

**【서지사항】****【서류명】** 특허출원서**【참조번호】** P20240518KR**【출원구분】** 특허출원**【출원인】****【명칭】** 연세대학교 산학협력단**【특허고객번호】** 2-2005-009509-9**【대리인】****【명칭】** 특허법인 지담**【대리인번호】** 9-2018-100261-1**【지정된변리사】** 김재흥**【포괄위임등록번호】** 2023-020263-8**【발명의 국문명칭】** 신약 후보 물질 도출을 위한 물리 지식 기반 3차원 분자 생성 방법 및 시스템**【발명의 영문명칭】** SYSTEM FOR GENERATING 3D MOLECULES BASED ON PHYSICAL KNOWLEDGE 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

**【공지에외적용대상증명서류의 내용】**

**【공개형태】** PIDiff 논문발표  
**【공개일자】** 2024.07.26

**【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】 42 면 0 원

【우선권주장료】                    0        건                                    0        원

|         |    |   |         |   |
|---------|----|---|---------|---|
| 【심사청구료】 | 12 | 항 | 778,000 | 원 |
|---------|----|---|---------|---|

【합계】 824,000원

【감면사유】 전담조직(50%감면)[1]

【감면후 수수료】 412,000 원

【첨부서류】 1. 공지에외적용대상(신규성상실의예외, 출원시의특례)규정을 적용받기 위한 증명서류\_1통

1 : 공지에외적용대상(신규성상실의예외, 출원시의특례)규정을 적용받기 위한 증명

서류

[PDF 파일 첨부](#)

## 【발명의 설명】

### 【발명의 명칭】

신약 후보 물질 도출을 위한 물리 지식 기반 3차원 분자 생성 방법 및 시스템{SYSTEM FOR GENERATING 3D MOLECULES BASED ON PHYSICAL KNOWLEDGE FOR DRUG CANDIDATE DISCOVERY AND METHOD THEREOF}

### 【기술분야】

【0001】 본 발명은 신약 후보 물질 도출을 위한 물리 지식 기반 3차원 분자 생성 방법 및 시스템에 관한 것으로, 더욱 상세하게는 물리화학적 원리를 기반으로 확산 모델 및 SE(3)-등변 신경망을 이용하여 단백질-리간드 간의 결합 자유 에너지를 최소화하는 3차원 분자를 생성하고, 최적화하는 기술이다.

### 【발명의 배경이 되는 기술】

【0002】 단백질과 리간드 간의 결합은 신약 개발에서 중요한 과정 중 하나로, 단백질의 활성 부위에 적합한 리간드를 설계하는 것이 핵심이다. 기존의 신약 후보 물질 도출 과정에서는 주로 기하학적 정보를 기반으로 한 분자 모델링이 사용되었다. 이러한 모델들은 1차원 문자열 또는 2차원 구조 기반의 예측 기법을 사용하여 분자의 결합 특성을 평가했으나, 단백질-리간드 간의 3차원 상호작용을 충분히 반영하지 못하는 한계를 가진다.

【0003】 특히, 기존의 기술들은 물리적 상호작용을 충분히 고려하지 않고, 단순히 분자의 기하학적 정보에 의존하여 분자 구조를 예측 및 설계했다. 이로 인

해 생성된 분자가 실제 단백질과 결합할 때 결합 자유 에너지가 충분히 최소화되지 않으며, 결합 안정성이 떨어지는 문제가 발생할 수 있다. 또한, 기존의 자기회귀적 방식은 분자의 각 부분을 순차적으로 생성하는 방식으로, 누적 오차가 발생할 가능성이 높고, 전체 분자의 구조적 일관성을 유지하는 데에도 한계가 있었다.

【0004】 특히, 기존의 1D 또는 2D 기반 분자 생성 방식은 3차원 공간에서의 상호작용을 반영하지 못하므로, 단백질의 결합 포켓 내에서 발생하는 원자 간 복잡한 상호작용을 제대로 모델링하지 못했다. 물리적 화학적 원리가 반영되지 않기 때문에, 생성된 분자가 실제 결합에서 안정성을 유지하지 못할 가능성이 있다.

【0005】 결과적으로, 이러한 한계들은 신약 후보 물질의 설계 과정에서 더 현실적인 3차원 분자 구조를 생성하고, 결합 안정성을 유지할 수 있는 기술의 필요성을 시사한다. 물리적 상호작용을 기반으로 단백질-리간드 간의 결합 자유 에너지를 최소화하며, 분자의 기하학적 특성을 고려한 최적화된 3차원 구조를 생성할 수 있는 새로운 접근 방식이 요구된다.

【0006】 따라서, 물리적 화학적 원리를 반영하여 결합 자유 에너지가 최소화된 3차원 분자 구조를 생성할 수 있는 시스템의 필요성이 있다. 또한, 단백질-리간드 간의 상호작용을 효율적으로 평가하고 최적화하여 구조 기반 약물 설계(SBDD)에 적용될 수 있는 기술의 개발이 시급하다.

【선행기술문헌】

【특허문헌】

【0007】(특허문헌 0001) 한국 공개특허공보 제10-2024-0082885호 (2024. 06. 11.)

(특허문헌 0002) 한국 등록특허공보 제10-2558546호 (2023. 07. 18.)

## 【발명의 내용】

### 【해결하고자 하는 과제】

【0008】본 발명이 이루고자 하는 기술적 과제는 물리적 상호작용을 기반으로 단백질-리간드 간의 결합 자유 에너지를 최소화하는 3차원 분자 구조를 효율적으로 생성하고, 이를 자동으로 평가 및 최적화할 수 있는 시스템 및 방법을 제공하는 것이다.

【0009】또한, 본 발명이 이루고자 하는 기술적 과제는 비자기회귀적 샘플링 방식과 SE(3)-등변 신경망을 적용하여 분자의 회전 및 이동에 관계없이 일관된 분자 구조를 유지하면서 최적화할 수 있는 시스템을 구현하는 것이다.

【0010】또한, 본 발명이 이루고자 하는 기술적 과제는 생성된 리간드가 실제 생체 환경에서 안정적으로 결합할 수 있도록 리간드 구조의 물리적 안정성을 개선하는 방법을 제공하는 것이다.

【0011】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】

【0012】 상기 기술적 과제를 달성하기 위하여, 본 발명의 실시예에 따른 신약 후보 물질 도출을 위한 물리 지식 및 확산 모델을 이용하여 리간드의 3차원 분자 구조를 생성하는 방법은, 단백질 포켓의 3차원 데이터를 입력받는 단계; 물리 지식 기반 인공 지능 모델을 이용하여 상기 단백질 포켓 데이터를 기초로 가우시안 노이즈를 주입하여 초기 상태의 리간드 구조를 설정하는 단계; 물리 지식 기반 인공 지능 모델을 이용하여 상기 초기 상태의 리간드 구조에서 노이즈를 제거하는 역방향 과정에서 SE(3)-등변성을 적용하여, 상기 리간드 구조의 결합 자유 에너지를 최소화하도록 최적화하는 단계; 평가 결과에 따라 상기 리간드의 구조를 최종 출력하는 단계를 포함하며, 상기 물리 지식 기반 인공 지능 모델은 확산 모델을 이용하여 구현될 수 있다.

【0013】 여기서, 상기 결합 자유 에너지를 최소화하도록 최적화하는 단계는, 렌너드-존스(Lennard-Jones) 잠재 에너지를 계산하여 단백질과 리간드의 결합 안정성을 확보할 수 있다.

【0014】 또한, 상기 결합 자유 에너지를 최소화하도록 최적화하는 단계는 비자기회귀적 샘플링 방식(Non-Autoregressive Sampling Method)을 이용하여 리간드의 전체 구조를 한 번에 생성할 수 있다.

【0015】 또한, 상기 비자기회귀적 샘플링 방식(Non-Autoregressive Sampling Method)에 의해 결합 각도와 결합 길이를 포함한 분자의 기하학적 일관성을 유지하면서 분자를 생성하도록 할 수 있다.

【0016】 또한, 상기 리간드 구조의 결합 친화도를 평가하는 단계와, 상기 결합 친화도 평가 단계에서 타겟 단백질과의 결합 친화도를 평가하여 분자의 안정성을 검증하는 평가 단계를 더 포함할 수 있다.

【0017】 또한, 상기 최종 출력 단계에서 결합 자유 에너지가 최소화된 리간드 구조를 신약 후보 물질로서 출력하는 것이다.

【0018】 상기 기술적 과제를 달성하기 위하여 본 발명의 실시예에 따른 신약 후보 물질 도출을 위한 물리 지식 및 확산 모델을 이용하여 리간드의 3차원 분자 구조를 생성하는 시스템은, 단백질 포켓의 3차원 데이터를 입력하는 데이터 입력부; 상기 데이터를 기초로 하여 가우시안 노이즈를 주입하여 초기 상태의 리간드 구조를 설정하며, 상기 초기 상태의 리간드 구조에서 노이즈를 제거하는 역방향 과정에서 SE(3)-등변성을 적용하여, 상기 리간드 구조의 결합 자유 에너지를 최소화하도록 최적화하는 리간드의 3차원 구조를 생성하는 물리 지식 기반 인공지능 모델 및 상기 생성된 리간드 구조의 결합 친화도 및 구조적 유사성 평가에 따라 리간드의 3차원 구조를 최종 출력하는 분자 구조 출력부를 포함하며, 상기 물리 지식 기반 인공지능 모델은 확산 모델을 이용하여 구현될 수 있다.

【0019】여기서, 상기 물리 지식 기반 인공지능 모델은, 렌너드-존스 (Lennard-Jones) 잠재 에너지를 계산하여 단백질과 리간드의 결합 안정성을 확보할 수 있다.

【0020】또한, 상기 물리 지식 기반 인공지능 모델은, 비자기회귀적 샘플링 방식(Non-Autoregressive Sampling Method)을 이용하여 리간드의 전체 구조를 한번에 생성할 수 있다.

【0021】또한, 상기 비자기회귀적 샘플링 방식(Non-Autoregressive Sampling Method)에 의해 결합 각도와 결합 길이를 포함한 분자의 기하학적 일관성을 유지하면서 분자를 생성할 수 있다.

【0022】또한, 상기 물리 지식 기반 인공지능 모델은, 상기 리간드 구조의 결합 친화도를 평가하고, 상기 결합 친화도 평가에서 타겟 단백질과의 결합 친화도를 평가하여 분자의 안정성을 검증할 수 있다.

【0023】또한, 상기 분자 구조 출력부에서 결합 자유 에너지가 최소화된 리간드 구조를 신약 후보 물질로서 출력할 수 있다.

### 【발명의 효과】

【0024】본 발명의 실시예에 따르면, 신약 후보 물질 도출을 위한 물리 지식 기반 3차원 분자 생성 방법 및 시스템은 물리 지식 기반 인공지능 모델(PIDiff)을 활용하여 단백질 포켓과 결합할 수 있는 리간드 구조를 자동으로 생성할 수 있고, 이로써, 타겟 단백질에 최적화된 리간드를 설계함으로써, 높은 결합 친화도와 안정

성을 갖춘 신약 후보 물질을 효과적으로 생성할 수 있다.

【0025】 또한, 본 발명의 실시예에 따르면, SE(3)-등변성을 적용하여 리간드가 3차원 공간 내에서 회전이나 평행 이동이 발생하더라도 결합 구조의 안정성을 유지할 수 있고, 이로써, 다양한 환경 변화에도 불구하고 결합 친화도를 유지하는 리간드를 생성할 수 있으며, 신약 개발 과정에서의 안정성과 재현성을 높일 수 있다.

【0026】 또한, 본 발명의 실시예에 따르면, 비자기회귀적 샘플링 방식을 통해 리간드 전체 구조를 한 번에 생성하므로, 기존 자기회귀적 모델이 가지는 누적 오류 문제를 효과적으로 방지하며, 생성된 분자 구조의 정확성과 효율성을 높여, 신약 개발 초기 단계에서의 분자 설계 시간을 단축할 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 물리 지식 기반의 3차원 분자 생성 방법 및 시스템을 도시한 개요도이다.

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

도 3은 본 발명의 실시예에 따른 물리 지식 기반 인공지능 모델의 생성 및

학습 방법을 도시한 도면이다.

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

도 5는 본 발명의 실시예에 따른 물리 지식 기반 인공지능 모델이 물리 지식 기반 최적화에 연관되는 수식을 도시한 참조도이다.

도 6은 본 발명의 실시예에 따른 물리 지식 기반 인공지능 모델(PIDiff)의 작동 방식을 도시한 도면이다.

도 7은 본 발명의 실시예에 따른 분자 생성 시스템의 작동 순서도를 도시한 도면이다.

도 8은 본 발명의 실시예에 따른 신약 후보 물질 도출을 위한 물리 지식 기반 3차원 분자 생성 시스템의 인공지능 모델(PIDiff)이 생성한 분자의 성능을 평가한 참조도이다.

도 9는 본 발명의 실시예에 따른 디스크립터 생성 방법 및 생성 장치를 구현하는 컴퓨팅 장치를 도시한 도면이다.

### 【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 부

분에 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 중간에 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 상호 호환적으로 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】 이하 첨부된 도면을 참고하여 본 발명의 실시예를 상세히 설명하기

로 한다.

【0034】 도 1은 신약 후보 물질 도출을 위한 물리 지식 기반의 3차원 분자 생성 방법 및 시스템을 도시한 개요도이다.

【0035】 구체적으로, 물리지식 기반 3D 분자 생성 시스템이 물리 지식 기반 인공지능 모델(PIDiff)(120)을 이용하여 단백질 포켓에 최적화된 리간드를 생성하는 개요를 나타낸다. 도 1에 도시된 바와 같이, 단백질 포켓(20) 데이터가 입력되고, 인공지능 모델(120)에 의해 처리되어 최종적으로 리간드(30)의 구조를 생성 및 최적화한다. 물리지식 기반 3D 분자 생성 시스템은 단백질 포켓(20)과 결합할 수 있는 최적의 리간드(30)의 구조를 자동으로 생성하며, 높은 결합 친화도와 안정성을 가지는 분자 구조를 설계한다.

【0036】 먼저, 단백질 포켓(20)은 단백질 내에서 리간드가 결합할 수 있는 활성 부위이다. 상기 단백질 포켓(20)의 구조 데이터가 물리 지식 기반 인공지능 모델(120)에 입력으로 제공된다. 상기 데이터는 단백질 내에서 리간드가 결합할 수 있는 활성 부위의 3차원 좌표 정보를 포함하며, 인공지능 모델의 초기 학습 및 최적화 과정에서 중요한 역할을 한다. 또한, 상기 데이터는 단백질의 구조적 특징과 결합 친화도에 대한 기본 정보를 제공하여 리간드가 타겟 부위에 최적화되도록 돕는다.

【0037】 물리 지식 기반 인공지능 모델(120)은 입력된 단백질 포켓 데이터를 기반으로, 리간드(30)의 구조를 생성 및 최적화하는 과정을 수행한다. 물리 지식 기반 인공지능 모델(120)은 물리화학적 상호작용을 고려하여 결합 자유 에너지를

최소화하면서 리간드를 설계하는 것을 목표로 한다.

【0038】 상기 과정에서 물리 지식 기반 인공지능 모델(120)은 단백질과 리간드 사이의 안정 상태에 대한 화학적 원칙을 반영한 확산 기반 생성 모델(Diffusion-Based Generative Model)로서 순방향 과정과 역방향 과정을 수행한다. 먼저, 순방향 과정(Forward Process)은 리간드에 무작위 노이즈를 주입하여 다양한 구조를 학습하며, 단백질과의 잠재적 결합 가능성을 탐색하는 과정이다. 한편, 역방향 과정(Reverse Process)은 노이즈를 제거하며 결합 자유 에너지가 최소화되는 방향으로 리간드의 구조를 최적화한다. 또한, SE(3)-등변 신경망(SE(3)-Equivariant Neural Networks)의 SE(3)-등변성을 적용하여, 분자가 회전 및 평행 이동에 대해 불변성을 유지하며 최적의 결합 상태를 형성하도록 한다.

【0039】 리간드(30)는 상기 물리 지식 기반 인공지능 모델(120)에 의해 최적화된 리간드의 3차원 구조이다. 상기 출력된 리간드(30)는 단백질 포켓(20)과의 높은 결합 친화도를 가지며, 화학적 안정성과 물리적 적합성을 만족하는 구조로 설계된다. 이로써, 상기 리간드(30)는 신약 후보 물질로서의 활용 가능성을 가지며, 본 발명의 물리 지식 기반 3D 분자 생성 시스템은 최적화된 리간드 구조를 자동으로 생성하여 신약 개발의 초기 단계에서 유용하게 사용될 수 있다.

【0040】 도 2는 본 발명의 실시예에 따른 물리 지식 기반 3D 분자 생성 시스템의 구성을 도시한 도면이다.

【0041】 도 2를 참조하면, 데이터 입력부(110), 물리 지식 기반 인공지능 모델(120) 및 분자 구조 출력부(130)를 포함할 수 있다.

【0042】 먼저, 데이터 입력부(110)는 물리적 특성 데이터를 입력받아 물리 지식 기반 인공지능 모델(120)에 제공하는 기능을 수행한다. 상기 데이터는 단백질의 구조 정보, 결합 자유 에너지 및 결합에 관련된 물리적·화학적 조건들을 포함할 수 있다.

【0043】 물리 지식 기반 인공지능 모델(120)은 상기 데이터 입력부(110)로부터 입력된 데이터에 따라 3차원 분자 구조를 생성한다. 상기 물리 지식 기반 인공지능 모델(120)은 결합 자유 에너지를 최소화하는 방향으로 분자의 3차원 구조를 최적화하며, 비자기회귀적 샘플링 방식(Non-Autoregressive Sampling Method)을 통해 전체 분자를 한 번에 생성한다. 또한, SE(3)-등변 신경망(SE(3)-Equivariant Neural Networks)에 의한 SE(3)-등변성을 유지하여, 회전 및 평행 이동에 대한 불변성을 확보하며, 생성된 분자가 단백질 포켓에 높은 결합 친화도를 가지도록 학습된다.

【0044】 분자 구조 출력부(130)는 상기 물리 지식 기반 인공지능 모델(120)이 생성한 3차원 분자 구조를 출력한다. 출력된 분자는 타겟 단백질과 높은 결합력을 가지는 최적화된 분자 구조로 생성되어 신약 후보 물질로 활용될 수 있다.

【0045】 도 3은 본 발명의 실시예에 따른 물리 지식 기반 인공지능 모델의 생성 및 학습 방법을 도시한 도면이다.

【0046】 단계(S121)에서 상기 물리 지식 기반 인공지능 모델(120)은, 벤치마크 데이터셋(예를 들어, CrossDocked2020 데이터셋 등)을 이용하여 단백질과 리간드의 3차원 구조 데이터를 수집한다. 상기 수집된 데이터는 단백질 포켓의 3차원 좌표와 각 원자의 위치 정보를 포함하며, 학습을 위해 데이터 정규화 및 전처리 과정이 수행된다. 데이터 전처리는 3D 좌표의 정규화와 단백질 포켓 내 원자 위치 정보를 포함하며, 이후 단계에서 인공지능 모델이 정확한 결합 구조를 학습하는 데 중요한 역할을 한다.

【0047】 단계(S122)에서 상기 물리 지식 기반 인공지능 모델(120)은, 확산 모델(Forward Process)을 이용하여 리간드의 초기 구조에 노이즈를 주입하며, 다양한 분자 구조를 탐색하고 학습을 시작한다. 상기 노이즈 주입 과정에서 리간드의 원자 위치와 타입에 가우시안 확률 분포 기반의 노이즈가 주입되어 초기 상태가 무작위로 설정된다. 단계(S122)는 모델이 여러 구조적 가능성을 학습하도록 하여, 최종적으로 단백질 포켓에 최적화된 리간드 구조를 생성할 수 있도록 한다.

【0048】 단계(S123)에서 상기 물리 지식 기반 인공지능 모델(120)은, 노이즈 주입에 의하여 무작위 상태로 설정된 리간드를 역방향 과정(Reverse Process)을 통해 결합 자유 에너지를 최소화하는 방향으로 최적화한다. 단계(S123)에서는 물리화학적 원리를 기반으로 결합 자유 에너지를 계산하고, 상기 결합 자유 에너지를 최소화하도록 리간드의 원자 위치를 조정한다. 상기 조정 과정에서 반데르발스 상호작용 등의 물리적 상호작용이 고려되며, 리간드와 단백질 사이의 안정적인 결합을 이루기 위해 필요한 최적의 위치를 학습한다.

【0049】 단계(S124)에서 상기 물리 지식 기반 인공지능 모델(120)은, 결합 친화도와 구조 안정성을 향상시키기 위하여, 손실 계산을 수행하고 모델 파라미터를 업데이트한다. 상기 손실 계산을 위한 손실 함수는 결합 자유 에너지 최소화와 실제 데이터와의 차이를 기초로 하여 정의된다. 이로써, 상기 물리 지식 기반 인공지능 모델(120)이 더 나은 결합 구조를 학습할 수 있도록 파라미터가 조정된다. 또한, 상기 파라미터 업데이트 과정을 반복하여 최적의 모델 상태에 도달하게 된다.

【0050】 단계(S125)에서 상기 물리 지식 기반 인공지능 모델(120)은, 상기 손실 계산을 이용한 학습이 완료된 후, 생성된 리간드가 단백질 포켓에 대해 가지는 결합 친화도를 평가한다. 상기 결합 친화도는 도킹 소프트웨어(예를 들어, AutoDock Vina 등)를 이용하여 측정되며, 상기 생성된 리간드가 타겟 단백질 포켓에 얼마나 강하게 결합할 수 있는지를 확인한다.

【0051】 단계(S126)에서 상기 물리 지식 기반 인공지능 모델(120)은, 상기 생성된 리간드의 안정성을 평가하기 위하여, 분자 동역학(MD) 시뮬레이션 및 RMSD (평균 제곱근 편차) 분석을 수행한다. 상기 분자 동역학(MD) 시뮬레이션은 리간드가 실제 환경에서 타겟 단백질과 결합한 상태에서 안정성을 유지할 수 있는지를 검증하며, RMSD 분석으로부터 리간드의 구조 변화를 수치화하여 평가한다. 이로써, 상기 물리 지식 기반 인공지능 모델(120)이 생성한 리간드가 타겟 단백질에 대해 물리적으로 안정적인 결합 상태를 유지하는지를 확인한다.

【0052】 단계(S127)에서 상기 물리 지식 기반 인공지능 모델(120)은, 평가 결과를 기초로 하여 조정 및 최종화 된다. 상기 평가 단계(S126)에서 얻은 결합 친

화도 및 안정성 결과를 반영하여 상기 물리 지식 기반 인공지능 모델(120)의 파라미터를 미세하게 조정하며, 이로써, 최종적으로 최적화된 3차원 리간드 구조를 생성할 수 있는 상기 물리 지식 기반 인공지능 모델(120)이 완성된다. 단계(S127)는 반복 학습 과정으로 상기 물리 지식 기반 인공지능 모델(120)의 성능을 지속적으로 개선하여 신약 후보 물질로 활용할 수 있는 리간드 구조를 생성하는 것을 목표로 한다.

【0053】 도 4는 본 발명의 실시예에 따른 물리 지식 기반 인공지능 모델이 리간드의 3차원 구조를 생성하는 방법을 도시한 도면이다.

【0054】 단계(S131)에서 상기 물리 지식 기반 인공지능 모델(120)은, 단백질의 3차원 구조 데이터를 불러와, 결합 포켓의 위치와 형태를 분석한다. 단백질 포켓 정보는 분자 결합 부위를 예측하는 데 필수적이며, 목표 결합 위치의 좌표를 포함한다. 상기 물리 지식 기반 인공지능 모델(120)은 상기 단백질 포켓 정보를 기초로 하여 리간드가 결합할 수 있는 이상적인 구조를 찾아내기 위한 기초 데이터를 확보한다.

【0055】 단계(S132)에서 상기 물리 지식 기반 인공지능 모델(120)은, 리간드의 초기 구조에 노이즈를 주입하여 모델이 다양한 구조적 가능성을 학습할 수 있도록 한다. 전방향 과정(Forward Process)로 불리는 상기 단계(S132)는 가우시안 확률 분포에 따라 노이즈를 주입하는 과정을 포함하며, 리간드의 원자 위치를 무작위로 변형한다. 이로써, 리간드가 초기 상태에서부터 여러 결합 가능성을 가지도록 설정하며, 다양한 구조적 특징을 학습할 수 있도록 한다.

【0056】 단계(S133)에서 상기 물리 지식 기반 인공지능 모델(120)은, 역방향 확산 과정에서 노이즈를 점진적으로 제거하여 리간드가 결합 자유 에너지가 최소화 된 상태로 점진적으로 이동하도록 유도한다. 결합 자유 에너지 최소화는 리간드가 단백질 포켓과 안정적인 결합을 형성할 수 있는 최적의 상태로 조정되기 위한 필수 과정이다. 상기 과정에서 가우시안 분포를 기초로 원래의 분자 구조가 복원되며, 확률적 예측을 이용하여 분자의 각 원자 좌표와 특성을 복원한다.

【0057】 여기서, 가우시안 확률 분포에 기초한 가우시안 프로세스는 노이즈가 점진적으로 제거되면서 원래 분자의 위치를 복원한다. 상기 물리 지식 기반 인공지능 모델(PIDiff)(120)의 역방향 과정에서 노이즈가 제거되는 단계는 가우시안 확률 분포에 따라 각 원자 좌표에 주입된 노이즈를 단계적으로 제거하며, 이로써, 분자의 3D 구조가 복원된다. 상기 노이즈 제거 과정은 확산 모델의 역방향 과정으로, 확률적 분포에 따라 원자 간의 거리와 위치를 복구하도록 한다. 이때, 확률 분포  $p_{\theta}(X_{m,t-1}|X_{m,t},X_p)$ 를 이용해 각 시점에서의 원자 좌표를 예측한다. 여기서  $X_m$  과  $X_p$ 는 각각 리간드와 단백질의 원자 좌표를 나타낸다.

【0058】 단계(S134)에서 상기 물리 지식 기반 인공지능 모델(120)은, 비자기 회귀적 샘플링 방식(Non-Autoregressive Sampling Method)을 사용하여 전체 분자 구조를 동시에 생성한다. 상기 비자기회귀적 샘플링 방식은 전체 분자를 한 번에 샘플링하여 동시에 모든 원자 좌표를 생성하므로, 효율적으로 분자를 생성할 수 있다. 상기 비자기회귀적 샘플링 방식은 원자 하나씩을 순차적으로 생성하는 종래의

자기회귀적(Autoregressive) 방식과 달리, 전체 구조를 동시에 샘플링하여, 누적 오차나 노출 편향을 줄인다.

【0059】 비자기회귀적 샘플링 방식은 자기회귀적 방식에 비해 진보성을 가진다. 자기회귀적 방식에서는 매번 이전 원자나 좌표에 의존하여 다음 원자를 샘플링해야 하므로, 전체 분자 생성 과정이 느려질 수 있다. 반면, 비자기회귀적 방식은 동시에 모든 원자 좌표를 샘플링하여 더 빠르게 분자를 생성할 수 있다.

【0060】 또한, 자기회귀적 방식에서는 이전 단계의 샘플링 오차가 다음 단계에 누적되어 최종적으로 생성되는 분자 구조의 정확도가 떨어질 수 있다. 반면, 상기 비자기회귀적 방식은 한 번에 전체 구조를 샘플링하므로 누적 오차가 발생하지 않아 더 정확한 구조를 생성할 수 있다. 특히, 분자의 복잡도가 높을수록 자기회귀적 방식에서는 오차가 누적될 가능성이 커지지만, 비자기회귀적 방식은 복잡한 화합물에도 적합하다. 상기 비자기회귀적 샘플링 방식은 결합 자유 에너지를 최소화하는 과정에서도 사용되며, 물리 화학적 원리를 통합하여 리간드-단백질 결합 안정성을 확보하는 데 중점을 둔다.

【0061】 또한, 단계(S134)에서 상기 물리 지식 기반 인공지능 모델(120)은, SE(3)-등변 신경망(SE(3)-Equivariant Neural Networks)에 의하여 SE(3)-등변성을 적용하여, 3차원 공간에서의 회전 및 이동에 대해 불변성을 유지하며, 분자가 어떠한 위치에서도 동일한 결합 친화도를 가지도록 한다. 이로써 공간적 변화에 따른 불확실성을 줄이며, 더 높은 결합 안정성을 확보할 수 있다.

【0062】 단계(S135)에서 상기 물리 지식 기반 인공지능 모델(120)은, 리간드

와 단백질 간의 상호작용을 최적화하기 위하여 결합 자유 에너지를 계산한다. 상기 결합 자유 에너지는 분자 간 상호작용에서 중요한 역할을 하며, 결합 에너지가 낮을수록 더 강한 결합을 형성할 수 있다. 모델은 이를 통해 리간드가 단백질 포켓에 안정적으로 결합할 수 있도록 에너지 상태를 조정한다.

【0063】 단계(S136)에서 상기 물리 지식 기반 인공지능 모델(120)은, 렌너드-존스 (Lennard-Jones) 잠재 에너지를 계산하여, 분자 간 거리와 상호작용을 기초로 결합의 안정성을 평가한다. 상기 렌너드-존스 잠재 에너지는 리간드와 단백질 사이의 상호작용 강도를 나타내며, 결합 거리와 관계된 반발 및 인력 요소를 포함한다. 이로써, 최적의 결합 간격을 유지하면서 결합 구조를 형성할 수 있다.

【0064】 단계(S137)에서 상기 물리 지식 기반 인공지능 모델(120)은, SE(3)-등변 신경망(SE(3)-Equivariant Neural Networks)의 SE(3)-등변성 적용하여 회전 및 평행 이동에 대해 분자 구조의 불변성을 유지하도록 최적화한다. SE(3)-등변성은 분자가 3D 공간 내에서 이동하더라도 동일한 결합 친화도와 안정성을 유지하도록 보장하는 특징으로, 이로써, 결합 구조가 외부의 물리적 변화에 강하게 작용할 수 있게 한다.

【0065】 상기 SE(3)-등변성을 적용한 역방향 과정(Equivariant Reverse Process)은 다음과 같은 방식으로 수행된다. 이 과정에서는 시간 단계  $t$ 에서  $t-1$ 로 이동하면서 리간드의 노이즈를 점진적으로 제거하여 원래의 안정적인 상태로 복원한다. 모델은 각 단계에서 SE(3)-등변성을 적용해, 분자가 3차원 공간에서의 회전이나 이동과 관계없이 동일한 물리적 결합 특성을 가지도록 보장한다.

【0066】 수식으로는 다음과 같이 표현된다:

$$\text{【0067】 } p_{\theta}(\mathcal{M}_{t-1} \mid \mathcal{M}_t, \mathcal{P}) \mathcal{N}(X_{t-1}^M; \mu_{\theta}(\mathcal{M}_t, t, \mathcal{P}), \sigma_t^2 I) \cdot \mathcal{C}(V_{t-1}^M \mid c_{\theta}(\mathcal{M}_t, t, \mathcal{P}))$$

【0068】 여기서,  $X^M$  와  $V^M$  는 각각 리간드 원자의 3D 좌표와 원자 타입을 나타내며, 이는 모델이 예측하는 주요 변수이다.

【0069】  $X^P$  와  $V^P$  는 단백질 원자의 3D 좌표와 원자 타입으로, 단백질-리간드 간의 결합 구조를 유지하는 데 사용된다.

【0070】  $\mathcal{M}$  은 가우시안 분포로, 각 원자의 좌표를 예측한다.

【0071】  $\mathcal{C}$  는 범주형 분포로, 각 원자의 유형을 예측한다.

【0072】  $\mu_{\theta}$  와  $c_{\theta}$  는 역방향 전이 커널의 매개변수로, 시간 단계  $t$ 에서  $t-1$ 로 이동할 때 SE(3)-등변성을 유지하면서 분자의 구조를 안정적인 상태로 복원하도록 한다.

【0073】 상기 과정에 의해 모델은 각 시간 단계에서 노이즈를 제거하고, 결합 자유 에너지가 최소화된 상태로 리간드 구조를 조정한다. 최종적으로, SE(3)-등변성을 유지하여 분자의 3D 좌표와 결합 특성이 외부 환경 변화에 영향을 받지 않도록 함으로써, 높은 결합 친화도와 안정성을 가진 분자 구조를 형성한다.

【0074】 단계(S138)에서 상기 물리 지식 기반 인공지능 모델(120)은, 결합 친화도를 평가하며, AutoDock Vina와 같은 도킹 소프트웨어를 사용하여 리간드가 단백질 포켓에 얼마나 강하게 결합할 수 있는지를 측정한다. 상기 결합 친화도는 생성된 분자의 유효성을 판단하는 중요한 기준이며, 리간드가 실제로 약물로서 기

능할 가능성을 평가한다.

【0075】 단계(S139)에서 상기 물리 지식 기반 인공지능 모델(120)은, 결합 안정성을 평가하기 위해 RMSD(평균 제곱근 편차) 분석과 분자 동역학(MD) 시뮬레이션을 수행한다. 분자 동역학(MD) 시뮬레이션을 이용하여 리간드가 실제 환경에서 결합을 유지할 수 있는지 확인하며, RMSD 분석으로 결합 전후의 구조적 변화를 수치화한다. 이로써, 생성된 리간드가 타겟 단백질과의 결합에서 물리적으로 안정성을 유지하는지를 평가한다.

【0076】 단계(S140)에서 상기 물리 지식 기반 인공지능 모델(120)은, 평가 결과를 바탕으로 물리 지식 기반 인공지능 모델(120)을 최종 조정하여, 타겟 단백질에 대해 최적화된 분자 구조를 생성한다. 상기 최적화된 분자 구조는 신약 후보 물질로 활용될 수 있으며, 높은 결합 친화도와 안정성을 기초로 하여 약물 개발에 사용될 수 있다.

【0077】 도 5는 본 발명의 실시예에 따른 물리 지식 기반 인공지능 모델이 물리 지식 기반 최적화에 연관되는 수식을 도시한 참조도이다.

【0078】 구체적으로, 단백질-리간드(Protein-Ligand) 상호작용에서 사용되는 힘장 방정식(Force Field Equation)과 단백질-리간드 원자 쌍 간의 합을 계산하는 경험적 힘장(Empirical Forcefield)에 대하여 도시하고 있다. 도 5의 수식들은 물리 지식 기반 인공지능 모델(PIDiff)이 분자 구조를 생성하고 최적화할 때 사용하는 물리적 원리와 관련이 있으며, 특히 결합 자유 에너지를 최소화하는 목표로 최적의 결합 구조를 형성하는 과정에서 중요한 역할을 한다.

【0079】 도 5의 (A1)의 힘장 방정식(Force Field Equation)은 리간드와 단백질 간의 상호작용을 에너지 함수로 나타낸 것으로, 결합 자유 에너지를 구성하는 요소들을 정의한다.

【0080】 도 5의 (A2)의 결합 항(Bonded Terms)에는 결합 신장(Bond Stretching), 각도 굽힘(Angle Bending), 이면각 회전(Dihedral Rotation)이 포함되며, 구체적으로, 결합 길이, 결합 각도, 그리고 회전 각도의 변화를 다루는 항목들이다.

【0081】 도 5의 (A3)의 비결합 항(Nonbonded Terms)에는 12-6 렌너드-존스 상호작용(Lennard-Jones Interactions)과 전기적 상호작용(Electrostatic Interactions)이 포함된다. 상기 렌너드-존스 상호작용은 원자 간의 반발 및 인력 관계를 나타내며, 원자 사이의 거리 변화에 따른 잠재 에너지 변화를 반영한다. 또한, 전기적 상호작용(Electrostatic Interactions)은 원자 간 전하 분포에 따른 전기적 상호작용을 나타내며, 마찬가지로 원자 간 거리와 관련이 있다.

【0082】 도 5의 (B)는 경험적 힘장(Empirical Forcefield)를 기반으로 한 단백질-리간드(Protein-Ligand)의 원자 간 상호작용 에너지를 계산하는 방식을 설명한다.

【0083】 도 5의 (B1)에서 반데르발스(Van der Waals (vdW)) 에너지는 다음과 같은 수식으로 표현된다:

【0084】 
$$E_{vdw} = \sum_{j \in P} \sum_{i \in M} \epsilon \left[ \left( \frac{d'_{ij}}{d_{ij}} \right)^{12} - 2 \left( \frac{d'_{ij}}{d_{ij}} \right)^6 \right]$$

【0085】 여기서  $d_{ij}$  는 단백질과 리간드의 특정 원자 쌍 간 거리,  $\epsilon$  은 반발 및 인력 상수를 나타낸다. 상기 식은 렌너드-존스 상호작용의 핵심 원리를 반영하여, 두 원자가 가까워질 때의 인력과 멀어질 때의 반발력을 모델링한다.

【0086】 도 5의 (B2)에서 총 에너지 계산  $L_{energy}$  은 반데르발스 상호작용을 기초로 하여, 에너지 함수의 도함수를 통해 각 원자의 위치 변화에 따른 에너지 기울기를 계산하여 최적의 에너지 상태를 찾는 과정을 거친다:

$$\text{【0087】 } L_{energy} = \sum_{j \in P} \sum_{i \in M} \frac{\partial \left( \epsilon \left[ \left( \frac{d'_{ij}}{d_{ij}} \right)^{12} - 2 \left( \frac{d'_{ij}}{d_{ij}} \right)^6 \right] \right)}{\partial d_{ij}}$$

【0088】 상기 수식을 통해 각 원자 쌍 간의 에너지 기울기를 계산하고, 최적화하여 최종적으로 안정적인 분자 구조를 형성한다.

【0089】 도 5에서 나타난 수식들은 상기 물리 지식 기반 인공지능(120)이 결합 자유 에너지를 계산하고 최적화하는 단계 및 렌너드-존스 잠재 에너지를 계산하는 단계에서 이용된다. 예를 들어, 도 4의 단계(S135)와 단계(S136)에서 상기 에너지 계산과 최적화 과정이 수행될 수 있다.

【0090】 예를 들어, 단계(S135)에서 상기 힘장 방정식을 사용하여 리간드와 단백질 간의 결합 자유 에너지를 계산하고, 최적화를 수행한다. 또한, 단계(S136)에서 상기 렌너드-존스 상호작용 수식이 사용되어, 각 원자 간의 거리와 잠재 에너지를 평가하여 최적화한다. 상기 단계들(S135, S136)로부터 최적화된 에너지는 단계(S137)에서 SE(3)-등변성을 적용한 최적화 과정에 의해, 외부 환경 변화에도 강

한 결합 구조를 생성하는 데 기여한다.

【0091】 도 6은 본 발명의 실시예에 따른 물리 지식 기반 인공지능 모델(PIDiff)의 작동 방식을 도시한 도면이다.

【0092】 구체적으로, 도 6은 물리 지식 기반 인공지능 모델(PIDiff)이 단백질과 리간드 사이의 안정 상태에 대한 화학적 원칙을 반영한 확산 기반 생성 모델(Diffusion-Based Generative Model)임을 나타낸다. 상기 물리 지식 기반 인공지능 모델은 단백질 원자의 위치를 고정된 상태에서 실제 분자(Real Molecule)의 원자 위치와 유형에 노이즈를 주입하여, 순방향 과정(Forward Process)을 통해 노이즈가 추가된 분자(Noised Molecule)로 변환하는 방식으로 작동한다. 이후, 역방향 과정(Reverse Process)을 거쳐 노이즈를 제거하여 다시 실제 분자(Real Molecule)로 되돌아가게 된다. 이로써, 물리적 안정성과 결합 친화도가 높은 3차원 분자 구조를 효과적으로 생성할 수 있음을 보여주며, 순방향 과정과 역방향 과정에 의해 결합 자유 에너지를 최적화하는 과정을 시각적으로 설명하고 있다.

【0093】 도 6에서 단백질 포켓(Protein Pocket)(101), 리간드의 원자들(Atoms of Ligand)(102) 및 3차원 분자 구조(103)를 도시하고 있다. 여기서, 최종적으로 생성된 3차원 분자 구조(103)는 실제 분자(Real Molecule)이며, 단백질 포켓(101)과 안정적으로 결합하는 신약 후보 물질로서의 잠재력을 가진다.

【0094】 도 6의 (1)은 초기 상태에서 노이즈가 주입된 분자 구조(Noised Molecule)를 나타낸다. 초기 상태에서, 리간드의 원자들(102)은 단백질 포켓(101) 내부에 노이즈가 주입된 무작위 분자 구조로 배치된다. 초기 상태에서는 물리 지식

기반 인공지능 모델(PIDiff)이 다양한 결합 가능성을 학습할 수 있도록 각 원자 위치에 가우시안 노이즈를 더하여 리간드의 구조를 변형시킨다. 상기 과정은 순방향 과정(Forward Process)이라고 하며, 단백질과 리간드 간의 다양한 상호작용 가능성을 고려하여 초기 상태를 설정하는 단계이다.

【0095】 도 6의 (5)에서부터 (1)까지의 방향에 해당하는 순방향 과정(Forward Process)에서는 리간드의 원자들(102)에 노이즈가 점진적으로 증가하면서 초기 상태의 구조가 더욱 무작위로 변형된다. 상기 순방향 과정은 물리 지식 기반 인공지능 모델이 리간드가 가질 수 있는 다양한 형태를 학습하도록 돕는다. 순방향 과정(Forward Process)은 주어진 단백질 포켓 내에서 리간드가 가능한 모든 형태를 탐색하도록 하여, 이후 역방향 과정(Reverse Process)에서의 최적화를 위한 기초 데이터를 제공한다.

【0096】 도 6의 (1)에서부터 (5)까지의 방향에 해당하는 역방향 과정(Reverse Process)는 순방향 과정(Forward Process)에서 주입된 노이즈를 점차 제거하여, 리간드가 결합 자유 에너지가 최소화되는 상태로 되돌아가는 과정을 의미한다. 상기 역방향 과정에서 물리 지식 기반 인공지능 모델은 SE(3)-등변 신경망(SE(3)-Equivariant Neural Networks)에 의하여 SE(3)-등변성을 유지하면서 원자 위치를 최적화하여 단백질 포켓 내에서 안정적인 결합을 형성하는 구조로 변환한다. 즉, 물리 지식 기반 인공지능 모델은 리간드가 단백질 포켓(101)과 최적의 결합을 이루는 상태로 수렴할 수 있도록 학습되며, 이로써 생성된 리간드는 물리적 안정성이 높다.

【0097】 여기서, 결합 자유 에너지 지형(Binding Free Energy Landscape)은 역방향 과정에서 리간드의 위치가 최적화되면서 결합 자유 에너지(Binding Free Energy)가 최소화되는 방향으로 이동하게 된다. 도 6 하단의 3D 그래프는 결합 자유 에너지와 반응 좌표(Reaction Coordinate) 사이의 관계를 나타낸다. 상기 3D 그래프는 물리 지식 기반 인공지능 모델이 노이즈 제거를 통해 결합 안정성이 높은 위치로 리간드를 이동시키는 과정을 시각화한 것이다. 여기서, 결합 자유 에너지가 낮을수록 단백질 포켓 내에서 리간드가 더 안정적으로 결합할 수 있음을 의미한다.

【0098】 도 6의 (5)에서 최종적으로 생성된 분자 구조(103)는 노이즈가 추가된 분자(Noised Molecule)에서 역방향 과정을 거쳐 노이즈가 제거되어 실제 분자(Real Molecule)로 변환된 것이다. 결합 자유 에너지가 최소화된 상태로, 상기 단백질 포켓(101) 내에서 높은 결합 친화도와 안정성을 가지는 상태로 생성된다. 상기 분자 구조(103)를 생성하는 방법은 상기 물리 지식 기반 인공지능 모델(PIDiff)이 물리 기반 최적화와 SE(3)-등변성을 적용하여 단백질과 리간드 간의 결합 친화도를 극대화하는 방법이다. 따라서, 최종적으로 생성된 분자 구조(103)는 신약 후보 물질로서의 활용 가능성이 높다.

【0099】 이로써, 본 발명의 물리 지식 기반 인공지능 모델(PIDiff)이 역방향 과정 동안 단순히 실제 분자의 원자 위치와 유형을 재구성하는데 그치지 않고, 리간드가 타겟 단백질에 안정적으로 결합하기 위해 필요한 물리화학적 원칙을 충족하도록 설계되었음을 확인할 수 있다.

【0100】 도 7은 본 발명의 실시예에 따른 분자 생성 시스템의 작동 순서도를 도시한 도면이다.

【0101】 구체적으로, 도 7은 상기 분자 생성 시스템(100)이 단백질 포켓 데이터 입력, 물리 기반 모델 처리, 리간드 구조 출력의 세 가지 주요 단계에 의해 최적화된 3차원 분자 구조를 생성하는 일련의 과정을 보여준다. 상기 분자 생성 시스템(100)은 단백질과의 결합 안정성을 극대화하기 위한 알고리즘을 기초로 하여, 고도화된 신약 개발 환경에서 활용될 수 있는 리간드 구조를 생성한다.

【0102】 먼저, 단계(S11)에서 상기 분자 생성 시스템(100)은, 데이터 입력부(110)로 단백질 포켓의 3차원 구조 데이터를 입력받는다. 상기 데이터는 단백질의 활성 부위인 리간드가 결합할 위치에 대한 정보와 관련된 좌표 데이터를 포함한다. 단백질 포켓의 구조는 리간드의 결합 특성을 결정하는 중요한 요소로, 정확한 3차원 좌표와 결합 부위의 물리적 특성이 입력되어야 한다. 따라서, 단계(S11)는 리간드가 결합할 목표 위치를 모델에 전달하는 초기 데이터 입력 단계이다.

【0103】 단계(S12)에서 상기 분자 생성 시스템(100)은, 상기 입력된 단백질 포켓 데이터를 기초로 하여 상기 물리지식 기반 인공지능 모델(120)이 작동하여 리간드의 3차원 구조를 생성하는 과정이 수행된다. 상기 물리지식 기반 인공지능 모델(120)은 단백질과 리간드 간의 상호작용에 영향을 미치는 물리적 요소를 고려하며, SE(3)-등변성 및 결합 자유 에너지 최소화 알고리즘을 적용하여 리간드 구조를 최적화한다. 상기 물리지식 기반 인공지능 모델(120)은 노이즈가 주입된 상태에서 역방향 과정에 의해 점진적으로 노이즈를 제거하여 결합 자유 에너지가 낮은 상태

로 리간드를 형성한다. 상기 형성 과정에서 단백질 포켓 내에서 리간드의 회전 및 평행 이동에 대해 불변성을 유지하도록 SE(3)-등변성이 적용되며, 리간드가 안정적으로 결합할 수 있는 구조가 생성된다.

【0104】 단계(S13)에서 상기 분자 생성 시스템(100)은, 분자 구조 출력부(130)로 상기 인공지능 모델에서 생성된 리간드의 최적화된 3차원 구조 데이터를 출력한다. 상기 출력 데이터는 단백질 포켓과 높은 결합 친화도를 가지며 물리적으로 안정적인 구조로서, 신약 개발 등에서 유용하게 활용될 수 있다.

【0105】 도 8은 본 발명의 실시예에 따른 신약 후보 물질 도출을 위한 물리 지식 기반 3차원 분자 생성 시스템의 인공지능 모델(PIDiff)이 생성한 분자의 성능을 평가한 참조도이다.

【0106】 도 8에서는 CrossDocked2020 데이터셋을 활용하여 상기 물리 지식 기반 인공지능 모델(PIDiff)(120)이 생성한 분자의 결합 친화도를 여러 기준 모델과 비교한 결과를 보여준다. 결합 친화도 평가(Binding Affinity Evaluation)는 예를 들어, Vina Score, Vina Min, VinaDock, High Affinity 등의 지표를 사용하여 측정되며, 각 지표는 생성된 분자가 단백질의 결합 포켓과 얼마나 안정적으로 결합할 수 있는지를 나타낸다. 상기 물리 지식 기반 인공지능 모델(PIDiff)(120)은 VinaScore, VinaMin, VinaDock 등의 지표에서 다른 모델보다 낮은 값을 기록하며, 비교 모델에 비하여 높은 결합 친화도와 안정성을 입증하였다. 또한, HighAffinity와 성공률(SR, Success Rate)에서도 가장 높은 성능을 나타냈다. 상기 결과는 상기 물리 지식 기반 인공지능 모델(PIDiff)(120)이 물리 기반 학습 방식을 이용하여 단

백질 포켓과의 결합 자유 에너지를 최소화한 3D 리간드 구조를 효과적으로 생성할 수 있음을 의미한다.

【0107】 도 9는 본 발명의 실시예에 따른 디스크립터 생성 방법 및 생성 장치를 구현하는 컴퓨팅 장치를 도시한 도면이다.

【0108】 도 1 ~ 도 8에 의해 설명된 본 발명의 실시예는 적어도 하나의 프로세서에 의해 동작하는 컴퓨팅 장치(900)로 구현될 수 있다.

【0109】 컴퓨팅 장치(900)는 프로세서(910), 메모리(920), 스토리지(930), 통신 인터페이스(940), 시스템 인터커넥트(950) 및 디스플레이(960)를 포함할 수 있다.

【0110】 프로세서(910)는, CPU(Central Processing Unit), MPU(Micro Processor Unit), MCU(Micro Controller Unit), GPU(Graphic Processing Unit) 및 APU(Application Processing Unit)을 포함한다.

【0111】 메모리(920)는 프로세서(910)와 상호작용하여 프로그램이 효율적으로 실행될 수 있도록 데이터를 저장하고 필요한 정보에 빠르게 접근할 수 있도록 하는 기능을 수행한다. 상기 메모리(920)는 레지스터, 캐시 메모리, 주 메모리, 읽기 전용 메모리, 가상 메모리, 비휘발성 메모리 중 적어도 하나를 포함한다.

【0112】 스토리지(930)은 데이터를 영구적으로 저장하고 관리하는 역할을 한다. 스토리지는 컴퓨팅 시스템이 꺼지거나 재부팅된 후에도 데이터를 보존하며, 운영 체제, 애플리케이션, 사용자 파일 등을 저장하는 데 사용된다.

스토리지(930)는, 하드 디스크 드라이브(HDD), 솔리드 스테이트 드라이브(SSD), 광학 디스크, 네트워크 스토리지 및 클라우드 스토리지 중 적어도 하나를 포함한다.

【0113】 통신 인터페이스(940)는 컴퓨팅 시스템 내부 및 외부의 다양한 장치들 간에 데이터를 주고받기 위한 경로를 제공한다. 상기 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.

【0114】 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.

【0115】 디스플레이(960)는 컴퓨팅 시스템의 출력 장치로서, 사용자에게 시각적인 정보를 제공하는 기능을 수행한다.

【0116】 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.

【0117】 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 파

일, 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.

【0118】 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.

【0119】 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.

【0120】 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.

【0121】 전술한 본 발명의 설명은 예시를 위한 것이며, 본 발명이 속하는 기술분야의 통상의 지식을 가진 자는 본 발명의 기술적 사상이나 필수적인 특징을 변경하지 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 구성요소들도 결합된 형태로 실시될 수 있다.

【0122】 본 발명의 범위는 후술하는 청구범위에 의하여 나타내어지며, 청구범위의 의미 및 범위 그리고 그 균등 개념으로부터 도출되는 모든 변경 또는 변형된 형태가 본 발명의 범위에 포함되는 것으로 해석되어야 한다.

**【부호의 설명】**

【0123】 20: 단백질 포켓

30: 리간드

110: 데이터 입력부

120: 물리 지식 기반 인공지능 모델

130: 분자 구조 출력부

**【청구범위】****【청구항 1】**

신약 후보 물질 도출을 위한 물리 지식 및 확산 모델을 이용하여 리간드의 3차원 분자 구조를 생성하는 방법에 있어서,

단백질 포켓의 3차원 데이터를 입력받는 단계;

물리 지식 기반 인공 지능 모델을 이용하여 상기 단백질 포켓 데이터를 기초로 가우시안 노이즈를 주입하여 초기 상태의 리간드 구조를 설정하는 단계;

물리 지식 기반 인공 지능 모델을 이용하여 상기 초기 상태의 리간드 구조에서 노이즈를 제거하는 역방향 과정에서 SE(3)-등변성을 적용하여, 상기 리간드 구조의 결합 자유 에너지를 최소화하도록 최적화하는 단계;

평가 결과에 따라 상기 리간드의 구조를 최종 출력하는 단계;를 포함하며,

상기 물리 지식 기반 인공 지능 모델은 확산 모델을 이용하여 구현되는 것인 리간드 분자 구조 생성 방법.

**【청구항 2】**

제1항에 있어서,

상기 결합 자유 에너지를 최소화하도록 최적화하는 단계는, 렌너드-존스 (Lennard-Jones) 잠재 에너지를 계산하여 단백질과 리간드의 결합 안정성을 확보하는 것인 리간드 분자 구조 생성 방법.

**【청구항 3】**

제2항에 있어서,

상기 결합 자유 에너지를 최소화하도록 최적화하는 단계는 비자기회귀적 샘플링 방식(Non-Autoregressive Sampling Method)을 이용하여 리간드의 전체 구조를 한 번에 생성하는 것인 리간드 분자 구조 생성 방법.

**【청구항 4】**

제3항에 있어서,

상기 비자기회귀적 샘플링 방식(Non-Autoregressive Sampling Method)에 의해 결합 각도와 결합 길이를 포함한 분자의 기하학적 일관성을 유지하면서 분자를 생성하도록 하는 것인 리간드 분자 구조 생성 방법.

**【청구항 5】**

제1항에 있어서,

상기 리간드 구조의 결합 친화도를 평가하는 단계와,

상기 결합 친화도 평가 단계에서 타겟 단백질과의 결합 친화도를 평가하여 분자의 안정성을 검증하는 평가 단계;

를 더 포함하는 것인 리간드 분자 구조 생성 방법.

**【청구항 6】**

제5항에 있어서,

상기 최종 출력 단계에서 결합 자유 에너지가 최소화된 리간드 구조를 신약

후보 물질로서 출력하는 것인 리간드 분자 구조 생성 방법.

### 【청구항 7】

신약 후보 물질 도출을 위한 물리 지식 및 확산 모델을 이용하여 리간드의 3차원 분자 구조를 생성하는 시스템에 있어서,

단백질 포켓의 3차원 데이터를 입력하는 데이터 입력부;

상기 데이터를 기초로 하여 가우시안 노이즈를 주입하여 초기 상태의 리간드 구조를 설정하며,

상기 초기 상태의 리간드 구조에서 노이즈를 제거하는 역방향 과정에서 SE(3)-등변성을 적용하여, 상기 리간드 구조의 결합 자유 에너지를 최소화하도록 최적화하는 리간드의 3차원 구조를 생성하는 물리 지식 기반 인공지능 모델; 및

상기 생성된 리간드 구조의 결합 친화도 및 구조적 유사성 평가에 따라 리간드의 3차원 구조를 최종 출력하는 분자 구조 출력부;를 포함하며,

상기 물리 지식 기반 인공지능 모델은 확산 모델을 이용하여 구현되는 것인

리간드 분자 구조 생성 시스템.

### 【청구항 8】

제7항에 있어서,

상기 물리 지식 기반 인공지능 모델은, 렌너드-존스(Lennard-Jones) 잠재 에너지를 계산하여 단백질과 리간드의 결합 안정성을 확보하는 것인 리간드 분자 구조 생성 시스템.

**【청구항 9】**

제8항에 있어서,

상기 물리 지식 기반 인공지능 모델은, 비자기회귀적 샘플링 방식(Non-Autoregressive Sampling Method)을 이용하여 리간드의 전체 구조를 한 번에 생성하는 것인 리간드 분자 구조 생성 시스템.

**【청구항 10】**

제9항에 있어서,

상기 비자기회귀적 샘플링 방식(Non-Autoregressive Sampling Method)에 의해 결합 각도와 결합 길이를 포함한 분자의 기하학적 일관성을 유지하면서 분자를 생성하도록 하는 것인 리간드 분자 구조 생성 시스템.

**【청구항 11】**

제7항에 있어서,

상기 물리 지식 기반 인공지능 모델은,

상기 리간드 구조의 결합 친화도를 평가하고,

상기 결합 친화도 평가에서 타겟 단백질과의 결합 친화도를 평가하여 분자의 안정성을 검증하는 것인 리간드 분자 구조 생성 시스템.

**【청구항 12】**

제11항에 있어서,

상기 분자 구조 출력부에서 결합 자유 에너지가 최소화된 리간드 구조를 신

약 후보 물질로서 출력하는 것인 리간드 분자 구조 생성 시스템.

## 【요약서】

### 【요약】

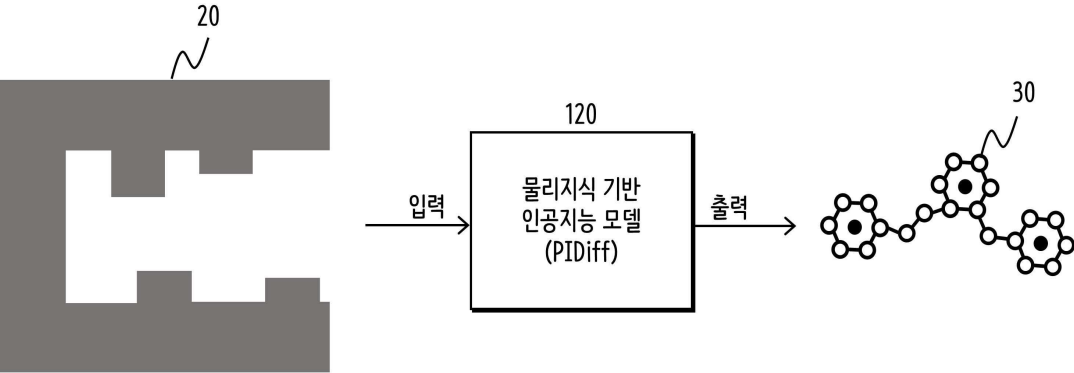
본 발명은 물리화학적 원리를 기반으로 확산 모델 및 SE(3)-등변 신경망을 이용하여 단백질-리간드 간의 결합 자유 에너지를 최소화하는 3차원 분자를 생성하고, 최적화하는 신약 후보 물질 도출을 위한 물리 지식 기반 3차원 분자 생성 방법 및 시스템에 관한 것으로, 그 구성은 단백질 포켓의 3차원 데이터를 입력하는 데이터 입력부, 상기 데이터를 기초로 하여 가우시안 노이즈를 주입하여 초기 상태의 리간드 구조를 설정하며, 상기 초기 상태의 리간드 구조에서 노이즈를 제거하는 역방향 과정에서 SE(3)-등변성을 적용하여, 상기 리간드 구조의 결합 자유 에너지를 최소화하도록 최적화하는 리간드의 3차원 구조를 생성하는 물리 지식 기반 인공지능 모델 및 상기 생성된 리간드 구조의 결합 친화도 및 구조적 유사성 평가에 따라 리간드의 3차원 구조를 최종 출력하는 분자 구조 출력부를 포함하며, 상기 물리 지식 기반 인공지능 모델은 확산 모델을 이용하여 구현되는 것이다.

### 【대표도】

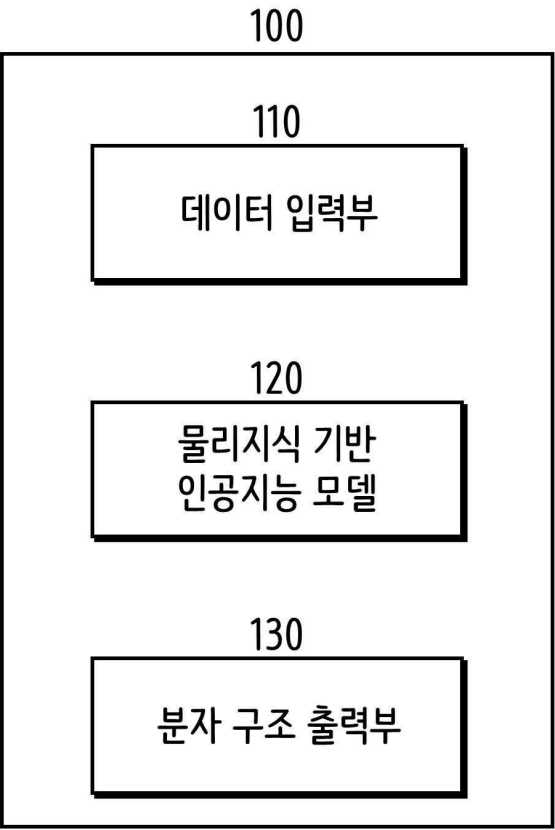
도 1

【도면】

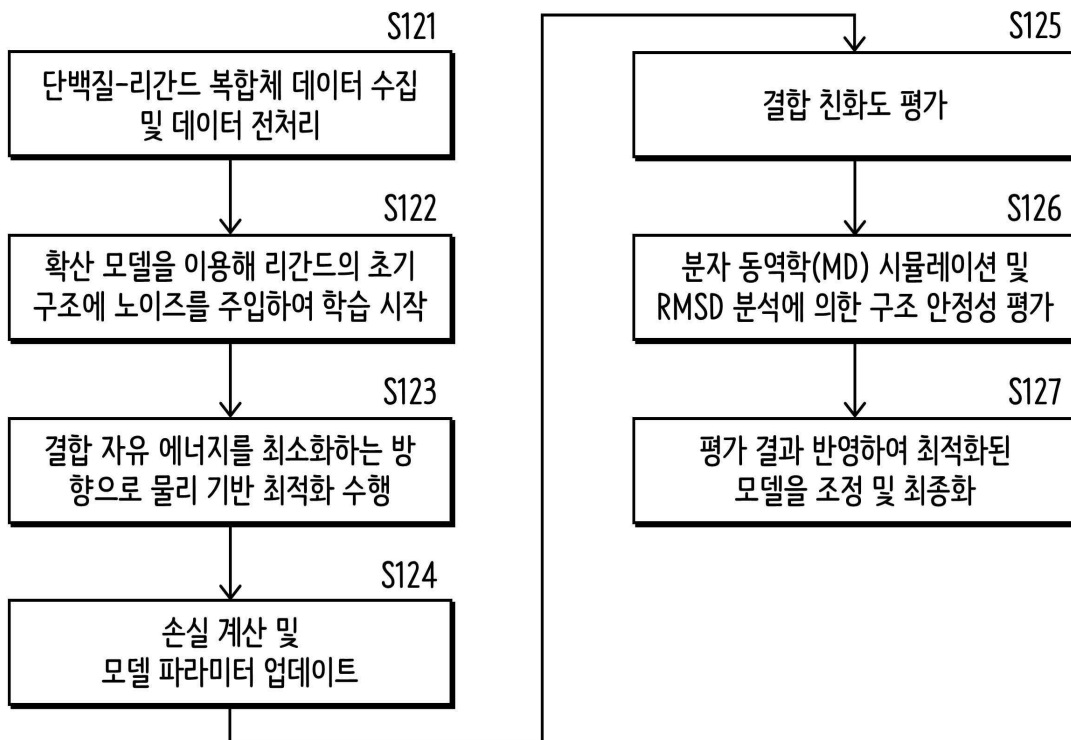
【도 1】



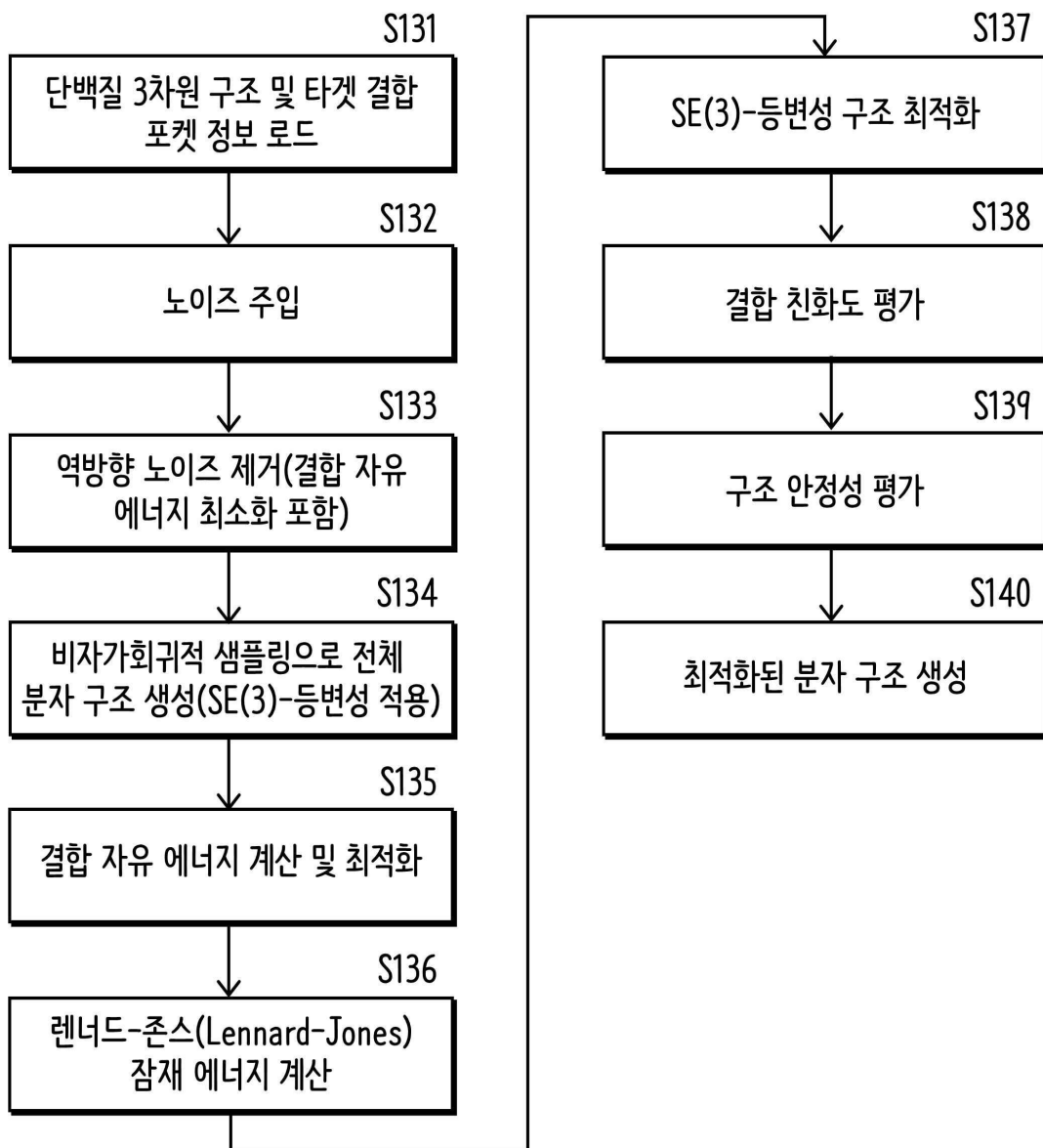
【도 2】



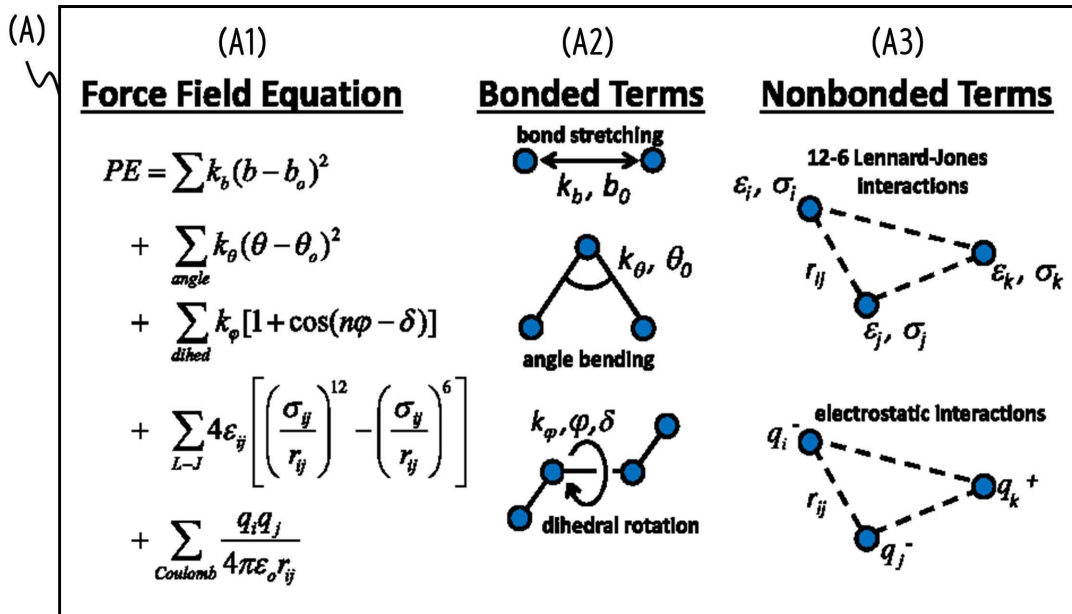
【도 3】



【도 4】



【㉔ 5】



(B)

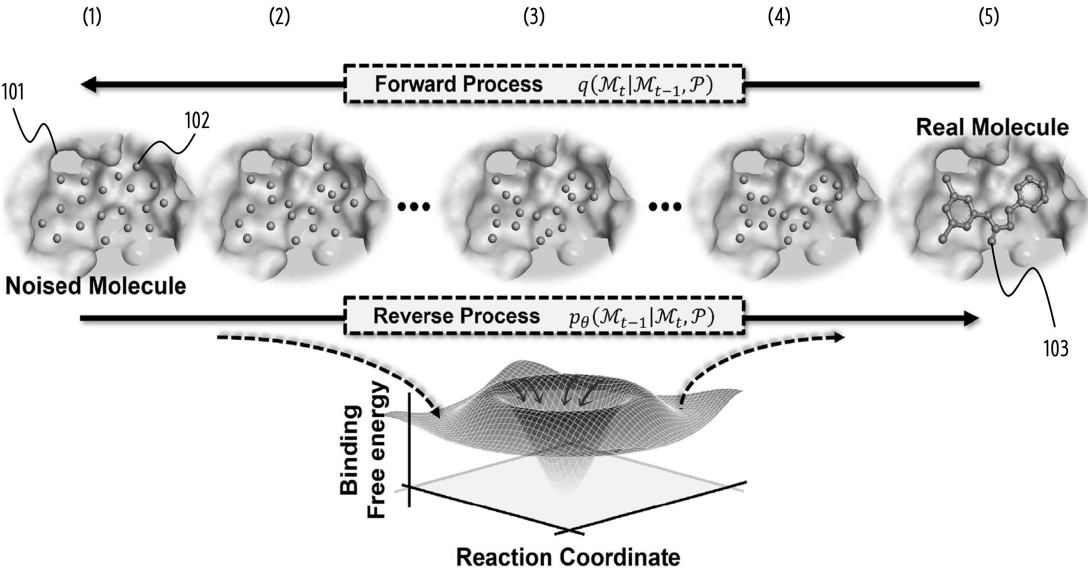
(B1)

$$E_{vdw} = \sum_{j \in P} \sum_{i \in M} \epsilon \left[ \left( \frac{d'_{ij}}{d_{ij}} \right)^{12} - 2 \left( \frac{d'_{ij}}{d_{ij}} \right)^6 \right]$$

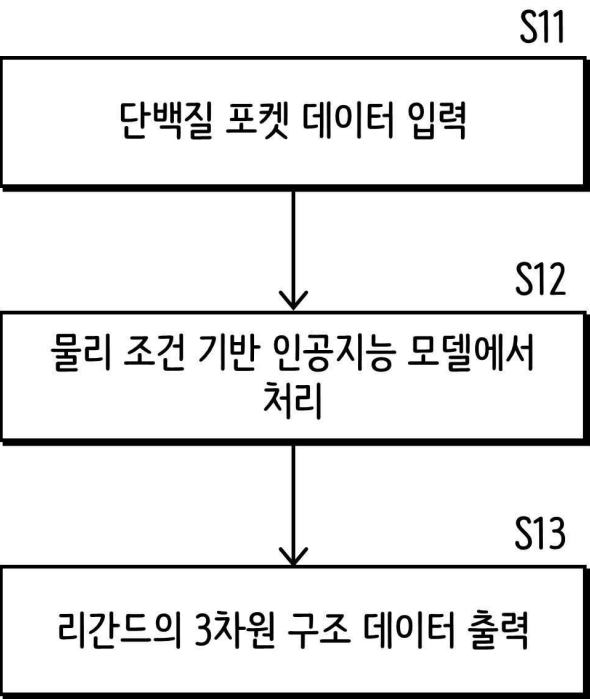
(B2)

$$L_{energy} = \sum_{j \in P} \sum_{i \in M} \frac{\partial \left( \epsilon \left[ \left( \frac{d'_{ij}}{d_{ij}} \right)^{12} - 2 \left( \frac{d'_{ij}}{d_{ij}} \right)^6 \right] \right)}{\partial d_{ij}}$$

【도 6】



【도 7】



| Method        | Metric | VinaScore (↓) |       | VinaMin (↓) |       | VinaDock (↓) |       | HighAffinity (↑) |       | VinaScore <sup>SA</sup> (↓) |       | SR(↑) | Valid <sup>PR</sup> (↑) |
|---------------|--------|---------------|-------|-------------|-------|--------------|-------|------------------|-------|-----------------------------|-------|-------|-------------------------|
|               |        | Avg.          | Med.  | Avg.        | Med.  | Avg.         | Med.  | Avg.             | Med.  | Avg.                        | Med.  |       |                         |
| TestSet       |        | -6.36         | -6.46 | -6.71       | -6.49 | -7.45        | -7.26 | -                | -     | -6.28                       | -6.34 | -     | 93.0%                   |
| AR            |        | -5.75         | -5.64 | -6.18       | -5.88 | -6.75        | -6.62 | 0.379            | 0.310 | -5.59                       | -5.48 | 74.7% | 68.2%                   |
| liGAN         |        | N/A           | N/A   | N/A         | N/A   | -6.33        | -6.2  | 0.211            | 0.111 | N/A                         | N/A   | 68.4% | N/A                     |
| GraphBP       |        | N/A           | N/A   | N/A         | N/A   | -4.80        | -4.70 | 0.142            | 0.067 | N/A                         | N/A   | 57.1% | N/A                     |
| Pocket2Mol    |        | -5.14         | -4.70 | -6.42       | 5.82  | -7.15        | -6.79 | 0.483            | 0.510 | -5.12                       | -5.48 | 88.7% | 69.5%                   |
| DiffSBDD      |        | 52.78         | 44.57 | 16.45       | 0.49  | -6.65        | -7.15 | 0.452            | 0.454 | 51.53                       | 43.46 | 83.0% | 1.5%                    |
| TargetDiff    |        | -5.47         | -6.30 | -6.64       | -6.83 | -7.80        | -7.91 | 0.579            | 0.625 | -5.31                       | -6.13 | 91.9% | 52.5%                   |
| ResGen        |        | 13.79         | 5.75  | -1.53       | -3.36 | -4.90        | -5.26 | 0.232            | 0.000 | 13.73                       | 5.75  | 40.7% | 14.6%                   |
| PIDiff [Ours] |        | -6.58         | -7.37 | -7.52       | -7.57 | -8.10        | -8.35 | 0.641            | 0.666 | -6.03                       | -6.89 | 100%  | 58.1%                   |

【도 9】

