



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2019-0005398
(43) 공개일자 2019년01월16일

(51) 국제특허분류(Int. Cl.)
G16C 10/00 (2019.01)

(52) CPC특허분류
G06F 19/706 (2013.01)
G06F 19/701 (2013.01)

(21) 출원번호 10-2017-0085981
(22) 출원일자 2017년07월06일
심사청구일자 2017년07월06일

(71) 출원인
부경대학교 산학협력단
부산광역시 남구 신선로 365 (용당동,
부경대학교)

(72) 발명자
유준
부산광역시 해운대구 세실로 174, 106동 1404호(
좌동, 삼성아파트)

이명기
부산광역시 남구 분포로 111, 134동 702호(
용호동, 엘지메트로시티)

주벨라 페타
싱가포르 우드랜즈 드라이브 16 포레스트빌 40
11-41

(74) 대리인
김정수

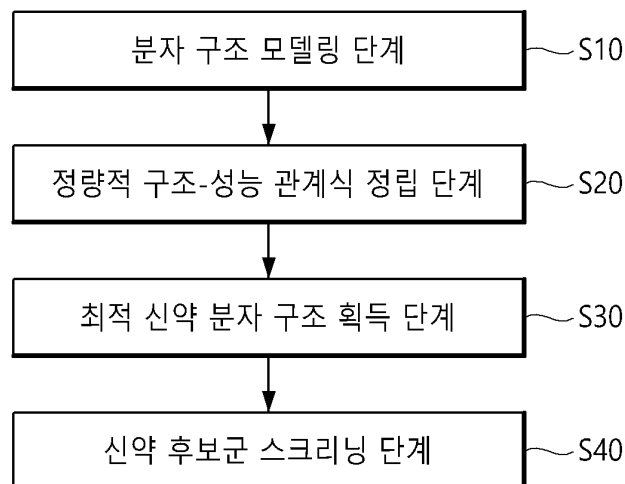
전체 청구항 수 : 총 8 항

(54) 발명의 명칭 정량적 구조-성능 관계식의 수치적 반전과 분자동역학 전산모사를 통한 표적신약의 스크리닝 방법

(57) 요약

본 발명에 따른 정량적 구조-성능 관계식의 수치적 반전과 분자동역학 전산모사를 통한 표적신약의 스크리닝 방법은 표적분자에 대한 시험 화합물 군의 분자 구조를 모델링하는 분자 구조 모델링 단계; 상기 시험 화합물 군의 분자 구조와 성능 사이의 정량적 구조-성능 관계식을 구하는 정량적 구조-성능 관계식 정립 단계; 상기 관계식의 수치적 반전을 통해 최적의 신약 분자 구조를 획득하는 최적 신약 분자 구조 획득 단계; 및 상기 최적의 신약 분자 구조와 유사한 신약 후보군을 스크리닝하는 신약 후보군 스크리닝 단계를 포함하는 것을 특징으로 한다.

대 표 도 - 도1



(52) CPC특허분류
G06F 19/703 (2013.01)

명세서

청구범위

청구항 1

표적분자에 대한 시험 화합물 군의 분자 구조를 모델링하는 분자 구조 모델링 단계;

상기 시험 화합물 군의 분자 구조와 성능 사이의 정량적 구조-성능 관계식을 구하는 정량적 구조-성능 관계식 정립 단계;

상기 관계식의 수치적 반전을 통해 최적의 신약 분자 구조를 획득하는 최적 신약 분자 구조 획득 단계; 및

상기 최적의 신약 분자 구조와 유사한 신약 후보군을 스크리닝하는 신약 후보군 스크리닝 단계를 포함하는 정량적 구조-성능 관계식의 수치적 반전과 분자동역학 전산모사를 통한 표적신약의 스크리닝 방법.

청구항 2

제 1항에 있어서,

상기 분자 구조 모델링 단계는,

상기 시험 화합물 군을 선택하는 화합물 선택 공정;

상기 시험 화합물 군의 화학적, 생물학적 실험 데이터를 입력받는 데이터 입력 공정; 및

상기 실험 데이터를 이용하여 상기 시험 화합물 군의 분자 구조를 모델링하여 최적화하는 분자 구조 모델링 공정을 포함하는 정량적 구조-성능 관계식의 수치적 반전과 분자동역학 전산모사를 통한 표적신약의 스크리닝 방법.

청구항 3

제 1항에 있어서,

상기 정량적 구조-성능 관계식 정립 단계는,

상기 분자 구조로부터 분자설명자를 계산하는 분자설명자 계산 공정; 및

상기 분자설명자를 사용하여 상기 정량적 구조-성능 관계식을 모델링하는 정량적 구조-성능 관계식 모델링 공정을 포함하는 정량적 구조-성능 관계식의 수치적 반전과 분자동역학 전산모사를 통한 표적신약의 스크리닝 방법.

청구항 4

제 3항에 있어서,

상기 정량적 구조-성능 관계식에서의 성능은 생물학적 활성(biological activity), 억제작용(inhibitory activity), 친유성(lipophilicity), 독성(toxicity), 대사 안정성(metabolic stability) 및 혈액-뇌 장벽 투과성(blood-brain barrier permeability) 중 어느 하나를 포함하는 정량적 구조-성능 관계식의 수치적 반전과 분자동역학 전산모사를 통한 표적신약의 스크리닝 방법.

청구항 5

제 3항에 있어서,

상기 정량적 구조-성능 관계식 모델링 공정은,

상기 계산된 분자설명자에 대해 유전 알고리즘을 적용하여 일부 분자설명자를 선택한 후, 상기 선택된 일부 분자 설명자를 사용하여 상기 정량적 구조-성능 관계식을 모델링하는 정량적 구조-성능 관계식의 수치적 반전과 분자동역학 전산모사를 통한 표적신약의 스크리닝 방법.

청구항 6

제 1항에 있어서,

상기 최적 신약 분자 구조 획득 단계는,

하기 [수학식 1] 또는 [수학식 2]에 의한 수치적 반전을 통해 최적의 신약 분자 구조를 획득하는 정량적 구조-성능 관계식의 수치적 반전과 분자동역학 전산모사를 통한 표적신약의 스크리닝 방법.

[수학식 1]

$$\mathbf{x}^* = \operatorname{argmax} \log \hat{k}_w$$

$$s.t \log \hat{k}_w = \mathbf{C} \hat{\mathbf{t}}$$

$$\hat{\mathbf{t}} = \mathbf{P} \mathbf{x}$$

$$\hat{\mathbf{t}}^T \mathbf{S}_t^{-1} \mathbf{t} \leq c_1$$

$$\| \mathbf{P} \hat{\mathbf{t}} - \mathbf{x} \| \leq c_2$$

$\mathbf{x}$: 신약의 분자설명자

$\mathbf{x}^*$: 위의 수치계획법 수학식으로부터 계산된 최적의 신약의 분자설명자

$\mathbf{C}$: 부분최소자승법(partial least squares, PLS)의 출력변수 로딩(loading) 행렬

$\mathbf{t}$: PLS 입력변수(여기서는 분자설명자, $\mathbf{x}$) 스코어(score) 벡터

$\mathbf{P}$: PLS 입력변수 로딩 행렬

$\wedge$: PLS 모델에 의해 예측된 값

$\mathbf{S}_t$: $\mathbf{t}$ 의 표본 공분산 행렬

c_1, c_2 : 적절한 상수

[수학식 2]

$$\mathbf{x}^* = \operatorname{argmax} \sqrt{(\log k_w - \log k_{w,ref})^2 + (\log k_i - \log k_{i,ref})^2}$$

$$s.t [\log \hat{k}_w, \log \hat{k}_w]^T = \mathbf{C} \hat{\mathbf{t}}$$

$$\hat{\mathbf{t}} = \mathbf{P} \mathbf{x}$$

$$\hat{\mathbf{t}}^T \mathbf{S}_t^{-1} \mathbf{t} \leq c_1$$

$$\| \mathbf{P} \hat{\mathbf{t}} - \mathbf{x} \| \leq c_2$$

$\mathbf{x}$: 신약의 분자설명자

$\mathbf{x}^*$: 위의 수치계획법 수학식으로부터 계산된 최적의 신약의 분자설명자

$\mathbf{C}$: 부분최소자승법(partial least squares, PLS)의 출력변수 로딩(loading) 행렬

$\mathbf{t}$: PLS 입력변수(여기서는 분자설명자, $\mathbf{x}$) 스코어(score) 벡터

$\mathbf{P}$: PLS 입력변수 로딩 행렬

$\wedge$: PLS 모델에 의해 예측된 값

$\mathbf{S}_t$: $\mathbf{t}$ 의 표본 공분산 행렬

c_1, c_2 : 적절한 상수

$\log k_{w,ref}$: 사용자가 지정한 친유성 값

$\log k_{i,ref}$: 사용자가 지정한 활성 값

청구항 7

제 1항에 있어서,

상기 신약 후보군 스크리닝 단계는,

상기 최적의 신약 분자 구조와 신약 후보군의 분자 구조의 유클리드 거리에 따라 신약 후보군에 등급을 부여하는 신약 후보군 등급부여 공정; 및

상기 신약 후보군 중 일정 등급 이상의 신약 후보군을 선택하는 신약 후보군 선택 공정을 포함하는 정량적 구조-성능 관계식의 수치적 반전과 분자동역학 전산모사를 통한 표적신약의 스크리닝 방법.

청구항 8

제 7항에 있어서,

상기 신약 후보군 스크리닝 단계 후,

상기 스크리닝 된 신약 후보군과 상기 표적분자의 복합물(complex)을 분자동역학 전산모사(molecular dynamics simulation)를 통해 검증하는 표적 신약 검증 단계를 더 포함하는 정량적 구조-성능 관계식의 수치적 반전과 분자동역학 전산모사를 통한 표적신약의 스크리닝 방법.

발명의 설명

기술 분야

[0001] 본 발명은 표적신약의 스크리닝 방법에 관한 것으로서, 보다 상세하게는 시험 화합물에 대한 정량적 구조-성능 관계식의 수치적 반전을 통해 최적 신약 분자 구조를 획득한 후, 신약 후보군과 표적분자의 복합물을 분자동역학 전산모사를 통해 표적 신약을 스크리닝할 수 있는 방법에 관한 것이다.

배경 기술

[0002] 수년 동안 신약개발은 화합물 자체의 개발 및 분석에 초점을 둔 구조기반방법이 지배적이었다. 최근에는 기술의 발달과 함께 표적기반 약물설계 및 가상 리간드 스크리닝과 같은 방법이, 신약개발 연구단계와 개발단계 초기에 특정 생물학적 표적에 대한 활성 화합물들을 컴퓨터를 이용하여 가상적으로 비교적 정확하게 예측하는 인실리코 예측 기술(in silico prediction technology)로써 유용하게 활용되고 있다.

[0003] 미국식품의약품안전국(FDA)에서는 제약회사들을 상대로 인실리코(in silico) 컴퓨터 모델 기술을 이용한 약물의 효능과 부작용 검증을 활용하도록 권고하고 있으며, 적절한 인실리코(in silico) 컴퓨터 모델 기술을 활용했을 경우, 기존의 신약승인 과정에서 요구되었던 임상시험(clinical test) 단계를 대폭적으로 줄일 수 있는 효과가 있다. 실제, FDA는 약물 적용 시, 실패여부 확인 방법을 10%만 개선하여도 약물개발 비용을 1억 내지 10억 달러까지 절감할 수 있는 것으로 추정하고 있다.

[0004] 따라서, 세계적 기술력을 확보한 선진제약기업에서는 전임상에 진입하기 전 대사검색, 흡수 스크리닝 실험데이터 등을 이용한 컴퓨터 예측시스템을 개발하고 있다. 대표적으로 2003년부터 다국적 제약회사인 노바티스(Novartis)사는 독성문제를 해결하기 위해 독스체크(ToxCheck)라는 프로그램을 개발한 것으로 알려져 있으나 외부에 공개되지 않아 그 기술력을 판단하기 어려운 상황이고, 최근 들어, 약물의 상호작용 및 독성을 평가해주는 기관/기업체가 다수 설립되는 것만으로도 예측모델 개발의 필요성을 뒷받침하고 있다.

[0005] 또한, 컴퓨터를 이용한 예측모델을 만들기 위하여, 필요한 실험 데이터 축적에 필요한 많은 투자가 미국, 독일, 영국, 일본 등을 중심으로 진행되고 있다.

선행기술문헌

특허문헌

- [0006] (특허문헌 0001) 미국공개특허 제2005/0240355호 (공개일자 2005.10.27)
 (특허문헌 0002) 한국등록특허 제10-1128425호(공개일자 2010.12.08)
 (특허문헌 0003) 한국공개특허 제10-2015-0100152호(공개일자 2015.09.02)

발명의 내용

해결하려는 과제

- [0007] 본 발명은 상기와 같은 필요에 의해 창출된 것으로서, 표적분자에 대한 시험 화합물의 정량적 구조-성능 관계식의 수치적 반전과 신약 후보군과 표적분자의 복합물에 대한 분자동역학 전산모사를 이용함으로써 약물 개발에서의 소모 비용을 상당히 줄이고 잘못된 합성과 불필요한 제약 테스트를 피할 수 있는 표적 신약의 스크리닝 방법을 제공하는데 그 목적이 있다.

과제의 해결 수단

- [0008] 상기한 목적을 달성하기 위해, 본 발명에 따른 정량적 구조-성능 관계식의 수치적 반전과 분자동역학 전산모사를 통한 표적신약의 스크리닝 방법은 표적분자에 대한 시험 화합물 군의 분자 구조를 모델링하는 분자 구조 모델링 단계, 상기 시험 화합물 군의 분자 구조와 성능 사이의 정량적 구조-성능 관계식을 구하는 정량적 구조-성능 관계식 정립 단계, 상기 관계식의 수치적 반전을 통해 최적의 신약 분자 구조를 획득하는 최적 신약 분자 구조 획득 단계 및 상기 최적의 신약 분자 구조와 유사한 신약 후보군을 스크리닝하는 신약 후보군 스크리닝 단계를 포함하는 것을 특징으로 한다.
- [0009] 또한, 상기 분자 구조 모델링 단계는 상기 시험 화합물 군을 선택하는 화합물 선택 공정, 상기 시험 화합물 군의 화학적, 생물학적 실험 데이터를 입력받는 데이터 입력 공정 및 상기 실험 데이터를 이용하여 상기 시험 화합물 군의 분자 구조를 모델링하여 최적화하는 분자 구조 모델링 공정을 포함할 수 있다.
- [0010] 또한, 상기 정량적 구조-성능 관계식 정립 단계는 상기 분자 구조로부터 분자설명자를 계산하는 분자설명자 계산 공정 및 상기 분자설명자를 사용하여 상기 정량적 구조-성능 관계식을 모델링하는 정량적 구조-성능 관계식 모델링 공정을 포함할 수 있다.
- [0011] 또한, 상기 정량적 구조-성능 관계식에서의 성능은 생물학적 활성(biological activity), 억제작용(inhibitory activity), 친유성(lipophilicity), 독성(toxicity), 대사 안전성(metabolic stability) 및 혈액-뇌 장벽 투과성(blood-brain barrier permeability) 중 하나 이상을 포함할 수 있다.
- [0012] 또한, 상기 정량적 구조-성능 관계식 모델링 공정은 상기 계산된 분자설명자에 대해 유전 알고리즘을 적용하여 일부 분자설명자를 선택한 후, 상기 선택된 일부 분자 설명자를 사용하여 상기 정량적 구조-성능 관계식을 모델링할 수 있다.
- [0013] 또한, 상기 최적 신약 분자 구조 획득 단계는, 하기 [수학식 1] 또는 [수학식 2]에 의한 수치적 반전을 통해 최적의 신약 분자 구조를 획득할 수 있다.

[0014] [수학식 1]

$$\begin{aligned} \mathbf{x}^* &= \operatorname{argmax} \log \hat{k}_w \\ s.t \quad & \log \hat{k}_w = \mathbf{C} \hat{\mathbf{t}} \\ & \hat{\mathbf{t}} = \mathbf{P} \mathbf{x} \\ & \hat{\mathbf{t}}^T \mathbf{S}_t^{-1} \mathbf{t} \leq c_1 \\ & \| \mathbf{P} \hat{\mathbf{t}} - \mathbf{x} \| \leq c_2 \end{aligned}$$

[0015]

[0016] $\mathbf{x}$: 신약의 분자설명자

[0017] $\mathbf{x}^*$: 위의 수치계획법 수학식으로부터 계산된 최적의 신약의 분자설명자

[0018] $\mathbf{C}$: 부분최소자승법(partial least squares, PLS)의 출력변수 로딩(loading) 행렬

[0019] $\mathbf{t}$: PLS 입력변수(여기서는 분자설명자, $\mathbf{x}$) 스코어(score) 벡터

[0020] $\mathbf{P}$: PLS 입력변수 로딩 행렬

[0021] $\hat{\cdot}$: PLS 모델에 의해 예측된 값

[0022] $\mathbf{S}_t$: $\mathbf{t}$ 의 표본 공분산 행렬

[0023] c_1, c_2 : 적절한 상수

[0024] [수학식 2]

$$\begin{aligned} \mathbf{x}^* &= \operatorname{argmax} \sqrt{(\log k_w - \log k_{w,ref})^2 + (\log k_i - \log k_{i,ref})^2} \\ s.t \quad & [\log \hat{k}_w, \log \hat{k}_i]^T = \mathbf{C} \hat{\mathbf{t}} \\ & \hat{\mathbf{t}} = \mathbf{P} \mathbf{x} \\ & \hat{\mathbf{t}}^T \mathbf{S}_t^{-1} \mathbf{t} \leq c_1 \\ & \| \mathbf{P} \hat{\mathbf{t}} - \mathbf{x} \| \leq c_2 \end{aligned}$$

[0025]

[0026] $\mathbf{x}$: 신약의 분자설명자

[0027] $\mathbf{x}^*$: 위의 수치계획법 수학식으로부터 계산된 최적의 신약의 분자설명자

[0028] $\mathbf{C}$: 부분최소자승법(partial least squares, PLS)의 출력변수 로딩(loading) 행렬

[0029] $\mathbf{t}$: PLS 입력변수(여기서는 분자설명자, $\mathbf{x}$) 스코어(score) 벡터

[0030] $\mathbf{P}$: PLS 입력변수 로딩 행렬

[0031] $\hat{\cdot}$: PLS 모델에 의해 예측된 값

[0032] $\mathbf{S}_t$: $\mathbf{t}$ 의 표본 공분산 행렬

[0033] c_1, c_2 : 적절한 상수

[0034] $\log k_{w,ref}$: 사용자가 지정한 친유성 값

[0035] $\log k_{i,ref}$: 사용자가 지정한 활성 값

[0036] 또한, 상기 신약 후보군 스크리닝 단계는 상기 최적의 신약 분자 구조와 신약 후보군의 분자 구조의 유클리드

거리에 따라 신약 후보군에 등급을 부여하는 신약 후보군 등급부여 공정 및 상기 신약 후보군 중 일정 등급 이상의 신약 후보군을 선택하는 신약 후보군 선택 공정을 포함할 수 있다.

[0037] 아울러, 상기 신약 후보군 스크리닝 단계 후, 상기 스크리닝 된 신약 후보군과 상기 표적분자의 복합물(complex)을 분자동역학 전산모사(molecular dynamics simulation)를 통해 검증하는 표적 신약 검증 단계를 더 포함할 수 있다.

발명의 효과

[0038] 본 발명에 따른 표적 신약 스크리닝 방법은 모든 약물 및 분자 표적에 적용이 가능하고, 모든 약물의 성능에 적용 가능하며, 인실리코 실험에 필요한 계산량 또한 매우 적어 신약개발에 필요한 시간과 인력을 상당히 절약하는 효과가 있다.

도면의 간단한 설명

[0039] 도 1은 본 발명에 따른 표적 신약 스크리닝 방법의 블록도이다.
 도 2는 본 발명에 따른 분자 구조 모델링 단계의 블록도이다.
 도 3은 본 발명에 따른 정량적 구조-성능 관계식 정립 단계의 블록도이다.
 도 4는 본 발명에 따른 신약 후보군 스크리닝 단계의 블록도이다.
 도 5는 본 발명에 따른 표적 신약 스크리닝 방법의 다른 블록도이다.
 도 6은 본 발명에 따른 정량적 구조-친유성 관계식(quantitative structure-lipophiicity relationships, QSLR) 모델의 성능을 나타내는 도이다.
 도 7은 본 발명에 따른 QSLR 모델의 적용가능성 영역을 나타내는 도이다.
 도 8은 본 발명에 따른 술폰 아마이드-CA(Carbonic anhydrase) IX 복합체의 분자동역학 전산모사(molecular dynamics simulation)를 나타내는 도이다.
 도 9는 본 발명에 따른 CA IX의 안정성을 나타내는 도이다.
 도 10은 본 발명에 따른 CA IX의 유연성을 나타내는 도이다.

발명을 실시하기 위한 구체적인 내용

[0040] 이하 본 발명을 더욱 상세하게 설명한다. 하기의 설명은 하나의 실시예에 대한 구체적 설명이므로, 비록 단정적, 한정적 표현이 있더라도 특허청구범위로부터 정해지는 권리범위가 제한되지 않는다. 또한, 하기에서 본 발명을 설명함에 있어, 관련된 공지기능 또는 구성에 대한 구체적인 설명이 본 발명의 요지를 불필요하게 흐릴 수 있다고 판단되는 경우에는 그 상세한 설명을 생략한다.

[0041] 도 1은 본 발명에 따른 표적 신약 스크리닝 방법의 블록도이다.

[0042] 본 발명의 일 실시예에 따른 표적 신약 스크리닝 방법은 도 1에 도시된 바와 같이, 분자 구조 모델링 단계(S10), 정량적 구조-성능 관계식 정립 단계(S20), 최적 신약 분자 구조 획득 단계(S30) 및 신약 후보군 스크리닝 단계(S40)를 포함한다. 이러한 본 발명은 컴퓨터 상에서 실행될 수 있는 형태로 구현될 수 있다.

[0043] 도 2는 본 발명에 따른 분자 구조 모델링 단계의 블록도이다.

[0044] 상기 분자 구조 모델링 단계(S10)는 표적분자에 대한 시험 화합물 군의 분자 구조를 모델링하는 단계이다. 표적 분자는 단백질, 효소, DNA, RNA 등 일 수 있으며, 시험화합물은 표적분자를 억제하거나 변성시키는 화합물이다.

[0045] 구체적으로, 상기 분자 구조 모델링 단계(S10)는 도 2에 도시된 바와 같이, 화합물 선택 공정(S11), 데이터 입력 공정(S12) 및 분자 구조 모델링 공정(S13)을 포함한다.

[0046] 화합물 선택 공정(S11)은 상기 시험 화합물 군을 선택하는 공정으로, 표적분자를 억제 또는 변형하는 시험화합물들을 선택하는 공정이다.

[0047] 또한, 데이터 입력 공정(S12)은 상기 시험 화합물 군의 화학적, 생물학적 실험 데이터를 입력받는 공정으로, 상기 화학적, 생물학적 실험 데이터는 끓는점, 어는점, 극성, 용해도, 반응성, 독성, 선택성 등 일 수 있다.

- [0048] 아울러, 분자 구조 모델링 공정(S13)은 상기 실험 데이터를 이용하여 상기 시험 화합물 군의 분자 구조를 모델링하고 양자화학을 사용하여 최적화하는 공정이다.
- [0049] 도 3은 본 발명에 따른 정량적 구조-성능 관계식 정립 단계의 블록도이다.
- [0050] 상기 정량적 구조-성능 관계식 정립 단계(S20)는 상기 시험 화합물 군의 분자 구조와 성능 사이의 정량적 구조-성능 관계식을 구하는 단계이다.
- [0051] 구체적으로, 상기 정량적 구조-성능 관계식 정립 단계(S20)는 도 3에 도시된 바와 같이, 분자설명자 계산 공정(S21) 및 정량적 구조-성능 관계식 모델링 공정(S22)을 포함한다.
- [0052] 분자설명자 계산 공정(S21)은 상기 분자 구조로부터 4000개 이상의 분자설명자를 계산하는 공정이다.
- [0053] 또한, 정량적 구조-성능 관계식 모델링 공정(S22)은 상기 분자설명자를 사용하여 상기 정량적 구조-성능 관계식을 모델링하는 공정이다.
- [0054] 구체적으로, 정량적 구조-성능 관계식 모델링 공정(S22)에서는 상기 분자설명자 계산 공정(S21)에서 계산된 분자설명자에 대해 유전 알고리즘(GA)을 적용하여 일부 분자설명자를 선택한 후, 상기 선택된 일부 분자 설명자를 사용하여 상기 정량적 구조-성능 관계식을 모델링할 수 있다.
- [0055] 여기서, 유전 알고리즘(GA)은 생물학적 체계에서 자연선택과 유전자의 다윈 진화론에 대한 직접적 유추에 기반을 둔 가장 인기 있는 최적화 알고리즘으로, 데이터 마이닝과 최적화 같은 다양한 문제들에 성공적으로 적용 가능하다.
- [0056] 본 발명에서는 종전의 정량적 구조-활성 관계식(quantitative structure-activity relationship, QSAR) 등에서 자주 사용하는 주로 2차원적인 분자설명자와는 달리, 이론적으로 계산이 가능한 모든 분자설명자를 사용하여 관계식을 모델링하기 때문에 보다 더 정확한 신약의 분자구조 묘사가 가능하다.
- [0057] 본 발명의 성능은 종전에 사용되던 생물학적 활성(biological activity) 이외에도 억제작용(inhibitory activity), 친유성(lipophilicity), 독성(toxicity), 대사 안전성(metabolic stability), 혈액-뇌 장벽 투과성(blood-brain barrier permeability) 등이 될 수 있고, 상기 복수의 성능과 분자 구조 사이의 관계식을 구하고 후속 단계에서 이를 사용할 수도 있다
- [0058] 본 발명에서는 4000여개 이상의 분자설명자를 사용하여 관계식을 모델링하기 때문에 회귀식 모델링과 함께 약의 성능 예측에 가장 큰 영향을 갖는 분자설명자 선택을 동시에 수행할 수 있다.
- [0059] 상기 최적 신약 분자 구조 획득 단계(S30)는 상기 정량적 구조-성능 관계식의 수치적 반전을 통해 최적의 신약 분자 구조를 획득하는 단계이다.
- [0060] 구체적으로, 상기 최적 신약 분자 구조 획득 단계(S30)는 하기의 [수학식 1]에 의한 수치적 반전을 통해 약의 성능(친유성)을 최대화 하는 최적의 신약 분자 구조를 획득할 수 있다.

수학식 1

$$\mathbf{x}^* = \operatorname{argmax} \log \hat{k}_w$$

$$s.t \log \hat{k}_w = \mathbf{C} \hat{\mathbf{t}}$$

$$\hat{\mathbf{t}} = \mathbf{P} \mathbf{x}$$

$$\hat{\mathbf{t}}^T \mathbf{S}_t^{-1} \mathbf{t} \leq c_1$$

$$\| \mathbf{P} \hat{\mathbf{t}} - \mathbf{x} \| \leq c_2$$

[0061]

[0062] $\mathbf{x}$: 신약의 분자설명자

- [0063] $\mathbf{x}^*$: 위의 수치계획법 수학식으로부터 계산된 최적의 신약의 분자설명자
- [0064] $\mathbf{C}$: 부분최소자승법(partial least squares, PLS)의 출력변수 로딩(loading) 행렬
- [0065] $\mathbf{t}$: PLS 입력변수(여기서는 분자설명자, $\mathbf{x}$) 스코어(score) 벡터
- [0066] $\mathbf{P}$: PLS 입력변수 로딩 행렬
- [0067] $\hat{\cdot}$: PLS 모델에 의해 예측된 값
- [0068] $\mathbf{S}_t$: $\mathbf{t}$ 의 표본 공분산 행렬
- [0069] c_1, c_2 : 적절한 상수
- [0070] 또한, 상기 최적 신약 분자 구조 획득 단계(S30)에서 하기의 [수학식 2]에 의한 수치적 반전을 통해 사용자가 지정한 성능, 즉 친유성 및 활성(activity, $\log k_i$)을 갖는 최적의 신약 분자 구조를 획득 할 수 있다.

수학식 2

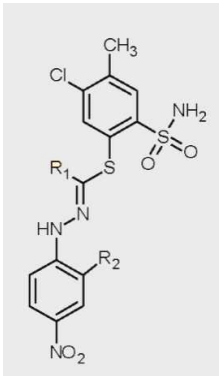
$$\begin{aligned} \mathbf{x}^* = \operatorname{argmax} & \sqrt{(\log k_w - \log k_{w,ref})^2 + (\log k_i - \log k_{i,ref})^2} \\ \text{s.t. } & [\log \hat{k}_w, \log \hat{k}_i]^T = \mathbf{C} \hat{\mathbf{t}} \\ & \hat{\mathbf{t}} = \mathbf{P} \mathbf{x} \\ & \hat{\mathbf{t}}^T \mathbf{S}_t^{-1} \hat{\mathbf{t}} \leq c_1 \\ & \|\mathbf{P} \hat{\mathbf{t}} - \mathbf{x}\| \leq c_2 \end{aligned}$$

- [0071]
- [0072] $\mathbf{x}$: 신약의 분자설명자
- [0073] $\mathbf{x}^*$: 위의 수치계획법 수학식으로부터 계산된 최적의 신약의 분자설명자
- [0074] $\mathbf{C}$: 부분최소자승법(partial least squares, PLS)의 출력변수 로딩(loading) 행렬
- [0075] $\mathbf{t}$: PLS 입력변수(여기서는 분자설명자, $\mathbf{x}$) 스코어(score) 벡터
- [0076] $\mathbf{P}$: PLS 입력변수 로딩 행렬
- [0077] $\hat{\cdot}$: PLS 모델에 의해 예측된 값
- [0078] $\mathbf{S}_t$: $\mathbf{t}$ 의 표본 공분산 행렬
- [0079] c_1, c_2 : 적절한 상수
- [0080] $\log k_{w,ref}$: 사용자가 지정한 친유성 값
- [0081] $\log k_{i,ref}$: 사용자가 지정한 활성 값
- [0082] 이처럼, 상기 최적 신약 분자 구조 획득 단계(S30)에서는 상기 [수학식1] 또는 [수학식 2]와 같이, 다양한 목적 함수를 사용하는 수치적 반전을 통해 사용자가 원하는 성능의 최적의 신약 분자 구조를 획득할 수 있다.
- [0083] 도 4는 본 발명에 따른 신약 후보군 스크리닝 단계의 블록도이다.
- [0084] 상기 신약 후보군 스크리닝 단계(S40)는 상기 최적의 신약 분자 구조와 유사한 신약 후보군을 스크리닝하는 단계이다.
- [0085] 구체적으로, 상기 신약 후보군 스크리닝 단계(S40)는 도 4에 도시된 바와 같이, 신약 후보군 등급부여 공정(S40) 및 신약 후보군 선택 공정(S41)을 포함한다.

- [0086] 신약 후보군 등급부여 공정(S40)은 상기 최적의 신약 분자 구조와 신약 후보군의 분자 구조의 유클리드 거리에 따라 신약 후보군에 등급을 부여하는 공정으로 유클리드 거리가 가까울수록 높은 등급을 부여한다.
- [0087] 또한, 신약 후보군 선택 공정(S41)은 상기 신약 후보군 중 일정 등급 이상의 신약 후보군을 선택하는 공정이다.
- [0088] 도 5는 본 발명에 따른 표적 신약 스크리닝 방법의 다른 블록도이다.
- [0089] 한편, 본 발명에 따른 정량적 구조-성능 관계식의 수치적 반전과 분자동역학 전산모사를 통한 표적신약의 스크리닝 방법은 도 5에 도시된 바와 같이, 상기 신약 후보군 스크리닝 단계(S40) 후, 표적 신약 검증 단계(S50)를 더 포함할 수 있다.
- [0090] 표적 신약 검증 단계(S50)는 상기 스크리닝 된 신약 후보군과 상기 표적분자의 복합물(complex)을 분자동역학 전산모사(molecular dynamics simulation)를 통해 검증하는 단계이다.
- [0091] 구체적으로, 상기 표적 신약 검증 단계(S50)에서는 상기 신약 후보군 선택 공정에서 선택된 일정 등급 이상의 신약 후보군과 표적분자의 복합물을 대상으로 해당 신약 후보를 검증함으로써 표적분자에 대한 최적의 표적신약을 스크리닝할 수 있다.
- [0092] 한편, 본 발명에서는 상기한 바와 같이, 분자동역학 전산모사를 이용하기 때문에 신약 후보군과 표적 분자의 고정된 분자구조가 아니라 다양한 구조 변화(conformational ensemble)를 고려할 수 있다.
- [0093] 이하, 본 발명의 내용을 실시예 및 실험예를 통하여 구체적으로 설명한다. 그러나, 이들은 본 발명을 보다 상세하게 설명하기 위한 것으로 본 발명의 권리범위가 이들에 의해 한정되는 것은 아니다.
- [0094] [실시예 : 탄산탈수효소 IX(CA IX, 분자표적)의 억제제를 위한 술폰 아마이드 유도체 설계 및 스크리닝]
- [0095] 본 발명을 탄산탈수효소 IX(CA IX, 분자표적)의 억제제를 위한 술폰 아마이드 유도체 설계에 적용하였다. 이 실시예에서는 친유성을 약의 성능으로 사용하였고 관계식 모델링은 부분최소자승법(partial least squares, PLS)를 사용하여 정량적 구조-친유성 관계(QSLR)를 모델링하였다.
- [0096] 액체크로마토그래피-질량 분석기(LC-MS)를 이용하여 미리 합성된 14개의 술폰 아마이드 화합물 이성질체의 친유성($\log k_w$) 값을 결정하였다(표 1).

표 1

[0097] 분석된 14개의 술폰 아마이드 화합물과 친유성 값.

구조	화합물	R ₁	R ₂	logk _w
	1EZ	Et	H	3.76
	2EZ	Pr	H	3.52
	3EZ	i-Pr	H	3.8
	4EZ	i-Pr	NO ₂	3.61
	5EZ	i-Bu	H	3.68
	6EZ	τ-Bu	H	3.64
	7EZ	n-C ₅ H ₁₁	H	3.54
	8EZ	Ph-CH ₂ -CH ₂	H	3.62
	9EZ	Ph	H	3.41
	10EZ	4-Cl-Ph	H	3.6
	11EZ	4-NO ₂ -Ph	H	3.56
	12EZ	4-Me-Ph	H	3.52
	13EZ	4-MeO-Ph	H	3.64
	14EZ	3-Br, 4-OH-Ph	H	3.49

[0098] 분자구조는 PM3 semiempirical Quantum Mechanics 방법을 사용하여 최적화하였으며, 4000 세트의 분자설명자 값을 계산하였다. 이 데이터를 학습군과 시험군으로 분리한 후 부분최소자승법과 결합된 유전알고리즘을 변수선택에 사용하여, 최종적으로 4 개의 분자설명자 (표 2)를 기반으로 한 정량적 구조-친유성 관계(QSLR) 모델을 개발하였다.

표 2

[0099]

QSLR 모델링을 위하여 선택된 분자설명자.

분자설명자	설명
R _{7e+}	Sanderson 전기 음성도에 의해 가중된 편차 7 의 R 최대 자기 상관
F10[C-C1]	n = 10의 위상 거리에서 (C-C1) 원자쌍의 주파수
H ⁶ⁱ	이온화 포텐셜에 의해 가중된 편차 6 의 H 자기 상관
MLOGP	소수성을 나타내는 모리구치 옥탄올-물 분배 계수

[0100]

도 6은 본 발명에 따른 정량적 구조-친유성 관계식(quantitative structure-lipophiicity relationships, QSLR) 모델의 성능을 나타내는 도이고, 도 7은 본 발명에 따른 QSLR 모델의 적용가능성 영역을 나타내는 도이다.

[0101]

정량적 구조-친유성 관계(QSLR) 모델은 적용가능한 범위 내에서 발견된 모든 화합물에 대해 도 6에 도시된 바와 같이, RMSEE(Root Mean Squared Error of Estimation)와 RMSEP(Root Mean Squared Error of Prediction)가 각각 0.67%와 0.97%로 매우 예측 가능하고, 또한 도 7에 도시된 바와 같이, 모든 학습 데이터(●로 표기)와 평가 데이터(○로 표기)가 관리 한계선 안에 위치하므로 안정적으로 나타났다.

[0102]

개발된 정량적 구조-친유성 관계(QSLR) 모델은 다음과 같이 제약조건을 갖는 수치 최적화를 사용하여 반전하였다.

$$\mathbf{x}^* = \operatorname{argmax} \log \hat{k}_w$$

$$s.t \log \hat{k}_w = \mathbf{C} \hat{\mathbf{t}}$$

$$\hat{\mathbf{t}} = \mathbf{P} \mathbf{x}$$

$$\hat{\mathbf{t}}^T \mathbf{S}_t^{-1} \mathbf{t} \leq c_1$$

$$\| \mathbf{P} \hat{\mathbf{t}} - \mathbf{x} \| \leq c_2$$

[0103]

[0104]

$\mathbf{x}$: 신약의 분자설명자

[0105]

$\mathbf{x}^*$: 위의 수치계획법 수학적식으로부터 계산된 최적의 신약의 분자설명자

[0106]

$\mathbf{C}$: 부분최소자승법(partial least squares, PLS)의 출력변수(여기서는 친유성) 로딩(loading) 행렬

[0107]

$\mathbf{t}$: PLS 입력변수(여기서는 분자설명자, $\mathbf{x}$) 스코어(score) 벡터

[0108]

$\mathbf{P}$: PLS 입력변수 로딩 행렬

[0109]

$\wedge$: PLS 모델에 의해 예측된 값

[0110]

$\mathbf{S}_t$: $\mathbf{t}$ 의 표본 공분산 행렬

[0111]

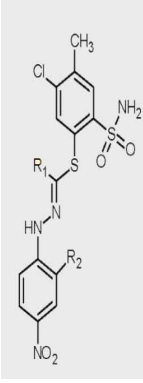
c_1, c_2 : 적절한 상수

[0112]

이를 통해 최적의 분자설명자를 얻었고 이전에 생성된 약물후보군의 데이터베이스에 대해 스크리닝했다. 약물후보는 이후 유클리드 거리에 따라 등급이 매겨졌다. 이 절차는 3.965의 최대 $\log k_w$ 값에 대해 11 개의 약물후보를 산출했다 (표 3).

표 3

[0113] 역 구조-성능 상관관계에 의해 설계된 이상 약물과 데이터베이스 내 후보 화합물간의 유클리드 거리.

x*				R _{7e+}	F10[C-C1]	H _{6i}	MLOGP	
Max. log kw = 3.965				0.025	5	1.040	4.529	
구조		R ₁	R ₂	데이터베이스로부터 측정된 X값				거리
C45		4-Cl-butyl	H	0.020	5	0.953	3.886	0.649
C46		4-Cl-butyl	NO ₂	0.022	5	1.205	3.790	0.757
C65		3-Cl,butoxy	H	0.015	6	0.967	3.959	1.153
C66		3-Cl,butoxy	NO ₂	0.017	6	1.241	3.885	1.206
C64		2-Cl,ethoxy	NO ₂	0.029	4	1.039	3.426	1.489
C63		2-Cl,ethoxy	H	0.021	4	0.539	3.499	1.520
C56		3-Cl,4-OH-phenyl	NO ₂	0.015	4	0.748	3.293	1.616
C55		3-Cl,4-OH-phenyl	H	0.024	4	0.500	3.366	1.626
C59		3-Cl,4-OH-butyl	H	0.018	4	1.474	3.147	1.760
C51		4-Me-butyl	H	0.023	3	1.206	3.886	2.107
C34		pentyl	NO ₂	0.017	3	0.756	3.715	2.178

[0114] 도 8은 본 발명에 따른 술폰 아마이드-CA(Carbonic anhydrase) IX 복합체의 분자동역학 전산모사(molecular dynamics simulation)를 나타내는 도이다.

[0115] 산출된 약물후보 화합물을 검증하기 위해 도 8에 도시된 바와 같이, 술폰 아마이드-CA IX 복합체의 분자동역학 전산모사(molecular dynamics simulation)를 사용하였다. 이때, 사람의 체액 안에서 복합체의 움직임을 모사하기 위해 복합체가 물 분자로 둘러싸여 12면체의 모양을 하고 있다고 가정하였다.

[0116] 분자동역학 전산모사(molecular dynamics simulation)를 통해, Zn²⁺ 이온은 활성 부위의 3 개의 히스티딘 잔기와 배위되어 유지되었다. 이것은 CA IX 효소의 기능이 모사 내에서 방해받지 않도록 보장하기 위해 수행되었다.

[0117] 도 9는 본 발명에 따른 CA IX의 안정성을 나타내는 도이다.

[0118] 효소의 안정성과 유연성은 RMSD(root mean square deviation)와 RMSF(root mean square fluctuation)를 계산하여 분석하였다. 이때 수소 결합, 소수성 및 친수성 상호작용도 분석되었다.

[0119] 도 9에 도시된 바와 같이, 11 개의 복합체 중 C65, C64, C63 리간드가 있는 복합체의 경우 5-(1-나프탈렌-1-일-1,2,3-트리아졸-4-일)티오펜-2-설폰아미드(5-(1-naphthalen-1-yl-1,2,3-triazol-4-yl)thiophene-2-sulfonamide, CA IX-9FK)와 비슷한 RMSD 값 (약 0.1 nm)이 계산되었지만, 다른 복합체(약 0.2 nm)의 경우 더 높은 값이 계산되었다.

[0120] 이러한 불일치는 9FK 술폰 아마이드의 크기가 작기 때문이며, 스크린된 화합물은 부피가 크기 때문에, 효소는 활성 위치 내에서 화합물을 수용할 수 있는 방식으로 그 형태와 일치해야 한다.

[0121] 도 10은 본 발명에 따른 CA IX의 유연성을 나타내는 도이다.

[0122] 모든 복합체에 대해 RMSF (root mean square fluctuation)를 계산하여 도 10 에 도시된 바와 같이, CA IX의 구조적 변동도 평가하였다.

[0123] 결과는 효소가 CA IX-9FK 시스템 내에서 가장 안정하다는 것을 보여주었다. 다른 모든 복합체는 유사한 패턴의 RMSF 모양을 공유한다. 그러나 대부분은 (i) 유연한 N 말단(잔기 9에서 잔기 20) 및 (ii) 유연한 루프(잔기 230 내지 240)의 두 개의 전형적인 영역에서 더 높은 RMSF 값을 나타내었다.

[0124] CA IX 구조 내에서 발견된 복잡한 수소 결합 네트워크를 분석하여 RMSF 값 차이의 원인을 파악하였다. 이를 위해, 모사를 통해 수소 결합 형성의 백분율이 계산되었다(표 4). 수소 결합이 W9-H68, A133-R136 및 R136-G139 결합 쌍 사이에 형성되었음을 알 수 있다.

표 4

[0125]

모사 중 형성된 수소결합의 백분율.

Bond pair	C45	C46	C65	C66	C64	C63	C56	C55	C59	C51	C34	9FK
W9 - H68	0.04	1.73	67.01	2.77	35	21.77	9.09	0.41	26.07	48.06	0	80.45
A133 - R136	43.55	49.71	49.04	36.07	48.71	38.48	8.12	44.04	12.07	52.3	2.69	40.09
R136 - G139	46.17	45.43	47.47	30.65	51.81	32.83	16.7	38.64	2.12	55.2	0.96	38.04
S237 - P234	56.47	44.01	64.93	30.11	33.26	0	67.82	63.49	67.11	67.17	46.47	92.14
S237 - G233	69.42	0	66.81	76.46	34.86	78.51	65.51	55.86	55.51	46.92	68.82	83.82
L249 - R252	46.75	66.86	3.66	66.4	70.65	20.36	69.95	85.47	68	89.04	80.53	82

[0126]

CA IX의 활성 부위 내에서 설계된 솔폰 아마이드 유도체의 안정성은 상호 작용에 상당한 영향을 미친다. 이것이 효소의 구조와 유연성의 변화가 고려된 이유이다. 그들의 유연성은 처음 22.06 ns 후 구조물에 구조물을 중첩할 때 RMSD (root mean square deviation) 값을 분석하여 관찰되었다. 결과는 모든 리간드가 활성 사이트에서 상당히 안정한 것으로 나타났다.

[0127]

CA IX와 설계된 솔폰 아마이드 유도체 및 9FK 사이의 상호작용 분석은 특정 아미노산 잔기 (표 5)와의 리간드 상호 작용의 백분율을 계산하고, 시뮬레이션 전반에 걸쳐 리간드의 0.35 nm 내에서 평균 원자 수 및 잔류물 수를 계산함으로써 달성되었다(표 6).

표 5

[0128]

CA IX와 약물후보군 간 상호작용.

Residue	C45	C46	C65	C66	C64	C63	C56	C55	C59	C51	C34	9FK
T200	60.47	92.22	78.73	n.a.	98.02	82.95	98.02	100	99.19	n.a.	99.34	100
E106	n.a.	100	n.a.	n.a.	n.a.	32.24	97.01	100	n.a.	n.a.	100	100
Q92	99.94	n.a.	38.33	84.48	99.42	93.23	95.43	62.05	n.a.	91.77	80.81	71.97
W210	n.a.	31.74	30.7	n.a.	81.86	n.a.	n.a.	65.91	100	96.02	99.8	64.09
L134	66.68	n.a.	53.46	n.a.	n.a.	n.a.	n.a.	n.a.	71.83	58.6	n.a.	59.09
P203	83.41	n.a.	62.94	n.a.	44.02	n.a.	n.a.	n.a.	n.a.	51.52	n.a.	46.94

표 6

[0129]

CA IX와 약물후보군 간 상호작용.

Measure	C45	C46	C65	C66	C64	C63	C56	C55	C59	C51	C34	9FK
Number of atoms within 0.35 nm of ligand	29.19 ± 3.46	35.78 ± 4.23	29.13 ± 3.06	22.28 ± 5.53	28.12 ± 3.56	24.01 ± 2.89	32.70 ± 4.07	35.06 ± 4.89	45.79 ± 4.23	23.49 ± 3.28	45.12 ± 4.35	18.39 ± 3.26
Number of protein residues within 0.35 nm of ligand	13.5 ± 1.54	14.5 ± 1.53	17.18 ± 1.70	10.08 ± 1.58	12.82 ± 1.76	13.56 ± 1.49	16.06 ± 1.52	16.66 ± 1.66	18.14 ± 1.60	13.09 ± 1.56	16.62 ± 1.49	9.41 ± 1.60

[0130]

결정학 연구에 의해 결정된 9FK 사이의 상호 작용은 모사 내내 유지된다. C65 리간드는 E106과의 상호 작용을 제외하고는 매우 유사한 특징을 나타낸다. 반대로 C66은 CA IX-9FK 복합체와 가장 다른 복합체이다. 리간드와 상호 작용하는 7 개의 잔기 중 2 개만 보존된다. 다른 복합체의 시뮬레이션을 위해 더 강한 소수성 및 친수성 상호 작용이 관찰되었다.

[0131]

상기한 바와 같이, 실시예에 따르면, 14 개의 새로운 약물후보가 역 구조-친유성 상관관계에 의해 제안되었고 이들 중 11 개에 대한 분자동역학 전산모사(molecular dynamics simulation)가 수행되었으며, 모두 기준 솔폰 아마이드 화합물보다 CA IX의 저해에 더 적합한 것으로 밝혀졌다.

[0132]

또한 치환 그룹과 활성 부위 사이의 모든 소수성 및 친수성 상호 작용을 면밀히 분석하였다. CA IX-리간드 복합체를 결정화하는 것이 효소의 복잡한 막 결합 구조로 인하여 매우 힘들기 때문에 미래 합성에 대한 통찰력과 지

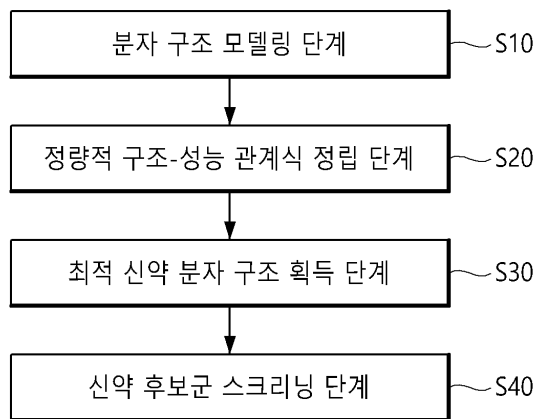
침을 제공할 수 있다.

[0133] 결과를 바탕으로, 두 개의 화합물 (i) (5-클로로-4-메틸-2-설파모일-페닐)(1E)-4-클로로-5-히드로시-N-(4-메틸아닐리노)펜탄이미도티오에이트 (5-chloro-4-methyl-2-sulfamoyl-phenyl)(1E)-4-chloro-5-hydroxy-N-(4-methylanilino)pentanimidothioate, C59) 및 (ii) (5-클로로-4-메틸-2-설파모일-페닐)(1E)-N-(4-메틸-2-니트로-아닐리노)헥사이미도티오에이트 (5-chloro-4-methyl-2-sulfamoyl-phenyl)(1E)-N-(4-methyl-2-nitro-anilino)hexanimidothioate, C34) 이 다음 단계의 생체외 및 생체내 실험을 위한 실제 합성에 특히 유망한 것으로 나타났다.

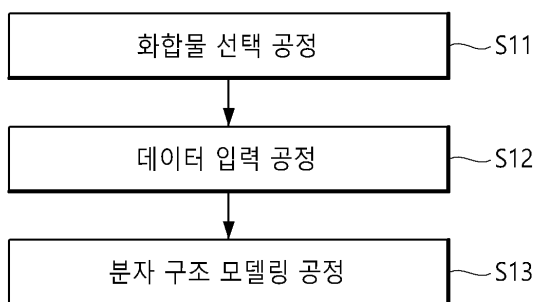
[0134] 상기한 바와 같이, 본 발명에 따른 표적신약 스크리닝 방법은 빠르고 효율적인 표적신약 개발에 기여할 수 있다. 즉, 정량적 구조-성능 관계식의 수치적 반전과 분자동역학 전산모사(molecular dynamics simulation)의 단순성 및 속도는 약물 개발에서의 소모 비용을 상당히 줄이고 잘못된 합성과 불필요한 제약 테스트를 피할 수 있게 한다.

도면

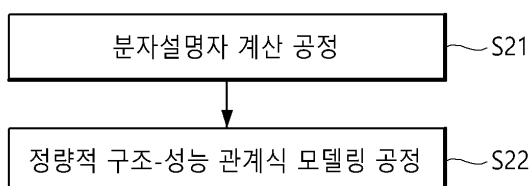
도면1



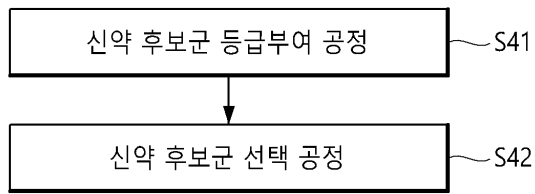
도면2



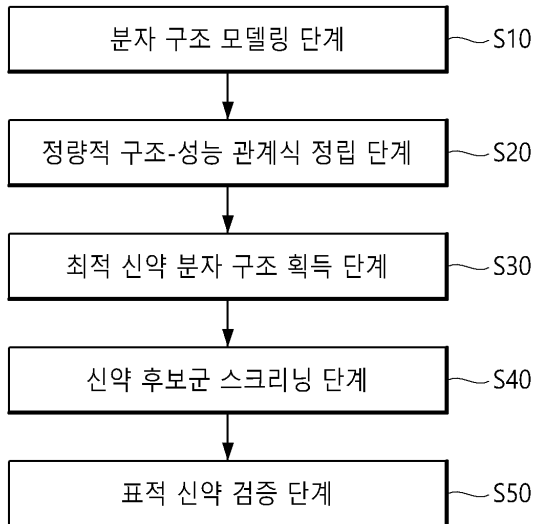
도면3



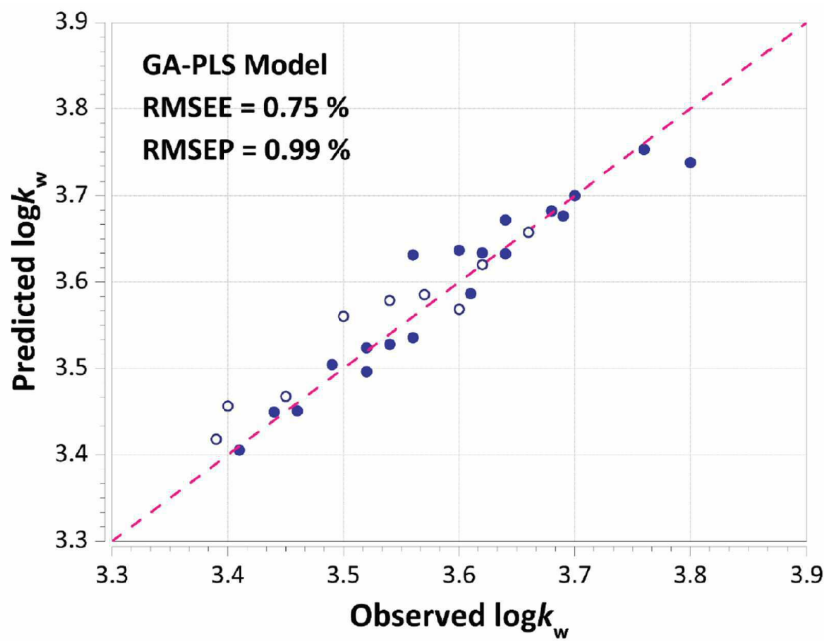
도면4



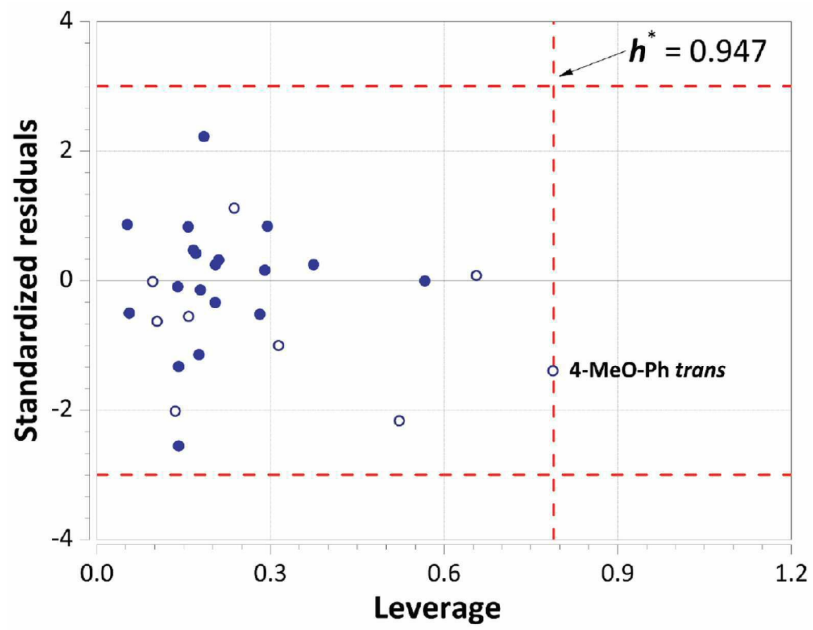
도면5



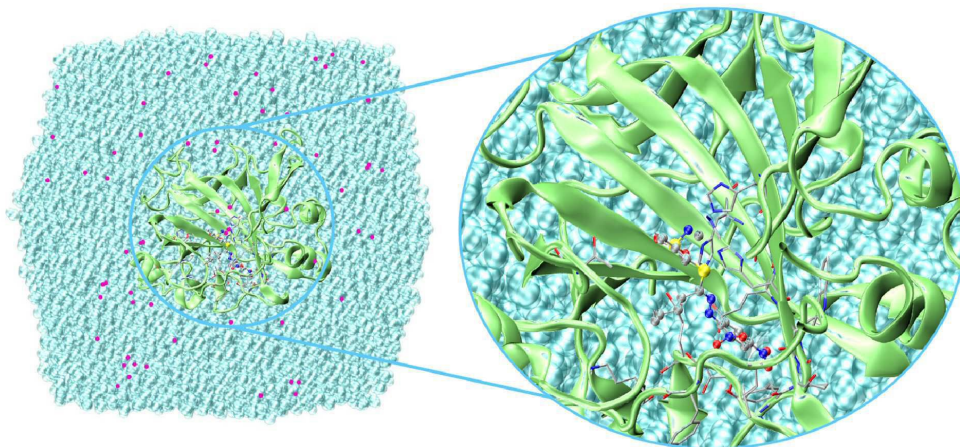
도면6



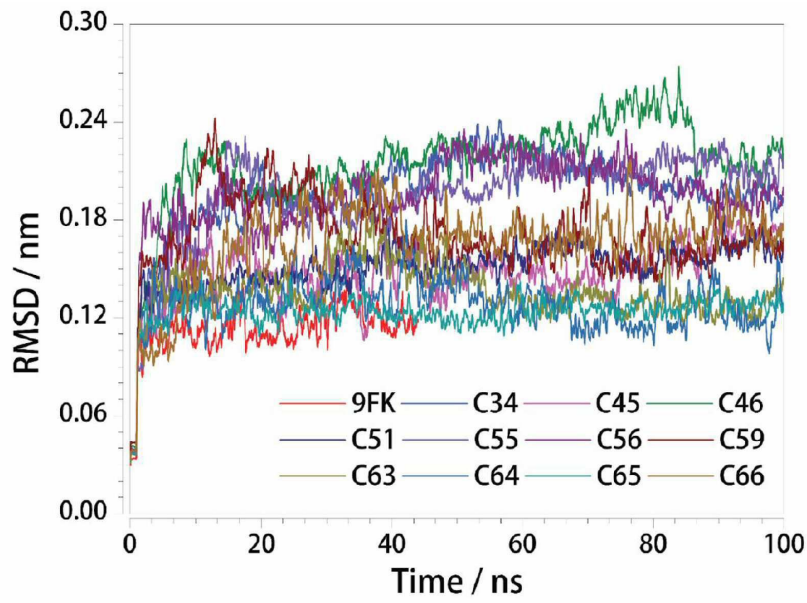
도면7



도면8



도면9



도면10

