

(12) 특허협력조약에 의하여 공개된 국제출원

(19) 세계지식재산권기구
국제사무국

(43) 국제공개일

2023년 12월 28일 (28.12.2023) WIPO | PCT



(10) 국제공개번호

WO 2023/249441 A1

(51) 국제특허분류:

G16C 60/00 (2019.01) G06N 3/08 (2006.01)
G16C 20/70 (2019.01) G06N 3/047 (2023.01)
G16C 20/50 (2019.01) G06N 7/01 (2023.01)

(21) 국제출원번호: PCT/KR2023/008705

(22) 국제출원일: 2023년 6월 22일 (22.06.2023)

(25) 출원언어: 한국어

(26) 공개언어: 한국어

(30) 우선권정보:
10-2022-0076504 2022년 6월 23일 (23.06.2022) KR

(71) 출원인: 고려대학교 산학협력단 (KOREA UNIVERSITY RESEARCH AND BUSINESS FOUNDATION)
[KR/KR]; 02841 서울특별시 성북구 안암로 145, Seoul (KR).

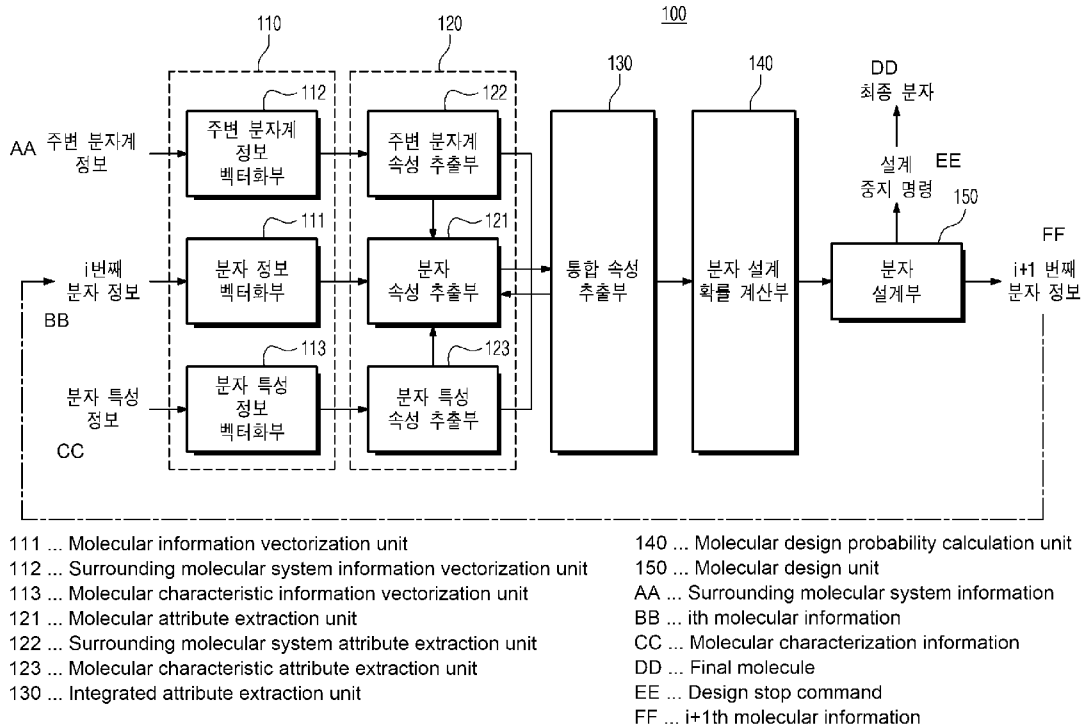
(72) 발명자: 박성남 (PARK, Sungnam); 06320 서울특별시 강남구 개포로 303, 102동 802호, Seoul (KR). 정준영 (JOUNG, Joonyoung); 02771 서울특별시 성북구 돌곶이로25길 62-4, Seoul (KR). 한민희 (HAN, Minhi); 02577 서울특별시 동대문구 왕산로 81, 512호, Seoul (KR). 정민석 (JEONG, Minseok); 12920 경기도 하남시 미사강변중앙로 120, 2504동 503호, Gyeonggi-do (KR). 최동훈 (CHOI, Dong Hoon); 06105 서울특별시 강남구 학동로38길 22, 101동 205호, Seoul (KR).

(74) 대리인: 강일신 등 (KANG, Il Shin et al.); 06753 서울특별시 서초구 바우피로 150, 2층, Seoul (KR).

(81) 지정국 (별도의 표시가 없는 한, 가능한 모든 종류의 국내 권리의 보호를 위하여): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CV, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR,

(54) Title: DEEP LEARNING-BASED MOLECULAR DESIGN SYSTEM, AND DEEP LEARNING-BASED MOLECULAR DESIGN METHOD

(54) 발명의 명칭: 딥러닝 기반의 분자 설계 시스템 및 딥러닝 기반의 분자 설계 방법



(57) Abstract: A deep learning-based molecular design system according to the present invention comprises: a vectorization unit which receives and vectorizes molecular information, surrounding molecular system information, and molecular characteristic information about an ith molecule; an attribute extraction unit which extracts molecular attributes from the vectorized molecular information, extracts surrounding molecular system attributes from the vectorized surrounding molecular system information, and extracts molecular characteristic attributes from the vectorized molecular characteristic information; an integrated attribute extraction unit which extracts integrated attributes of the ith molecule using an integrated attribute extraction algorithm, which is a neural network algorithm that receives the molecular attributes, surrounding molecular system attributes, and molecular characteristic attributes as inputs; a molecular

[다음 쪽 계속]

HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, MG, MK, MN, MU, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.

- (84) 지정국 (별도의 표시가 없는 한, 가능한 모든 종류의 역내 권리의 보호를 위하여): ARIPO (BW, CV, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SC, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 유라시아 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 유럽 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, ME, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

공개:

— 국제조사보고서와 함께 (조약 제21조(3))

design probability calculation unit which extracts a molecular design probability vector for molecular design on the basis of the i th molecule using a molecular design probability calculation algorithm, which is a neural network algorithm that receives the integrated attributes as inputs; and a molecular design unit which extracts molecular information about the $i+1$ th molecule on the basis of the molecular design probability vector or outputs a design stop command to output a final molecule, wherein i is an integer greater than or equal to 1.

(57) 요약서: 본 발명인 딥러닝 기반의 분자 설계 시스템은, i 번째 분자의 분자정보, 주변분자계정보, 및 분자특성정보를 수신하고 벡터화하는 벡터화부, 벡터화된 분자정보에서 분자속성을 추출하고 벡터화된 주변분자계정보에서 주변분자계속성을 추출하고 벡터화된 분자특성정보에서 분자특성속성을 추출하는 속성추출부, 분자속성, 주변분자계속성, 및 분자특성속성을 입력으로 수신하는 신경망 알고리즘인 통합속성 추출알고리즘을 이용하여 i 번째 분자의 통합속성을 추출하는 통합속성추출부, 통합속성을 입력으로 수신하는 신경망 알고리즘인 분자설계확률 계산알고리즘을 이용하여 i 번째 분자를 기초로 분자설계를 위한 분자설계확률벡터를 추출하는 분자설계확률계산부 및 분자설계확률벡터에 기초하여 $i+1$ 번째 분자의 분자정보를 추출하거나 설계중지명령을 출력하여 최종분자를 출력하는 분자설계부를 포함하고, i 는 1보다 크거나 같은 정수이다

명세서

발명의 명칭: 딥러닝 기반의 분자 설계 시스템 및 딥러닝 기반의 분자 설계 방법

기술분야

- [1] 본 발명은 딥러닝 기반의 분자 설계 시스템 및 딥러닝 기반의 분자 설계 방법에 관한 것이다. 구체적으로, 특정한 분자특성을 가질 뿐만 아니라 주변분자계의 영향을 고려하여 분자를 설계하기 위한 딥러닝 기반의 분자 설계 시스템 및 딥러닝 기반의 분자 설계 방법에 관한 것이다.

배경기술

- [2] 목적에 맞는 소재를 개발하기 위해 많은 소재분자들이 개발되고 있다. 일반적으로, 연구자의 경험과 이론을 바탕으로 특정한 분자특성을 지닐 것으로 예측되는 소재분자들을 개발하고자 하지만, 연구자의 경험과 이론의 한계로 인해 원하는 분자특성을 가지는 소재분자들을 개발하기에는 어려움이 있다.
- [3] 이에, 다양한 시행착오를 통해 원하는 분자특성을 가지는 소재분자를 개발하나 많은 시간과 비용이 소요되는 등 다양한 문제점이 발생되고 있다.
- [4] 한편, 최근에는 머신러닝 또는 딥러닝 기술을 이용하여 원하는 분자특성을 갖는 소재분자를 설계하고자 하는 다양한 시도가 있으나, 소재분자의 주변환경을 고려하지 못하여, 분자설계의 정확도가 떨어진다.
- [5] 이에, 시간과 비용을 단축시킬 수 있을 뿐만 아니라 주변환경을 고려하여 원하는 특성을 갖는 분자를 정확히 설계하는 기술이 필요한 실정이다.
- [6] 본 발명은 교육부의 이공분야 대학중점연구소지원사업(과제고유번호: 1345347024, 과제번호: 2019R1A6A1A11044070, 연구과제명: π -전자 기반 에너지 환경 혁신소재 연구, 과제관리기관: 한국연구재단, 과제수행기관: 고려대학교 산학협력단, 연구기간: 2022.03.01. ~ 2023.02.28. 기여율: 50%)과 개인기초연구(과기정통부)(과제고유번호: 1711153079, 과제번호: 2022R1A2C1003627, 연구과제명: 딥러닝 기반 분자 특성 예측 및 새로운 분자구조 생성, 과제관리기관: 한국연구재단, 과제수행기관: 고려대학교 산학협력단, 연구기간: 2022.03.01 ~ 2023.02.28, 기여율: 50%)의 일환으로 수행한 연구로부터 도출된 것이다. 한편, 본 발명의 모든 측면에서 한국 정부의 재산 이익은 없다..

발명의 상세한 설명

기술적 과제

- [7] 본 발명이 해결하고자 하는 기술적 과제는 주변환경(또는, 주변분자계)을 고려하여 원하는 분자특성을 가지는 분자를 설계하기 위한 딥러닝 기반의 분자 설계 시스템 및 딥러닝 기반의 분자 설계 방법에 관한 것이다.

과제 해결 수단

- [8] 본 발명의 한 실시예에 따른 딥러닝 기반의 분자 설계 시스템은 i 번째 분자의 분자정보, 주변분자계정보, 및 분자특성정보를 수신하고 벡터화하는 벡터화부, 상기 벡터화된 분자정보에서 분자속성을 추출하고 벡터화된 주변분자계정보에서 주변분자계속성을 추출하고 벡터화된 분자특성정보에서 분자특성속성을 추출하는 속성추출부, 분자속성, 주변분자계속성, 및 분자특성속성을 입력으로 수신하는 신경망 알고리즘인 통합속성 추출알고리즘을 이용하여 i 번째 분자의 통합속성을 추출하는 통합속성추출부, 통합속성을 입력으로 수신하는 신경망 알고리즘인 분자설계확률 계산알고리즘을 이용하여 i 번째 분자를 기초로 분자설계를 위한 분자설계확률벡터를 추출하는 분자설계확률계산부 및 분자설계확률벡터에 기초하여 $i+1$ 번째 분자의 분자정보를 추출하거나 설계중지명령을 출력하여 최종분자를 출력하는 분자설계부를 포함하고, i 는 1보다 크거나 같은 정수이다.
- [9] 또한, 본 발명의 한 실시예에 따른 벡터화부는, i 번째 분자의 분자정보를 SMILES(Simplified Molecular-Input Line-Entry System)표현으로 수신하고, 분자 핑거프린트(Molecular Fingerprint), 분자설명자(Molecular Descriptor), 화학구조식에 대한 이미지, 분자그래프(Molecular Graph), 분자좌표(Molecular Coordinates), 및 SMILES코드 중 적어도 하나의 표현방법을 이용하여 벡터화하는 분자정보벡터화부, i 번째 분자의 주변분자계정보를 SMILES(Simplified molecular-Input Line-Entry System)표현으로 수신하고, 분자핑거프린트(Molecular Fingerprint), 분자 설명자(Molecular Descriptor), 화학구조식에 대한 이미지, 분자그래프(Molecular Graph), 분자좌표(Molecular Coordinates), 및 SMILES코드 중 적어도 하나의 표현 방법을 이용하여 벡터화하는 주변분자계정보벡터화부 및 i 번째 분자의 분자특성정보를 문자열 또는 실수값 집합의 형태로 입력받고, 토큰화(tokenization), 정규화(normalization), 및 원-핫 인코딩(one-hot encoding) 중 적어도 하나의 표현방법을 이용하여 벡터화하는 분자특성정보벡터화부를 포함한다.
- [10] 또한, 본 발명의 한 실시예에 따른 속성추출부는, 벡터화된 i 번째 분자의 분자정보를 입력으로 수신하는 신경망 알고리즘인 분자속성 추출알고리즘을 이용하여 i 번째 분자의 분자속성을 추출하는 분자속성추출부, 벡터화된 i 번째 분자의 주변분자계정보를 입력으로 수신하는 신경망 알고리즘인 주변분자계속성 추출알고리즘을 이용하여 i 번째 분자의 주변분자계속성을 추출하는 주변분자계속성추출부 및 벡터화된 i 번째 분자의 분자특성정보를 입력으로 수신하는 신경망 알고리즘인 분자특성속성 추출알고리즘을 이용하여 i 번째 분자의 분자특성속성을 추출하는 분자특성속성추출부를 포함한다.
- [11] 또한, 본 발명의 한 실시예에 따른 분자정보는 화학구조식에 대한 정보를 포함하고, 주변분자계정보는 하나 이상의 용매에 대한 정보를 포함하고, 분자특성정보는 분자의 구조적, 화학적, 물리적, 분광학적, 전기화학적, 반응성 중 적어도 하나 이상에 대한 정보를 포함한다.

- [12] 또한, 본 발명의 한 실시예에 따른 첫번째 분자의 분자정보는 화학구조식이 없거나 사용자에게 의해 제공되는 어느 하나의 화학구조식에 대한 정보를 포함한다.
- [13] 또한, 본 발명의 한 실시예에 따른 분자설계부는, 분자설계확률벡터를 구성하는 어느 하나의 원소를 이용하여 산출된 확률값에 따라 $i+1$ 번째 분자를 설계하기 위한 $i+1$ 번째 분자의 분자정보를 추출하고, $i+1$ 번째 분자의 분자정보는 i 번째 분자를 구성하는 어느 하나의 원자에 한개의 원자를 결합하거나, i 번째 분자를 구성하는 원자 사이를 연결하는 결합을 추가하여 설계된 $i+1$ 번째 분자의 화학구조식에 대한 정보를 포함한다.
- [14] 또한, 본 발명의 한 실시예에 따른 분자설계부는, 분자설계확률벡터를 구성하는 어느 하나의 원소를 이용하여 산출된 확률값에 따라 설계중지명령을 출력하여 i 번째 분자를 최종분자로 결정한다.
- [15]
- [16] *
- [17] 또한, 본 발명의 한 실시예에 따른 분자속성 추출알고리즘, 주변분자계속성 추출알고리즘, 분자특성속성 추출알고리즘, 통합속성 추출알고리즘, 및 분자설계확률 계산알고리즘은 적어도 하나 이상의 은닉계층(Hidden Layer)을 포함하는 신경망 알고리즘이다.
- [18] 또한, 본 발명의 한 실시예에 따른 딥러닝 기반의 분자 설계 방법은 벡터화부에 의해 i 번째 분자의 분자정보, 주변분자계정보, 및 분자특성정보를 수신하고 벡터화하는 단계, 속성추출부에 의해 벡터화된 분자정보에서 분자속성을 추출하고 벡터화된 주변분자계정보에서 주변분자계속성을 추출하고 벡터화된 분자특성정보에서 분자특성속성을 추출하는 단계, 통합속성추출부에 의해 분자속성, 주변분자계속성, 및 분자특성속성을 입력으로 수신하는 신경망 알고리즘인 통합속성 추출알고리즘을 이용하여 i 번째 분자의 통합속성을 추출하는 단계, 분자설계확률계산부에 의해 통합속성을 입력으로 수신하는 신경망 알고리즘인 분자설계확률 계산알고리즘을 이용하여 i 번째 분자를 기초로 분자설계의 진행을 위한 분자설계확률벡터를 출력하는 단계 및 분자설계부에 의해 분자설계확률벡터에 기초하여 $i+1$ 번째 분자의 분자정보를 추출하거나 설계중지명령을 출력하여 최종분자를 출력하는 단계를 포함하고, i 는 1보다 크거나 같은 정수이다.
- [19] 또한, 본 발명의 한 실시예에 따른 딥러닝 기반의 분자 설계 방법을 실행시키는 프로그램이 기록된 컴퓨터로 판독가능한 기록매체를 포함한다.

발명의 효과

- [20] 본 발명에 따른 딥러닝 기반의 분자 설계 시스템 및 딥러닝 기반의 분자 설계 방법은 주변분자계를 고려하여 원하는 분자특성을 가지는 분자를 설계하여 분자설계의 정확도를 높일 수 있다.
- [21] 또한, 본 발명에 따른 딥러닝 기반의 분자 설계 시스템 및 딥러닝 기반의 분자 설계 방법은 사용자에게 의해 입력된 정보에 기초하여 원하는 분자특성을 가지는

분자를 설계하므로, 분자설계과정에서의 시행착오를 줄일 수 있을 뿐만 아니라 소요되는 시간 및 개발 비용을 감축시킬 수 있는 효과가 있다.

도면의 간단한 설명

- [22] 도 1은 본 발명의 한 실시예에 따른 딥러닝 기반의 분자 설계 시스템의 구성에 관한 도면이다.
- [23] 도 2a는 본 발명의 한 실시예에 따른 속성추출부의 구현예에 관한 도면이다. 도 2b는 본 발명의 한 실시예에 따른 통합속성추출부의 구현예에 관한 도면이다. 도 2c는 본 발명의 한 실시예에 따른 분자설계확률계산부의 구현예에 관한 도면이다. 도 2d는 본 발명의 한 실시예에 따른 분자설계부의 구현예에 관한 도면이다.
- [24] 도 3a은 본 발명의 한 실시예에 따른 딥러닝 기반 분자 설계 시스템에서 최종분자를 설계하는 구현예에 관한 도면이다. 도 3b는 본 발명의 다른 실시예에 따른 딥러닝 기반 분자 설계 시스템에서 최종분자를 설계하는 구현예에 관한 도면이다.
- [25] 도 4는 본 발명의 한 실시예에 따른 벤젠을 첫번째 분자로하여 분자설계확률벡터에 따라 최종분자를 설계하는 과정에 관한 도면이다.
- [26] 도 5a는 본 발명의 한 실시예에 따른 분자정보, 주변분자계정보, 및 분자특성정보에 기초하여 최종분자를 설계한 결과에 관한 도면이다. 도 5b는 본 발명의 다른 실시예에 따른 분자정보, 주변분자계정보, 및 분자특성정보에 기초하여 최종분자를 설계한 결과에 관한 도면이다. 도 5c는 본 발명의 다른 실시예에 따른 분자정보, 주변분자계정보, 및 분자특성정보에 기초하여 최종분자를 설계한 결과에 관한 도면이다. 도 5d는 본 발명의 다른 실시예에 따른 분자정보, 주변분자계정보, 및 분자특성정보에 기초하여 최종분자를 설계한 결과에 관한 도면이다.
- [27] 도 6은 본 발명의 한 실시예에 따른 딥러닝 기반의 분자 설계 방법에 관한 흐름도이다.

발명의 실시를 위한 형태

- [28] 이하, 첨부한 도면을 참고로 하여 본 발명의 여러 실시 예들에 대하여 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자가 용이하게 실시할 수 있도록 상세히 설명한다. 본 발명은 여러 가지 상이한 형태로 구현될 수 있으며 여기에서 설명하는 실시 예들에 한정되지 않는다.
- [29] 본 발명을 명확하게 설명하기 위해서 설명과 관계없는 부분은 생략하였으며, 명세서 전체를 통하여 동일 또는 유사한 구성요소에 대해서는 동일한 참조 부호를 붙이도록 한다. 따라서 앞서 설명한 참조 부호는 다른 도면에서도 사용할 수 있다.
- [30] 또한, 도면에서 나타난 각 구성의 크기 및 두께는 설명의 편의를 위해 임의로 나타내었으므로, 본 발명이 반드시 도시된 바에 한정되지 않는다. 도면에서 여러 층 및 영역을 명확하게 표현하기 위하여 두께를 과장되게 나타낼 수 있다.

- [31] 또한, 설명에서 "동일하다"라고 표현한 것은, "실질적으로 동일하다"는 의미일 수 있다. 즉, 통상의 지식을 가진 자가 동일하다고 납득할 수 있을 정도의 동일함일 수 있다. 그 외의 표현들도 "실질적으로"가 생략된 표현들일 수 있다.
- [32] 또한, 설명에서 어떤 부분이 어떤 구성요소를 '포함'한다고 할 때, 이는 특별히 반대되는 기재가 없는 한 다른 구성요소를 제외하는 것이 아니라 다른 구성요소를 더 포함할 수 있는 것을 의미한다. 본 명세서에서 사용되는 '~부'는 적어도 하나의 기능이나 동작을 처리하는 단위로서, 예를 들어 소프트웨어, FPGA 또는 하드웨어 구성요소를 의미할 수 있다. '~부'에서 제공하는 기능은 복수의 구성요소에 의해 분리되어 수행되거나, 다른 추가적인 구성요소와 통합될 수도 있다. 본 명세서의 '~부'는 반드시 소프트웨어 또는 하드웨어에 한정되지 않으며, 어드레싱할 수 있는 저장 매체에 있도록 구성될 수도 있고, 하나 또는 그 이상의 프로세서들을 재생시키도록 구성될 수도 있다. 이하에서는 도면을 참조하여 본 발명의 실시예에 대해서 구체적으로 설명하기로 한다.
- [33] 도 1은 본 발명의 한 실시예에 따른 딥러닝 기반의 분자 설계 시스템의 구성에 관한 도면이다.
- [34] 본 발명의 한 실시예에 따른 딥러닝 기반의 분자 설계 시스템(100)은 벡터화부(110), 속성추출부(120), 통합속성추출부(130), 분자설계확률계산부(140), 및 분자설계부(150)를 포함할 수 있다.
- [35] 벡터화부(110)는 분자정보벡터화부(111), 주변분자계정보벡터화부(112), 및 분자특성정보벡터화부(113)를 포함할 수 있다. 속성추출부(120)는 분자속성추출부(121), 주변분자계속성추출부(122), 및 분자특성속성추출부(123)를 포함할 수 있다.
- [36] 벡터화부(110)는 i (단, i 는 1보다 크거나 같은 정수)번째 분자의 분자정보, 주변분자계정보, 및 분자특성정보를 수신하고 벡터화할 수 있다.
- [37] 구체적으로, 분자정보벡터화부(111)는 i 번째 분자의 분자정보를 SMILES(Simplified Molecular-Input Line-Entry System)표현으로 수신하고, 분자핑거프린트(Molecular Fingerprint), 분자설명자(Molecular Descriptor), 화학구조식에 대한 이미지, 분자그래프(Molecular Graph), 분자좌표(Molecular Coordinates), 및 SMILES코드 중 적어도 하나의 표현방법을 이용하여 벡터화할 수 있다.
- [38] 이때, SMILES(Simplified Molecular-Input Line-Entry System)란 화학물질의 구성원소, 결합의 종류, 방향족성(Aromaticity), 브랜치의 유무 등 화학적 구조 정보를 ASCII 코드의 문자열로 표현하는 방법을 의미한다.
- [39] 주변분자계정보벡터화부(112)는 상술한 분자정보벡터화부(111)와 동일하게 i 번째 분자의 주변분자계정보를 SMILES(Simplified molecular-Input Line-Entry System)표현으로 수신하고, 분자핑거프린트(Molecular Fingerprint), 분자설명자(Molecular Descriptor), 화학구조식에 대한 이미지, 분자그래프(Molecular Graph), 분자좌표(Molecular Coordinates), 및 SMILES코드 중 적어도 하나의 표현방법을 이용하여 벡터화할 수 있다.

- [40] 분자특성정보벡터화부(113)는 i 번째 분자의 분자특성정보를 문자열 또는 실수 값 집합의 형태로 입력받고, 토큰화(tokenization), 정규화(normalization), 및 원-핫 인코딩(one-hot encoding) 중 적어도 하나의 표현방법을 이용하여 벡터화할 수 있다.
- [41] 이때, 분자정보는 분자의 화학구조식에 관한 정보를 포함할 수 있다. 예를 들어, i 번째 분자의 분자정보는 i 번째 분자의 화학구조식에 관한 정보를 포함할 수 있으며, 첫번째 분자의 분자정보는 화학구조식에 관한 정보가 없거나 사용자에게 의해 제공되는 특정한 어느 하나의 분자의 화학구조식에 관한 정보를 포함할 수 있다.
- [42] 또한, 주변분자계정보는 분자가 설계되는 주변환경(이하, 주변분자계라 명명함)인 하나 이상의 용매(Solvents)에 대한 정보를 포함할 수 있다.
- [43] 구체적으로, 주변분자계가 기체상인 경우 주변분자계가 없거나 기체분자에 대한 정보를 포함할 수 있다. 주변분자계가 액체상인 경우 주변분자계는 단일용매 또는 공용매(Cosolvent)와 같은 복수의 용매에 대한 정보를 포함할 수 있다. 주변분자계가 고체상인 경우 단일용매 또는 공용매(Cosolvent), 매질(Matrix), 호스트(Host)와 같은 복수의 용매에 대한 정보를 포함할 수 있다.
- [44] 또한, 분자특성정보는 분자의 구조적, 화학적, 물리적, 분광학적, 전기화학적, 반응성 중 적어도 하나 이상에 대한 정보를 포함할 수 있다.
- [45] 예를 들어, 분자특성정보는 분자의 구조적, 화학적, 물리적, 분광학적, 전기화학적, 반응성 중 어느 하나에 대한 정보만을 포함할 수 있다. 또는, 분자특성정보는 분자의 구조적, 화학적, 물리적, 분광학적, 전기화학적, 반응성 중 적어도 2개 이상의 정보를 포함할 수 있다.
- [46] 분자속성추출부(121)는 i 번째 벡터화된 분자정보에서 분자속성을 추출할 수 있다.
- [47] 분자속성추출부(121)는 신경망 알고리즘의 형태인 분자속성 추출알고리즘을 미리 저장할 수 있다. 분자속성추출부(121)는 벡터화된 i 번째 분자의 분자정보를 신경망 알고리즘 형태인 분자속성 추출알고리즘에 입력하여 i 번째 분자의 분자속성을 추출할 수 있다.
- [48] 주변분자계속성추출부(122)는 i 번째 벡터화된 주변분자계정보에서 주변분자계속성을 추출할 수 있다.
- [49] 주변분자계속성추출부(122)는 신경망 알고리즘의 형태인 주변분자계속성 추출알고리즘을 미리 저장할 수 있다. 주변분자계속성추출부(122)는 벡터화된 i 번째 분자의 주변분자계정보를 신경망 알고리즘 형태인 주변분자계속성 추출알고리즘에 입력하여 i 번째 분자의 주변분자계속성을 추출할 수 있다.
- [50] 분자특성속성추출부(123)는 i 번째 벡터화된 분자특성정보에서 분자특성속성을 추출할 수 있다.
- [51] 분자특성속성추출부(123)는 신경망 알고리즘의 형태인 분자특성속성 추출알고리즘을 미리 저장할 수 있다. 분자특성속성추출부(123)는 벡터화된 i 번째 분자

의 분자특성정보를 신경망 알고리즘 형태인 분자특성속성 추출알고리즘에 입력하여 i 번째 분자의 분자특성속성을 추출할 수 있다.

- [52] 한편, 분자속성추출부(121)는 사용되는 분자속성 추출알고리즘에 따라서 주변 분자계속성추출부(122)에서 추출된 i 번째 분자의 주변분자계속성, 분자특성속성추출부(123)에서 추출된 i 번째 분자의 분자특성속성, 이하 서술할 통합속성추출부(130)에서 추출된 i 번째 분자의 통합속성을 추가로 입력받아 i 번째 분자의 분자속성을 추출할 수 있다.
- [53] 분자속성추출부(121)에서 i 번째 분자의 분자속성을 추출하는 과정, 주변분자계속성추출부(122)에서 i 번째 분자의 주변분자계속성을 추출하는 과정, 및 분자특성속성추출부(123)에서 분자특성속성을 추출하는 과정은 이하 도 2a에서 구체적으로 서술하기로 한다.
- [54] 통합속성추출부(130)는 i 번째 분자의 분자속성, i 번째 분자의 주변분자계속성, 및 i 번째 분자의 분자특성속성을 이용하여 i 번째 분자의 통합속성을 추출할 수 있다.
- [55] 구체적으로, 통합속성추출부(130)는 신경망 알고리즘의 형태인 통합속성 추출알고리즘을 미리 저장할 수 있다. 통합속성추출부(130)는 속성추출부(120)로부터 제공된 i 번째 분자의 분자속성, i 번째 분자의 주변분자계속성, 및 i 번째 분자의 분자특성속성을 신경망 알고리즘 형태인 통합속성 추출알고리즘에 입력하여 i 번째 분자의 통합속성을 추출할 수 있다.
- [56] 통합속성추출부(130)에서 i 번째 분자의 통합속성을 추출하는 과정은 이하 도 2b에서 구체적으로 서술하기로 한다.
- [57] 분자설계확률계산부(140)는 i 번째 분자의 통합속성을 이용하여 i 번째 분자를 기초로 분자설계를 위한 분자설계확률벡터를 출력할 수 있다.
- [58] 구체적으로, 분자설계확률계산부(140)는 신경망 알고리즘의 형태인 분자설계확률 계산알고리즘을 미리 저장할 수 있다. 분자설계확률계산부(140)는 통합속성추출부(130)로부터 제공된 i 번째 분자의 통합속성을 신경망 알고리즘 형태인 분자설계확률 계산알고리즘에 입력하여 i 번째 분자를 기초로 분자설계를 위한 분자설계확률벡터를 추출할 수 있다.
- [59] 분자설계확률계산부(140)에서 i 번째 분자를 기초로 분자설계를 위한 분자설계확률벡터를 추출하는 과정은 아래 도 2c에서 구체적으로 서술하기로 한다.
- [60] 분자설계부(150)는 분자설계확률계산부(140)에서 추출된 분자설계확률벡터를 구성하는 원소를 이용하여 산출된 확률값에 따라 $i+1$ 번째 분자를 설계하기 위한 $i+1$ 번째 분자의 분자정보를 추출할 수 있다.
- [61] 이때, $i+1$ 번째 분자의 분자정보는 i 번째 분자를 구성하는 어느 하나의 원자에 한 개의 원자를 결합하거나, i 번째 분자를 구성하는 원자 사이를 연결하는 결합을 추가하여 설계된 $i+1$ 번째 분자의 화학구조식에 대한 정보를 포함한다.

- [62] 또는, 분자설계부(150)는 분자설계확률계산부(140)에서 추출된 분자설계확률 벡터를 구성하는 원소를 이용하여 산출된 확률값에 따라 설계중지명령을 출력하여 i 번째 분자를 최종분자로 결정하고 출력할 수 있다.
- [63] 분자설계부(150)에서 분자설계확률벡터를 이용하여 $i+1$ 번째 분자의 분자정보를 추출하거나 설계중지명령을 출력하여 최종분자를 결정하는 과정은 아래 도 2d에서 구체적으로 서술하기로 한다.
- [64] 상술한 분자설계부(150)에서 분자설계확률벡터에 기초하여 $i+1$ 번째 분자의 분자정보가 추출된 경우, 분자정보벡터화부(111)로 입력될 수 있으며, 분자설계부(150)에서 설계중지명령이 출력될 때까지 상술한 과정을 반복하여 분자를 설계하여 최종분자를 결정할 수 있다.
- [65] 도 1에서 상술한 바와 같이, 본 발명의 한 실시예에 따른 딥러닝 기반의 분자 설계 시스템(100)은 주변분자계를 고려하면서 특정한 분자특성을 지닌 최종분자를 설계함으로써 개발시간과 비용을 크게 줄일 수 있다.
- [66]
- [67] 도 2a는 본 발명의 한 실시예에 따른 속성추출부의 구현예에 관한 도면이다. 도 2b는 본 발명의 한 실시예에 따른 통합속성추출부의 구현예에 관한 도면이다. 도 2c는 본 발명의 한 실시예에 따른 분자설계확률계산부의 구현예에 관한 도면이다. 도 2d는 본 발명의 한 실시예에 따른 분자설계부의 구현예에 관한 도면이다.
- [68] 도 2a 내지 도 2d를 참고하면, 본 발명의 한 실시예에 따른 속성추출부(120), 통합속성추출부(130), 분자설계확률계산부(140), 및 분자설계부(150)에서 구현되는 분자속성 추출알고리즘, 주변분자계속성 추출알고리즘, 분자특성속성 추출알고리즘, 통합속성 추출알고리즘, 및 분자설계확률 계산알고리즘은 적어도 하나 이상의 은닉층(Hidden Layer)을 포함하는 신경망 알고리즘일 수 있다.
- [69] 본 발명의 한 실시예에 따른 분자속성추출부(121)에서 분자의 분자정보를 추출하는 과정, 주변분자계속성추출부(122)에서 주변분자계속성을 추출하는 과정, 분자특성속성추출부(123)에서 분자특성속성을 추출하는 과정은 서로 독립적으로 수행될 수 있다.
- [70] 이하, 도 2a에서는 본 발명의 주변분자계속성추출부(122)를 예로 들어 설명한다.
- [71] 도 2a를 참고하면, 주변분자계속성추출부(122)에 미리 저장된 주변분자계속성 추출알고리즘은 하나 이상의 은닉계층(Hidden Layer)을 포함하는 신경망 알고리즘의 형태이며 멀티 레이어 퍼셉트론(Multi-Layer Perceptron, MLP)으로 구현될 수 있다.
- [72] 이때, 주변분자계속성추출부(122)의 주변분자속성 추출알고리즘에 입력되는 i 번째 분자의 주변분자계정보의 벡터화형식에 따라 주변분자계속성 추출알고리즘은 상술한 멀티 레이어 퍼셉트론(Multi-Layer Perceptron, MLP)외에 추가적인 알고리즘이 적용될 수 있다.

- [73] 예를 들어, 주변분자계속성추출부(122)의 주변분자계속성 추출알고리즘에 입력되는 i 번째 분자의 주변분자계정보의 벡터화형식이 이미지형식인 경우 추가적인 알고리즘은 CNN(Convolutional Neural Network)일 수 있다. 또는, 문자열형식인 경우 추가적인 알고리즘은 RNN(Recurrent Neural Network)일 수 있다. 또는, 그래프형식인 경우 추가적인 알고리즘은 GCN(Graph Convolutional Network)일 수 있다.
- [74] 한편, 상술한 추가적인 알고리즘이 적용되기 전에 상술한 멀티 레이어 퍼셉트론(Multi-Layer Perceptron, MLP)이 먼저 적용될 수 있고, 또는, 상술한 추가적인 알고리즘이 적용된 후에 상술한 멀티 레이어 퍼셉트론(Multi-Layer Perceptron, MLP)이 적용될 수 있다.
- [75] 또는, 상술한 추가적인 알고리즘이 서로 조합되어 상술한 멀티 레이어 퍼셉트론(Multi-Layer Perceptron, MLP)의 전, 후에 적용될 수 있다.
- [76] 즉, 주변분자계속성 추출알고리즘은 멀티 레이어 퍼셉트론(Multi-Layer Perceptron, MLP) 또는 멀티 레이어 퍼셉트론(Multi-Layer Perceptron, MLP)과 추가적인 알고리즘의 조합 또는 멀티 레이어 퍼셉트론(Multi-Layer Perceptron, MLP)과 추가적인 알고리즘의 조합의 조합으로 구현될 수 있다.
- [77] 주변분자계속성추출부(122)는 벡터화된 i 번째 분자의 주변분자계정보를 신경망 알고리즘 형태인 상술한 주변분자계속성 추출알고리즘에 입력하여 i 번째 분자의 주변분자계속성을 추출할 수 있다.
- [78] 분자속성추출부(121)에서 i 번째 분자의 분자속성을 추출하는 과정과 분자특성속성추출부(123)에서 i 번째 분자의 분자특성속성을 추출하는 과정은 상술한 주변분자계속성추출부(122)에서 i 번째 분자의 주변분자계속성을 추출하는 과정과 실질적으로 동일하거나 유사하므로, 중복되는 내용은 생략하기로 한다.
- [79] 한편, 분자속성추출부(121)에서 i 번째 분자의 분자속성을 추출하는 과정은 주변분자계속성추출부(122)에서 추출된 i 번째 분자의 주변분자계속성과 분자특성속성추출부(123)에서 추출된 i 번째 분자의 분자특성속성, 이하 도 2b에서 서술할 통합속성추출부(130)에서 추출된 i 번째 분자의 통합속성을 추가로 입력받아 i 번째 분자의 분자속성을 추출할 수 있다. 도 2b를 참고하면, 통합속성추출부(130)에 미리 저장된 통합속성 추출알고리즘은 하나 이상의 은닉계층(Hidden Layer)을 포함하는 신경망 알고리즘의 형태이며, 적어도 하나 이상의 멀티 레이어 퍼셉트론(Multi-Layer Perceptron, MLP)으로 구현될 수 있다.
- [80] 통합속성추출부(130)는 속성추출부(120)로부터 제공된 i 번째 분자의 분자속성, i 번째 분자의 주변분자계속성, 및 i 번째 분자의 분자속성을 신경망 알고리즘 형태인 상술한 통합속성 추출알고리즘에 입력하여 i 번째 분자의 통합속성을 추출할 수 있다.
- [81] 도 2c를 참고하면, 분자설계확률계산부(140)에 미리 저장된 분자설계확률 계산 알고리즘은 하나 이상의 은닉계층(Hidden Layer)을 포함하는 신경망 알고리즘의

형태이며 멀티 레이어 퍼셉트론(Multi-Layer Perceptron, MLP)으로 구현될 수 있다.

[82] 이때, 분자설계확률계산부(140)의 분자설계확률 계산알고리즘은 상술한 멀티 레이어 퍼셉트론(Multi-Layer Perceptron, MLP)외에 추가적인 알고리즘이 적용될 수 있다.

[83] 예를 들어, 분자설계확률계산부(140)의 분자설계확률 계산알고리즘에는 상술한 멀티 레이어 퍼셉트론(Multi-Layer Perceptron, MLP)외에 RNN(Recurrent Neural Network)형식의 추가적인 알고리즘이 적용될 수 있다.

[84] 한편, 상술한 추가적인 알고리즘이 적용되기 전에 상술한 멀티 레이어 퍼셉트론(Multi-Layer Perceptron, MLP)이 먼저 적용될 수 있고, 또는, 상술한 추가적인 알고리즘이 적용된 후에 상술한 멀티 레이어 퍼셉트론(Multi-Layer Perceptron, MLP)이 적용될 수 있다.

[85] 분자설계확률계산부(140)는 통합속성추출부(130)로부터 제공된 i 번째 분자의 통합속성을 신경망 알고리즘 형태인 상술한 분자설계확률 계산알고리즘에 입력하여 i 번째 분자를 기초로 분자설계를 위한 분자설계확률벡터를 추출할 수 있다.

[86] 이때, 적어도 하나 이상의 원소가 상기 분자설계확률벡터를 구성할 수 있다. 분자설계확률벡터를 구성하는 각각의 원소는 i 번째 분자를 구성하는 어느 하나의 원자에 한 개의 원자를 결합하여 $i+1$ 번째 분자를 설계하기 위한 확률값, i 번째 분자를 구성하는 원자 사이를 연결하는 결합을 추가하여 $i+1$ 번째 분자를 설계하기 위한 확률값, 및 설계중지명령을 출력하여 i 번째 분자를 최종분자로 결정하기 위한 확률값을 의미할 수 있다.

[87] 도 2d를 참고하면, 분자설계부(150)는 분자설계확률계산부(140)에서 추출된 분자설계확률벡터를 구성하는 원소를 이용하여 산출된 확률값에 따라 $i+1$ 번째 분자를 설계하기 위한 $i+1$ 번째 분자의 분자정보를 추출할 수 있다.

[88] 구체적으로, 도 2c에서 상술한 바와 같이, 분자설계부(150)는 분자설계확률계산부(140)에서 추출된 분자설계확률벡터를 구성하는 어느 하나의 원소를 선택하고 확률값을 산출할 수 있다.

[89] 분자설계부(150)는 상술한 확률값에 따라 i 번째 분자를 구성하는 어느 하나의 원자에 한 개의 원자를 결합하거나, i 번째 분자를 구성하는 원자 사이를 연결하는 결합을 추가하여 $i+1$ 번째 분자를 설계하기 위한 $i+1$ 번째 분자의 분자정보를 추출할 수 있다.

[90] 분자설계부(150)는 상술한 확률값에 따라 설계중지명령을 출력하여 i 번째 분자를 최종분자로 결정하고 출력할 수 있다.

[91]

[92] 도 3a는 본 발명의 한 실시예에 따른 딥러닝 기반 분자 설계 시스템에서 최종분자를 설계하는 구현예에 관한 도면이다. 도 3b는 본 발명의 다른 실시예에 따른 딥러닝 기반 분자 설계 시스템에서 최종분자를 설계하는 구현예에 관한 도면이다.

- [93] 먼저, 도 3a를 참고하여 본 발명의 한 실시예에 따른 딥러닝 기반 분자 설계 시스템에서 최종분자를 설계하는 구현예를 설명한다.
- [94] 분자정보벡터화부(111)에는 i번째 분자의 분자정보가 수신될 수 있으며 벡터화될 수 있다. 주변분자계정보벡터화부(112)에는 i번째 분자의 주변분자계정보가 수신될 수 있으며 벡터화될 수 있다.
- [95] 이때, 분자정보벡터화부(111)에서 벡터화된 i번째 분자의 분자정보와 주변분자계정보벡터화부(112)에서 벡터화된 i번째 분자의 주변분자계정보는 분자그래프(Molecular Graph)의 표현방법을 이용하여 벡터화될 수 있다.
- [96] 분자특성정보벡터화부(113)에서 i번째 분자의 분자특성정보가 수신될 수 있으며 벡터화될 수 있다.
- [97] 분자정보벡터화부(111)에서 벡터화된 i번째 분자의 분자정보는 분자속성추출부(121)로 입력될 수 있다. 이때, i번째 분자의 분자정보는 32개, 64개, 128개, 128개, 256개, 256개의 노드(또는, 요소)로 구성된 6계층의 GCN(Graph Convolutional Network)을 순차적으로 통과하고 각 GCN(Graph Convolutional Network)의 출력값으로 총 6개의 i번째 분자의 분자속성이 추출될 수 있다.
- [98] 주변분자계정보벡터화부(112)에서 벡터화된 i번째 분자의 주변분자계정보는 주변분자계속성추출부(122)로 입력될 수 있다. 이때, i번째 분자의 주변분자계정보는 128개, 128개, 128개, 128개, 128개, 256개의 노드(또는, 요소)로 구성된 GCN(Graph Convolutional Network) 및, 32개의 노드(또는, 요소)로 구성된 멀티 레이어 퍼셉트론(Multi-Layer Perceptron, MLP)을 순차적으로 통과하여 i번째 분자의 주변분자계속성이 추출될 수 있다.
- [99] 분자특성정보벡터화부(113)에서 벡터화된 i번째 분자의 분자특성정보는 분자특성속성추출부(123)로 입력될 수 있다. 이때, i번째 분자의 분자특성정보는 32개의 노드(또는, 요소)로 구성된 멀티 레이어 퍼셉트론(Multi-Layer Perceptron, MLP)을 통과하여 i번째 분자의 분자특성속성이 추출될 수 있다.
- [100] 분자속성추출부(121)에서 추출된 i번째 분자의 분자속성, 주변분자계속성추출부(122)에서 추출된 i번째 분자의 주변분자계속성, 및 분자특성속성추출부(123)에서 추출된 i번째 분자의 분자특성속성은 통합속성추출부(130)에 입력되어 서로 연결(Concatenate)될 수 있다.
- [101] 통합속성추출부(130)에 입력된 i번째 분자의 분자속성, 주변분자계속성, 및 분자특성속성은 256개의 노드(또는, 요소)로 구성된 멀티 레이어 퍼셉트론(Multi-Layer Perceptron, MLP)을 통과하여 i번째 분자의 통합속성이 추출될 수 있다.
- [102] 통합속성추출부(130)에서 추출된 i번째 분자의 통합속성은 분자설계확률계산부(140)으로 입력될 수 있다.
- [103] 분자설계확률계산부(140)로 입력된 i번째 분자의 통합속성은 512개의 노드(또는, 요소)로 구성된 멀티 레이어 퍼셉트론(Multi-Layer Perceptron, MLP) 및 512개의 노드(또는, 요소)로 구성된 RNN(Recurrent Neural Network)을 통과하여 i번째 분자를 기초로 분자설계를 위한 분자설계확률벡터가 추출될 수 있다.

- [104] 분자설계확률계산부(140)에서 추출된 분자설계확률벡터는 분자설계부(150)으로 입력될 수 있다.
- [105] 분자설계부(150)는 입력된 분자설계확률벡터를 구성하는 각 요소를 가중치로 하여 확률값을 계산하고, 상기 확률값에 기초하여 분자설계확률벡터를 구성하는 어느 하나의 요소를 선택할 수 있다. 분자설계부(150)는 선택한 요소에 따라 $i+1$ 번째 분자를 설계하기 위한 $i+1$ 번째 분자의 분자정보를 추출하거나 설계중지명령을 출력할 수 있다.
- [106] 분자설계부(150)에서 $i+1$ 번째 분자를 설계하기 위한 $i+1$ 번째 분자의 분자정보가 추출된 경우, 추출된 $i+1$ 번째 분자의 분자정보는 분자정보벡터화부(111)에 재입력되어 상술한 과정이 반복되며, 분자설계부(150)에서 설계중지명령이 출력될 때까지 분자설계가 진행된다.
- [107] 한편, 분자설계부(150)에서 설계중지명령이 출력된 경우, i 번째 분자를 최종분자로 결정하고 출력할 수 있다.
- [108] 이하, 도 3b를 참고하여 본 발명의 다른 실시예에 따른 딥러닝 기반 분자 설계 시스템에서 최종분자를 설계하는 구현예를 설명한다.
- [109] 도 3b는 상술한 도 3a와 비교하여 분자특성속성추출부(123)가 제외된다.
- [110] 분자정보벡터화부(111)에는 i 번째 분자의 분자정보가 수신될 수 있으며 벡터화될 수 있다. 주변분자계정보벡터화부(112)에는 i 번째 분자의 주변분자계정보가 수신될 수 있으며 벡터화될 수 있다.
- [111] 이때, 분자정보벡터화부(111)에서 벡터화된 i 번째 분자의 분자정보와 주변분자계정보벡터화부(112)에서 벡터화된 i 번째 분자의 주변분자계정보는 분자그래프(Molecular Graph)의 표현방법을 이용하여 벡터화될 수 있다.
- [112] 분자특성정보벡터화부(113)에서 i 번째 분자의 분자특성정보가 수신될 수 있으며 벡터화될 수 있다.
- [113] 분자정보벡터화부(111)에서 벡터화된 i 번째 분자의 분자정보는 분자속성추출부(121)로 입력될 수 있다. 이때, i 번째 분자의 분자정보는 32개, 64개, 128개, 128개, 256개, 256개의 노드(또는, 요소)로 구성된 6계층의 GCN(Graph Convolutional Network)을 각각 통과하고 총 6개의 i 번째 분자의 분자속성이 추출될 수 있다.
- [114] 주변분자계정보벡터화부(112)에서 i 번째 분자의 주변분자계정보는 주변분자계속성추출부(122)로 입력될 수 있다. 이때, i 번째 분자의 주변분자계정보는 128개, 128개, 128개, 128개, 128개, 256개의 노드(또는, 요소)로 구성된 GCN(Graph Convolutional Network) 및, 5개의 노드(또는, 요소)로 구성된 1개의 멀티 레이어 퍼셉트론(Multi-Layer Perceptron, MLP)을 순차적으로 통과하여 i 번째 분자의 주변분자계속성이 추출될 수 있다.
- [115] 분자특성정보벡터화부(113)에서 벡터화된 i 번째 분자의 분자특성정보 및 주변분자계속성추출부(122)에서 추출된 i 번째 분자의 주변분자계속성은 통합속성추출부(130)에 입력되어 서로 연결(Concatenate)될 수 있다.

- [116] 통합속성추출부(130)에 입력되어 연결된 i 번째 분자의 분자특성정보 및 주변분자계속성은 32개, 64개, 128개, 128개, 256개, 256개의 노드(또는, 요소)로 구성된 6개의 멀티 레이어 퍼셉트론(Multi-Layer Perceptron, MLP)을 각각 통과할 수 있다.
- [117] 6개의 멀티 레이어 퍼셉트론(Multi-Layer Perceptron, MLP)을 각각 통과한 출력값은 GCN(Graph Convolutional Network)을 각각 통과하여 추출된 총 6개의 i 번째 분자의 분자속성과 더해진(Sum) 후, 다음 계층의 GCN(Graph Convolutional Network)에 입력되거나 전부 연결(Concatenate)된 후 256개의 노드(또는, 요소)로 구성된 1개의 멀티 레이어 퍼셉트론(Multi-Layer Perceptron, MLP)을 통과하여 i 번째 분자의 통합속성이 추출될 수 있다.
- [118] 통합속성추출부(130)에서 추출된 i 번째 분자의 통합속성은 분자설계확률계산부(140)으로 입력될 수 있다.
- [119]
- [120] *
- [121] 분자설계확률계산부(140)로 입력된 i 번째 분자의 통합속성은 512개의 노드(또는, 요소)로 구성된 멀티 레이어 퍼셉트론(Multi-Layer Perceptron, MLP) 및 512개의 노드(또는, 요소)로 구성된 RNN(Recurrent Neural Network)을 통과하여 i 번째 분자를 기초로 분자설계를 위한 분자설계확률벡터가 추출될 수 있다.
- [122] 분자설계확률계산부(140)에서 추출된 분자설계확률벡터는 분자설계부(150)으로 입력될 수 있다.
- [123] 분자설계부(150)는 입력된 분자설계확률벡터를 구성하는 각 요소를 가중치로 하여 확률값을 계산하고, 상기 확률값에 기초하여 분자설계확률벡터를 구성하는 어느 하나의 요소를 선택할 수 있다. 분자설계부(150)는 선택한 요소에 따라 $i+1$ 번째 분자를 설계하기 위한 $i+1$ 번째 분자의 분자정보를 추출하거나 설계중지명령을 출력할 수 있다.
- [124] 분자설계부(150)에서 $i+1$ 번째 분자를 설계하기 위한 $i+1$ 번째 분자의 분자정보가 추출된 경우, 추출된 $i+1$ 번째 분자의 분자정보는 분자정보벡터화부(111)에 재입력되어 상술한 과정이 반복되며, 분자설계부(150)에서 설계중지명령이 출력될 때까지 분자설계가 진행된다.
- [125] 한편, 분자설계부(150)에서 설계중지명령이 출력된 경우, i 번째 분자를 최종분자로 결정하고 출력할 수 있다.
- [126]
- [127] 도 4는 본 발명의 한 실시예에 따른 벤젠을 첫번째 분자로하여 분자설계확률벡터에 따라 최종분자를 설계하는 과정에 관한 도면이다.
- [128] 도 4를 참고하면, 첫번째 분자의 분자정보 즉, 첫번째 분자가 벤젠으로 입력된 경우, 분자설계확률계산부(140)에서 분자설계확률벡터를 추출하고 분자설계부(150)에서 분자설계확률벡터를 구성하는 원소를 이용하여 최종분자를 설계할 수 있다.

- [129] 예를 들어, 분자설계확률계산부(140)는 첫번째 분자를 기초로 분자설계를 위한 분자설계확률벡터를 추출할 수 있다.
- [130] 분자설계부(150)는 분자설계확률벡터를 구성하는 원소를 이용하여 산출된 확률값에 따라 2번째 분자를 설계하기 위한 2번째 분자의 분자정보를 추출할 수 있다.
- [131] 도 4를 참고하면 분자설계부(150)에 의해 분자설계확률벡터를 구성하는 원소에 대한 확률값을 산출하여, 어느 하나의 확률값에 기초하여 다음분자가 설계되는 경우를 실선화살표로 표시하였고, 다음분자가 설계되지 않는 경우는 점선화살표로 표시하였다.
- [132] 분자설계부(150)는 분자설계확률벡터를 구성하는 원소를 이용하여 확률값을 산출하고, 상기 확률값 중 가장 큰 확률값에 대응하는 분자정보에 따라 다음분자를 설계할 수 있다.
- [133] 또는, 분자설계부(150)는 분자설계확률벡터를 구성하는 원소를 이용하여 확률값을 산출하고, 상기 확률값을 가중치로 하여 다음분자를 설계할 수 있다.
- [134] 최종적으로 51.3%에 대응하는 설계중지명령이 출력된 경우 분자설계부(150)는 분자설계를 중지하고 최종분자를 출력할 수 있다.
- [135]
- [136] 도 5a는 본 발명의 한 실시예에 따른 분자정보, 주변분자계정보, 및 분자특성정보에 기초하여 최종분자를 설계한 결과에 관한 도면이다. 도 5b는 본 발명의 다른 실시예에 따른 분자정보, 주변분자계정보, 및 분자특성정보에 기초하여 최종분자를 설계한 결과에 관한 도면이다. 도 5c는 본 발명의 또 다른 실시예에 따른 분자정보, 주변분자계정보, 및 분자특성정보에 기초하여 최종분자를 설계한 결과에 관한 도면이다.
- [137] 도 5a를 참고하면, 첫번째 분자의 분자정보는 화학구조식이 없고, 주변분자계정보는 톨루엔(toluene)에 대한 정보를 포함하고, 분자특성정보는 최대흡광파장에 대한 정보를 포함하도록 설정하여 최종분자를 설계한 결과에 대한 도면이다. 또한, 도 5a에서는 상술한 분자설계를 10000번 이상 반복하여 최종분자를 설계한 결과에 해당한다.
- [138]
- [139] *
- [140] 딥러닝 기반의 분자 설계 시스템(100)에서 분자특성정보에 포함된 최대흡광파장을 400nm로 설정하여 분자설계를 수행한 경우, 최대흡광파장이 400nm를 가지는 최종분자의 비율은 최대흡광파장이 400nm를 가지는 비교군(데이터베이스)에 비해 400nm를 중심으로 밀집됨을 알 수 있다.
- [141] 또한, 딥러닝 기반의 분자 설계 시스템(100)에서 분자특성정보에 포함된 최대흡광파장을 500nm로 설정하여 분자설계를 수행한 경우, 최대흡광파장이 500nm를 가지는 최종분자의 비율은 최대흡광파장이 500nm를 가지는 비교군(데이터베이스)에 비해 500nm를 중심으로 밀집됨을 알 수 있다.

- [142] 또한, 딥러닝 기반의 분자 설계 시스템(100)에서 분자특성정보에 포함된 최대 흡광파장을 600nm로 설정하여 분자설계를 수행한 경우, 최대흡광파장이 600nm를 가지는 최종분자의 비율은 최대흡광파장이 600nm를 가지는 비교군(데이터베이스)에 비해 600nm를 중심으로 밀집됨을 알 수 있다.
- [143] 또한, 딥러닝 기반의 분자 설계 시스템(100)에서 분자특성정보에 포함된 최대 흡광파장을 700nm로 설정하여 분자설계를 수행한 경우, 최대흡광파장이 700nm를 가지는 최종분자의 비율은 최대흡광파장이 700nm를 가지는 비교군(데이터베이스)에 비해 700nm를 중심으로 밀집됨을 알 수 있다.
- [144] 또한, 딥러닝 기반의 분자 설계 시스템(100)에서 분자특성정보에 포함된 최대 흡광파장을 800nm로 설정하여 분자설계를 수행한 경우, 최대흡광파장이 800nm를 가지는 최종분자의 비율은 최대흡광파장이 800nm를 가지는 비교군(데이터베이스)에 비해 800nm를 중심으로 밀집됨을 알 수 있다.
- [145] 즉, 본 발명의 한 실시예에 따른 딥러닝 기반의 분자 설계 시스템(100)은 주변분자계를 고려하여 원하는 분자특성을 가지는 분자를 정확도가 높게 설계할 수 있다.
- [146] 도 5b를 참고하면, 첫번째 분자의 분자정보는 벤젠에 대한 화학구조식에 대한 정보를 포함하고, 주변분자계정보는 톨루엔에 대한 정보를 포함하고, 분자특성정보는 최대흡광파장 및 최대발광파장에 대한 정보를 포함하도록 설정하여 최종분자를 설계한 결과에 대한 도면이다. 또한, 도 5b에서는 상술한 분자설계를 10000번 이상 반복하여 최종분자를 설계한 결과에 해당한다.
- [147] 딥러닝 기반의 분자 설계 시스템(100)에서 분자특성정보에 포함된 최대흡광파장을 400nm로 설정하고, 최대발광파장을 450nm로 설정하여 분자설계를 수행한 경우 최종분자의 비율은 최대흡광파장이 400nm이고 최대발광파장이 450nm에 밀집됨을 알 수 있다.
- [148] 또한, 딥러닝 기반의 분자 설계 시스템(100)에서 분자특성정보에 포함된 최대 흡광파장을 400nm로 설정하고, 최대발광파장을 500nm로 설정하여 분자설계를 수행한 경우 최종분자의 비율은 최대흡광파장이 400nm이고 최대발광파장이 500nm에 밀집됨을 알 수 있다.
- [149] 또한, 딥러닝 기반의 분자 설계 시스템(100)에서 분자특성정보에 포함된 최대 흡광파장을 500nm로 설정하고, 최대발광파장을 600nm로 설정하여 분자설계를 수행한 경우 최종분자의 비율은 최대흡광파장이 500nm이고 최대발광파장이 600nm에 밀집됨을 알 수 있다.
- [150] 또한, 딥러닝 기반의 분자 설계 시스템(100)에서 분자특성정보에 포함된 최대 흡광파장을 600nm로 설정하고, 최대발광파장을 650nm로 설정하여 분자설계를 수행한 경우 최종분자의 비율은 최대흡광파장이 600nm이고 최대발광파장이 650nm에 밀집됨을 알 수 있다.

- [151] 즉, 본 발명의 한 실시예에 따른 딥러닝 기반의 분자 설계 시스템(100)은 주변분자계를 고려하여 둘 이상의 원하는 분자특성을 가지는 분자를 정확도가 높게 설계할 수 있다.
- [152] 도 5c를 참고하면, 첫번째 분자의 분자정보는 화학구조식이 포함되지 않고, 주변분자계정보는 톨루엔에 대한 정보를 포함하고, 분자특성정보는 최대흡광파장(370nm), 흡차반치전폭(4600 cm^{-1}), 물흡광계수(4.5), 최대발광파장(450nm), 발광반치전폭(3000 cm^{-1}), 발광양자수율(0.5), 발광수명(1.45ns)에 대한 정보를 모두 포함하도록 설정하여 최종분자를 설계한 도면이다.
- [153] 도 5c와 같이, 딥러닝 기반의 분자 설계 시스템(100)에 상술한 바와 같이 분자의 분자정보, 주변분자계정보, 및 7개의 분자특성정보를 동시에 함께 입력하더라도 입력된 분자특성정보를 중심으로 밀집된 비율을 가지는 최종분자가 설계됨을 알 수 있다.
- [154] 즉, 본 발명의 한 실시예에 따른 딥러닝 기반의 분자 설계 시스템(100)은 주변분자계를 고려하여 다양한 분자특성을 가지는 분자를 정확도가 높게 설계할 수 있다.
- [155] 도 5d를 참고하면, 첫번째 분자의 분자정보는 화학구조식이 포함되지 않고, 분자특성정보는 최대흡광파장(370nm), 흡차반치전폭(4700 cm^{-1}), 물흡광계수(3.6), 최대발광파장(550nm), 발광반치전폭(3800 cm^{-1}), 발광양자수율(0.01), 발광수명(2.0ns)에 대한 정보를 모두 포함하고 주변분자계정보는 물(H₂O)에 대한 정보를 포함하는 경우와 톨루엔에 대한 정보를 포함하는 경우로 설정하여 최종분자를 각각 설계한 도면이다.
- [156] 도 5d를 참고하면, 주변분자계정보가 물에 대한 정보를 포함하는 경우와 주변분자계정보가 톨루엔에 대한 정보를 포함하는 경우 서로 다른 최종분자를 설계함을 알 수 있다.
- [157] 구체적으로, 주변 분자계가 물인 경우 용매의 극성이 크기 때문에 스토크스 이동(Stokes shift)이 비교적 작은 분자로도 이를 수 있으나, 주변 분자계가 톨루엔인 경우 용매의 극성이 작기 때문에 분자 내에서 주개(donor)-받개(acceptor)의 거리를 상대적으로 더 멀게 분자가 설계되었음을 확인할 수 있다.
- [158] 즉, 본 발명의 한 실시예에 따른 딥러닝 기반의 분자 설계 시스템(100)은 주변분자계를 고려하여 원하는 분자특성을 가지는 분자를 정확도가 높게 설계할 수 있음을 알 수 있다.
- [159] 도 6은 본 발명의 한 실시예에 따른 딥러닝 기반의 분자 설계 방법에 관한 흐름도이다.
- [160] 단계(S10)에서 i번째 분자의 분자정보, 주변분자계정보, 및 분자특성정보를 수신하고 벡터화할 수 있다.

- [161] 구체적으로, 벡터화부(110)는 i (단, i 는 1보다 크거나 같은 정수)번째 분자의 분자정보, 주변분자계정보, 및 분자특성정보를 수신하고 벡터화할 수 있다.
- [162] 단계(S11)에서 벡터화된 분자정보에서 분자속성을 추출하고 벡터화된 주변분자계정보에서 주변분자계속성을 추출하고 벡터화된 분자특성정보에서 분자특성속성을 추출할 수 있다.
- [163] 구체적으로, 분자속성추출부(121)는 벡터화된 i 번째 분자의 분자정보를 신경망 알고리즘 형태인 분자속성 추출알고리즘에 입력하여 i 번째 분자의 분자속성을 추출할 수 있다.
- [164] 주변분자계속성추출부(122)는 벡터화된 i 번째 분자의 주변분자계정보를 신경망 알고리즘 형태인 주변분자계속성 추출알고리즘에 입력하여 i 번째 분자의 주변분자계속성을 추출할 수 있다.
- [165] 분자특성속성추출부(123)는 벡터화된 i 번째 분자의 분자특성정보를 신경망 알고리즘 형태인 분자특성속성 추출알고리즘에 입력하여 i 번째 분자의 분자특성속성을 추출할 수 있다.
- [166] 단계(S12)에서 분자속성, 주변분자계속성, 및 분자특성속성을 입력으로 수신하는 신경망 알고리즘인 통합속성 추출알고리즘을 이용하여 상기 i 번째 분자의 통합속성을 추출할 수 있다.
- [167] 구체적으로, 통합속성추출부(130)는 속성추출부(120)로부터 제공된 i 번째 분자의 분자속성, i 번째 분자의 주변분자계속성, 및 i 번째 분자의 분자속성을 신경망 알고리즘 형태인 통합속성 추출알고리즘에 입력하여 i 번째 분자의 통합속성을 추출할 수 있다.
- [168] 단계(S13)에서 통합속성을 입력으로 수신하는 신경망 알고리즘인 분자설계확률 계산알고리즘을 이용하여 i 번째 분자를 기초로 분자설계의 진행을 위한 분자설계확률벡터를 추출할 수 있다.
- [169] 구체적으로, 분자설계확률계산부(140)는 통합속성추출부(130)로부터 제공된 i 번째 분자의 통합속성을 신경망 알고리즘 형태인 분자설계확률 계산알고리즘에 입력하여 i 번째 분자를 기초로 분자설계를 위한 분자설계확률벡터를 추출할 수 있다.
- [170] 단계(S14)에서 분자설계확률벡터에 기초하여 $i+1$ 번째 분자의 분자정보를 추출하거나 설계중지명령을 출력하여 최종분자를 출력할 수 있다.
- [171] 구체적으로, 분자설계부(150)는 분자설계확률계산부(140)에서 추출된 분자설계확률벡터를 구성하는 원소를 이용하여 산출된 확률값에 따라 $i+1$ 번째 분자를 설계하기 위한 $i+1$ 번째 분자의 분자정보를 추출할 수 있다.
- [172] 또는, 분자설계부(150)는 분자설계확률계산부(140)에서 추출된 분자설계확률벡터를 구성하는 원소를 이용하여 산출된 확률값에 따라 설계중지명령을 출력하여 i 번째 분자를 최종분자로 결정하고 출력할 수 있다.
- [173] 지금까지 참조한 도면과 기재된 발명의 상세한 설명은 단지 본 발명의 예시적인 것으로서, 이는 단지 본 발명을 설명하기 위한 목적에서 사용된 것이지 의미

한정이나 특허청구범위에 기재된 본 발명의 범위를 제한하기 위하여 사용된 것은 아니다. 그러므로 본 기술 분야의 통상의 지식을 가진 자라면 이로부터 다양한 변형 및 균등한 타 실시 예가 가능하다는 점을 이해할 것이다. 따라서, 본 발명의 진정한 기술적 보호 범위는 첨부된 특허청구범위의 기술적 사상에 의해 정해져야 할 것이다.

- [174] 이상에서 설명된 실시예들은 하드웨어 구성요소, 소프트웨어 구성요소, 및/또는 하드웨어 구성요소 및 소프트웨어 구성요소의 조합으로 구현될 수 있다. 예를 들어, 실시예들에서 설명된 장치, 방법 및 구성요소는, 예를 들어, 프로세서, 콘트롤러, ALU(Arithmetic Logic Unit), 디지털 신호 프로세서(Digital Signal Processor), 마이크로컴퓨터, FPGA(Field Programmable Gate Array), PLU(Programmable Logic Unit), 마이크로프로세서, 또는 명령(instruction)을 실행하고 응답할 수 있는 다른 어떠한 장치와 같이, 하나 이상의 범용 컴퓨터 또는 특수 목적 컴퓨터를 이용하여 구현될 수 있다.
- [175] 처리 장치는 운영 체제 및 상기 운영 체제 상에서 수행되는 하나 이상의 소프트웨어 애플리케이션을 수행할 수 있다. 또한, 처리 장치는 소프트웨어의 실행에 응답하여, 데이터를 접근, 저장, 조작, 처리 및 생성할 수도 있다. 이해의 편의를 위하여, 처리 장치는 하나가 사용되는 것으로 설명된 경우도 있지만, 해당 기술 분야에서 통상의 지식을 가진 자는 처리 장치가 복수 개의 처리 요소(Processing Element) 및/또는 복수 유형의 처리요소를 포함할 수 있음을 이해할 것이다.
- [176] 예를 들어, 처리 장치는 복수 개의 프로세서 또는 하나의 프로세서 및 하나의 콘트롤러를 포함할 수 있다. 또한, 병렬 프로세서(Parallel Processor) 와 같은, 다른 처리 구성(Processing configuration)도 가능하다. 소프트웨어는 컴퓨터 프로그램(Computer Program), 코드(code), 명령(instruction), 또는 이들 중 하나 이상의 조합을 포함할 수 있으며, 원하는 대로 동작하도록 처리 장치를 구성하거나 독립적으로 또는 결합적으로(collectively) 처리 장치를 명령할 수 있다.
- [177] 소프트웨어 및/또는 데이터는, 처리 장치에 의하여 해석되거나 처리 장치에 명령 또는 데이터를 제공하기 위하여, 어떤 유형의 기계, 구성요소(component), 물리적 장치, 가상 장치(virtual equipment), 컴퓨터 저장 매체 또는 장치에 구체화(embody) 될 수 있다. 소프트웨어는 네트워크로 연결된 컴퓨터 시스템 상에 분산되어서, 분산된 방법으로 저장되거나 실행될 수도 있다. 소프트웨어 및 데이터는 하나 이상의 컴퓨터 판독 가능 기록 매체에 저장될 수 있다.
- [178] 실시예에 따른 방법은 다양한 컴퓨터 수단을 통하여 수행될 수 있는 프로그램 명령 형태로 구현되어 컴퓨터 판독 가능 매체에 기록될 수 있다. 컴퓨터 판독 가능 매체는 프로그램 명령, 데이터 파일, 데이터 구조 등을 단독으로 또는 조합하여 포함할 수 있다. 매체에 기록되는 프로그램 명령은 실시예를 위하여 특별히 설계되고 구성된 것들이거나 컴퓨터 소프트웨어 당업자에게 공지되어 사용 가능한 것일 수도 있다.

- [179] 컴퓨터 판독 가능 기록 매체의 예에는 하드 디스크, 플로피 디스크 및 자기 테이프와 같은 자기 매체(magnetic media), CDROM, DVD와 같은 광기록 매체(optical media) 및 롬(ROM), 램(RAM), 플래시 메모리 등과 같은 프로그램 명령을 저장하고 수행하도록 특별히 구성된 하드웨어 장치가 포함된다. 프로그램 명령의 예에는 컴파일러에 의해 만들어지는 것과 같은 기계어 코드뿐만 아니라 인터프리터 등을 사용해서 컴퓨터에 의해서 실행될 수 있는 고급 언어 코드를 포함한다. 상기된 하드웨어 장치는 실시예의 동작을 수행하기 위해 하나 이상의 소프트웨어 모듈로서 작동하도록 구성될 수 있으며, 그 역도 마찬가지이다.
- [180] 이상과 같이 실시예들이 비록 한정된 실시예와 도면에 의해 설명되었으나, 해당 기술분야에서 통상의 지식을 가진 자라면 상기의 기재로부터 다양한 수정 및 변형이 가능하다. 예를 들어, 설명된 기술들이 설명된 방법과 다른 순서로 수행되거나, 및/또는 설명된 시스템, 구조, 장치, 회로 등의 구성요소들이 설명된 방법과 다른 형태로 결합 또는 조합되거나, 다른 구성요소 또는 균등물에 의하여 대치되거나 치환되더라도 적절한 결과가 달성될 수 있다. 그러므로, 다른 구현들, 다른 실시예들 및 청구범위와 균등한 것들도 후술하는 청구범위의 범위에 속한다.

청구범위

- [청구항 1] i 번째 분자의 분자정보, 주변분자계정보, 및 분자특성정보를 수신하고 벡터화하는 벡터화부;
- 상기 벡터화된 분자정보에서 분자속성을 추출하고 상기 벡터화된 주변분자계정보에서 주변분자계속성을 추출하고 상기 벡터화된 분자특성정보에서 분자특성속성을 추출하는 속성추출부;
- 상기 분자속성, 상기 주변분자계속성, 및 상기 분자특성속성을 입력으로 수신하는 신경망 알고리즘인 통합속성 추출알고리즘을 이용하여 상기 i 번째 분자의 통합속성을 추출하는 통합속성추출부;
- 상기 통합속성을 입력으로 수신하는 신경망 알고리즘인 분자설계확률 계산알고리즘을 이용하여 상기 i 번째 분자를 기초로 분자설계를 위한 분자설계확률벡터를 추출하는 분자설계확률계산부; 및
- 상기 분자설계확률벡터에 기초하여 $i+1$ 번째 분자의 분자정보를 추출하거나 설계중지명령을 출력하여 최종분자를 출력하는 분자설계부를 포함하고,
- 상기 i 는 1보다 크거나 같은 정수인,
- 딥러닝 기반의 분자 설계 시스템.
- [청구항 2] 제1 항에 있어서,
- 상기 벡터화부는,
- 상기 i 번째 분자의 분자정보를 SMILES(Simplified Molecular-Input Line-Entry System)표현으로 수신하고, 분자핑거프린트(Molecular Fingerprint), 분자설명자(Molecular Descriptor), 화학구조식에 대한 이미지, 분자그래프(Molecular Graph), 분자좌표(Molecular Coordinates), 및 SMILES코드 중 적어도 하나의 표현방법을 이용하여 벡터화하는 분자정보벡터화부;
- 상기 i 번째 분자의 주변분자계정보를 상기 SMILES(Simplified molecular-Input Line-Entry System)표현으로 수신하고, 상기 분자핑거프린트(Molecular Fingerprint), 상기 분자설명자(Molecular Descriptor), 상기 화학구조식에 대한 이미지, 분자그래프(Molecular Graph), 상기 분자좌표(Molecular Coordinates), 및 상기 SMILES코드 중 적어도 하나의 표현방법을 이용하여 벡터화하는 주변분자계정보벡터화부; 및
- 상기 i 번째 분자의 분자특성정보를 문자열 또는 실수값 집합의 형태로 입력받고, 토큰화(tokenization), 정규화(normalization), 및 원-핫 인코딩(one-hot encoding) 중 적어도 하나의 표현방법을 이용하여 벡터화하는 분자특성정보벡터화부를 포함하는,
- 딥러닝 기반의 분자 설계 시스템.
- [청구항 3] 제2 항에 있어서,
- 상기 속성추출부는,

상기 벡터화된 i 번째 분자의 분자정보를 입력으로 수신하는 신경망 알고리즘인 분자속성 추출알고리즘을 이용하여 상기 i 번째 분자의 분자속성을 추출하는 분자속성추출부;

상기 벡터화된 i 번째 분자의 주변분자계정보를 입력으로 수신하는 신경망 알고리즘인 주변분자계속성 추출알고리즘을 이용하여 상기 i 번째 분자의 주변분자계속성을 추출하는 주변분자계속성추출부; 및

상기 벡터화된 i 번째 분자의 분자특성정보를 입력으로 수신하는 신경망 알고리즘인 분자특성속성 추출알고리즘을 이용하여 상기 i 번째 분자의 분자특성속성을 추출하는 분자특성속성추출부를 포함하는,

딥러닝 기반의 분자 설계 시스템.

[청구항 4]

제1 항에 있어서,

상기 분자정보는 화학구조식에 대한 정보를 포함하고,

상기 주변분자계정보는 하나 이상의 용매에 대한 정보를 포함하고,

상기 분자특성정보는 상기 분자의 구조적, 화학적, 물리적, 분광학적, 전기화학적, 반응성 중 적어도 하나 이상에 대한 정보를 포함하는,

딥러닝 기반의 분자 설계 시스템.

[청구항 5]

제4 항에 있어서,

상기 첫번째 분자의 분자정보는 상기 화학구조식이 없거나 사용자에게 의해 제공되는 어느 하나의 화학구조식에 대한 정보를 포함하는,

딥러닝 기반의 분자 설계 시스템.

[청구항 6]

제1 항에 있어서,

상기 분자설계부는,

상기 분자설계확률벡터를 구성하는 어느 하나의 원소를 이용하여 산출된 확률값에 따라 상기 $i+1$ 번째 분자를 설계하기 위한 상기 $i+1$ 번째 분자의 분자정보를 추출하고,

상기 $i+1$ 번째 분자의 분자정보는 상기 i 번째 분자를 구성하는 어느 하나의 원자에 한개의 원자를 결합하거나, 상기 i 번째 분자를 구성하는 원자 사이를 연결하는 결합을 추가하여 설계된 상기 $i+1$ 번째 분자의 화학구조식에 대한 정보를 포함하는,

딥러닝 기반의 분자 설계 시스템.

[청구항 7]

제1 항에 있어서,

상기 분자설계부는,

상기 분자설계확률벡터를 구성하는 어느 하나의 원소를 이용하여 산출된 확률값에 따라 상기 설계중지명령을 출력하여 상기 i 번째 분자를 상기 최종분자로 결정하는,

딥러닝 기반의 분자 설계 시스템.

[청구항 8]

제3 항에 있어서,

상기 분자속성 추출알고리즘, 상기 주변분자계속성 추출알고리즘, 상기 분자특성속성 추출알고리즘, 상기 통합속성 추출알고리즘, 및 상기 분자 설계확률 계산알고리즘은 적어도 하나 이상의 은닉계층(Hidden Layer)을 포함하는 상기 신경망 알고리즘인,

딥러닝 기반의 분자 설계 시스템.

[청구항 9]

벡터화부에 의해 i 번째 분자의 분자정보, 주변분자계정보, 및 분자특성정보를 수신하고 벡터화하는 단계;

속성추출부에 의해 상기 벡터화된 분자정보에서 분자속성을 추출하고 상기 벡터화된 주변분자계정보에서 주변분자계속성을 추출하고 상기 벡터화된 분자특성정보에서 분자특성속성을 추출하는 단계;

통합속성추출부에 의해 상기 분자속성, 상기 주변분자계속성, 및 상기 분자특성속성을 입력으로 수신하는 신경망 알고리즘인 통합속성 추출알고리즘을 이용하여 상기 i 번째 분자의 통합속성을 추출하는 단계;

분자설계확률계산부에 의해 상기 통합속성을 입력으로 수신하는 신경망 알고리즘인 분자설계확률 계산알고리즘을 이용하여 상기 i 번째 분자를 기초로 분자설계의 진행을 위한 분자설계확률벡터를 출력하는 단계; 및 분자설계부에 의해 상기 분자설계확률벡터에 기초하여 $i+1$ 번째 분자의 분자정보를 추출하거나 설계중지명령을 출력하여 최종분자를 출력하는 단계를 포함하고,

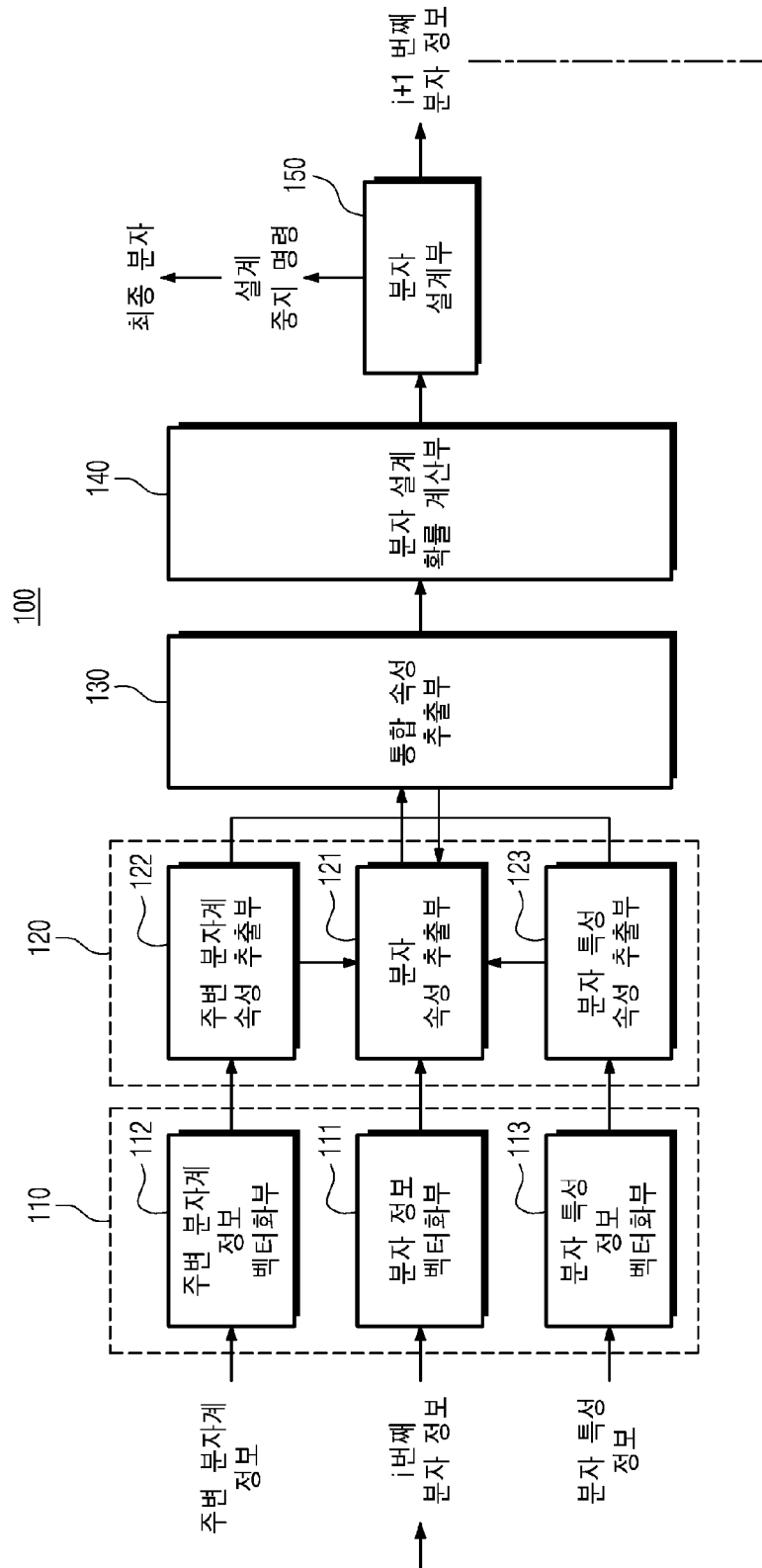
상기 i 는 1보다 크거나 같은 정수인,

딥러닝 기반의 분자 설계 방법.

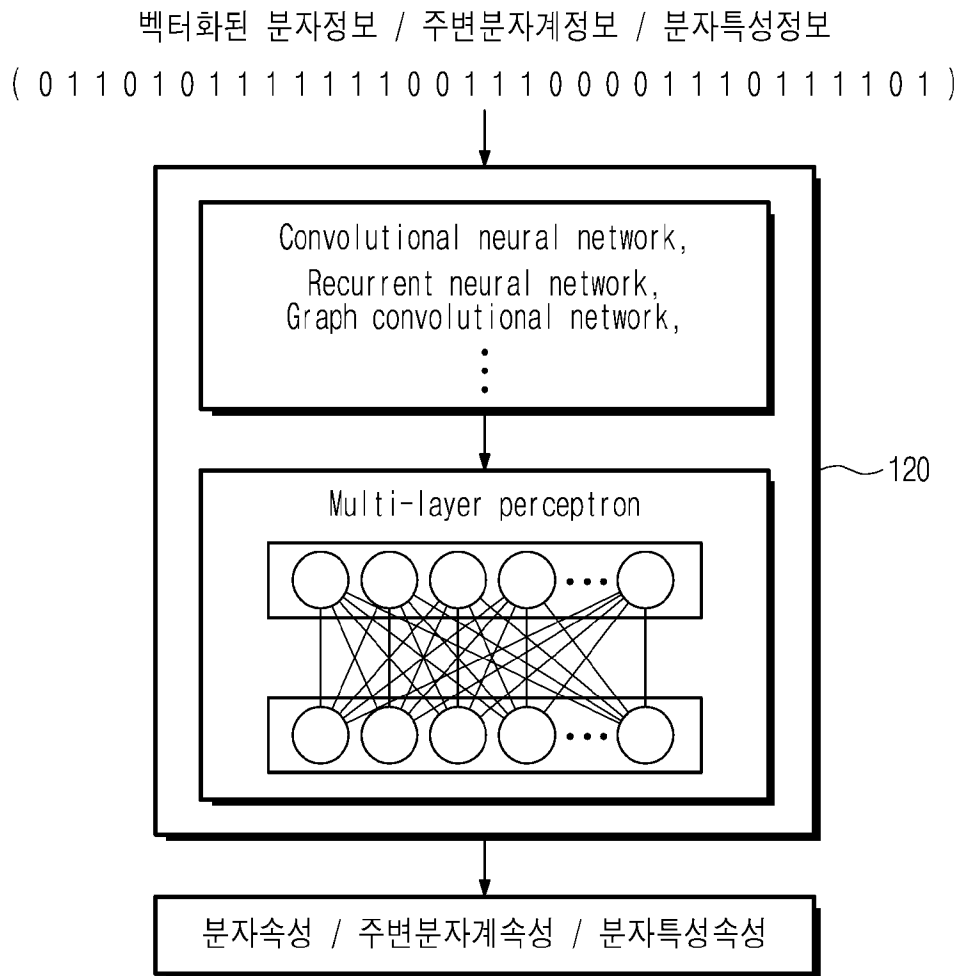
[청구항 10]

제9 항의 딥러닝 기반의 분자 설계 방법을 실행시키는 프로그램이 기록된 컴퓨터로 판독가능한 기록매체.

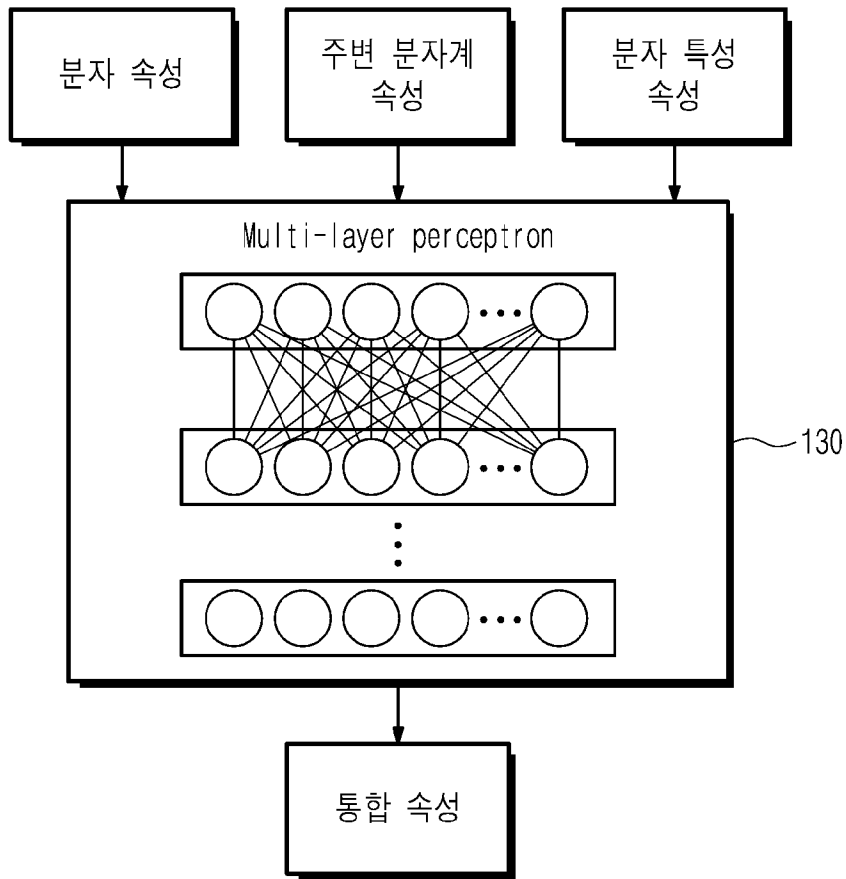
[도 1]



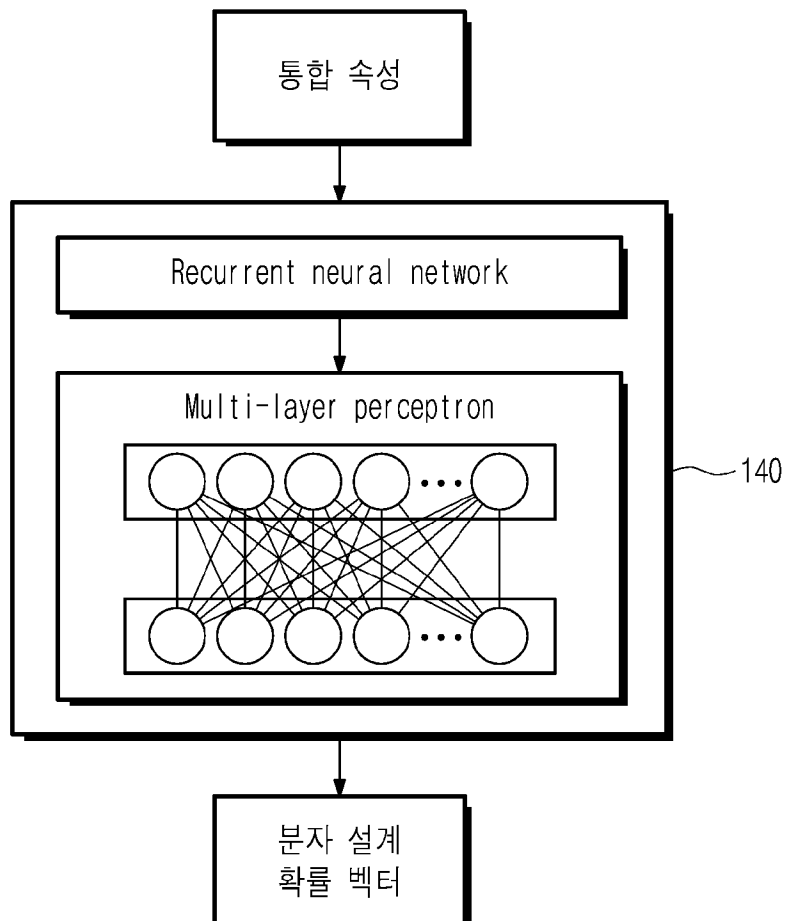
[도2a]



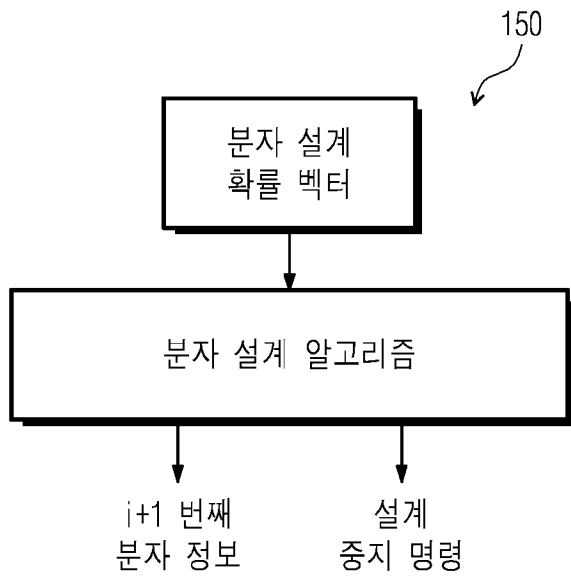
[도2b]



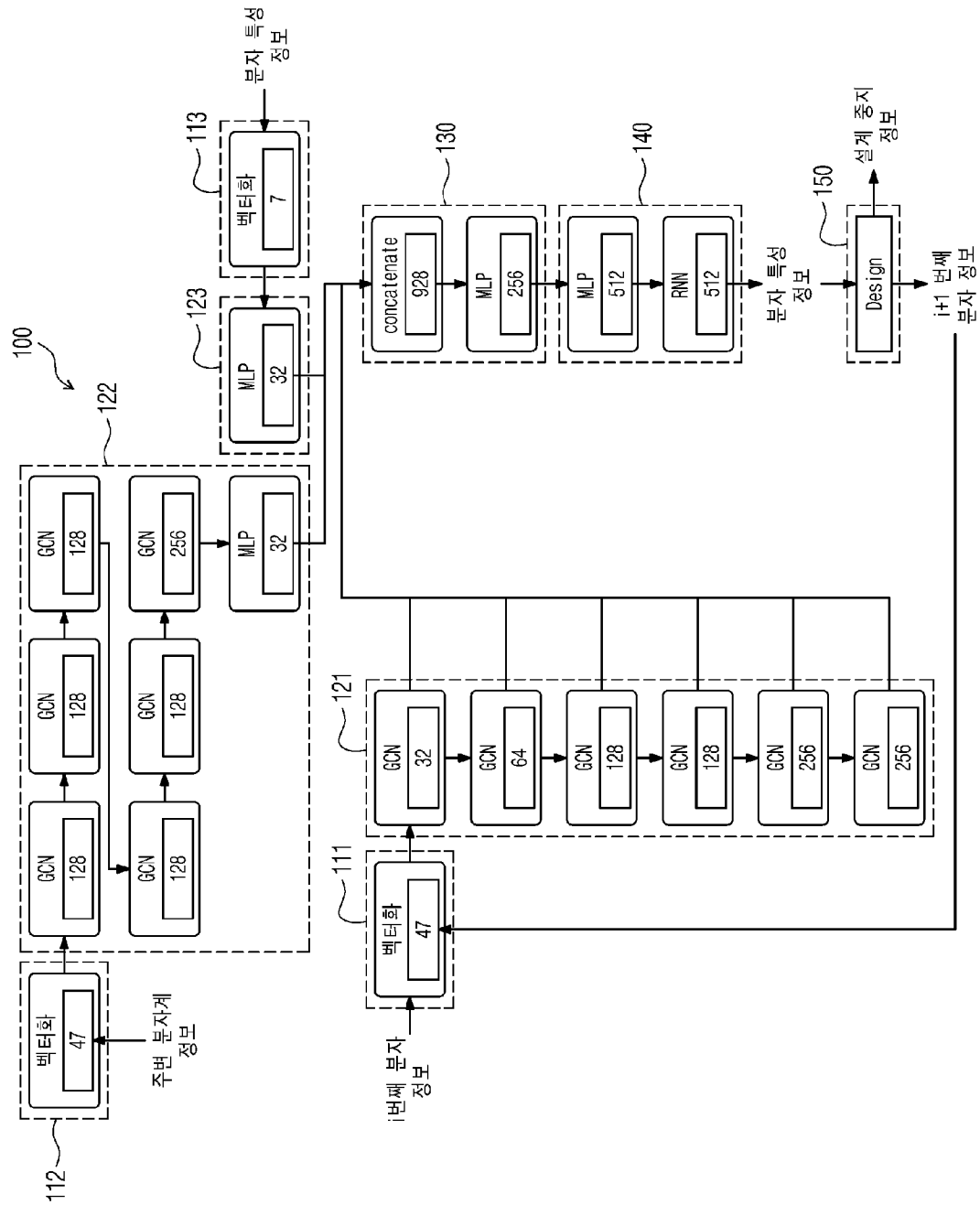
[도2c]



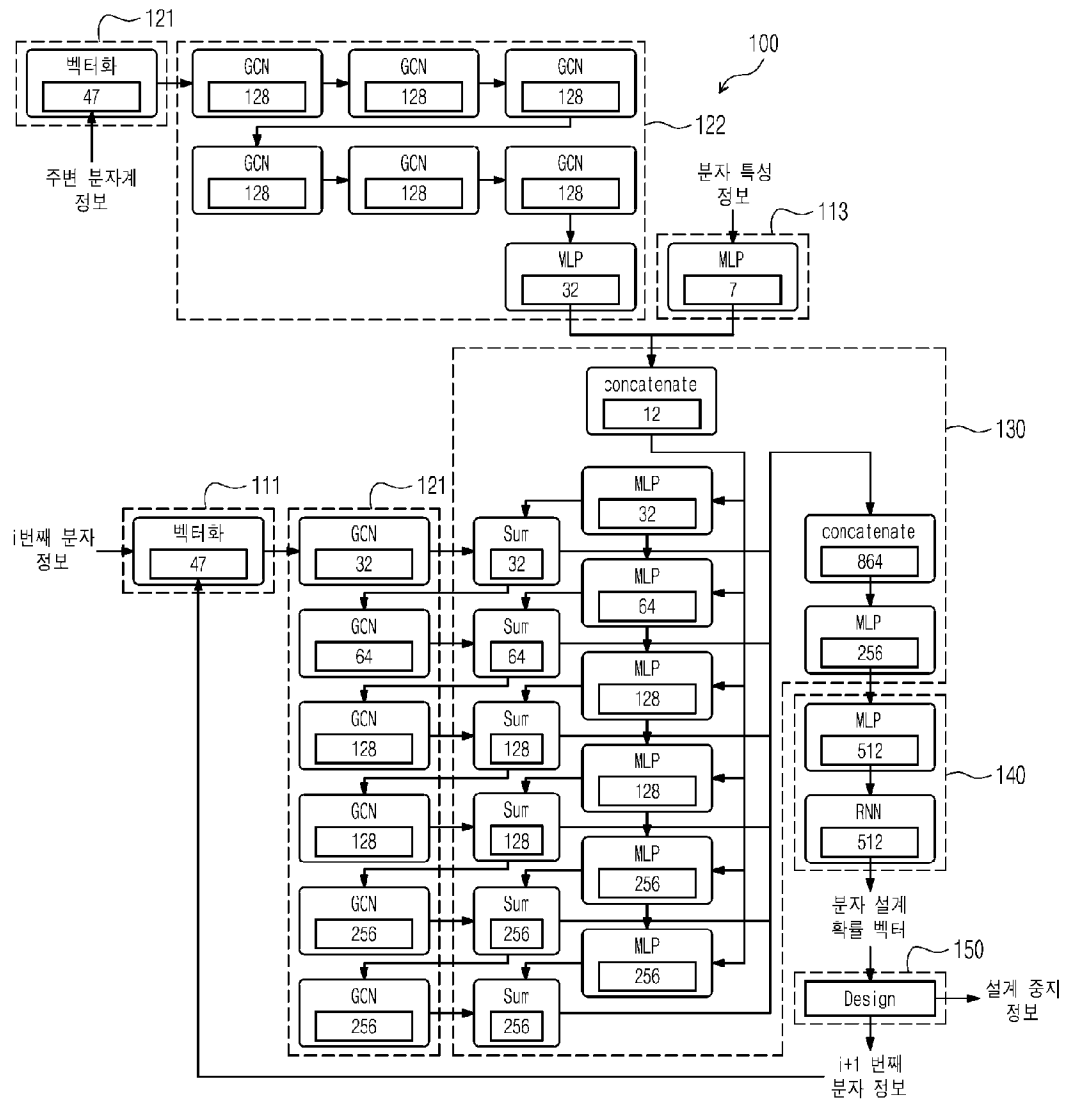
[도2d]



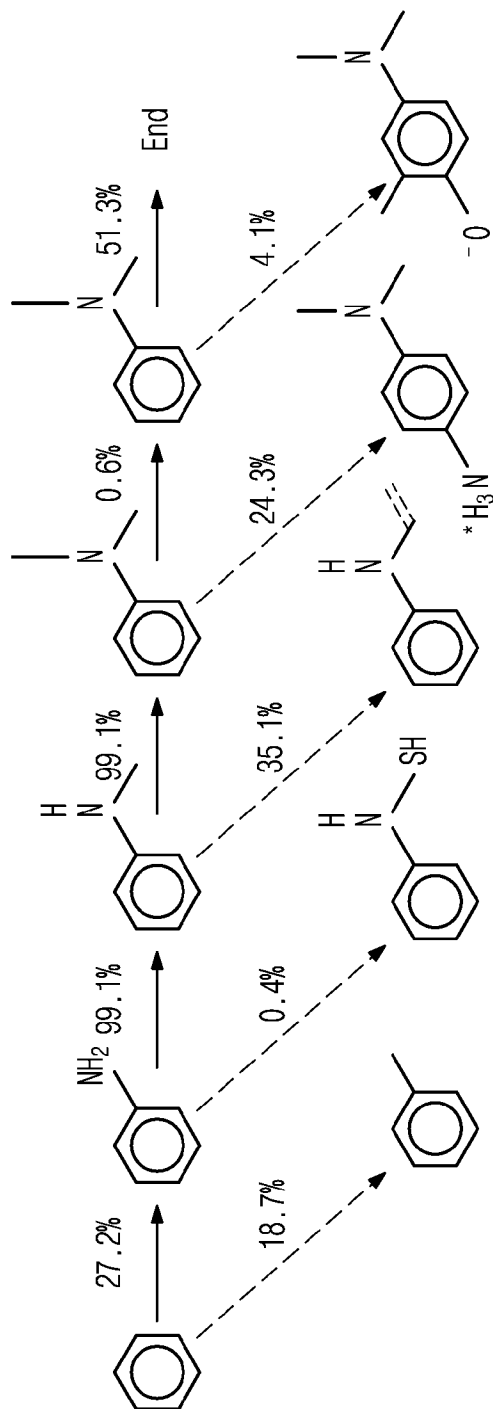
[도 3a]



[도3b]

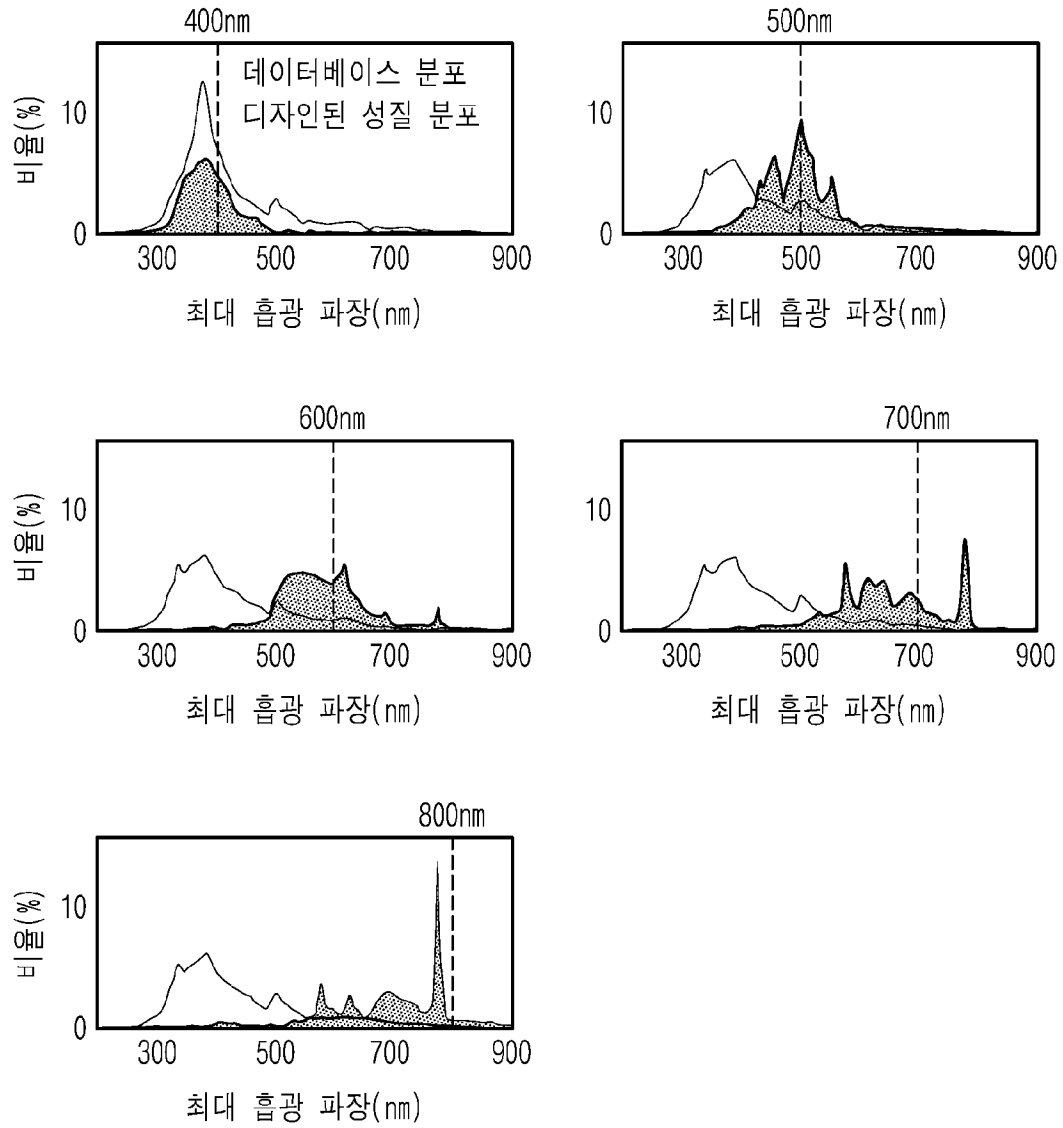


[도4]



[도5a]

최대 흡광 파장을 조건으로 분자 설계

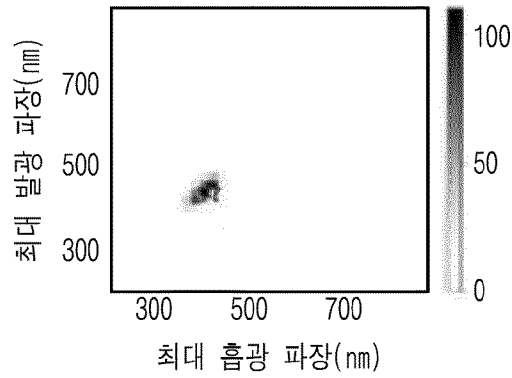


[도5b]

최대 흡광 파장과 최대 발광 파장을 조건으로 분자 설계

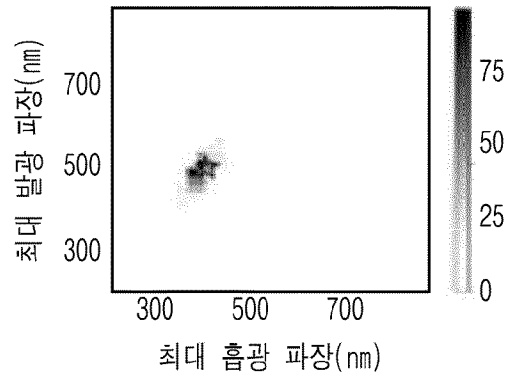
최대 흡광 파장: 400nm

최대 발광 파장: 450nm



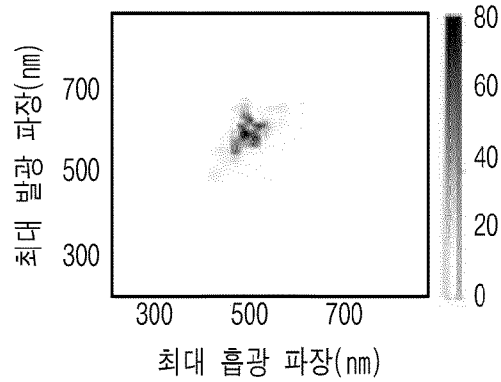
최대 흡광 파장: 400nm

최대 발광 파장: 500nm



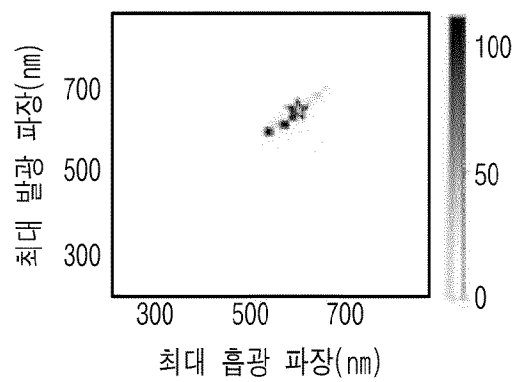
최대 흡광 파장: 500nm

최대 발광 파장: 600nm

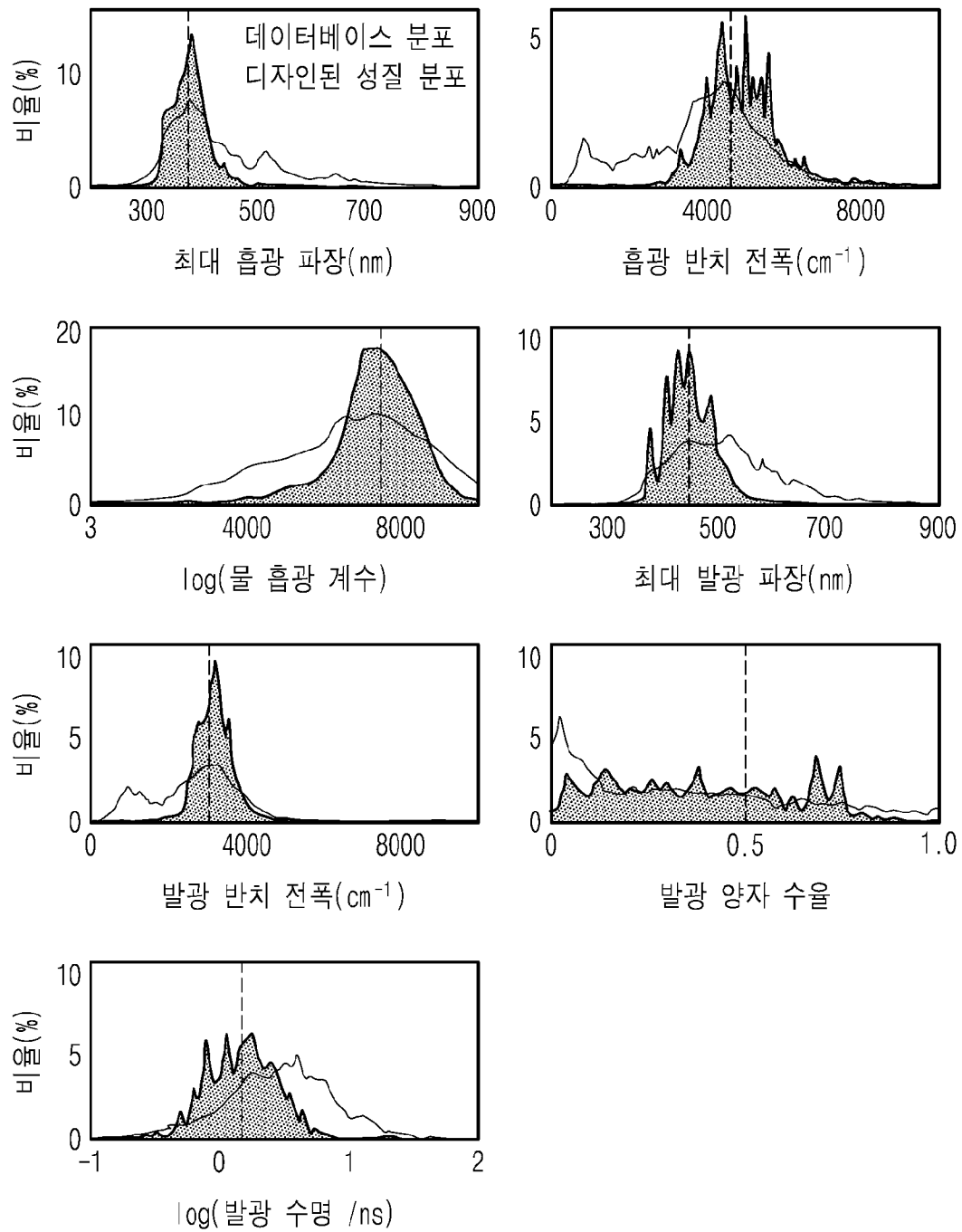


최대 흡광 파장: 600nm

최대 발광 파장: 650nm

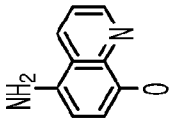
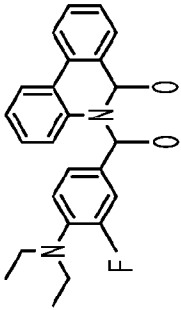
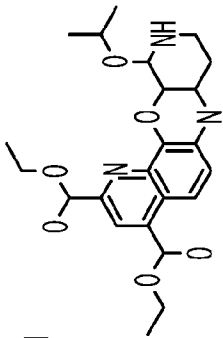
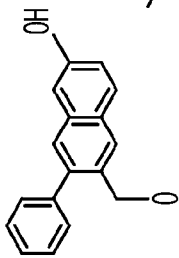
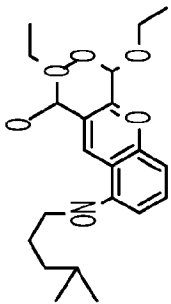


[도5c]

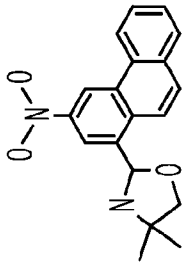
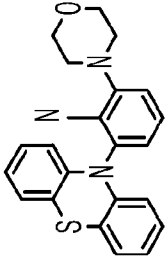
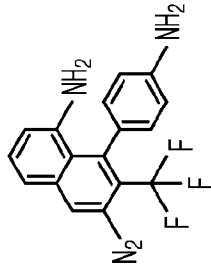
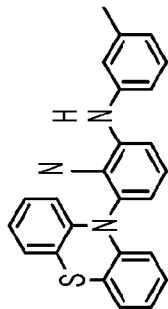
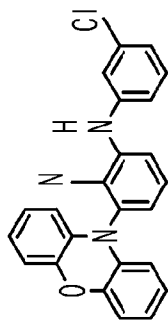


[도5d]

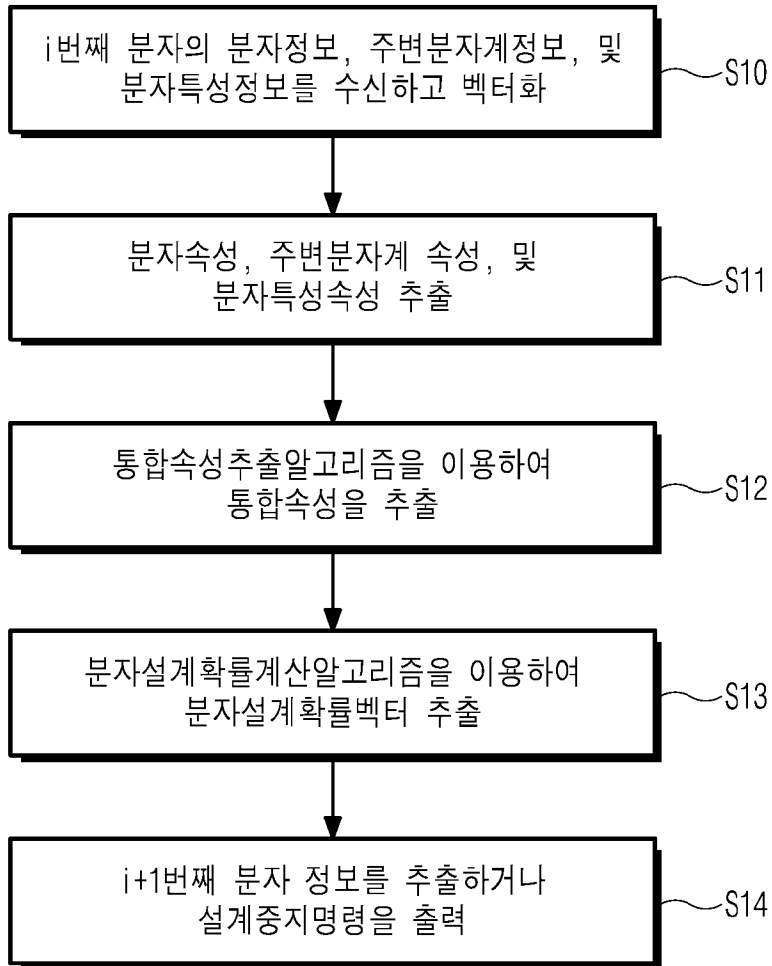
· 주변 분자계 = 물



· 주변 분자계 = 톨루엔



[도6]



INTERNATIONAL SEARCH REPORT

International application No.

PCT/KR2023/008705

A. CLASSIFICATION OF SUBJECT MATTER

G16C 60/00(2019.01)i; **G16C 20/70**(2019.01)i; **G16C 20/50**(2019.01)i; **G06N 3/08**(2006.01)i; **G06N 3/047**(2023.01)i;
G06N 7/01(2023.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G16C 60/00(2019.01); G06N 20/00(2019.01); G16B 5/20(2019.01); G16C 10/00(2019.01); G16C 20/50(2019.01);
 G16C 20/70(2019.01)

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Korean utility models and applications for utility models: IPC as above
 Japanese utility models and applications for utility models: IPC as above

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

eKOMPASS (KIPO internal) & keywords: 딥러닝(deep-learning), 분자 설계(molecular designing), 벡터(vector), 확률(probability), 알고리즘(algorithm)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	KR 10-2407120 B1 (HITS INC.) 10 June 2022 (2022-06-10) See claims 1, 3-6 and 8-10; and paragraphs [0003] and [0017].	1-10
A	US 2021-0304854 A1 (PEPTILOGICS, INC.) 30 September 2021 (2021-09-30) See claims 1-15.	1-10
A	KR 10-2019-0130446 A (SAMSUNG ELECTRONICS CO., LTD.) 22 November 2019 (2019-11-22) See claims 1-12; and figures 1-7.	1-10
A	US 2019-0304568 A1 (BOARD OF TRUSTEES OF MICHIGAN STATE UNIVERSITY) 03 October 2019 (2019-10-03) See claims 1-9.	1-10
A	SOUSA, T. et al. Generative Deep Learning for Targeted Compound Design. J. Chem. Inf. Model. 2021, vol. 61, pp. 5343-5361. See entire document.	1-10



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:

“A” document defining the general state of the art which is not considered to be of particular relevance
 “D” document cited by the applicant in the international application
 “E” earlier application or patent but published on or after the international filing date
 “L” document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
 “O” document referring to an oral disclosure, use, exhibition or other means
 “P” document published prior to the international filing date but later than the priority date claimed

“T” later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

“X” document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

“Y” document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

“&” document member of the same patent family

Date of the actual completion of the international search

25 September 2023

Date of mailing of the international search report

25 September 2023

Name and mailing address of the ISA/KR

Korean Intellectual Property Office
Government Complex-Daejeon Building 4, 189 Cheongsaro, Seo-gu, Daejeon 35208

Facsimile No. +82-42-481-8578

Authorized officer

Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/KR2023/008705

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
KR	10-2407120	B1	10 June 2022	None			
US	2021-0304854	A1	30 September 2021	CN	114270376	A	01 April 2022
				EP	3987460	A1	27 April 2022
				EP	3987460	A4	24 August 2022
				US	11049590	B1	29 June 2021
				US	11462304	B2	04 October 2022
				US	2021-0249108	A1	12 August 2021
				US	2021-0295955	A1	23 September 2021
				US	2021-0366581	A1	25 November 2021
				WO	2021-162874	A1	19 August 2021
KR	10-2019-0130446	A	22 November 2019	US	11721413	B2	08 August 2023
				US	2019-0325983	A1	24 October 2019
US	2019-0304568	A1	03 October 2019	US	2021-0027862	A1	28 January 2021
				WO	2019-191777	A1	03 October 2019

A. 발명이 속하는 기술분류(국제특허분류(IPC)) G16C 60/00(2019.01)i; G16C 20/70(2019.01)i; G16C 20/50(2019.01)i; G06N 3/08(2006.01)i; G06N 3/047(2023.01)i; G06N 7/01(2023.01)i		
B. 조사된 분야		
조사된 최소문헌(국제특허분류를 기재) G16C 60/00(2019.01); G06N 20/00(2019.01); G16B 5/20(2019.01); G16C 10/00(2019.01); G16C 20/50(2019.01); G16C 20/70(2019.01)		
조사된 기술분야에 속하는 최소문헌 이외의 문헌 한국등록실용신안공보 및 한국공개실용신안공보: 조사된 최소문헌란에 기재된 IPC 일본등록실용신안공보 및 일본공개실용신안공보: 조사된 최소문헌란에 기재된 IPC		
국제조사에 이용된 전산 데이터베이스(데이터베이스의 명칭 및 검색어(해당하는 경우)) eKOMPASS(특허청 내부 검색시스템) & 키워드: 딥러닝(deep-learning), 분자 설계(molecular designing), 벡터(vector), 확률(probability), 알고리즘(algorithm)		
C. 관련 문헌		
카테고리*	인용문헌명 및 관련 구절(해당하는 경우)의 기재	관련 청구항
A	KR 10-2407120 B1 (주식회사 히츠) 2022.06.10 청구항 1, 3-6, 8-10; 단락 [0003], [0017]	1-10
A	US 2021-0304854 A1 (PEPTILOGICS, INC.) 2021.09.30 청구항 1-15	1-10
A	KR 10-2019-0130446 A (삼성전자주식회사) 2019.11.22 청구항 1-12; 도면 1-7	1-10
A	US 2019-0304568 A1 (BOARD OF TRUSTEES OF MICHIGAN STATE UNIVERSITY) 2019.10.03 청구항 1-9	1-10
A	SOUSA, T. 등, 'Generative Deep Learning for Targeted Compound Design', J. Chem. Inf. Model., 2021, 61권, 페이지 5343-5361 전문	1-10
☐ 추가 문헌이 C(계속)에 기재되어 있습니다. ☒ 대응특허에 관한 별지를 참조하십시오.		
* 인용된 문헌의 특별 카테고리: “A” 특별히 관련이 없는 것으로 보이는 일반적인 기술수준을 정의한 문헌 “D” 본 국제출원에서 출원인이 인용한 문헌 “E” 국제출원일보다 빠른 출원일 또는 우선일을 가지나 국제출원일 이후에 공개된 선출원 또는 특허 문헌 “L” 우선권 주장에 의문을 제기하는 문헌 또는 다른 인용문헌의 공개일 또는 다른 특별한 이유(이유를 명시)를 밝히기 위하여 인용된 문헌 “O” 구두 개시, 사용, 전시 또는 기타 수단을 언급하고 있는 문헌 “P” 우선일 이후에 공개되었으나 국제출원일 이전에 공개된 문헌 “T” 국제출원일 또는 우선일 후에 공개된 문헌으로, 출원과 상충하지 않으며 발명의 기초가 되는 원리나 이론을 이해하기 위해 인용된 문헌 “X” 특별한 관련이 있는 문헌. 해당 문헌 하나만으로 청구된 발명의 신규성 또는 진보성이 없는 것으로 본다. “Y” 특별한 관련이 있는 문헌. 해당 문헌이 하나 이상의 다른 문헌과 조합하는 경우로 그 조합이 당업자에게 자명한 경우 청구된 발명은 진보성이 없는 것으로 본다. “&” 동일한 대응특허문헌에 속하는 문헌		
국제조사의 실제 완료일 2023년09월25일(25.09.2023)	국제조사보고서 발송일 2023년09월25일(25.09.2023)	
ISA/KR의 명칭 및 우편주소 대한민국 특허청 (35208) 대전광역시 서구 청사로 189, 4동 (둔산동, 정부대전청사) 팩스 번호 +82-42-481-8578	심사관 허주형 전화번호 +82-42-481-5373	

국제조사보고서에서 인용된 특허문헌	공개일	대응특허문헌	공개일
KR 10-2407120 B1	2022/06/10	없음	
US 2021-0304854 A1	2021/09/30	CN 114270376 A	2022/04/01
		EP 3987460 A1	2022/04/27
		EP 3987460 A4	2022/08/24
		US 11049590 B1	2021/06/29
		US 11462304 B2	2022/10/04
		US 2021-0249108 A1	2021/08/12
		US 2021-0295955 A1	2021/09/23
		US 2021-0366581 A1	2021/11/25
		WO 2021-162874 A1	2021/08/19
KR 10-2019-0130446 A	2019/11/22	US 11721413 B2	2023/08/08
		US 2019-0325983 A1	2019/10/24
US 2019-0304568 A1	2019/10/03	US 2021-0027862 A1	2021/01/28
		WO 2019-191777 A1	2019/10/03