

【서지사항】

【서류명】	특허출원서
【참조번호】	P20240517KR
【출원구분】	특허출원
【출원인】	
【명칭】	연세대학교 산학협력단
【특허고객번호】	2-2005-009509-9
【대리인】	
【명칭】	특허법인 지담
【대리인번호】	9-2018-100261-1
【지정된변리사】	김재홍
【포괄위임등록번호】	2023-020263-8
【발명의 국문명칭】	신약 후보 물질 도출을 위한 파마코포어(Pharmacophore) 기반의 3차원 분자 생성 방법 및 시스템
【발명의 영문명칭】	PHARMACOPHORE-BASED 3D MOLECULE GENERATION SYSTEM FOR DRUG CANDIDATE DISCOVERY AND METHOD THEREOF
【발명자】	
【성명】	박상현
【성명의 영문표기】	PARK, Sang Hyun
【주민등록번호】	670101-1XXXXXX
【우편번호】	03626
【주소】	서울특별시 서대문구 세무서8길, 101동 1502호
【발명자】	

【성명】	최승연
【성명의 영문표기】	CHOI, Seung Yeon
【주민등록번호】	970828-1XXXXXX
【우편번호】	07213
【주소】	서울특별시 영등포구 당산로45길 7-3, 101동 808호
【발명자】	
【성명】	서상민
【성명의 영문표기】	SEO, Sang Min
【주민등록번호】	930507-1XXXXXX
【우편번호】	63272
【주소】	제주특별자치도 제주시 고마로 44, 801호
【출원언어】	국어
【심사청구】	청구
【이 발명을 지원한 국가연구개발사업】	
【과제고유번호】	1711198526
【과제번호】	00229822
【부처명】	과학기술정보통신부
【과제관리(전문)기관명】	한국연구재단
【연구사업명】	원천기술개발사업
【연구과제명】	[통합이지바로/주관] 난치성 질환 극복을 위한 인공지능 기반의 다중 약물 적응증 최적화 플랫폼 개발 및 혁신신약 발굴 (1/2단계)(1/2)

【과제수행기관명】 연세대학교

【연구기간】 2024.01.01 ~ 2024.12.31

【취지】 위와 같이 특허청장에게 제출합니다.

대리인 특허법인 지담 (서명 또는 인)

【수수료】

【출원료】 0 면 46,000 원

【가산출원료】 43 면 0 원

【우선권주장료】 0 건 0 원

【심사청구료】 14 항 880,000 원

【합계】 926,000원

【감면사유】 전담조직(50%감면)[1]

【감면후 수수료】 463,000 원

【발명의 설명】

【발명의 명칭】

신약 후보 물질 도출을 위한 파마코포어(Pharmacophore) 기반의 3차원 분자 생성 방법 및 시스템{PHARMACOPHORE-BASED 3D MOLECULE GENERATION SYSTEM FOR DRUG CANDIDATE DISCOVERY AND METHOD THEREOF}

【기술분야】

【0001】 본 발명은 신약 후보 물질 도출을 위한 파마코포어(Pharmacophore)를 기반의 3차원 분자 생성 방법 및 시스템에 관한 것으로, 더욱 상세하게는 단백질-리간드 상호작용을 고려하여 타겟 단백질의 결합 포켓과 결합 가능한 3차원 리간드 분자 구조를 효율적으로 설계하고 최적화하는 기술이다.

【발명의 배경이 되는 기술】

【0002】 기존의 신약 설계 및 개발 과정은 타겟 단백질과 결합할 수 있는 후보 리간드(Ligand)를 탐색하고 설계하는 데 많은 시간과 자원이 소요되며, 높은 비용이 발생하는 문제가 있었다. 특히, 기존 기술들은 주로 리간드를 2차원 또는 1차원 구조로 표현하여 분석했는데, 이러한 방식은 단백질-리간드 간의 상호작용(Protein-Ligand Interaction)을 충분히 반영하지 못했다. 이로 인해 결합 친화도가 낮거나 불안정한 리간드가 설계되었으며, 이는 신약 개발의 효율성을 크게 저하시켰다.

【0003】 전통적으로 사용된 신약 개발 모델인 "열쇠-자물쇠 모델"은 단백질과 리간드 간의 단순한 구조적 일치에 의존하여 결합을 예측했다. 이러한 접근은 단백질-리간드 상호작용을 기하학적(Geometry) 또는 위상학적(Topology) 관점에서만 설명했으며, 실제 복잡한 화학적 상호작용을 충분히 반영하지 못했다. 특히, 리간드의 특정 약리학적 작용점(Pharmacophore)을 고려하지 않거나, 단백질 결합 포켓과 리간드 간의 비공유 결합 및 유연성을 설명하지 못하는 한계가 있었다. 이러한 접근으로 생성된 리간드는 실제 생체 환경에서 예측한 약리학적 효능에 도달하지 못할 가능성이 높고, 결합 친화도가 낮거나 예측이 부정확한 신약 후보 물질이 도출될 위험이 컸다.

【0004】 상기 문제를 해결하기 위해 최근에서는 3차원 리간드 생성 모델을 개발하여 파마코포어 특징을 반영한 리간드 구조를 설계하려는 시도가 이루어지고 있다. 3차원 리간드 모델은 단백질-리간드 상호작용을 보다 정밀하게 분석할 수 있도록 설계되었으나, 기존 모델들도 여전히 파마코포어(Pharmacophore)를 정확히 인식하지 못하거나, 파마코포어와 단백질 결합 포켓(Binding Pocket) 간의 복잡한 상호작용을 충분히 반영하지 못하는 문제가 있었다. 또한, 모델 학습 효율성이 떨어져 신약 개발 과정에서 실시간으로 활용하기 어렵다는 한계가 존재한다.

【0005】 따라서, 단백질-리간드 간의 복잡한 화학적 상호작용을 정확하게 모델링할 수 있는 기술에 대한 요구가 증가하고 있다. 특히, 파마코포어를 명확하게 인식하고 이를 기반으로 단백질과의 결합력을 극대화할 수 있는 신약 후보 물질을 효율적으로 설계할 수 있는 기술이 필요하다. 이를 위해서는 기존의 2차원 또는 1

차원 방식에서 벗어나, 리간드의 3차원 구조를 정밀하게 생성하고, 단백질의 특정 결합 부위에 최적화된 리간드 분자를 설계할 수 있는 새로운 기술이 개발이 필요하다.

【선행기술문헌】

【특허문헌】

【0006】(특허문헌 0001) 한국 등록특허공보 제10-2472724호 (2022. 11. 25.)

(특허문헌 0002) 한국 공개특허공보 제10-2024-0084664호 (2024. 06. 14.)

(특허문헌 0003) 한국 등록특허공보 제10-2558546호 (2023. 07. 18.)

【발명의 내용】

【해결하고자 하는 과제】

【0007】본 발명이 이루고자 하는 기술적 과제는 단백질과 리간드의 결합 상호작용에서 중요한 파마코포어를 인식하고, 파마코포어를 기초로 하여 단백질과 높은 결합 친화도를 갖는 3차원 리간드를 효율적으로 생성하는 방법 및 시스템을 제공하는 것이다.

【0008】또한, 본 발명이 이루고자 하는 기술적 과제는 생성된 리간드가 실제 생체 환경에서 안정적으로 결합할 수 있도록 리간드 구조의 물리적 안정성을 개선하는 방법을 제공하는 것이다.

【0009】 또한, 본 발명이 이루고자 하는 기술적 과제는 파마코포어와 결합 포켓 간의 상호작용을 정확히 반영하여, 리간드의 결합 예측 정확도를 향상시키는 시스템을 제공하는 것이다.

【0010】 본 발명이 이루고자 하는 기술적 과제는 이상에서 언급한 기술적 과제로 제한되지 않으며, 언급되지 않은 또 다른 기술적 과제들은 아래의 기재로부터 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자에게 명확하게 이해될 수 있을 것이다.

【과제의 해결 수단】

【0011】 상기 기술적 과제를 달성하기 위하여 본 발명의 실시예에 따른 신약 후보 물질 도출을 위한 파마코포어(Pharmacophore) 및 확산 생성 모델을 이용하여 리간드의 3차원 분자 구조를 생성하는 방법은, 단백질의 결합 포켓 데이터를 수집하여 입력하는 단계; 리간드의 파마코포어를 인식 및 식별하는 단계; 상기 결합 포켓 데이터에 기초하여 파마코포어를 고정된 기준으로 설정하고, 나머지 리간드 구조를 형성하기 위해 SE(3)-동치 신경망(SE(3)-Equivalent Neural Network)과 비자기회귀적(Non-Autoregressive) 확산 생성 모델을 적용하여 리간드의 3차원 구조를 생성하는 단계; 상기 생성된 리간드의 결합 친화도 및 구조적 유사성을 평가하는 단계; 상기 평가 결과에 따라 리간드의 3차원 구조를 최적화하여 최종 리간드를 생성하는 단계를 포함한다.

【0012】 여기서, 상기 파마코포어 인식 및 식별 단계는, PLIP(Protein-Ligand Interaction Profiler)를 이용하여 단백질과의 상호작용이 높은 리간드 부위로 인식 및 식별될 수 있다.

【0013】 또한, 상기 비자기회귀적(Non-Autoregressive) 확산 생성 모델은 정방향 과정(Forward Process)에 의하여 리간드에 노이즈를 추가하고, 역방향 과정(Reverse Process)에 의하여 점진적으로 노이즈를 제거하여 결합 포켓과 최적화된 3차원 리간드 구조를 형성할 수 있다.

【0014】 또한, 상기 SE(3)-동치 신경망(SE(3)-Equivalent Neural Network)은 리간드의 회전 및 이동에 영향을 받지 않도록 설계되어, 단백질 결합 포켓에 적합한 3차원 구조를 형성할 수 있다.

【0015】 또한, 상기 파마코포어 인식 및 식별 단계는, 단백질-리간드 복합체로부터 수집한 결합 정보를 이용하여 결합에 중요한 리간드 원자의 인덱스를 상호작용 집합으로 정의하고, 상기 상호작용 집합 내 원자의 위치 및 결합 상태를 파마코포어로 선정할 수 있다.

【0016】 또한, 상기 파마코포어 인식 및 식별 단계는, 리간드 문자를 다수의 조각으로 분해한 후, 각 조각의 원자 집합(A_i)과 원자 간의 엣지 정보(E_i)를 사용하여 단백질과 결합 가능성이 높은 조각을 선택하고, 상기 조각을 병합하여 파마코포어를 형성할 수 있다.

【0017】 또한, 상기 파마코포어 인식 및 식별 단계는, 파마코포어로 정의된 조각들 간의 상호작용을 평가하여 최종적으로 단백질과의 결합 친화도를 극대화하는 파마코포어 리스트를 생성하고, 상기 파마코포어 리스트를 기반으로 비자기회귀적 확산 생성 모델을 이용하여 리간드의 나머지 구조를 최적화할 수 있다.

【0018】 상기 기술적 과제를 달성하기 위하여 본 발명의 실시예에 따른 신약 후보 물질 도출을 위한 파마코포어(Pharmacophore) 및 확산 생성 모델을 이용하여 리간드의 3차원 문자 구조를 생성하는 시스템은, 단백질의 결합 포켓 데이터와 파마코포어 데이터를 입력하는 데이터 입력부; 상기 데이터를 기초로 하여 리간드의 파마코포어를 인식 및 식별하며, 상기 결합 포켓 데이터에 기초하여 파마코포어를 고정된 기준으로 설정하고, 나머지 리간드 구조를 형성하기 위해 SE(3)-동치 신경망(SE(3)-Equivalent Neural Network)과 비자기회귀적(Non-Autoregressive) 확산 생성 모델을 적용하여 리간드의 3차원 구조를 생성하는 파마코포어 기반 인공지능 모델 및 상기 생성된 리간드의 결합 친화도 및 구조적 유사성을 평가에 따라 리간드의 3차원 구조를 출력하는 문자 구조 출력부를 포함할 수 있다.

【0019】 여기서, 상기 파마코포어 기반 인공지능 모델은, PLIP(Protein-Ligand Interaction Profiler)를 이용하여 단백질과의 상호작용이 높은 리간드 부위로 인식 및 식별할 수 있다.

【0020】 또한, 상기 비자기회귀적(Non-Autoregressive) 확산 생성 모델은, 정방향 과정(Forward Process)에 의하여 리간드에 노이즈를 추가하고, 역방향 과정(Reverse Process)에 의하여 점진적으로 노이즈를 제거하여 결합 포켓과 최적화된

3차원 리간드 구조를 형성할 수 있다.

【0021】 또한, 상기 SE(3)-동치 신경망(SE(3)-Equivalent Neural Network)은, 리간드의 회전 및 이동에 영향을 받지 않도록 설계되어, 단백질 결합 포켓에 적합한 3차원 구조를 형성할 수 있다.

【0022】 또한, 상기 파마코포어 기반 인공지능 모델은, 단백질-리간드 복합체로부터 수집한 결합 정보를 이용하여 결합에 중요한 리간드 원자의 인덱스를 상호작용 집합으로 정의하고, 상기 상호작용 집합 내 원자의 위치 및 결합 상태를 파마코포어로 선정할 수 있다.

【0023】 또한, 상기 파마코포어 기반 인공지능 모델은, 리간드 분자를 다수의 조각으로 분해한 후, 각 조각의 원자 집합(A_i)과 원자 간의 엣지 정보(E_i)를 사용하여 단백질과 결합 가능성이 높은 조각을 선택하고, 상기 조각을 병합하여 파마코포어를 형성할 수 있다.

【0024】 또한, 상기 파마코포어 기반 인공지능 모델은, 파마코포어로 정의된 조각들 간의 상호작용을 평가하여 최종적으로 단백질과의 결합 친화도를 극대화하는 파마코포어 리스트를 생성하고, 상기 파마코포어 리스트를 기반으로 비자기회귀적 확산 생성 모델을 이용하여 리간드의 나머지 구조를 최적화할 수 있다.

【발명의 효과】

【0025】 본 발명의 실시예에 따르면, 신약 후보 물질 도출을 위한 파마코포어 기반의 3차원 문자 생성 방법 및 시스템은 파마코포어를 기반으로 한 인공지능

모델을 통해 결합 친화도가 높은 리간드 문자 구조를 효율적으로 생성할 수 있으므로, 약물 개발 시간을 크게 단축시킬 수 있다.

【0026】 또한, 본 발명의 실시예에 따르면, 문자 동역학 시뮬레이션을 통해 생성된 리간드 문자의 생체 내 안정성을 검증할 수 있어, 실제 약물로서의 사용 가능성을 높일 수 있다.

【0027】 본 발명의 효과는 상기한 효과로 한정되는 것은 아니며, 본 발명의 상세한 설명 또는 특허청구범위에 기재된 발명의 구성으로부터 추론 가능한 모든 효과를 포함하는 것으로 이해되어야 한다.

【도면의 간단한 설명】

【0028】 도 1은 신약 후보 물질 도출을 위한 파마코포어 기반의 3차원 문자 생성 방법 및 시스템을 도시한 개요도이다.

도 2는 본 발명의 실시예에 따른 파마코포어 기반 3D 문자 생성 시스템의 구성을 도시한 도면이다.

도 3은 본 발명의 실시예에 따른 파마코포어 기반 인공지능 모델의 생성 및 학습 방법을 도시한 도면이다.

도 4는 본 발명의 실시예에 따른 파마코포어 기반 인공지능 모델이 리간드의 3차원 구조를 생성하는 방법을 도시한 도면이다.

도 5는 본 발명의 실시예에 따른 3D 파마코포어 인식 및 식별 알고리즘을 도시한 도면이다.

도 6은 본 발명의 실시예에 따른 파마코포어 기반 인공지능 모델의 작동 방식을 도시한 도면이다.

도 7은 본 발명의 실시예에 따른 파마코포어 기반 인공지능 모델에 의한 결합 문자를 시각화하여 도시한 도면이다.

도 8은 본 발명의 실시예에 따른 문자 생성 시스템이 3차원 구조의 리간드 문자 생성하는 과정을 도시한 도면이다.

도 9는 본 발명의 실시예에 따른 문자 생성 시스템의 PharDiff 모델(파마코포어 기반 인공지능 모델)이 생성한 문자의 성능을 결합 친화도와 구조적 유사성 측면에서 평가한 참조도이다.

도 10은 본 발명의 실시예에 따른 디스크립터 생성 방법 및 생성 장치를 구현하는 컴퓨팅 장치를 도시한다.

【발명을 실시하기 위한 구체적인 내용】

【0029】 이하에서는 첨부한 도면을 참조하여 본 발명을 설명하기로 한다. 그러나 본 발명은 여러 가지 상이한 형태로 구현될 수 있으며, 따라서 여기에서 설명하는 실시예로 한정되는 것은 아니다. 그리고 도면에서 본 발명을 명확하게 설명하기 위해서 설명과 관계없는 부분은 생략하였으며, 명세서 전체를 통하여 유사한 부분에 대해서는 유사한 도면 부호를 붙였다.

【0030】 명세서 전체에서, 어떤 부분이 다른 부분과 "연결(접속, 접촉, 결합)"되어 있다고 할 때, 이는 "직접적으로 연결"되어 있는 경우뿐 아니라, 그 중

간에 다른 부재를 사이에 두고 "간접적으로 연결"되어 있는 경우도 포함한다. 또한, 어떤 부분이 어떤 구성요소를 "포함"한다고 할 때, 이는 특별히 반대되는 기재가 없는 한 다른 구성요소를 제외하는 것이 아니라 다른 구성요소를 더 구비할 수 있다는 것을 의미한다.

【0031】 본 명세서에서 사용한 용어는 단지 특정한 실시예를 설명하기 위해 사용된 것으로, 본 발명을 한정하려는 의도가 아니다. 단수의 표현은 문맥상 명백하게 다르게 뜻하지 않는 한, 복수의 표현을 포함한다. 본 명세서에서, "포함하다" 또는 "가지다" 등의 용어는 명세서상에 기재된 특징, 숫자, 단계, 동작, 구성요소, 부품 또는 이들을 조합한 것이 존재함을 지정하려는 것이지, 하나 또는 그 이상의 다른 특징들이나 숫자, 단계, 동작, 구성요소, 부품 또는 이들을 조합한 것들의 존재 또는 부가 가능성을 미리 배제하지 않는 것으로 이해되어야 한다.

【0032】 본 명세서에서, "모듈"은 하드웨어, 소프트웨어 또는 펌웨어로 구성된 유닛을 포함하며, 예컨대 로직, 논리 블록, 부품, 또는 회로 등의 용어와 상호호환적으로 사용될 수 있다. 모듈은 일체로 구성된 부품 또는 하나 또는 그 이상의 기능을 수행하는 최소 단위 또는 그 일부가 될 수 있다. 예컨대 모듈은 ASIC(application-specific integrated circuit)으로 구성될 수 있다.

【0033】 본 발명의 실시예에서 언급하는 “리간드(Ligand)”란, 단백질의 결합 포켓에 결합하여 생리적 반응을 유도하거나 억제하는 분자를 의미한다. 리간드는 표적 단백질과의 선택적 결합을 통해 약물의 효능을 나타내며, 약물 개발에서 중요한 구성 요소이다.

【0034】 본 발명의 실시예에서 언급하는 “파마코포어(Pharmacophore)”란, 특정 생체 분자와 결합하기 위해 리간드가 가져야 할 필수적인 화학적 작용점을 의미한다. 파마코포어는 결합의 선택성과 친화도를 결정하는 중요한 특징을 포함하며, 리간드가 표적 단백질에 적합하게 결합할 수 있도록 하는 구조적 요소이다.

【0035】 이하 첨부된 도면을 참고하여 본 발명의 실시예를 상세히 설명하기로 한다.

【0036】 도 1은 신약 후보 물질 도출을 위한 파마코포어(Pharmacophore) 기반의 3차원 분자 생성 방법 및 시스템을 도시한 개요도이다. 구체적으로, 파마코포어 기반 인공지능 모델(PharDiff, 120)을 이용하여 단백질의 결합 포켓 정보를 기반으로 최적화된 리간드 구조를 생성하는 개요를 도시한다. 즉, 도 1은 단백질의 결합 포켓 정보가 입력되고, 상기 정보가 인공지능 모델을 통해 처리되어, 최종적으로 결합에 적합한 리간드가 출력되는 과정을 나타낸다.

【0037】 도 1에 도시된 바와 같이, 상기 단백질 포켓(20)은 단백질 표면에 리간드가 결합할 수 있는 특정 부위를 나타내며, 약물 개발에서 중요한 타겟이 된다. 상기 단백질 포켓의 구조적 및 화학적 정보는 파마코포어 기반 인공지능 모델(120)에 입력되며, 리간드가 해당 부위에 맞춰 최적화된 형태로 생성되도록 한다.

【0038】 파마코포어 기반 인공지능 모델(120)은 상기 단백질 포켓 정보와 파마코포어를 활용하여 리간드의 3차원 구조를 생성하는 인공지능 모델이다. 상기 파

마코포어 기반 인공지능 모델(120)은 단백질과 리간드 간의 상호작용을 분석하고, 단백질 포켓에 최적화된 리간드 구조를 설계할 수 있다. 상기 파마코포어 기반 인공지능 모델(120)은 결합에 필요한 특정 작용점인 파마코포어를 기준으로, 리간드가 상기 단백질 포켓과 잘 맞도록 나머지 구조를 형성합니다. 상기 파마코포어 기반 인공지능 모델(120)은 확산 모델과 SE(3)-동치 신경망을 적용하여 3D 구조 생성 과정에서 리간드의 회전 및 이동에 영향을 받지 않도록 설계된다.

【0039】 리간드(30)는 상기 파마코포어 기반 인공지능 모델(120)에서 생성된 3차원 구조로, 단백질의 결합 포켓에 적합한 형태로 생성된 결과물이다. 상기 리간드 (Ligand)는 약물의 효능을 발휘하는 주요 분자로, 단백질과의 결합을 통해 생리적 반응을 유도하거나 억제하는 역할을 한다. 상기 파마코포어 기반 인공지능 모델(120)에 의하여 생성된 리간드는 단백질 결합에 필요한 모든 필수 요소를 포함하여 결합 친화도가 높고 안정적인 구조로 설계된다.

【0040】 또한, 종래 기술의 단순한 기계적 맞춤에 의존하는 결합 모델과는 달리, 본 발명의 PharDiff 모델은 리간드와 단백질 사이의 복잡한 화학적 상호작용을 고려함으로써, 더 높은 결합 친화도를 지닌 신약 후보 물질을 도출할 수 있다.

【0041】 도 2는 본 발명의 실시예에 따른 파마코포어 기반 3D 분자 생성 시스템의 구성을 도시한 도면이다.

【0042】 도 2를 참조하면, 데이터 입력부(110), 파마코포어 기반 인공지능 모델(120) 및 분자 구조 출력부(130)를 포함할 수 있다.

【0043】 먼저, 데이터 입력부(110)는 타겟 단백질과 결합 포켓의 구조적 데이터를 수집하고 이를 시스템에 입력한다. 상기 데이터에는 단백질의 결합 포켓에 대한 3D 좌표 및 화학적 특징이 포함되고, 또한, 상기 데이터는 리간드가 결합할 수 있는 최적의 공간을 찾기 위해 사용된다. 또한, 상기 데이터 입력부(110)는 단백질-리간드 상호작용에 필수적인 파마코포어를 인식하여 후속 단계에 반영한다.

【0044】 파마코포어 기반 인공지능 모델(120)은 입력된 결합 포켓 정보와 파마코포어를 기반으로 최적의 리간드 구조를 생성하기 위해 특별히 설계된 확산 생성 모델이다. 상기 파마코포어 기반 인공지능 모델(120)은 초기 상태에서 무작위로 추가된 노이즈를 점차 제거하며 원하는 리간드 구조로 수렴한다. 특히, SE(3)-동치를 갖춘 신경망 구조를 사용하여 문자의 회전이나 평행 이동에 영향을 받지 않고 리간드의 3D 구조를 생성한다. 상기 생성 과정은 파마코포어를 고정한 상태에서 나머지 문자 구조를 구성하는 방식으로 진행된다.

【0045】 문자 구조 출력부(130)는 생성된 리간드의 3차원 구조를 출력한다. 상기 문자 구조 출력부(130)는 생성된 리간드 구조가 타겟 단백질과 결합할 수 있는지를 평가하며, 결합 친화도와 구조적 안정성을 최종적으로 검토한 후 출력할 수 있다. 이로써, 약물 개발에서 활용할 수 있는 안정적인 문자 구조를 제공한다. 특히, 출력된 구조는 문자 동역학 시뮬레이션을 통해 실험적 검증을 거친 후 최종적으로 안정성이 높은 리간드 구조로 도출된다.

【0046】 도 3은 본 발명의 실시예에 따른 파마코포어 기반 인공지능 모델의 생성 및 학습 방법을 도시한 도면이다.

【0047】 단계(S121)에서 상기 파마코포어 기반 인공지능 모델(120)은, 단백질-리간드 간의 상호작용(Protein-Ligand Interaction) 데이터 및 파마코포어(Pharmacophore) 데이터를 수집하고 준비한다. 상기 단백질-리간드 상호작용 데이터는 단백질 결합 포켓(Binding Pocket)의 구조적 특징과 화학적 정보를 포함하며, 상기 파마코포어 데이터는 리간드 내에서 단백질과의 결합에 중요한 작용점을 식별하는 데 사용된다. 단계(S121)에서 수집된 데이터는 이후의 모델 학습 및 리간드 생성 과정의 기초가 된다.

【0048】 단계(S122)에서 상기 파마코포어 기반 인공지능 모델(120)은, SE(3)-동치 신경망(SE(3)-Equivalent Neural Network)을 설계하고 하이퍼파라미터를 설정하여 모델을 초기화한다. 상기 SE(3)-동치 신경망은 리간드의 3차원 구조가 회전이나 이동에 영향을 받지 않도록 학습을 진행할 수 있는 모델 구조를 의미하며, 단백질-리간드 상호작용의 공간적 관계를 정확하게 반영할 수 있도록 한다. 또한, 상기 하이퍼파라미터는 학습률, 배치 크기 등의 모델 학습 효율을 조절하는 매개변수로, 최적의 모델 성능을 보장하기 위해 초기 단계에서 설정된다.

【0049】 단계(S123)에서 상기 파마코포어 기반 인공지능 모델(120)은, 비자기회귀적(Non-Autoregressive) 확산 생성 모델을 적용하여 리간드 구조 생성을 시작한다. 상기 확산 모델은 초기 상태에서 리간드의 구조에 무작위 노이즈를 추가한 후, 점진적으로 노이즈를 제거하여 목표하는 구조로 수렴하는 방식으로 작동한다.

상기 과정에서 상기 파마코포어가 고정된 상태로 유지되며, 나머지 구조는 결합 포켓에 최적화된 형태로 형성된다.

【0050】 단계(S124)에서 상기 파마코포어 기반 인공지능 모델(120)은, 손실 최적화 과정을 통해 모델을 학습한다. 상기 단계(S124)에서 손실 함수는 리간드 구조와 목표 구조 간의 차이를 최소화하는 방향으로 설정되며, 이로써 생성된 리간드가 단백질 결합 포켓에 안정적으로 결합할 수 있도록 학습이 진행된다. 상기 파마코포어 기반 인공지능 모델(120)은 손실 값을 반복적으로 줄여가면서, 최종적으로 결합 친화도와 구조적 안정성이 높은 리간드 구조를 생성하도록 학습된다.

【0051】 단계(S125)에서 상기 파마코포어 기반 인공지능 모델(120)은, 생성된 리간드 구조를 결합 친화도와 구조적 안정성 측면에서 평가하고 최적화한다. 상기 단계(S125)에서는 생성된 리간드의 결합 강도와 안정성이 평가되며, 소프트웨어(예를 들어, AutoDock Vina)를 사용하여 단백질-리간드 결합의 에너지를 계산한다. 평가된 결과를 기초로 하여 리간드의 결합 친화도와 안정성이 부족한 경우 추가적인 최적화 과정을 거쳐, 최종적으로 단백질 포켓에 적합한 리간드 구조가 생성된다.

【0052】 도 4는 본 발명의 실시예에 따른 파마코포어 기반 인공지능 모델(PharDiff)이 리간드의 3차원 구조를 생성하는 방법을 도시한 도면이다.

【0053】 먼저, 단계(S131)에서 상기 파마코포어 기반 인공지능 모델(120)은, 데이터 수집부를 통해 단백질과 리간드 간의 결합 포켓 정보를 인식한다. 상기 결합 포켓은 리간드와의 상호작용을 위한 타겟 부위로, 단백질의 특정 3D 구조 및 화

학적 성질을 반영한다.

【0054】 단계(S132)에서 상기 파마코포어 기반 인공지능 모델(120)은, 단백질 결합 포켓에 적합한 파마코포어(Pharmacophore)를 인식한다. 상기 파마코포어는 리간드 내에서 결합에 필수적인 화학적 작용점들로 구성되어 있으며, 단백질과의 결합 강도와 선택성에 영향을 끼친다. 상기 파마코포어 기반 인공지능 모델(120)은 상기 파마코포어를 고정된 기준점으로 인식하여, 이후 리간드 생성 과정에서 상기 파마코포어 부위를 유지하면서 나머지 구조를 최적화한다.

【0055】 단계(S133)에서 상기 파마코포어 기반 인공지능 모델(120)은, 비자기회귀적(Non-Autoregressive) 확산 생성 모델의 초기 상태로서 리간드 구조에 무작위 노이즈를 추가한다. 초기 노이즈는 상기 비자기회귀적(Non-Autoregressive) 확산 생성 모델의 핵심 개념으로, 리간드가 목표하는 3D 구조로 수렴하는 과정의 시작점 역할을 한다. 상기 초기 노이즈가 추가됨으로써 리간드는 초기 불명확한 형태로 시작되고, 이후 단계에서 상기 노이즈를 점차 제거하여 최종 구조를 형성하게 된다.

【0056】 단계(S134)에서 상기 파마코포어 기반 인공지능 모델(120)은, SE(3)-동치 신경망(SE(3)-Equivalent Neural Network)을 적용하여 리간드 구조가 회전이나 이동에 영향을 받지 않도록 한다. 상기 SE(3)-동치 신경망은 리간드의 3D 구조가 공간적 위치나 방향에 관계없이 단백질과의 결합 가능성을 유지할 수 있도록 설계되어 있다. 이로써, 리간드가 다양한 각도와 위치에서 단백질과 결합할 때 결합 친화도를 안정적으로 유지할 수 있다.

【0057】 단계(S135)에서 상기 파마코포어 기반 인공지능 모델(120)은, 노이즈를 점진적으로 제거하여 리간드의 3D 구조를 생성한다. 상기 확산 모델은 초기 노이즈를 점차 줄여가며, 상기 파마코포어를 중심으로 리간드의 나머지 구조를 형성한다. 이 과정에서 단백질 결합 포켓에 최적화된 리간드 구조가 점진적으로 생성되며, 상기 파마코포어 기반 인공지능 모델(120)은 결합 친화도가 높은 형태로 리간드를 완성한다.

【0058】 단계(S136)에서 상기 파마코포어 기반 인공지능 모델(120)은, 수집된 결합 포켓(Binding Pocket) 정보를 리간드 생성 과정에 반영한다. 상기 결합 포켓 정보는 리간드의 최종 구조에서 단백질과 결합하는 부위에 반영되며, 상기 파마코포어와 결합 포켓의 상호작용이 리간드의 전체 구조에 구현된다. 이로써, 리간드의 결합력이 강화된 3D 구조가 만들어진다.

【0059】 단계(S137)에서 상기 파마코포어 기반 인공지능 모델(120)은, 결합 친화도와 구조적 안정성을 평가하고 최적화한다. 상기 생성된 리간드는 단백질과의 결합 친화도와 구조적 안정성 측면에서 평가된다. 상기 결합 친화도 평가를 통해 리간드가 단백질과 얼마나 강하게 결합할 수 있는지 측정하며, 안정성 평가에서는 리간드 구조가 생체 환경에서 얼마나 잘 유지될 수 있는지를 검토한다.

【0060】 단계(S138)에서 상기 파마코포어 기반 인공지능 모델(120)은, 생성된 리간드를 문자 동역학 시뮬레이션을 통해 검증한다. 상기 문자 동역학 시뮬레이션은 리간드와 단백질 간의 상호작용이 실제 생체 환경에서 안정적으로 유지될 수 있는지를 테스트한다. 상기 테스트 과정에서 리간드의 위치와 결합 상태가 얼마나

안정적인지를 확인하여 검증을 진행한다.

【0061】 단계(S139)에서 상기 파마코포어 기반 인공지능 모델(120)은, 리간드-단백질 복합체의 시각적 분석을 수행한다. 상기 시각적 분석을 통해 생성된 리간드가 단백질 결합 포켓과의 상호작용에서 정확하게 맞물리는지 확인한다. 이로써, 리간드의 결합 자세와 결합 부위를 시각적으로 검토하여, 약물 설계의 적합성을 판단할 수 있다.

【0062】 단계(S140)에서 상기 파마코포어 기반 인공지능 모델(120)은, 최적화된 리간드의 3차원 구조를 최종적으로 출력한다. 상기 출력된 리간드의 3차원 구조는 타겟 단백질에 결합할 수 있도록 최적화된 형태로, 약물 후보 물질로서의 활용 가능성을 지닌다. 상기 출력된 리간드의 3차원 구조는 생체 내에서 높은 결합력과 안정성을 지닌 리간드로, 신규 약물 개발의 효율을 높인다.

【0063】 도 5는 본 발명의 실시예에 따른 3D 파마코포어 인식 및 식별 알고리즘을 도시한 도면이다.

【0064】 본 발명의 3D 파마코포어 인식 및 식별 알고리즘은 단백질-리간드 상호작용을 분석하여, 단백질과 리간드 간의 결합을 형성하는 데 중요한 파마코포어(Pharmacophore)를 인식하고 추출하는 과정을 수행한다. 상기 알고리즘은 단백질-리간드 쌍의 집합으로부터 파마코포어 리스트를 생성하며, 다양한 상호작용 프로파일을 이용하여 결합에 중요한 화학적 단서를 도출한다. 구체적으로, PLIP(Protein-Ligand Interaction Profiler) 도구를 활용하여 단백질-리간드 복합체에서 상호작용 정보를 추출하고, 상기 상호작용 정보를 기초로 하여 3D 파마코포

어 인식 및 식별 알고리즘을 적용하여 분자 내 중요한 파마코포어를 식별한다.

【0065】 3D 파마코포어 인식 및 식별 알고리즘은 먼저 단백질과 리간드가 결합한 복합체에서 유효한 상호작용을 추출하는 과정을 시작한다. 여기서, PLIP(Protein-Ligand Interaction Profiler) 도구를 활용하여 단백질과 리간드 간의 결합 상태를 분석하고, 안정적인 결합을 형성하는 상호작용 정보를 수집한다. 상기 상호작용 정보는 PLIP 도구를 통해 얻어지며, 결합에 중요한 리간드 원자의 인덱스를 B 집합으로 정의하여 기록한다. 여기서 $B = \{a_{ik}\}$ 는 I개의 탐지된 상호작용에 관여하는 K개의 원자를 포함한다.

【0066】 상기 3D 파마코포어 인식 및 식별 알고리즘은 3차원으로 표현된 리간드 분자를 여러 개의 조각 세트($M = \{(F_i, E)\}$)로 분해한다. 각 조각(F_i)은 리간드의 특정 부분을 나타내며, 원자 집합(A_i)과 해당 원자 간의 엣지 정보(E_i)로 구성된다. 두 조각 간에는 겹치는 엣지 세트(E)가 존재하며, 각각의 조각들이 리간드 내에서 어떻게 연결되어 있는지를 나타낸다.

【0067】 상기 3D 파마코포어 인식 및 식별 알고리즘은 B 집합에 정의된 상호작용 원자의 인덱스를 기초로하여, 상호작용에 관여하는 리간드의 특정 조각을 식별한다. 여기서, 각 조각(F_i)의 원자 집합(A_i)을 순회하면서, B 집합의 원자 인덱스와 일치하는 원자를 포함한 조각을 수집한다. 이때 상호작용을 형성하는 원자를 포함한 모든 조각들이 수집되고, 병합이 이루어진다. 상기 병합 과정은 각각의 상호작용에 대해 독립적으로 수행되며, 결합에 중요한 3D 파마코포어를 형성하게 된다.

【0068】 상기 병합 과정에 의해 결합에 중요한 조각들이 하나의 파마코포어

로 정의된다. 파마코포어는 리간드가 단백질과 상호작용할 수 있도록 하는 핵심 작용점들로 구성되며, 최종적으로 Pharmacophore List에 추가된다. 상기 과정은 각각의 단백질-리간드 쌍에 대해 개별적으로 수행되며, 상호작용 유형에 따라 여러 파마코포어가 생성될 수 있다.

【0069】 상기 3D 파마코포어 인식 및 식별 알고리즘에서 파마코포어 인식 및 식별 과정에 의하여 생성된 파마코포어는 확산 모델에서 리간드 생성의 기준점으로 활용된다. 확산 과정에서는 이 중 하나의 파마코포어가 무작위로 선택되어 리간드의 초기 구조 형성에 영향을 미치며, 이후 점진적으로 노이즈를 제거하여 결합 친화도가 높은 최적의 리간드 구조를 완성하게 된다.

【0070】 상기 3D 파마코포어 인식 및 식별 알고리즘의 상세한 구조 및 파마코포어 식별 과정은 아래에서 설명하도록 하겠다.

【0071】 도 5에 도시된 바와 같이, 입력은 단백질-리간드 쌍의 집합 $A = \{(p_i, m_i)\}_{i=1}^N$ 이다. 여기서, p_i 는 단백질을 의미하고, m_i 는 리간드를 의미한다.

【0072】 출력은 각 단백질-리간드 쌍에 대해 생성된 파마코포어 리스트 (Pharmacophore List)이다.

【0073】 결합 프로파일 생성(GetBinding) 단계에서는 단백질-리간드 상호작용 프로파일러가 실행되어, 단백질과 리간드 간의 상호작용 정보가 수집된다. 상기 수집된 정보는 파마코포어 추출을 위한 초기 데이터로 활용된다.

【0074】 3D 구조 분할(TreeFrag) 단계에서는 3D 분할(Fragmentizing) 과정이 실행되며, 리간드의 구조가 여러 개의 작은 단편으로 나뉜다. 상기 과정은 파마코포어의 구조적 특징을 파악하는 데 사용된다.

【0075】 파마코포어 생성(GetPharm) 단계에서는 서로 연결된 노드를 결합하여 하나의 파마코포어로 묶는다. 상기 묶는 과정은 약물의 특정 화학적 작용점을 기반으로 결합되는 영역을 의미한다.

【0076】 초기화 및 리스트 생성(TotalPharList 및 PharList) 단계에서는 TotalPharList와 PharList가 빈 리스트로 초기화된다. 상기 리스트는 각각 모든 파마코포어 리스트와 현재 단백질-리간드 쌍의 파마코포어 리스트를 저장하기 위한 리스트이다.

【0077】 도 5의 5-9번 단계는 결합 정보 처리(GetBinding) 단계이다. 단백질-리간드 쌍에 대한 결합 정보를 처리하며, 상호작용하는 노드를 추출하고 결합된 원자 집합을 통해 파마코포어 후보를 선정한다.

【0078】 도 5의 10-15번 단계는 교차 리스트 생성(IntersectionList)하는 단계이다. 상기 추출된 결합 정보로부터, 교차된 원자 목록을 생성하고, 파마코포어를 결정하는 데 필요한 정보를 축적한다.

【0079】 도 5의 16번 단계는 파마코포어 결정(GetPharm) 단계이다. IntersectionList를 사용하여 최종적인 파마코포어를 결정하고, PharList에 추가한다.

【0080】 도 5의 17-21번 단계는 전체 리스트 업데이트하고 저장하는 단계이다. 상기 생성된 파마코포어를 TotalPharList에 추가하여 모든 단백질-리간드 쌍에 대해 파마코포어 리스트를 완성한다.

【0081】 상기 알고리즘은 3차원 공간에서 단백질-리간드 상호작용을 기반으로 중요한 파마코포어를 효율적으로 추출할 수 있다. 3D 구조를 분할하고 다수의 결합 정보를 교차 분석함으로써, 생물학적 중요성이 높은 영역을 정확하게 식별할 수 있다. 이로써, 상기 알고리즘은 각 단계에서 단백질-리간드 상호작용의 중요 부위를 파악하고, 이를 기초로 하여 리간드의 핵심 작용점을 설정함으로써 PharDiff 모델의 정확성과 효율성을 높이는 역할을 한다.

【0082】 도 6은 본 발명의 실시예에 따른 파마코포어 기반 인공지능 모델(PharDiff)의 작동 방식을 도시한 도면이다.

【0083】 구체적으로, 파마코포어 기반 인공지능 모델(PharDiff)이 비자기회귀적(Non-Autoregressive) 확산 생성 모델을 이용하여 파마코포어(Pharmacophore)와 나머지 문자 구조를 활용한 3차원 문자를 생성하는 과정을 도시한다.

【0084】 상기 파마코포어 기반 인공지능 모델(PharDiff)은 파마코포어(101)를 고정시키고, 나머지 부분인 추론 대상 하위 구조인 원자들(102) 및 표적 단백질(103) 간의 상호작용을 기초로 하여 3차원 문자를 생성한다.

【0085】 이때, 상기 원자들(102)은 파마코포어(101)를 제외한 문자의 나머지 부분으로, 도면에서 점으로 나타난 원자들의 집합에 해당하고, 문자를 구성하는 여

러 원자들의 집합으로 이루어진 구조를 나타낸다. 상기 원자들(102)은 정방향 과정(Forward Process)과 역방향 과정(Reverse Process)에서 노이즈를 추가 및 제거하는 과정의 주요 대상이 되며, 파마코포어와의 상호작용을 고려하여 최적의 3차원 문자 구조를 도출하도록 한다. 또한, 상기 원자들(102)은 문자의 일부로서 여러 원자들로 구성된 화학적 구조를 나타내며, 단일 원자가 아닌 여러 원자들의 조합을 통해 문자 전체의 물리적 및 화학적 특성을 나타낸다.

【0086】 여기서, 정방향 과정(Forward Process)은 모델이 데이터를 처리하면서 점진적으로 노이즈(Noise)를 추가하여 문자의 불확실성(Uncertainty)을 증가시키는 과정이다. 한편, 역방향 과정(Reverse Process)은 노이즈를 제거하여 문자의 명확한 구조를 복원하는 과정이다. 정방향 과정에서는 노이즈를 추가하여 다양한 가능성을 열어두며, 역방향 과정에서는 상기 노이즈를 제거하여 최적의 문자 구조를 도출하도록 한다. 이로써, 정방향 및 역방향 과정은 파마코포어(101)를 고정시키면서 나머지 문자의 원자 좌표와 결합 특성을 효율적으로 예측할 수 있다.

【0087】 먼저, 도 6의 (5)로부터 (1) 방향에 대응하는 정방향 과정(Forward Process)은 파마코포어(101)를 제외한 문자의 하위 구조인 원자(102)에 노이즈를 추가하여, 다양한 문자 가능성을 고려하는 과정이다. 상기 정방향 과정은 점진적으로 문자 구조에 불확실성을 더해가며, 원자들의 좌표와 결합 정보를 학습한다. 입력 데이터에 노이즈를 추가하는 과정에서 확률 분포 수식이 사용된다. 상기 과정의 수식은 다음과 같다:

【0088】 $q_\phi(T_{t-1} | T_t, C, P)$

【0089】 q_ϕ : 정방향 과정의 확률 분포

【0090】 ϕ : 정방향 과정의 확률 분포를 제어하는 파라미터(모델의 학습 가능한 변수)

【0091】 T_t : t 시점에서의 분자의 하위 구조인 원자들(102)

【0092】 C: 고정된 파마코포어(Pharmacophore, 101)

【0093】 P: 표적 단백질(Target Protein, 103)

【0094】 상기 정방향 과정에서는 분자의 하위 구조인 원자들(T , 102)에 노이즈가 점진적으로 추가되며, 파마코포어(C , 101)와 표적 단백질(P , 103)은 고정된 상태로 유지된다. 상기 과정은 시간 t 에서 시간 $t-1$ 로 진행되면서, 점차 더 많은 노이즈가 원자들(T , 102)에 추가된다.

【0095】 도 6의 (1)로부터 (5) 방향에 대응하는 역방향 과정(Reverse Process)은 상기 정방향 과정에서 추가된 노이즈를 점진적으로 제거하여, 원자들(102)이 점차적으로 명확한 구조를 가지도록 만드는 과정이다. 노이즈가 제거되면서 분자가 점차 명확한 3차원 구조로 생성되는 과정에서 확률 분포 수식이 사용된다. 이를 설명하는 수식은 다음과 같다:

【0096】 $p_\theta(T_{t-1} | T_t, C, P)$

【0097】 p_θ : 역방향 과정의 확률 분포

【0098】 θ : 역방향 과정의 확률 분포를 제어하는 파라미터(모델의 학습 가능

능한 변수)

【0099】 상기 역방향 과정에서는 노이즈가 단계별로 제거되며, 최종적으로 문자의 3차원 구조가 명확해진다. 상기 역방향 과정에서 파마코포어(C, 101)과 표적 단백질(P, 103)은 고정된 상태로 유지되며, 하위 구조인 원자들(T, 102)에서만 노이즈를 점차 제거하여 최종적으로 고정된 파마코포어(101)와의 결합이 최적화된 문자 구조로 완성된다.

【0100】 도 6의 (5)는 역방향 과정이 완료되어 최적화된 문자 구조가 생성된 상태를 도시한다. 도6의 (5)에 도시된 바와 같이, 파마코포어(101)는 고정된 상태로 표적 단백질(103)과의 결합을 유지하며, 원자들(102)은 명확한 3차원 문자 구조를 가진다. 상기 3차원 문자 구조는 단백질-리간드 상호작용에서 안정적인 결합을 형성할 수 있는 최적의 상태로 생성된다.

【0101】 상기 파마코포어 기반 인공지능 모델(PharDiff)은 비자기회귀적(Non-Autoregressive) 확산 생성 모델을 사용하여, 파마코포어(Pharmacophore)를 고정한 상태에서 문자의 나머지 부분을 예측한다. 또한, 3차원 문자 구조는 상기 과정에서 SE(3)-동치 신경망(SE(3)-Equivariant Neural Networks)을 활용하여, 문자 구조의 회전 및 이동에도 불변성을 유지할 수 있도록 설계될 수 있다. 이로써, 상기 파마코포어 기반 인공지능 모델(PharDiff)은 단백질-리간드 상호작용을 효과적으로 모델링할 수 있도록 최적화된다. 마찬가지로 최종적으로 생성된 문자는 약물 작용성을 유지하면서도 단백질과의 결합력을 극대화할 수 있는 최적의 구조를 갖출 수 있다.

【0102】 도 7은 본 발명의 실시예에 따른 파마코포어 기반 인공지능 모델에 의한 결합 분자를 시각화하여 도시한 도면이다. 구체적으로, 도 7은 상기 파마코포어 기반 인공지능 모델에 의해 생성된 문자와 참조 문자 사이의 결합을 비교한 것이다. 도면에서 실선과 점선으로 표시된 결합이 단백질과 리간드 간의 상호작용을 나타내며, 상기 상호작용이 일어나는 부분을 파마코포어(Pharmacophore)로 식별할 수 있다.

【0103】 도 7의 (E1)은 단백질과 결합된 참조 결합 문자(Reference Binding Molecule)의 결합 구조를 나타내며, 상기 문자는 단백질과의 결합에 필요한 파마코포어를 포함하고 있다. 상기 파마코포어는 결합에 필요한 화학적 상호작용을 가능하게 하여, 상기 PharDiff 모델이 학습하고 추출해야 할 정보로 작용한다.

【0104】 도 7의 (E2) 및 (E3)은 상기 파마코포어 기반 인공지능 모델(PharDiff)에 의해 생성된 리간드 문자 구조의 예시를 각각 나타낸다. 각각의 생성된 리간드는 단백질과의 결합 포켓에 적합한 형태로 설계되었으며, 참조 문자와 유사한 위치에 파마코포어를 포함하고 있다. 상기 PharDiff 모델은 파마코포어를 정확하게 식별하여 참조 문자와 유사한 결합 특성을 유지할 수 있도록 생성된 리간드의 구조를 최적화한다.

【0105】 도 7에서 점선은 단백질과 리간드 간의 비공유 결합(Intermolecular Interaction)을 나타내며, 결합에 중요한 역할을 하는 화학적 상호작용을 설명한다. 상기 상호작용 부위는 상기 파마코포어 기반 인공지능 모델(PharDiff)이 참조 결합 문자와 유사한 결합을 재현할 수 있음을 보여주며, 상기 파마코포어 기

반 인공지능 모델이 단백질과의 결합 친화도를 최적화하여 약리학적 효능을 높일 수 있는 리간드를 생성할 수 있음을 입증한다.

【0106】 따라서, 도 7은 상기 파마코포어 기반 인공지능 모델이 단백질과의 결합을 위한 리간드 생성에 있어 참조 문자의 중요한 결합 요소를 정확히 재현할 수 있음을 시작적으로 입증하며, 파마코포어 기반 AI 모델의 효율성을 강조할 수 있다. 이로써, 본 발명은 약물 후보 물질의 설계 과정에서 결합 친화성을 높이는 데 기여하며, 특히 신약 개발에서 중요한 역할을 할 수 있다.

【0107】 도 8은 본 발명의 실시예에 따른 문자 생성 시스템이 3차원 구조의 리간드 문자 생성하는 과정을 도시한 도면이다.

【0108】 단계(S11)에서 상기 문자 생성 시스템(100)은, 단백질 포켓 데이터를 입력받는다. 상기 데이터는 리간드가 결합할 수 있는 단백질의 특정 부위를 포함한 3차원 좌표와 화학적 특성을 포함한다. 상기 단백질 결합 포켓 정보는 리간드 와의 상호작용을 위한 목표 지점으로, 상기 문자 생성 시스템(100)의 다음 단계에서 리간드 생성에 필요한 핵심 데이터를 제공한다.

【0109】 단계(S12)에서 상기 문자 생성 시스템(100)은, 파마코포어 기반 인공지능 모델(PharDiff)을 이용하여 입력된 단백질 포켓 데이터를 처리한다. 단계(S12)에서는 PharDiff 모델이 단백질 포켓에 최적화된 리간드 구조를 설계하기 위해 입력 데이터를 분석하고, 파마코포어를 식별하여 리간드의 특정 결합 특성을 반영한다. 상기 PharDiff 모델은 단백질과 리간드 간의 결합 친화도를 높이기 위해 파마코포어를 기준으로 나머지 문자 구조를 형성하며, 최종적으로 리간드의 3차원

구조를 생성할 준비를 완료한다.

【0110】 단계(S13)에서 상기 문자 생성 시스템(100)은, 리간드의 3차원 구조 데이터를 출력한다. 상기 출력 데이터는 단백질 결합 포켓과의 결합에 최적화된 리간드의 구조적 정보를 포함하며, 약물 개발 및 연구에 활용될 수 있다. 상기 출력된 리간드 구조는 단백질과의 결합 친화도가 높은 형태로, 약리학적 효능을 극대화하기 위해 설계된다.

【0111】 도 9는 본 발명의 실시예에 따른 문자 생성 시스템의 PharDiff 모델이 생성한 문자의 성능을 결합 친화도와 구조적 유사성 측면에서 평가한 참조도이다.

【0112】 도 9의 T1에서는 PharDiff 모델이 생성한 문자의 결합 친화도를 여러 기준 모델과 비교한 결과를 보여준다. 결합 친화도 평가(Binding Affinity Evaluation)는 예를 들어, Vina Score, Vina Min, High Affinity 등의 지표를 사용하여 측정되며, 각 지표는 생성된 문자가 단백질의 결합 포켓과 얼마나 안정적으로 결합할 수 있는지를 나타낸다.

【0113】 도 9의 T1에 나타난 바와 같이 상기 PharDiff 모델은 결합 친화도 측면에서 경쟁 모델들을 능가하는 성능을 보인다. Vina Score 및 기타 지표에서 높은 점수를 획득하며, 결합 친화도가 뛰어난 문자를 생성할 수 있는 PharDiff의 효율성을 입증한다. 상기 결과는 PharDiff 모델이 단백질-리간드 간의 상호작용에 필수적인 파마코포어를 정확히 식별하고 최적화된 리간드 구조를 생성할 수 있음을 뒷받침한다.

【0114】 도 9의 T2에서는 PharDiff 모델이 생성한 문자의 구조적 형태를 참조 문자와 비교하여 평가하였다. 평가를 위하여 결합 길이 분포의 유사성을 측정하였으며, Jensen-Shannon 발산을 사용하여 구조적 유사성을 정량적으로 평가하였다.

【0115】 T2에서 나타난 바와 같이, Jensen-Shannon 발산 값이 낮을수록 구조적 유사성(Structural Conformation Similarity)이 높음을 의미하며, PharDiff는 참조 문자와 유사한 결합 길이 분포를 유지하는 데 있어 탁월한 성과를 보였다. 상기 평가 결과는 PharDiff 모델이 실제 문자의 기하학적 구조와 유사한 문자 구조를 생성할 수 있는 능력을 입증하며, 약물 설계에서 요구되는 구조적 안정성과 유사성을 충족할 수 있음을 보여준다.

【0116】 도 10은 본 발명의 실시예에 따른 디스크립터 생성 방법 및 생성 장치를 구현하는 컴퓨팅 장치를 도시한다.

【0117】 도 1 ~ 도 8에 의해 설명된 본 발명의 실시예는 적어도 하나의 프로세서에 의해 동작하는 컴퓨팅 장치(900)로 구현될 수 있다.

【0118】 컴퓨팅 장치(900)는 프로세서(910), 메모리(920), 스토리지(930), 통신 인터페이스(940), 시스템 인터커넥트(950) 및 디스플레이(960)를 포함할 수 있다.

【0119】 프로세서(910)은, CPU(Central Processing Unit), MPU(Micro Processor Unit), MCU(Micro Controller Unit), GPU(Graphic Processing Unit) 및 APU(Application Processing Unit)을 포함한다.

【0120】 메모리(920)는 프로세서(910)와 상호작용하여 프로그램이 효율적으로 실행될 수 있도록 데이터를 저장하고 필요한 정보에 빠르게 접근할 수 있도록 하는 기능을 수행한다. 메모리(920)는 레지스터, 캐시 메모리, 주 메모리, 읽기 전용 메모리, 가상 메모리, 비휘발성 메모리 중 적어도 하나를 포함한다.

【0121】 스토리지(930)는 데이터를 영구적으로 저장하고 관리하는 역할을 한다. 스토리지는 컴퓨팅 시스템이 꺼지거나 재부팅된 후에도 데이터를 보존하며, 운영 체제, 애플리케이션, 사용자 파일 등을 저장하는 데 사용된다. 상기 스토리지(930)는 하드 디스크 드라이브(HDD), 솔리드 스테이트 드라이브(SSD), 광학 디스크, 네트워크 스토리지 및 클라우드 스토리지 중 적어도 하나를 포함한다.

【0122】 통신 인터페이스(940)는 컴퓨팅 시스템 내부 및 외부의 다양한 장치들 간에 데이터를 주고받기 위한 경로를 제공한다. 통신 인터페이스(940)는 USB(Universal Serial Bus), PCIe(Peripheral Component Interconnect Express), SATA(Serial ATA), Ethernet, Wi-Fi, Thunderbolt 및 HDMI(High-Definition Multimedia Interface) 중 적어도 하나의 통신 방식을 지원할 수 있다.

【0123】 시스템 인터커넥트(950)는 컴퓨팅 시스템 내부에서 다양한 구성 요소들 간의 데이터와 신호를 주고받는 역할을 한다. 시스템 인터커넥트(950)는, 버스(Bus), 포인트-투-포인트(Point-to-Point), 크로스바 스위치(Crossbar Switch), 네트워크-온-칩(Network-on-Chip, NoC) 중 적어도 하나의 방식을 지원할 수 있다.

【0124】 디스플레이(960)는 컴퓨팅 시스템의 출력 장치로서, 사용자에게 시각적인 정보를 제공하는 기능을 수행한다.

【0125】 전술한 구성에 의하여, 본 발명의 실시예에 따른 프로그램은, 프로세서(910)에 의해 실행되는 명령어들에 기초하여 실행되며, 메모리(920) 또는 스토리지(930)에 저장될 수 있다.

【0126】 전술한 본 발명의 실시예에 따른 방법은 다양한 컴퓨터 구성요소를 통하여 실행될 수 있는 프로그램 명령 형태로 구현되어 컴퓨터 판독 가능한 기록매체에 기록될 수 있다. 컴퓨터 판독 가능한 기록매체는 프로그램 명령어, 데이터 파일, 데이터 구조 등을 단독으로 또는 조합하여 포함할 수 있다. 컴퓨터 판독 가능한 기록매체에 기록되는 프로그램 명령은 본 발명의 실시예를 위하여 특별히 설계되고 구성된 것이거나, 컴퓨터 소프트웨어 분야의 통상의 기술자에게 공지되어 사용가능한 것일 수 있다. 컴퓨터 판독 가능한 기록매체는, 하드디스크, 플로피디스크, 자기테이프 등의 자기기록 매체, CD-ROM, DVD 등의 광기록 매체, 플롭티컬디스크 등의 자기-광 매체, ROM, RAM, 플래시 메모리 등과 같이, 프로그램 명령을 저장하고 수행하도록 구성된 하드웨어를 포함한다. 프로그램 명령은, 컴파일러에 의해 만들어지는 기계어 코드, 인터프리터를 사용하여 컴퓨터에서 실행될 수 있는 고급언어 코드를 포함한다. 하드웨어는 본 발명에 따른 방법을 처리하기 위하여 하나 이상의 소프트웨어 모듈로서 작동하도록 구성될 수 있고, 그 역도 마찬가지이다.

【0127】 본 발명의 실시예에 따른 방법은 프로그램 명령 형태로 전자장치에서 실행될 수 있다. 전자장치는 스마트폰이나 스마트패드 등의 휴대용 통신 장치,

컴퓨터 장치, 휴대용 멀티미디어 장치, 휴대용 의료 기기, 카메라, 웨어러블 장치, 가전 장치를 포함한다.

【0128】 본 발명의 실시예에 따른 방법은 컴퓨터 프로그램 제품에 포함되어 제공될 수 있다. 컴퓨터 프로그램 제품은 상품으로서 판매자 및 구매자 간에 거래될 수 있다. 컴퓨터 프로그램 제품은 기기로 읽을 수 있는 기록매체의 형태로, 또는 어플리케이션 스토어를 통해 온라인으로 배포될 수 있다. 온라인 배포의 경우에, 컴퓨터 프로그램 제품의 적어도 일부는 제조사의 서버, 어플리케이션 스토어의 서버, 또는 중계 서버의 메모리와 같은 저장 매체에 적어도 일시 저장되거나, 임시적으로 생성될 수 있다.

【0129】 본 발명의 실시예에 따른 구성요소, 예컨대 모듈 또는 프로그램 각각은 단수 또는 복수의 서브 구성요소로 구성될 수 있으며, 이러한 서브 구성요소들 중 일부 서브 구성요소가 생략되거나, 또는 다른 서브 구성요소가 더 포함될 수 있다. 일부 구성요소들(모듈 또는 프로그램)은 하나의 개체로 통합되어, 통합되기 이전의 각각의 해당 구성요소에 의해 수행되는 기능을 동일 또는 유사하게 수행할 수 있다. 본 발명의 실시예에 따른 모듈, 프로그램 또는 다른 구성요소에 의해 수행되는 동작들은 순차적, 병렬적, 반복적 또는 휴리스틱하게 실행되거나, 적어도 일부 동작이 다른 순서로 실행되거나, 생략되거나, 또는 다른 동작이 추가될 수 있다.

【0130】 전술한 본 발명의 설명은 예시를 위한 것이며, 본 발명이 속하는 기술분야의 통상의 지식을 가진 자는 본 발명의 기술적 사상이나 필수적인 특징을 변

경하지 않고서 다른 구체적인 형태로 쉽게 변형이 가능하다는 것을 이해할 수 있을 것이다. 그러므로 이상에서 기술한 실시예들은 모든 면에서 예시적인 것이며 한정적이 아닌 것으로 이해해야만 한다. 예를 들어, 단일형으로 설명되어 있는 각 구성요소는 분산되어 실시될 수도 있으며, 마찬가지로 분산된 것으로 설명되어 있는 구성 요소들도 결합된 형태로 실시될 수 있다.

【0131】 본 발명의 범위는 후술하는 청구범위에 의하여 나타내어지며, 청구 범위의 의미 및 범위 그리고 그 균등 개념으로부터 도출되는 모든 변경 또는 변형된 형태가 본 발명의 범위에 포함되는 것으로 해석되어야 한다.

【부호의 설명】

【0132】 20: 단백질 포켓

30: 리간드

110: 데이터 입력부

120: 파마코포어 기반 인공지능 모델

130: 문자 구조 출력부

【청구범위】

【청구항 1】

신약 후보 물질 도출을 위한 파마코포어(Pharmacophore) 및 확산 생성 모델을 이용하여 리간드의 3차원 분자 구조를 생성하는 방법에 있어서,

단백질의 결합 포켓 데이터를 수집하여 입력하는 단계;

리간드의 파마코포어를 인식 및 식별하는 단계;

상기 결합 포켓 데이터에 기초하여 파마코포어를 고정된 기준으로 설정하고, 나머지 리간드 구조를 형성하기 위해 SE(3)-동치 신경망(SE(3)-Equivalent Neural Network)과 비자기회귀적(Non-Autoregressive) 확산 생성 모델을 적용하여 리간드의 3차원 구조를 생성하는 단계;

상기 생성된 리간드의 결합 친화도 및 구조적 유사성을 평가하는 단계;

상기 평가 결과에 따라 리간드의 3차원 구조를 최적화하여 최종 리간드를 생성하는 단계; 를 포함하는 리간드 분자 구조 생성 방법.

【청구항 2】

제1항에 있어서,

상기 파마코포어 인식 및 식별 단계는, PLIP(Protein-Ligand Interaction Profiler)를 이용하여 단백질과의 상호작용이 높은 리간드 부위로 인식 및 식별되는 것인 리간드 분자 구조 생성 방법.

【청구항 3】

제1항에 있어서,

상기 비자기회귀적(Non-Autoregressive) 확산 생성 모델은 정방향 과정(Forward Process)에 의하여 리간드에 노이즈를 추가하고, 역방향 과정(Reverse Process)에 의하여 점진적으로 노이즈를 제거하여 결합 포켓과 최적화된 3차원 리간드 구조를 형성하는 것인 리간드 문자 구조 생성 방법.

【청구항 4】

제1항에 있어서,

상기 SE(3)-동치 신경망(SE(3)-Equivalent Neural Network)은 리간드의 회전 및 이동에 영향을 받지 않도록 설계되어, 단백질 결합 포켓에 적합한 3차원 구조를 형성하는 것인 리간드 문자 구조 생성 방법.

【청구항 5】

제2항에 있어서,

상기 파마코포어 인식 및 식별 단계는, 단백질-리간드 복합체로부터 수집한 결합 정보를 이용하여 결합에 중요한 리간드 원자의 인덱스를 상호작용 집합으로 정의하고, 상기 상호작용 집합 내 원자의 위치 및 결합 상태를 파마코포어로 선정하는 것인 리간드 문자 구조 생성 방법.

【청구항 6】

제5항에 있어서,

상기 파마코포어 인식 및 식별 단계는, 리간드 분자를 다수의 조각으로 분해한 후, 각 조각의 원자 집합(A_i)과 원자 간의 엣지 정보(E_i)를 사용하여 단백질과 결합 가능성이 높은 조각을 선택하고, 상기 조각을 병합하여 파마코포어를 형성하는 것인 리간드 분자 구조 생성 방법.

【청구항 7】

제6항에 있어서,

상기 파마코포어 인식 및 식별 단계는, 파마코포어로 정의된 조각들 간의 상호작용을 평가하여 최종적으로 단백질과의 결합 친화도를 극대화하는 파마코포어 리스트를 생성하고, 상기 파마코포어 리스트를 기반으로 비자기회귀적 확산 생성 모델을 이용하여 리간드의 나머지 구조를 최적화하는 것인 리간드 분자 구조 생성 방법.

【청구항 8】

신약 후보 물질 도출을 위한 파마코포어(Pharmacophore) 및 확산 생성 모델을 이용하여 리간드의 3차원 분자 구조를 생성하는 시스템에 있어서,

단백질의 결합 포켓 데이터와 파마코포어 데이터를 입력하는 데이터 입력부;

상기 데이터를 기초로 하여 리간드의 파마코포어를 인식 및 식별하며,

상기 결합 포켓 데이터에 기초하여 파마코포어를 고정된 기준으로 설정하고, 나머지 리간드 구조를 형성하기 위해 SE(3)-동치 신경망(SE(3)-Equivalent Neural Network)과 비자기회귀적(Non-Autoregressive) 확산 생성 모델을 적용하여 리간드

의 3차원 구조를 생성하는 파마코포어 기반 인공지능 모델; 및

상기 생성된 리간드의 결합 친화도 및 구조적 유사성을 평가에 따라 리간드의 3차원 구조를 출력하는 문자 구조 출력부;를 포함하는 리간드 문자 구조 생성 시스템.

【청구항 9】

제8항에 있어서,

상기 파마코포어 기반 인공지능 모델은, PLIP(Protein-Ligand Interaction Profiler)를 이용하여 단백질과의 상호작용이 높은 리간드 부위로 인식 및 식별하는 것인 리간드 문자 구조 생성 시스템.

【청구항 10】

제8항에 있어서,

상기 비자기회귀적(Non-Autoregressive) 확산 생성 모델은, 정방향 과정(Forward Process)에 의하여 리간드에 노이즈를 추가하고, 역방향 과정(Reverse Process)에 의하여 점진적으로 노이즈를 제거하여 결합 포켓과 최적화된 3차원 리간드 구조를 형성하는 것인 리간드 문자 구조 생성 시스템.

【청구항 11】

제8항에 있어서,

상기 SE(3)-동치 신경망(SE(3)-Equivalent Neural Network)은, 리간드의 회전 및 이동에 영향을 받지 않도록 설계되어, 단백질 결합 포켓에 적합한 3차원 구

조를 형성하는 것인 리간드 문자 구조 생성 시스템.

【청구항 12】

제9항에 있어서,

상기 파마코포어 기반 인공지능 모델은, 단백질-리간드 복합체로부터 수집한 결합 정보를 이용하여 결합에 중요한 리간드 원자의 인덱스를 상호작용 집합으로 정의하고, 상기 상호작용 집합 내 원자의 위치 및 결합 상태를 파마코포어로 선정하는 것인 리간드 문자 구조 생성 시스템.

【청구항 13】

제12항에 있어서,

상기 파마코포어 기반 인공지능 모델은, 리간드 문자를 다수의 조각으로 분해한 후, 각 조각의 원자 집합(A_i)과 원자 간의 엣지 정보(E_i)를 사용하여 단백질과 결합 가능성이 높은 조각을 선택하고, 상기 조각을 병합하여 파마코포어를 형성하는 것인 리간드 문자 구조 생성 시스템.

【청구항 14】

제13항에 있어서,

상기 파마코포어 기반 인공지능 모델은, 파마코포어로 정의된 조각들 간의 상호작용을 평가하여 최종적으로 단백질과의 결합 친화도를 극대화하는 파마코포어 리스트를 생성하고, 상기 파마코포어 리스트를 기반으로 비자기회귀적 확산 생성 모델을 이용하여 리간드의 나머지 구조를 최적화하는 것인 리간드 문자 구조 생성

시스템.

【요약서】

【요약】

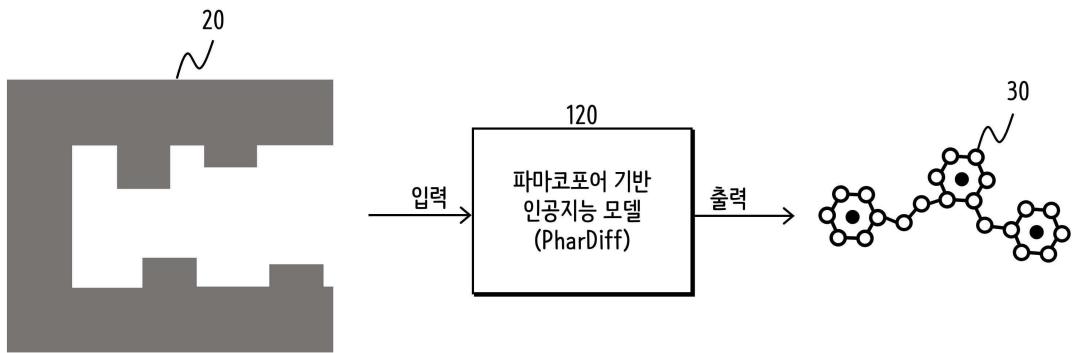
본 발명은 단백질-리간드 상호작용을 고려하여 타겟 단백질의 결합 포켓과 결합 가능한 3차원 리간드 분자 구조를 효율적으로 설계하고 최적화하는 신약 후보 물질 도출을 위한 파마코포어(Pharmacophore)를 기반의 3차원 분자 생성 방법 및 시스템에 관한 것으로, 그 구성은 단백질의 결합 포켓 데이터를 수집하여 입력하는 단계, 리간드의 파마코포어를 인식 및 식별하는 단계, 상기 결합 포켓 데이터에 기초하여 파마코포어를 고정된 기준으로 설정하고, 나머지 리간드 구조를 형성하기 위해 SE(3)-동치 신경망(SE(3)-Equivalent Neural Network)과 비자기회귀적(Non-Autoregressive) 확산 생성 모델을 적용하여 리간드의 3차원 구조를 생성하는 단계, 상기 생성된 리간드의 결합 친화도 및 구조적 유사성을 평가하는 단계, 상기 평가 결과에 따라 리간드의 3차원 구조를 최적화하여 최종 리간드를 생성하는 단계를 포함한다.

【대표도】

도 1

【도면】

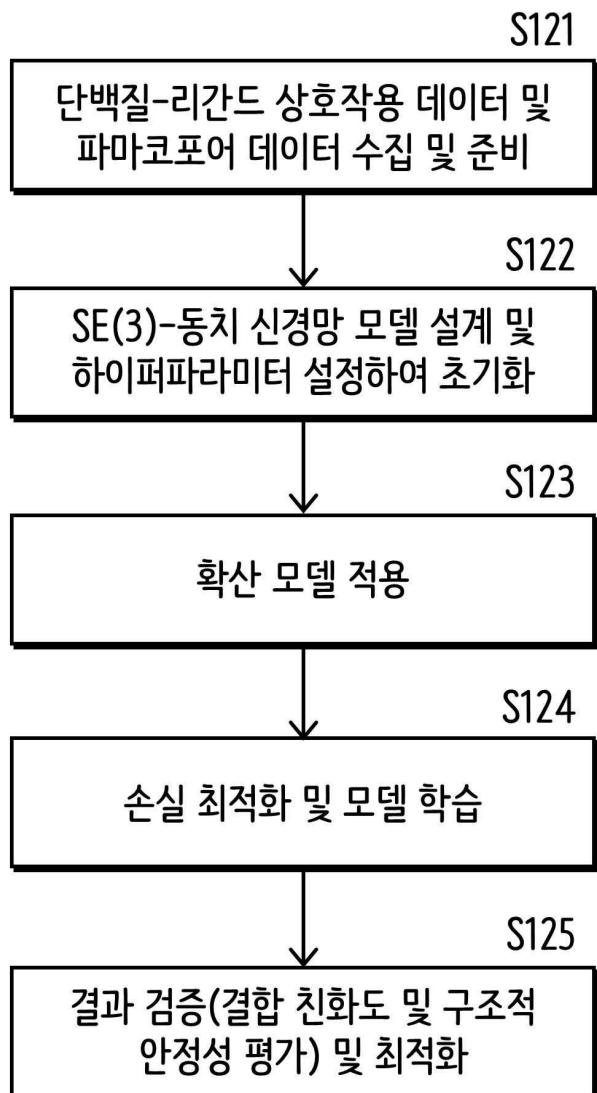
【도 1】



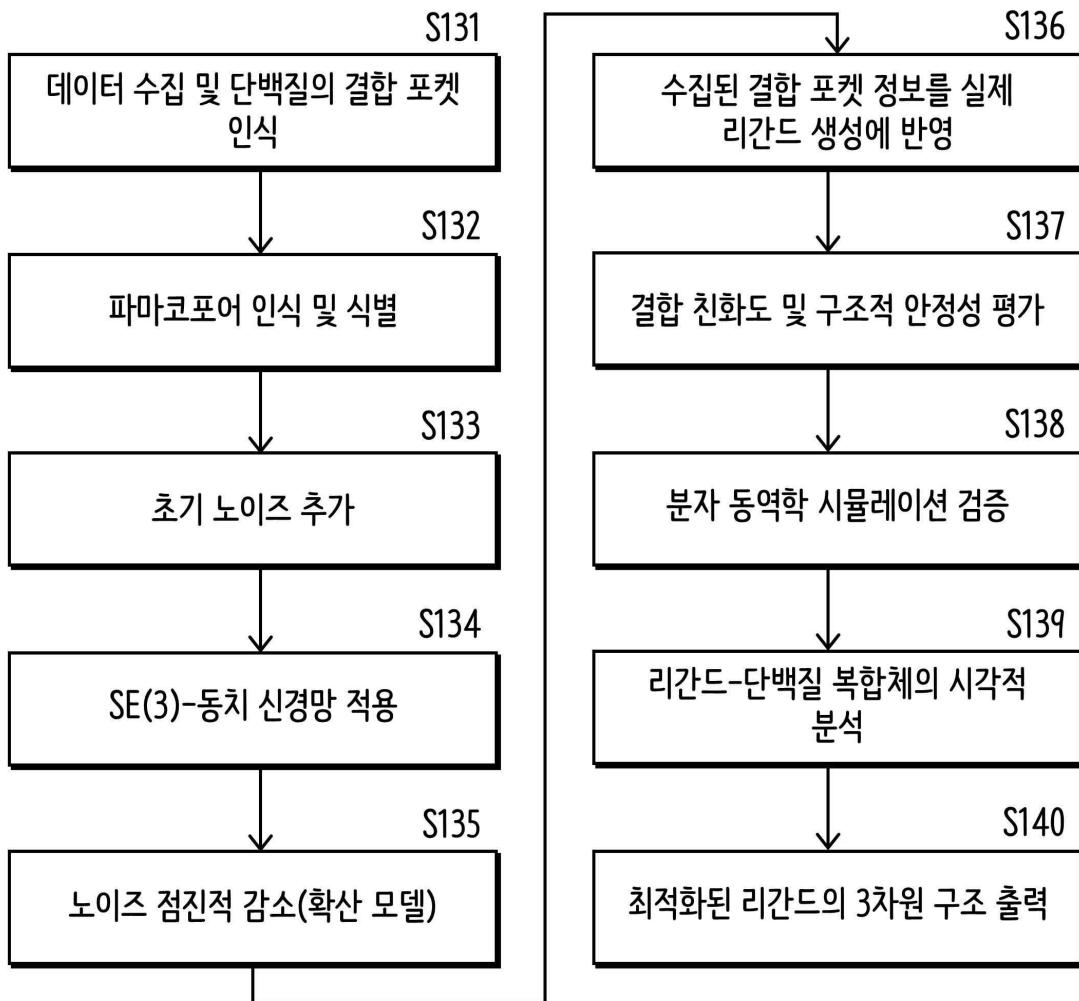
【도 2】



【도 3】



【도 4】



【도 5】

Algorithm 1 3D Pharmacophore Decomposition

Input: Set of Protein-Ligand Pair $\mathcal{A} : \{(\mathbf{p}_i, \mathbf{m}_i)\}_{i=1}^N$

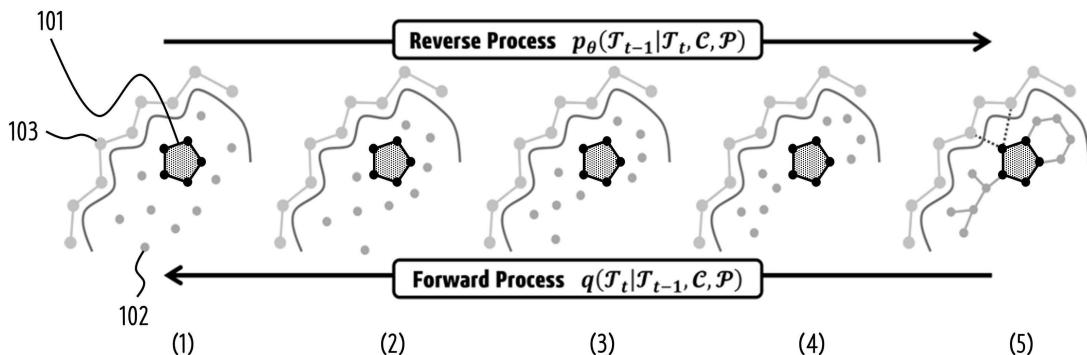
Output: Pharmacophore List

```

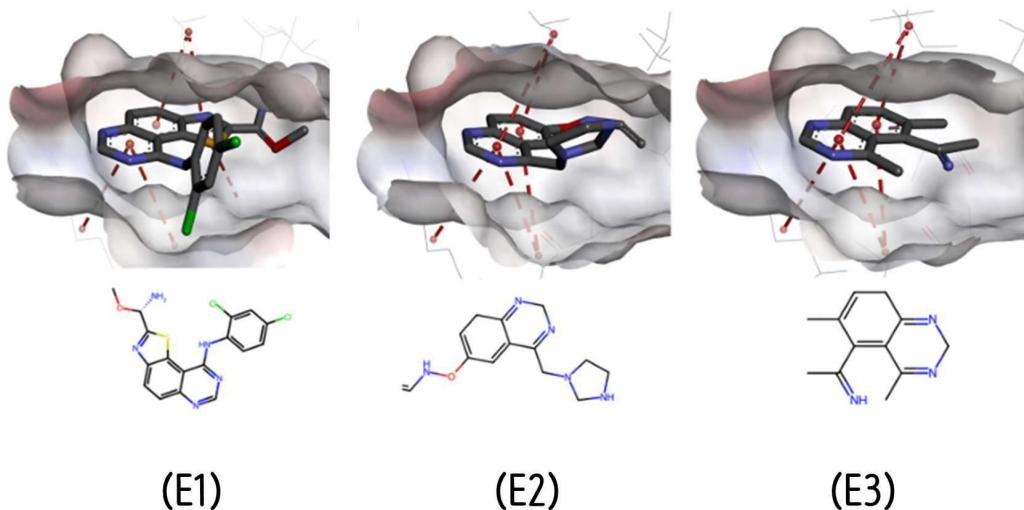
1: Consider only one interaction case
2: GetBinding  $\leftarrow$  Executes interaction profiler
3: TreeFrag  $\leftarrow$  Executes 3D Fragmentizing
4: GetPharm  $\leftarrow$  Merges multiple nodes into a fragment
5: TotalPharList  $\leftarrow$  Empty list
6: for  $(\mathbf{p}, \mathbf{m}) \in \mathcal{C}$  do
7:   PharList  $\leftarrow$  Empty list
8:    $\{a_k\}_{k=1}^K \leftarrow \text{GetBinding}(\mathbf{p}, \mathbf{m})$ 
9:    $\{\{(A_j, E_j), \mathcal{E}\}\}_{j=1}^J \leftarrow \text{TreeFrag}(\mathbf{m})$ 
10:  for  $k \in K$  do
11:    IntersectionList  $\leftarrow$  empty list
12:    for  $j \in J$  do
13:      if  $a_k \in A_j$  then
14:        IntersectionList.append( $F_j$ )
15:      end if
16:    end for
17:    Pharmacophore  $\leftarrow \text{GetPharm}(\text{IntersectionList})$ 
18:    PharList.append(Pharmacophore)
19:  end for
20:  TotalPharList.append(PharList)
21: end for

```

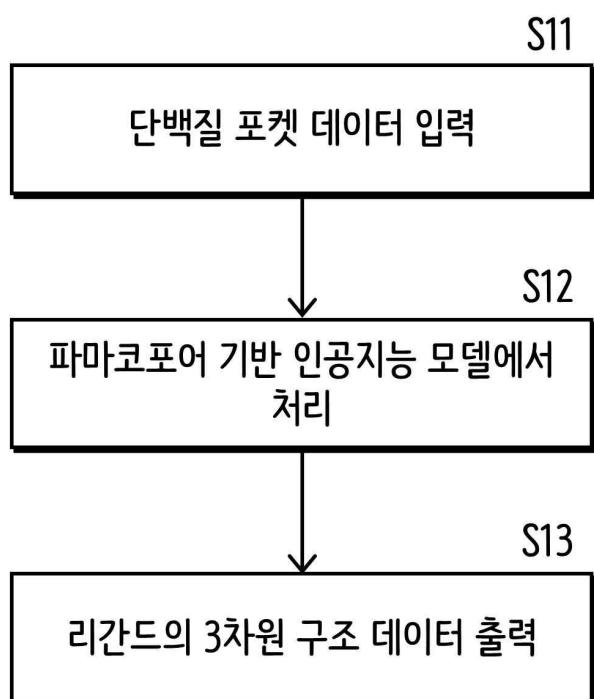
【도 6】



【도 7】



【도 8】



【도 9】

Method	Metric		Vina Dock (↓)		Vina Min (↓)		Vina Score (↓)		High Affinity (↑)		Vina Score ^{SA} (↓)		SR(↑)
	Avg.	Med.	Avg.	Med.	Avg.	Med.	Avg.	Med.	Avg.	Med.	Avg.	Med.	
TestSet	-7.45	-7.26	-6.71	-6.49	-6.36	-6.46	-	-	-6.28	-6.34	-	-	
AR	-6.75	-6.62	-6.18	-5.88	-5.75	-5.64	0.379	0.310	-5.59	-5.48	74.7%		
liGAN	-6.33	-6.2	N/A	N/A	N/A	N/A	0.211	0.111	N/A	N/A	68.4%		
GraphBP	-4.80	-4.70	N/A	N/A	N/A	N/A	0.142	0.067	N/A	N/A	57.1%		
Pocket2Mol	-7.15	-6.79	-6.42	5.82	-5.14	-4.70	0.483	0.510	-5.12	-5.48	88.7%		
DiffSBDD	-6.65	-7.15	16.45	0.49	52.78	44.57	0.452	0.454	51.53	43.46	83.0%		
DrugGPS	-3.73	-4.72	6.33	-2.56	28.18	11.76	0.123	0.000	27.32	11.13	48.1%		
TargetDiff	-7.80	-7.91	-6.64	-6.83	-5.47	-6.30	0.579	0.625	-5.31	-6.13	91.9%		
ResGen	-4.90	-5.26	-1.53	-3.36	13.79	5.75	0.232	0.000	13.73	5.75	40.7%		
PharDiff (w/o G)	-7.79	-7.87	-6.68	-6.79	-5.37	-6.18	0.588	0.612	-5.14	-5.89	89.1%		
PharDiff	-8.16	-8.00	-7.20	-7.01	-6.64	-7.08	0.622	0.636	-5.93	-6.49	98.1%		

(T2) ↵	C-C	C=C	C-N	C=N	C-O	C=O	C:C	
	AR	0.609	0.620	0.474	0.635	0.492	0.558	0.451
liGAN	0.601	0.665	0.634	0.749	0.656	0.661	0.497	
Pocket2Mol	0.496	0.561	0.416	0.629	0.454	0.516	0.416	
DiffSBDD	0.395	0.327	0.353	0.399	0.356	0.420	0.339	
DrugGPS	0.519	0.614	0.557	0.690	0.610	0.674	0.535	
TargetDiff	0.369	0.505	0.363	0.550	0.421	0.461	0.263	
ResGen	0.426	0.334	0.317	0.399	0.382	0.470	0.402	
PharDiff	0.389	0.293	0.324	0.333	0.369	0.412	0.184	

【도 10】

900

