

Doc Code: PA..

Document Description: Power of Attorney

PTO/AIA/82A (07-13)

Approved for use through 11/30/2014. OMB 0651-0051

U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE

Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

TRANSMITTAL FOR POWER OF ATTORNEY TO ONE OR MORE REGISTERED PRACTITIONERS

NOTE: This form is to be submitted with the Power of Attorney by Applicant form (PTO/AIA/82B) to identify the application to which the Power of Attorney is directed, in accordance with 37 CFR 1.5, unless the application number and filing date are identified in the Power of Attorney by Applicant form. If neither form PTO/AIA/82A nor form PTO/AIA82B identifies the application to which the Power of Attorney is directed, the Power of Attorney will not be recognized in the application.

Application Number	
Filing Date	
First Named Inventor	Sanghyun PARK
Title	TRAINING METHOD AND DEVICE FOR MOLECULAR GENERATION MODEL
Art Unit	
Examiner Name	
Attorney Docket Number	PB23-0097

SIGNATURE of Applicant or Patent Practitioner

Signature	/Andrew S. Park/	Date (Optional)	December 28, 2023
Name	Andrew S. Park	Registration Number	47,841
Title (if Applicant is a juristic entity)			
Applicant Name (if Applicant is a juristic entity)			

NOTE: This form must be signed in accordance with 37 CFR 1.33. See 37 CFR 1.4(d) for signature requirements and certifications. If more than one applicant, use multiple forms.



*Total of 2 forms are submitted.

This collection of information is required by 37 CFR 1.131, 1.32, and 1.33. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.11 and 1.14. This collection is estimated to take 3 minutes to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**

If you need assistance in completing the form, call 1-800-PTO-9199 and select option 2.

POWER OF ATTORNEY BY APPLICANT

I hereby revoke all previous powers of attorney given in the application identified in either the attached transmittal letter or the boxes below.

Application Number	Filing Date

(Note: The boxes above may be left blank if information is provided on form PTO/AIA/82A.)

☒ I hereby appoint the Patent Practitioner(s) associated with the following Customer Number as my/our attorney(s) or agent(s), and to transact all business in the United States Patent and Trademark Office connected therewith for the application referenced in the attached transmittal letter (form PTO/AIA/82A) or identified above:

67326

OR

☐ I hereby appoint Practitioner(s) named in the attached list (form PTO/AIA/82C) as my/our attorney(s) or agent(s), and to transact all business in the United States Patent and Trademark Office connected therewith for the patent application referenced in the attached transmittal letter (form PTO/AIA/82A) or identified above. (Note: Complete form PTO/AIA/82C.)

Please recognize or change the correspondence address for the application identified in the attached transmittal letter or the boxes above to:

☒ The address associated with the above-mentioned Customer Number

OR

☐ The address associated with Customer Number:

OR

Firm or Individual Name					
Address					
City	State	Zip			
Country					
Telephone	Email				

I am the Applicant (if the Applicant is a juristic entity, list the Applicant name in the box):

UIF (University Industry Foundation), Yonsei University

- ☐ Inventor or Joint Inventor (title not required below)
- ☐ Legal Representative of a Deceased or Legally Incapacitated Inventor (title not required below)
- ☒ Assignee or Person to Whom the Inventor is Under an Obligation to Assign (provide signer's title if applicant is a juristic entity)
- ☐ Person Who Otherwise Shows Sufficient Proprietary Interest (e.g., a petition under 37 CFR 1.46(b)(2) was granted in the application or is concurrently being filed with this document) (provide signer's title if applicant is a juristic entity)

SIGNATURE of Applicant for Patent

The undersigned (whose title is supplied below) is authorized to act on behalf of the applicant (e.g., where the applicant is a juristic entity).

Signature	<i>Si Hyun E. Kim</i>	Date (Optional)	<i>Sep. 24, 2023</i>
Name	<i>Si Hyun E. Kim</i>		
Title	<i>President</i>		

NOTE: Signature - This form must be signed by the applicant in accordance with 37 CFR 1.33. See 37 CFR 1.4 for signature requirements and certifications. If more than one applicant, use multiple forms.

☐ Total of _____ forms are submitted.

This collection of information is required by 37 CFR 1.131, 1.32, and 1.33. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.11 and 1.14. This collection is estimated to take 3 minutes to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.

If you need assistance in completing the form, call 1-800-PTO-9199 and select option 2.

Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

Application Data Sheet 37 CFR 1.76		Attorney Docket Number	PB23-0097
		Application Number	
Title of Invention	TRAINING METHOD AND DEVICE FOR MOLECULAR GENERATION MODEL		
<p>The application data sheet is part of the provisional or nonprovisional application for which it is being submitted. The following form contains the bibliographic data arranged in a format specified by the United States Patent and Trademark Office as outlined in 37 CFR 1.76.</p> <p>This document may be completed electronically and submitted to the Office in electronic format using the Electronic Filing System (EFS) or the document may be printed and included in a paper filed application.</p>			

Secrecy Order 37 CFR 5.2:

<input type="checkbox"/>	Portions or all of the application associated with this Application Data Sheet may fall under a Secrecy Order pursuant to 37 CFR 5.2 (Paper filers only. Applications that fall under Secrecy Order may not be filed electronically.)
--------------------------	---

Inventor Information:

Inventor 1					Remove
Legal Name					
Prefix	Given Name	Middle Name	Family Name	Suffix	
	Sanghyun		PARK		
Residence Information (Select One) <input type="radio"/> US Residency <input checked="" type="radio"/> Non US Residency <input type="radio"/> Active US Military Service					
City	Seoul	Country of Residence ⁱ	KR		
Mailing Address of Inventor:					
Address 1		204-3701, 300 Omok-ro, Yangcheon-gu			
Address 2					
City	Seoul	State/Province			
Postal Code		08004	Country ⁱ	KR	
Inventor 2					Remove
Legal Name					
Prefix	Given Name	Middle Name	Family Name	Suffix	
	Jonghwan		CHOI		
Residence Information (Select One) <input type="radio"/> US Residency <input checked="" type="radio"/> Non US Residency <input type="radio"/> Active US Military Service					
City	Incheon	Country of Residence ⁱ	KR		
Mailing Address of Inventor:					
Address 1		103-409, 4 Bongo-daero 691beon-gil, Gyeyang-gu			
Address 2					
City	Incheon	State/Province			
Postal Code		21090	Country ⁱ	KR	
Inventor 3					Remove
Legal Name					

Application Data Sheet 37 CFR 1.76		Attorney Docket Number	PB23-0097
		Application Number	
Title of Invention	TRAINING METHOD AND DEVICE FOR MOLECULAR GENERATION MODEL		

Prefix	Given Name	Middle Name	Family Name	Suffix
	Sangmin		SEO	
Residence Information (Select One) <input type="radio"/> US Residency <input checked="" type="radio"/> Non US Residency <input type="radio"/> Active US Military Service				
City	Seoul	Country of Residence ⁱ	KR	

Mailing Address of Inventor:				
Address 1	29 Yeonhui-ro 14-gil, Seodaemun-gu			
Address 2				
City	Seoul	State/Province		
Postal Code	03724	Country ⁱ	KR	
All Inventors Must Be Listed - Additional Inventor Information blocks may be generated within this form by selecting the Add button. <div style="text-align: right;">Add</div>				

Correspondence Information:

Enter either Customer Number or complete the Correspondence Information section below. For further information see 37 CFR 1.33(a).				
<input type="checkbox"/> An Address is being provided for the correspondence information of this application.				
Customer Number	67326			
Email Address	info@kpatent.com	Add Email	Remove Email	
Email Address	uspto@kpatent.com	Add Email	Remove Email	
Email Address	uspto.pks@gmail.com		Remove Email	

Application Information:

Title of the Invention	TRAINING METHOD AND DEVICE FOR MOLECULAR GENERATION MODEL			
Attorney Docket Number	PB23-0097	Small Entity Status Claimed <input checked="" type="checkbox"/>		
Application Type	Nonprovisional			
Subject Matter	Utility			
Total Number of Drawing Sheets (if any)		Suggested Figure for Publication (if any)		

Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

Application Data Sheet 37 CFR 1.76		Attorney Docket Number	PB23-0097
		Application Number	
Title of Invention	TRAINING METHOD AND DEVICE FOR MOLECULAR GENERATION MODEL		

Filing By Reference:

Only complete this section when filing an application by reference under 35 U.S.C. 111(c) and 37 CFR 1.57(a). Do not complete this section if application papers including a specification and any drawings are being filed. Any domestic benefit or foreign priority information must be provided in the appropriate section(s) below (i.e., "Domestic Benefit/National Stage Information" and "Foreign Priority Information").

For the purposes of a filing date under 37 CFR 1.53(b), the description and any drawings of the present application are replaced by this reference to the previously filed application, subject to conditions and requirements of 37 CFR 1.57(a).

Application number of the previously filed application	Filing date (YYYY-MM-DD)	Intellectual Property Authority or Country

Publication Information:

☐ Request Early Publication (Fee required at time of Request 37 CFR 1.219)

☐ **Request Not to Publish.** I hereby request that the attached application not be published under 35 U.S.C. 122(b) and certify that the invention disclosed in the attached application **has not and will not** be the subject of an application filed in another country, or under a multilateral international agreement, that requires publication at eighteen months after filing.

Representative Information:

Representative information should be provided for all practitioners having a power of attorney in the application. Providing this information in the Application Data Sheet does not constitute a power of attorney in the application (see 37 CFR 1.32). Either enter Customer Number or complete the Representative Name section below. If both sections are completed the customer Number will be used for the Representative Information during processing.

Please Select One:	<input checked="" type="radio"/> Customer Number	<input type="radio"/> US Patent Practitioner	<input type="radio"/> Limited Recognition (37 CFR 11.9)
Customer Number	67326		

Domestic Benefit/National Stage Information:

This section allows for the applicant to either claim benefit under 35 U.S.C. 119(e), 120, 121, 365(c), or 386(c) or indicate National Stage entry from a PCT application. Providing benefit claim information in the Application Data Sheet constitutes the specific reference required by 35 U.S.C. 119(e) or 120, and 37 CFR 1.78.

When referring to the current application, please leave the "Application Number" field blank.

Prior Application Status			Remove
Application Number	Continuity Type	Prior Application Number	Filing or 371(c) Date (YYYY-MM-DD)
Additional Domestic Benefit/National Stage Data may be generated within this form by selecting the Add button.			Add

Application Data Sheet 37 CFR 1.76		Attorney Docket Number	PB23-0097
		Application Number	
Title of Invention	TRAINING METHOD AND DEVICE FOR MOLECULAR GENERATION MODEL		

Foreign Priority Information:

This section allows for the applicant to claim priority to a foreign application. Providing this information in the application data sheet constitutes the claim for priority as required by 35 U.S.C. 119(b) and 37 CFR 1.55. When priority is claimed to a foreign application that is eligible for retrieval under the priority document exchange program (PDX)ⁱ the information will be used by the Office to automatically attempt retrieval pursuant to 37 CFR 1.55(i)(1) and (2). Under the PDX program, applicant bears the ultimate responsibility for ensuring that a copy of the foreign application is received by the Office from the participating foreign intellectual property office, or a certified copy of the foreign priority application is filed, within the time period specified in 37 CFR 1.55(g)(1).

Application Number	Country ⁱ	Filing Date (YYYY-MM-DD)	Access Code ⁱ (if applicable)
10-2023-0043303	KR	2023-04-03	7278

Additional Foreign Priority Data may be generated within this form by selecting the **Add** button.

Statement under 37 CFR 1.55 or 1.78 for AIA (First Inventor to File) Transition Applications

☐ This application (1) claims priority to or the benefit of an application filed before March 16, 2013 and (2) also contains, or contained at any time, a claim to a claimed invention that has an effective filing date on or after March 16, 2013.

NOTE: By providing this statement under 37 CFR 1.55 or 1.78, this application, with a filing date on or after March 16, 2013, will be examined under the first inventor to file provisions of the AIA.

Application Data Sheet 37 CFR 1.76		Attorney Docket Number	PB23-0097
		Application Number	
Title of Invention	TRAINING METHOD AND DEVICE FOR MOLECULAR GENERATION MODEL		

Authorization or Opt-Out of Authorization to Permit Access:

When this Application Data Sheet is properly signed and filed with the application, applicant has provided written authority to permit a participating foreign intellectual property (IP) office access to the instant application-as-filed (see paragraph A in subsection 1 below) and the European Patent Office (EPO) access to any search results from the instant application (see paragraph B in subsection 1 below).

Should applicant choose not to provide an authorization identified in subsection 1 below, applicant **must opt-out** of the authorization by checking the corresponding box A or B or both in subsection 2 below.

NOTE: This section of the Application Data Sheet is **ONLY** reviewed and processed with the **INITIAL** filing of an application. After the initial filing of an application, an Application Data Sheet cannot be used to provide or rescind authorization for access by a foreign IP office(s). Instead, Form PTO/SB/39 or PTO/SB/69 must be used as appropriate.

1. Authorization to Permit Access by a Foreign Intellectual Property Office(s)

A. Priority Document Exchange (PDX) - Unless box A in subsection 2 (opt-out of authorization) is checked, the undersigned hereby **grants the USPTO authority** to provide the European Patent Office (EPO), the Japan Patent Office (JPO), the Korean Intellectual Property Office (KIPO), the State Intellectual Property Office of the People's Republic of China (SIPO), the World Intellectual Property Organization (WIPO), and any other foreign intellectual property office participating with the USPTO in a bilateral or multilateral priority document exchange agreement in which a foreign application claiming priority to the instant patent application is filed, access to: (1) the instant patent application-as-filed and its related bibliographic data, (2) any foreign or domestic application to which priority or benefit is claimed by the instant application and its related bibliographic data, and (3) the date of filing of this Authorization. See 37 CFR 1.14(h)(1).

B. Search Results from U.S. Application to EPO - Unless box B in subsection 2 (opt-out of authorization) is checked, the undersigned hereby **grants the USPTO authority** to provide the EPO access to the bibliographic data and search results from the instant patent application when a European patent application claiming priority to the instant patent application is filed. See 37 CFR 1.14(h)(2).

The applicant is reminded that the EPO's Rule 141(1) EPC (European Patent Convention) requires applicants to submit a copy of search results from the instant application without delay in a European patent application that claims priority to the instant application.

2. Opt-Out of Authorizations to Permit Access by a Foreign Intellectual Property Office(s)

☐ A. Applicant **DOES NOT** authorize the USPTO to permit a participating foreign IP office access to the instant application-as-filed. If this box is checked, the USPTO will not be providing a participating foreign IP office with any documents and information identified in subsection 1A above.

☐ B. Applicant **DOES NOT** authorize the USPTO to transmit to the EPO any search results from the instant patent application. If this box is checked, the USPTO will not be providing the EPO with search results from the instant application.

NOTE: Once the application has published or is otherwise publicly available, the USPTO may provide access to the application in accordance with 37 CFR 1.14.

Application Data Sheet 37 CFR 1.76		Attorney Docket Number	PB23-0097
		Application Number	
Title of Invention	TRAINING METHOD AND DEVICE FOR MOLECULAR GENERATION MODEL		

Applicant Information:

Providing assignment information in this section does not substitute for compliance with any requirement of part 3 of Title 37 of CFR to have an assignment recorded by the Office.			
Applicant 1			Remove
<p>If the applicant is the inventor (or the remaining joint inventor or inventors under 37 CFR 1.45), this section should not be completed. The information to be provided in this section is the name and address of the legal representative who is the applicant under 37 CFR 1.43; or the name and address of the assignee, person to whom the inventor is under an obligation to assign the invention, or person who otherwise shows sufficient proprietary interest in the matter who is the applicant under 37 CFR 1.46. If the applicant is an applicant under 37 CFR 1.46 (assignee, person to whom the inventor is obligated to assign, or person who otherwise shows sufficient proprietary interest) together with one or more joint inventors, then the joint inventor or inventors who are also the applicant should be identified in this section.</p>			
Clear			
<input checked="" type="radio"/> Assignee	<input type="radio"/> Legal Representative under 35 U.S.C. 117	<input type="radio"/> Joint Inventor	
<input type="radio"/> Person to whom the inventor is obligated to assign.		<input type="radio"/> Person who shows sufficient proprietary interest	
If applicant is the legal representative, indicate the authority to file the patent application, the inventor is:			
Name of the Deceased or Legally Incapacitated Inventor: <input type="text"/>			
If the Applicant is an Organization check here. <input checked="" type="checkbox"/>			
Organization Name	UIF (University Industry Foundation), Yonsei University		
Mailing Address Information For Applicant:			
Address 1	50 Yonsei-ro, Seodaemun-gu		
Address 2			
City	Seoul	State/Province	
Country	KR	Postal Code	03722
Phone Number		Fax Number	
Email Address			
Additional Applicant Data may be generated within this form by selecting the Add button. Add			

Assignee Information including Non-Applicant Assignee Information:

Providing assignment information in this section does not substitute for compliance with any requirement of part 3 of Title 37 of CFR to have an assignment recorded by the Office.

Application Data Sheet 37 CFR 1.76		Attorney Docket Number	PB23-0097
		Application Number	
Title of Invention	TRAINING METHOD AND DEVICE FOR MOLECULAR GENERATION MODEL		

Assignee 1			
Complete this section if assignee information, including non-applicant assignee information, is desired to be included on the patent application publication. An assignee-applicant identified in the "Applicant Information" section will appear on the patent application publication as an applicant. For an assignee-applicant, complete this section only if identification as an assignee is also desired on the patent application publication.			
Remove			
If the Assignee or Non-Applicant Assignee is an Organization check here. <input checked="" type="checkbox"/>			
Organization Name	UIF (University Industry Foundation), Yonsei University		
Mailing Address Information For Assignee including Non-Applicant Assignee:			
Address 1	50 Yonsei-ro, Seodaemun-gu		
Address 2			
City	Seoul	State/Province	
Country ⁱ	KR	Postal Code	03722
Phone Number		Fax Number	
Email Address			
Additional Assignee or Non-Applicant Assignee Data may be generated within this form by selecting the Add button. Add			

Signature:[Remove](#)

NOTE: This Application Data Sheet must be signed in accordance with 37 CFR 1.33(b). However, if this Application Data Sheet is submitted with the **INITIAL** filing of the application **and** either box A or B is **not** checked in subsection 2 of the "Authorization or Opt-Out of Authorization to Permit Access" section, then this form must also be signed in accordance with 37 CFR 1.14(c).

This Application Data Sheet **must** be signed by a patent practitioner if one or more of the applicants is a **juristic entity** (e.g., corporation or association). If the applicant is two or more joint inventors, this form must be signed by a patent practitioner, **all** joint inventors who are the applicant, or one or more joint inventor-applicants who have been given power of attorney (e.g., see USPTO Form PTO/AIA/81) on behalf of **all** joint inventor-applicants.

See 37 CFR 1.4(d) for the manner of making signatures and certifications.

Signature	/Andrew S. Park/		Date (YYYY-MM-DD)	2023-12-28	
First Name	Andrew	Last Name	Park	Registration Number	47841
Additional Signature may be generated within this form by selecting the Add button. Add					

Application Data Sheet 37 CFR 1.76	Attorney Docket Number	PB23-0097
	Application Number	
Title of Invention	TRAINING METHOD AND DEVICE FOR MOLECULAR GENERATION MODEL	

This collection of information is required by 37 CFR 1.76. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 23 minutes to complete, including gathering, preparing, and submitting the completed application data sheet form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**

Privacy Act Statement

The Privacy Act of 1974 (P.L. 93-579) requires that you be given certain information in connection with your submission of the attached form related to a patent application or patent. Accordingly, pursuant to the requirements of the Act, please be advised that: (1) the general authority for the collection of this information is 35 U.S.C. 2(b)(2); (2) furnishing of the information solicited is voluntary; and (3) the principal purpose for which the information is used by the U.S. Patent and Trademark Office is to process and/or examine your submission related to a patent application or patent. If you do not furnish the requested information, the U.S. Patent and Trademark Office may not be able to process and/or examine your submission, which may result in termination of proceedings or abandonment of the application or expiration of the patent.

The information provided by you in this form will be subject to the following routine uses:

- 1 The information on this form will be treated confidentially to the extent allowed under the Freedom of Information Act (5 U.S.C. 552) and the Privacy Act (5 U.S.C. 552a). Records from this system of records may be disclosed to the Department of Justice to determine whether the Freedom of Information Act requires disclosure of these records.
- 2 A record from this system of records may be disclosed, as a routine use, in the course of presenting evidence to a court, magistrate, or administrative tribunal, including disclosures to opposing counsel in the course of settlement negotiations.
- 3 A record in this system of records may be disclosed, as a routine use, to a Member of Congress submitting a request involving an individual, to whom the record pertains, when the individual has requested assistance from the Member with respect to the subject matter of the record.
- 4 A record in this system of records may be disclosed, as a routine use, to a contractor of the Agency having need for the information in order to perform a contract. Recipients of information shall be required to comply with the requirements of the Privacy Act of 1974, as amended, pursuant to 5 U.S.C. 552a(m).
- 5 A record related to an International Application filed under the Patent Cooperation Treaty in this system of records may be disclosed, as a routine use, to the International Bureau of the World Intellectual Property Organization, pursuant to the Patent Cooperation Treaty.
- 6 A record in this system of records may be disclosed, as a routine use, to another federal agency for purposes of National Security review (35 U.S.C. 181) and for review pursuant to the Atomic Energy Act (42 U.S.C. 218(c)).
- 7 A record from this system of records may be disclosed, as a routine use, to the Administrator, General Services, or his/her designee, during an inspection of records conducted by GSA as part of that agency's responsibility to recommend improvements in records management practices and programs, under authority of 44 U.S.C. 2904 and 2906. Such disclosure shall be made in accordance with the GSA regulations governing inspection of records for this purpose, and any other relevant (i.e., GSA or Commerce) directive. Such disclosure shall not be used to make determinations about individuals.
- 8 A record from this system of records may be disclosed, as a routine use, to the public after either publication of the application pursuant to 35 U.S.C. 122(b) or issuance of a patent pursuant to 35 U.S.C. 151. Further, a record may be disclosed, subject to the limitations of 37 CFR 1.14, as a routine use, to the public if the record was filed in an application which became abandoned or in which the proceedings were terminated and which application is referenced by either a published application, an application open to public inspections or an issued patent.
- 9 A record from this system of records may be disclosed, as a routine use, to a Federal, State, or local law enforcement agency, if the USPTO becomes aware of a violation or potential violation of law or regulation.



ELECTRONIC PAYMENT RECEIPT

APPLICATION #
18/399,450

RECEIPT DATE / TIME
12/28/2023 07:08:46 PM Z ET

ATTORNEY DOCKET #
PB23-0097

Title of Invention

TRAINING METHOD AND DEVICE FOR MOLECULAR GENERATION MODEL

Application Information

APPLICATION TYPE Utility - Nonprovisional Application
under 35 USC 111(a)

PATENT # -

CONFIRMATION # 4354

FILED BY SY MOON

PATENT CENTER # 63769961

AUTHORIZED BY Andrew Park

CUSTOMER # 67326

FILING DATE -

CORRESPONDENCE
ADDRESS -

FIRST NAMED
INVENTOR Sanghyun PARK

Payment Information

PAYMENT METHOD
CARD / 2160

PAYMENT TRANSACTION ID
E2023BRJ10159069

PAYMENT AUTHORIZED BY
SY MOON

FEE CODE	DESCRIPTION	ITEM PRICE(\$)	QUANTITY	ITEM TOTAL(\$)
2053	NON-ENGLISH TRANSLATION	56.00	1	56.00
2111	UTILITY PATENT APPL. SEARCH FEE	280.00	1	280.00
4011	BASIC FILING FEE- UTILITY	64.00	1	64.00
2311	EXAMINATION OF ORIGINAL PATENT APPLICATION	320.00	1	320.00
TOTAL AMOUNT:				\$720.00

by the applicant, and including page counts, where applicable. It serves as evidence of receipt similar to a Post Card, as described in MPEP 503.

New Applications Under 35 U.S.C. 111

If a new application is being filed and the application includes the necessary components for filing date (see 37 CFR 1.53(b)-(d) and MPEP 506), a Filing Receipt (37 CFR 1.54) will be issued in due course and the date shown on this Acknowledgement Receipt will establish the filing date of the application.

National Stage of an International Application under 35 U.S.C. 371

If a timely submission to enter the national stage of an international application is compliant with the conditions of 35 U.S.C. 371 and other applicable requirements a Form PCT/DO/EO/903 indicating acceptance of the application as a national stage submission under 35 U.S.C. 371 will be issued in addition to the Filing Receipt, in due course.

New International Application Filed with the USPTO as a Receiving Office

If a new international application is being filed and the international application includes the necessary components for an international filing date (see PCT Article 11 and MPEP 1810), a Notification of the International Application Number and of the International Filing Date (Form PCT/RO/105) will be issued in due course, subject to prescriptions concerning national security, and the date shown on this Acknowledgement Receipt will establish the international filing date of the application.

FIG. 1

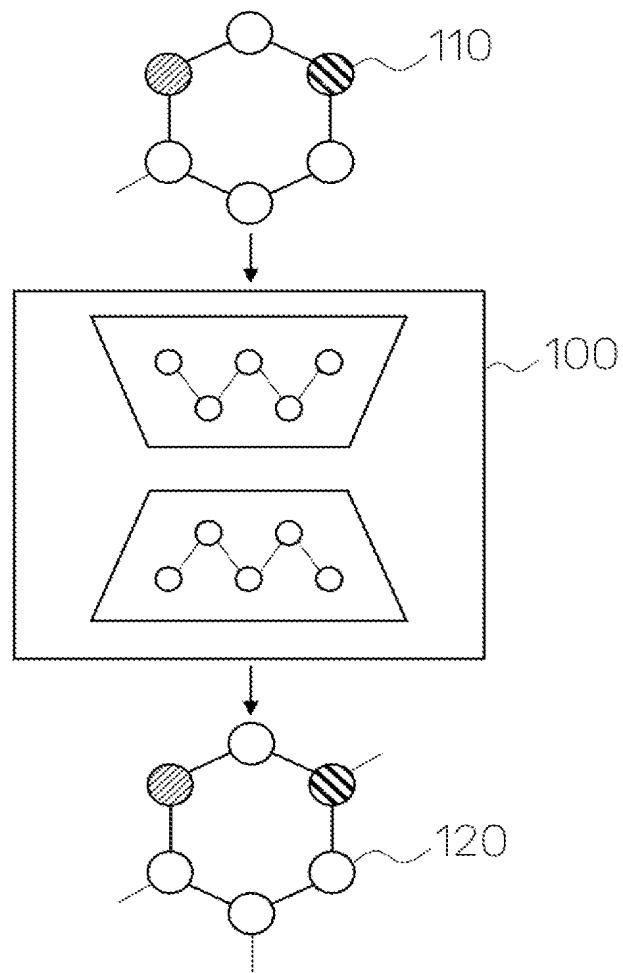


FIG. 2A

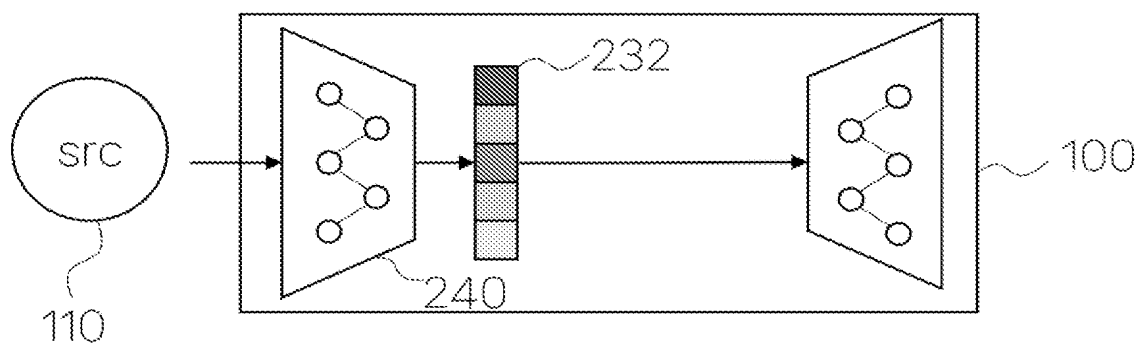


FIG. 2B

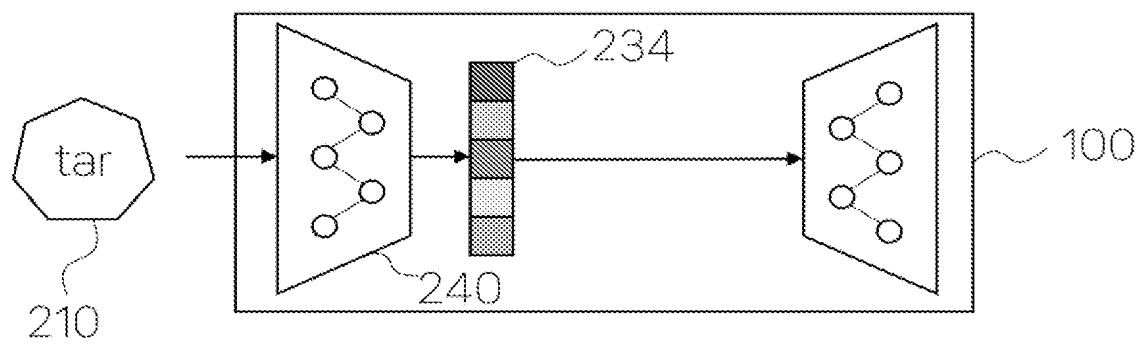


FIG. 2C

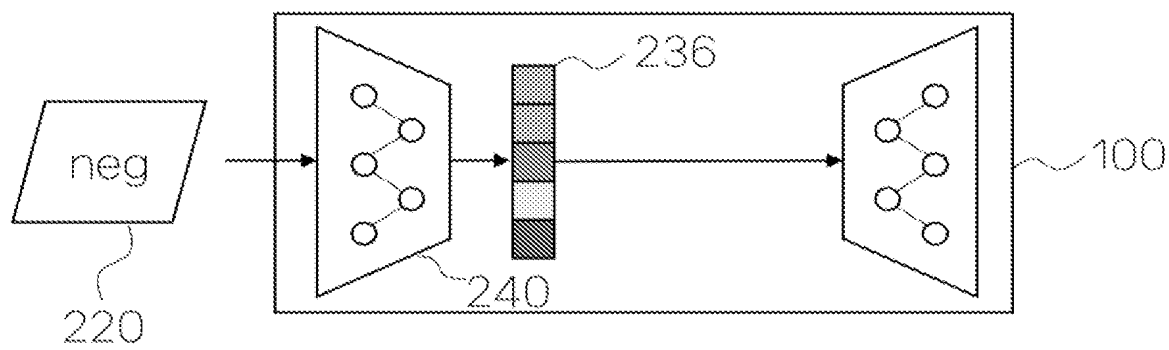


FIG. 3

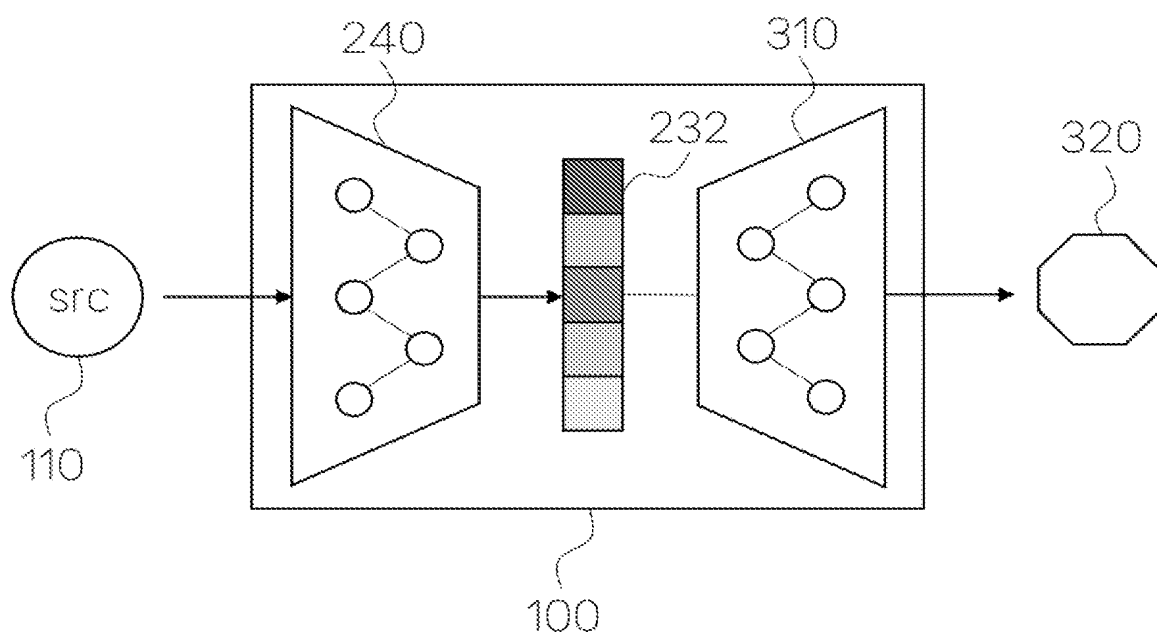
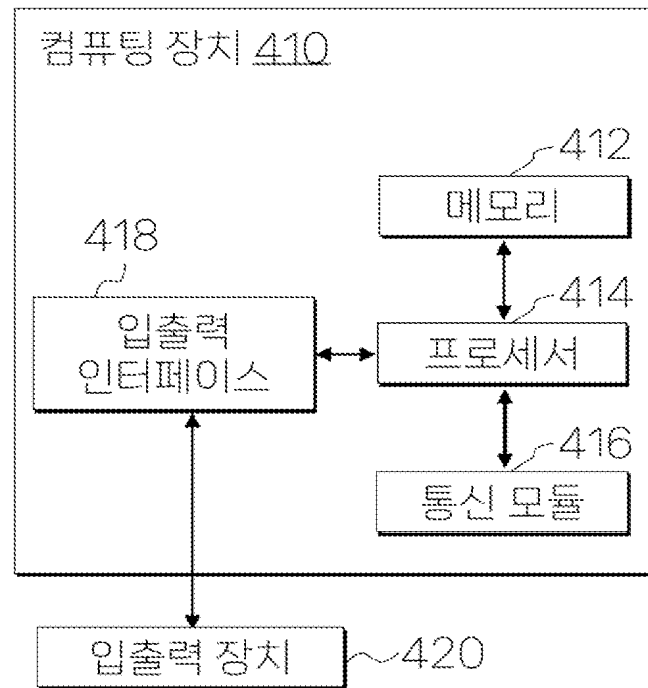


FIG. 4



DRD2

Model	0.40	0.45	0.50	0.55	0.60	0.65	0.70
COMA	0.48	0.32	0.21	0.12	0.07	0.03	0.02
VJTNN	0.23	0.13	0.07	0.03	0.02	0.01	0.01
VJTNN+GAN	0.22	0.12	0.06	0.03	0.02	0.01	0.01
CORE	0.18	0.08	0.04	0.02	0.01	0.01	0.01
HierG2G	0.17	0.07	0.03	0.01	0.01	0.01	0.01
HierG2G+BT	0.16	0.06	0.03	0.01	0.01	0.01	0.01
JTVAE	0.10	0.05	0.02	0.01	0.01	0.01	0.01
UGMMT	0.09	0.04	0.02	0.01	0.01	0.01	0.01

FIG. 5B

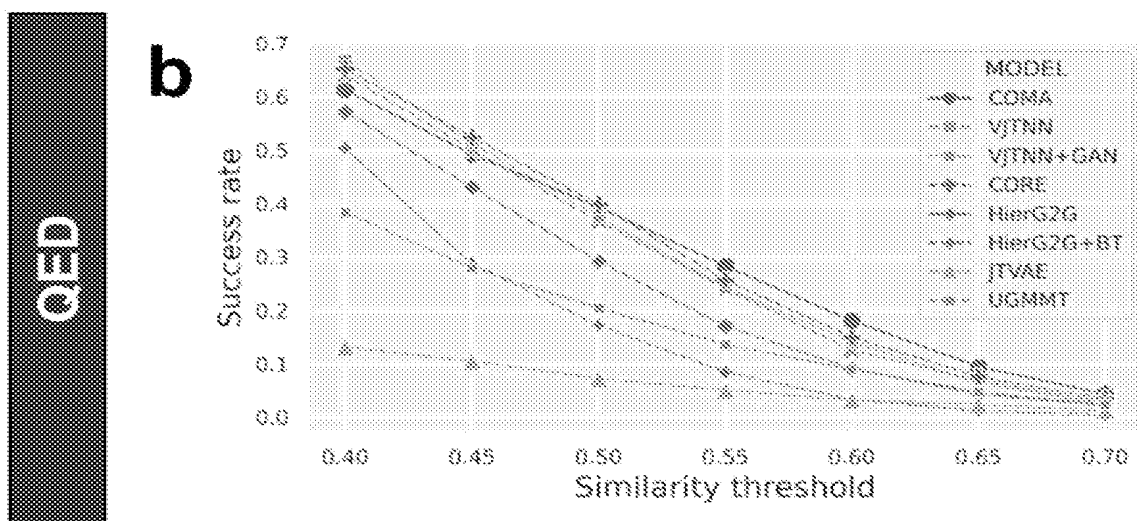


FIG. 5C

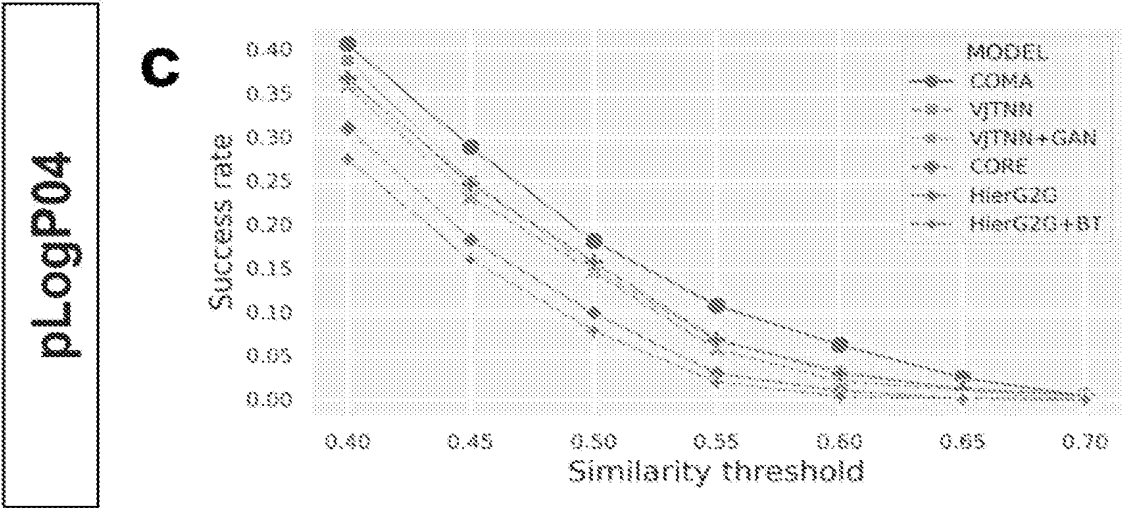


FIG. 5D

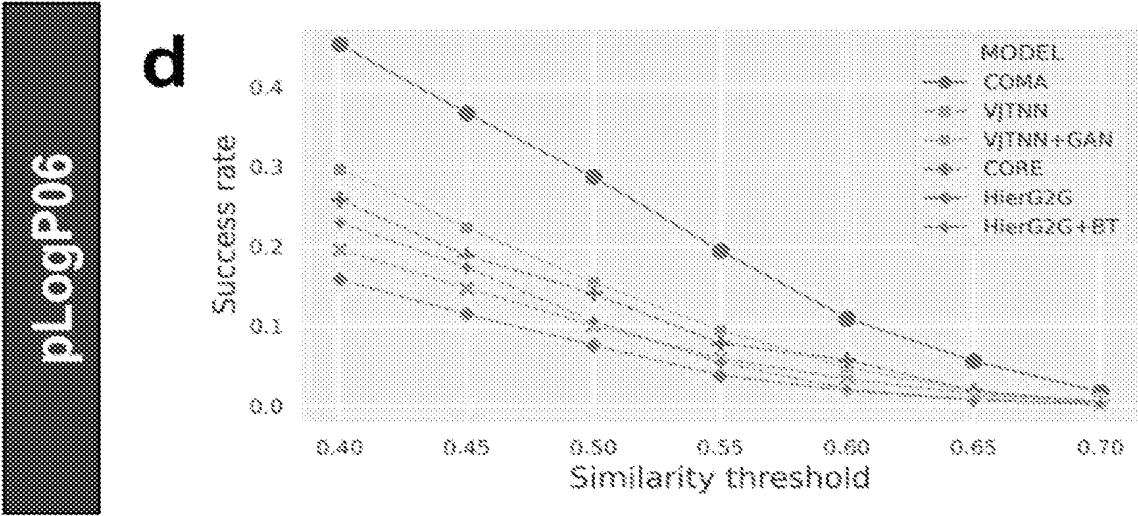


FIG. 6A

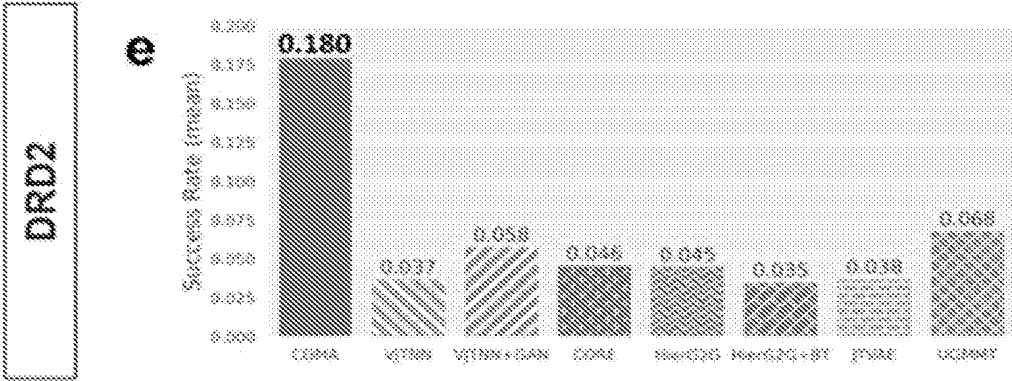


FIG. 6B

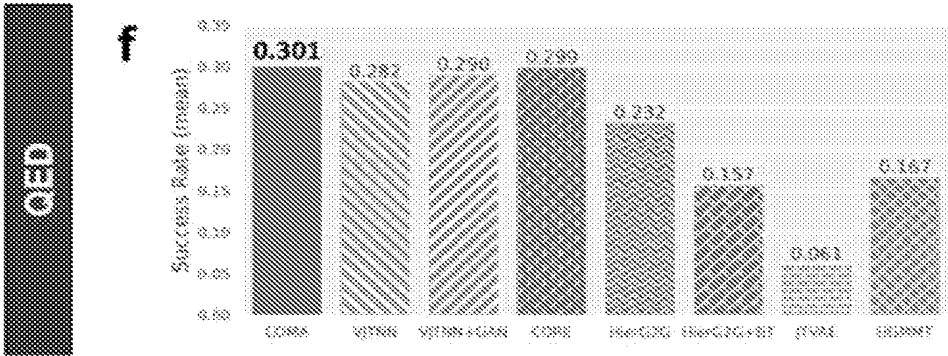


FIG. 6C

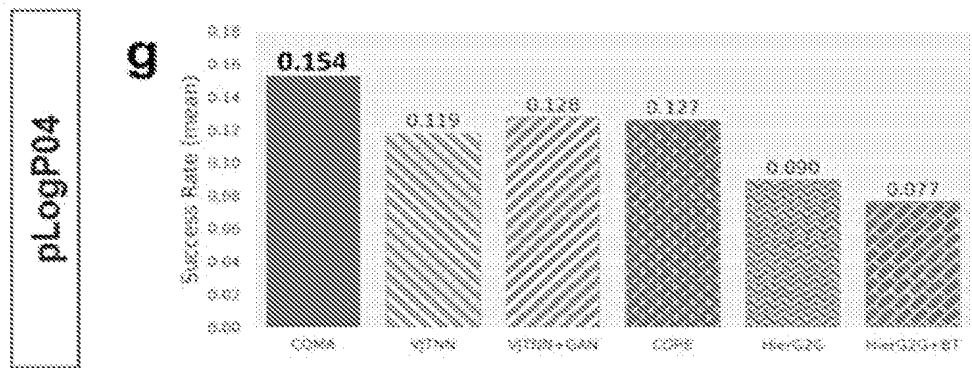


FIG. 6D

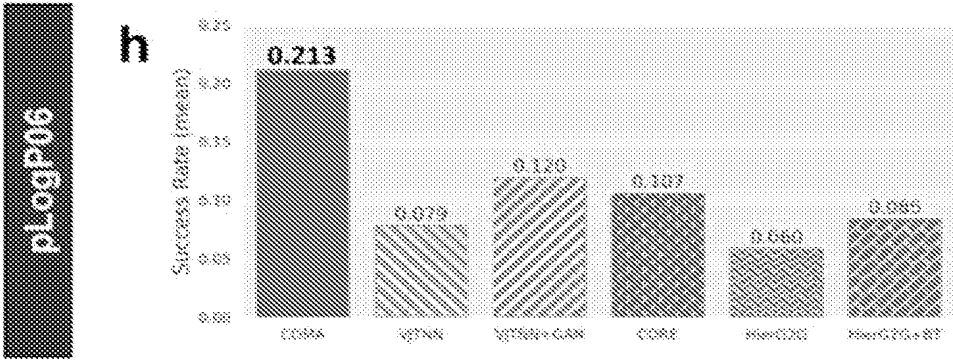


FIG. 7A

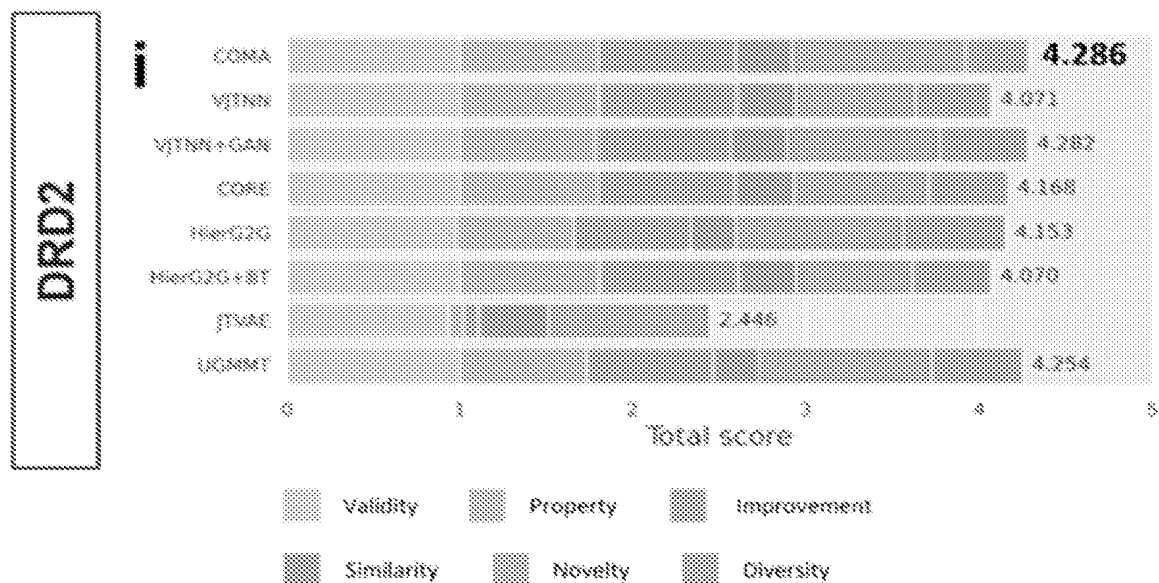


FIG. 7B

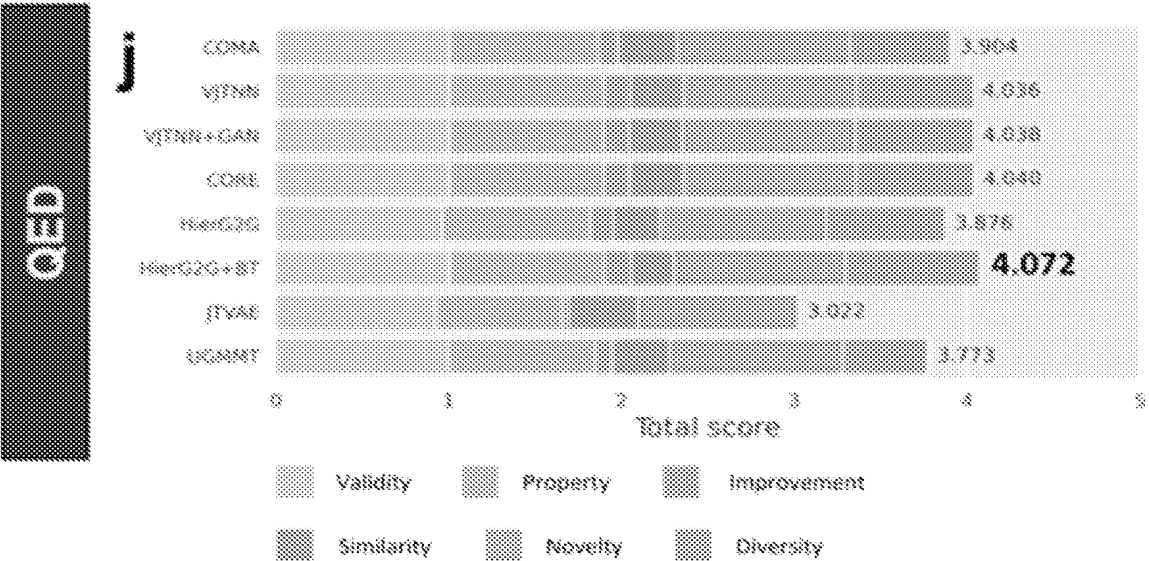


FIG. 7C

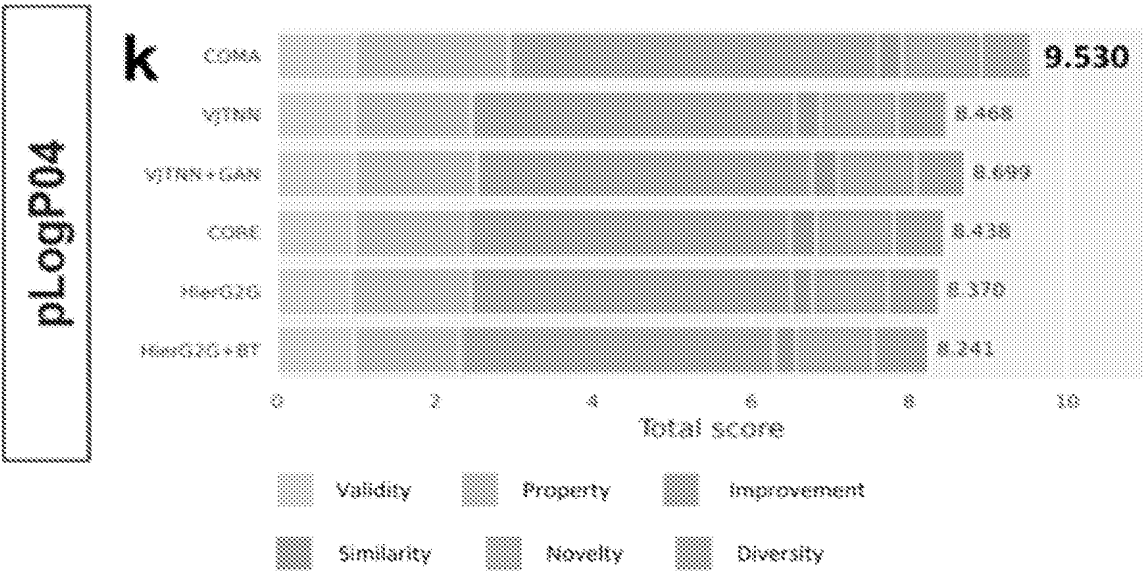


FIG. 7D

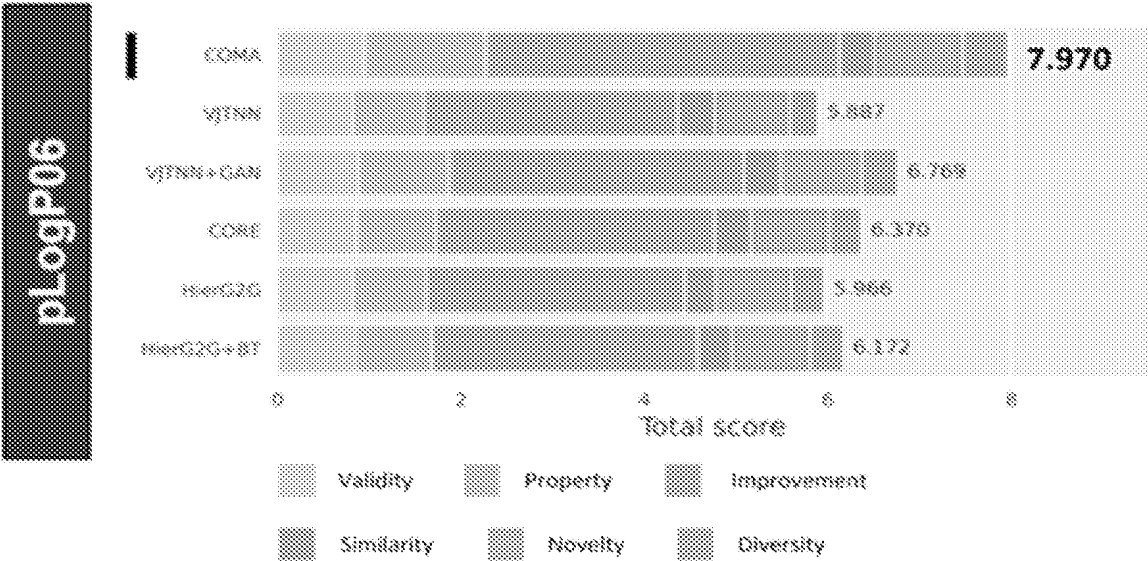


FIG. 8A

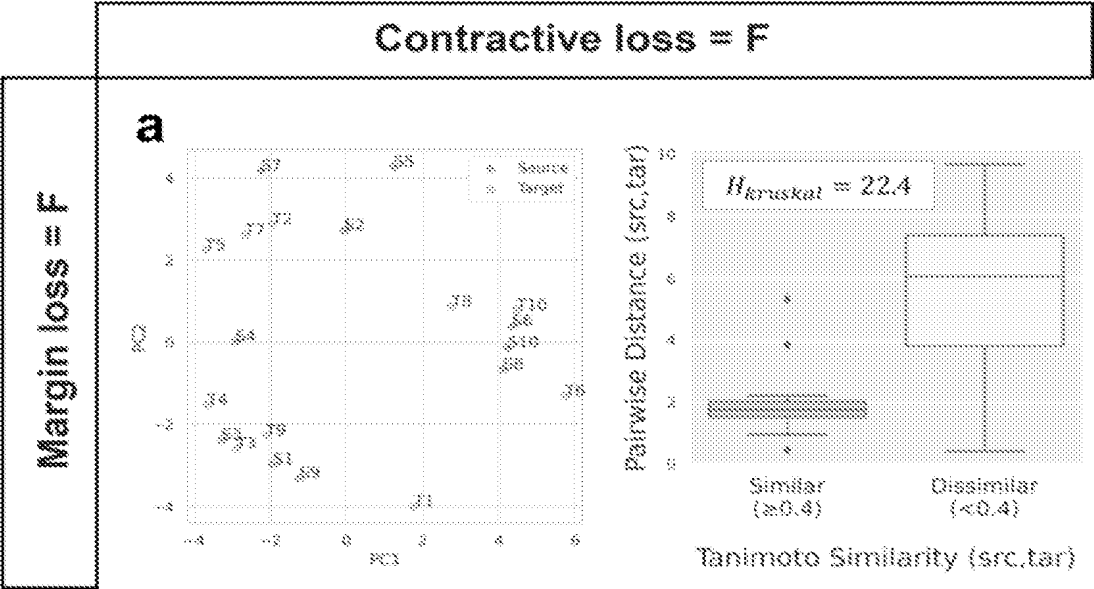


FIG. 8B

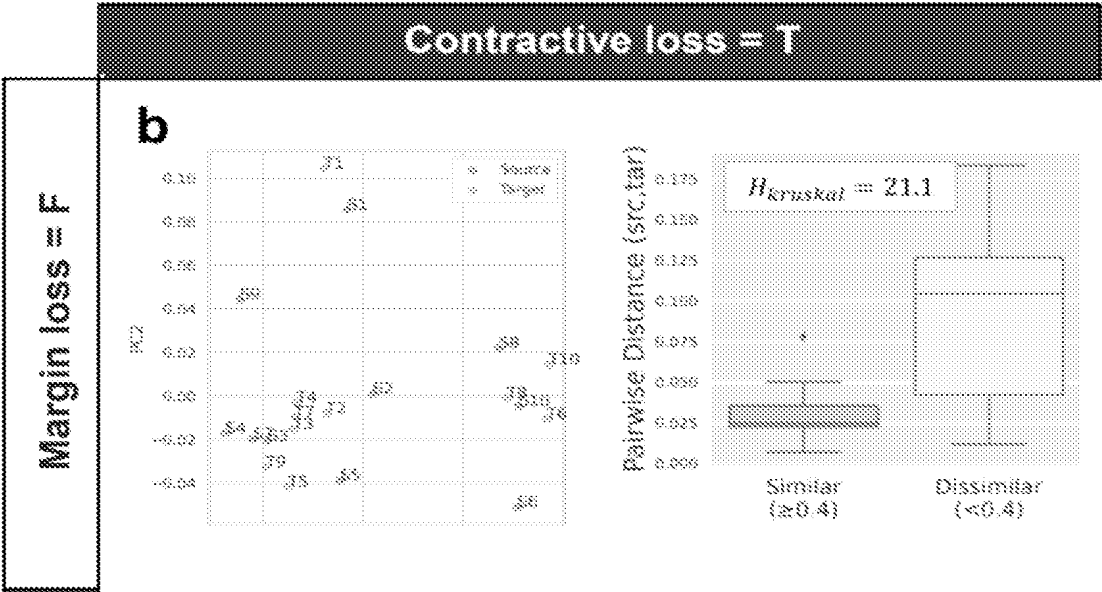


FIG. 8C

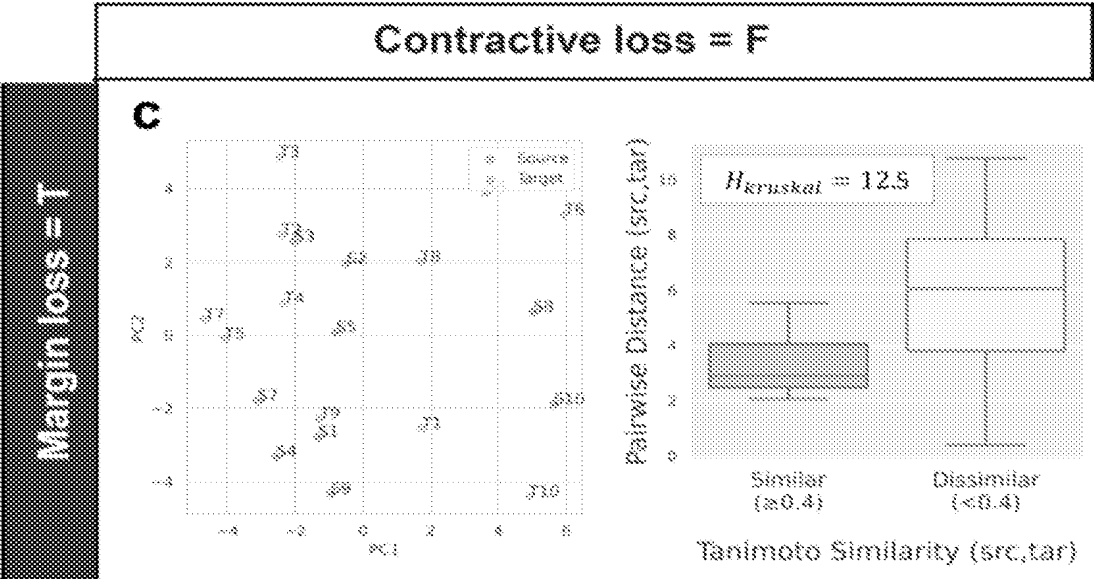


FIG. 8D

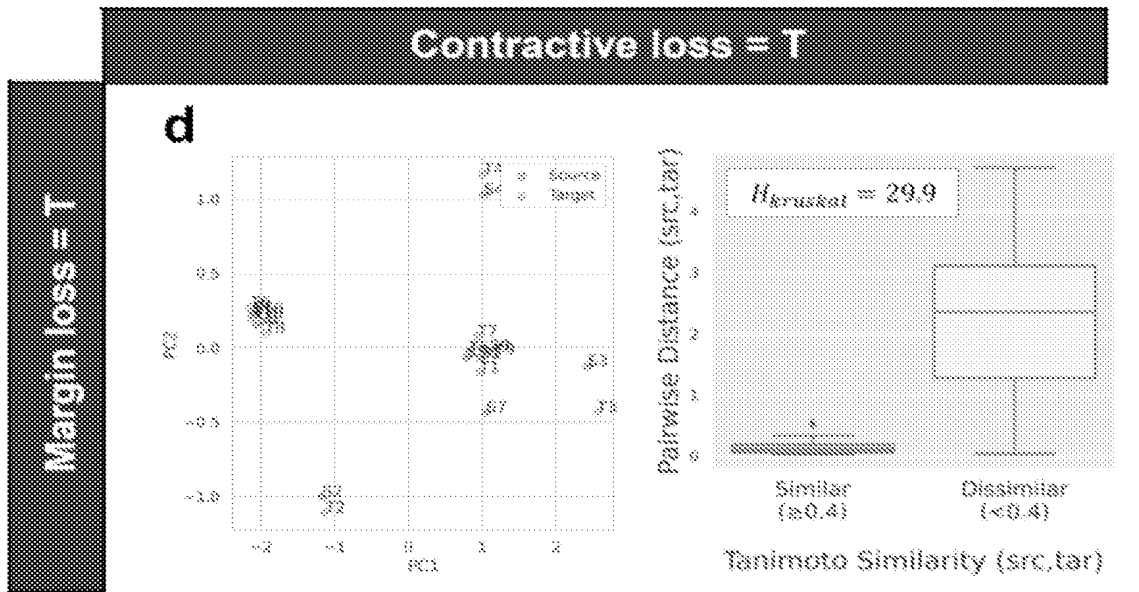


FIG. 9

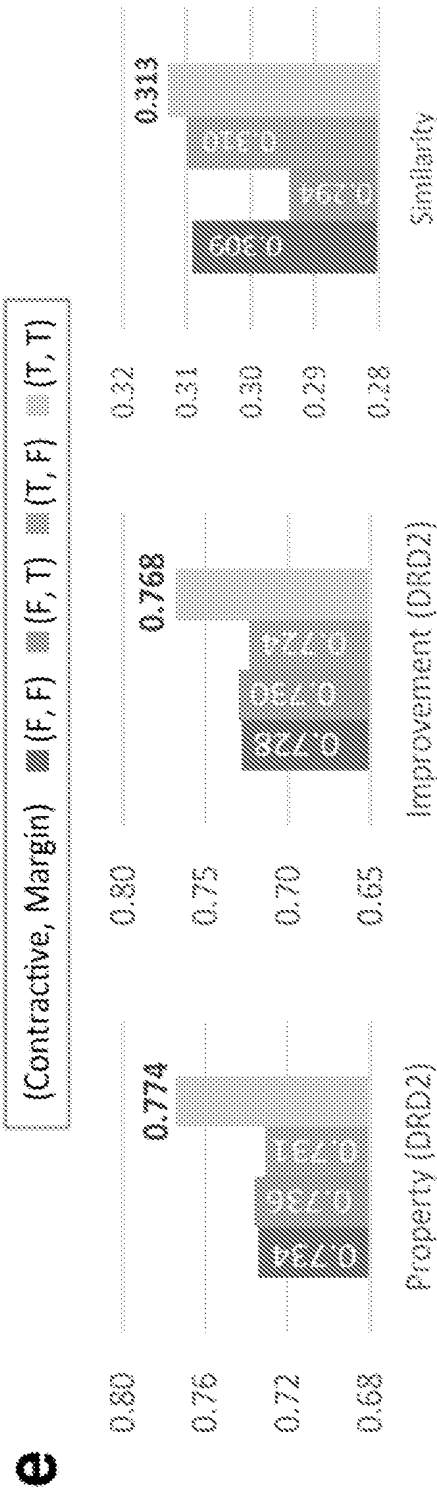


FIG. 10

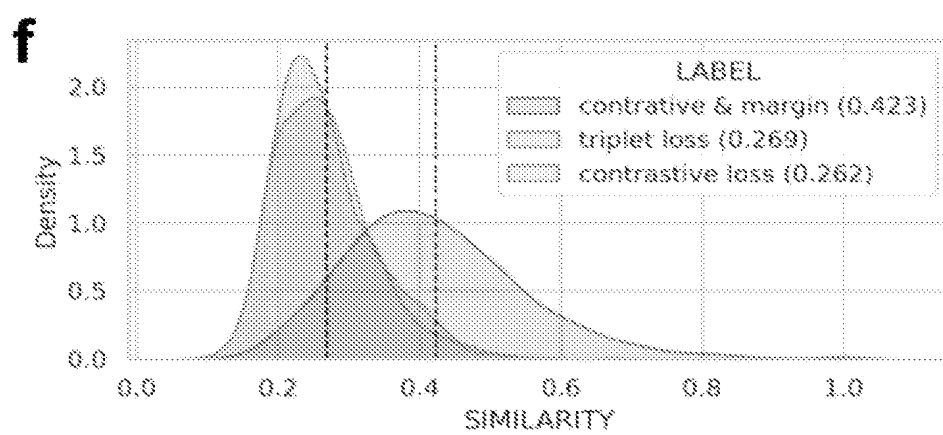


FIG.11A

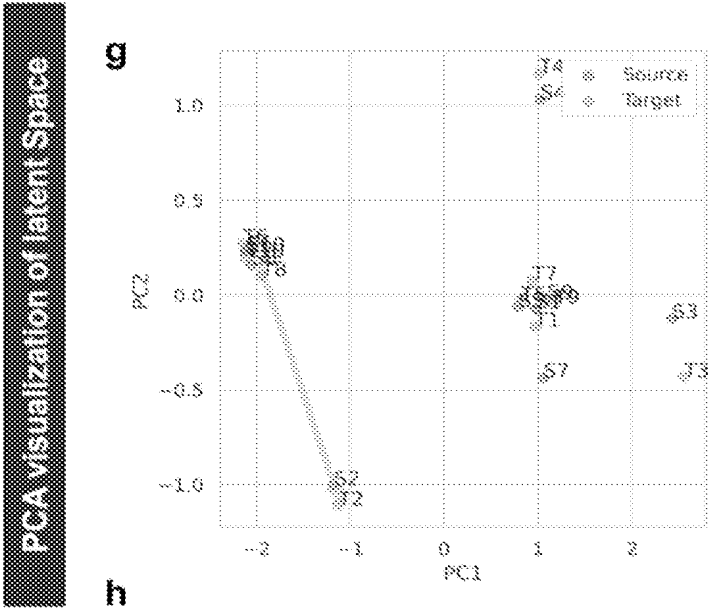


FIG. 11B

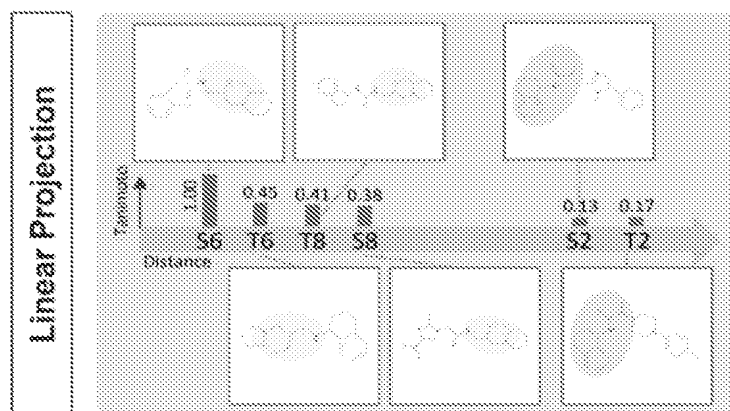


FIG. 12

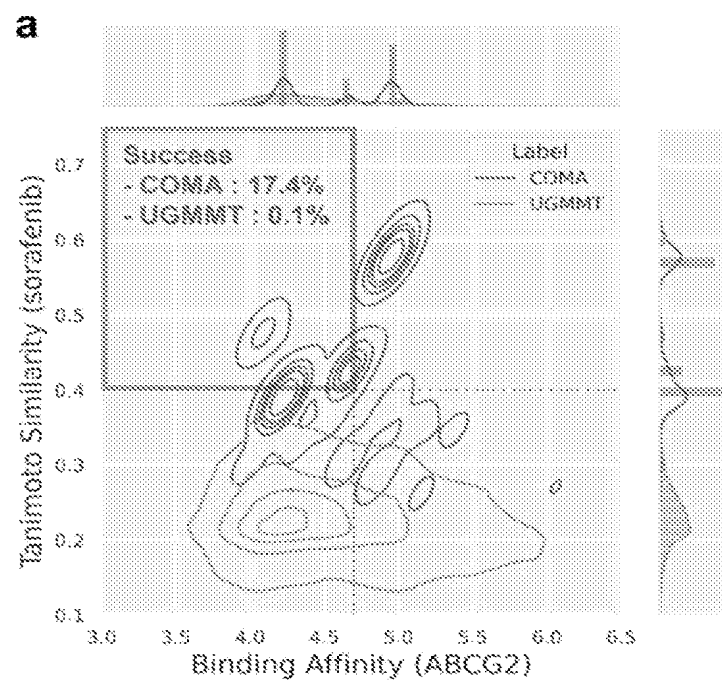


FIG. 13

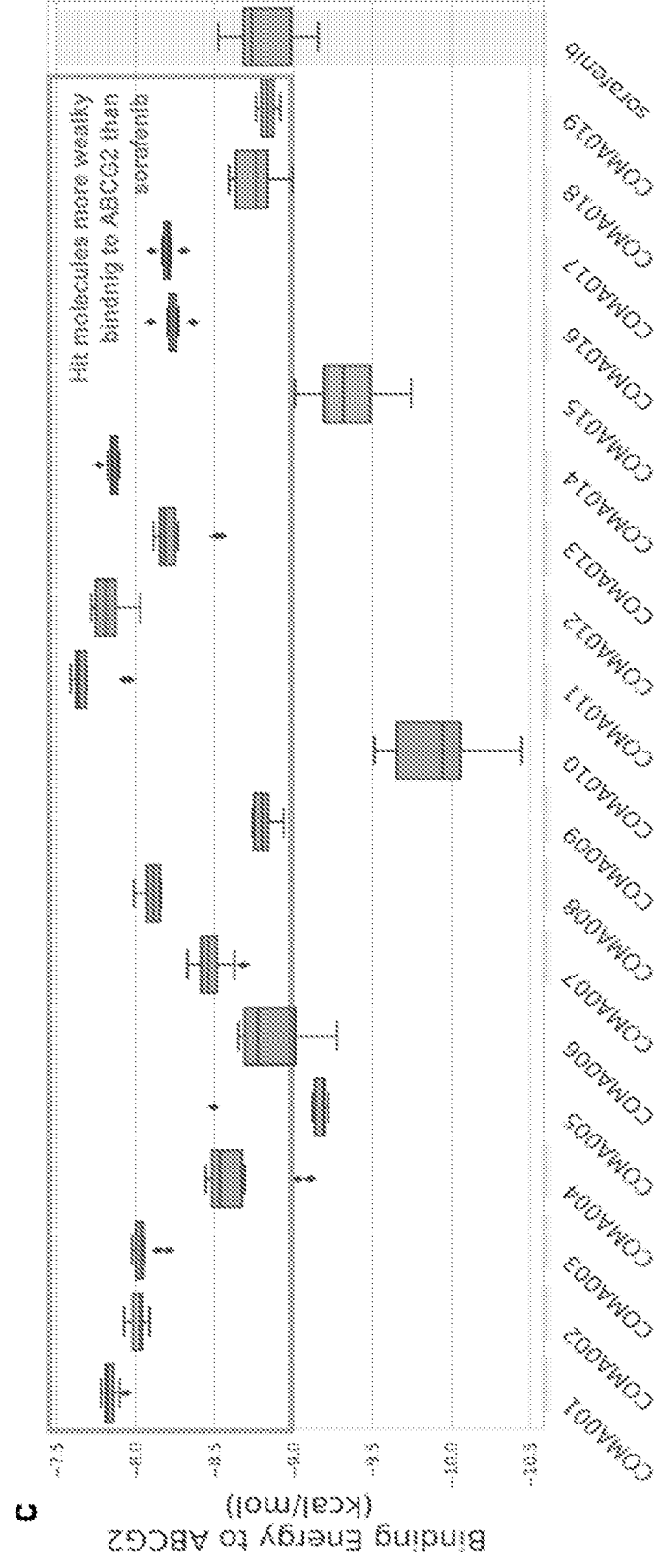


FIG. 14A

d

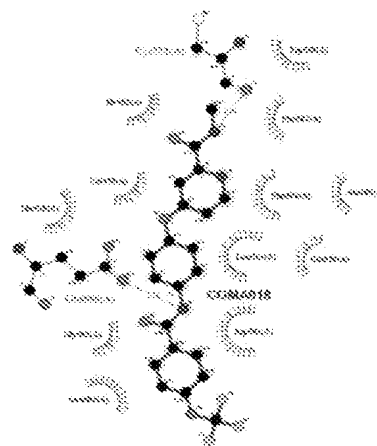
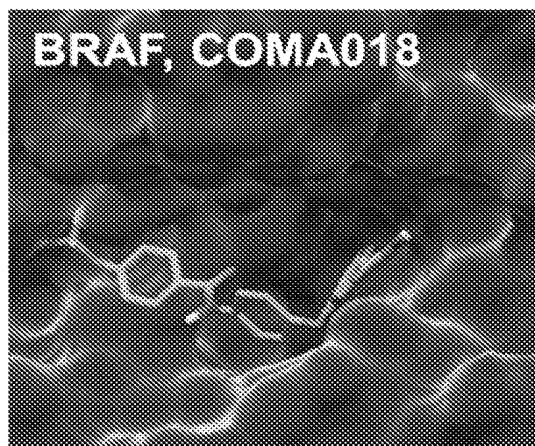


FIG. 14B

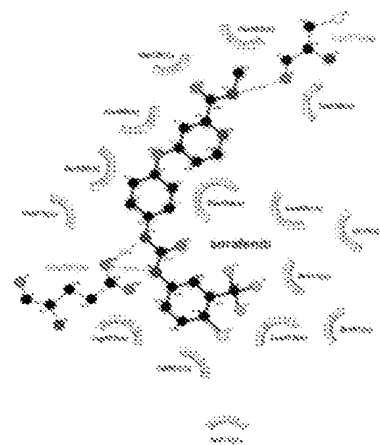
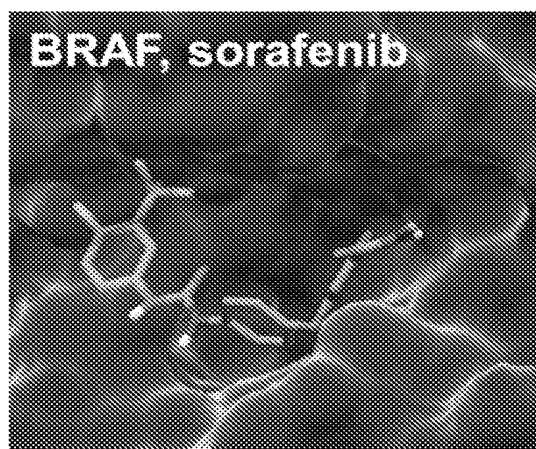


FIG. 15

b

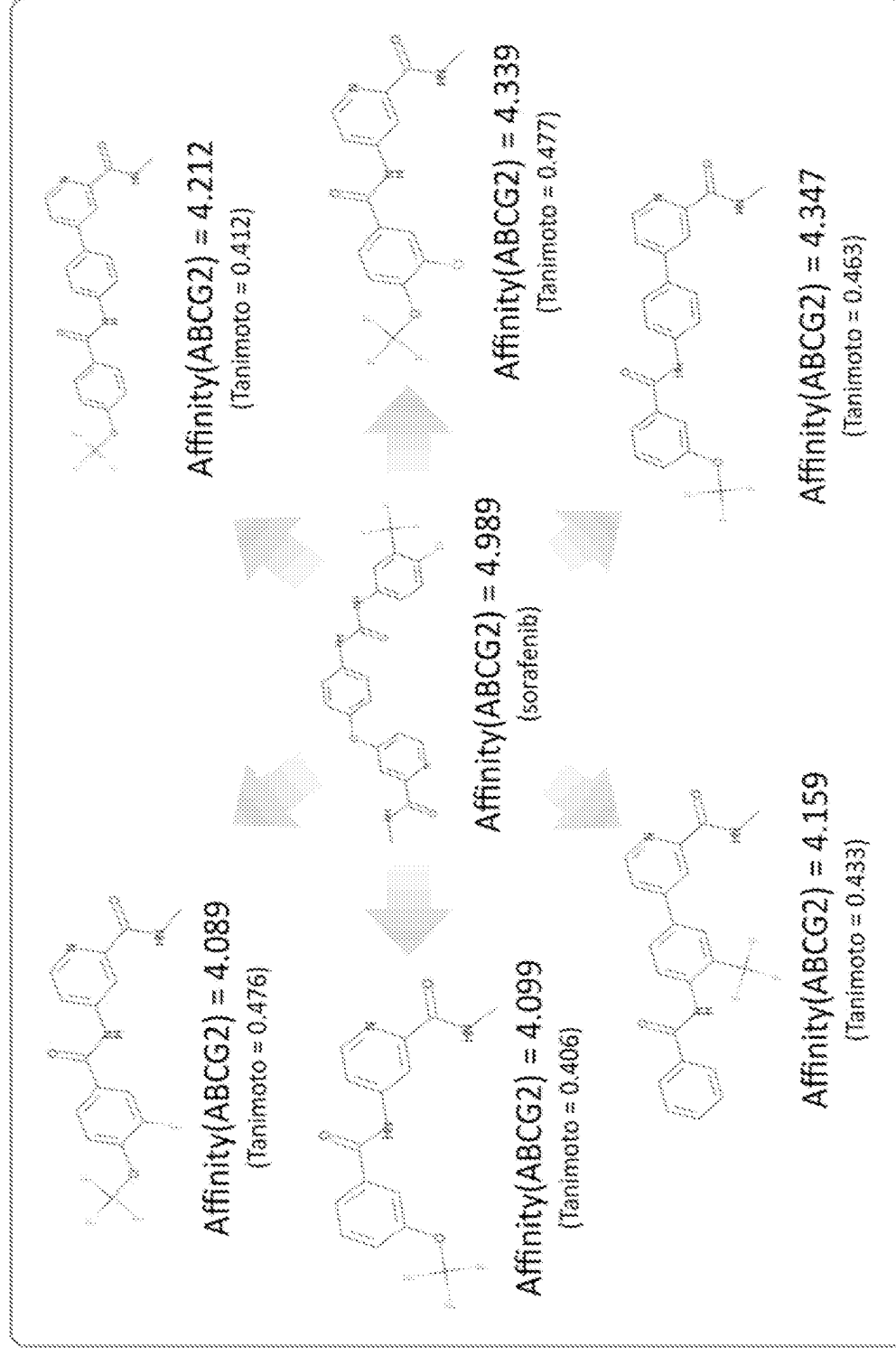
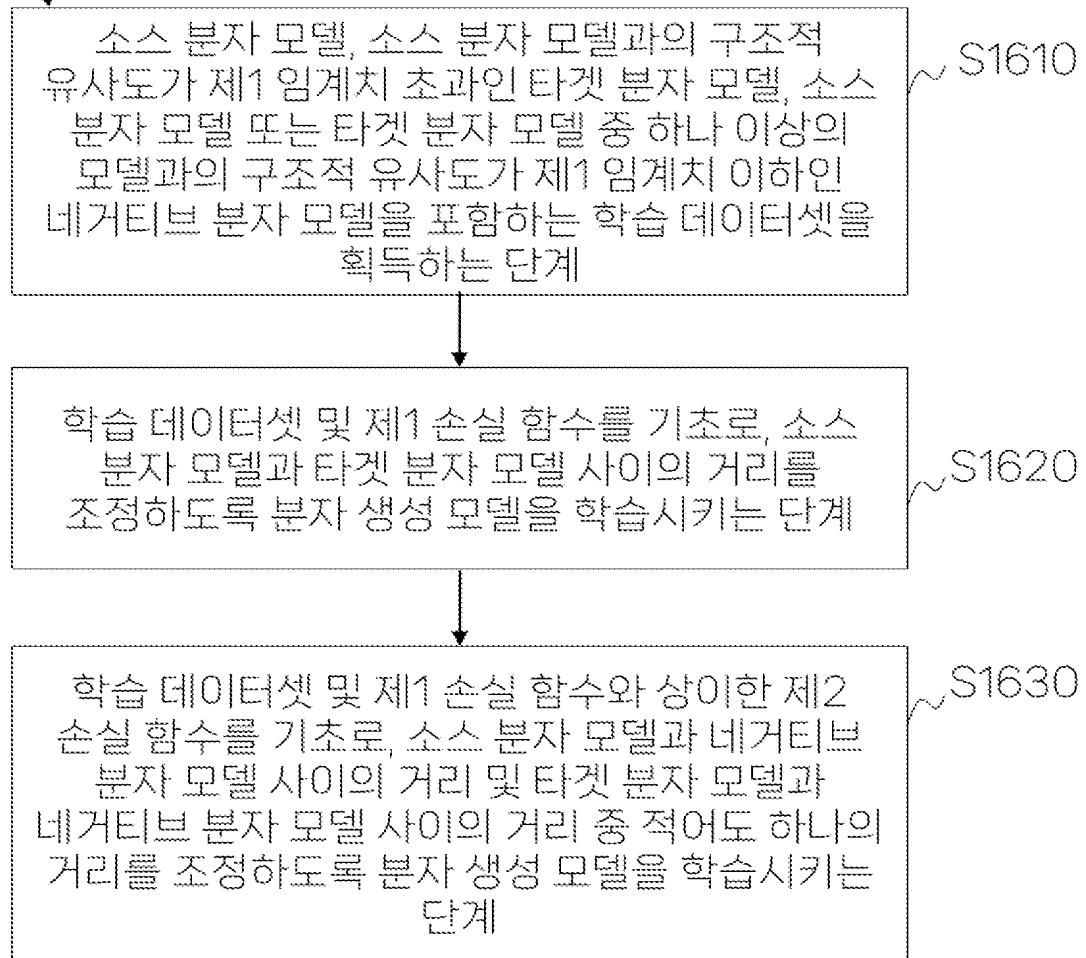


FIG. 16

1600



Under the Paperwork Reduction Act of 1995 no persons are required to respond to a collection of information unless it displays a valid OMB control number.

UTILITY PATENT APPLICATION TRANSMITTAL (Only for new nonprovisional applications under 37 CFR 1.53(b))	Attorney Docket No.	PB23-0097
	First Named Inventor	Sanghyun PARK
	Title	TRAINING METHOD AND DEVICE FOR MOLECULAR GENERATION MODEL
	Priority Mail Express® Label No.	

APPLICATION ELEMENTS <i>See MPEP chapter 600 concerning utility patent application contents.</i>	Commissioner for Patents P.O. Box 1450 Alexandria, VA 22313-1450
--	---

1. <input checked="" type="checkbox"/> Fee Transmittal Form (PTO/SB/17 or equivalent)	ACCOMPANYING APPLICATION PAPERS
2. <input checked="" type="checkbox"/> Applicant asserts small entity status. See 37 CFR 1.27	
3. <input type="checkbox"/> Applicant certifies micro entity status. See 37 CFR 1.29. Applicant must attach form PTO/SB/15A or B or equivalent.	
4. <input checked="" type="checkbox"/> Specification [Total Pages <u>39</u>] Both the claims and abstract must start on a new page. (See MPEP § 608.01(a) for information on the preferred arrangement)	
5. <input checked="" type="checkbox"/> Drawing(s) (35 U.S.C. 113) [Total Sheets <u>32</u>]	
6. Inventor's Oath or Declaration [Total Pages <u>3</u>] (including substitute statements under 37 CFR 1.64 and assignments serving as an oath or declaration under 37 CFR 1.63(e)) a. <input checked="" type="checkbox"/> Newly executed (original or copy) b. <input type="checkbox"/> A copy from a prior application (37 CFR 1.63(d))	
7. <input checked="" type="checkbox"/> Application Data Sheet * See note below. See 37 CFR 1.76 (PTO/AIA/14 or equivalent)	
8. CD-ROM or CD-R in duplicate, large table, or Computer Program (Appendix) <input type="checkbox"/> Landscape Table on CD	
9. Nucleotide and/or Amino Acid Sequence Submission (if applicable, items a. – c. are required) a. <input type="checkbox"/> Computer Readable Form (CRF) b. <input type="checkbox"/> Specification Sequence Listing on: i. <input type="checkbox"/> CD-ROM or CD-R (2 copies); or ii. <input type="checkbox"/> Paper c. <input type="checkbox"/> Statements verifying identity of above copies	
10. <input type="checkbox"/> Assignment Papers (cover sheet & document(s)) Name of Assignee _____	
11. <input type="checkbox"/> 37 CFR 3.73(c) Statement (when there is an assignee) <input checked="" type="checkbox"/> Power of Attorney	
12. <input type="checkbox"/> English Translation Document (if applicable)	
13. <input checked="" type="checkbox"/> Information Disclosure Statement (PTO/SB/08 or PTO-1449) <input checked="" type="checkbox"/> Copies of citations attached	
14. <input type="checkbox"/> Preliminary Amendment	
15. <input type="checkbox"/> Return Receipt Postcard (MPEP § 503) (Should be specifically itemized)	
16. <input type="checkbox"/> Certified Copy of Priority Document(s) (if foreign priority is claimed)	
17. <input type="checkbox"/> Nonpublication Request Under 35 U.S.C. 122(b)(2)(B)(i). Applicant must attach form PTO/SB/35 or equivalent.	
18. <input type="checkbox"/> Other: _____ _____ _____ _____	

***Note:** (1) Benefit claims under 37 CFR 1.78 and foreign priority claims under 1.55 **must** be included in an Application Data Sheet (ADS).
(2) For applications filed under 35 U.S.C. 111, the application must contain an ADS specifying the applicant if the applicant is an assignee, person to whom the inventor is under an obligation to assign, or person who otherwise shows sufficient proprietary interest in the matter. See 37 CFR 1.46(b).

19. CORRESPONDENCE ADDRESS				
<input checked="" type="checkbox"/> The address associated with Customer Number: <u>67326</u> OR <input type="checkbox"/> Correspondence address below				
Name				
Address				
City	State	Zip Code		
Country	Telephone	Email		

Signature	/Andrew S. Park/	Date	December 28, 2023
Name (Print/Type)	Andrew S. Park	Registration No. (Attorney/Agent)	47,841

A Federal agency may not conduct or sponsor, and a person is not required to respond to, nor shall a person be subject to a penalty for failure to comply with an information collection subject to the requirements of the Paperwork Reduction Act of 1995, unless the information collection has a currently valid OMB Control Number. The OMB Control Number for this information collection is 0651-0032. Public burden for this form is estimated to average 12 minutes per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the information collection. Send comments regarding this burden estimate or any other aspect of this information collection, including suggestions for reducing this burden to the Chief Administrative Officer, United States Patent and Trademark Office, P.O. Box 1450, Alexandria, VA 22313-1450 or email InformationCollection@uspto.gov. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. If filing this completed form by mail, send to: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.

If you need assistance in completing the form, call 1-800-PTO-9199 and select option 2.

Privacy Act Statement

The Privacy Act of 1974 (P.L. 93-579) requires that you be given certain information in connection with your submission of the attached form related to a patent application or patent. The United States Patent and Trademark Office (USPTO) collects the information in this record under authority of 35 U.S.C. 2. The USPTO's system of records is used to manage all applicant and owner information including name, citizenship, residence, post office address, and other information with respect to inventors and their legal representatives pertaining to the applicant's/owner's activities in connection with the invention for which a patent is sought or has been granted. The applicable Privacy Act System of Records Notice for the information collected in this form is COMMERCE/PAT-TM-7 Patent Application Files, available in the Federal Register at 78 FR 19243 (March 29, 2013). <https://www.govinfo.gov/content/pkg/FR-2013-03-29/pdf/2013-07341.pdf>

Routine uses of the information in this record may include disclosure to: 1) law enforcement, in the event that the system of records indicates a violation or potential violation of law; 2) a Federal, state, local, or international agency, in response to its request; 3) a contractor of the USPTO having need for the information in order to perform a contract; 4) the Department of Justice for determination of whether the Freedom of Information Act (FOIA) requires disclosure of the record; 5) a Member of Congress submitting a request involving an individual to whom the record pertains, when the individual has requested the Member's assistance with respect to the subject matter of the record; 6) a court, magistrate, or administrative tribunal, in the course of presenting evidence, including disclosures to opposing counsel in the course of settlement negotiations; 7) the Administrator, General Services Administration (GSA), or their designee, during an inspection of records conducted by GSA under authority of 44 U.S.C. 2904 and 2906, in accordance with the GSA regulations and any other relevant (i.e., GSA or Commerce) directive, where such disclosure shall not be used to make determinations about individuals; 8) another federal agency for purposes of National Security review (35 U.S.C. 181) and for review pursuant to the Atomic Energy Act (42 U.S.C. 218(c)); 9) the Office of Personnel Management (OPM) for personnel research purposes; and 9) the Office of Management and Budget (OMB) for legislative coordination and clearance.

If you do not furnish the information requested on this form, the USPTO may not be able to process and/or examine your submission, which may result in termination of proceedings, abandonment of the application, and/or expiration of the patent.

Additional Uses

Additional USPTO uses of the information in this record may include disclosure to: 1) the International Bureau of the World Intellectual Property Organization, if the record is related to an international application filed under the Patent Cooperation Treaty; 2) the public i) after publication of the application pursuant to 35 U.S.C. 122(b), ii) after issuance of a patent pursuant to 35 U.S.C. 151, iii) if the record was filed in an application which became abandoned or in which the proceedings were terminated and which application is referenced by either a published application, an application open to public inspections, or an issued patent, or iv) without publication of the application or patent under the specific circumstances provided for by 37 CFR 1.14(a)(1)(v)-(vii); and/or 3) the National Archives and Records Administration, for inspection of records.

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)	Application Number			
	Filing Date			
	First Named Inventor	Sanghyun PARK		
	Art Unit			
	Examiner Name			
	Attorney Docket Number	PB23-0097		

U.S.PATENTS							Remove	
Examiner Initial*	Cite No	Patent Number	Kind Code ¹	Issue Date	Name of Patentee or Applicant of cited Document	Pages,Columns,Lines where Relevant Passages or Relevant Figures Appear		
	1							
If you wish to add additional U.S. Patent citation information please click the Add button.							Add	
U.S.PATENT APPLICATION PUBLICATIONS							Remove	
Examiner Initial*	Cite No	Publication Number	Kind Code ¹	Publication Date	Name of Patentee or Applicant of cited Document	Pages,Columns,Lines where Relevant Passages or Relevant Figures Appear		
	1	20220318596	A1	2022-10-06	Mohammad Reza SARSHOGH, et al.			
	2	20220238191	A1	2022-07-28	Jie FAN, et al.			
	3	20220262494	A1	2022-08-18	Jim ABRAHAM, et al.			
If you wish to add additional U.S. Published Application citation information please click the Add button.							Add	
FOREIGN PATENT DOCUMENTS							Remove	
Examiner Initial*	Cite No	Foreign Document Number ³	Country Code ²	Kind Code ⁴	Publication Date	Name of Patentee or Applicant of cited Document	Pages,Columns,Lines where Relevant Passages or Relevant Figures Appear	T ⁵
	1							<input type="checkbox"/>

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)	Application Number		
	Filing Date		
	First Named Inventor	Sanghyun PARK	
	Art Unit		
	Examiner Name		
	Attorney Docket Number	PB23-0097	

If you wish to add additional Foreign Patent Document citation information please click the Add button

NON-PATENT LITERATURE DOCUMENTS				<input type="button" value="Remove"/>
--	--	--	--	---------------------------------------

Examiner Initials*	Cite No	Include name of the author (in CAPITAL LETTERS), title of the article (when appropriate), title of the item (book, magazine, journal, serial, symposium, catalog, etc), date, pages(s), volume-issue number(s), publisher, city and/or country where published.	T ⁵
	1	JONGHWAN CHOI, ET AL., "COMA: Efficient Structure-constrained Molecular Generation using Contractive and Margin Losses", Journal of Cheminformatics 15, Article number: 8 (2023), January 19, 2023, 56pages.	<input type="checkbox"/>
	2	GABRIEL A. PINHEIRO, ET AL., "SMICLR : Contrastive Learning on Multiple Molecular Representations for Semisupervised and Unsupervised Representation Learning", Journal of Chemical Information and Modeling, August 31, 2022, 14pages.	<input type="checkbox"/>

If you wish to add additional non-patent literature document citation information please click the Add button

EXAMINER SIGNATURE			
---------------------------	--	--	--

Examiner Signature		Date Considered	
--------------------	--	-----------------	--

*EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609. Draw line through a citation if not in conformance and not considered. Include copy of this form with next communication to applicant.

¹ See Kind Codes of USPTO Patent Documents at www.USPTO.GOV or MPEP 901.04. ² Enter office that issued the document, by the two-letter code (WIPO Standard ST.3). ³ For Japanese patent documents, the indication of the year of the reign of the Emperor must precede the serial number of the patent document. ⁴ Kind of document by the appropriate symbols as indicated on the document under WIPO Standard ST.16 if possible. ⁵ Applicant is to place a check mark here if English language translation is attached.

**INFORMATION DISCLOSURE
STATEMENT BY APPLICANT**
(Not for submission under 37 CFR 1.99)

Application Number	
Filing Date	
First Named Inventor	Sanghyun PARK
Art Unit	
Examiner Name	
Attorney Docket Number	PB23-0097

CERTIFICATION STATEMENT

Please see 37 CFR 1.97 and 1.98 to make the appropriate selection(s):

☐ That each item of information contained in the information disclosure statement was first cited in any communication from a foreign patent office in a counterpart foreign application not more than three months prior to the filing of the information disclosure statement. See 37 CFR 1.97(e)(1).

OR

☐ That no item of information contained in the information disclosure statement was cited in a communication from a foreign patent office in a counterpart foreign application, and, to the knowledge of the person signing the certification after making reasonable inquiry, no item of information contained in the information disclosure statement was known to any individual designated in 37 CFR 1.56(c) more than three months prior to the filing of the information disclosure statement. See 37 CFR 1.97(e)(2).

☐ See attached certification statement.

☐ The fee set forth in 37 CFR 1.17 (p) has been submitted herewith.

☒ A certification statement is not submitted herewith.

SIGNATURE

A signature of the applicant or representative is required in accordance with CFR 1.33, 10.18. Please see CFR 1.4(d) for the form of the signature.

Signature	/Andrew S. Park/	Date (YYYY-MM-DD)	2023-12-28
Name/Print	Andrew S. Park	Registration Number	47841

This collection of information is required by 37 CFR 1.97 and 1.98. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 1 hour to complete, including gathering, preparing and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1450, Alexandria, VA 22313-1450. **DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**

Privacy Act Statement

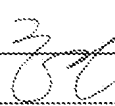
The Privacy Act of 1974 (P.L. 93-579) requires that you be given certain information in connection with your submission of the attached form related to a patent application or patent. Accordingly, pursuant to the requirements of the Act, please be advised that: (1) the general authority for the collection of this information is 35 U.S.C. 2(b)(2); (2) furnishing of the information solicited is voluntary; and (3) the principal purpose for which the information is used by the U.S. Patent and Trademark Office is to process and/or examine your submission related to a patent application or patent. If you do not furnish the requested information, the U.S. Patent and Trademark Office may not be able to process and/or examine your submission, which may result in termination of proceedings or abandonment of the application or expiration of the patent.

The information provided by you in this form will be subject to the following routine uses:

1. The information on this form will be treated confidentially to the extent allowed under the Freedom of Information Act (5 U.S.C. 552) and the Privacy Act (5 U.S.C. 552a). Records from this system of records may be disclosed to the Department of Justice to determine whether the Freedom of Information Act requires disclosure of these records.
2. A record from this system of records may be disclosed, as a routine use, in the course of presenting evidence to a court, magistrate, or administrative tribunal, including disclosures to opposing counsel in the course of settlement negotiations.
3. A record in this system of records may be disclosed, as a routine use, to a Member of Congress submitting a request involving an individual, to whom the record pertains, when the individual has requested assistance from the Member with respect to the subject matter of the record.
4. A record in this system of records may be disclosed, as a routine use, to a contractor of the Agency having need for the information in order to perform a contract. Recipients of information shall be required to comply with the requirements of the Privacy Act of 1974, as amended, pursuant to 5 U.S.C. 552a(m).
5. A record related to an International Application filed under the Patent Cooperation Treaty in this system of records may be disclosed, as a routine use, to the International Bureau of the World Intellectual Property Organization, pursuant to the Patent Cooperation Treaty.
6. A record in this system of records may be disclosed, as a routine use, to another federal agency for purposes of National Security review (35 U.S.C. 181) and for review pursuant to the Atomic Energy Act (42 U.S.C. 218(c)).
7. A record from this system of records may be disclosed, as a routine use, to the Administrator, General Services, or his/her designee, during an inspection of records conducted by GSA as part of that agency's responsibility to recommend improvements in records management practices and programs, under authority of 44 U.S.C. 2904 and 2906. Such disclosure shall be made in accordance with the GSA regulations governing inspection of records for this purpose, and any other relevant (i.e., GSA or Commerce) directive. Such disclosure shall not be used to make determinations about individuals.
8. A record from this system of records may be disclosed, as a routine use, to the public after either publication of the application pursuant to 35 U.S.C. 122(b) or issuance of a patent pursuant to 35 U.S.C. 151. Further, a record may be disclosed, subject to the limitations of 37 CFR 1.14, as a routine use, to the public if the record was filed in an application which became abandoned or in which the proceedings were terminated and which application is referenced by either a published application, an application open to public inspections or an issued patent.
9. A record from this system of records may be disclosed, as a routine use, to a Federal, State, or local law enforcement agency, if the USPTO becomes aware of a violation or potential violation of law or regulation.

Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

**DECLARATION (37 CFR 1.63) FOR UTILITY OR DESIGN APPLICATION USING AN
APPLICATION DATA SHEET (37 CFR 1.76)**

Title of Invention	TRAINING METHOD AND DEVICE FOR MOLECULAR GENERATION MODEL
<p>As the below named inventor, I hereby declare that:</p> <p>This declaration is directed to: <input checked="" type="checkbox"/> The attached application, or <input type="checkbox"/> United States application or PCT international application number _____ filed on _____</p> <p>The above-identified application was made or authorized to be made by me.</p> <p>I believe that I am the original inventor or an original joint inventor of a claimed invention in the application.</p> <p>I hereby acknowledge that any willful false statement made in this declaration is punishable under 18 U.S.C. 1001 by fine or imprisonment of not more than five (5) years, or both.</p> <p style="text-align: center;">WARNING:</p> <p>Petitioner/applicant is cautioned to avoid submitting personal information in documents filed in a patent application that may contribute to identity theft. Personal information such as social security numbers, bank account numbers, or credit card numbers (other than a check or credit card authorization form PTO-2038 submitted for payment purposes) is never required by the USPTO to support a petition or an application. If this type of personal information is included in documents submitted to the USPTO, petitioners/applicants should consider redacting such personal information from the documents before submitting them to the USPTO. Petitioner/applicant is advised that the record of a patent application is available to the public after publication of the application (unless a non-publication request in compliance with 37 CFR 1.213(a) is made in the application) or issuance of a patent. Furthermore, the record from an abandoned application may also be available to the public if the application is referenced in a published application or an issued patent (see 37 CFR 1.14). Checks and credit card authorization forms PTO-2038 submitted for payment purposes are not retained in the application file and therefore are not publicly available.</p>	
LEGAL NAME OF INVENTOR	
Inventor: <u>Sanghyun PARK</u>  Date (Optional): <u>2023.12.28</u>	
Signature: _____	
<p>Note. An application data sheet (PTO/SB/14 or equivalent), including naming the entire inventive entity, must accompany this form or must have been previously filed. Use an additional PTO/AIA/01 form for each additional inventor.</p>	

This collection of information is required by 35 U.S.C. 115 and 37 CFR 1.63. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.11 and 1.14. This collection is estimated to take 1 minute to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.

If you need assistance in completing the form, call 1-800-PTO-9199 and select option 2.

Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

**DECLARATION (37 CFR 1.63) FOR UTILITY OR DESIGN APPLICATION USING AN
APPLICATION DATA SHEET (37 CFR 1.76)**

Title of
Invention

TRAINING METHOD AND DEVICE FOR MOLECULAR GENERATION MODEL

As the below named inventor, I hereby declare that:

This declaration
is directed to:



The attached application, or



United States application or PCT international application number _____
filed on _____.

The above-identified application was made or authorized to be made by me.

I believe that I am the original inventor or an original joint inventor of a claimed invention in the application.

I hereby acknowledge that any willful false statement made in this declaration is punishable under 18 U.S.C. 1001
by fine or imprisonment of not more than five (5) years, or both.

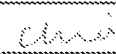
WARNING:

Petitioner/applicant is cautioned to avoid submitting personal information in documents filed in a patent application that may contribute to identity theft. Personal information such as social security numbers, bank account numbers, or credit card numbers (other than a check or credit card authorization form PTO-2038 submitted for payment purposes) is never required by the USPTO to support a petition or an application. If this type of personal information is included in documents submitted to the USPTO, petitioners/applicants should consider redacting such personal information from the documents before submitting them to the USPTO. Petitioner/applicant is advised that the record of a patent application is available to the public after publication of the application (unless a non-publication request in compliance with 37 CFR 1.213(a) is made in the application) or issuance of a patent. Furthermore, the record from an abandoned application may also be available to the public if the application is referenced in a published application or an issued patent (see 37 CFR 1.14). Checks and credit card authorization forms PTO-2038 submitted for payment purposes are not retained in the application file and therefore are not publicly available.

LEGAL NAME OF INVENTOR

Inventor: Jonghwan CHOI

Date (Optional): 2022.12.28

Signature: 

Note: An application data sheet (PTO/SB/14 or equivalent), including naming the entire inventive entity, must accompany this form or must have been previously filed. Use an additional PTO/AIA/01 form for each additional inventor.

This collection of information is required by 35 U.S.C. 116 and 37 CFR 1.63. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.11 and 1.14. This collection is estimated to take 1 minute to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.

If you need assistance in completing the form, call 1-800-PTO-9199 and select option 2.

Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

**DECLARATION (37 CFR 1.63) FOR UTILITY OR DESIGN APPLICATION USING AN
APPLICATION DATA SHEET (37 CFR 1.76)**

Title of
Invention

TRAINING METHOD AND DEVICE FOR MOLECULAR GENERATION MODEL

As the below named inventor, I hereby declare that:

This declaration
is directed to:



The attached application, or



United States application or PCT international application number _____
filed on _____.

The above-identified application was made or authorized to be made by me.

I believe that I am the original inventor or an original joint inventor of a claimed invention in the application.

I hereby acknowledge that any willful false statement made in this declaration is punishable under 18 U.S.C. 1001
by fine or imprisonment of not more than five (5) years, or both.

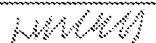
WARNING:

Petitioner/applicant is cautioned to avoid submitting personal information in documents filed in a patent application that may contribute to identity theft. Personal information such as social security numbers, bank account numbers, or credit card numbers (other than a check or credit card authorization form PTO-2038 submitted for payment purposes) is never required by the USPTO to support a petition or an application. If this type of personal information is included in documents submitted to the USPTO, petitioners/applicants should consider redacting such personal information from the documents before submitting them to the USPTO. Petitioner/applicant is advised that the record of a patent application is available to the public after publication of the application (unless a non-publication request in compliance with 37 CFR 1.213(a) is made in the application) or issuance of a patent. Furthermore, the record from an abandoned application may also be available to the public if the application is referenced in a published application or an issued patent (see 37 CFR 1.14). Checks and credit card authorization forms PTO-2038 submitted for payment purposes are not retained in the application file and therefore are not publicly available.

LEGAL NAME OF INVENTOR

Inventor: Sangmin SEO

Date (Optional): 23.11.23

Signature: 

Note: An application data sheet (PTO/SB/14 or equivalent), including naming the entire inventive entity, must accompany this form or must have been previously filed. Use an additional PTO/AIA/01 form for each additional inventor.

This collection of information is required by 35 U.S.C. 115 and 37 CFR 1.63. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.11 and 1.14. This collection is estimated to take 1 minute to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1480, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.

If you need assistance in completing the form, call 1-800-PTO-9199 and select option 2



ELECTRONIC ACKNOWLEDGEMENT RECEIPT

APPLICATION #
18/399,450

RECEIPT DATE / TIME
12/28/2023 07:08:46 PM Z ET

ATTORNEY DOCKET #
PB23-0097

Title of Invention

TRAINING METHOD AND DEVICE FOR MOLECULAR GENERATION MODEL

Application Information

APPLICATION TYPE Utility - Nonprovisional Application
under 35 USC 111(a)

PATENT # -

CONFIRMATION # 4354

FILED BY SY MOON

PATENT CENTER # 63769961

FILING DATE -

CUSTOMER # 67326

FIRST NAMED
INVENTOR Sanghyun PARK

CORRESPONDENCE
ADDRESS -

AUTHORIZED BY Andrew Park

Documents

TOTAL DOCUMENTS: 11

DOCUMENT	PAGES	DESCRIPTION	SIZE (KB)
PB23-0097_ADS.pdf	9	Application Data Sheet	2174 KB
PB23-0097_POA.pdf	2	Power of Attorney	263 KB
PB23-0097_DEC.pdf	3	Oath or Declaration filed	1446 KB
PB23-0097_SPEC_kr.pdf	39	-	1012 KB
PB23-0097_SPEC_kr- SPECNO.pdf	(1-32) 32	Specification - Not in English	743 KB
PB23-0097_SPEC_kr- CLM.pdf	(33-38) 6	Claims	465 KB

PB23-0097__SPEC__kr-ABST.pdf	(39-39)	1	Abstract	443 KB
PB23-0097__Drawings__kr.pdf		32	Drawings-only black and white line drawings	3536 KB
PB23-0097__aia0015.pdf		2	Transmittal of New Application	447 KB
PB23-0097__SB08.pdf		4	Information Disclosure Statement (IDS) Form (SB08)	1234 KB
1__JonghwanCHOI.pdf		56	Non Patent Literature	3576 KB
2__Gabriel.pdf		14	Non Patent Literature	2817 KB

Digest

DOCUMENT	MESSAGE DIGEST(SHA-512)
PB23-0097__ADS.pdf	3DE5EBE6A120F48A9500E9AB61B913DEEB37A6E5154427C60D28F36CD5D18E94D23F42EA590ED000005EA157C0A81C5A015F005BB9CF9B2124987DA8B54F58B7
PB23-0097__POA.pdf	F7D3ADB6F6C422A76BD81A00EEC213AFC4E5BF7599F20486F2B1E5B0478186BFBF81294AECCDD62519613D52A22793D2339D78F40BA546218CA28D90582ECA50
PB23-0097__DEC.pdf	B7BC1FA36D654CB58067D6A3190048E65BCB7F21F25F65877099FA9A1F799A4A8842E785D2ED1188896850A2F12B0394CDFBA2F687125EE9F755439CB1B91E6D
PB23-0097__SPEC__kr.pdf	04434A461FC21DE937994679F80E31BBB6D7C8E177F8C330F02F031C71CF0EA4F5926F9E30B6A5734947B068BF87A3F78146FCC2C76C2D19CFD36C3F49218221
PB23-0097__SPEC__kr-SPECNO.pdf	BCA322ABF46C4A668DD22DC30D1CD96DEAEBE2E78CBD33B5EB218B0A719041749235DF946CB415E9AE9875E472C9AC320E66789730AD9E6D8E7B16C9AB4AD15E

PB23-0097__SPEC__kr-CLM.pdf	24B5772744C05B32CB89FF54565AB41125848302E94F7179FC 8B90AEC916F4FFFB7B23F167E32DBD92A54AF73725D0B384E 7F7E0057C2B7C0366F6295AFB4880
PB23-0097__SPEC__kr-ABST.pdf	E266B36316A2A0C49365F08E010CB1E0D53F93B19F6087B517 00950375A30D2F71DE7C5110BA38C6DB582C0508F6E2706167 1DFD6B633F2ADFEF6EDE1B21171B
PB23-0097__Drawings__kr.pdf	511BC30A9C8C66BCBC7D4A9A0E6807EA7E9DF9D015F12C82 EE215DE07E41A75412779851ECD6C41F453077CEBCE8716C9 47803C2D708DC95354C3CF83DD8A25E
PB23-0097__ala0015.pdf	9D23508B20CA05771250DE2564DEC24D5251F0800567361480 FC617090AC5BAB0DC82C57C50E0745DEF4F8AEF805FD3E96 66BDE6E09E18F20482AB4AD96A9206
PB23-0097__SB08.pdf	88FD9D912C32994C3584E13A4F6553E5CC2F2478A42D77CDB D3512A44D309C6B084C26E347A990FB25AE134BFA083889D8 05445CE252CB2EF6A9BCDF67837775
1__JonghwanCHOI.pdf	FE47B9E334C21ECC967832D143CB01BD7726352C04447A65A 85103BF31DE77DC22D79C4A288B49AACA1012F513A58E13D8 3D56568CB35AF2F929F4E97C19E921
2__Gabriel.pdf	307913F58B5A75C819D900B7DA9DE5359EC507FE86F5BC595 ECBE2F2BC6C1F0FA5CF5B2CAFFFFF2122BDCB08F2F362DCD C044B573DF7774C30574C9DA9588BB12

This Acknowledgement Receipt evidences receipt on the noted date by the USPTO of the indicated documents, characterized by the applicant, and including page counts, where applicable. It serves as evidence of receipt similar to a Post Card, as described in MPEP 503.

New Applications Under 35 U.S.C. 111

If a new application is being filed and the application includes the necessary components for filing date (see 37 CFR 1.53(b)-(d) and MPEP 506), a Filing Receipt (37 CFR 1.54) will be issued in due course and the date shown on this Acknowledgement Receipt will establish the filing date of the application

National Stage of an International Application under 35 U.S.C. 371

If a timely submission to enter the national stage of an international application is compliant with the conditions of 35 U.S.C. 371 and other applicable requirements a Form PCT/DO/EO/903 indicating acceptance of the application as a national stage submission under 35 U.S.C. 371 will be issued in addition to the Filing Receipt, in due course.

New International Application Filed with the USPTO as a Receiving Office

If a new international application is being filed and the international application includes the necessary components for an

international filing date (see PCT Article 11 and MPEP 1810), a Notification of the International Application Number and of the International Filing Date (Form PCT/RO/105) will be issued in due course, subject to prescriptions concerning national security, and the date shown on this Acknowledgement Receipt will establish the international filing date of the application.

UNITED STATES PATENT APPLICATION

OF

Sanghyun PARK, Jonghwan CHOI and Sangmin SEO

FOR

TRAINING METHOD AND DEVICE FOR MOLECULAR GENERATION
MODEL

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application claims priority under 35 U.S.C. § 119 to Korean Patent Application No. 10-2023-0043303 filed on April 3, 2023, in the Korean Intellectual Property Office, the disclosures of which are incorporated by reference herein in their entireties.

TECHNICAL FIELD

[0002] 본 개시는 분자 생성 모델을 위한 학습 방법 및 장치에 관한 것이다.

BACKGROUND

[0003] 구조 제약 분자 생성은 목표 지향 분자 최적화 연구에서 어려운 문제에 해당한다. 구조 제약 분자 생성의 목표는 소스 약물의 분자 구조와 유사하면서도 향상된 표적 화학적 특성을 가진 새로운 분자를 생성하는 것이다. 유기 화학의 전통적인 접근 방식은 특정 질병에 대한 잠재적 활성을 가진 새로운 약물 후보를 식별하기 위해 먼저 소스 약물의 구조 중에서 표적 생물학적 실체와 결합하는 주요 영역을 조사하고, 해당 부분을 제외한 나머지 부분에 대해 치환 가능한 분자 모티프 조합을 고려하여 분자 구조의 식별을 포함한다. 그러나 이러한 **brute-force-like** 접근법은 1030-1060 범위로 추정되는 약물과 같은 화학 공간의 크기가 크기 때문에 상당한 전문 지식과 막대한 비용이 필요하다.

[0004] 이러한 비효율성 문제를 해결하기 위해 다양한 컴퓨터 지원 약물 설계 방법, 특히 인공지능(AI) 기술 기반 응용 프로그램이 제안되었다. 그러나, 종래의 다양한 인공지능 기술 기반 분자 생성 모델은 특정한 화학적 속성을 만족하는 분자를 생성하는 데에는 효과적이거나, 동시에 소스 분자와 유사한 구조를 갖는 분자를 생성하는 것을 동시에 달성하기에는 아직 개선의 여지가 남아있다.

SUMMARY

[0005] 본 개시의 일 실시예에 따르면, 분자 생성 모델을 위한 학습 방법은 소스 분자 모델, 소스 분자 모델과의 구조적 유사도가 제 1 임계치 초과인 타겟 분자 모델, 소스 분자 모델 또는 타겟 분자 모델 중 하나 이상의 모델과의 구조적 유사도가 제 1 임계치 이하인 네거티브 분자 모델을 포함하는 학습 데이터셋을 획득하는 단계, 학습 데이터셋 및 제 1 손실 함수를 기초로, 소스 분자 모델과 타겟 분자 모델 사이의 거리를 조정하도록 분자 생성 모델을 학습시키는 단계 및 학습 데이터셋 및 제 1 손실 함수와 상이한 제 2 손실 함수를 기초로, 소스 분자 모델과 네거티브 분자 모델 사이의 거리 및 타겟 분자 모델과 네거티브 분자 모델 사이의 거리 중 적어도 하나의 거리를 조정하도록 분자 생성 모델을 학습시키는 단계를 포함할 수 있다.

[0006] 일 실시예에 따르면, 학습 데이터셋 및 제 1 손실 함수를 기초로, 소스 분자 모델과 타겟 분자 모델 사이의 거리를 조정하도록 분자 생성 모델을 학습시키는 단계는 학습 데이터셋 및 제 1 손실 함수를 기초로, 소스 분자 모델과 타겟 분자 모델 사이의 거리가 가까워지도록 분자 생성 모델을 학습시키는 단계를 포함할 수 있다.

[0007] 일 실시예에 따르면, 학습 데이터셋 및 제 1 손실 함수와 상이한 제 2 손실 함수를 기초로, 소스 분자 모델과 네거티브 분자 모델 사이의 거리 및 타겟 분자 모델과 네거티브 분자 모델 사이의 거리 중 적어도 하나의 거리를 조정하도록 분자 생성 모델을 학습시키는 단계는 학습 데이터셋 및 제 1 손실 함수와 상이한 제 2 손실 함수를 기초로, 소스 분자 모델과 네거티브 분자 모델 사이의 거리 및 타겟 분자 모델과 네거티브 분자 모델 사이의 거리 중 적어도 하나의 거리가 멀어지도록 분자 생성 모델을 학습시키는 단계를 포함할 수 있다.

[0008] 일 실시예에 따르면, 학습 데이터셋 및 보상 함수를 기초로, 소스 분자 모델로부터 소스 분자 모델과의 구조적 유사도가 제 2 임계치 초과인 분자 모델이 출력되도록, 분자 생성 모델을 학습시키는 단계를 더 포함할 수 있다.

[0009] 일 실시예에 따르면, 학습 데이터셋 및 보상 함수를 기초로, 소스 분자 모델로부터 소스 분자 모델과의 구조적 유사도가 제 2 임계치 초과인 분자 모델이 출력되도록, 분자 생성 모델을 학습시키는 단계는 분자 생성 모델에 소스 분자 모델을 입력하여 출력 분자 모델을 획득하는 단계 및 출력 분자 모델과 소스 분자 모델을 비교한 결과 출력 분자 모델과 소스 분자 모델 사이의 구조적 유사도가 제 2 임계치를 초과하는지 여부를 기초로, 출력 분자 모델과 연관된 포지티브 가중치 또는 네거티브 가중치를 산출하여 분자 생성 모델에 부여하는 단계를 포함할 수 있다.

[0010] 일 실시예에 따르면, 출력 분자 모델과 소스 분자 모델을 비교한 결과 출력 분자 모델과 소스 분자 모델 사이의 구조적 유사도가 제 2 임계치를 초과하는지 여부를 기초로, 출력 분자 모델과 연관된 포지티브 가중치 또는 네거티브 가중치를 산출하여 분자 생성 모델에 부여하는 단계는 출력 분자 모델과 소스 분자 모델을 비교한 결과 출력 분자 모델과 소스 분자 모델 사이의 구조적 유사도가 제 2 임계치를 초과하는지 여부 및 출력 분자 모델의 화학적 속성 스코어가 소스 분자 모델의 화학적 속성 스코어를 초과하는지 여부를 기초로, 출력 분자 모델과 연관된 포지티브 가중치 또는 네거티브 가중치를 산출하여 분자 생성 모델에 부여하는 단계를 포함할 수 있다.

[0011] 일 실시예에 따르면, 학습 데이터셋 및 보상 함수를 기초로, 소스 분자 모델로부터 소스 분자 모델과의 구조적 유사도가 제 2 임계치 초과인 분자 모델이 출력되도록, 분자 생성 모델을 학습시키는 단계는 학습 데이터셋 및 보상 함수를 기초로, 소스 분자 모델로부터, 소스 분자 모델과의 구조적 유사도가 제 2 임계치 초과이고, 소스 분자 모델의 화학적 속성 스코어보다 큰 화학적 속성 스코어를 갖는 분자 모델이 출력되도록, 분자 생성 모델을 학습시키는 단계를 포함할 수 있다.

[0012] 일 실시예에 따르면, 타겟 분자 모델의 화학적 속성 스코어는 소스 분자 모델의 화학적 속성 스코어보다 크게 구성될 수 있다.

[0013] 본 개시의 다른 실시예에 따르면, 분자 생성 모델을 위한 학습 방법을

실행시키도록 컴퓨터로 판독 가능한 기록매체에 기록된 컴퓨터 프로그램이 제공될 수 있다.

[0014] 본 개시의 또 다른 실시예에 따르면, 분자 생성 모델을 위한 학습 장치는 분자 생성 모델과 연관된 데이터를 저장하는 메모리 및 메모리와 연결되어 분자 생성 모델을 학습시키는 적어도 하나의 프로세서를 포함하고, 적어도 하나의 프로세서는 소스 분자 모델, 소스 분자 모델과의 구조적 유사도가 제 1 임계치 초과인 타겟 분자 모델, 소스 분자 모델 또는 타겟 분자 모델 중 하나 이상의 모델과의 구조적 유사도가 제 1 임계치 이하인 네거티브 분자 모델을 포함하는 학습 데이터셋을 획득하는 것 학습 데이터셋 및 제 1 손실 함수를 기초로, 소스 분자 모델과 타겟 분자 모델 사이의 거리를 조정하도록 분자 생성 모델을 학습시키는 것과 학습 데이터셋 및 제 1 손실 함수와 상이한 제 2 손실 함수를 기초로, 소스 분자 모델과 네거티브 분자 모델 사이의 거리 및 타겟 분자 모델과 네거티브 분자 모델 사이의 거리 중 적어도 하나의 거리를 조정하도록 분자 생성 모델을 학습시키는 것을 실행하도록 구성된 명령어들을 포함할 수 있다.

BRIEF DESCRIPTION OF THE FIGURES

[0015] 도 1 은 본 개시의 일 실시예에 따른 분자 생성 모델의 동작을 나타내는 아키텍처이다.

[0016] 도 2A 내지 2C 는 본 개시의 일 실시예에 따른 학습 데이터셋을 이용하여 분자 생성 모델을 학습시키는 과정의 일부를 나타내는 아키텍처이다.

[0017] 도 3 은 본 개시의 일 실시예에 따른 입력된 임의의 분자 모델을 이용하여 분자 생성 모델을 학습시키는 과정의 다른 일부를 나타내는 아키텍처이다.

[0018] 도 4 는 본 개시의 일 실시예에 따른 컴퓨팅 장치(410)의 내부 구성을 나타내는 블록도이다.

[0019] 도 5A 내지 7D 는 본 개시의 일 실시예에 따라 0.40 내지 0.70 범위의 여러 구조적 유사도 임계값에 대한 성공률을 평가한 결과를 나타낸다.

[0020] 도 8A 내지 8D 은 본 개시의 일 실시예에 따른 학습 방법의 이점을 입증하기 위하여 DRD2 벤치마크 데이터셋에 대한 절제 실험을 수행한 결과를 나타낸다.

[0021] 도 9 는 본 개시의 일 실시예에 따른 제 1 손실 함수(Contractive loss) 및 제 2 손실 함수(Margin loss)가 있거나 없는 각 훈련된 모델에 대해 속성, 개선도 및 유사도라는 세 가지 지표를 평가한 결과를 나타낸다.

[0022] 도 10 은 본 개시의 일 실시예에 따른 손실 함수들(contractive & margin)의 평균 구조적 유사도를 평가한 결과를 나타낸다.

[0023] 도 11A 및 11B 은 본 개시의 일 실시예에 따른 분자 생성 모델의 선형 프로젝션 분석 수행 결과를 나타낸다.

[0024] 도 12 는 본 개시의 일 실시예에 따라 소라페닙을 소스 분자 모델로 사용하여 10,000 개의 분자 모델을 학습하고 생성한 후 성공률을 비교한 결과를 나타낸다.

[0025] 도 13 은 본 개시의 일 실시예에 따라 분자 생성 모델로부터 육안으로도 소라페닙과 유사한 구조의 분자 모델이 생성된 결과를 나타낸다.

[0026] 도 14A 및 14B 는 본 개시의 일 실시예에 따른 실험실에서 히트 후보가 소라페닙보다 ABCG2 에 대한 결합 에너지가 높은 지를 확인하기 위해 AutoDock Vina 를 사용하여 결합 에너지를 비교한 결과를 나타낸다.

[0027] 도 15 는 본 개시의 일 실시예에 따른 실험실에서 히트 후보가 소라페닙만큼 BRAF 에 대한 결합 친화력이 강한 지를 확인하기 위해 Chimera 를 사용하여 수용체-리간드 복합체의 3D 구조에 대한 그래픽을 도시한 결과 및 LigPlot Plus 를 사용하여 수용체-리간드 복합체의 2D 구조에 대한 그래픽을 도시한 결과를

나타낸다.

[0028] 도 16 은 본 개시의 일 실시예에 따른 문자 생성 모델을 위한 학습 방법의 흐름도이다.

DETAILED DESCRIPTION

[0029] 이하, 본 개시의 실시를 위한 구체적인 내용을 첨부된 도면을 참조하여 상세히 설명한다. 다만, 이하의 설명에서는 본 개시의 요지를 불필요하게 흐릴 우려가 있는 경우, 널리 알려진 기능이나 구성에 관한 구체적 설명은 생략하기로 한다.

[0030] 첨부된 도면에서, 동일하거나 대응하는 구성요소에는 동일한 참조부호가 부여되어 있다. 또한, 이하의 실시예들의 설명에 있어서, 동일하거나 대응되는 구성요소를 중복하여 기술하는 것이 생략될 수 있다. 그러나 구성요소에 관한 기술이 생략되어도, 그러한 구성요소가 어떤 실시예에 포함되지 않는 것으로 의도되지는 않는다.

[0031] 개시된 실시예의 이점 및 특징, 그리고 그것들을 달성하는 방법은 첨부되는 도면과 함께 후술되어 있는 실시예들을 참조하면 명확해질 것이다. 그러나 본 개시는 이하에서 개시되는 실시예들에 한정되는 것이 아니라 서로 다른 다양한 형태로 구현될 수 있으며, 단지 본 실시예들은 본 개시가 완전하도록 하고, 본 개시가 통상의 기술자에게 발명의 범주를 완전하게 알려주기 위해 제공되는 것일 뿐이다.

[0032] 본 명세서에서 사용되는 용어에 대해 간략히 설명하고, 개시된 실시예에 대해 구체적으로 설명하기로 한다. 본 명세서에서 사용되는 용어는 본 개시에서의 기능을 고려하면서 가능한 현재 널리 사용되는 일반적인 용어들을 선택하였으나, 이는 관련 분야에 종사하는 기술자의 의도 또는 관례, 새로운 기술의 출현 등에 따라 달라질 수 있다. 또한, 특정한 경우는 출원인이 임의로 선정한 용어도 있으며, 이 경우 해당되는 발명의 설명 부분에서 상세히 그 의미를 기재할 것이다.

따라서 본 개시에서 사용되는 용어는 단순한 용어의 명칭이 아닌, 그 용어가 가지는 의미와 본 개시의 전반에 걸친 내용을 토대로 정의되어야 한다.

[0033] 본 명세서에서의 단수의 표현은 문맥상 명백하게 단수인 것으로 특정하지 않는 한, 복수의 표현을 포함한다. 또한, 복수의 표현은 문맥상 명백하게 복수인 것으로 특정하지 않는 한, 단수의 표현을 포함한다. 명세서 전체에서 어떤 부분이 어떤 구성요소를 '포함'한다고 할 때, 이는 특별히 반대되는 기재가 없는 한 다른 구성요소를 제외하는 것이 아니라 다른 구성요소를 더 포함할 수 있음을 의미한다.

[0034] 또한, 명세서에서 사용되는 '부'라는 용어는 소프트웨어 또는 하드웨어 구성요소를 의미하며, '부'는 어떤 역할들을 수행한다. 그렇지만 '부'는 소프트웨어 또는 하드웨어에 한정되는 의미는 아니다. '부'는 어드레싱할 수 있는 저장 매체에 있도록 구성될 수도 있고 하나 또는 그 이상의 프로세서들을 재생시키도록 구성될 수도 있다. 따라서, 일 예로서 '부'는 소프트웨어 구성요소들, 객체지향 소프트웨어 구성요소들, 클래스 구성요소들 및 태스크 구성요소들과 같은 구성요소들과, 프로세스들, 함수들, 속성들, 프로시저들, 서브루틴들, 프로그램 코드의 세그먼트들, 드라이버들, 펌웨어, 마이크로코드, 회로, 데이터, 데이터베이스, 데이터 구조들, 테이블들, 어레이들 또는 변수들 중 적어도 하나를 포함할 수 있다. 구성요소들과 '부'들은 안에서 제공되는 기능은 더 작은 수의 구성요소들 및 '부'들로 결합되거나 추가적인 구성요소들과 '부'들로 더 분리될 수 있다.

[0035] 본 개시의 일 실시예에 따르면 '부'는 프로세서 및 메모리로 구현될 수 있다. '프로세서'는 범용 프로세서, 중앙 처리 장치(CPU), 마이크로프로세서, 디지털 신호 프로세서(DSP), 제어기, 마이크로제어기, 상태 머신 등을 포함하도록 넓게 해석되어야 한다. 몇몇 환경에서는, '프로세서'는 주문형 반도체(ASIC), 프로그램 가능 로직 디바이스(PLD), 필드 프로그램가능 게이트 어레이(FPGA) 등을 지칭할

수도 있다. '프로세서'는, 예를 들어, DSP 와 마이크로프로세서의 조합, 복수의 마이크로프로세서들의 조합, DSP 코어와 결합한 하나 이상의 마이크로프로세서들의 조합, 또는 임의의 다른 그러한 구성들의 조합과 같은 처리 디바이스들의 조합을 지칭할 수도 있다. 또한, '메모리'는 전자 정보를 저장 가능한 임의의 전자 컴포넌트를 포함하도록 넓게 해석되어야 한다. '메모리'는 임의 액세스 메모리(RAM), 판독-전용 메모리(ROM), 비-휘발성 임의 액세스 메모리(NVRAM), 프로그램가능 판독-전용 메모리(PROM), 소거-프로그램가능 판독 전용 메모리(EPROM), 전기적으로 소거가능 PROM(EEPROM), 플래쉬 메모리, 자기 또는 광학 데이터 저장장치, 레지스터들 등과 같은 프로세서-판독가능 매체의 다양한 유형들을 지칭할 수도 있다. 프로세서가 메모리로부터 정보를 판독하고/하거나 메모리에 정보를 기록할 수 있다면 메모리는 프로세서와 전자 통신 상태에 있다고 불린다. 프로세서에 집적된 메모리는 프로세서와 전자 통신 상태에 있다.

[0036] 도 1 은 본 개시의 일 실시예에 따른 분자 생성 모델(100)의 동작을 나타내는 아키텍처이다. 도시된 바와 같이, 분자 생성 모델(100)은 입력되는 소스 분자 모델(110)로부터 최종 분자 모델(120)을 획득할 수 있다. 여기서, 최종 분자 모델(120)은 소스 분자 모델(110)에서 물리적 및/또는 화학적 속성이 변형된 분자 모델을 지칭할 수 있다. 한편, 분자 생성 모델(100)과 연관된 동작은 컴퓨팅 장치의 적어도 하나의 프로세서에 의해 수행되며, 컴퓨팅 장치의 하드웨어적 구성은 도 4 에서 상세히 후술된다.

[0037] 소스 분자 모델(110)은 최종 분자 모델(120)과 구조적으로 유사할 수 있다. 즉, 분자 생성 모델(100)은 소스 분자 모델(110)과 최종 분자 모델(120) 사이의 구조적 유사도가 사전 결정된 임계치를 초과하도록 구성될 수 있다. 이 경우, 구조적 유사도는 타니모토(Tanimoto) 유사도 측정 방식에 따라 측정될 수 있다.

[0038] 소스 분자 모델(110)은 최종 분자 모델(120)과 화학적 속성 스코어는

유사하지 않을 수 있다. 구체적으로, 최종 분자 모델(120)의 화학적 속성 스코어는 소스 분자 모델(110)의 화학적 속성 스코어보다 크게 구성될 수 있다. 이는 최종 분자 모델(120)의 화학적 속성이 소스 분자 모델(110)과 비교하여 보다 개선된 것을 의미할 수 있다. 또한, 화학적 속성 스코어는 소스 분자 모델(110) 및/또는 최종 분자 모델(120)과 같은 분자 모델의 특정한 화학적 반응 정도를 의미할 수 있다. 예를 들어, 분자 모델이 약물인 경우, 화학적 속성 스코어는 인체 내에서 분자 모델이 특정 세포와 반응하는 정도를 의미할 수 있다.

[0039] 프로세서는 소스 분자 모델(110)로부터 화학적 속성이 향상되고 분자 구조가 유사한 최종 분자 모델(120)이 출력되도록 분자 생성 모델(100)을 학습시킬 수 있다. 이를 위해, 프로세서는 복수의 분자 구조 모델 샘플들이 포함된 학습 데이터셋을 이용하여 분자 생성 모델(100)이 입력되는 분자 모델(여기서, 소스 분자 모델(110))의 화학적 속성 및 분자 구조를 학습하도록 분자 생성 모델(100)을 학습시킬 수 있다. 도 2A 내지 2C 및 3에서는 이와 같이 프로세서가 학습 데이터셋을 구성하고, 해당 학습 데이터셋을 이용하여 분자 생성 모델(110)을 학습시키는 과정이 상세히 후술된다.

[0040] 도 2A 내지 2C는 본 개시의 일 실시예에 따른 학습 데이터셋(110, 210, 220)을 이용하여 분자 생성 모델(100)을 학습시키는 과정의 일부를 나타내는 아키텍처이다. 도시된 바와 같이, 학습 데이터셋은 소스 분자 모델(110), 타겟 분자 모델(210) 및 네거티브 분자 모델(220) 중 적어도 하나를 포함할 수 있다. 이 경우, 타겟 분자 모델(210)은 소스 분자 모델(110)과 구조적 유사도가 제 1 임계치를 초과하도록 구성될 수 있다. 또한, 네거티브 분자 모델(220)은 소스 분자 모델(110)과의 구조적 유사도가 제 1 임계치 이하 및 타겟 분자 모델(210)과의 구조적 유사도가 제 1 임계치 이하가 되도록 구성될 수 있다. 추가적으로 또는 대안적으로, 네거티브 분자 모델(220)은 소스 분자 모델(110)과의 구조적 유사도가 제 1

임계치보다 작은 제 2 임계치 이하이고, 타겟 분자 모델(210)과의 구조적 유사도가 제 1 임계치보다 작은 제 2 임계치 이하가 되도록 구성될 수 있다.

[0041] 프로세서는 학습 데이터셋(110, 210, 220)을 획득할 수 있다. 그리고 나서, 프로세서는 학습 데이터셋(110, 210, 220) 및 제 1 손실 함수를 기초로, 소스 분자 모델(110)과 타겟 분자 모델(210) 사이의 거리를 조정하도록 분자 생성 모델(100)을 학습시킬 수 있다. 구체적으로, 프로세서는 학습 데이터셋(110, 210, 220) 및 제 1 손실 함수를 기초로, 소스 분자 모델(110)과 타겟 분자 모델(210) 사이의 거리가 짧아지도록 분자 생성 모델(100)을 학습시킬 수 있다. 예를 들어, 프로세서는 소스 분자 모델(110)을 분자 생성 모델(100)에 입력시켜, 제 1 벡터(232)를 획득할 수 있다. 또한, 프로세서는 타겟 분자 모델(210)을 분자 생성 모델(100)에 입력하여, 제 2 벡터(234)를 획득할 수 있다. 그리고 나서, 프로세서는 제 1 벡터(232)와 제 2 벡터(234) 사이의 거리를 좁힐 수 있다. 즉, 프로세서는 제 1 손실 함수를 기초로 소스 분자 모델(110)과 타겟 분자 모델(210) 사이의 구조적 유사도를 학습할 수 있다.

[0042] 프로세서는 학습 데이터셋(110, 210, 220) 및 제 1 손실 함수와는 상이한 제 2 손실 함수를 기초로, 소스 분자 모델(110)과 네거티브 분자 모델(220) 사이의 거리를 조정하도록 분자 생성 모델(100)을 학습시킬 수 있다. 구체적으로, 프로세서는 학습 데이터셋(110, 210, 220) 및 제 2 손실 함수를 기초로, 소스 분자 모델(110)과 네거티브 분자 모델(220) 사이의 거리가 짧아지도록 분자 생성 모델(100)을 학습시킬 수 있다. 예를 들어, 프로세서는 네거티브 분자 모델(220)을 분자 생성 모델(100)에 입력하여, 제 3 벡터(236)를 획득할 수 있다. 그리고 나서, 프로세서는 제 1 벡터(232)와 제 3 벡터(236) 사이의 거리를 넓힐 수 있다. 이에 따라, 제 1 벡터(232)와 제 3 벡터(236) 사이의 거리는 제 1 벡터(232)와 제 2 벡터(234) 사이의 거리보다 멀게 구성될 수 있다. 즉, 프로세서는 제 2 손실 함수를 기초로 소스 분자 모델(110)과 네거티브 분자 모델(220) 사이의 구조적 비유사도를 학습할 수

있다.

[0043] 프로세서는 학습 데이터셋(110, 210, 220) 및 제 2 손실 함수를 기초로, 타겟 분자 모델(210)과 네거티브 분자 모델(220) 사이의 거리를 조정하도록 분자 생성 모델(100)을 학습시킬 수 있다. 구체적으로, 프로세서는 학습 데이터셋(110, 210, 220) 및 제 2 손실 함수를 기초로, 타겟 분자 모델(210)과 네거티브 분자 모델(220) 사이의 거리를 조정하도록 분자 생성 모델(100)을 학습시킬 수 있다. 예를 들어, 제 2 벡터(234)와 제 3 벡터(236) 사이의 거리를 넓힐 수 있다. 이에 따라, 제 2 벡터(234)와 제 3 벡터(236) 사이의 거리는 제 1 벡터(232)와 제 2 벡터(234) 사이의 거리보다 멀게 구성될 수 있다. 즉, 프로세서는 제 2 손실 함수를 기초로 타겟 분자 모델(210)과 네거티브 분자 모델(220) 사이의 구조적 비유사도를 학습할 수 있다.

[0044] 한편, 도시된 바와 같이 제 1 벡터(232), 제 2 벡터(234) 및 제 3 벡터(236)는 분자 생성 모델(100)의 제 1 모듈(240)에 각각 소스 분자 모델(110), 타겟 분자 모델(210) 및 네거티브 분자 모델(220)을 입력함으로써 획득될 수 있다. 이 경우, 제 1 모듈(240)은 적어도 하나의 인코더를 포함할 수 있다.

[0045] 도 3은 본 개시의 일 실시예에 따른 입력된 임의의 분자 모델(여기서, 소스 분자 모델(110))을 이용하여 분자 생성 모델(100)을 학습시키는 과정의 다른 일부를 나타내는 아키텍처이다. 도 3에서 분자 생성 모델(100)이 학습되는 과정은 도 2A 내지 2C에서 분자 생성 모델(100)이 학습되는 과정 이후에 수행될 수 있다. 추가적으로 또는 대안적으로, 분자 생성 모델(100)이 학습되는 과정은 도 2A 내지 2C의 분자 생성 모델(100)이 학습되는 과정과 적어도 일부 병렬적으로 수행될 수도 있다.

[0046] 프로세서는 학습 데이터셋(예: 학습 데이터셋(110, 210, 220)) 및 보상 함수를 기초로 입력된 임의의 분자 모델로부터 임의의 분자 모델과의 구조적 유사도가 미리 결정된 임계치를 초과하는 분자 모델이 출력되도록 분자 생성

모델(100)을 학습시킬 수 있다. 즉, 보상 함수는 분자 생성 모델(100)로부터 출력된 분자 모델(여기서, 출력 분자 모델(320))의 물리적 구조 및/또는 화학적 속성을 기초로 분자 생성 모델(100)에 포지티브 가중치를 부여하여 출력된 분자 모델의 구조적 유사도 및 화학적 속성에 대한 학습을 강화할 수 있다. 또한, 보상 함수는 분자 생성 모델(100)로부터 출력된 분자 모델의 물리적 구조 및/또는 화학적 속성을 기초로 분자 생성 모델(100)에 네거티브 가중치를 부여하여 출력된 분자 모델의 구조적 비유사도 및 화학적 속성에 대한 학습을 강화할 수 있다. 이하에서는, 소스 분자 모델(110)이 입력된 경우를 예시로 분자 생성 모델(100)이 학습되는 과정이 상세히 후술된다.

[0047] 먼저, 프로세서는 학습 데이터셋에 포함된 임의의 분자 모델을 분자 생성 모델(100)에 입력하여 분자 모델에 대응하는 벡터를 획득할 수 있다. 예를 들어, 프로세서는 학습 데이터셋에 포함된 소스 분자 모델(110)을 분자 생성 모델(100)(구체적으로는, 제 1 모듈(240))에 입력하여 제 1 벡터(232)를 획득할 수 있다. 그리고 나서, 획득된 제 1 벡터(232)를 제 2 모듈(310)에 입력하여 출력 분자 모델(320)을 획득할 수 있다.

[0048] 프로세서는 획득된 출력 분자 모델(320)을 소스 분자 모델(110)과 비교할 수 있다. 구체적으로, 프로세서는 출력 분자 모델(320)과 소스 분자 모델(110) 사이의 구조적 유사도를 산출할 수 있다. 추가적으로 또는 대안적으로, 프로세서는 출력 분자 모델(320)의 화학적 속성 스코어를 산출할 수 있다.

[0049] 프로세서는 출력 분자 모델(320)과 소스 분자 모델(110) 사이의 구조적 유사도가 미리 결정된 임계치를 초과하는 경우, 출력 분자 모델(320)과 연관된 포지티브 가중치를 산출할 수 있다. 이 경우, 출력 분자 모델(320)과 연관된 포지티브 가중치는 출력 분자 모델(320)의 구조적 유사도 및/또는 화학적 속성 스코어를 기초로 결정될 수 있다. 그리고 나서, 프로세서는 산출된 포지티브

가중치를 분자 생성 모델(100)에 부여할 수 있다. 반면, 프로세서는 출력 분자 모델(320)과 소스 분자 모델(110) 사이의 구조적 유사도가 미리 결정된 임계치 이하인 경우, 출력 분자 모델(320)과 연관된 네거티브 가중치를 산출할 수 있다. 이 경우, 출력 분자 모델(320)과 연관된 네거티브 가중치는 출력 분자 모델(320)의 구조적 유사도 및/또는 화학적 속성 스코어를 기초로 결정될 수 있다. 그리고 나서, 프로세서는 산출된 네거티브 가중치를 분자 생성 모델(100)에 부여할 수 있다. 즉, 프로세서는 출력 분자 모델(320)이 입력된 소스 분자 모델(110)과 구조적으로 유사한 경우에는 분자 생성 모델(100)에 리워드를 부여하고, 출력 분자 모델(320)이 입력된 소스 분자 모델(110)과 구조적으로 비유사한 경우에는 분자 생성 모델(100)에 페널티를 부여함으로써 소스 분자 모델(110)의 구조에 대한 학습을 강화할 수 있다.

[0050] 유사하게, 프로세서는 출력 분자 모델(320)의 화학적 속성 스코어가 미리 결정된 임계치를 초과하는 경우, 출력 분자 모델(320)과 연관된 포지티브 가중치를 산출할 수 있다. 이 경우에도, 출력 분자 모델(320)과 연관된 포지티브 가중치는 출력 분자 모델(320)의 화학적 속성 스코어를 기초로 결정될 수 있다. 그리고 나서, 프로세서는 산출된 포지티브 가중치를 분자 생성 모델(100)에 부여할 수 있다. 반면, 프로세서는 출력 분자 모델(320)의 화학적 속성 스코어가 미리 결정된 임계치 이하인 경우, 출력 분자 모델(320)과 연관된 네거티브 가중치를 산출할 수 있다. 이 경우 또한, 출력 분자 모델(320)과 연관된 네거티브 가중치는 출력 분자 모델(320)의 화학적 속성 스코어를 기초로 결정될 수 있다. 즉, 프로세서는 출력 분자 모델(320)이 입력된 소스 분자 모델(110)과 화학적 반응 레벨이 유사하거나 및/또는 개선된 경우에는 분자 생성 모델(100)에 리워드를 부여하고, 출력 분자 모델(320)이 입력된 소스 분자 모델(110)보다 화학적 반응 레벨이 낮은 경우에는 분자 생성 모델(100)에 페널티를 부여함으로써, 분자 생성 모델(100)의 화학적 속성에 대한 학습을 강화할 수 있다.

[0051] 한편, 프로세서는 포지티브 가중치 및/또는 네거티브 가중치를 산출함에 있어서, 구조적 유사도 및 화학적 속성 스코어를 동시에 고려할 수도 있다. 예를 들어, 출력 분자 모델(320)의 화학적 속성 스코어가 미리 결정된 임계치를 초과하고, 출력 분자 모델(320)과 소스 분자 모델(110) 사이의 구조적 유사도가 미리 결정된 임계치를 초과하는 경우, 프로세서는 출력 분자 모델(320)의 구조적 유사도 및/또는 화학적 속성 스코어를 기초로 포지티브 가중치를 산출하여 분자 생성 모델(100)에 부여할 수 있다. 다른 예를 들어, 출력 분자 모델(320)의 화학적 속성 스코어가 미리 결정된 임계치 미만이거나, 출력 분자 모델(320)과 소스 분자 모델(110) 사이의 구조적 유사도가 미리 결정된 임계치 이하인 경우, 프로세서는 출력 분자 모델(320)의 구조적 유사도 및/또는 화학적 속성 스코어를 기초로 네거티브 가중치를 산출하여 분자 생성 모델(100)에 부여할 수 있다. 마찬가지로, 화학적 속성 스코어와 구조적 유사도가 각각 모두 해당 임계치의 미만인 경우에도 프로세서는 네거티브 가중치를 산출하여 분자 생성 모델(100)에 부여할 수 있다.

[0052] 도 4는 본 개시의 일 실시예에 따른 컴퓨팅 장치(410)의 내부 구성을 나타내는 블록도이다. 여기서, 컴퓨팅 장치(410)는 본 개시에 따른 분자 생성 모델을 위한 학습 방법을 실행하기 위한 장치를 지칭할 수 있다. 컴퓨팅 장치(410)는 머신러닝을 위한 프로그램 등이 실행 가능하고 유/무선 통신이 가능한 임의의 컴퓨팅 장치로서, 데스크탑, 스마트폰, 태블릿, 노트북 등을 포함할 수 있다.

[0053] 컴퓨팅 장치(410)는 메모리(412), 프로세서(414), 통신 모듈(416) 및 입출력 인터페이스(418)를 포함할 수 있다. 도 4는 도시되지 않았으나, 컴퓨팅 장치(410)는 통신 모듈(416)을 이용하여 네트워크를 통해 정보 및/또는 데이터를 통신할 수 있도록 구성될 수 있다. 또한, 입출력 장치(420)는 입출력 인터페이스(418)를 통해 컴퓨팅 장치(410)에 정보 및/또는 데이터를 입력하거나 컴퓨팅 장치(410)로부터 생성된 정보 및/또는 데이터를 출력하도록 구성될 수 있다.

[0054] 메모리(412)는 비-일시적인 임의의 컴퓨터 판독 가능한 기록매체를 포함할 수 있다. 일 실시예에 따르면, 메모리(412)는 RAM(random access memory), ROM(read only memory), 디스크 드라이브, SSD(solid state drive), 플래시 메모리(flash memory) 등과 같은 비소멸성 대용량 저장 장치(permanent mass storage device)를 포함할 수 있다. 다른 예로서, ROM, SSD, 플래시 메모리, 디스크 드라이브 등과 같은 비소멸성 대용량 저장 장치는 메모리와는 구분되는 별도의 영구 저장 장치로서 컴퓨팅 장치(410)에 포함될 수 있다. 또한, 메모리(412)에는 운영체제와 적어도 하나의 프로그램 코드(예를 들어, 컴퓨팅 장치(410)에 설치되어 구동되는 머신러닝을 위한 프로그램을 실행하기 위한 코드)가 저장될 수 있다.

[0055] 이러한 소프트웨어 구성요소들은 메모리(412)와는 별도의 컴퓨터에서 판독 가능한 기록매체로부터 로딩될 수 있다. 이러한 별도의 컴퓨터에서 판독가능한 기록매체는 이러한 컴퓨팅 장치(410) 및 외부 서버에 직접 연결가능한 기록 매체를 포함할 수 있는데, 예를 들어, 플로피 드라이브, 디스크, 테이프, DVD/CD-ROM 드라이브, 메모리 카드 등의 컴퓨터에서 판독 가능한 기록매체를 포함할 수 있다. 다른 예로서, 소프트웨어 구성요소들은 컴퓨터에서 판독 가능한 기록매체가 아닌 통신 모듈(416)을 통해 메모리(412)에 로딩될 수도 있다. 예를 들어, 적어도 하나의 프로그램은 개발자들 또는 어플리케이션의 설치 파일을 배포하는 파일 배포 시스템이 네트워크를 통해 제공하는 파일들에 의해 설치되는 컴퓨터 프로그램에 기반하여 메모리(412)에 로딩될 수 있다.

[0056] 프로세서(414)는 기본적인 산술, 로직 및 입출력 연산을 수행함으로써, 컴퓨터 프로그램의 명령을 처리하도록 구성될 수 있다. 명령은 메모리(412) 또는 통신 모듈(416)에 의해 프로세서(414)로 제공될 수 있다. 예를 들어, 프로세서(414)는 메모리(412)와 같은 기록 장치에 저장된 프로그램 코드에 따라 수신되는 명령을 실행하도록 구성될 수 있다.

[0057] 통신 모듈(416)은 네트워크를 통해 컴퓨팅 장치(410)와 외부 서버가 서로 통신하기 위한 구성 또는 기능을 제공할 수 있으며, 컴퓨팅 장치(410) 및/또는 외부 서버가 다른 사용자 단말 또는 다른 시스템(일례로 별도의 클라우드 시스템 등)과 통신하기 위한 구성 또는 기능을 제공할 수 있다. 일례로, 컴퓨팅 장치(410)의 프로세서(414)가 메모리(412) 등과 같은 기록 장치에 저장된 프로그램 코드에 따라 생성한 요청 또는 데이터는 통신 모듈(416)의 제어에 따라 네트워크를 통해 외부 서버로 전달될 수 있다. 역으로, 외부 서버의 프로세서의 제어에 따라 제공되는 제어 신호나 명령이 네트워크를 거쳐 컴퓨팅 장치(410)의 통신 모듈(416)을 통해 컴퓨팅 장치(410)에 수신될 수 있다.

[0058] 컴퓨팅 장치(410)의 입출력 인터페이스(418)는 입출력 장치(420)와의 상호 작용을 위한 수단일 수 있다. 구체적으로, 입출력 인터페이스(418)는 터치스크린 등과 같이 입력과 출력을 수행하기 위한 구성 또는 기능이 하나로 통합된 장치와의 인터페이스를 위한 수단일 수 있다. 이 경우, 입출력 장치(420)는 오디오 센서 및/또는 이미지 센서를 포함한 카메라, 키보드, 마이크로폰, 마우스 등의 입력 장치를 포함할 수 있다. 또한, 입출력 장치(420)는 디스플레이, 스피커, 햅틱 피드백 디바이스(haptic feedback device) 등과 같은 출력 장치를 포함할 수 있다.

[0059] 도 4에서는 입출력 장치(420)가 컴퓨팅 장치(410)에 포함되지 않도록 도시되어 있으나, 이에 한정되지 않으며, 컴퓨팅 장치(410)와 하나의 장치로 구성될 수 있다. 또한, 도 4에서는 입출력 인터페이스(418)가 프로세서(414)와 별도로 구성된 요소로서 도시되었으나, 이에 한정되지 않으며, 입출력 인터페이스(418)가 프로세서(414)에 포함되도록 구성될 수 있다.

[0060] 컴퓨팅 장치(410)는 도 4의 구성요소들보다 더 많은 구성요소들을 포함할 수 있다. 그러나, 대부분의 종래기술적 구성요소들을 명확하게 도시할 필요성은 없다. 일 실시예에 따르면, 컴퓨팅 장치(410)는 상술된 입출력 장치(420) 중

적어도 일부를 포함하도록 구현될 수 있다. 또한, 컴퓨팅 장치(410)는 트랜시버(transceiver), GPS(Global Positioning system) 모듈, 카메라, 각종 센서, 데이터베이스 등과 같은 다른 구성요소들을 더 포함할 수 있다. 예를 들어, 컴퓨팅 장치(410)가 스마트폰인 경우, 일반적으로 스마트폰이 포함하고 있는 구성요소를 포함할 수 있으며, 예를 들어, 가속도 센서, 자이로 센서, 마이크 모듈, 카메라 모듈, 각종 물리적인 버튼, 터치패널을 이용한 버튼, 입출력 포트, 진동을 위한 진동기 등의 다양한 구성요소들이 컴퓨팅 장치(410)에 더 포함되도록 구현될 수 있다.

[0061] 컴퓨팅 장치(410) 및 외부 서버 각각에서 머신러닝을 위한 프로그램이 동작하는 동안에, 프로세서(414)는 각각 입출력 인터페이스(418)와 연결된 입출력 장치(420)를 통해 입력되거나 선택된 수치 데이터, 텍스트, 이미지, 영상 등을 수신할 수 있으며, 수신된 수치 데이터, 텍스트, 이미지 및/또는 영상 등을 메모리(412)에 저장하거나 통신 모듈(416) 및 네트워크를 통해 서로에게 제공할 수 있다.

[0062] 외부 서버의 프로세서는 복수의 사용자 단말 및/또는 복수의 외부 시스템으로부터 수신된 정보 및/또는 데이터를 관리, 처리 및/또는 저장하도록 구성될 수 있다. 일 실시예에 따르면, 프로세서는 컴퓨팅 장치(410)로부터 수신된 사용자 입력 및 해당 사용자 입력에 따른 로그 데이터를 관리, 처리 및/또는 저장할 수 있다. 추가적으로 또는 대안적으로, 프로세서는 네트워크와 연결된 별도의 클라우드 시스템, 데이터베이스 등으로부터 컴퓨팅 장치(410)의 머신러닝에 이용되는 알고리즘을 실행하기 위한 프로그램 등을 저장 및/또는 업데이트하도록 구성될 수 있다.

[0063] 이하에서는, 본 개시의 문자 생성 모델(도 5A 내지 15에서 “COMA”로 표시됨)을 위한 학습 방법을 구현하기 위하여 수행된 실험례가 상세히 후술된다.

[0064] 문자 생성 모델의 개요

[0065] 본 개시의 문자 생성 모델은 SMILES 문자열을 인코딩 및 디코딩하기

위한 GRU(Gated Recurrent Unit) 기반의 VAE(Variational Autoencoder)로, ASCII 코드를 사용하여 분자 구조를 나타낼 수 있다. 여기서, 각 ASCII 코드는 분자 구조에 포함된 원자, 원자 간 결합 유형 및/또는 결합 구조(예: 가지 구조, 고리 구조 등)를 나타낼 수 있다.

[0066] 분자 생성 모델의 인코더(즉, 제 1 모듈)는 제 1 손실 함수를 기초로 유사한 구조를 가진 두 개의 분자를 잠재 공간에서 서로 가까운 지점에 삽입하는 반면, 제 2 손실 함수를 기초로 서로 다른 구조를 가진 두 개의 분자를 잠재 공간에서 가능한 멀리 배치할 수 있다. 즉, 분자 생성 모델의 디코더는 인코더로부터 출력된 잠재 벡터(예를 들어, 도 2A 내지 2C 의 제 1 벡터(232))에서 유효한 SMILES 문자열을 생성하도록 학습된다. 또한, 입력된 분자 모델(예를 들어, 도 3 의 소스 분자 모델(110))보다 개선된 화학적 속성을 가진 SMILES 문자열을 선택적으로 생성하기 위하여 보상 함수를 이용한 강화 학습이 적용될 수 있다.

[0067] 성능 평가

[0068] 분자 생성 모델의 성능을 평가하기 위하여 네 개의 벤치마크 데이터셋(DRD2, QED, pLogP04 및 pLogP06)이 이용되었다. DRD2 를 이용한 학습 목적은 Tanimoto 유사도가 0.4 이상이라는 조건 하에서 소스 분자 모델보다 도파민 수용체 D2 에 대하여 더 활성인 새로운 분자를 생성하는 것이고, QED 를 이용한 학습 목적은 Tanimoto 유사도가 0.4 이상이라는 조건 하에서 소스 분자 모델보다 더 (기존의) 약물과 유사한 새로운 분자를 생성하는 것이다. QED 스코어의 경우 범위는 [0,1]이며 값이 클수록 약물과의 유사도가 더 높음을 나타낸다. 마지막으로, pLogP04 및 pLogP06 작업의 목표는 각각 0.4 및 0.6 을 소스 분자 모델과의 구조적 유사도 임계치로 하여 페널티 logP 의 스코어를 향상시키는 것이다. 여기서, 페널티 logP 의 스코어는 logP 의 스코어에서 분자 구조 내 고리(ring)의 크기 값과 합성 접근성 점수를 차감한 것을 나타낸다.

[0069] 비교 모델

[0070] 본 개시의 분자 구조 비교 모델은 JTVAE, VJTNN, VJTNN+GAN, CORE, HierG2G, HierG2G+BT 및 UGMMT 의 7 가지 최신 모델과 비교되었다. JTVAE 는 베이지안 최적화 방법을 사용하여 분자 특성을 최적화하는 그래프 기반 분자 생성 모델을 나타낸다. VJTNN 은 뉴럴 어텐션 기능이 추가된 JTVAE 의 최신 버전 모델이고, VJTNN+GAN 은 적대적 훈련이 있는 보다 최신 버전의 모델에 해당한다. CORE 는 copy-and-refine 전략을 사용하여 분자를 생성하는 VJTNN+GAN 의 개선된 버전이다. HierG2G 는 계층적 인코딩 방식을 사용하는 그래프 기반 생성 모델이다. HierG2G+BT 는 데이터 증대를 위한 back-translation 단계를 추가한 HierG2G 의 개선된 버전이다. UGMMT 는 비지도 학습 방식을 사용하여 훈련되는 SMILES 기반 생성 모델이다.

[0071] 평가 지표

[0072] 구조에 제약이 있는 분자 생성의 다양한 평가 지표를 사용하여 분자 생성 모델과 다른 모델들을 평가했다. 먼저 각 벤치마크 작업의 학습 데이터셋으로 모든 모델을 교육하고 테스트 데이터셋의 각 소스 분자에 대해 분자를 20 번 생성한 다음 생성된 분자를 7 개의 지표로 평가했다.

[0073] *유효성(Validity): 테스트 데이터에서 생성된 유효한 SMILES 문자열의 비율

[0074] *Novelty: 훈련 데이터에 없는 유효한 SMILES 문자열의 비율

[0075] *속성(Property): 유효한 SMILES 문자열의 속성 스코어 평균

[0076] *개선(Improvement): 생성된 SMILES 문자열과 소스 SMILES 문자열 간의 속성 스코어 차이의 평균

[0077] *유사도(Similarity): 생성된 SMILES 문자열과 소스 SMILES 문자열 간의 Tanimoto 구조적 유사도의 평균

[0078] *다양성(Diversity): 생성된 SMILES 문자열 간의 Tanimoto 쌍 별 비유사도의 평균

[0079] *성공률(Success rate): 화학적 속성(여기서, 약물적 속성) 개선과 구조적 유사도 기준을 모두 만족하는 유효하고 새로운 SMILES 문자열의 비율

[0080] 성공률 비교

[0081] 도 5A 내지 7D 는 본 개시의 일 실시예에 따라 0.40 내지 0.70 범위의 여러 구조적 유사도 임계값에 대한 성공률을 평가한 결과를 나타낸다. 성공률은 모델이 세 가지 제약 조건인 Novelty, 화학적 속성의 개선 및 구조적 유사도라는 조건을 동시에 충족하는 유효 분자를 얼마나 많이 생성하는지 측정하는 데 가장 중요한 척도이기 때문이다.

[0082] 도 5A 내지 7D 를 참고하면, 본 개시의 분자 생성 모델은 여러 임계값 조건에서 기본 모델과 동등하거나 더 나은 성능을 보였고, 분자 생성 모델이 기본 모델보다 구조적 유사도 제약 조건(0.55 내지 0.70)에서 다른 조건들을 만족하는 분자를 보다 우수하게 생성할 수 있음을 확인했다.

[0083] 도 6A 내지 6D 를 참고하면, 정량적 비교를 위해 구조적 유사도의 임계값에 대한 평균 성공률을 계산하고, DRD2, QED, pLogP04 및 pLogP06 에 대해 본 개시의 분자 생성 모델의 평균 스코어가 각각 0.180, 0.301, 0.154 및 0.213 임을 확인했다. 기본 모델과 비교할 때 분자 생성 모델의 스코어는 최신 모델과 비교하여 0.002 내지 0.240 으로 더 높아 구조에 제약이 있는 분자 생성에 더 적합한 모델임을 확인했다.

[0084] 전반적인 성능

[0085] 도 7A 내지 7D 는 나머지 6 개의 지표는 본 개시의 분자 생성 모델 및 다른 모델들의 특성을 나타낸다. 분자 생성 모델의 유효성과 Novelty 는 모든 벤치마크 데이터셋에서 기본 모델을 능가했다. JTVAE 의 경우, DRD2 및 QED

작업에서 분자 생성 모델보다 우수한 구조적 유사도를 보였지만 화학적 속성 스코어를 동시에 향상시키지 못하여 성공률이 낮았다. 전반적인 평가를 위해 각 모델의 총 유효성, 속성, 개선도, 유사도, Novelty 및 다양성 스코어를 계산했다. QED 를 제외하고 분자 생성 모델이 가장 높은 스코어를 나타내 상술한 성공률 분석과 일치함을 확인하였다. 이러한 실험 결과는 제 1 손실 함수, 제 2 손실 함수 및 보상 함수를 이용한 본 개시의 학습 방법이 구조적 제약이 있는 분자 생성에 효과적임을 입증한다.

[0086] 제 1 손실 함수 및 제 2 손실 함수에 대한 애블레이션(ablation) 연구

[0087] 도 8A 내지 8D 은 본 개시의 일 실시예에 따른 학습 방법의 이점을 입증하기 위하여 DRD2 벤치마크 데이터셋에 대한 절제 실험을 수행한 결과를 나타낸다. 도시된 바와 같이, 제 1 손실 함수(여기서, Contractive loss) 및 제 2 손실 함수(여기서, Margin loss)를 모두 이용한 학습을 통해 높은 구조적 유사도를 달성할 수 있었다. Kruskal-Wallis H 검정을 사용하여 제 1 손실 함수 및 제 2 손실 함수를 모두 이용한 학습이 구조적 유사도에 대하여 통계적으로 유의하게 개선됨을 확인하였다.

[0088] 성능 비교

[0089] 도 9 는 본 개시의 일 실시예에 따른 제 1 손실 함수(여기서, Contractive loss) 및 제 2 손실 함수(여기서, Margin loss)가 있거나 없는 각 훈련된 모델에 대해 속성, 개선도 및 유사도라는 세 가지 지표를 평가한 결과를 나타낸다. 유사도에는 눈에 띄는 차이가 없었지만 제 1 손실 함수 및 제 2 손실 함수를 모두 사용한 경우에만 높은 속성 및 개선 스코어가 관찰되었다. 이러한 결과는 제 1 손실 함수 및 제 2 손실 함수가 구조적 제약이 있는 분자 생성에서 중요한 역할을 한다는 것을 나타낸다.

[0090] 도 10 은 본 개시의 일 실시예에 따른 손실 함수들(여기서, contractive

& margin)의 평균 구조적 유사도를 평가한 결과를 나타낸다. DRD2 데이터셋의 유사한 분자 생성 작업에서 손실 함수들의 조합이 종래의 triplet loss 와 contrastive loss 를 능가한다는 것을 확인했다. 분자 생성 모델은 소스 분자 모델에 대해 평균 유사도가 0.423 인 표적 분자를 생성한 반면, 종래의 triplet loss 및 contrastive loss 를 사용한 경우, 평균 유사도가 각각 0.269 및 0.262 인 표적 분자를 생성한 것으로 나타났다.

[0091] 잠재 공간 분석

[0092] 도 11A 및 11B 는 본 개시의 일 실시예에 따른 분자 생성 모델의 선형 프로젝션 분석 수행 결과를 나타낸다. 분자 생성 모델의 장점은 제 1 손실 함수 및 제 2 손실 함수를 이용하여 구조적 유사도 측면의 성능을 높인다는 것이다. 이를 위해, 분자 생성 모델의 잠재 공간에서 구조가 유사한 분자는 서로 가깝게 만들고 구조적으로 다른 분자는 서로 멀리 떨어지도록 설계되었다. 선형 프로젝션 분석은 상술한 통계 분석에서 사용된 데이터와 동일한 데이터가 사용되었다. 잠재 공간에서 한 점 S6 을 선택 하고 그 점에서 시작하여 임의의 방향으로 화살표를 그리고 화살표 위에 있는 6 개의 점에 해당하는 분자 구조를 비교한 결과, 시작점에 인접한 지점은 Tanimoto 유사도가 높고 먼 지점은 유사도 스코어가 낮음을 확인했다. 따라서 제안한 방법이 의도한 만큼 효과적이라고 판단하였다.

[0093] 구체적인 실험례: 소라페닙(sorafenib) 내성에 대한 약물 발견

[0094] 구조적 제약이 있는 분자 생성은 기존 약물과 유사한 새로운 분자를 생성하여 약물을 이용한 화학적 치료법에 내성이 있는 환자를 위한 약물 후보를 발견하는 데 사용할 수 있다. 약물 후보는 기존 약물의 pharmacophore 특성을 손실하지 않고 약물 내성과 관련된 화학적 특성을 감소시켜 얻을 수 있다. 본 실험례에서 소라페닙 내성 간암 환자에서 화학적 치료법의 치료 효과를 향상시키기 위해 간세포 암종(HCC)에 대한 표적 항암제인 소라페닙에 분자 생성 모델을

적용하였다.

[0095] 소라페닙 내성과 ABC 수송체 사이의 연관성

[0096] 소라페닙은 Raf/Mek/Erk 경로에서 종양 세포에서 세포 증식 및 혈관신생을 억제하는 단백질 키나아제의 억제제이다. 소라페닙의 중간 정도의 치료 효과와 은폐된 약물 내성으로 인해, 소라페닙의 대안으로 사용될 수 있는 신약 후보 물질의 발견은 중요한 연구 과제에 해당한다. 소라페닙 내성과 관련된 의심되는 메커니즘 중 하나는 세포에서 약물을 끌어내는 ATP 결합 카세트(ABC) 운반체이다. 소라페닙을 포함한 다중 표적 티로신 키나아제 억제제(TKI)는 ABC 수송체 기질로 작용하기 때문에, ABC 수송체는 소라페닙이 치료 표적 단백질에 결합하기 전에 HCC 종양 세포에서 소라페닙을 빼내는 것으로 분석된다. 따라서, 소라페닙의 치료 표적 단백질에 대한 친화력의 손실 없이 ABC 수송체 단백질에 대한 소라페닙의 결합 친화도를 감소시킨다면 간세포 암종 환자에서 소라페닙 내성을 완화함과 동시에 화학적 치료법의 효과를 높일 수 있다.

[0097] ABCG2 에 대한 결합 선호도 최적화

[0098] 소라페닙과 유사한 적응 발견을 위한 분자 생성 모델의 proof-of-concept 를 수행하기 위해, 본 실험실에서 소라페닙의 표적 키나아제인 세린/트레오닌-단백질 키나아제 B-raf (BRAF)에 대한 친화력 손실 없이 ABCG2(ABC subfamily G member 2)의 단백질에 대한 결합 친화도 스코어를 줄이면서 소라페닙의 하위 구조를 보존하는 것이 목표로 채택되었다. 이를 위해, ChEMBL 데이터베이스에서 16,000 여개의 SMILES 문자열을 선별 하고 분자 생성 모델 및 UGMMT 에 대한 학습 데이터셋이 구성되었다. 여기서, UGMMT 는 최신 SMILES 기반 모델이기 때문에 선택되었다.

[0099] 도 12 는 본 개시의 일 실시예에 따라 소라페닙을 소스 분자 모델로 사용하여 10,000 개의 분자 모델을 학습하고 생성한 후 성공률을 비교한 결과를

나타낸다. 성공률은 표적 분자 조건(Tanimoto 유사도 > 0.4 및 ABCG2 에 대한 친화도 스코어 < 4.7)을 만족하는 신규 분자 모델의 비율로 정의되었다. 분자 생성 모델(여기서, COMA)은 높은 성공률(0.174)을 보인 반면 UGMMT 는 낮은 성공률(0.001)을 보였다. UGMMT 가 분자 생성 모델보다 ABCG2 에 대한 결합 친화도를 더 많이 감소시켰지만 UGMMT 가 소라페넵과 유사한 분자를 생성하지 못했기 때문에 UGMMT 가 낮은 성공률을 보였다.

[00100] 도 13 은 본 개시의 일 실시예에 따라 분자 생성 모델로부터 육안으로도 소라페넵과 유사한 구조의 분자 모델이 생성된 결과를 나타낸다. 분자 생성 모델에 의해 생성된 분자들은 모두 표적 분자 조건(Tanimoto 유사도 > 0.4 및 ABCG2 에 대한 친화도 스코어 < 4.7)을 만족함을 확인했다.

[00101] 도 14A 및 14B 는 본 개시의 일 실시예에 따른 실험례에서 히트 후보가 소라페넵보다 ABCG2 에 대한 결합 에너지가 높은 지를 확인하기 위해 AutoDock Vina 를 사용하여 결합 에너지를 비교한 결과를 나타낸다. 도킹된 포즈는 AutoDock Vina 1.2.3 을 사용하여 평가되었고 Chimera 1.16 및 LigPlot Plus 2.2.5 를 사용하여 시각화되었다. 리간드(ligand)를 준비하기 위해 먼저 분자 생성 모델에 의해 생성된 10,000 개의 분자 중 19 개의 고유한 분자를 중복 분자 모델 제거를 통해 추출했고, Open Babel 3.1.1 을 사용하여 분자의 3 차원 좌표를 생성했다. 그리고 나서, pH 7.4 에서 분자를 양성자화하고 AutoDock 용 Python 라이브러리인 meeko 0.3.0 을 사용하여 pdbqt 파일을 만들었다. ABCG2 및 BRAF 수용체를 준비하기 위해 ABCG2 및 BRAF 에 대한 PDB 데이터베이스에서 각각 6VXH 및 1UWH 를 포함한 3D 구조 파일을 다운로드하고 ADFR 소프트웨어 1.0 을 사용하여 수소를 확인했다. 박스 중심과 크기를 정의하기 위해 Chimera 를 활용하고 수용체 리간드 쌍당 20 개의 포즈를 생성하기 위해 AutoDock Vina 를 실행했다. 그리고 나서, 수용체-리간드 쌍당 결합 에너지 스코어가 가장 높은 최상의 포즈를 선택하여 소라페넵과 비교했다.

도 14A 및 14B를 참고하면, 15개의 분자가 소라페닙보다 ABCG2에 대해 더 높은 결합 에너지를 가짐을 확인할 수 있다. 따라서 이러한 분자는 소라페닙의 대안으로 히트 후보 분자가 될 수 있다.

[00102] 도 15는 본 개시의 일 실시예에 따른 실험실에서 히트 후보가 소라페닙만큼 BRAF에 대한 결합 친화력이 강한 지를 확인하기 위해 Chimera를 사용하여 수용체-리간드 복합체의 3D 구조에 대한 그래픽을 도시한 결과 및 LigPlot Plus를 사용하여 수용체-리간드 복합체의 2D 구조에 대한 그래픽을 도시한 결과를 나타낸다. 도 15를 참고하면, 히트 후보 분자들이 BRAF에서 소라페닙에 대한 결합 포켓과 잘 맞는다는 것을 확인할 수 있다. 또한 생성된 히트 후보 분자와 소라페닙이 구조 분석 도구와 기본 매개변수를 사용하여 반 데르 발스 반경을 기반으로 공통 원자간 접촉을 가짐을 확인했다. 또한, LigPlot Plus에 의해 그려진 2D 플롯을 통해 분자가 Glu500(A) 및 Cys531(A)를 포함한 아미노산 잔기와 수소 결합을 갖고 BRAF에서 소라페닙과 상호 작용한다는 것을 확인할 수 있다.

[00103] 합성 접근성 평가

[00104] Scifinder-n의 역합성 분석을 이용하여 분자 생성 모델에 의해 생성된 분자의 합성 가능성을 평가하였다. 대부분의 분자는 두 단계로 합성될 수 있다. 이는 생성된 분자가 기존 약물 소라페닙과 유사하기 때문에 좋은 합성성을 보장할 수 있었기 때문이며, 분자 생성 모델과 같은 구조적 제약이 있는 분자 생성 모델이 목표 지향적 약물 발견을 위한 실제 작업에 효과적인 도구가 될 것임을 시사한다. 즉, *in silico* 분석 결과는 분자 생성 모델에 의해 생성된 소라페닙 유도체가 약물 내성이 높은 환자에서 소라페닙의 대체 약물 후보가 될 수 있음을 나타낸다.

[00105] 결론

[00106] 구조적 제약이 있는 분자 생성을 위한 AI 기반 생성 모델은 효과적인 약물 발견을 위한 솔루션일 뿐만 아니라 화학자 및 약리학자를 위한 강력하고 설명

가능한 도구가 될 수 있다. 기존의 구조 제약 분자 생성 모델은 화학적 속성 개선, 신규성, 소스 분자와의 높은 유사도를 동시에 만족시키는 분자를 생산하는데 한계가 있다. 본 개시의 분자 생성 모델은 두 가지 훈련 단계를 통해 높은 속성 개선과 높은 구조적 유사도를 모두 달성했다. 또한, 유사도 제약 및 속성 개선에 있어 분자 생성 모델은 DRD2, QED, plogP04 및 plogP06의 4 가지 벤치마크 데이터셋에서 다양한 최신 모델을 능가하는 성과를 나타냈다.

[00107] 구현 세부 정보

[00108] 본 개시의 분자 생성 모델은 Python 3.6 과 PyTorch 1.10.1 및 RDKit 2021.03.5 를 비롯한 여러 오픈 소스 도구를 사용하여 구현되었다. SMILES kekulization, SMILES 유효성 검사, Tanimoto 유사도 계산, QED 추정 에는 화학정보 학용 오픈 소스 도구 인 RDKit 이 사용되었다. 오픈 소스 기계 학습 프레임워크인 PyTorch 는 분자 생성 모델의 신경망을 구성하고 훈련하는 데 사용되었다. 모든 실험은 64GB 메모리와 GeForce RTX 3090 이 장착된 Ubuntu 18.04.6 LTS 에서 수행되었다.

[00109] 타니모토(Tanimoto) 유사도

[00110] 범위가 0 에서 1 인 Tanimoto 유사도는 Morgan 지문으로 표시되는 원자 쌍 및 위상 비트맵과 같은 분자 구조를 비교한다. 본 개시의 실험례에서 Morgan 지문은 반경이 2 이고 2048 비트인 RDKit 을 사용하여 생성된 이진 벡터이다. 해당 지문 벡터 $FP(x) = (p_1, p_2, \dots, p_{2048})$ 및 $FP(y) = (q_1, q_2, \dots, q_{2048})$ 이 있는 두 개의 SMILES 문자열 x 및 y 에 대해 Tanimoto 유사도 스코어는 수학식 1 에 따라 산출되었다.

[00111] 【수학식 1】

[00112]

$$\mathcal{T}(x, y) = \frac{\sum_{i=1}^{2048} p_i q_i}{\sum_{j=1}^{2048} (p_j + q_j - p_j q_j)}.$$

[00113] 결합 친화도 예측

[00114] ABCG2 및 BRAF 에 대한 결합 친화도 스코어를 예측하는 것은 소라페닙 내성에 대한 COMA 적용에 매우 중요하다. 본 개시의 실험례에서 가상 스크리닝을 위한 PyTorch 기반 라이브러리인 DeepPurpose 는 460 만 쌍 이상의 분자에 대한 정확하고 높은 처리량 친화도 예측에 사용되었다. 또한, UGMMT 및 COMA 에 대한 훈련 데이터셋을 생성하고 COMA 에서 강화 학습의 보상을 계산하기 위해 측정된 결합 친화도의 공개 데이터베이스인 BindingDB 에서 사전 훈련된 메시지 전달 및 컨볼루션 신경망으로 예측 모델을 활용했다.

[00115] 벤치마크 데이터셋

[00116] 이 연구에서는 표 1 에 제시된 종래에 제공된 4 개의 벤치마크 데이터셋과 소라페닙 내성에 대한 원본 데이터셋을 사용했다.

[00117] 【표 1】

[00118]

		DRD2	QED	pLogP04	pLogP06	Sorafenib
Number of Unique Items	Triplets (Src,Tar,Neg)	688040	1766120	1973800	1495400	4612380
	Pairs (Src,Tar)	34402	88306	98690	74770	230619
	Src	18490	38723	57856	67718	13840
	Tar	3141	13202	44759	69762	2340
	Neg	21632	51923	99066	132397	16180
Range of Tanimoto Similarity	(Src,Tar)	0.40 – 0.83	0.40 – 0.80	0.40 – 1.00	0.60 – 1.00	0.40 – 1.00
	(Src,Neg)	0.00 – 0.30	0.00 – 0.30	0.00 – 0.30	0.00 – 0.49	0.03 – 0.30
	(Tar,Neg)	0.00 – 0.30	0.00 – 0.30	0.00 – 0.30	0.00 – 0.49	0.03 – 0.30
Range of Property	Src	0.00 – 0.05	0.70 – 0.80	-62.52 – 1.66	-32.33 – 3.89	4.90 – 8.37
	Tar	0.50 – 1.00	0.90 – 0.95	-42.76 – 4.17	-30.63 – 5.48	3.39 – 4.70
	Difference (Tar - Src)	0.45 – 1.00	0.10 – 0.25	1.00 – 64.36	1.00 – 23.79	N/A

[00119] DRD2 데이터셋에는 ZINC 데이터베이스에서 파생된 DRD2 활동

스코어와 함께 34,000 여개의 분자 쌍(소스 및 타겟)이 포함되어 있다. DRD2 활동 스코어의 범위는 0 에서 1 까지이며 종래의 서포트 벡터 머신(Support Vector Machine) 회귀 모델을 사용하여 평가되었다. DRD2 데이터셋의 각 쌍에 대해 SMILES 문자열 쌍은 Tanimoto 유사도가 0.4 이상이고 소스 및 타겟 SMILES 문자열의 DRD2 스코어가 각각 0.05 미만 및 0.5 보다 큰 속성 제약 조건을 충족했다. QED 데이터셋에는 QED 스코어가 있는 ZINC 데이터베이스에서 파생된 88,000 여개의 분자 쌍이 포함되어 있다. QED 스코어의 범위는 0 에서 1 까지이며 RDKit 을 사용하여 계산되었다. QED 데이터셋의 각 쌍에 대해 두 SMILES 문자열 간의 Tanimoto 유사도는 0.4 이상이었고 소스 및 타겟의 QED 스코어는 각각 [0.7, 0.8] 및 [0.9, 1.0] 범위에 있었다. 페널티 logP04 및 페널티 logP06 데이터셋에는 각각 페널티 logP 스코어와 함께 ZINC 데이터베이스에서 파생된 98,000 여개 및 74,000 여개의 분자 쌍이 포함되어 있다. 페널티 logP 스코어 범위는 -63.0 에서 5.5 이다. 페널티가 적용된 logP04 데이터셋의 각 쌍에 대해 두 SMILES 문자열 간의 Tanimoto 유사도는 0.4 이상이였다. 페널티가 부여된 logP06 의 경우 유사도 임계값이 0.6 으로 설정되었다.

[00120] COMA 응용 사례를 소개하기 위해 소라페닙 유사 분자 생성을 위한 데이터셋을 구성했다. ABCG2 의 활성이 간세포 암종에서 소라페닙 내성과 관련된다는 관찰에 기초하여, 이 적용은 ABCG2 에 대해 더 낮은 결합 친화도를 갖는 소라페닙 유사 분자를 생성하는 동시에 표적 키나아제 BRAF 에 대한 친화도 수준을 보존하는 것을 목표로 했다. 이 데이터셋 에는 ChEMBL 데이터베이스에서 파생된 23 만여개의 분자 쌍이 ABCG2 및 BRAF 에 대한 결합 친화도 스코어와 함께 포함되어 있다. DeepPurpose 를 사용하여 평가한 결합 친화도 스코어는 pKd 였다. ABCG2 데이터셋의 각 쌍에 대해 두 분자 간의 Tanimoto 유사도는 0.4 이상이였고 소스와 타겟의 ABCG2 친화도 값은 각각 [4.9, 8.4] 및 [3.3, 4.7] 범위였다. BRAF 의

경우 소스와 대상 모두 결합 친화도가 6.0 보다 컸다.

[00121] 도 16 은 본 개시의 일 실시예에 따른 분자 생성 모델을 위한 학습 방법(1600)의 흐름도이다. 방법(1600)은 컴퓨팅 장치의 적어도 하나의 프로세서(예: 프로세서(414))에 의해 수행될 수 있다. 한편 도시된 바와 같이, 방법(1600)은 소스 분자 모델, 소스 분자 모델과의 구조적 유사도가 제 1 임계치 초과인 타겟 분자 모델, 소스 분자 모델 또는 타겟 분자 모델 중 하나 이상의 모델과의 구조적 유사도가 제 1 임계치 이하인 네거티브 분자 모델을 포함하는 학습 데이터셋을 획득하는 단계(S1610)로 개시될 수 있다.

[00122] 프로세서는 학습 데이터셋 및 제 1 손실 함수를 기초로, 소스 분자 모델과 타겟 분자 모델 사이의 거리를 조정하도록 분자 생성 모델을 학습시킬 수 있다(S1620). 예를 들어, 프로세서는 학습 데이터셋 및 제 1 손실 함수를 기초로, 소스 분자 모델과 타겟 분자 모델 사이의 거리가 가까워지도록 분자 생성 모델을 학습시킬 수 있다.

[00123] 프로세서는 학습 데이터셋 및 제 1 손실 함수와 상이한 제 2 손실 함수를 기초로, 소스 분자 모델과 네거티브 분자 모델 사이의 거리 및 타겟 분자 모델과 네거티브 분자 모델 사이의 거리 중 적어도 하나의 거리를 조정하도록 분자 생성 모델을 학습시킬 수 있다(S1630). 예를 들어, 프로세서는 학습 데이터셋 및 제 1 손실 함수와 상이한 제 2 손실 함수를 기초로, 소스 분자 모델과 네거티브 분자 모델 사이의 거리 및 타겟 분자 모델과 네거티브 분자 모델 사이의 거리 중 적어도 하나의 거리가 멀어지도록 분자 생성 모델을 학습시킬 수 있다.

[00124] 추가적으로, 프로세서는 학습 데이터셋 및 보상 함수를 기초로, 소스 분자 모델로부터 소스 분자 모델과의 구조적 유사도가 제 2 임계치 초과인 분자 모델이 출력되도록, 분자 생성 모델을 학습시킬 수도 있다. 예를 들어, 프로세서는 분자 생성 모델에 소스 분자 모델을 입력하여 출력 분자 모델을 획득하는 단계 및

출력 분자 모델과 소스 분자 모델을 비교한 결과 출력 분자 모델과 소스 분자 모델 사이의 구조적 유사도가 제 2 임계치를 초과하는지 여부를 기초로, 출력 분자 모델과 연관된 포지티브 가중치 또는 네거티브 가중치를 산출하여 분자 생성 모델에 부여할 수 있다. 다른 예를 들어, 프로세서는 출력 분자 모델과 소스 분자 모델을 비교한 결과 출력 분자 모델과 소스 분자 모델 사이의 구조적 유사도가 제 2 임계치를 초과하는지 여부 및 출력 분자 모델의 화학적 속성 스코어가 소스 분자 모델의 화학적 속성 스코어를 초과하는지 여부를 기초로, 출력 분자 모델과 연관된 포지티브 가중치 또는 네거티브 가중치를 산출하여 분자 생성 모델에 부여할 수도 있다. 또 다른 예를 들어, 프로세서는 학습 데이터셋 및 보상 함수를 기초로, 소스 분자 모델로부터, 소스 분자 모델과의 구조적 유사도가 제 2 임계치 초과이고, 소스 분자 모델의 화학적 속성 스코어보다 큰 화학적 속성 스코어를 갖는 분자 모델이 출력되도록, 분자 생성 모델을 학습시킬 수 있다.

[00125] 한편, 분자 생성 모델은 타겟 분자 모델의 화학적 속성 스코어가 소스 분자 모델의 화학적 속성 스코어를 초과하도록 구성될 수 있다.

[00126] 본 개시의 다양한 수정예들이 통상의 기술자들에게 쉽게 자명할 것이고, 본원에 정의된 일반적인 원리들은 본 개시의 취지 또는 범위를 벗어나지 않으면서 다양한 변형예들에 적용될 수도 있다. 따라서, 본 개시는 본원에 설명된 예들에 제한되도록 의도된 것이 아니고, 본원에 개시된 원리들 및 신규한 특징들과 일관되는 최광의의 범위가 부여되도록 의도된다.

[00127] 비록 예시적인 구현예들이 하나 이상의 독립형 컴퓨터 시스템의 맥락에서 현재 개시된 주제의 양태들을 활용하는 것을 언급할 수도 있으나, 본 주제는 그렇게 제한되지 않고, 오히려 네트워크나 분산 컴퓨팅 환경과 같은 임의의 컴퓨팅 환경과 연계하여 구현될 수도 있다. 또 나아가, 현재 개시된 주제의 양상들은 복수의 프로세싱 칩들이나 디바이스들에서 또는 그들에 걸쳐 구현될 수도 있고, 스토리지는

복수의 디바이스들에 걸쳐 유사하게 영향을 받게 될 수도 있다. 이러한 디바이스들은 PC 들, 네트워크 서버들, 및 핸드헬드 디바이스들을 포함할 수도 있다.

[00128] 본 명세서에서는 본 개시가 일부 실시예들과 관련하여 설명되었지만, 본 발명이 속하는 기술분야의 통상의 기술자가 이해할 수 있는 본 개시의 범위를 벗어나지 않는 범위에서 다양한 변형 및 변경이 이루어질 수 있다는 점을 알아야 할 것이다. 또한, 그러한 변형 및 변경은 본 명세서에서 첨부된 특허청구의 범위 내에 속하는 것으로 생각되어야 한다.

WHAT IS CLAIMED IS:

1. 소스 분자 모델, 상기 소스 분자 모델과의 구조적 유사도가 제 1 임계치 초과인 타겟 분자 모델, 상기 소스 분자 모델 또는 상기 타겟 분자 모델 중 하나 이상의 모델과의 구조적 유사도가 상기 제 1 임계치 이하인 네거티브 분자 모델을 포함하는 학습 데이터셋을 획득하는 단계;

상기 학습 데이터셋 및 제 1 손실 함수를 기초로, 상기 소스 분자 모델과 상기 타겟 분자 모델 사이의 거리를 조정하도록 분자 생성 모델을 학습시키는 단계; 및

상기 학습 데이터셋 및 상기 제 1 손실 함수와 상이한 제 2 손실 함수를 기초로, 상기 소스 분자 모델과 상기 네거티브 분자 모델 사이의 거리 및 상기 타겟 분자 모델과 상기 네거티브 분자 모델 사이의 거리 중 적어도 하나의 거리를 조정하도록 상기 분자 생성 모델을 학습시키는 단계

를 포함하는, 적어도 하나의 프로세서에 의해 수행되는 분자 생성 모델을 위한 학습 방법.

2. 제 1 항에 있어서,

상기 학습 데이터셋 및 제 1 손실 함수를 기초로, 상기 소스 분자 모델과 상기 타겟 분자 모델 사이의 거리를 조정하도록 분자 생성 모델을 학습시키는 단계는,

상기 학습 데이터셋 및 제 1 손실 함수를 기초로, 상기 소스 분자 모델과 상기 타겟 분자 모델 사이의 거리가 가까워지도록 상기 분자 생성 모델을 학습시키는 단계

를 포함하는, 적어도 하나의 프로세서에 의해 수행되는 분자 생성 모델을 위한 학습 방법.

3. 제 1 항에 있어서,

상기 학습 데이터셋 및 상기 제 1 손실 함수와 상이한 제 2 손실 함수를 기초로, 상기 소스 분자 모델과 상기 네거티브 분자 모델 사이의 거리 및 상기 타겟 분자 모델과 상기 네거티브 분자 모델 사이의 거리 중 적어도 하나의 거리를 조정하도록 상기 분자 생성 모델을 학습시키는 단계는,

상기 학습 데이터셋 및 상기 제 1 손실 함수와 상이한 제 2 손실 함수를 기초로, 상기 소스 분자 모델과 상기 네거티브 분자 모델 사이의 거리 및 상기 타겟 분자 모델과 상기 네거티브 분자 모델 사이의 거리 중 적어도 하나의 거리가 멀어지도록 상기 분자 생성 모델을 학습시키는 단계

를 포함하는, 적어도 하나의 프로세서에 의해 수행되는 분자 생성 모델을 위한 학습 방법.

4. 제 1 항에 있어서,

상기 학습 데이터셋 및 보상 함수를 기초로, 상기 소스 분자 모델로부터 상기 소스 분자 모델과의 구조적 유사도가 제 2 임계치 초과인 분자 모델이 출력되도록, 상기 분자 생성 모델을 학습시키는 단계

를 더 포함하는, 적어도 하나의 프로세서에 의해 수행되는 분자 생성 모델을 위한 학습 방법.

5. 제 4 항에 있어서,

상기 학습 데이터셋 및 보상 함수를 기초로, 상기 소스 분자 모델로부터 상기 소스 분자 모델과의 구조적 유사도가 제 2 임계치 초과인 분자 모델이 출력되도록, 상기 분자 생성 모델을 학습시키는 단계는,

상기 분자 생성 모델에 상기 소스 분자 모델을 입력하여 출력 분자 모델을 획득하는 단계; 및

상기 출력 분자 모델과 상기 소스 분자 모델을 비교한 결과 상기 출력 분자 모델과 상기 소스 분자 모델 사이의 구조적 유사도가 상기 제 2 임계치를 초과하는지 여부를 기초로, 상기 출력 분자 모델과 연관된 포지티브 가중치 또는 네거티브 가중치를 산출하여 상기 분자 생성 모델에 부여하는 단계

를 포함하는, 적어도 하나의 프로세서에 의해 수행되는 분자 생성 모델을 위한 학습 방법.

6. 제 5 항에 있어서,

상기 출력 분자 모델과 상기 소스 분자 모델을 비교한 결과 상기 출력 분자 모델과 상기 소스 분자 모델 사이의 구조적 유사도가 상기 제 2 임계치를 초과하는지 여부를 기초로, 상기 출력 분자 모델과 연관된 포지티브 가중치 또는 네거티브 가중치를 산출하여 상기 분자 생성 모델에 부여하는 단계는,

상기 출력 분자 모델과 상기 소스 분자 모델을 비교한 결과 상기 출력 분자 모델과 상기 소스 분자 모델 사이의 구조적 유사도가 상기 제 2 임계치를 초과하는지 여부 및 상기 출력 분자 모델의 화학적 속성 스코어가 상기 소스 분자 모델의 화학적 속성 스코어를 초과하는지 여부를 기초로, 상기 출력 분자 모델과 연관된 포지티브 가중치 또는 네거티브 가중치를 산출하여 상기 분자 생성 모델에 부여하는 단계

를 포함하는, 적어도 하나의 프로세서에 의해 수행되는 분자 생성 모델을 위한 학습 방법.

7. 제 4 항에 있어서,

상기 학습 데이터셋 및 보상 함수를 기초로, 상기 소스 분자 모델로부터 상기 소스 분자 모델과의 구조적 유사도가 제 2 임계치 초과인 분자 모델이 출력되도록, 상기 분자 생성 모델을 학습시키는 단계는,

상기 학습 데이터셋 및 상기 보상 함수를 기초로, 상기 소스 분자 모델로부터, 상기 소스 분자 모델과의 구조적 유사도가 상기 제 2 임계치 초과이고, 상기 소스 분자 모델의 화학적 속성 스코어보다 큰 화학적 속성 스코어를 갖는 분자 모델이 출력되도록, 상기 분자 생성 모델을 학습시키는 단계

를 포함하는, 적어도 하나의 프로세서에 의해 수행되는 분자 생성 모델을 위한 학습 방법.

8. 제 1 항에 있어서,

상기 타겟 분자 모델의 화학적 속성 스코어는 상기 소스 분자 모델의 화학적 속성 스코어보다 큰, 적어도 하나의 프로세서에 의해 수행되는 분자 생성 모델을 위한 학습 방법.

9. 제 1 항에 따른 분자 생성 모델을 위한 학습 방법을 실행시키도록 컴퓨터로 판독 가능한 기록매체에 기록된 컴퓨터 프로그램.

10. 분자 생성 모델과 연관된 데이터를 저장하는 메모리; 및

상기 메모리와 연결되어 상기 분자 생성 모델을 학습시키는 적어도 하나의 프로세서

를 포함하고, 상기 적어도 하나의 프로세서는

소스 분자 모델, 상기 소스 분자 모델과의 구조적 유사도가 제 1 임계치 초과인 타겟 분자 모델, 상기 소스 분자 모델 또는 상기 타겟 분자 모델 중 하나 이상의 모델과의 구조적 유사도가 상기 제 1 임계치 이하인 네거티브 분자 모델을 포함하는 학습 데이터셋을 획득하는 것,

상기 학습 데이터셋 및 제 1 손실 함수를 기초로, 상기 소스 분자 모델과 상기 타겟 분자 모델 사이의 거리를 조정하도록 분자 생성 모델을 학습시키는 것과

상기 학습 데이터셋 및 상기 제 1 손실 함수와 상이한 제 2 손실 함수를 기초로, 상기 소스 분자 모델과 상기 네거티브 분자 모델 사이의 거리 및 상기 타겟 분자 모델과 상기 네거티브 분자 모델 사이의 거리 중 적어도 하나의 거리를 조정하도록 상기 분자 생성 모델을 학습시키는 것을 실행하도록 구성된 명령어들을 포함하는, 분자 생성 모델을 위한 학습 장치.

11. 제 2 항에 따른 분자 생성 모델을 위한 학습 방법을 실행시키도록 컴퓨터로 판독 가능한 기록매체에 기록된 컴퓨터 프로그램.

12. 제 3 항에 따른 분자 생성 모델을 위한 학습 방법을 실행시키도록 컴퓨터로 판독 가능한 기록매체에 기록된 컴퓨터 프로그램.

13. 제 4 항에 따른 분자 생성 모델을 위한 학습 방법을 실행시키도록 컴퓨터로 판독 가능한 기록매체에 기록된 컴퓨터 프로그램.

14. 제 5 항에 따른 분자 생성 모델을 위한 학습 방법을 실행시키도록 컴퓨터로 판독 가능한 기록매체에 기록된 컴퓨터 프로그램.

15. 제 6 항에 따른 분자 생성 모델을 위한 학습 방법을 실행시키도록 컴퓨터로 판독 가능한 기록매체에 기록된 컴퓨터 프로그램.

16. 제 7 항에 따른 분자 생성 모델을 위한 학습 방법을 실행시키도록 컴퓨터로 판독 가능한 기록매체에 기록된 컴퓨터 프로그램.

17. 제 8 항에 따른 분자 생성 모델을 위한 학습 방법을 실행시키도록 컴퓨터로 판독 가능한 기록매체에 기록된 컴퓨터 프로그램.

ABSTRACT

본 개시의 일 실시예에 따르면, 분자 생성 모델을 위한 학습 방법은 소스 분자 모델, 소스 분자 모델과의 구조적 유사도가 제 1 임계치 초과인 타겟 분자 모델, 소스 분자 모델 또는 타겟 분자 모델 중 하나 이상의 모델과의 구조적 유사도가 제 1 임계치 이하인 네거티브 분자 모델을 포함하는 학습 데이터셋을 획득하는 단계, 학습 데이터셋 및 제 1 손실 함수를 기초로, 소스 분자 모델과 타겟 분자 모델 사이의 거리를 조정하도록 분자 생성 모델을 학습시키는 단계 및 학습 데이터셋 및 제 1 손실 함수와 상이한 제 2 손실 함수를 기초로, 소스 분자 모델과 네거티브 분자 모델 사이의 거리 및 타겟 분자 모델과 네거티브 분자 모델 사이의 거리 중 적어도 하나의 거리를 조정하도록 분자 생성 모델을 학습시키는 단계를 포함할 수 있다.