

## 정량적 구조-활성관계(QSAR) 모델에 대한 연구 동향

김길한<sup>○</sup> 류기곤 김연주 김현철

고려대학교 컴퓨터교육학과

gilly@comedu.korea.ac.kr, gon0121@korea.ac.kr, teddy95@korea.ac.kr, hkim@comedu.korea.ac.kr

### A Survey on Quantitative Structure-Activity Relationship(QSAR)

#### Models

Gilhan Kim<sup>○</sup> Kigon Lyu Yeonjoo Kim Hyeoncheol Kim

Dept. of Computer Science Education, Korea University

#### 서론

현대산업의 발전으로 기존 생태계에 존재하지 않은 신규 물질들이 빠르고 다양하게 생산 가능해짐에 따라, 매년 인간에게 얼마나 위해한지 규명되지 않은 신규 화학물질들의 종류가 급격히 증가하고 있다 이에 따른 신규 화학물질에 대한 국민노출도 빠르게 증가하고 있어 위해성을 고려한 화학물질의 안전성 관리가 중요한 사회문제로 대두되고 있다 전통적인 물질기반(invitro/vivo) 동물시험을 통한 위해성 평가/승인 방법은 매년 늘어나는 방대한 신규물질들의 위해성을 평가하는데 상당한 비용과 시간을 필요로 한다. 또한 유럽 국가들과 OECD를 중심으로 동물시험에 대한 윤리적 문제가 제기됨에 따라 기존의 위해성 평가방법을 대체할 기술이 절실히 요구되고 있다 이에, OECD[1]는 컴퓨터를 활용하여 경험적으로 알려진 기존 유해물질의 분자구조 정보를 이용해 신규 혹은 알려지지 않은 화학구조의 물리화학적 성질을 예측할 수 있는 정량적 구조-활성관계(QSAR)를 국제적 위해성 평가 및 검증 자료로 이용하기 위해 QSAR 표준화 정책을 제시하고 있으며 회원국을 중심으로 QSAR 사용을 의무화하고 있다 이에 본 논문에서는 OECD 수준의 QSAR에 대한 구체적 설명과 앞으로 QSAR의 발전 방향 및 중심 이슈들에 대해 논의한다.

#### 본론

QSAR(Quantitative Structure-Activity Relationship)은 분자 시스템의 특징적인 구조 혹은 화학적 특성(descriptor)과 분자 성질 사이에 존재하는 관련성을 통계적으로 모델화하여 분자 성질을 예측할 수 있는 모델로써, 구조가 유사한 물질은 성질도 유사할 것이라는 전제 하에 물질의 생물학적 활성(activity)을 예측하는 모델이다[2]. QSAR의 일반적인 목적은 크게 두 가지로 첫 번째는 구조와 활성사이의 관계를 이해하여, 화학물의 생물학적 활성을 결정짓는 화학적 속성에 대한 정보를 제공하는 것이고, 두 번째는 테스트되지 않은 혹은 아직 알려지지 않은 생물학적 활성을 예측하기 위함이다[3]. QSAR은 OECD와 선진국을 중심으로 동물시험 없이 신속하고 경제적으로 인간에게 유해한 물질을 구별할 수 있다는 장점이 부각되어, 국제적으로 제약 및 화학 산업분야에 신뢰된 대체검증시험 방법으로 권장되어 사용되고 있다. 또한 실제 시험자료가 없는 독성분야 연구 구조적으로 유사한 물질그룹 관리 기존화학물질의 평가 우선순위 설정 등에 유용하나 도구로써 활용 가능할 것으로 기대되고 있다 이에, OECD에서는 QSAR을 국제적 위해성평가시험 자료로 활용하기 위해 QSAR 원칙 5가지[4]와, QSAR 모델 검증을 위한 QSAR 가이드라인[5], 최근에는 QSAR application "Toolbox 1.0[6]"을 함께 제시하고 있다.

일반적으로 QSAR에서는 분자의 화학적인 성질이나 특징을 나타내는 표현자(descriptor)를 이용하여 분자구조를 표현한다 이때 분자구조를 평면으로 간주하는지 혹은 입체적인 구조로 간주하여 연구하는가에 따라 표현자(descriptor)를 1D, 2D, 3D 그리고 4D, 5D의 입력정보로 표현한다[7]. QSAR의 기본인 2D-QSAR 알고리즘은 대부분 통계적 접근 방법(MLR, PCA&PCR, PLS 등)을 이용하여 예측모델을 만들어 분석/예측하거나, 지금까지 알려진 Rule을 기반으로 Expert system을 운영하고 있다. 최근

QSAR의 알고리즘에 대한 연구는 기계학습 기반의 알고리즘(ANN, SVM, GA, GPM 등)들을 이용한 QSAR 모델에 대한 연구로써 기존의 통계기반 알고리즘 보다 더 효과적으로 고성능의 예측모델을 만들 수 있다는 연구결과들이 보고되고 있다 하지만, ANN이나 SVM과 같은 black-box형 기계학습 기반의 알고리즘으로부터 학습된 예측 모델의 경우 학습 알고리즘으로부터 학습된 결과나 예측과정을 사용자가 명백히 알 수 없다는 알고리즘적 특징으로 인해 우수한 성능을 보이지만 OECD에서 제시하는 QSAR 원칙 3(an unambiguous algorithm)에 위배되고 원칙 5에 적합하지 못한 문제를 갖고 있다[8,9]. 이에, 기계학습 기반의 알고리즘을 이용하여 예측 모델의 성능을 높이면 OECD 원칙을 준수 할 수 있는 기술에 대한 연구가 진행되어야 할 것이다 3D-QSAR 알고리즘의 경우, 기본적으로 3D 생화학물질의 구조정보를 얻은 후에는 2D-QSAR 알고리즘에서 사용한 알고리즘들을 이용하여 분자들의 활성값을 예측한다. 3D-QSAR 알고리즘은 3차원 공간에 화학구조를 정렬하고 분자를 설계하기까지의 기법들에 따라 CoMFA(Comparative Molecular Field Analysis)[10], CoMSIA(Comparative Molecular Similarity Indices Analysis)[11], HQSAR(Hologram QSAR)[12]로 분류된다.

#### 결론

물질과 활성간의 관계를 신속하게 예측할 수 있는 OECD 수준의 QSAR 모델은 앞으로 국제적 위해성 대체 시험자료로써 광범위하게 활용될 수 있을 것으로 기대되며 오늘날 국제사회서 QSAR은 더 이상 선택이 아닌 필수 기술이라 말할 수 있을 것이다 국제적 활용 가능한 QSAR 모델을 개발하기 위해서는 반드시 OECD QSAR 표준화 정책 가운데 QSAR 원칙을 준수하여 개발검증되어야 할 것이며, OECD 원칙을 준수한 기계학습기반의 알고리즘에 대한 연구도 진행되어야 할 것이다 또한 최근 Omics 기술이 부각됨에 따라 유전체학, 단백질체학, 대사체학 등의 다학제기술을 접목시켜 독성에 대한 생체지표물질을 발굴하는 독성유전체 기술에 대한 관심이 높아지고 있는 가운데 독성유전체(Toxicogenomics) 기술은 기존의 독성평가방법과는 다르게 세포나 생물체의 유전자 발현패턴에 기초하여 유전체기술의 독성을 평가한다. 따라서, 현재 생화학물질의 구조정보만은 이용하여 물질과 활성간의 관계를 예측하는 QSAR 모델을 "Omics" 영역에서 발생할 수 있는 유전자 발현패턴에 따른 활성을 예측하는 QSAR in Omics"에 대한 연구로 확장하여 진행되어야 할 것이다

#### 참고문헌

- [1] [http://www.oecd.org/document/23/0,2340,en\\_2649\\_201185\\_33957015\\_1\\_1\\_1\\_1,00.html](http://www.oecd.org/document/23/0,2340,en_2649_201185_33957015_1_1_1_1,00.html).
- [2] Crum-Brown, A. and Fraser, T. R., On the connection between chemical constitution and physiological action. Part I. On the physiological action of the salts of the ammonium bases, derived from Strychnia, Brucia, Thebaia, Codeia, Morphia, and Nicotia, Transactions of the Royal Society of Edinburgh; 25, 151, 1989;
- [3] Eriksson, L., Jaworska, J., Worth, A.P., Cronin, M.T.D., McDowell, R.M. and Gramatica, P., Methods for reliability and uncertainty assessment and for applicability evaluations of classification-and regression-based QSARs, Environmental Health Perspectives; 111(10), 1361-1375, 2003;
- [4] <http://www.oecd.org/dataoecd/33/37/37849783.pdf>
- [5] <http://www.oecd.org/dataoecd/55/35/38130292.pdf>
- [6] <http://www.oecd.org/dataoecd/5/4/40324067.pdf>
- [7] Livingstone, D.J., The characterization of chemical structures using molecular properties. A survey, Journal of Chemical Information and Computer Sciences; 40, 195-209, 2000;
- [8] Marini F., Roncaglioni A., Novič M., Variable selection and interpretation in structure-affinity correlation modeling of estrogen receptor binders. J. Chem. Inf. Model; 45(6), 1507-1519, 2005;
- [9] H. Liu, E. Papa, P. Gramatica., QSAR Prediction of Estrogen Activity for a Large Set of Diverse Chemicals under the Guidance of OECD Principles. Chemical Research in Toxicology; 19, 1540-1548, 2006;
- [10] R.D. Cramer, D.E. Patterson and J.D. Bunce., Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins J. Am. Chem. Soc.; 110, 5959, 1998;
- [11] G. Klebe, U. Abraham and T. Mietaner., Molecular similarity indices in a comparative analysis (CoMSIA) of drug molecules to correlate and predict their biological activity, J. Med. Chem.; 37, 4130-4146, 1994;
- [12] Lowis, S.P. et al., Chemotherapy for spinal cord astrocytoma: Can natural history be modified? Childs Nerv. Syst., 14, 317-321, 1998;