증하는 것이며 이를 High Throughput Screening(HTS) 이라고 한다. 물론 이런 방법을 통해 많은 신약선도물질이 발굴되었고 일부는 신약으로까지 개발된 것들이 있으나 이 HTS의 최대의 약점은 약효검증에 사용되는 화합물들의 다양성을 확보하기가 어렵다는 점이다[2]. 즉 약으로 사용될 수 있는 화합물이란 화학적인 방법으로 합성 가능한 것들이어야 하며 초기 선도물질탐색에 이용되는 화합물은 가능한 한 구조적, 화학적인 유사성이 적은 화합물이어야 하는데, 이러한 화합물을 수십, 수백만 개를 확보하는 것은 실제로는 불가능에 가까운 일이기 때문이다. 이러한 이유로 Genome Project에 따른 신규 질병타겟의 발굴은 쉬워진 반면 신약발굴 비율은 오히려 감소하고 있다고 볼 수 있다.

## 2.2 신약개발 방법의 변화

신약후보물질 발굴을 가속화 하기 위해 HTS를 대체하기 위한 방법으로 연구되기 시작한 것이 컴퓨터를 이용해 비 실험적인 방법으로 HTS를 수행하는 virtual HTS이다. 이 방법은 1970년대 부터 연구되기 시작한 전산화학(Computational Chemistry)의 발전에 힘입어 1990년대 말부터 분자설계 및 단백질 구조예측 분야가 획기적인 발전을 하면서 실용화 되게 되어 in silico structure-based design 분야가 정착되

게 되었다. 신약탐색의 전제조건은 질병타겟이 되는 단백질의 기능이 밝혀진 후 이 단백질의 구조가 3차원적으로 규명되어야 하나, 아직은 질병타겟 단백질의 기능을 밝히는 속도에 비해 구조를 규명하는 속도는 현저히 떨어지는 상황이다. 단백질의 3차원 구조를 실험적으로 규명하는 방법으로는 X-ray crystallography를 이용하는 방법과 NMR spectroscopy를 이용하는 방법이 있으나 그림 2에서 보는 바와 같이 생물학적인 방법에 의해 대량의 단백질을 얻어서 실험을 해야 하는 관계로 상당한 비용과 어려움이 따르고 있다. 그림 3에서 보면 단백질의 3차원 구조가 규명되는 숫자는 증가추세에 있으나 지금까지 알려지지 않았던 새로운 구조를 갖는 단백질의 구조 규명 속도는 오히려 감소하는 추세에 있다. 따라서 아무리 Genome Project에 의해 신규 질병타겟이 발견된다 해도 구조에 대한 정보가 얻어지지 않는다면 신약개발은 과거에 전통적인 시행착오에 의한 방법을 따를 수 밖에는 없게 되는 것이다. 이를 극복하기 위한 방법이 바로 in silico structure-based design과 virtual ligand screening인 것이다[3].

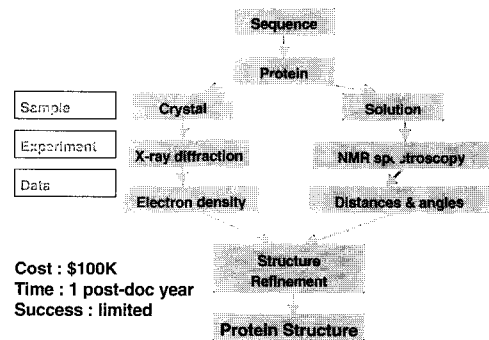
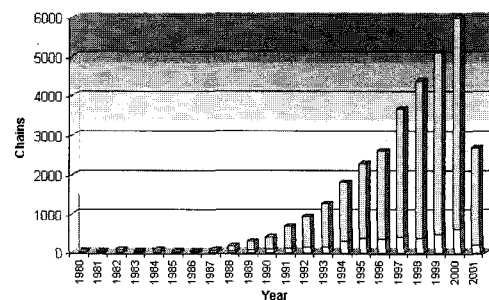
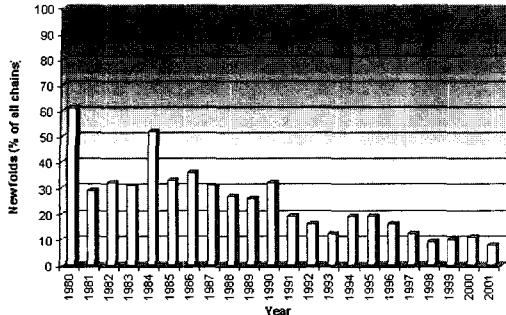


그림 2 단백질 3차 구조 규명을 위한 실험적인 방법



(a) 전체 단백질의 3차 구조 규명



(b) 신규 구조의 연도별 규명 정도

그림 3 단백질 3차 구조 규명 추세

### 2.2.1 in silico structure-based design

이 방법은 지금까지 알려진 단백질 구조 정보로 단백질의 구조를 컴퓨터를 이용한 계산에 의해 예측하는 것으로 과거에는 단백질의 전체구조를 예측하는데 주력하였으나 최근의 추세는 단백질의 약효작용점이라고 예상되는 구역의 구조만을 예측하는데 주력하고 있다. 일반적으로 이 약효작용점을 Drugable binding pocket이라고 부르며 한 개의 단백질에는 이미 잘 알려진 binding pocket이 있는가 하면 전혀 알려지지 않은 binding pocket이 존재하기도 한다. 따라서 컴퓨터를 이용한 예측기술을 이용한다면 알려지지 않은 binding pocket도 쉽게 찾아낼 수 있다[그림 4].

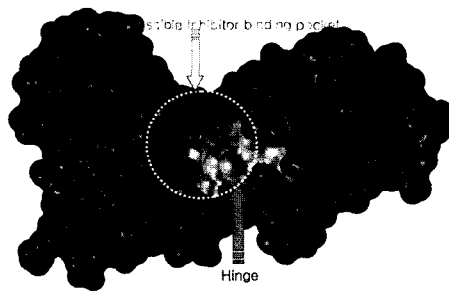


그림 4 컴퓨터로 예측된 골다공증 관련 단백질의 3차원 구조와 예측된 drug binding pocket

### 2.2.2 Virtual ligand screening

이 방법은 질병타겟 단백질의 구조가 알려진 경우 (실험적인 방법이나 컴퓨터를 이용한 예측을 통해) drug binding pocket에 컴퓨터를 이용해 가상으로 약

효물질로 예상되는 화합물과 단백질을 결합시켜 보는 가상실험을 말한다. 보통 생물학에서 스크리닝이라 함은 여러 물질 가운데 약효가 있는 물질을 선별하기 위한 실험을 말하는데, 앞서 설명한 HTS가 바로 이러한 스크리닝을 한번에 대량으로 수행하는 것이다. 우리가 가상실험이라 함은 단지 어떤 물질이 약효가 있는지 없는지를 예측하기 위해 컴퓨터를 통한 수치모델로 수행하기 때문에 붙여진 이름이며 실제 수행 방법은 실험적인 스크리닝과는 약간 차이를 두고 있다.

Virtual ligand screening을 수행하기 위해서는 먼저 단백질의 drug binding pocket을 정확히 규명하는 것이 선행되어야 하고 이 binding pocket 내에서도 실제로 약효물질과 정확한 상호작용을 하는 Pharmacophore를 찾아 내어야 한다. Pharmacophore는 원자수준에서 작용하는 수소결합, 정전기적 상호작용과 van der Waals 작용을 하는 구역으로 약효물질의 정량적 약효를 결정하는 핵심체를 말한다[그림 5].

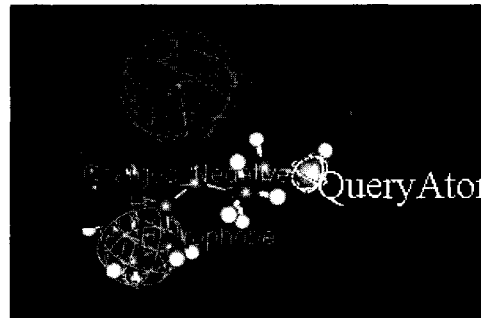


그림 5 분자 내에 존재하는 Pharmacophore의 예

Virtual screening의 다음 단계는 앞서 결정된 단백질 내의 Pharmacophore와 작용할 수 있을 것으로 예상되는 화합물을 컴퓨터 검색을 통해 찾아내는 것이고, 대상 화합물들의 데이터베이스를 구축한 후 한번에 수십만 개에서 수백만 개의 화합물의 3차 구조를 검색하여 새로운 신약선도물질을 찾아내는 것이 바로 virtual HTS이다.

Virtual HTS의 핵심기술은 앞에서 설명한 3차원 공간상에 존재하는 Pharmacophore에 잘 맞는 화합물의 3차원 구조를 어떻게 하면 빠르고 정확하게 검색해 낼 수 있는가 이다. 일반적으로 한 개의 유기화합물은 정적인 구조가 아니라 주변환경에 따라 평균

30개 이상의 동작용 구조를 갖기 때문에 10만개의 화합물 데이터베이스를 구축한 경우에 검색대상에 포함시켜야 할 3차 구조는 300만개 이상이 되게 된다. 보통 virtual HTS를 위해 구축하는 화합물 데이터베이스는 100만개 이상이며 우리나라의 경우 460만개의 데이터베이스를 보유하고 있으며 cluster computer를 이용하여 약 1.5억 개의 3차 구조를 12시간 이내에 검색할 수 있는 능력을 보유하고 있다.

최근의 연구 결과에 의하면 virtual screening 방법을 이용하는 경우 신약선도물질을 발견할 확률이 약 5~25%을 보인다고 보고 되고 있으며 이 수치는 전통적인 HTS 방법이 약 0.05% 성공률을 나타내는 것에 비하면 대단히 좋은 결과라 할 수 있다.

### 3. 결론

신약개발분야에 있어서 정보기술의 활용은 약 5년 정도의 짧은 역사를 갖고 있다. 그러나 발전 속도는 타 생물정보학 분야에 비해 훨씬 빠르게 진행되고 있다. 그 이유는 이미 축적된 생물학적 화학적 연구 결과의 데이터베이스가 충분히 축적된 상태이고 다른 생물분야에 비하면 좀 더 쉽게 수학적 모델을 개발하기 쉽기 때문이 아닌가 생각된다. 현재 활용 가능한 기술은 본 내용에서 설명한 단백질 구조예측, 가상탐색, 가상 HTS 정도이나 현재 활발하게 연구되고 있는 분야인 단백질 상호작용 예측, 단백질-펩타이드 상호작용, 단일염기변이(SNP)에 의한 약효예측 등의 발전과 더불어 향후 5년 이내에 또 다른 신약개발의 전기가 이루어 질 수 있을 것으로 기대 된다. 또한 현재는 논의의 대상으로 되고 있는 전임상, 임상실험을 컴퓨터상에서 수행할 수 있도록 할 수 있는 단백질 네트워크연구와 e-cell, e-body 연구가 결실을 맺는

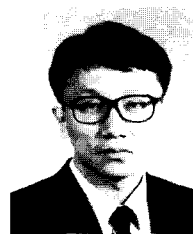
다면 명실상부한 가상신약개발(Virtual Drug Development)이 가능해 질 것이며, 현재와 같은 생물기술과 정보기술의 발전속도라면 공상과학 영화에서나 가능한 개인의 특성에 맞는 맞춤형 약의 시대도 예측해 볼 수 있을 것이다.

### 참고문헌

[1] Robeson, B. L., "Pharmatech, Business Briefing," 2002.  
 [2] McGovern, S. L., Caselli, E., Gregorieff, N and Shoichet, B. K., "J. Med. Chem.," 45, 1712-1722, 2002.  
 [3] van Gusteren, W. F., Weiner, P. K., and Wilkinson, A. J.(eds), "Computer simulation of biomolecular system: Theoretical and experimental application," 3rd eds, Kluwer Academic Publisher, Dordrecht, 1997.

---

#### 김 승 목



1980. 3~1984. 2 연세대학교 화학과 졸업(학사)  
 1984. 3~1988. 8 한국과학기술원 화학과 졸업(박사)  
 1988. 9~1990. 12 생명공학연구소 단백질화학실 Post Doc.  
 1991. 1~1997. 8 생명공학연구소 생체분자구조 RU/유전체사업단 선임연구원  
 1999. 8~1999. 12 생명공학연구소 Genome Center Bioinformatics Team 팀장  
 2000. 1~현재 (주)아이디알 대표이사  
 E mail : smkim@idrtech.com

---