

신약개발과 생물정보

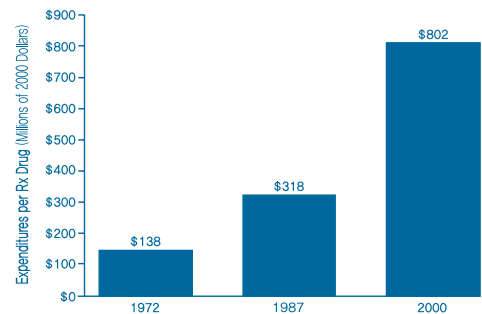
글 _ 노 경 태 교수 · 연세대학교 생명공학과 · ktno@yonsei.ac.kr

생물정보학의 발전은 신약의 탐색(Discovery) 및 개발(Development)에 지대한 영향을 미치며 그 영향력은 미래 산업의 방향 및 구조를 바꾸기에 충분하다. 의약 관련 산업은 생물산업의 거의 70%를 차지하며 생물정보 연구에 많은 연구비를 지원하고 있다. 따라서 생물정보의 다양한 기술들이 신약개발 과정에 매우 중요하게 사용되고 있다. 생물정보학의 궁극적인 목적은 모든 생명체 그중에서도 특히 인체에 관한 모든 정보를 확보하고 이를 가공, 분석 그리고 각 생명체의 생리학적인 현상들을 정확하게 모사/예측하는 것이다. 이러한 정확한 생명체들에 관한 정보 및 모델은 현대의 신약개발 연구에 없어서는 안 될 필수 요소로 등장하였다. 본 글에서는 신약개발에 사용되는 생물정보와 그들의 역할에 대해서 간단히 설명하고 일부 내용은 현재 신약개발에서 가장 중요한 분야로 간주되는 약물동력학(Pharmacokinetics) 연구를 생물정보를 바탕으로 하는 컴퓨터 시뮬레이션으로 대체해 나가는 최근의 연구 동향에 관하여 소개 하려 한다.

많은 새로운 혁신적인 기술들 예를 들면 조합화학(combinatorial chemistry), 초고속 검사(high-throughput screening), 기능유전체학(functional genomics), 단백질체학(Proteomics) 등의 도입에도 불구하고 신약을 탐색하고 개발하는 연구에는 지난 10년간의 자료에 의하면 한 가지 신약 당 약 8억\$라는 엄청난 자금이 소요된다. 또한 약을 개발하는 비용은 매년 기하급수적으로 증가하고 있다. 이는 일부 임상 단계에서의 신약후보물질의 높은 탈락율에 기인한다. 세계적인 관리 기관에 의해 신약(New Chemical Entity)으로 인증 받는 수가 최근 몇 년간 줄고 있는 상황이다. 이런 현상의 결과로서 여러 치명적인 질병이 아직 치료되지 힘든 상태로 방치되고 있다. 이러한 문제를 해결하는 길은 신약의 개발비를 낮춤으로 제한된 재원으로 더 많은 약을 개발할 수 있는 능력을 배양하는 것이다. 이러한 이유로 생물정보를 이용한 신약개발 연구가 더욱 중

요하게 되었다.

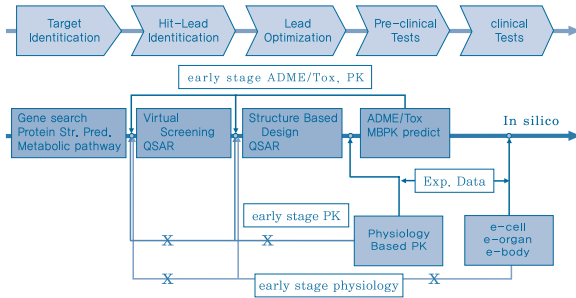
인간게놈프로젝트의 결과로 인간이나 병원체의 유전자에 관한 많은 정보들이 알려지고 있다. 이 정보는 생체 조절 단백질의 기능과 구조의 정보로 바뀌고 또한 여러 생물학적 실험을 통하여 질병 치료의 대상 단백질들이 선정되고 있으며 그 수가 종래의 500개 정도에서 3,000~10,000로 증가할 것으로 예상되고 있다. 이렇게 질병 대상 단백질(Target)의 증가는 새로운 신약개발의 방법론을 필요로 한다.



<그림 1> 신약을 개발하기 위하여 투자되는 경비가 매년 기하급수적으로 증가하고 있다.

다음의 <그림 2>는 생물정보와 신약개발의 관계를 설명 할 때 항상 등장하는 도표로 신약개발의 과정을 시간에 따라 관련기술을 순차적으로 표시하였다. 이 도표는 연구의 전략에 따라 또는 보는 시각에 따라 차이점을 보이며 또한 점차 새로운 기술들이 개발 소개 되면서 조금씩 그 형태들이 바뀌어 가고 있다.

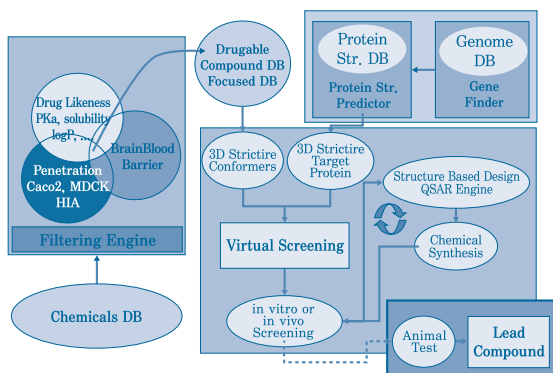
그림에서 보듯이 유전자 탐색, 단백질 구조의 비교 및 대사경로 등의 정보를 이용하면 신약개발 대상 target을 찾는 데 큰 도움이 된다. 일단 대상 target이 결정되면 가상탐색(Virtual Screening : VS)이나 Quantitative Structure Activity Relation(QSAR) 방법을 이용하여 Target 단백질과 결합할 수 있는 화합물을 찾아내게 된다.



〈그림 2〉 신약개발의 과정(맨 위의 화살표)을 단계별로 순차적으로 표시하였다. 또한 각 단계에서 사용되는 생물정보 및 컴퓨터 이용 방법들(가운데 적색 화살표)을 열거하였다.

이때 얻어진 화합물들을 Hit라 부른다. 이 Hit를 단백질과 잘 결합할 수 있는 상태로 구조를 최적화 하여야 하는데 이 과정을 최적화(Optimization)라 한다. 이 최적화 과정에서는 대상화합물의 생리활성(Biological Activity)뿐 아니라 이 화합물의 인체에서의 약물동력학(Pharmacokinetics: PK)적 성질 및 인체에 미치는 독성을 미리 예측하는 것이 필요하게 된다. 즉 신약탐색 단계에서 Lead 물질을 도출하려면 화합물의 두 성질 즉 i) 생리활성과 ii) 약물동력학적/독성 성질이 만족되어야 한다. 최근 들어 컴퓨터를 이용한 약물동력학적 성질 및 독성의 예측은 신약개발에 있어서 없어서는 안 될 도구로 간주 되고 있으며, 이러한 기술 개발을 위한 연구에 많은 인력 및 연구비가 투지되고 있다. 또한 관련 전문 회사 및 상업화된 S/W 개발에 많은 투자가 이루어지고 있다.

〈그림 3〉은 컴퓨터를 이용한 신약개발의 과정 및 관련 기술 그리고 기술들 간의 연계를 보여 주고 있다. 먼저 기술의 단위로 화합물 DB가 필요한데 현재 상업적으로 구매 가능한 화합물은 대략 8,000,000개에 이른다. 이런 화합물 및 합성 가능한 화합물의 DB가 필요하다.

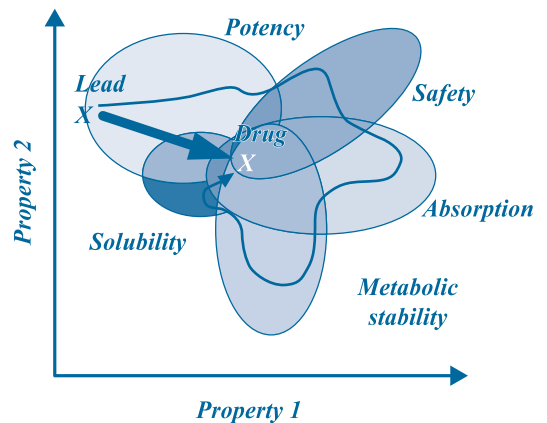


〈그림 3〉 컴퓨터를 이용한 신약탐색 과정, Lead 화합물 도출 과정을 관련 기술 및 이들 간의 관계로 표시하였다.

현재 KISTI에서 운영하는 화합물 DB도 이러한 목적으로 사용이 가능한 DB이다.

이 DB에 있는 모든 화합물이 약으로서의 성질을 만족하는 것은 아니다. 화합물중 약의 기본적인 성질을 만족하는 화합물을 골라내는 것도 중요한 과정의 하나로 여러 종류의 여과(Filtering) S/W들이 개발되어 있다. 또한 약 개발의 타겟 발굴을 위하여 유전체/유전자 DB 및 단백질의 3차 구조/기능 DB가 필요하다. 또한 실험적으로 얻어진 단백질의 3차 구조가 없을 경우 이를 예측하는 것(Protein Folding/Structure Prediction)도 매우 중요한 기술이다. 골라진 화합물(Focused DB)과 타겟 단백질과의 결합 구조 및 결합 강도를 예측하는 연구, Virtual Screening(VS)은 이미 많은 방법들이 제시되어 있으며 신약개발 연구에 가장 효율적으로 많이 사용되는 방법으로 기술상 많은 진전이 있는 분야이다. VS으로부터 얻어진 화합물은 실제 이 물질을 확보하여 생리활성도 실험을 수행하게 된다. 이 실험의 결과로부터 얻어진 활성도를 바탕으로 더 좋은 활성도를 얻기 위하여 QSAR 및 구조바탕의 최적화 과정을 수행하게 된다.(도표에서 원형 화살표로 표시)

이 반복되는 과정 동안 제안된 기준치를 만족하는 화합물을 얻게 되면 이물질을 가지고 동물 실험을 하게 되며 이 결과 실험상 기준을 만족하는 화합물을 얻게 되면 이를 Lead 화합물이라 부른다. 현재 지금까지 언급한 단계가 컴퓨터를 이용한 신약탐색이 많이 사용되고 있다.



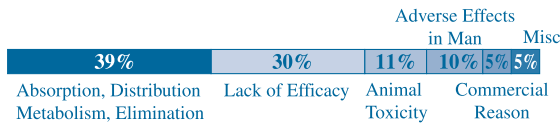
〈그림 4〉 신약개발의 과정 중 만족해야 될 성질을 보여주고 있다.

이제부터는 지금까지 언급된 기술 중 최근 매우 활발한 기술의 개발이 진행되고 있는 PK의 컴퓨터를 이용한 예측에 관하여 좀 자세하게 설명하기로 한다. 그림 4는 Lead 화합물이 약(Drug)으로 되기 위하여 만족해야 될

성질들을 보여 주고 있다. 이 모든 성질들이 만족되어야 실제 약으로서의 개발이 가능해지며 그렇지 못한 경우 신약으로서의 허가가 불가능하다.

어떤 화합물이 약으로 되려면 치료효과(Potency)는 필수적으로 갖추어야하며 인체에 부작용을 주지 말아야하며(Safety/Toxicity), 인체에서 좋은 PK의 성질을 만족해야 하며(Absorption, Distribution, Metabolism, Excretion) 기타 약으로서의 물리 화학적 물성을 만족해야한다.

<그림 5>는 1997년도에 “Drug Discovery Today”에 발표된 신약개발 단계에서의 실패를 원인별로 분류해 놓은 것이다. 이 표에 의하면 ADME 및 Animal Toxicity의 문제점으로 실패한 경우가 전체 실패의 50%에 해당한다. 이러한 문제점을 해결하기 위하여 그림 2에서 보듯이 개발의 초기 단계에 ADME/Tox를 미리 예측하여 문제가 될 소지가 있는 화합물을 초기 단계에서부터 걸러내는 작업이 필요하다. 이 걸러내는 과정(Screening)을 위하여 컴퓨터를 이용한 ADME/Tox의 예측기술의 개발에 많은 연구가 집중되고 있다.

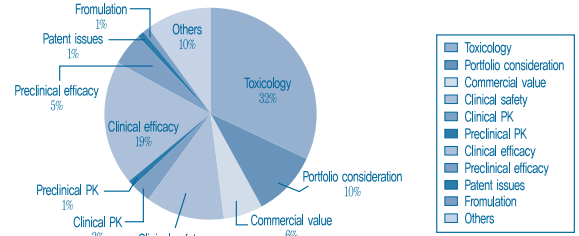


<그림 5> 신약개발 단계에서 실패하는 요인들을 정리한 것 “ADME : Single Largest Cause of Attrition in Drug Development Kennedy,T., Drug Discovery Today, 1997, 2, 436-444”

또한 <그림 6>은 2002년도의 보고서로 ADME에 의한 실패율은 매우 감소하였고 독성의 문제는 매우 크게 증가한 것을 보여 주고 있다. <그림 5>에서 보여준 ADME의 문제점에 의한 실패는 이미 컴퓨터를 이용한 예측 및 초기 단계의 in vitro(실험관) 실험의 도입으로 2002년도의 자료에는 많이 감소했음을 보여주고 있다.

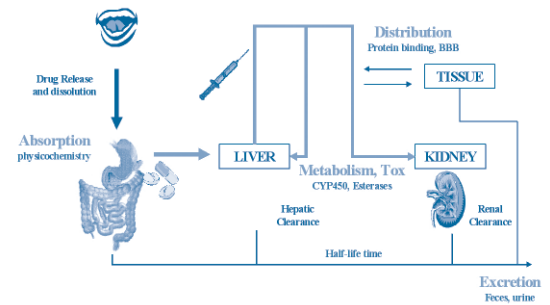
따라서 컴퓨터를 이용한 ADME 성질의 예측은 이미 신약개발에 있어서 없어서는 안 될 주요 기술로 자리잡아가고 있다. 또한 아직 독성에 의한 실패가 매우 높게 나타나있는 것을 볼 때 컴퓨터를 이용한 독성예측을 할 수 있는 기술의 개발이 중요한 연구 분야로 남아 있는 것을 알 수 있다.

ADME란 외부의 화합물, 예를 들면 약이 입을 통하여



<그림 6> 2002년도 자료로서 신약개발 단계에서의 실패 요인들을 정리한 것이다.

또는 주사기를 통하여 몸속에 들어가서 지나가는 경로 및 반응 그리고 흡수 및 배출 되는 과정에 관한 것이다. ADME란 Absorption, Distribution, Metabolism, 그리고 Excretion 인데 먼저 어떻게 약이 장을 통하여 흡수되고 인체에 각 부위에 전달되며 간에 의해서 대사가 되고 마지막으로 어떻게 배출되는가에 관한 것이다. 즉 약이란 흡수가 잘되고 원하는 신체의 부위에 전달이 되어 약효를 나타내며 간에 의해 분해가 되지 않아야 오랜 시간 약효가 지속될 것이며 또한 마지막으로 어떤 형태로 배출 되는가도 중요하다. 아래의 <그림 7>은 이 ADME의 과정을 도식으로 나타내었다.



<그림 7> 약이 경구로 투여된 후 거치는 과정을 도식하였다. 먼저 소장에서 흡수되고 간에서 분해과정이 있으며 혈액을 따라 돌며 여러 인체의 부위로 전달되고 궁극적으로는 방출이 되는 것을 보여주고 있다.

컴퓨터를 이용하여 ADME를 예측하기 위해서는 가장 먼저 필요한 것이 DB의 구축이다. 유전체 정보와 달리 ADME 분야 DB의 구축은 매우 어려운 일이다. 이 분야의 실험 데이터는 대량 생산이 힘들며 여러 실험실 간의 실험 조건이 매우 다르며 또한 자료 하나하나의 신빙성을 검증하는 과정이 필요하다. 또한 필요한 컴퓨터 기술로서는 DB 운영 시스템, 여러 linear/nonlinear regression 및 minimization 그리고 유전자 알고리즘 등 다양한 방법이 사용된다. 또한 참여하는 연구 인력의 범위가 수학, 컴퓨터, 화학, 생물, 약학 및 의학 등

거의 학문의 모든 분야를 포함하고 있다. 또한 이 ADME를 서로 연계하고 시간의 개념을 도입하게 되면 in silico Pharmacokinetic 연구가 가능해 지는데 이는 대표적인 Systems Biology의 분야로 A/D/M/E의 각 구성 성분을 시간 및 물질의 이동을 매개로 한 개의 시스템으로 통합할 수 있게 된다. 먼저 ADME의 각 현상을 실험 결과를 만족시킬 수 있도록 훈련과정(또는 regression)을 통하여 분자의 표현자(descriptors)들로 수학적으로 표시하는 과정이 필요하다.

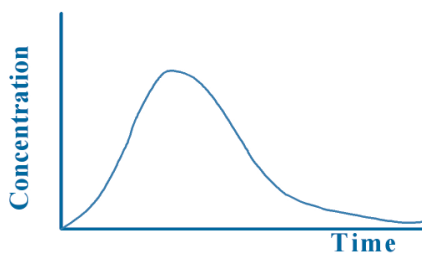
$$A=A(\text{Mol.Str.}), D=D(\text{Mol.Str.}), \dots$$

위의 A/D/M/E의 각 예측 모듈을 시간의 함수로 통합할 수 있다.

$$\text{Conc.in Blood}(t)=f(A,D,M,E,t)$$

마지막으로 분자의 구조 정보만으로 이 분자의 혈중 농도를 아래와 같이 시간의 함수로 나타낼 수 있게 되면 이는 신약 개발에 있어 매우 귀중한 연구 도구로 사용될 수 있을 것이다. 분자의 구조로부터 <그림 8>과 같은 결과를 보여줄 수 있게 된다.

$$\text{Conc.in Blood}(t)=f(\text{Mol.Str})$$



<그림 8> 혈액속의 화합물의 농도를 이와 같이 나타내는 것이 PK의 중요 목적이다.

현재 ADME 및 Toxicity를 예측 하는 상업화된 프로그램을 개발 및 판매하는 회사들은 매우 많으나 그 중 국외의 10개사와 국내 1개 그리고 개발된 상품을 정리해 보았다.

현재 ADME 예측의 기술은 아직은 초기 상용화 단계로 많은 문제점을 가지고 있어 이 분야의 연구 및 개발자에는 많은 기회가 있다고 생각된다. 또한 어떤 생물정보 관련 산업보다 훨씬 큰 시장을 확보하고 있고 빠른

성장이 예상되는 분야이다. 각자의 S/W 들은 예측할 수 있는 분야가 제한적이고 또한 각자의 전문성이 따로 있다.<표2> 따라서 국외의 제약사 등에서는 여러 S/W의 결과들을 토대로 판단하는 과정을 연구 내용 속에 포함하고 있다.

회 사	상 품
Bayer Technology Service	PK-Map™ / PK-SIM™
Accelrys	C2-ADME / TOPKAT / DS ADME
Pharma Algorithms	ADME Boxes v2.2
Tripos	Volsurf™
LION	pkEXPRESS™
Inpharmatica	Admensa™
Bio-Rad	KnowItAll ²
Strand Genomics	truPK
Schrodinger	Qikprop™
Simulations Plus	QMRPlus
BMD(Korea)	PreADME™

<표1> ADME 관련 국내·외 전문 회사 및 상업화된 S/W를 정리하였음.

	A	D	M	E	Tox	D.L.	PK
Bayer Tech. Ser.	0	0	0				
Accelrys	0	0	0	0	0	0	
Pharma Algorithms	0	0	0	0			
Tripos	0	0	0				
LION	0	0	0	0	0		0
Inpharmatica	0	0		0			
Bio-Rad	0	0	0	0	0		
Strand Genomics	0	0	0	0			
Schrodinger	0	0	0	0			
Simulations Plus	0	0	0			0	0
BMD(Korea)	0	0				0	

<표2> 각 개발 회사별 S/W의 ADME/Tox/PK/DL 분야별 예측 가능 분야를 정리하였다.

생물정보학이 직접 응용되는 산업 분야의 육성은 생물정보학을 전공하는 인력의 진로 및 지속적인 연구 재원의 확보 그리고 생물정보학의 발전을 위해서는 필수 불가결하다. 그런 의미에서 ADME 및 Tox 그리고 PK의 예측은 생물정보학 연구자들에게 매우 흥미로운 분야라 할 수 있다. 또한 이 분야 연구를 통하여 여러 학문의 분야가 공동의 목적을 달성하기 위한 융합의 기회가 될 수 있을 것이다. 