

## 정량적인 구조-활성상관(QSAR) 기법에 의한 새로운 농약의 개발. III. 3D QSAR 기법들과 컴퓨터를 이용한 분자설계(CAMD)

### 성 낙 도

충남대학교 농업생명과학대학 응용생물화학부

**요약** : 새로운 농약을 탐색하고 개발하는데 있어서 고효율 유기합성(HTOS) 기술과 고효율 검색(HTS) 기술 등의 발전과 더불어 컴퓨터 화학을 이용한 분자설계(CAMD) 방법으로 보편화되고 있는 비교 분자장 분석(CoMFA)과 비교 분자 유사성 지수분석(CoMSIA) 등, 3D QSAR 기법들을 위시하여 분자 홀로그램 구조-활성관계(HQSAR) 분석방법 등, QSAR 기법들을 요약하고 그 활용 사례들을 간략하게 소개하였다.(2003년 1월 16일 접수, 2003년 3월 14일 수리)

**Key words** : Computer-assisted molecular design (CAMD), comparative molecular field analysis (CoMFA), comparative molecular similarity indices analysis (CoMSIA), hologram quantitative structure-activity relationship (HQSAR), 3D QSAR.

### 서 론

근래 새로운 농약의 개발에 관한 연구는 post-genome 시대의 도래와 더불어 컴퓨터 화학을 이용한 분자설계(CAMD; Computer-assisted molecular design 또는 CADD; Computer-assisted drug design) 기술(Borher 등, 1997)과 조합화학(combinatorial chemistry)을 이용한 고효율 유기합성(HTOS; high throughput organic synthesis) 기술(Gallop 등, 1994) 및 합성한 물질의 생물활성을 효율적으로 검색하는 고효율 검색(HTS; high throughput screening) 기술(Wexler 등, 1996) 등의 발전에 편승하여 고도로 향상되고 있다.

그 주체로서 QSAR 기법(Hansch 등, 1995)은 새로운 약제의 개발에 소요되는 오랜 기간과 천문학적인 비용을 획기적으로 절감할 수 있게 하였으나 생체계의 복잡성과 불확실성을 극복하고 해결하여야 하는 수많은 난제들이 아직도 상존하고 있다. 그리고 CADD 분야는 수용체와 기질 분자가 최적화되는 과정에서 리간드와 수용체에 대한 정보의 양에 의존적이므로 각각의 정보에 따라 합리적인 접근방식(Eto, 1992)으로 분자를 설계하는 데에는 ligand-based approach 방법과 receptor-based approach 방법(Reddy 등, 1999) 등의 2가

지 접근 방식이 있다.

본 총설(III)에서는 컴퓨터 화학을 이용한 분자설계 방법으로서 ligand-based approach 방법인 CoMFA (Cramer 등, 1988a)와 CoMSIA(Klebe 등, 1994) 기법을 위시한 3D QSAR 기법과 HQSAR (Heritage 등, 1999) 기법에 관한 분자 상호 작용장(interaction field)에 관한 (kubinyi, 1993b) 개괄적인 개념과 활용상 절차를 소개하고 그에 따른 몇 가지 저자의 연구 사례와 QSAR에 관련한 다변량 분석(multivariate analysis) (Stahle 등, 1988)에 대하여 간략하게 요약하였다.

### 3D QSAR 기법의 발전과정

소위 classical QSAR 즉, 2D QSAR 기법(Kubinyi, 1993a)에서는 분자들의 물리-화학적 성질인 제한된 설명 인자들과 생리적인 반응 사이의 관계식을 유도하였으나, 컴퓨터의 발전으로 분자의 3차원적인 구조를 토대로 한 3D-QSAR 기법이 발전하였다(Waterbeemd, 1992). 기질분자와 수용체와의 반응에서 3D QSAR 기법을 활용하는 데에는 기질분자와 수용체에 관련한 각각의 정보에 따라 2가지의 합리적인 접근 방법으로서의 분자설계(rational molecular design) 방식이 있다. 첫째로 ligand-based approach 방법은 기질 분자에 관련한 정보만을 알고 있음으로서 2D 및 3D QSAR 분석, 작용점의 가정과 자료(data base) 탐색 및 분자 유

\*연락처

Table 1. Development process of three dimensional (3D) QSAR methodology

Year	Authors	Comments
1979	Wise <i>et al.</i>	DYLOMMS (Dynamic lattice oriented molecular modeling system): Dynamic lattice model
1988	Cramer <i>et al.</i>	CoMFA (Comparative molecular field analysis): Steric, hydrophobic & electrostatic field.
1994	Klebe <i>et al.</i>	CoMSIA (Comparative molecular similarity analysis) : Steric, hydrophobic, electrostatic field. H-bond acceptor & H-bond doner.
1996	Silverman <i>et al.</i>	CoMMA (Comparative molecular moment analysis): Utilizes moment of the molecular mass and charge distributions
1997	Hopfinger <i>et al.</i>	4D QSAR: Incorporates conformational and alignment. Grid cell (spatial) occupancy measures of atoms as descriptors.
2002	Vedani <i>et al.</i>	5D QSAR: Multiple representation of induced-fit hypotheses.

사성 등을 탐색하는 방법이고 둘째로, receptor-based approach 방법(Reddy 등, 1999)은 앞의 경우와 반대로 수용체의 구조와 성질 등에 대하여만 알고 있을 경우에 시도하는 방법으로서 수용체-기질간 상호작용, 수용체에 기초한 3D database 탐색, dynamics, docking 및 lead design 등을 통하여 기질분자를 설계하는 방법이다. 그러나 어떤 방법으로 분자 설계가 이루어지더라도 CADD 방법으로는 극복할 수 없어서 설계 상 제한적인 수밖에 없는 요소(그림 1)들로는 합성적이나 경제적인 가능성, 분자 성질의 영역과 다양성 및 생물학적인 유효성(Waterbeemd, 1996) 등이 있다.

이와 같이 컴퓨터를 이용한 약제설계(CADD) 방법

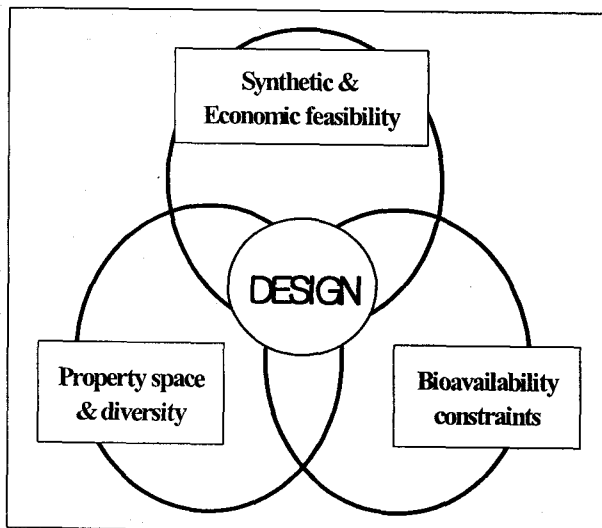


Fig. 1. Limitations in molecular design. A compromise must be made between synthetic, economic, molecular diversity and bioavailability aspects.

(Borher 등, 1997)과 분자설계(CAMD) 또는 분자 모델링(CMC, 1996)이라 함은 축적된 자료에 기초하여 원하는 물성을 가지는 분자를 경험적으로 유추하는 과정을 의미한다. 그리고 이를 활용하는 목적에 따라 단백질이나 용매효과 등을 다루기 위한 분자 역학적(molecular mechanics) 방법과 분자의 물리-화학적 성질을 다루고자 할 경우에 이용하는 양자 역학적인(quantum mechanics) 방법이 있다. 분자 역학적 방법에는 분자 동역학적 방법, 통계 역학적 방법 및 형태 탐색방법 등이 있고 양자 역학적인 방법에는 반 경험적 방법과 이론적인(*Ab initio*) 방법 등이 있다.

컴퓨터 화학을 이용한 분자 설계기법이 발전함에 따른 3D QSAR 기법들의 발전과정을 표 1에 정리하였다. 1979년에는 DYLOMMS(Wise 등, 1983)에 기초하여 dynamic lattice를 모델로 하는 개념이 알려졌으며 1988년에는 CoMFA(Cramer 등, 1988)에 관한 논리가 발표되었다.

이어서 CoMSIA(Klebe 등, 1994) 및 CoMMA(Silverman 등, 1996) 등을 위시하여 1997년에는 4D QSAR 기법(Hopfinger 등, 1997) 에컨데, 유연한 생물 조절제를 표현하는 평준화된 수용체 모델의 용도와 약리 작용점의 가정 등, 3D QSAR 개념이 지닌 2가지 한계점을 극복하기 위한 방법 그리고 local induced fit에 대한 고려는 물론, 양성자화 된 상태와 지향성 및 형태의 집합체에 의하여 리간드 분자를 묘사하기 위한 2가지 목적으로 4D QSAR 개념이 발전되었으며 이후, 리간드에 대한 복식 표현 개념의 5D QSAR 기법(Vedani 등, 2002) 등이 알려지게 되었다. 이들 3D QSAR 기법들 중에서 CoMFA기법에 대하여 알아보자.

**비교 분자장 분석 (CoMFA)**

3차원(3D)적인 분자 구조를 토대로 한 3D-QSAR 기법 중, CoMFA 기법(Cramer 등, 1988a)은 QSAR과 분자설계에 있어서 보편화 된 유용한 도구로써 입체장(steric field)과 정전기장(electrostatic field)에 대한 등고도(contour map)를 통하여 기질 분자의 물리 화학적 성질과 생물활성 사이의 상관성을 연구하는 방법이다.

CoMFA 분석을 실행하기 위해서는 먼저, 기질 분자의 구조식들을 그려 넣고 분자 내 원자 사이의 반발력이 최소화되어 전체 에너지가 제일 낮은 가장 안정한 형태를 가진 분자를 simulated annealing 방법(Keer 등, 1994) 및 GA(Genetic algorithm) 방법(David 등, 1994) 등으로 최적화한다. 회전 가능한 모든 결합을 일정한 각도씩 회전함에 따른 계통적인 탐색을 재 실시하여 기질 분자의 최저 에너지 형태를 계산한다. 그리고 분자내 모든 원자의 부분 전하들을 Gasteiger-Marsili 방법(Gasteiger 등, 1980)을 사용하여 계산하고 자료중의 분자와 분자의 기본 골격으로 구성된 주형(template) 분자 사이의 반응점 요소들을 최소자승 접합에 따르는 점침 법칙에 따라 3차원 공간상에 atom fit 또는 flexible fit 등의 방식으로 정렬(alignment)시킨다. 정렬된 분자들을 대개 2Å 공간을 가지는 3차 원 입방 격자 중에 놓은 다음에 그림 2와 같이 화합물들의 형태를 포함하는 열(row)과 성질을 나타내는 행(columns)으로 표시되는 QSAR table을 작성한다.

수용체 부위에 상응하는 기질 분자와의 상호작용 에너지를 계산하기 위하여 기질분자 주변에 methyl-기

(C:  $sp^3$ ; 반경; 1.53Å 및 하전; +1.0)와 같은 입방 격자점을 생성시키고 입체장(van der Waals)과 정전기장(Coulombic potential)을 계산한다. 그러나 분자역학에서 정전기적 상호작용에는 coulombic potential 그리고 분산력(dispersion)을 설명하기 위하여 Lennard-Jones potential을 자주 사용한다. 따라서 분자간 에너지 합수로서 비공유 결합 상호작용 에너지(E)는 다음 식과 같이 탐침원자(probe atom)와 기질 분자내 각 원자간 상호작용 에너지들의 합으로 계산된다. 여기에서  $E_{vdw}$ 는 van der Waals 상호작용들에 기인한 에너지이고  $E_{el}$ 은 정전기적인 에너지 그리고  $E_{hb}$ 는 수소결합의 생성에 따르는 에너지이다.

$$E = \sum E_{vdw} + \sum E_{el} + \sum E_{hb}$$

이어서 예측 통계모델에 대한 기초를 마련하기 위하여 격자 원자와 기질 화합물간의 정전기적 및 입체적인 상호작용 에너지와 생물활성에 대하여 계산되는 CoMFA field를 설명인자(Tong 등, 1998)로 선택하고 PLS (partial least squares) 회귀분석(Wold 등, 1984)을 통하여 그 계수(PLS coefficient)를 등고도(contour map) 모양으로 나타내어 기질분자 구조상의 활성요소를 확인하고 예측한다.

가장 작은 예측의 표준 오차에 상응하는 주성분(principale component)의 적정수는 leave-one-out (LOO) 교차확인 과정으로 결정되며 이 과정에서 각 화합물은 제거되기도 하고 나머지 화합물들로부터 설정된 CoMFA 모델에 의하여 예측되기도 한다. 또한, 주성분의 적정수를 사용하여 3D QSAR 관계를 추출하기 위해서는 3D 공간상에 정렬된 기질 분자들의 생물활성 값에 대하여 PLS라는 통계분석(Stahle 등, 1988)을 실시한다.

PLS 분석에는 교차확인 방법과 bootstrapping 분석 방법이 있다(Cramer 등, 1988b). 교차확인 방법은 분석 결과의 예측 타당성을 가늠할 수 있게 하는 방법으로 모델로부터 화합물을 하나씩 제거하여 모델을 재 설정하고 target property를 예측하는 재설정과 예측 순환 과정을 모든 성질 값들이 정확하게 예측될 때까지 계속한다. 그리고 bootstrapping 방법은 원래의 자료 설정을 참 집단으로 가정하고 그것으로부터 새로운 자료를 무작위로 N번 발생함으로써 각각을 통계 분석하여 모델의 안정성(stability)를 예측하는 방법이다. 따라서 RMS (root mean square) 오류인 PR-

ESS (predictive residual sum of squares)는 실측된 것과 예측된 독립변수에 관한 제곱편차의 합이며 SSD (sum

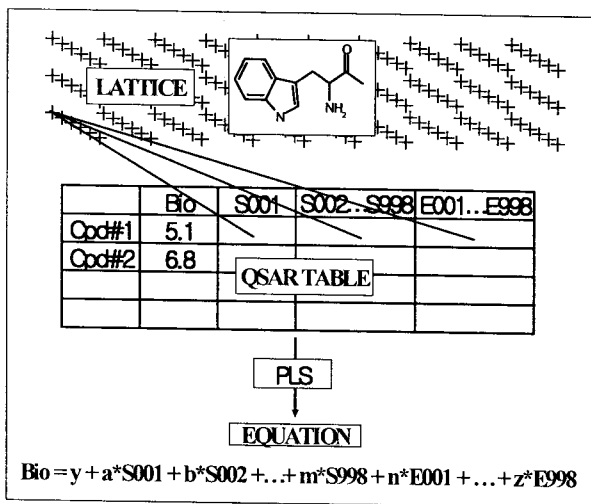


Fig. 2. Process of comparative field analysis(CoMFA) as 3D QSAR.

Table 2. Advantages and disadvantages of CoMFA method<sup>a)</sup>

Advantages	Disadvantages
May be applied to a mixed series.	More complicated to run.
No parameter needs to be estimated.	Requires a superposition rule.
Includes 3D structural information.	Difficult to extrapolate into unexplored region.
The results can be displayed graphically.	Most suitable for the study of drug-receptor interactions.
Can describe steric effects well.	Requires weighting or scaling of some parameters.
Can describe H-bonding effects well.	Less useful information is obtainable from the coefficients of variables in the correlation equation, etc.
Does not need to worry about chance correlation.	
Does not need to assume which parts of the molecule might contribute to the potency, etc.	

<sup>a)</sup>Kim, 1993.

of the squared deviation)는 모든 독립변수의 평균값으로부터 각 독립변수의 제곱 편차의 합을 뜻 하는 모델의 평가 기준이 된다. 여기에서 잠재 변수인 적정한 조성의 수는 가장 작은 표준 편차와 가장 높은 LOO와 no validation으로 자료를 분석하고 교차 확인된  $q^2$  값으로 결과의 신빙성을 예측하며  $q^2$  값은  $q^2=1-\text{PRESS}/\text{SSD}$ 의 관계에 있다.

즉, 예측 정도는 결정계수,  $r^2$  값과 교차확인,  $q^2$  값으로 각각 확인하는데  $r^2$  값은 모델의 적합 정도를 그리고  $q^2$  값(predictive correlation coefficient)은 모델의 예측 능력의 척도이다. 교차확인 시험은 signal과 noise를 구분하며 자료에서 제외시켜야 하는 화합물(outlier)에 대한 정보에 따라 outlier를 제거 한 다음에 다시 모델을 설정하고  $q^2$  값으로 결과의 신빙성을 예측한다. 특히,  $r^2$  값이 1에 가까운 값을 나타내지 않을 경우에는 분자를 재 정렬시키거나 모델을 재 설정한다. 따라서  $r^2$ 는 내부 예측성이고  $q^2$ 은 외부 예측성이다. 그러므로  $r^2$  값은 높운데  $q^2$  값이 낮은 경우에 모델은 안정성(stability)이 낮고 외부 예측성도 역시 낮다고 말하며 다시 그에 따른 교차확인 시험으로 outlier를 선정하고 최적의 잠재변수인 조성(component)의 수를 결정한다.

일반적으로  $q^2 > 0.5$  그리고  $r^2 > 0.9$ 일 경우에 상관성과 예측성이 양호한 모델로 간주된다(Cramer 등, 1988a). 이들  $q^2$  값과  $r^2$  값으로부터 통계 결과가 양호한 수준이라고 판단되면 최종 모델은 최적값의 조성수로 재 설정하고 no validation test를 수행하여 설명인자와 구조간의 적절한 관계식을 유도한다. 그리고 그 결과, 얻어지는 입체장(steric field), 정전기장(electrostatic field) 및 소수성장(hydrophobic field) 등, 3

가지 물리-화학적 성질에 대한 CoMFA field의 PLS 계수를 등고도로 나타낸다. CoMFA 등고도를 통하여 새로운 화합물을 설계하고 대상으로 하는 기질 화합물의 예측값(prediction)을 계산한다. 마지막으로 모델이 실험값과 일치하지 않고 편차(error)가 큰 경우에는 모델을 수정하거나 새로운 모델을 설정하여 만족할 만한 결과를 얻을 때까지 전술한 바와 같은 과정을 반복한다.

이해를 돕기 위하여 표 2에는 CoMFA기법의 장점과 단점에 대하여 간략히 정리하였다(Kim, 1993). 이에 따르면 장점으로는 다루고자 하는 화합물의 구조가 동일하지 않아도 적용이 가능하며 3차원 구조에 대한 입체적인 정보를 영상이나 등고도 형식으로 볼 수 있고 파라미터의 도입이 요구되지 않을 뿐 더러 chance correlation을 걱정할 필요가 없다. 반면에 단점으로는 운영이 복잡하며 가장 안정한 형태를 3차원 공간상에 정렬시켜야 하는 절차가 요구될 뿐만 아니라, 외삽(extrapolate)에 의한 예측이 불가능하다는 점 등이다.

이와 같이 CoMFA는 실질적인 활성 구조를 찾아내는 과정이 쉽지가 않고 3차원 공간상에 기질 분자의 정렬에 따라 그 결과가 크게 좌우되는 단점이 있으나 분자의 공통된 특성을 입체장과 정 전기장의 등고도 형식의 그림으로 보여줌으로써 pharmacophore를 규명하는데 도움을 줄 뿐 아니라, CoMFA field의 특성을 이해함으로써 새로운 물질을 예측하는데 큰 도움을 얻을 수 있다.

저자에 의하여 식물성 병원균에 대한 살균활성을 CoMFA기법으로 분석한 사례로는 다음과 같은 보고들이 있다. 벼 도열병균(*Pyricularia oryzae*)과 토마토 역병균(*Phytophthora infestans*)에 대한 2-thienyl 및 2-furyl 치환 bis-aromatic  $\alpha,\beta$ -불포화 케톤 유도체들

의 살균활성에 관한 CoMFA 분석 결과(성 등, 1998), 두 균주는 모두 입체성 효과가 정 전기적인 효과보다 큰 영향을 미쳤으며  $\beta$ -탄소의 양하전이 두 균주를 선택적으로 작용하는 반응점 임을 알았다. 또한, FPTase (farnesyl protein transferase) 저해활성에 대하여는 비 치환체가 가장 강한 활성을 보이는 공통점을 나타내었으며 atom fit 방법에 의한 정렬시에 교차 확인 값과 상관성이 가장 좋았고 정 전기장과 입체장이 비슷한 정도로 기여하는 경향이었다(Yu 등, 2002).

감수성 및 저항성 고추역병균(*Phytophthora capsici*)에 대한 phenylisoxazole 유도체들의 살균 활성에 관한 CoMFA 분석결과, 교차 확인값과 상관성이 매우 양호하였으며 두 균주에 대한 선택성은 benzyl-기의 *ortho*, *meta*-위치에 입체적으로 큰 치환체 그리고 *meta*, *para*-위치에는 음하전이 생성되는 조건이었다(성 등, 2002b). 그리고 벼(*Oryza sativa* L.)와 논피(*Echinochloa crus-galli*)에 대한 제초활성에 관한 CoMFA 분석 사례로는 cyclic imide (CyI)계 광역동성 제초제로서 *N*-치환 phenyl-3,4,5,6-tetrahydrophthalimide 유도체들(성 등, 2003a)과 quinolinc carboxylic acid형 제초제로서 3-phenyl-5-(3,7-dichloro-8-quinolinyl)-1,2,4-oxadiazole 유도체들이 있다(성 등, 2003b).

### 비교분자 유사성 지수분석 (CoMSIA)

CoMSIA 기법(Klebe 등, 1994)은 CoMFA 기법에서와 같이 분자들의 가장 안정한 형태를 탐색하여 3차원 공간상에 정렬시키고 입체장, 소수성장, 정 전기장, 수소결합 주계장 및 수소결합 받게장 등, 물리-화학적 성질 장의 비교를 통하여 기질 화합물에 대한 생물활성의 기여를 입체적으로 표시함으로써 기질 분자의 물리-화학적 성질과 생리활성 사이의 상관성을 토대로 하는 분자설계 수단이다. 그러므로 CoMFA장 계산에 구축된 동일한 grid를 CoMSIA장 계산에 사용되며 보통 탐침(probe) 원자를 사용하여 정렬되기 전의 분자들에 대하여 일정한 공간상의 격자점들(grid points)에서 유사성 지수들이 계산된다.

분자내 원자들과 탐침 원자 사이의 거리 의존성에 대하여는 Gaussian 함수가 사용된다. 왜냐하면, Gaussian 함수가 다르기 때문에 분자 표면의 안쪽과 밖쪽 그리고 모든 격자점에서 유사성 지수들이 계산될 수 있다. 또한, CoMSIA는 격자(lattice) 내 중첩된 분자들의 배향 변화에는 민감하지 않고 CoMSIA에 의하

여 얻어진 상관 결과들은 그래프 형태의 지도로 해석될 수 있다(Klebe, 1998). 또한, CoMSIA 방법은 5가지 서로 다른 성질 장들 예컨대 입체장, 소수성장, 정전기장, 수소결합 주계장 및 수소결합 받게장 등이 서로 다른 물리-화학적인 성질들에 초점이 맞추어져 있음에도 그 장들은 각각에 대하여 고도로 상호 의존적이며 소수성(Ghose 등, 1999)은 매개 변수화에 따라 표기된다.

CoMSIA에서도 CoMFA의 경우와 같이 수용체 부위에 상응하는 기질 분자와의 상호작용 에너지를 계산하기 위하여 기질 분자 주변에 탐침 원자로서 methyl-기(C:  $sp^3$  반경; 1.53Å 및 소수성, 수소결합 주계와 받게 성질에 관한 하전; +1.0)와 같은 입방 격자점을 생성시키고 모든 격자점에서 입체장(van der Waals)과 정 전기장(Coulombic potential)을 위치하여 소수성장, 수소결합 주계장과 받게장을 각각 계산한다. 따라서 CoMFA는 3차원 공간에 일정한 간격으로 놓여있는 격자 원자의 입체장과 정 전기장 및 소수성장으로 표현되나 CoMFA의 입체장은 CoMSIA에서 입체장과 소수성장으로 그리고 CoMFA의 정 전기장은 CoMSIA에서 정전기장과 수소결합 주계와 받게로 각각 표현되며 분자의 유사성 지수를 계산하기 위하여 다음 식이 이용(Malinowski 등, 1980)된다.

$$AF_{i,kq}(j) = -\sum_{k=1}^N W_{\text{prob},k} W_{ik} e^{-\alpha r_{ik}^2}$$

즉, data set 분자,  $j$  중의 모든 원자들,  $i$ 가 격자점,  $q$ 에서 격자의 교차점에 대칭적으로 놓여진 탐침 원자와 분자사이의 유사성 지수,  $AF_{i,k}$ 는 위식으로 계산된다. 여기에서  $i$ 는 분자,  $j$ 의 모든 원자에 대한 지수 합이고  $W_{\text{prob},k}$ 는 +1의 전하를 가지는 반경 1Å, 소수성, 수소결합 주계, 수소결합 받게의 탐침 원자이다. 그리고  $W_{ik}$ 는 원자,  $i$ 의 물리-화학적 성질의 실질적인 값,  $k$ 이며 또한,  $r_{ik}$ 는 격자점,  $q$ 에서 탐침 원자와 시험 분자 중 원자,  $i$  사이의 상호간 거리이며  $\alpha$ 는 attenuation factor이다.

결국, 생물활성과 조합된 성질 장의 계산 결과들은 PLS table에 포함되어 PLS 방법들은 CoMSIA 모델을 유도하는데 적용되며 적절한 조성을 선택하고 모델의 통계적인 의미를 확인하기 위하여 leave-one-out-(LOO) 교차확인 방법으로 PLS 논리에 따른다(Cramer 등, 1988b). 그리고 CoMSIA 결과들은 성질 장을 사용한 장(field)에 관한 기여정도를 나타낸 지도 형태(contour

plot)로 해석되어 도식적으로 표기된다. 그러므로 CoMSIA 분석 결과들에 대한 모델의 통계적인 평가는 CoMFA에 대하여 기술된 바와 같은 방식으로 마무리 된다.

저자에 의한 CoMSIA 분석 사례로는 metalaxyl 살균제 저항성과 감수성 고추역병균(*Phytophthora capsici*) 들에 대한 2-*N*-benzyl-5-phenoxy-3-isothiazolone 유도체의 살균활성에 관한 CoMSIA와 HQSAR 분석에서 교차 확인값( $q^2$ )과 상관성( $r^2$ )은 대략 비슷한 결과를 보였다. 비교 분자장은 정 전기장과 소수성장이 동일한 비율로 가장 큰 영향을 미쳤고 입체장의 영향은 아주 작은 편이었다. 그러나 수소-결합 받개는 7:1의 비율로 저항성 군주가 감수성 군주에 비하여 매우 크게 영향을 미치는 것으로 나타났다. CoMSIA 등고도 분석결과, 두 군주에 대한 선택성 요소는 소수성장과 수소-결합 받개이었으나 두 분자장의 기여도가 1:2의 비율로 소수성장의 기여도가 크므로 phenyl 고리상 meta-위치에 도입되는 소수 성장이 선택성 요소로 판단되었다(성 등, 2002a).

벼(*Oryza sativa* L.)와 논피(*Echinochloa crus-galli*)의 줄기와 뿌리에 대한 부위 별 제초활성에 관한 HQSAR 분석 사례로는 quinclorac계 제초제로서 3-phenyl-5-(3,7-dichloro-8-quinolinyl)-1,2,4-oxadiazole 유도체들이 있다(성 등, 2003b). 이 경우에는 HQSAR 모델이 CoMFA 모델보다 양호한 상관성과 예측성을 보였으며 논피에 대한 선택성 조건은 입체적으로 큰 치환기로서

phenyl 고리상에 양하전을 생성하는 전자 끌개가 치환되어야 할 것으로 설명되었다. 그러므로 합성된 바 없는 새로운 화합물로서 2,6-dichloro 및 2,4, 6-trichloro-치환체 등은 두 초종에 대하여 선택성과 고활성이 예측되는 화합물로 설계되었다.

### 분자 홀로그램 QSAR

분자 홀로그램 (H) QSAR(Lewis, 1997) 기법은 3D 구조의 결정과 분자정렬 및 추정되는 결합형태 등에 대한 필요성이 배제된 상태에서 2D 화합물의 구조적 특징들과 생물활성 사이의 예측관계를 PLS 기법으로 분석하는 방법으로 자료중 분자들은 일련의 유일한 구조적인 조각(bin)으로 나뉘어져 분자 홀로그램(hologram) 형태로 정돈된다(Heritage 등, 1999). 따라서 분자 홀로그램의 bin들은 분자의 조성과 위상학적인 정보를 함축하고 있는 구조적인 변수이다. 분자 홀로그램은 두벌의 2D 지문으로부터 비롯된 분자조각(fragment) 수를 포함하는 정수의 선형 배열이다. 설명 인자로서 파라미터는 홀로그램의 길이( $L$ )와 분자 조각의 크기( $M$ : 최소 및  $N$ : 최대)이고 분자 조각을 구분하기 위하여 원자, 결합, 수소원자, 원자의 혼성화 척도인 연결성 및 키랄성 등이 사용된다. 표 3에는 HQSAR 기법에서 활용되는 분자 조각에 관한 파라미터들의 정의를 요약하였다(Tong 등, 1998).

그림 3에는 분자 홀로그램의 생성을 설명하는 모식도를 나타내었다. 즉, 분자들은 몇 개의 분자 조각으

Table 3. Definition of fragment parameters in HQSAR methodology<sup>a)</sup>

Parameters	Definitions
Atoms	The <i>atoms</i> parameter enables fragments to be distinguished based on elemental atom types.
Bonds	The <i>bonds</i> parameter enables fragments to be distinguished based orders; for example, in the absence of hydrogen, allowing butane to be distinguished from 2-butane.
Connections	The <i>connections</i> parameter provides a measure of atomic hybridization states within fragment; that is, <i>connections</i> causes HQSAR to keep track of how many connections are made to constituent atoms and the bond order of those connections.
Hydrogens	By default, HQSAR ignores the hydrogen atoms during fragment generation. The <i>hydrogens</i> parameter overrides this behavior.
Chirality	The <i>chirality</i> parameter enables fragments to be distinguished based on atomic and bond stereochemistry. Thus, stereochemistry allows <i>cis</i> double bond to be distinguished from their <i>trans</i> counterparts, and <i>R</i> -enantiomers to be distinguished from <i>S</i> -enantiomers at all chiral centers.

<sup>a)</sup>Tong et al., 1998.

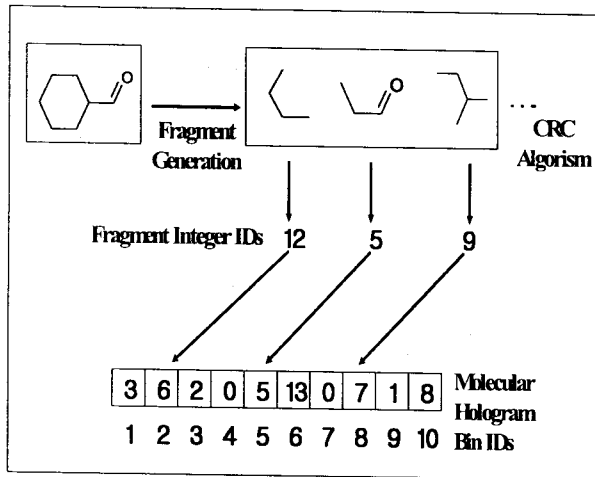


Fig. 3. Illustration of schematic generation of a molecular hologram.

로 나누어지고 일정한 길이의 분자 홀로그램 형태로 점유되는 bin은 기호 구성과 형태적인 분자에 대한 정보를 기호화한 구조적인 설명 인자들이 된다. 여기에서 data set의 구조에 대한 분자 홀로그램의 계산은  $R \times L$  차원의 행렬( $X$ )로 주어지며  $R$ 은 data set내 화합물의 수이고  $L$ 은 홀로그램 지문 중 bin들의 수를 조절하는 분자 홀로그램의 길이이다. 분자의 독특한 조각들은 CRC (cyclic redundancy check) 연산에 의하여 양의 정수로 표시된다. 또한, QSAR 목적으로 생물활성,  $Y$ 의 매트릭스가 만들어지며 PLS 분석은 직교설명 변수들을 확인하기 위하여 적용되는 것으로 고유  $L$  변수들의 선형 조합이다. 그리고 LOO cross-validation 과정은 적절한 예측 모델을 따르는 주 조성 (principal component)의 수를 결정하기 위하여 적용되며 (Cramer 등, 1988b)  $q^2$ 는 이 방법에 기초한 예측된 값이고 (Malinowski 등, 1980)  $r^2$ 은 관측된 활성값과 계산된 관측값 사이의 Pearson 상관계수이다. 적절한 모델은 각 화합물의 상응하는 생물활성에 대한 분자 홀로그램 bin값과 연관된 PLS 분석 결과식으로 확인되며 생성된 QSAR 모델은 다음 식으로 나타낸다

$$\text{Activity}_i = C_0 + \sum_{l=1}^L X_{il} C_l$$

여기에서  $X_{il}$ 은 bin,  $l$ 이나 또는 위치에서 화합물,  $i$ 의 분자 홀로그램에 대한 점유값이고  $C_l$ 는 PLS 분석으로부터 유도된 bin에 대한 계수이다. 그리고  $L$ 은 홀로그램의 길이이고  $\text{Activity}_i$ 는 화합물,  $i$ 의 생물 활성이며  $C_0$ 는 상수이다. 홀로그램의 길이는 PLS 계산으로 반응점 탐색에 이용된다. 따라서 홀로그램과 생

물 활성간의 상관성을 계산한 PLS 계수를 공간상에 구조로 나타냄으로서 반응점들에 대한 정보를 얻을 수 있으며 화합물의 활성에 관련된 분자 조각들의 유형을 확인함으로써 분자들의 활성이 예측된다.

실제, CoMFA 및 CoMSIA 기법의 분석 자료를 토대로 조성의 수는 계산에 사용된 화합물 수의 절반의 수와 사용되는 홀로그램의 길이의 절반보다 작은 수를 조성의 수로 사용하여 LOO 방법과 교차확인 등과 같은 PLS 방법으로 계산된다. 이상과 같이 HQSAR 기법은 각 분자에 대한 구조적인 조각의 생성과 홀로그램중에 분자 조각들의 암호화 그리고 생물활성 자료의 상관분석이 이루어지는 3가지 주된 과정으로 이루어진다. 이와 같은 CoMFA 및 CoMSIA 등, 3D QSAR 기법들은 2D QSAR (Kubinyi, 1993a)에 비하여 비교 우위적으로 양호한 개념이 아니라, 상호 보완적으로 각 경우마다 장단점을 가지고있는 (Kim, 1993) 방법이다.

HQSAR 분석 방법을 활용한 저자의 연구 사례로 metalaxyl 살균제 저항성과 감수성 고추역병균 (*Phytophthora capsici*)들에 대한 2-*N*-benzyl-5-phenoxy-3-isothiazolone 유도체의 살균 활성에 관한 HQSAR 분석에서 분자 조각은 4~7이었고 길이는 두 군주에 대하여 각각 59 및 53이었다. 또한, 주성분 수는 2~6이었고 교차 확인값( $q^2$ )은 저항성 그리고 상관성( $r^2$ )은 감수성 군주가 큰 편으로 저항성 군주에 대하여 예측능력이 높은 경향이였다 (성 등, 2002a).

### QSAR 관련 다변량 분석

QSAR 기법을 실행하는 데에 쓰이는 통계적인 분석 방법에는 단순 또는 다중 회귀분석, 주성분 회귀 분석, 부분 최소자승법(PLS) 및 유전자 기능 추정법 등, 대략 5가지의 다변량 분석 (multivariate analysis) 방법들이 있다. Hansch식과 같은 형태의 classical 2D QSAR 기법에서는 단순 또는 다중 회귀분석 방법 (Dillon 등, 1984)을 사용한다.

단순 회귀분석 방법의 장점은, 가장 중요한 설명 인자를 찾아내는데 유용하며 다중 회귀분석 방법은 한 분자의 생물 활성에 대하여 5배수 이상의 설명 인자와 조합으로 직선 관계식을 얻는 방법이다. 흔히 쓰이는 단순 또는 다중 회귀분석 방법에서는 화합물에 따른 물리-화학 파라미터인 설명 인자의 중요성이 처음부터 뚜렷하게 판명되는 경우는 거의 없다. 그러므로 1개의 설명 인자에 대하여 최소한 4~5개 가능

하면 5개 이상의 화합물이 있을 때에는 chance correlation의 영향이 거의 제거된다. 그리고 상관계수의 자승값인 결정계수( $r^2$ )는 부호에 상관없이 두 변량간 연관성의 정도를 표시 할 경우에 사용하는 값으로 그 범위는  $0 \leq r^2 \leq +1$ 이며  $r^2$  값이 0.99~1.00인 경우는 매우 양호, 0.95~0.99 사이의 값을 가질 경우에는 양호, 0.90~0.95 범위는 대체로 양호 그리고 0.90~ 이하인 경우에는 불량한 수준으로 판정한다. 특히, 2D QSAR 기법에서는 독립변수 즉, 설명 인자 1개당 기질 분자가 5개 이상이 요구되는 thumb rule을 만족시켜야 chance correlation coefficient (r)을 피할 수 있고 변수들 사이에는 독립적이어야 함을 correlation matrix로부터 확인하여야 한다.

주성분 회귀분석(PCR; Principal component regression analysis) 방법은, 비 독립변수를 주성분으로 변환한 다음에 활성과 각각의 성분과의 다중 선형 관계식을 유도하는 방법이다. 이 방법은 설명 인자를 변형 인자로 이용하기 때문에 결과의 해석이 어려우나 예측 능력이 좋은 방법이기도 하다. PLS는 비 독립변수와 독립변수를 모두 전환하여 직선 관계식의 유도에 이용하는 방법(Wold 등, 1984)으로 설명 인자들 사이의 상관성이 많은 경우에 효과적으로 활용하는 PCR 방법과 유사한 방법이다. 그리고 유전자 기능 추정법(GFA; Genetic function approximation)(David 등, 1994)은, 진화의 원리를 적용한 방법으로 모델의 수가 적은 경우에는 높은 예측 능력을 나타내지만 각 세대에서도 교차확인(cross-validation)을 하는데 많은 시간이 요구되는 단점이 있다.

이외에 Free-Wilson 방법(de novo method) (Free 등, 1964)은 약물의 구조에 관한 특성이 복수개의 정성적인 자료로 표현되었을 경우에 이것을 binary scale(0 또는 1)의 특성 수량으로 변환하여 설명 변수로 사용하는 수량화 이론이다. 이 방법은 다음 식과 같이 약물의 구조상 특징을 수량화하고 생물 활성(BA)과의 기여도에 대한 관련성을 정량적으로 찾아내려는 QSAR 기법에 활용되는 통계적인 방법이다.

$$\log BA_i = \sum_j a_{jk} X_{jk} + \mu$$

즉, 모든 화합물의 생물 활성은 각 위치(j)에 있는 각 치환기(k)에 의한 생물 활성의 기여( $a_{jk}$ )와 모체의 기여도( $\mu$ )의 합(BA)으로 표현되며  $X_{jk}$ 는 치환기,  $X_k$ 가 j 위치에 있을 때의 값이다. 이와 같이 수식적으로

표현된 구조-활성 상관을 QSAR 모델이라고 한다. 그러므로 모델을 만드는 일은 구조와 생물 활성 사이에 통계적인 상관관계를 귀납하는 일이다.

한편, 설명 인자가 많은 CoMFA 기법을 활용하는 경우에는 PCR과 PLS 방법이 활성과 설명 인자간의 상관 관계식을 유도하는데 사용된다. 그리고 전술한 바와 같이 회귀 모델로서 적합성 여부 즉, 신뢰도는 일반적으로 상관계수의 제곱값인 Pearson 상관성 계수,  $r^2$  값으로 평가하며  $r^2$  값은 0.90이상 그리고 예측성 평가는 교차확인 된  $q^2$  값이 0.50 이상이어야 비교적 적합한 모델이라고 판단한다.

결국, QSAR 연구는 화합물의 구조와 성질을 함축하는 설명 인자들의 생성과 그리고 설명 인자들과 생물 활성 사이의 관계를 통계적으로 회귀분석하는 2가지 과정을 포함하고 있다. 그리고 3D QSAR 기법들은 PLS라는 통계분석 방법과 computer graphics의 발달로 빠르게 발전하고 있으며 분자 모델링 도구로는 Sybyl (Tripos)과 Cerius 2 (Accelrys) 프로그램 등이 사용된다.

## 감사의 글

본 연구는 한국과학재단 우수연구센터(ERC) 지원계획(No. R11-2002-100-03002-0)으로 이루어진 것으로 지원에 감사드립니다.

## 인용문헌

- Accelrys : www. accelrys. com, 200 Wheeler Rd., S. tower, Burlington, MA. U. S. A.
- Borher, D. C., D. S. Fullerton and K. Yoshioka (1997) Computer-assisted drug design, ACS symposium series 112. pp. 259. American chemical society, Washington, DC.
- CMC (1996) Computational medicinal chemistry, CMC modeling guide, Computer aided molecular design, QSAR, Faculty of pharmacy Utrecht university; www:cmc.pharm.uu.nl, Netherlands.
- Cramer, R. D., D. E. Patterson and J. D. Bunce (1988a) Comparative molecular field analysis (CoMFA), I. Effect of shape on binding of steroids to carrier proteins, J. Am. Chem. Soc., 110:5959~5967.



- Cramer, R. D. III., J. D. Bunce and D. E. Patterson (1988b) Cross-validation, bootstrapping, and partial least squares compared with multiple regression in conventional qsar studies. *Quant. Struct. Act. Relat.* 7:18~25.
- David, R. and A. J. Hofinger (1994) Application of genetic function approximation to quantitative structure-activity relationships and quantitative structure-property relationships. *J. Chem. Inf. Comput. Sci.* 34:854~866.
- Dillon, W. R. and M. Goldstein (1984) *Multivariate analysis: Methods and applications.* pp.209., John Wiley & Sons, New York.
- Eto, M. (1992) Rational approaches to finding agrochemical leads, Ch. 4, *In Rational Approaches to Structure, Activity, and Ecotoxicology of Agrochemicals* (ed. Draber, W. and T. Fujita) CRC Press, London.
- Free, S. M. and J. W. Wilson (1964) A mathematical contribution to structure-activity studies. *J. Med. Chem.* 7:359~366.
- Gallop, M. A., R. W. Barrett, W. Dow., S. P. A. Fodor and E. M. Gordon (1994) Applications of combinatorial technologies to drug discovery. 1. Background and peptide combinatorial libraries, *J. Med. Chem.* 37:1233~1251.
- Gasteiger, J. and M. Marsili (1980) Iterative partial equalization of orbital electro-negativity; a rapid access to atomic charges. *Tetrahedron*, 36:3219~3228.
- Ghose, A. K., V. N. Viswanadhan, J. J. Wendoloski (1999) A knowledge-based approach in designing combinatorial or medicinal chemistry libraries for drug discovery. 1. A qualitative and quantitative characterization of known drug databases. *J. Comb. Chem.* 1:55~68.
- Hansch, C. and A. Leo (1995) *Exploring QSAR: Fundamentals and Applications in Chemistry and Biology.* ACS Professional Reference Book, American Chemical Society, Washington, D.C.
- Heritage T. W. and D. R. Lewis (1999) Molecular hologram QSAR. Ch. 4., *In Rational drug design: Novel Methodology and Practical Applications* (ed. Parrill, A. L. and M. R. Reddy), ACS Symposium Series 719, American Chemical Society. Washington, DC.
- Hopfinger, A. J., S. Wang, J. S. Tokarski, B. Jin, M. Albuquerque, P. J. Madhav and C. Duraiswami (1997) Construction of 3D QSAR models using the 3D QSAR analysis formalim. *J. Am. Chem. Soc.* 119: 10509~10524.
- Keer, I. D., R. Sankaramakrishnan, O. S. Smart and M. S. P. Sansom (1994) Parallel helix bundles and ionchannels: molecular modeling via simulated annealing and restrained molecular dynamics, *Biophysical J.* 67: 1501~1515.
- Kim, K. H. (1993) Comparison of classical and 3D QSAR. pp.619~642, *In 3D QSAR Drug Design Theory, Methods and Applications* (ed. Kubinyi, H.) ESCOM. Leiden.
- Klebe, G., U. Abraham and T. Mietzner (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.
- Klebe, G. (1998) Comparative molecular similarity indices analysis: CoMSIA. *Perspect, Drug Discovery Des.* 12:87~104.
- Kubinyi, H. (1993a) QSAR, Hansch analysis and related approaches, *In Methods and Principles in Medicinal Chemistry*(ed. Mannhold, R., L. P. Krogsgaard and H. Timmerman), Vol. 1., VCH, Weinheim.
- Kubinyi, H. (1993b) *3D QSAR in Drug Design; Theory, Methods and Applications.* ESCOM, Leiden.
- Lewis, D. R. (1997) HQSAR. A new, highly predictive QSAR technique. *Tripos Technical Notes*, Vol. 1, No. 5.
- Malinowski, E. R. and D. G. Howery (1980) *Factor Analyses in Chemistry.* Wiley, New York.
- Reddy, M. R. and A. L. Parrill (1999) Overview of rational drug design. *In Rational Drug Design*(ed. Reddy, M. R. and Parrill). Ch. 1., ACS Symposium Series 719, American Chemical Society, Washington, DC.
- Silverman, B. D. and D. E. Platt (1996) Comparative molecular moment analysis (CoMMA): 3D-QSAR with out molecular superposition. *J. Med. Chem.* 39:2129~2140.
- Stahle, L. and S. Wold (1988) *Multivariate data analysis*

- and experimental design in biomedical research. *Progr. Med. Chem.* 25:292~334.
- Tong, W., D. R. Lewis, R. Perkins, Y. Chen, W. J. Welsh,, D. W. Goddette, T. W. Heritage and D. M. Sheehan (1998) Evaluation of quantitative structure-activity relationship methods for large-scale prediction of chemicals binding to the estrogen receptor. *J. Chem. Inf. Comput. Sci.* 38:669~677.
- Tripos: www. tripos. com, Sybyl program (Ver. 6.9). 1699, S. Hanely Rd., Suite 303, St. Louis, MO. 63144, U. S. A.
- Yu, S. J. and N. D. Sung (2002) Comparative molecular field analysis (CoMFA) on the farnesyl protein transferase (FPTase) inhibition activity of bis-aromatic  $\alpha, \beta$ -unsaturated ketones. *Agric. Chem. Biotechnol.* 45 (1):1~6.
- Vedani, A. and M. Dobler (2002) 5D-QSAR: The key for simulating induced fit. *J. Med. Chem.*, 45:2139~2149.
- Waterbeemd, H. V. D. (1992) The history of drug research: From Hansch to the present, *Quant. Struct. Act. Relat.* 11:200~204.
- Waterbeemd, H. V. D. (1996) Design of bioactive compounds, Ch.1., *In Structure-Property Correlations in Drug Research* (ed. Waterbeemd, H. V. D.). Academic Press, R. G. Landes Co. Austin.
- Wexler, R. R., W. J. Greenlee, J. D. Irvin, M. R. Goldberg, K. Prendergast, R. D. Smith and P. B. Timmermans (1996) Nonpeptide angiotensin II receptor antagonists: The next generation in antihypertensive therapy, *J. Med. Chem.*, 39:625~656.
- Wise, M., R. D. Cramer, D. Smith and I. Exman (1983) Progress in three-dimensional drug design: The use of real time colour graphics and computer postulation of bioactive molecules in DYLOMMS. pp.145~146. *In Quantitative Approaches to Drug Design* (ed. Dearden, J. C.), Elsevier, Amsterdam.
- Wold, S., C. Albano, W. J. I. Dunn, U. Edlund,, K. Esbensen, P. Geladi, S. Hellberg, E. Johansson, W. Lindberg and M. Sjostrom (1984) Multivariate data analysis in chemistry. *In Chemometrics: Mathematics and Statistics in Chemistry* (ed Kowalski, B.) Reidel, Dordrecht, Netherlands.
- 성낙도, 유성재, 임치환, 赤松美紀 (1998) 비스 방향족  $\alpha, \beta$ -불포화 케톤 유도체 중 2-thienyl 및 2-furyl 치환체의 항균활성에 관한 비교분자장 분석(CoMFA), *한국농약과학회지* 2(2):16~21.
- 성낙도, 김기현 (2002a) 고추역병균에 대한 2-N-benzyl-5-phenoxy-3-isothiazolone 유도체의 살균활성에 관한 비교분자 유사성 지수분석(CoMSIA)과 홀로그래프 구조-활성관계(HQSAR). *한국농약과학회지* 6(3): 209~217.
- 성낙도, 이희철 (2002b) 고추 역병균에 대한 3-phenyl-isoxazole과 3-phenyl-2,5-dihydrois-oxazol-5-one 유도체들의 살균활성에 관한 비교 분자장 분석(CoMFA). *한국농약과학회지* 6(2):72~79.
- 성낙도, 옥환석, 송중환, 이용구 (2003a) N-치환 phenyl-3,4,5,6-tetrahydrophthalimide와 N-치환 phenyl-3,4-dimethylmalcimine 유도체의 생장 저해 활성에 관한 비교 분자장 분석(CoMFA). *한국농약과학회지* 투고 중.
- 성낙도, 이상호, 송중환, 김형래 (2003b) 제초성 3-phenyl-5-(3,7-dichloro-8-quinoliny)-1,2,4-oxadiazole 유도체들의 생장 저해활성에 관한 비교 분자장 분석(CoMFA)과 분자 홀로그래프 구조-활성관계(HQSAR). *한국농약과학회지* 투고 중

---

**Development of new agrochemicals by quantitative structure-activity relationship (QSAR) methodology.**

**III. 3D QSAR methodologies and computer-assisted molecular design (CAMD)**

Nack-Do Sung\* (*Division of Applied Biology & Chemistry, College of Agriculture & Life Sciences, Chungnam National University, Daejeon 305-764, Korea.*)

**Abstract** : According to improvement of HTOS (high throughput organic synthesis) and HTS (high throughput screening) technique, the CoMFA (comparative molecular field analysis), CoMSIA (comparative molecular similarity indices analysis) and molecular HQSAR (hologram quantitative structure-activity relationship) analysis techniques as methodology of computer assisted molecular design (CAMD) were introduced generally and summarized for some application cases.

---

\*Corresponding author (Fax : +82-42-825-3306, E-mail : ndsung@cnu.ac.kr)