

정량적 구조 활성관계(QSAR) 모델개발 동향과 농약

옥환석*

한국과학기술정보연구원 Reseat 프로그램

(2011년 3월 10일 접수, 2011년 3월 22일 수리)

Developing Trend of QSAR Modeling and Pesticides

Hwan-Suck Ock*

Reseat Program, Korea Institute of Science and Technology Information, Daejeon 305-806, Korea

서 론

물질의 화학구조의 특성 또는 물리화학 정수와 생분해성 생물농축성·각종 독성 엔드 포인트 등과 같은 생물학적 활성과의 관계를 구조 활성 관계(structure activity relationship, SAR)라고 한다. 또 생물활성의 크기 변화와 화학구조의 특성차이를 자유에너지의 변화로 나타내어 정량적으로 예측할 수 있는 기술을 구조 활성 관계라고 하며, 최근에는 자유에너지 변화로 표현하는 것보다는 생물학적 활성의 크기를 정량적으로 예측할 수 있는 구조 활성 관계 모두를 가리키는 경우를 정량적 구조 활성 관계(quantitative structure activity relationship, QSAR)라고 한다. 그리고 정량적-정성적(quantitative-qualitative) 구조 활성 관계를 모두 (Q)SAR로 표기하며, 정성적인 구조 활성 관계를 SAR 라고 표기하여 구별하기도 한다.

QSAR은 리간드(ligands) 또는 화합물의 생리활성이 분자의 구조 또는 성질과 관계가 있으며 구조상으로 유사한 분자는 유사한 생리활성을 나타낸다는 의약 화학의 일반적인 원리에 기초를 두고 있다. 기본 구조가 같은 화합물에서 생리활성은 그 기본 구조에 결합하고 있는 치환기 등에 따라 다르고 그러한 치환기 혹은 화합물의 물리화학적 성질과 생리활성의 차이는 컴퓨터가 일반적으로 숫자로서 작동되기 때문에 수식으로 표현하는 것이다.

수용체와 리간드의 상호작용에서 중요한 인자는 소수성, 전자효과, 수소결합상호작용이며, 용매화 에너지도 중요하게 작용한다. QSAR 수식은 일반적으로 전자 효과 파라미터인 Hammett의 시그마(σ)상수, 소수성 파라미터인 Hansch의

소수성 치환기정수(π) 그리고 입체 효과의 파라미터인 Taft의 입체 인자(E_s)를 수식화하고 이식을 이용하여 보다 더 생리 활성이 높은 화합물을 예측한다. 그리고 이러한 분자구조상의 정보를 분자 설명인자로서 코드화하고 알려지지 않은 리간드의 활성을 예측하기 위하여 알려진 리간드의 설명자와 생물활성의 수학적 관계를 QSAR 모델로 정의한다.

또한 효율적인 합성과 조합화학을 통해 만들어지는 라이브러리의 화합물 구조와 관련된 정보와 함께 생물학적 데이터는 초고속 고효율 약물 스크리닝(high throughput screening, HTS)을 통하여 기하급수적으로 발생하고 있다. 이러한 모든 정보를 화학, 생물학, 수학, 전산학 등의 학문이 융합되어 처리하고 분석하여 컴퓨터상에서 정보화되고 규칙화하여 데이터베이스를 구축하고 활용하는 것을 Chemistry와 Informatics의 합성어인 화학정보학(Chemoinformatics)이라고 하는 과정을 거쳐 내장된 컴퓨터 지원 약물설계(Computer Aided Drug Design, CADD) 또는 컴퓨터 지원 분자설계(computer-assisted molecular design, CAMD)연구가 광범위하게 이루어지고 있으며, QSAR도 이 연구의 한 분야이다.

또한 라이브러리의 화합물 구조와 관련된 정보와 생물학적 데이터를 분석하는 요인 분석(factor analysis, FA)은 행동 과학자에 의해 발견되었으며, 초기 역사는 1901년에 Pearson에 의해 발표된 직교회귀선에 관한 논문을 그 원류로 Harman는 기록하고 있다. 그리고 첫 번째 실질적인 발달은 1933년에 Hotelling에 의해 완성되었으나 화학계에서는 주목을 받지 못하였다. 그러나 1970년대에 화학과 컴퓨터 과학의 획기적 조화로 분석화학의 한 분야로 화학실험에서 얻은 데이터를 다루고 해석하며 통계학적 수학적 모형을 통하여 중요한 화학적 정보를 추출하는 Chemometrics 기술이 태동하면서 번

*연락처 : Tel. +82-42-869-0685, Fax. +82-42-869-0699

E-mail: hs123@reseat.re.kr

영하게 되었다. 요인분석은 변수들 간의 상관관계를 고려하여 서로 유사한 변수들 끼리 묶어주는 방법내지는 많은 변수로 구성된 데이터가 몇 개의 요인에 의해 영향을 받는가를 알아보는 것이라고도 할 수 있으며, chemometric이 가지고 있는 가장 유력한 기술의 하나라고 알려져 있다. 특히 Chemometrics 기술에는 화학, 통계학, 응용수학, 산업공학 등의 학문이 근간을 이루고 있으며 이 기술의 핵심은 정확한 성분예측을 위한 모델의 개발이라 할 수 있고 현재는 부분최소자승법(partial least squares, PLS)을 주로 사용하고 있다.

QSAR은 이미 존재하거나 가상으로 설정한 분자의 결합 친화성이나 독성가능성과 같은 측정량을 추정하기 위하여 실질적 모형을 만드는 컴퓨터 연구 분야로서 현재의 유망한 약물을 확인하고 새로운 화합물질(new chemical entity, NCE)의 분자설계에 이용할 수 있다. 수용체 구조는 모르나 활성을 가진 일련의 화합물에 대한 정보를 가지고 있을 경우 이 화합물을 통하여 수용체부위에 기하학적으로 대응하는 3차원 공간에서의 작용기의 위치에 따라 특징을 나타내는 약물분자구조를 결정한다. 따라서 이런 정보를 바탕으로 활성과 불활성을 나타내는 화합물들의 구조적인 특성들을 통계적으로 분석하여 활성에 영향을 미치는 주요 인자를 찾아 활성이 높은 새로운 화합물을 설계하는 리간드기반 접근(ligand-based approach) 또는 리간드기반 분자설계(ligand-based drug design, LBDD)에도 이용된다. X-ray 회절 연구에서 가능하게 되었던 단백질의 입체 구조가 증가하면서 구조기반 설계(structure-based design, SBD)는 상호작용 또는 분자역학(MM)과 동적 최적화(MD)와 더불어 자동화된 도킹을 사용하여 고분자 결합 위치(pocket)의 위상에 구조를 알맞게 한 높은 친화력을 가진 작은 분자 리간드를 확인하는 유망한 도구로 1980년대 부상하였다(Markus A. Lill, 2007).

효율적인 화합물 라이브러리를 구축하고, 구축된 화합물 라이브러리로부터 생물학적 약효 활성 구조에 기초하여 새로운 리간드들을 빠르게 찾아내는 가상 탐색에도 이용된다. 화합물들의 생물학적 이용가능성(bioavailability)을 향상시키는 중요한 물성들을 예측하기 위하여 가상실험 조건에서의 컴퓨터 프로그래밍(in silico)은 흡착, 분포, 대사, 제거, 독성(adsorption, distribution, metabolism, elimination, toxicity, ADMET)을 예측하고 새로운 구조의 신물질을 설계하는 신생설계(De novo design) 뿐만 아니라 화학정보학(Chemoinformatics) 분야를 망라하는 제약, 농약산업 등에서 필수적인 방법으로 활용되고 있는 핵심첨단기술이다.

1863년 Cros가 알코올의 물에 대한 용해도와 포유류에 대한 독성은 반비례하는 것으로 관찰되었고 1868년 Crum-Brown

과 Fraser가 알칼로이드의 생리활성에 관한 연구를 시작하여 1894년 자물쇠와 열쇠의 유추법인 Fischer 공식이후 Hansch, Fujita, Free와 Wilson의 공헌으로 QSAR은 신약개발 과정에 실질적으로 기여하고 널리 이용되는 도구로 발전하였다.

OECD를 중심으로 동물시험에 대한 윤리적 문제가 제기되면서, 위해성 평가방법을 대체할 기술이 절실히 요구되고 있다. 따라서 정량적 구조-활성관계(QSAR)를 국제적 위해성 평가 및 검증 자료로 이용하기 위해 QSAR 표준화 정책을 제시하고 있어 농약관련 연구자나 종사자들도 이에 대해 이해할 필요가 있다. 따라서 필자는 약효 검증대상 기질화합물들이나 분자 내 특정 치환기의 이화학적 성질과 활성형태 그리고 생물활성간의 상관성과 생리활성 발현 메커니즘에 대한 단서를 이해하는 데 도움이 될 수 있도록 QSAR 기법을 소개하고 환경독성이나 생체측적 등의 예측과 같은 활용방법에 쉽게 접근할 수 있는 방안을 소개하고자한다.

QSAR의 발전

1863년 Cros가 알코올의 물에 대한 용해도와 포유류에 대한 독성은 반비례하는 것으로 관찰한 이래 1868년 Crum-Brown과 Fraser에 의하여 알칼로이드의 생리활성에 관한 연구를 시작하였다. 1894년 자물쇠와 열쇠 유추법인 Fischer 공식이 발표되고 1899년 Charles Richet는 유기 화합물에 대한 이화학적 성질과 독성간의 관계를 처음으로 보고하였다.

1901년 Meyer와 Overton은 마취작용에 관한 리포이드이론(lipoid theory)인 Meyer Overton 가설을 발표하여 생리활성 크기에 지질과 물 분자간의 분배계수인 소수성(hydrophobicity)이 중요한 의미를 가진다는 사실을 처음으로 확인하였다.

1930년대에 Hammett는 벤조산(R-H) 유도체 치환기(R-X) 변화에 따르는 산해리 상수에 대한 비의 대수를 σ 상수로 정의함으로써 반응에 기초한 자유 에너지 직선관계(LFER)로부터 전자적 성질인 평형상수(K)는 물론, 산해리상수(pKa)와 정량적인 관계가 있다는 Hammett 방정식 (1)을 제안하였다. 이때 σ 상수는 방향족 화합물에 국한하여 사용할 수 있는 치환기 상수로서 비치환-phenyl 고리의 H 원자를 기준으로 치환기의 전자 수수정도에 관한 상수이고 ρ 는 반응감도를 나타내는 비례 상수이다.

$$\log K_{R-X} - \log K_{R-H} = \rho\sigma \quad (1)$$

Hansch는 자유에너지선형관계(Linear free energy relationship, LFER) 식에 기초하여 소수성을 포함시켜 화합물의 생물학적인 유용성을 설명하였다. 따라서 약물과 수용체간의 상호작용

으로 생물활성 $\log 1/C$ 는 (2)식이 성립하고 이 식은 크게 전자 전달효과(electronic effect, ρ)와 입체효과(steric effect, E_s) 그리고 소수성효과(hydrophobic effect, π)의 3가지 형태로 분류되는 이화학적 성질에 대한 선형 대수함수로 설명함으로써 QSAR 기법의 초 열역학적 접근(extra-thermodynamic approach)이 약물설계에 활용될 수 있게 되었다.

$$\log 1/C = a\pi^2 + b\pi + \rho\sigma + cE_s + d \quad (2)$$

QSAR식은 보통 다중선형회귀(Multiple linear regression, MLR)와 같은 통계방법을 사용하여 이화학적 성질을 독립변수로 생물활성(biological activity, BA)을 종속변수로 표현함으로써 본격적으로 2D-QSAR 연구기반을 구축하였다. 이때 BA항은 실험적으로 측정되는 기질물질의 농도($\log 1/C$), p_{150} , IC_{50} , 그리고 K_m 등과 같은 척도인 반면 우변 항의 독립변수들은 수용체와 상호작용에 관여하는 기질분자의 위상학적인 상수나 물리화학적 성질을 나타내는 전자 전달효과와 입체효과 그리고 소수성 효과와 같은 형태의 설명인자들이다.

$$BA = a_1x_1 + a_2x_2 + \dots + a_nx_n \quad (3)$$

1988년 Cramer 등은 Hansch 식에서 언급된 3가지 이화학적 성질을 비교분자 장으로 표현하는 3D-QSAR인 CoMFA 기법을 제안하여 현재 가장 보편적으로 활용되는 방법으로 인식되기에 이르렀다. 1997년에는 화합물의 구조적 특징과 생물활성간의 예측관계를 부분 최소자승(partial least square, PLS) 방법으로 분석하는 분자홀로그램(HQSAR)기법 그리고 Hopfinger 등에 의하여 3D-QSAR 기법의 2가지 한계반응 점의 가상모형과 유연한 생물 조절제를 표현하는데 있어서 평준화된 수용체 모델의 유용성 등을 보강한 리간드 분자의 양성자화 상태, 다중형태 그리고 지향성이 고려된 개념의 4D-QSAR 기법이 알려졌다. 최근에 Vedani 등은 유도적응(induced-fit) 가설의 다중적인 설명뿐만 아니라 리간드의 위상학적인 복식표현 개념의 5D-QSAR 기법과 QSAR소프트웨어에서 정량화를 위한 결합 형태와 6D-QSAR의 확인을 위해 유연한 도킹(Yeti 소프트웨어)을 겸하는 약물과 화학제품이 촉발시키는 역효과의 가상실험에서 컴퓨터 프로그램(*in silico*) 시뮬레이션에 대한 개념이 등장하기에 이르렀다.

따라서 모든 QSAR 식은 방법과 표현은 다르다 할지라도 구조상의 요소와 생물활성 크기에 대한 2가지 변화들 간의 대응 논리를 근간으로 하고 있으며, QSAR 기법은 활성 메커니즘을 규명하려는 생물학적 실험과 기질분자-수용체간의 구조와 상호작용에 관한 화학적 이론 그리고 이들 두 요인 분석

방법, 통계학적 경험의 3가지 유형을 망라하는 최적화 수단으로 QSAR 기본개념은 다음 식과 같이 표현된다. 이때 x_i 는 구조적인 성질들로 최소자승 회귀분석으로 얻어지는 값이고 a_i 는 x_i 의 계수이며 m 은 화합물의 수이다.

$$BA = \sum_{i=1}^m a_i x_i + Const \quad (4)$$

모든 화학반응의 속도 상수와 온도와의 관계는 Arrhenius 식 [$kr = A \exp(-E/RT)$]으로 잘 정리되어 있으며, Hammett 관계식들은 모두 1차식 ($\log k = a \log K + b$)의 관계를 나타낸다. 따라서 온도가 일정한 조건에서 이들 두 관계식을 짝지어 정리하면 다음 식과 같이 자유에너지 변화인 ΔF^\ddagger 와 ΔF° 사이에는 직선관계가 성립한다. 그러므로 Hammett 식이 성립할 때, 자유에너지 직선관계($\Delta G = -RT \ln K$)가 존재한다고 하며, 이 식의 성립조건은 한 계열의 반응 메커니즘이 같고 구조가 대단히 유사한 경우라고 말할 수 있다. 따라서 자유에너지 직선관계(Linear free energy relationship, LFER)는 유사성의 정량적인 모델이라고 해석될 수 있다.

$$\Delta F^\ddagger = a \Delta F^\circ + Constant \quad (5)$$

이상과 같이 화합물이 생체 내 흡수된 후 활성발현은 작용점에 이르는 수송과정과 작용점에서 일어나는 상호작용 그리고 그 결과로 인하여 관측되는 생물활성을 발현하는 과정 등 3가지 과정으로 구분되며, QSAR 기법은 구조상 요소의 변화와 활성 변화에 대응하여 발전되고 있다. 신농약 개발의 모태로서 약물화학은 생명관련 산업의 중추적인 학문분야이며, 물리·유기화학을 기본 바탕으로 하는 약제 개발수단(Hansch, 1995)으로서 QSAR 기법을 이해하고자 함은 약리 효과가 가장 큰 화합물을 탐색하는 데에 따른 비용과 시간 및 노력 등을 최소화함으로써 약물개발 연구에 있어서 가장 효율적인 방법이기 때문이다.

QSAR 기법의 3가지 특징적인 유용성은 첫째 생물활성 실험의 질적인 신뢰성과 자료에 대한 합리성을 추구하며, 둘째 실험 결과의 모순점은 물론, 실험값과 모델에 의하여 예측된 값들 사이의 편차를 알 수 있고 투약 량 범위를 평가하여 진단하며 그리고 셋째 분자설계 시 독성은 최소화되 활성은 극대화할 수 있을 뿐만 아니라, 아울러 그 성질들을 예측하는 일이다.

수많은 자료를 분석하고 검토하는 데에는 필수적인 컴퓨터 기술이 발달하였고, 많은 화합물의 합성과 그에 따른 생물활성을 검증하는데 있어서 실험 량을 줄여야 할 필요성이 대

두되었으며, 생물활성 메커니즘과 수용체 구조와 성질 및 대사 등에 관련한 지식수준의 량과 질이 크게 향상됨에 따른 학문적인 체계를 확립하고 합리성을 추구하여야 할 필요성이 증대되었다. 이상과 같이 구체적이고 체계적일 뿐 아니라, 신속하고 정확하게 기질과 수용체간의 관계를 이해할 수 있도록 함으로서 새로운 농약을 개발하는데 있어서 QSAR 기법 도입은 필수적이다.

QSAR 모델 개발

일반적으로 QSAR 모델을 만드는 것은 Fig. 1과 같이 훈련 데이터 세트를 사용하여 데이터 세트 준비, 분자 설명자 선택과 생성, 수학적 또는 통계 모델 유도, 모형 훈련 그리고 시험 데이터 세트에 훈련 테스트세트와 테스트 모델시험을 사용하여 정당성을 입증하는 여러 단계의 체계적인 과정이

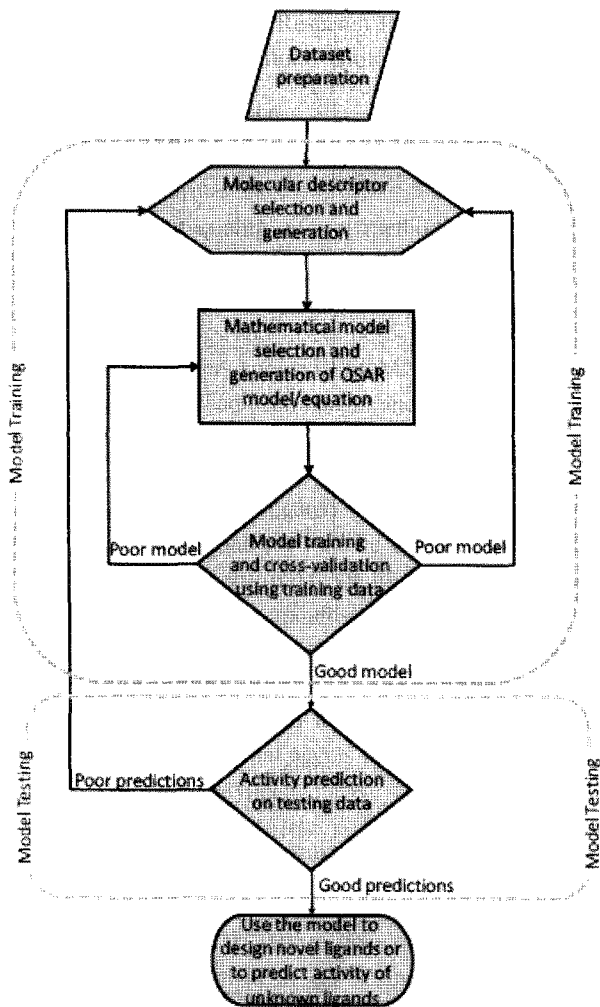


Fig. 1. A general scheme of a QSAR model development which includes systematic training and testing processes.

있다. 1단계 데이터 세트 준비에는 확실한 QSAR 모델을 개발하기 위해서 데이터의 품질에 유의하는 것이 중요하다. 데이터는 동일한 생물 검정 협약에서 나와야 하고 데이터가 같지 않고 연구실간 차이를 피하기 위하여 한 실험실 또는 공급원에서 만든 자료를 모아 사용하는 것이 좋다. QSAR 모델을 만드는 두 번째 단계는 데이터 세트에 리간드에 대한 분자 설명자를 선택하여 만드는 것으로 이용할 수 있는 설명자는 많이 있고 그 중에서 몇 가지가 활성과 현저한 상관관계가 있다. 따라서 구조적 변형과 정보를 가장 잘 포착한 적절한 설명자의 선택이 확고한 QSAR 모델을 유도하는데 중요하다.

분자설명자(molecular descriptor)는 데이터 세트에 있는 모든 리간드를 정의하고 생성된 후에는 통계적이고 수학적 모형으로 생물활성과의 관계를 탐색하기 위해 결정하는 것이다. 예를 들면 다중선형회귀(multiple linear regression, MLR)나 부분최소제곱(partial least square, PLS)과 같은 선형접근 그리고 신경네트워크 또는 지원벡터머신(support vector machine)과 같은 비선형방법은 상관관계나 사상함수(mapping function)로 사용될 수 있다. 일단 모델이 선택되면 남겨둔 나머지는 테스트 화합물로 사용되고 알려진 데이터 세트로부터 무작위로 선정한 화합물의 하위세트를 포함하는 훈련 데이터 세트로 훈련시킨다. 모델 훈련에 단일잔류교차검정(leave-one-out cross-validation, LOOCV)과 같은 확인 방법은 QSAR 모델의 통계적인 안정성을 유지하기 위하여 실행하며, 훈련 과정은 만족한 훈련 성과가 주어질 때까지 반복한다. 마지막으로 테스트 과정은 시험 장치에 있는 화합물의 활성 값을 예측하기 위하여 이용된 훈련한 모델에서 실행한다.

QSAR의 유형

QSAR의 유형은 한 분자의 생물학적 효과를 설명하기 위하여 용해도나 해리상수 같은 단 하나의 이화학적 성질에 기초를 두는 1D-QSAR, Hansch, Fujita, Free와 Wilson이 단 하나의 원자와 작용기의 물리 화학적 성질 그리고 그들의 생물활성에 대한 기여를 고려해서 화합물의 연결성을 암묵적으로 포함시킨 2D-QSAR, Hansch-Fujita식에 치환기의 길이와 폭과 같은 3D 구조상 설명자를 포함하는 3D-QSAR로 기질 화합물의 구조를 평면 구조로 간주하느냐 아니면 입체적인 3차원구조로 연구하느냐에 따라 2D-QSAR, 3D-QSAR로 각각 구분된다. 그러나 4D, 5D 그리고 6D-QSAR 기법은 이들 단계의 하위단계 기법의 제한점을 개선 향상한 구분이다(Table 1).

Table 1. Classification of QSAR approaches based on their dimensionality

| Dimension | Method | Protein |
|-----------|---|----------|
| 1D-QSAR | Affinity is correlated with global molecular properties of ligands, that is one value per property and ligand (pKa, log P, etc.) | No |
| 2D-QSAR | Affinity is correlated with structural patterns (connectivity, 2D pharmacophore, etc.) without consideration of an explicit 3D representation of these properties | No |
| 3D-QSAR | Affinity is correlated with the three-dimensional structure of the ligands | Possible |
| 4D-QSAR | Ligands are represented as an ensemble of configurations | Possible |
| 5D-QSAR | As 4D-QSAR + explicit representation of different induced-fit models | Yes |
| 6D-QSAR | As 5D-QSAR + representation of different solvation scenarios | Yes |

2D-QSAR 방법

Free-Wilson과 Hansch-Fujita 모형과 같은 종래의 2D-QSAR 방법은 정량적 예측을 실행하기 위하여 2차원 분자의 치환기 또는 분획 그리고 그들의 물리 화학적인 성질을 이용한다. 여기서는 최근에 개발한 분획과 유사성을 기초로 하는 2차원 QSAR 방법의 개관을 소개한다.

Hologram QSAR(HQSAR)

분자분획을 기초로 하는 방법 중 하나는 Tripos에서 개발한 홀로그램 QSAR(HQSAR)이다. HQSAR는 2차원 분자

분획을 기초로 하는 방법으로 분자 정렬이 필요하지 않으며, 수동으로 하지 않고 많은 자료를 자동분석 한다. HQSAR 방법에서 첫 단계는 분자 분획의 카운트를 포함하는 분자 3차원 입체상을 재생하는 홀로그램을 생성하는 방법이다. Fig. 2에서 묘사된 대로 분자들은 1에서부터 홀로그램의 총 길이인 L까지의 범위로 몇 개 분획으로 나누어지고 일정길이의 분자 홀로그램 형태로 점유되는 분획(bin)은 기호구성과 형태적인 분자에 대한 정보를 기호화한 구조적인 설명인자들이 된다. 이 분자분획들은 주기적 덧셈셈 검사(cyclic redundancy check, CRC) 연산에 의하여 양의 정수로 표시된다. QSAR 목적으로 생물활성 매트릭스가 만들어 지며, 부분 최소 제곱(PLS)

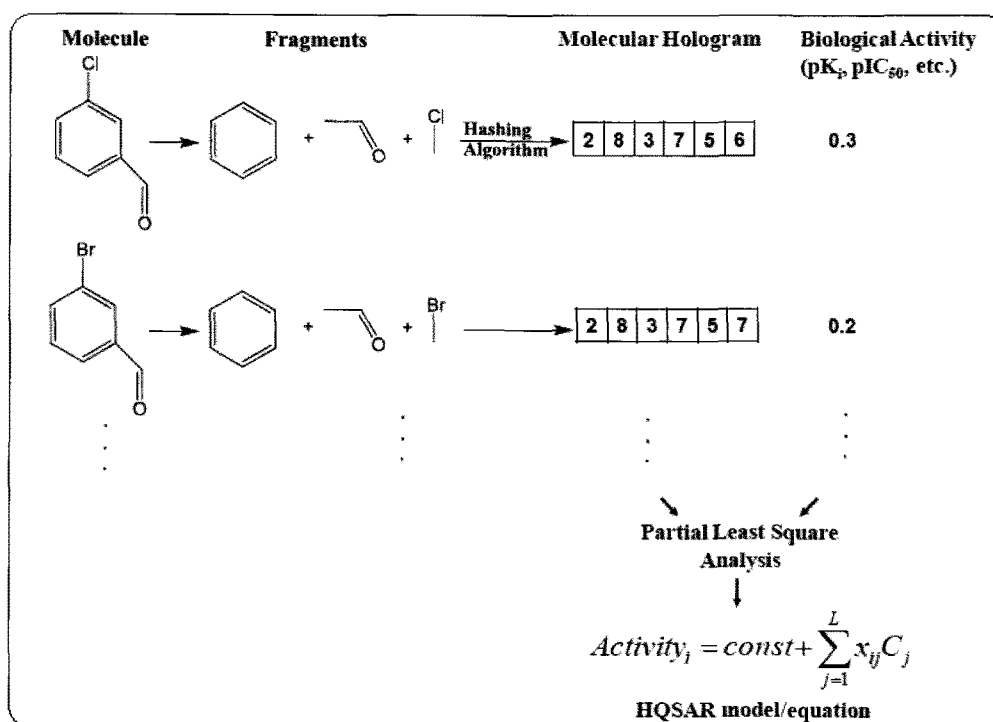


Fig. 2. Hologram-QSAR (HQSAR) model development, which includes molecular hologram generation and partial least square analysis to derive a final predictive HQSAR equation.

분석은 직교 설명변수들을 확인하기 위하여 적용되는 것으로 고유 L 변수들의 선형조합이다. 따라서 분자분획 양 또는 홀로그램 bin들은 수학 방정식의 형식에 준하는 종속변수로 생물 활성도와 관련되고 LOOCV은 최적수량의 설명변수 또는 최적모형을 만드는 구성요소를 확인하기 위하여 수행된다. 이 때 표준 PLS 분석을 이용해서 뒤에 오는 회귀방정식은 홀로그램 bin 값 또는 이에 준하는 생물 활성도와 함께 구성요소를 관련시켜 유도한다.

$$BA_i = const + \sum_{j=1}^L x_{ij} C_j \quad (6)$$

여기서 BA_i 는 i 번째 화합물의 생물활성도이고, x_{ij} 는 i 번째 위치 또는 bin j 위치 화합물의 분자 홀로그램에 대한 점유 값이다. C_j 는 PLS 분석에서 유도된 bin j에 대한 상수이고, L은 홀로그램 길이이다.

HQSAR의 한 가지 결점은 분획을 찾아내는 과정에서 일어나는 분획 충돌 문제라고 부르는 현상이며, 홀로그램의 길이를 감소시키는 것은 bin이 같은 bin에서 다른 분획을 가지고 있는 것이 원인이다. 홀로그램 길이는 사용자 정의가 가능한 매개변수인 홀로그램에서 bin 수와 홀로그램 길이 조절이 bin 점유모양이 변화하는 원인이 될 수 있다. 홀로그램의 길이는 PLS 계산으로 반응 점 탐색에 이용된다. 홀로그램과 생물활성간의 상관성을 계산한 PLS계수를 공간상에 구조로 나타냄으로써 반응 점들에 대한 정보를 얻을 수 있으며, 화합물의 활성에 관련된 분자 분획들의 유형을 확인함으로써 분자들의 활성을 예측할 수 있다. HQSAR기법은 각 분자에 대한 구조적인 분획의 생성과 홀로그램 중에 분자 분획들의 암호화 그리고 생물활성자료의 상관분석이 이루어지는 3가지 주된 과정으로 이루어진다.

Fragment-Based QSAR(FB-QSAR)

최근 Du 등은 분자분획에 기초하여 Hansch-Fujita 선형 자유 에너지 방정식과 Free-Wilson 방정식을 혼합 사용하는 2D-QSAR 방법을 소개하였다. 특히 분자 분획은 리간드에서 처음으로 생성되고 리간드 i와 수용체간의 자유 에너지를 묶는 전체결합은 모든 분획에서 기여도 $\Delta g_{i,\alpha}$ 의 합계로 간주된다.

$$\Delta G_i^o = \sum_{\alpha=1}^M b_{\alpha} \Delta g_{i,\alpha} \quad (7)$$

여기서 $\Delta g_{i,\alpha}$ 는 분획의 자유 에너지 기여도이고 $F_{i\alpha}$ 와 b_{α} 는

각 분획에 대한 무게상수이며, 자유 에너지결합 $\Delta g_{i,\alpha}$ 는 분획의 이화학적 성질 상태로 기술된다.

$$\Delta G_{i,\alpha} = \sum_{l=1}^L a_l p_{i,\alpha,l} \quad (8)$$

$p_{i,\alpha,l}$ 는 분자에서 분획 $F_{i,\alpha}$ 의 l 번째 성질이고 m_i 와 a_i 는 분획의 i번째 성질의 상수이다.

정량적 결과는 새로운 방법인 반복 이중 최소자승법(iterative double least square, IDLS) 실행절차는 예측능력을 강화하고 예측 잠재력을 탐색하기 위해 필요하다는 점을 입증하였다.

Fragment-Similarity Based QSAR(FS-QSAR)

최근에는 기존 Free-Wilson 방법의 중요한계를 해결하기 위하여 선형회귀방정식에 분획유사성 개념을 도입한 분획유사성기반 QSAR(FS-QSAR) 방법이 소개되었다. 유사성 개념은 유일한 해결책이라고 할 수 없는 이화학적 성질을 이용하는 대신에 종래의 Free-Wilson 방정식을 개량하기 위하여 처음으로 적용하였다. 이 접근방법에서 분획 유사성 추정 은 유사성에 의해 수행되는데 각각의 원자의 부분적인 전하와 각 분획에 있는 그들 원자의 연결정보를 가지고 있는 BCUT-행렬에서 산출되는 가장 낮거나 가장 높은 고유의 값을 사용했다. FS-QSAR의 새로운 방정식은 다음과 같다.

$$-\log K_i = const + \sum_{j=1}^N [\max_{k=1}^{P_j} \{Sim(F_{jk}, F_{jg})\}] \times A_j^{MSF} \quad (9)$$

여기서 N = 치환기위치 총계, P_j = j번째 치환기 위치에 가능한 치환기의 총계, \max = 최대 기능은 유사성 점수 중 최대 점수를 선택했다. F_{jk} = j번째 치환기 위치에 k번째 분획(연수 때 알려진 분획), F_{jg} = j번째 치환기 위치에 주어진 분획(미지 화합물 또는 점검분획), $Sim[F_{jk}, F_{jg}]$ = 분획 유사성 기능은 F_{jk} 에 F_{jg} 를 비교하여 유사성 점수를 계산한다. A_j^{MSF} = j 번째 치환기위치 에서 가장 유사한 분획 상수.

방정식 (9)에서 사용된 유사성 기능은 다음과 같이 정의된다.

$$Sim(F_{jk}, F_{jg}) = e^{-|EV(F_{jk}) - EV(F_{jg})|} \quad (10)$$

$EV(F_{jk})$ = 분획(F_{jk})의 BCUT모체의 가장 낮거나 가장 높은 고유 값.

알고리즘은 83 COX2 아날로그와 85 triaryl bis sulfone 아날로그를 포함하여 다른 데이터 세트에 개발되고 시험되었으며, 통계 모델에서 화합물 무작위 선택으로 생성된 5개의

다른 테스트 세트에 반복적으로 시험되고 평균제곱 상관계수 r^2 는 5번 이상의 시험세트로 COX2 아날로그에 대해 0.62와 bis sulfone 아날로그에 대해 0.68이었다. Free-Wilson 방법 비교 역시 COX2 데이터세트는 0.46 그리고 bis sulfone 데이터세트에서는 0.42로 r^2 평균값을 달성하는 시험이었다. 더구나 BCUT 유사성 함수는 Tanimoto 상수(Tc) 방법, 종래의 2차원 분자 유사성함수로 대체되고, 평균 r^2 는 COX2와 bis sulfone 아날로그 모두 0.62이었다. 회귀방정식으로 유사개념을 소개해서 본래의 Free-Wilson 방법의 중요한 한계를 해결했기 때문에 FS-QSAR 방법은 종래의 2D-QSAR 방법과 비교한 효과적인 예측능력이 입증되었다. 따라서 FS-QSAR의 예측 정확도는 보다 높은 차원의 다른 QSAR 방법만큼은 높지 않을지도 모르지만 객관적이고 유일하며 재생 가능한 2D-QSAR 모델을 제공한다.

Top Priority Fragment QSAR

Casalegno 등(2006)은 무지개송어에 대한 살충제의 수생 독성효과를 예측하기 위해 중요원인이 되는 화합물 분획의 하나를 추정하여 독성원인의 우선순위를 정하는 방법인 분자 분획을 기반으로 한 QSAR을 소개하였다. 회귀방정식으로 유사성 개념을 소개해서 본래의 Free-Wilson 방법의 중요한 한계를 해결했기 때문에 FS-QSAR 방법은 종래의 2D-QSAR 방법과 비교하여 효과적인 예측 능력이 입증되었다. 그러나 FS-QSAR의 예측 정확도는 다른 더 높은 차원의 QSAR 방법만큼은 높지 않을 지라도 객관적이고 유일하고 재생 가능한 2D-QSAR 모델을 제공한다.

기타 Fragment-Related QSAR 연구

최근에는 생물학적 관심사항 적용에 더하여 일부 새로운 분획기반 QSAR 방법이 개발되었다. Zhokhova 등(2007)은 표지 원자로 분획 설명자로 사용하는 방법을 도입하면서 그들이 수행한 QSAR/QSPR(구조적 정량특성관계) 연구에 적용하였다. 그들이 접근한 방법은 QSAR/QSPR 모델을 구축하고 표지 원자로 분획 설명자를 생성하기 위해 사용한 NASAWIN 프로그램에서 신속다단계멀티선형회귀(FSMLR)와 3개 층의 인공신경네트워크(ANN) 법으로 수행하였다. Andrade 등(2008)은 안티결핵제제로 일련의 결핵치료제를 연구하기 위한 기타 2D-QSAR 프로그램과 HQSAR를 활용하였다. 또한 2차원 분자 설명자 선발과 발생에는 DRAGON 5.4 버전, BuildQSAR 및 PIROUETTE 프로그램을 활용하였다.

3D-QSAR

3D-QSAR 방법은 2차원의 예측 정확도를 개선하기 위하여 개발되었으나 계산적으로 더 복잡하다. 일반적으로 3D-QSAR 방법에는 정렬에 의존하는 방법과 의존하지 않는 방법 두 가지가 있으며, 두 방법 모두 연구를 위한 원형으로 리간드의 실험적 또는 계산적으로 유도된 생물활성 구조가 필요하다. 3차원 형태는 믿을 수 있는 3D-QSAR 모델을 만드는 중요한 요인의 하나이기도하고 주된 결점이기도 하다.

Comparative Molecular Field Analysis (CoMFA) and Comparative Molecular Similarity Indices Analysis (CoMSIA)

3차원적인 분자구조를 기초로 한 3D-QSAR기법 중에서 Cramer 등(1988)에 의해 소개된 비교 분자 장 분석(comparative molecular field analysis, CoMFA) 기법은 QSAR과 분자설계에서 보편화된 방법이다. 분자장을 생성하기 위하여 첨가되는 리간드의 입체장(steric field)과 정전기장(electrostatic field)에 대한 등고선 지도(contour map)를 통하여 기질분자의 이화학적 성질과 생물활성 사이의 상관성을 연구하는 방법이다. CoMFA 방법은 주관적인 분자정렬과 생물활성 구조 문제와 같은 어려움으로 알려진 3D-QSAR 문제에 당면하여 4D-QSAR 5D-QSAR, 등과 같은 다차원(mD)-QSAR 뿐만 아니라 CoMSIA, SOMFA, CoMMA 같은 다 같은 3D-QSAR 방법의 발달을 위한 기초를 다졌다.

CoMFA 분석을 실행하기 위해서는 먼저 기질분자의 구조식들을 그려 넣고 분자내 원자사이의 반발력이 최소화되어 전체 에너지가 가장 낮은 안정한 상태의 분자를 모의담금질(simulated annealing)과 유전학적 알고리즘(genetic algorithm, GA) 방법 등으로 최적화한다.

회전 가능한 모든 결합을 일정한 각도씩 회전함에 따른 계통적인 탐색을 재 실시하여 기질분자의 최저에너지형태를 계산한다. 그리고 분자 내 모든 원자의 부분전하는 Gasteiger-Marsili 방법을 사용하여 계산하고 자료중의 분자와 분자의 기본골격으로 구성된 주형(template)분자간의 반응요인들을 최소자승 집합에 따르는 겹침 법칙에 따라 3차원 공간에서 원자적응(atom fit) 또는 유연적응(flexible fit) 방식으로 정렬(alignment)시킨다. 정렬된 분자들을 대개 2Å 공간을 가지는 3차원 입자 입자격자 중에 놓은 다음에 Fig. 3과 Fig. 4와 같이 화합물들의 형태를 포함하는 열(row)과 성질을 나타내는 행(column)으로 표시하는 QSAR 표를 작성한다. Receptor 부위에 상응하는 기질 분자와 상호작용 에너지를 계산하기 위하여 기질

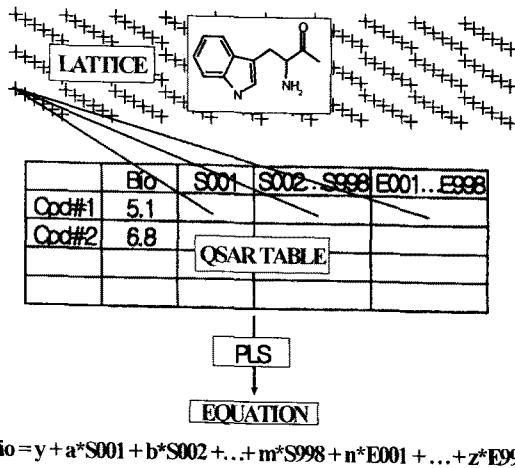


Fig. 3. Process of comparative field analysis (CoMFA) as 3D QSAR.

분자 주변에 Methyl-기(C:sp³, 반경 1.5 Å, 하전 +1.0)와 같은 일방 격자점을 생성시키고 van der Waals의 입체장과 정전기장(Coulombic potential)을 계산한다.

그러나 분자역학에서 정전기적 상호작용에는 정전기장 그리고 분자력을 설명하기 위하여 Lennard-Jones potential을 자주 사용한다. 따라서 분자 간 에너지함수로서 비 공유결합 상호작용 에너지(E)는 다음식과 같이 탐침원자와 기질분자 내 각 원자간 상호작용 에너지들의 합으로 계산된다.

$$E = \sum E_{vdw} + \sum E_{el} + \sum E_{hb} \quad (11)$$

E_{vdw} 는 van der Waals 상호작용에 기인한 에너지이고 E_{el} 는 정전기적인 에너지이며 그리고 E_{hb} 는 수소결합의 생성에 따르는 에너지이다. 이어서 예측통계에 대한 기초를 마련하기 위하여 격자원자와 기질 화합물간의 정전기적 그리고 입체적인 상호작용 에너지와 생물활성에 대하여 계산되는 CoMFA 필드 설명인자로 선택하고 PLS(partial least square)회귀분석을 통하여 그 계수(PLS coefficient)를 등고선 지도 모양으로 나타내어 기질분자 구조상의 활성요소를 확인하고 예측한다. 최종 예측하는 3D-QSAR 방정식은 다음과 같다.

$$BA_i = const + \sum_{j=1}^n a_{ij} F_{ij} \quad (12)$$

BA_i 는 i 번째 화합물의 생물활성이고 a_{ij} 는 화합물 i의 j 번째 자장 에너지(입체 정전기) 상수이며, F_{ij} 은 화합물 i의 j번째 자장 에너지이다.

CoMFA기법의 장점은 다루고자하는 호합물의 구조가 동일하지 않아도 적용이 가능하며, 3차원 구조에 대한 입체적

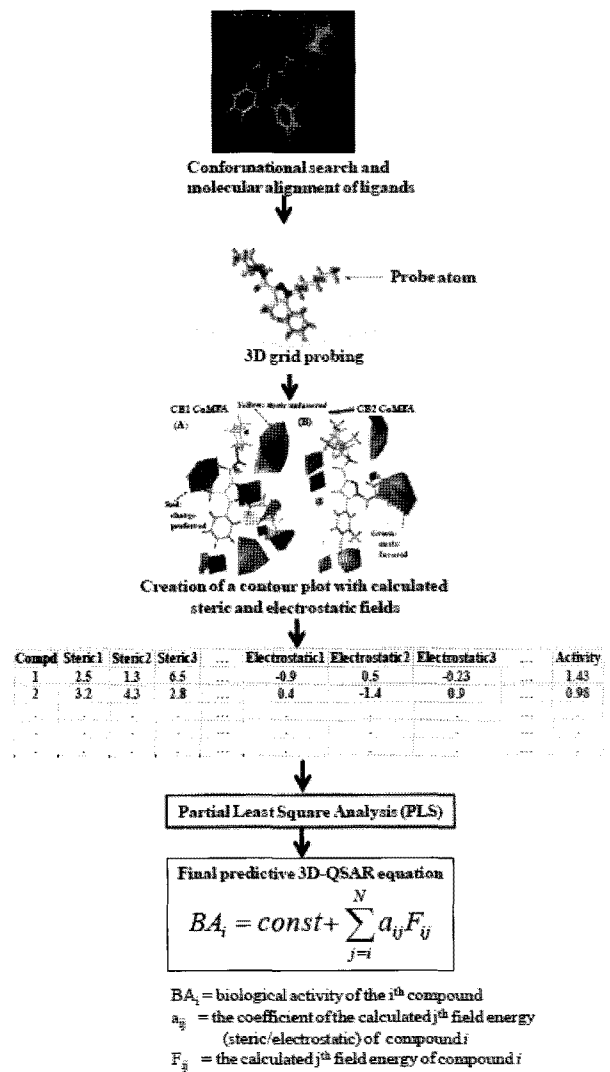


Fig. 4. A general CoMFA workflow.

정보를 영상이나 등고선 지도 형식으로 볼 수 있고 파라미터의 도입이 필요하지 않으며, chance correlation을 걱정할 필요가 없다. 반면에 단점은 운영이 복잡하며 가장 안전한 형태를 3차원 공간상에서 정렬시켜야하는 절차가 요구될 뿐만 아니라 외삽(extrapolate)에 의한 예측이 불가능하다는 점 등이다. 이와 같이 CoMFA는 실질적인 활성구조를 찾아내는 과정이 쉽지 않고 3차원 공간상에 기질 분자의 정렬에 따라 그 결과가 크게 좌우되는 단점이 있으나 분자의 공통된 특성을 입체장과 정전기장의 등고선 지도 형식의 그림으로 보여줌으로서 약물 분자 구조 또는 약물특이분자단(pharmacophore) 규명에 기여할 뿐만 아니라 CoMFA 자장 특성을 이해함으로써 새로운 물질을 예측하는데 기여할 수 있다.

Klebe 등(1994)에 의하여 유사성 지수분석(Comparative Molecular Similarity Indices Analysis, CoMSIA)으로 명명된 또 다른 3D-QSAR 방법은 격자점에 따라서 탐침원자

사용은 CoMFA와 유사하지만 추가 분자 장은 CoMSIA 접근에서 실행되었다. CoMSIA 기법은 CoMFA 기법처럼 분자들이 가장 안정한 형태를 탐색하여 3차원 공간상에 정렬시키고 입체 장 소수성장 정전기장 수소결합 주계장과 수소결합 받게장 등 이화학적 성질 비교를 통하여 기질화합물에 대한 생물활성 기여를 입체적으로 표시함으로써 기질분자의 물리화학적 성질과 생물활성사이의 상관성을 기초로 하는 분자설계 수단이다. CoMFA장 계산에서 구축된 동일한 격자를 CoMSIA장 계산에 사용되며 보통 탐침원자를 사용하여 정렬되기 전 분자들에 대하여 일정공간상의 격자점들에서 유사성지수를 계산한다. Guassi분자의 함수가 다르기 때문에 분자 내 원자들과 탐침원자 사이의 거리의존가 다르기는 Guassi분자의 함수가 사용된다. 분자 표면의 안쪽과 바깥쪽 그리고 모든 격자점에서 유사성지수들이 계산될 수 있다.

CoMSIA장에서도 CoMFA장의 경우와 마찬가지로 수용체 부위에 상응하는 기질 분자와의 상호작용 에너지를 계산하기 위하여 기질분자 주변에 탐침원자로 Methyl-기($C:sp^3$; 반경; 1.53 Å 그리고 하전; +1.0)와 같은 일방 격자점을 생성시키고 모든 격자점에서 van der Waals 입체장과 쿨롱전하 (Coulombic potential)를 위시하여 소수성장 수소결합 주계장과 받게장을 계산한다. 특히, 정전기, 입체, 소수성, 수소결합 수용체(HBA), 그리고 수소결합 주는(HBD) 성질은 가우스 거리함수(Gaussian distance function)를 사용하여 생성된다. Lennard-Jones와 Coulombic 함수 대신에 그런 Gaussian 유형 잠재함수(potential function)를 사용하여 분자 장을 계산하기 위해 격자점에 정확한 정보를 제공한다.

CoMFA는 3차원공간에서 일정한 간격으로 놓여있는 격자원자의 입체장과 정전기장 그리고 표수성장으로 표현되나 CoMFA의 입체 장은 CoMFA에서 입체장과 소수성장으로 CoMFA의 정전기장은 CoMFA에서 정전기장과 수소결합 주계장과 받게장으로 각각 표현되며, 분자의 유사성지수를 계산하기 위하여 다음식이 이용된다.

$$AF_{kq(j)} = - \sum_{j=1}^N W_{prob,k} W_{ik} e^{-ar_{iq}^2} \quad (13)$$

여기서 i 는 분자 j 의 모든 원자에 대한 지수 합이고 $W_{prob,k}$ 는 +1 전하를 가지는 반경 1 Å 소수성, 수소결합 주계와 받게의 탐침원자이며, W_{ik} 는 원자 i 의 물리화학적 성질의 실질적인 값이 k 이며, 또한 riq 는 격자점 q 에서 탐침원자의 시험분자 중 원자 i 사이의 상호간 거리이며 a 는 희석인자(attenuation factor) 이다.

CoMFA와 CoMSIA 두 방법 모두의 중요결점은 모든 분자가 정렬되어야 하고 그런 정렬이 마지막 CoMFA나 CoMSIA 모델과 예측에 영향을 미친다는 것이다. 좋은 정렬은 필요하고 그런 정렬의 질은 주관적이고, 시간이 걸릴 수 있고 CoMFA 또는 CoMSIA 모델은 때때로 재생할 수 없다. 그럼에도 불구하고 몇몇 CoMFA 또는 CoMSIA 모델은 많은 약물 디자인과 분자설계 연구를 위해 개발되고 있다.

Topomer CoMFA

Cramer 등(1988)은 기능은 같으나 골격이 다른 분자구조 (lead scaffold hopping)에 대한 최적화한 구조적 변화뿐만 아니라 생물 활성도를 최적화할 수 있는 중요한 R-그룹을 예측하는 빠른 분자분해를 기초로 하는 Topomer CoMFA라고 명명된 새로운 3D-QSAR 방법을 소개하였다. 이 방법은 치환체나 R-그룹과 같은 동정에 분자분해 소스로서 화합물라이브리리 수집에도 활용한다. CoMFA와는 달리 Topomer CoMFA 방법은 3D 리간드 이형태체(conformer)의 주관적 배열이 필요 없고 자동 배열규칙을 이용한다. Topomer는 수용체 위치 또는 다른 리간드와 관련 없이 2차원 구조를 기초로 하여 생성되었다 그 후 CoMFA 분석은 3차원격자주위의 탐침원자로 계산되는 정전기장과 입체장에서 수행되고 예측 모델을 만들기 위해 단일잔류교차확인과 함께 부분최소자승(PLS)법이 수행되었다 Topomer CoMFA는 분해를 통해 리간드의 생물 활성도를 최적화하는 가능성을 가지고 있고 선도화합물 최적화와 R-group 실질적 선별 연구에서 사용되고 있다.

Self-Organizing Molecular Field Analysis (SOMFA)

Robinson 등(1999)은 분자모양과 정전기전하를 기초로 하는 자기조직 분자 장 분석(Self-Organizing Molecular Field Analysis, SOMFA) 방법을 도입하였다. CoMFA, MS-WHIM (molecular surface-weighted holistic invariant molecular) 그리고 몇 가지 다른 스테로이드 자료와 같은 타 방법과 비교하여 SOMFA는 오차의 제곱평균제곱근(root mean square error, RMSE)인 가장 작은 표준편차를 갖는다. 요약하면, SOMFA는 SOMFA 모형이 부분최소자승(PLS) 같이 비중 있는 통계적인 요소 없이 개념적으로 간단하기 때문에 격자와 분자 정렬의 필요성 이용이라는 점에서 CoMFA와 유사하나 통계적으로 CoMFA처럼 엄격하지 않다.

정렬하지 않는 3D-QSAR 방법

지난 수십 년 동안은 정렬에 의존하지 않는 다른 3D-QSAR

방법이 소개되었다. 그 예로 분자표면특성자기상관(autocorrelation of molecular surfaces properties, AMSP), 분자 순간분석 비교(comparative molecular moment analysis, CoMMA), WHIM(weighted holistic invariant molecular), 분자표면특성에 유래한 MS-WHIM 그리고 GRIND(GRid-INdependent descriptors)를 포함한다.

Autocorrelation of Molecular Surfaces Properties (AMSP)

Wagener 등(1995)은 van der Waals 표면 및 개별 원자에 리간드의 물리적 특성을 설정하기 위한 분자표면특성자기상관(autocorrelation of molecular surfaces properties, AMSP) 방법을 각각 소개하였다. 그것은 분자 표면의 뚜렷한 점에 분자 특성의 공간 자체 상관에 기초한 3차원 설명인자를 사용한다. 분자 표면 점에는 지속적으로 표면을 가지도록 무작위로 배부되고 자기 상관계수는 각각의 거리에서 점의 다양한 쌍에 특성 값의 생성물을 합하여 얻어진다. 벡터는 분자 표면에서 특성 배분의 압축 표현이며, 요약하자면 자기상관 벡터의 이점은 단지 공간 거리가 사용되고 분자 표면의 묘사를 압축하였기 때문에 치환과 회전을 위해 불변이라고 보여 진다.

Comparative Molecular Moment Analysis (CoMMA)

Silverman 등(1996)은 4중 극자 모멘트와 같은 전하의 0, 1, 2차 공간 모멘트와 관성모멘트와 같은 질량분포를 계산하는 비교분자 순간분석(comparative molecular moment analysis, CoMMA) 방법을 도입하였다. 분자 모멘트 설명자는 분자구조에 대한 설명자와 분자전하에 대한 설명자 그리고 구조와 전하 둘 다에 관한 설명자의 3가지 다른 범주에서 분류할 수 있다. CoMFA에는 CoMMA의 대등한 통계적인 성과에도 불구하고 발표된 CoMMA 응용이 한정된 것은 에 대해 설명될 수 있는 약간 제한이 있다. 한 가지 이유는 이들 설명자의 값이 주요한 관성축, 그들의 쌍극자 모멘트가 0인 대칭 분자에 대한 무한한 동등성에 대한 쌍극자의 중심과 질량 중심 사이 치환을 측정하는 것이다.

Weighted Holistic Invariant Molecular Descriptor-Based QSAR

가중 전체 불변 분자(Weighted Holistic Invariant Molecular, WHIM) 설명자는 분자 크기, 모양, 대칭 그리고 분자 표면 점 배위의 분포 같은 3차원 분자 정보를 포함하고 있다. 분자 표면(Molecular surface, MS)-WHIM는 WHIM을 기초로

한 분자표면 특성에서 직접 유도된 3D 설명자이다.

이 표현에서 원자는 질량, van der Waals 용량, 원자의 전기 음성도, Kier와 Hall의 전기위상학적(electrotopological)지표, 원자 분극성 그리고 분자 정전기장전위(molecular electrostatic potential)로 가중 될 수 있다. WHIM접근의 중요한 이점은 3차 분자구조의 치환과 회전이 변하지 않는 3D-QSAR 설명자를 제공하는 것이다. MS-WHIM에서 분자표면 점과 관련된 특성은 통계적 매개변수를 계산하기 위하여 각기 다른 가중치 식으로 이용된다. 특히 단일 값과 분자정전기전하(molecular electrostatic potential, MEP)는 Connolly 분자표면 각 점에서 계산되고 그들은 무게로 간주된다. 단일 값은 분자 표면 모양에 관한 정보를 가지고 분자정전기전하(MEP)는 전자밀도 분포에 관한 정전기 정보를 제공한다. WHIM/MS-WHIM 설명자는 3차원 분자배향에 불변이다 그러나 두 방법 모두 다른 3D-QSAR 방법과 같이 리간드-수용체 공동결정(co-crystal) 구조가 관심사의 표적으로 알려져 있지 않을 경우에 주관적 일지도 미지의 리간드구조가 필요하다.

Grid-Independent Descriptors (GRIND)-Based QSAR

Pastor 등(2000)은 이해와 해석이 용이하고 정렬하지 않는 설명자를 제공하기 위해서 격자독립 설명자를 도입하였다. 분자상호작용장 (molecular interaction field, MIF)을 계산하기 위하여 산소탐침(카르보닐기 산소)과 N1 탐침(아미드 질소) 같은 특수한 탐침을 이용한다. 격자 각 마디에서 탐침과 표적리간드 간 에너지(E)는 다음과 같이 계산 된다.

$$E = \sum E_{es} + \sum E_{hb} + \sum E_{ij} \quad (13)$$

여기서 E_{es} 가 정전기 에너지, E_{hb} 는 수소결합에너지이고, E_{ij} 는 Lennard-Jones의 전위 에너지이다.

이 방법에서는, 정전기 상호작용, 소수성 상호작용, 수소결합 수용체와 수소결합 주체 장은 가상 수용체 부위(virtual receptor site, VRS)를 정의하는 위치의 고정을 얻는 것이라 여겨진다. VRS 지역은 그들 지역이 더 이상 3차원 공간에서 그들의 위치에 좌우되지 않기 위해 자동 또는 교차 상호상관 변형을 통해 GRIND로 암호화 한다. 바꿔 말하면 장의 자기상관 설명자는 계산되고 다만 분자상호작용 에너지의 가장 높은 생성물은 저장되고. 반면에 다른 것은 버려진다. 이 차이는 GRIND의 가역성에 있고 설명자는 ALMOND이라고 하는 다른 관련 프로그램을 사용하여 3차원 공간에서 후에 계획될 수 있다. GRIND의 통계적인 성과는 다른 방법과 비

숫하지만 장점은 정렬을 하지 않아도 되고 해석하기 쉽다. 그러나 리간드의 생물활성 구조는 VRS를 유도하는 귀중한 정보이다. 그런 정보의 제한은 다른 3D 방법과 같이 마지막 예측 모델에 영향을 미칠 수도 있다.

Multi-Dimensional (mD) QSAR Methods

다차원 QSAR(multi-dimensional QSAR, mD-QSAR) 방법은 근본적으로 3D-QSAR 방법의 연장이다. 이 방법은 Table 1에서 보는 바와 같이 3D-QSAR 방법의 결점을 보완하기 위하여 물리적 특징, 특성 또는 새로운 차원을 추가적으로 통합하는 것이다.

한 예로 QSAR 모델을 생성하는 동안 분자구조와 정렬을 시도하는 Hopfinger 등(1997)에 의한 4D-QSAR이다. 수용체 기하학이 알려진 수용체 의존(RD)방법 뿐만 아니라 수용체 독립(RI) 방법으로 방법을 사용가능하게 하고 동시에 몇 가지 CoMFA 특징을 통합하는 형태적 Boltzmann 표본추출인 4 차원을 소개한다. 4D-QSAR 방법은 정렬 문제를 해결하지 못하고 있으나 각각의 예비 정렬의 빠른 평가를 할 수 있다. 그런 4D-QSAR 실시는 XMAP 프로그램에서 찾아볼 수 있다. 최근에는 5D-와 6D-QSAR은 그것의 용매화 상태 뿐만 아니라 수용체의 다중 묘사를 위해 사용 될 수 있다. 5D-QSAR 방법에서는 Vedani 등은 유도적합가설의 다중 묘사 즉 각각의 리간드 위상기하학에 대한 수용체 결합부위 적응을 다섯 번째 차원으로 소개하였다. 바꾸어 말하면 그들은 유전 연산법을 사용하여 최적화된 유사 원자론적 수용체 대리의 조직단위를 생성하였다. 결합 에너지는 다음과 같이 산출되었다

$$E_{\text{binding}} \approx E_{\text{ligand-receptor}} - E_{\text{solvationligand}} - T\Delta S - E_{\text{Internalstrain}} - E_{\text{Inducedfit}} \quad (14)$$

Eligand-receptor는 리간드-수용체 상호작용 힘의 장 에너지이고, $E_{\text{solvationligand}}$ 는 리간드 용매화 에너지이며, TS는 수용체결합의 엔트로피리간드에 있는 변화이다. $E_{\text{internalstrain}}$ 은 수용체 결합에서 리간드 내부 에너지 변화이고, $E_{\text{induced fit}}$ 는 수용체 대리를 적응시키기 위해 필요한 에너지 활용이다.

5D-QSAR은 새로운 분자의 결합 친화성이 다른 더 낮은 차원 모형보다 더 정확하게 예측되는 것으로 결론지었다. 비교분자 장 분석(CoMFA), 비교분자 유사성 지수분석(CoMSIA), 3D-QSAR 도구인 Topomer CoMFA, 자기조직 분자 장 분석(SOMFA), 분자 순간분석 비교(COMMA), 분자표면 특성 자체상관(AMSP), 전체가중 일괄 불변분자(WHIM), 설

명자 기반 QSAR, 격자독립 설명자(GRIND)-기반 QSAR, 4D-QSAR, 5D-QSAR 그리고 6D-QSAR 방법 같은 3D-와 nD-QSAR 방법뿐만 아니라 fragment-similarity-based QSAR (FS-QSAR), fragment-based QSAR(FB-QSAR), Hologram QSAR(HQSAR) 그리고 top priority fragment QSAR와 같은 최근 fragment-based QSAR 방법을 제시한다. 분자 구조 정보는 분자 설명자에서 암호로 고쳐 쓰고 QSAR 모형은 알려지지 않은 리간드의 활성을 예측하기 위한 설명자와 알려진 리간드의 생물활성도 간의 수학적 상관관계로 정의한다. 비록 풍부한 실험적인 자료가 결합과정에서 표적 단백질의 활성역할을 강조한다 하더라도 QSAR 연구는 작은 분자 리간드의 특성에 자주 제약을 받는다.

4D-QSAR

3D-QSAR에서는 예측 모델을 유도하기 위해서는 특히 표적 단백질의 구조상 정보가 없을 경우 리간드 분자의 생물활성 구조와 상대적 배향을 모두 확인하여야 한다. 고분자 표적의 3차원 구조가 알려졌을 경우 예를 들면 도킹(docking)을 이용하여 본연의 목적을 위해 이용될 수 있다. 구조와 배향의 확실한 확인이 매우 어렵기 때문에 확실히 리간드 구조와 배향에 기초를 두고 있는 3 D-QSAR 모델은 생물활성 으로부터 벗어나 목적을 예측하기 위해 어떤 사용도 거의 도움이 되지 않는 경우가 있다. 정렬독립적인 설명자(alignment-independent descriptors, AID)에 기초한 QSAR은 명쾌하게 정렬 문제점을 해결하기 위하여 제시되고 있으며, 분자 표면에서 예상되는 소수성 성질과 같은 화합물의 3차원적 특성은 화합물의 물리화학적장의 순간 확장과 같은 위치 독립적인 특성으로 변형된다. 설명자에 기초를 두고 있는 QSAR 모델의 질을 위해서는 구조형태의 선택이 중요하다. 4D-QSAR 개념은 분자와 공간 다양성을 포함하고 다른 구조, 배향, 호반체(tautomers), 입체 이성체 또는 양자화상태의 각 분자를 나타내는 것으로 정렬 문제점에 접근한다. 진정한 결합형태 또는 생물활성 구조는 QSAR 개념의 기초를 이루는 연산법(algorithm)에 의해 확인된다.

일반적으로 4차원 개념의 2가지 다른 유형이 개발되었으며, QSAR의 한 종류는 구조적으로 보통 1,000 또는 그 이상의 유사한 형태의 큰 집단을 사용한다. CoMFA와 같은 3D-QSAR에서 리간드에서 더욱 작은 형태적 변화는 리간드탐침 상호작용과 QSAR 시뮬레이션의 결과에 따라서 깊은 충격이 있을 수 있다. 전반적인 결합 형태의 주위에 구조적공간을 표본 추출 하는 것은 선택한 리간드 배열에 QSAR 결과의 선명한

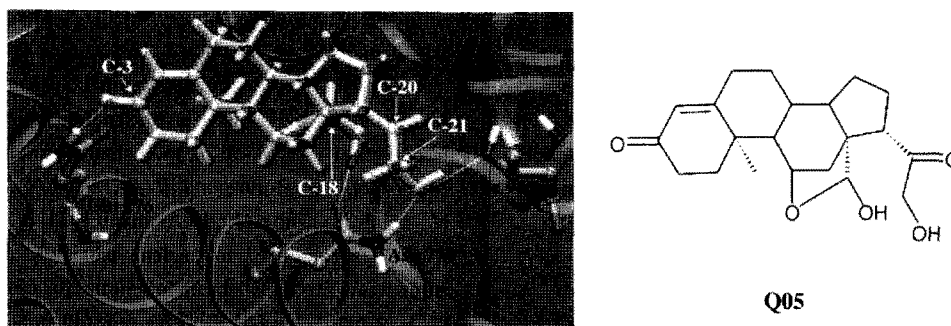


Fig. 5. The comparison of the in silico docked aldosterone (Q05) with the experimentally determined structure (Bledsoe et al., 2005) displays an rms deviation of 0.36 Å.

도움을 감소시킨다. 그러나 이 4D-QSAR 접근 유형은 화합물의 전반적인 결합 형태에서 불확실하여 취급 능력이 없어 보인다. 의약 화학연구에서는 같은 종류가 아닌 일련의 화합물 예를 들면 핵 수용체 단백질에 관한 구조적 정보가 쓸데없는 예를 들면 GPCRs(G protein-coupled receptor) 프로젝트에서 또는 약제의 독성 검사와 잡다한 단백질 예를 들면 시토크롬 P450 효소를 포함하는 환경화학제 한 분자의 결합형태의 확인은 거의 모두 정확하다.

4D-QSAR 접근의 두 번째 종류는 다양한 리간드 배열의 작은 세트는 QSAR 모델을 만들기 위해 충분한 평균평방근편차와 대체 정렬규약 상으로 중요한 결합 형태는 독자적인 대안을 제안하고 있다. 단일 시뮬레이션 내에서 구조적 대안의 신속한 생성은 필요한 QSAR 시뮬레이션의 격증 때문에 모든 종래의 3D-QSAR 의 모든 리간드의 교체 결합형태의 일련의 역시 검사 이상으로 확실한 장점이 있다.

전체적인 효과는 다른 구조, 방향, 호변체, 입체 이성체 및 양자화상태를 포함할 수 있다. 근본적인 연산법(algorithm)은 결합 대리 또는 그것들의 조합에 대한 가장 높은 상호작용 에너지를 가진 상대적 배치를 선택한다(Boltzmann 가중 전체효과). 그것들의 실제생물활성은 근본적인 최적화 연산에 의해 확인하는 것으로 생물활성 이형태체와 리간드정렬의 선택과 관련된 편향을 감소시킬 수 있다. 특히 4D-QSAR는 X-ray 결정학조차 가끔 명백하게 양자화상태를 결정할 수 없기 때문에 가장 가능한 호변이성의 모양 확인에서 중요한 역할을 할 수 있다. 4D-QSAR은 3D-QSAR의 실현가능한 연장으로 정렬 과정 중 불확실성을 설명할 수 있다.

4D-QSAR 기법은 단 하나의 시뮬레이션에서 다른 리간드 형상에 대하여 명쾌하게 설명할 수 있다. 최근에는 이것은 성공적으로 자동화된 도킹 절차에 따라 확인한 평균 4개의 다른 결합 자세로 각기 작은 분자를 나타내는 시토크롬 P450 3A4에 구조상으로 다양한 화합물의 결합을 가장하기 위하여

적용하고 있다(Fig. 5).

5D-QSAR

CADD(computer aided drug design) 개념에서 단백질 유연성에 대하여 설명할 필요성은 요즘에는 최첨단이라고 생각하는 작은 분자 리간드가 도킹하지만 결합부위의 유연성을 허용하는 유연한 단백질이 도킹하는 새로운 구조를 기초로 하는 방법이 최근에 강조되고 있다. QSAR의 분야에 대한 이 원리의 적용은 아직 초기단계이다. 명백한 방법으로 유도적응을 가상하는 것은 각 리간드 분자에 대한 결합위치 표면모델의 위상적인 적응이 유도되었다는 것이다. 여기에 결합 위치 모델의 표면은 그것에 결합하는 리간드의 크기와 기하학에 따라 경미하게 줄어들거나 확장할 수 있다. 유도적합의 정확한 크기와 메커니즘의 확인으로 진짜 표적 단백질의 구조 없이는 불가능하고 예를 들면 크기는 입체, 정전기, 수소결합, 또는 친지질성 잠재력에 달려있는 다른 유도적응 규약이 QSAR에 대체 각분(5D-QSAR)으로 주어진다.

에너지 벌칙은 유도적합의 크기에 비례해서 적용된다. 주계 또는 받게 같은 수소 결합 속성은 수용체 모델에서 튕겨져 떨어지고 부근에 있는 리간드 분자의 수소 결합 성질에 따른다. 화합물은 유도적응의 중요성으로 결합위치의 다른 지역 아래(sub-pockets)에 결합하고 따라서 단백질의 다른 상호작용 장을 경험할 수도 있다. 2개의 공간에 분리된 전자각은 유도적응 단백질 운동에 기인하는 다른 지역아래의 다른 물리화화적인 본질을 나타내는 안과 외각 사이 특성의 분포에서 핵 수용체에 있는 호전성(agonism)과 길항작용에 포함된 그들과 같이 큰 구조 변화 뿐 아니라 국부적으로 단백질 적응의 시뮬레이션을 모두 허용한다. 이것은 작용제와 길항제의 단지 작용제상태의 연구에서 보다 더 큰 구조변화가 예상될 수 있는 상황의 동시 시뮬레이션을 가능하게 한다. 그런 연구의

필요조건은 결합 형태에서 중요한 오버랩과 리간드 가운데 길항작용의 일관성 있는 메커니즘의 존재이다. 예를 들면 neuroaminidase에 대한 풍부한 구조상 데이터는 결합 지역을 분획 내는 아미노산의 수소결합 속성은 리간드를 가진 분자 내 또는 분자사이의 수소 결합에 관련된 것이다. 결과적으로 결합 위치의 상호작용 장(interacting field) 그리고 기하학(topology)은 모두 각 개개의 리간드에 대해 변화할 수 있다. 결합 위치 대응에서 이것은 리간드의 기하학에 입체 조정과 리간드와 수용체 모델사이 인력 또는 반발작용 때문에 대리에 관한 물리 화학적인 성질을 변경하는 성분을 결합해서 달성될 수 있다.

6D-QSAR

Vedani A 등(2005)은 에스트로겐 수용체에 결합된 106개의 다양한 분자에서 얻은 결과($q^2 = 0.903$; $p^2 = 0.885$)는 우리의 접근이 약물과 화학물질과 관계있는 내분비선 혼란의 잠재력의 확인을 위해 적당하다는 것을 제외하고 약물과 화학약품으로 촉발되는 역효과의 정량화를 위한 결합 형태와 6D-QSAR (QSAR 소프트웨어 사용)의 확인을 위해 유연한 도킹(Yeti 소프트웨어 사용)과 가상실험에서의 *in silico* 시뮬레이션에 대한 개념을 제시하였다.

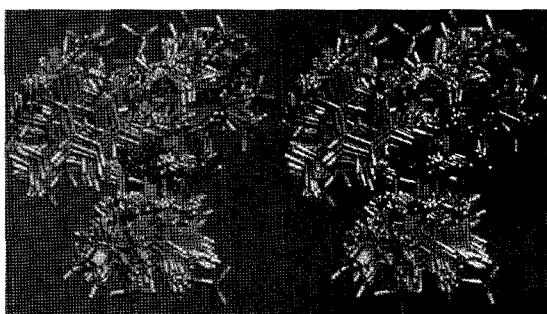


Fig. 6. Stereoview of the superimposed 48 ligands as obtained from flexible docking to the hMR.



Fig. 7. Stereoview of the binding of tetrahydrogestrinone (THG, "The Clear") to the hMR(human mineralocorticoid receptor).

리간드는 다른 유도적합모델(5D-QSAR)의 종합 효과뿐만 아니라 구조, 방향, 양자화상태 및 거울상체(4D-QSAR)의 종합적 효과와 동시에 용매화 모델(6D-QSAR)을 고려하여 으로 표현된다(Fig. 7).

이것은 용매 접근 가능한 지역(pocket) 위치 그리고 크기가 유전 산법으로 최적화한 표면부분이 용해성질로 나타내어질 때 명확하게 이루어질 수 있다. 리간드 탈 용매화와 용제 휘발성분제거(solvent stripping)를 포함하는 용매화 기간은 또한 대항종류 내의 각 다른 모델과 관계없이 암묵적 접근의 결합지역의 다양한 용제접근성을 반영하여 표준화할 수 있다. Quasar에서는, 결합 에너지는 방정식에 따라 산출한다.

$$E_{binding} = E_{ligand-receptor} - E_{ligand desolvation} - E_{ligand strain} - T\Delta S - E_{induced fit}$$

$$E_{binding} = E_{ligand-receptor} - E_{ligand desolvation} - E_{ligand strain} - T\Delta S - E_{induced fit} \quad (15)$$

다차원 QSAR 소프트웨어 Raptor의 특징은 근본적으로 다른 평가기록계산 기능(scoring function)과 결합지역(binding pocket)의 위상 기하학 뿐만 아니라 부분적인 특성까지도 바꾸는 유도적합 메커니즘에 따라 두 전자각에 의한 수용체 결합 부위를 표현한다. 평가기록계산 기능은 리간드결합에 관한 소수성에 대한 기간(GHphob), 위상 기하학적인 적응비용(GIF) 그리고 엔트로피변화(TS) 뿐만 아니라 수소 결합에 대한 지향성기간(GHbond)을 포함하는 리간드-단백질 상호작용을 평가하기 위해 Raptor에 의해 사용하였다.

$$E_{binding} = E_{ligand-receptor} - T\Delta S - E_{induced fit}$$

$$E_{ligand-receptor} = E_{hydrogen bonding (shell-1)} + E_{hydrophobic (shell-1)} + E_{hydrogen bonding(shell-2)} + E_{hydrophobic(shell-2)} \quad (16)$$

2D-QSAR와 3D-QSAR

일반적으로 2D-QSAR 방법은 대표적인 치환기가 치환된 한 부류의 화합물을 합성하여 생리작용을 측정된 다음 각 화합물에 대한 생물활성과 치환기들에 관한 설명인자들과의 관계를 회귀분석방법으로 계산하여 합성한 전체 화합물에 대하여 상관성이 높은 QSAR식을 얻는다. 이 식으로부터 여러 설명인자인 독립변수들의 종류와 계수의 가중요인의 크기를 감안하여 치환기와 생리활성사이의 상관관계에 미치는 경향을 파악한다.

3D-QSAR 방법은 먼저 기질분자의 생물활성과 물리화학

적인 성질에 관한 자료를 만들고 안정한 형태를 결정하여 3차원공간에서 겹침 법칙에 따라 정렬한다. 설명인자를 계산하고 통계 분석하여 QSAR 모델을 구축한 다음에 결과를 해석하고 확인하며 예측하는 일련의 과정을 거친다. 따라서 3D-QSAR은 2D-QSAR 방법에서 가장 취약한 점으로 인정되는 구조적 해석상의 문제점을 크게 보완하였다는 것이다. 일반적으로 3D-QSAR 방법이 예측하는 품질은 분자 정렬 또는 중첩의 품질과 같은 몇몇 요인 및 리간드 생물활성 구조상의 정보에 달려 있다. 특히 분자 중첩은 주관적이고 대응 수용체 리간드 복합물에 구조상 정보가 없을 때 리간드 생물활성 구조는 항상 불분명하게 남아 있다.

종래의 CoMFA 결과는 모델이 다양하고 주관적일 수 있는 분자의 정렬의 방향에 달려 있기 때문에 수시로 재생할 수 없고 비록 여러 가지 개선된 방법과 다른 절차가 3D-QSAR 방법의 중요한 제한 즉 주관적인 분자 정렬과 활성 구조 문제를 극복하기 위하여 각종 개량된 방법이 소개되더라도 대다수는 아직도 수동간섭과 중첩이 필요하다. 앞으로도 2차원 분획에 기초를 둔 QSAR 방법은 분획에 기초하거나 2D-QSAR 방법이 간단하고 활기차고 또 주관적이고 시간이 걸리는 분자 정렬이 필요하지 않으며 추정하는 결합구조 또는 3차원 구조의 결정 때문에 다차원 QSAR 방법 이상의 장점이 있다. 그러나 단점은 Hansch-Fujita 방법과 같은 2D-QSAR 방법의 몇 가지는 유일하지 않은 다양한 해법이 제시되고 총체적으로 예측하는 질적 요구가 컴퓨터로 더 복잡하고 지나치게 요구하는 몇 가지 다차원 방법처럼 좋지 않을 수도 있다.

고 찰

QSAR는 이미 존재하거나 가상분자의 결합 친화성 또는 독성가능성과 같은 양을 추정하기 위하여 실질적 모형을 만드는 컴퓨터 연구의 분야이다. 비록 풍부한 실험적인 자료가 결합과정에서 표적 단백질의 활성역할을 강조한다 하더라도 QSAR 연구는 작은 분자 리간드의 특성에 좌우된다. 분자의 기하학적 구조를 최적화시키고 에너지 계산을 통하여 가장 안정한 구조를 예측할 수 있을 뿐만 아니라, 분자의 여러 가지 성질을 규명할 수 있다. 분자의 최적화된 구조 및 에너지 계산을 하는 이론적 방법에는 크게 분자역학적 방법과 양자역학적 방법이 있다.

분자역학(Molecular mechanics)은 하나의 작은 분자의 성질은 보다 복잡한 분자의 이와 유사한 작용기의 성질과 비슷하다는 기본적인 가정 하에 이론을 전개하는 것으로서 실험을 통해서 알 수 있는 많은 수치와 기존의 알려진 결합길

이, 결합각 등을 새로운 분자에 적용하는 계산 방법이다. 여기에 쓰이는 결합길이, 결합각 그리고 여기에 더하여 여러 가지 종류의 진동운동의 힘 상수 등을 통칭하여 파라미터라고 하고 기존의 알려진 결합길이, 결합각 등을 사용해서 분자의 에너지를 계산하며, 파라미터들로부터 에너지를 실지로 계산하는데, 필요한 식들과 파라미터를 모두 통칭하여 역장 (force field)라고 한다.

분자역학의 가장 근본적인 가정 중 하나는 분자 전체의 에너지는 여러 가지 부분의 향으로 나누어 이들의 합으로 표시될 수 있다는 것이다. 다시 말하면, 분자 전체의 에너지는 결합 신축운동(bond stretching) 에너지, 변각운동(angle bending) 에너지, 뒤틀림각 상호작용(torsional interaction) 에너지, 정전기 상호작용(electrostatic interaction) 에너지, 반데르발스 상호작용(van der Waals interaction) 에너지 등의 합으로 표시된다. 분자 궤도 (Molecular Orbital, MO)계산을 통해 양자역학(Quantum mechanics)의 기본법칙을 수학적 변환과 근사적 방법으로 사용하여 원자 및 분자계의 특성을 예측하는 방법이다. 일반적으로 QSAR에서는 분자의 화학적인 성질이나 특징을 나타내는 표현자(descriptor)를 이용하여 분자구조를 표현한다. 이러한 양자오인들을 생체역학의 Or연구에서 다변량분석방법(Multivariate Analysis Methods in Biomechanistic Activity Studies, MAMBAs) 등의 변수들 간의 상관관계를 고려하여 서로 유사한 변수들 끼리 묶어주는 방법 요인분석과정을 거쳐 모델화한다. 그러나 정량적 구조 활성 관계(QSARs)의 개발을 위한 공식적인 지침서가 없기 때문에 오류의 위험이 있을 수 있어 검증과정을 거칠 필요가 있다. 이 때 분자구조를 평면으로 간주하는지 혹은 입체적인 구조로 간주하여 연구하는가에 따라 표현자(descriptor)를 1D, 2D, 3D 그리고 4D, 5D, 6D의 입력정보로 표현한다.

QSAR의 기본인 2D-QSAR 알고리즘은 대부분 통계적 접근 방법(MLR, PCA&PCR, PLS 등)을 이용하여 예측모형을 만들어 분석 예측하고 있으나 최근에는 기계학습 기반의 알고리즘(ANN, SVM, GA, GPM 등)들을 이용한 QSAR 모델에 대한 연구로써, 기존의 통계기반 알고리즘 보다 더 효과적으로 고성능의 예측모형을 만들 수 있다는 연구결과들이 보고되고 있다.

이렇게 만들어진 모델은 장단점이 있기 때문에 연구자는 그 방법을 모델링하기 위한 적합한 방법을 선택해야 하고 때로는 인간의 선입견뿐만 아니라 3D-QSAR에 대한 입력으로 정확한 리간드 정렬의 중요성을 고려하면 어쩌면 이 단계와 관련되어 다차원 접근이 신약개발 목적을 위해 고전적인 3D-QSAR에 유망한 대안일 수도 있다.

Hansch, Fujita, Free와 Wilson에 의해 근대적인 QSAR 기법이 알려졌으며, 국내에서는 1984년 성(2002)이 처음으로 농약학계에 소개하여 재초활성은 발아 전보다 발아 후에 전자 밀도에 의한 전자효과와 입체효과에 의존적이라는 것을 밝혔고 계속해서 CoMFA(Cramer 등, 1988), CoMSIA(Klebe 등, 1994) 그리고 HQSAR(Heritage 등 1999)등 3D-QSAR기법에 의한 연구(성, 2003)가 보고 되고 receptor-based approach 방식이 활성화되고 있으며, Molecular docking 연구가 많이 이루어지고 있다.

성 등(2003 a, b)은 새로운 물질탐색에 활용하기 위하여 살균제는 FRAC(Fungicide Resistance Action Committee), 살충제는 IRAC(Insecticide Resistance Action Committee) 그리고 제초제는 HRAC(Herbicide Resistance Action Committee)의 작용기작 분류 기준에 따라 시판 농업용 살균제 133품목, 살충제 152품목 그리고 제초제 245품목을 선정하여 구조를

확인하였다. 이들 약제들의 작용기작 분류를 보면 제초제는 제2광합성저해제가 49성분으로 가장 많았으며 ALS저해제 42성분 그리고 cell division 저해제 21성분으로 이들 세 유형의 제초제가 거의 반을 차지하고 있다. 살균제는 ergosterol 생합성저해제 demethylation inhibitor(DMI)가 32, fungal respiration 중 complex III 저해제가 8, multi site activity 저해제가 13 그리고 작용이 규명되지 않은 품목이 11성분 등이었다. 그리고 살충제는 AChE저해제가 90성분 chitin biosynthesis 저해제가 10성분 등이었다. 한 성분 당 소수성, 쌍극자능률, 최고점유분자궤도(HOMO)에너지와 최저비점유분자궤도합수(LUMO) 에너지, 원자궤절, 분극율, van der Waals분자표면적, van der Waals 분자부피, 분자량(molecular weight) 그리고 수화에너지 등 10종의 물리화학 파라미터를 hyperchem 프로그램으로 계산하였다. 물리화학적 파라미터는 물론 분포 범위 값을 변별력을 높이기 위하여 임의로 85%를 기준으로

Table 2. Range (85%) of physicochemical parameters for calculated according to herbicides, PS-II and ALS inhibitor

| Parameters(unit) | Herbicides | PS-II | ALS |
|--------------------------------|--------------|--------------|---------------|
| OLog P | -0.90~4.50 | -0.10~3.60 | -1.40~2.00 |
| CLog P | -0.80~3.3.7 | -0.52~3.37 | -0.40~2.20 |
| DM (Debye) | 1.80~12.22 | 2.20~7.70 | 5.00~17.60 |
| MR (cm ³ /mol) | 53.0~104.0 | 57.0~90.0 | 80.0~106.0 |
| Pol (Å ³) | 19.0~37.0 | 21.0~32.0 | 28.0~40.0 |
| HOMO (eV) | -8.98~-7.34 | -9.90~-8.54 | -9.58~-7.50 |
| LUMO (eV) | -2.76~0.40 | -0.80~0.60 | -3.10~-0.82 |
| Vol. (cm ³) | 558.0~995.0 | 595.0~856.0 | 727.0~1042.0 |
| Mass (amu) | 202.0~430.0 | 213.0~340.0 | 289.0~487.0 |
| Surface area (Å ²) | 194.0~356.0 | 208.0~300.0 | 255.0~370.0 |
| Hy.E (Kcal/mol) | -10.16~114.7 | -6.96~111.16 | -15.51~113.13 |

Table 3. The range (85%) of physicochemical parameters for fungicides and sterol biosynthesis inhibitors (DMIs) and for insecticides and AChE inhibitors

| Parameter(unit) | Fungicide | DMI | Insecticide | AChE |
|--------------------------------|----------------|----------------|----------------|---------------|
| OLog P | -1.10~3.80 | -0.20~2.50 | 0.00~5.50 | -0.03~4.40 |
| CLog P | -1.09~3.75 | -0.09~2.51 | -0.26~4.34 | -0.03~3.21 |
| DM (Debye) | 1.60~7.60 | 2.10~5.50 | 1.20~10.20 | 2.01~10.20 |
| MR (cm ³ /mol) | 60.0~107.0 | 82.0~104.9 | 48.0~121.0 | 50.0~97.0 |
| Pol (Å ³) | 22.0~40.0 | 28.0~37.0 | 17.0~43.0 | 18.0~34.0 |
| HOMO (eV) | -10.330~-6.820 | -10.070~-7.360 | -10.180~-7.170 | -9.960~-7.580 |
| LUMO (eV) | -2.650~0.180 | -1.460~0.020 | -2.930~0.100 | -3.020~-0.340 |
| Vol. (cm ³) | 564.0~990.0 | 760.0~952.0 | 550.0~1086.0 | 572.0~955.0 |
| Mass (amu) | 210~412 | 292~377 | 183~477 | 193~368 |
| Surface area (Å ²) | 203.0~358.0 | 263.0~337.0 | 190.0~384.0 | 202.0~341.0 |
| Hy.E (Kcal/mol) | -14.03~122.64 | -9.04~56.74 | -11.71~110.73 | -11.71~109.70 |

자료화한 내용은 Table 1과 Table 2와 같다. 이들 제조제 살균제 살충제별로 설정한 특정한 물리화학적 파라미터들의 범위 값은 다양한 화합물들이 작물보호제로서 제조제 살균제 살충제로 활용가능성을 예측하는데 유용할 것으로 생각된다.

또한 화학공간(chemical space)이란 용어는 신약개발에서 널리 주목 받고 있다. 화학 공간은 일반적으로 작은 유기 분자들로 제한되고 모든 중요한 화합물 전체를 수집하는 것으로 정의할 수 있다. 그 범위는 무려 10^{18} 에서 10^{180} 분자 사이라고 추정되고 있다. Fig. 8에서 보는 바와 같이 화학 공간의

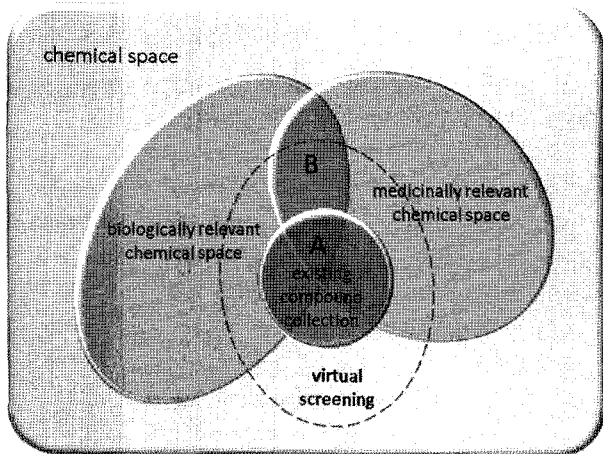


Fig. 8. Regions of biologically and medicinally relevant chemical space within the continuum of chemical space.

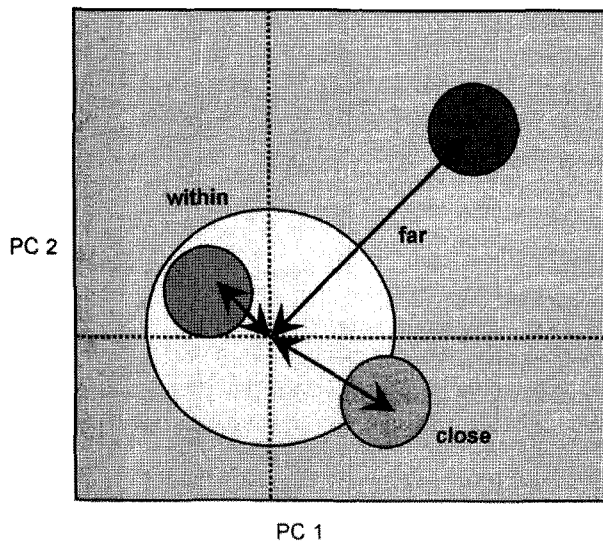


Fig. 9. A graphical illustration of the domain of applicability in principal component (PC) space. The QSAR model training set is represented by the yellow circle. Query molecules are coloured as follows: within the training space (green), close to model space (orange) and far (red). Query compounds predicted further from training model space would be expected to be less reliably predicted.

연속체 내에 생물과 약 관련 화학 공간 중첩 부분에서 단지 화학 공간의 일부가 약 (A)을 발견으로 이끌 기존의 화합물 수집에 의해 샘플링 되었다. 가상 선별지역이 약제가 발견될 (B) 가능성이 있는 새로운 곳으로 탐색되지 않은 화학 공간으로 확장시키는 기회가 있는 유일한 곳이다. 따라서 QSAR을 이용하여 화학공간에서의 유용한 화합물을 발견하는 연구가 기대된다.

문제는 신규화합물(NCEs)의 개발가능 특성의 고려사항은 신약을 개발하면서 점점 중요하게 여기게 되었다. 신약을 개발 과정에서 훨씬 이전에 신규화합물을 선별하기위해 생체, 시험관 그리고 컴퓨터에 의한 가상실험 방법에 대해 수요증가를 이어지는 ADMET(흡수, 분포, 물질 대사, 배설물 그리고 독성)에 대한 차이를 이유로 약의 60%가 실패한다는 사실이다. 따라서 정량적 구조 활성 관계(QSARs)는 생체시험이나 실험실적시험보다 처리량이 많기 때문에 빠르고 값싼 대안으로 나타내기 때문에 화합물설계와 진행 과정에서 중요한 구성요소가 되었다.

또한 QSAR은 기본적으로 컴퓨터 수단을 사용하여 알려진 활성을 가진 분자의 집합으로 부터 결정된 수학 방정식이다. 구조와 활동 사이 관계의 정확한 모양은 다양한 통계적인 방법을 사용하여 결정될 수 있고 계산된 분자 설명자와 이 방정식은 새로운 분자의 활성을 예측하려 할 때 이용된다. 따라서 활성영역의 영역은 모델을 구축하는 데 사용된 화합물이 얼마나 유사한 기반으로 특정 분자의 예측에 불확실성을 예측할 수 있는지가 정량 구조 활동 관계(QSAR)에서 중요한 개념이다(Fig. 9).

약물의 전임상적 독성중점, 전임상적 부작용 그리고 물질 대사 예측에 대한 가상실험에서의 컴퓨터 프로그래밍기법을 적용 사용하는 것은 과학계와 일반대중에게도 높은 관심을 사고 있다.

결론

지금까지 QSAR을 유형별로 한 분자의 생물학적 효과를 설명하기 위하여 용해도나 해리상수 단 하나의 이화학적 성질에 기초를 두는 1D-QSAR에서부터 2D-QSAR, 3D-QSAR 그리고 4D, 5D, 6D-QSAR 기법들을 검토하였다. 역시 3D, 4D, 5D 및 6D 모델이 최선의 예측을 만든다.

모든 QSAR 방법에는 장단점이 있기 때문에, 연구자는 그 방법을 모델링하기 위한 적합한 방법을 선택해야 한다. 때때로 중요한 인간의 선입견뿐만 아니라 3D-QSAR에 대한 입력으로 정확한 리간드 정렬의 중요성을 고려하면 어쩌면 이

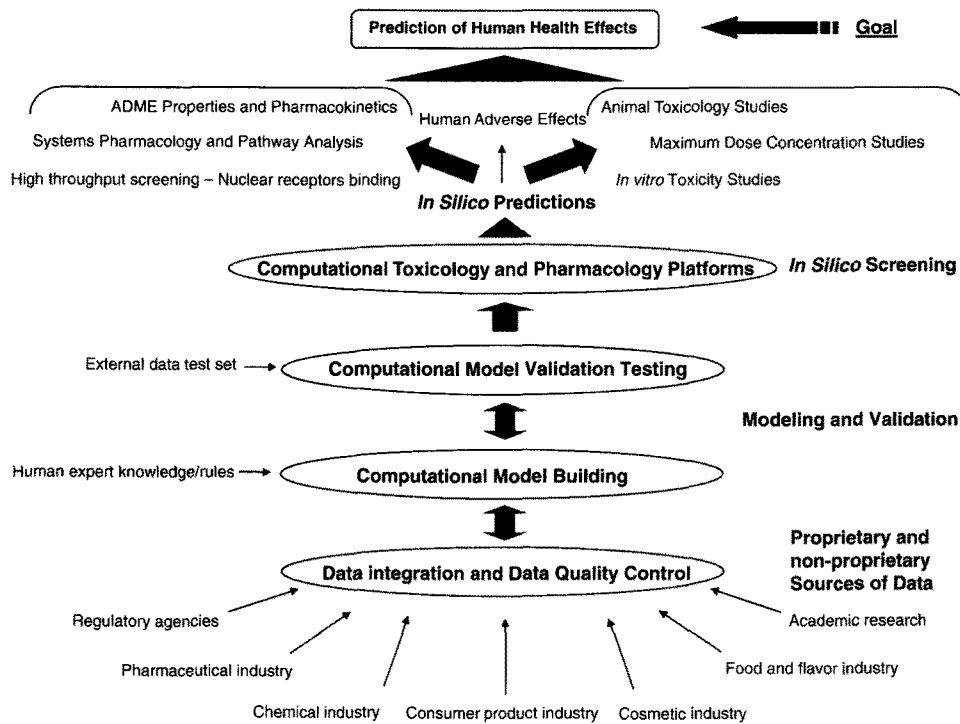


Fig. 10. The path of work flow for the use of drug and chemical toxicity databases and models; starting from the source of data to the goal of predicting human health effects.

단계와 관련 되어 mQSAR 접근이 약 발견 목적을 위해 고전적인 3D-QSAR에 유망한 대안을 제공한다.

mQSAR는 고분자 구조에 도킹하는 리간드와 협력하여 양을 정하는 결합 친화성을 위해 가장 적합 된다는 것을 보일 것이다. 전형적인 3D-QSAR의 유용한 확장은 다차원(multidimensional QSAR, mQSAR) 방법이 새로운 화합물질의 친화력을 예측하는데 신뢰도를 높일 것 같다. mQSAR이 자유도의 비율을 생물학적 데이터·포인트(data taints)의 수로 늘릴 수 있으므로 그것은 QSAR에 이들 추가 차원을 신중하게 포함하는 것이 절대 필요하다. 자유도의 숫자를 증가시키는 것이 mQSAR의 필연적인 결과는 아니다. 예를 들면, 3D 격자(예를 들면 CoMFA에서)상에서 결합부위집합에서 표면집합으로 움직이는 것으로써 이 숫자를 감소시킬 수 있다.

따라서 국제적으로 제약 및 화학 산업분야에 신뢰할 수 있는 동물실험 대체검증시험 방법으로 권장되어 사용되고 있고 또한 실제 시험자료가 없는 독성분야 연구, 구조적으로 유사한 물질그룹의 관리, 기존 화합물질의 평가 우선순위 설정 등에 유용한 도구로써 활용 가능할 것으로 기대되고 있어 이해의 폭이 넓어지기를 기대한다.

> 인 / 용 / 문 / 헌

성낙도 (2002), 정량적인 구조-활성상관(QSAR) 기법에 의한 새로운 농약의 개발 I. 기본 개념과 QSAR 기법의 유형, 농약과학회지, 6(3), 166~174.

성낙도 (2002), 정량적인 구조-활성상관(QSAR) 기법에 의한 새로운 농약의 개발 II. 자유에너지 직선관계(LFER)와 설명인자들, 농약과학회지 6(4), 231~243.

성낙도 (2003) 정량적인 구조-활성상관 (QSAR) 기법에 의한 새로운 농약의 개발. III. 3D QSAR 기법들과 컴퓨터를 이용한 분자 설계(CAMD), 농약과학회지 7(1), 1~11.

성낙도 (2003) 정량적인 구조-활성상관(QSAR) 기법에 의한 새로운 농약의 개발 -IV. 국내의 연구 동향과 전망, Journal of the Korean Society for Applied Biological Chemistry, 46(3), pp. 155~164.

성낙도, 송선섭 (2003a), 제초제의 활성성분에 대한 물리화학 파라미터의 범위, 한국농약과학회지 7(1), 58~65.

성낙도, 송선섭 (2003b) 작물보호제로서 살균제와 살충제의 활성성분에 대한 물리-화학 파라미터의 범위, J. Korean Soc. for Agric. Chem. Biotechnol. 46(4), 280~284.

Akinori Hirashim, Eiichi Kuwano, Morifusa Eto (2003), Comparative receptor surface analysis of octopaminergic antagonists for the locust neuronal octopamine receptor, Computational Biology and Chemistry 27:531~540.

Andrade, C.; Salum, L.B.; Castilho, M.; Pasqualoto, K.; Ferreira, E.; Andricopulo, A. (2008), Fragment-based and classical

- quantitative structure-activity relationships for a series of hydrazides as antituberculosis agents. *Mol. Divers.* 12, 47~59.
- Casalegno M.; Sello G. Benfenati E. (2006), Top-Priority Fragment QSAR Approach in Predicting Pesticide Aquatic Toxicity. *Chem. Res. Toxicol.* 19, 1533~1539.
- Chunyan Zhao, Elena Boriani, Antonio Chana, Alessandra Roncagliani, Emilio Benfenati (2008), A new hybrid system of QSAR models for predicting bioconcentration factors (BCF), *Chemosphere* 73:1701~1707.
- Cramer, R.; Patterson, D.; Bunce, J. 1988, Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins. *J. Am. Chem. Soc.* 110, 5959~5967.
- D. J. de Ridder, L. Villacorte, A.R.D. Verliefde, J.Q.J.C. Verberk, S.G.J. Heijman, G.L. Amy, J.C. van Dijk (2010), Modeling equilibrium adsorption of organic micropollutants onto activated carbon, *Water Research* 44:3077~3086
- Frank E. Blaney, Dorica Naylor, and John Woods (1993), MAMBAs: A real-time graphics environment for QSAR, *J. Mol. Graphics* 11: September 157~165.
- Hopfinger, A.J.; Wang, S.; Tokarski, J.S.; Jin, B.; Albuquerque, M.; Madhav, P.J.; Duraiswami, C. (1997), Construction of 3D-QSAR models using the 4D-QSAR analysis formalism. *J. Am. Chem. Soc.* 119, 10509~10524.
- Joop Hermens (1995), Prediction of environmental toxicity based on structure-activity relationships using mechanistic information, *The Science of the Total Environment* 171: 235~242.
- Klebe, G.; Abraham, U.; Mietzner, T. (1994), Molecular similarity indices in a comparative analysis (CoMSIA) of drug molecules to correlate and predict their biological activity. *J. Med. Chem.* 37, 4130~4146.
- Kyaw Zeyar Myint, Xiang-Qun Xie, (2010) Recent Advances in Fragment-Based QSAR and Multi-Dimensional QSAR Methods, *Int. J. Mol. Sci.*, 11:3846~3866.
- Linnan He, Peter C. Jurs (2005), Assessing the reliability of a QSAR model's predictions, *Journal of Molecular Graphics and Modelling* 23:503~523.
- Luis G. Valerio Jr. (2009), In silico toxicology for the pharmaceutical sciences, *Toxicology and Applied Pharmacology* 241:356~370.
- Mark T.D. Cronin, T. Wayne Schultz (2003), Pitfalls in QSAR, *Journal of Molecular Structure (Theochem)* 622:39~51.
- Markus A. Lill (2007), Multi-dimensional QSAR in drug discovery, *Drug Discovery Today* 12(23/24), pp. 1013~1017.
- Markus Boehm (2011), Virtual Screening of Chemical Space: From Generic Compound Collections to Tailored Screening Libraries, pp. 5.
- Matthew S. Sigman, Jeremie J. Miller (2009), Examination of the Role of Taft-Type Steric Parameters in Asymmetric Catalysis, *J. Org. Chem.* 74:7633~7643.
- Nicholas E. Jewell, David B. Turner, Peter Willett, Graham J. Sexton (2001), Automatic generation of alignments for 3D QSAR analyses, *Journal of Molecular Graphics and Modelling* 20:111~121.
- Ourania Peristera, Morena Spreafico, Martin Smiesko, Beat Ernst, Angelo Vedani (2009), Mixed-model QSAR at the human mineralocorticoid receptor: Predicting binding mode and affinity of anabolic steroids, *Toxicology Letters* 189:219~24.
- Ourania Peristera, Morena Spreafico, Martin Smiesko, Beat Ernst, Angelo Vedani. (2009) Mixed-model QSAR at the human mineralocorticoid receptor: Predicting binding mode and affinity of anabolic steroids, *Toxicology Letters* 189, 219~224.
- Pastor, M.; Cruciani, G.; McLay, I.; Pickett, S.; Clementi, S. (2000), GRid-INdependent descriptors (GRIND): A novel class of alignment-independent three-dimensional molecular descriptors. *J. Med. Chem.* 43, 3233~3243.
- Robinson, D.D.; Wimm, P.J.; Lyne, P.D.; Richards, W.G. (1999), Self-organizing molecular field analysis: A tool for structure-activity studies. *J. Med. Chem.* 42, 573~583.
- Shane Weaver, M. Paul Gleeson (2008), The importance of the domain of applicability in QSAR modeling, *Journal of Molecular Graphics and Modelling* 26:1315~1326.
- Shunlai Li, Yan Zheng (2006), Self-Organizing Molecular Field Analysis on a New Series of COX-2 Selective Inhibitors: 1,5-Diarylimidazoles, *Int. J. Mol. Sci.* 7, 220~229.
- Silverman, B.D.; Platt, D.E. (1996), Comparative molecular moment analysis (CoMMA): 3D-QSAR without molecular superposition. *J. Med. Chem.* 39, 2129~2140.
- Supratik Kar, Kunal Roy (2010), QSAR modeling of toxicity of diverse organic chemicals to *Daphnia magna* using 2D and 3D descriptors, *Journal of Hazardous Materials* 177 : 344~351.
- Vivek VYAS, Anurekha JAIN, Avijeet JAIN, Arun GUPTA (2008), Virtual Screening: A Fast Tool for Drug Design, *Sci Pharm.* 76:333~360.
- Wagener, M.; Sadowski, J.; Gasteiger, J. (1995), Autocorrelation of molecular surface properties for modeling corticosteroid binding globulin and cytosolic ah receptor activity by neural networks. *J. Am. Chem. Soc.* 117, 7769~7775.
- Yovani Marrero Ponce (2004), Total and local (atom and atom type) molecular quadratic indices: significance interpretation, comparison to other molecular descriptors, and QSPR/QSAR applications *Bioorganic & Medicinal Chemistry* 12:6351~6369.
- Zhokhova, N.; Baskin, I.; Palyulin, V.; Zefirov, A.; Zefirov, N. (2007), Fragmental descriptors with labeled atoms and their application in QSAR/QSPR studies. *Doklady Chem.* 417, 282~284.