



- (51) **International Patent Classification:**
G06N 99/00 (2019.01) *G06N 7/00* (2006.01)
- (21) **International Application Number:**
PCT/US2021/013451
- (22) **International Filing Date:**
14 January 2021 (14.01.2021)
- (25) **Filing Language:** English
- (26) **Publication Language:** English
- (30) **Priority Data:**
62/961,112 14 January 2020 (14.01.2020) US
- (71) **Applicant: FLAGSHIP PIONEERING INNOVATIONS VI, LLC** [US/US]; 55 Cambridge Parkway, 8th Floor, Cambridge, MA 02142 (US).
- (72) **Inventors: ESER, Umut;** Flagship Pioneering Innovations VI, LLC, 55 Cambridge Parkway, 8th Floor, Cambridge, MA 02142 (US). **WOLF, Fabian, Alexander;** Flagship Pioneering Innovations VI, LLC, 55 Cambridge Parkway, 8th Floor, Cambridge, MA 02142 (US). **PLUGIS, Nicholas, McCarthney;** Flagship Pioneering Innovations VI, LLC, 55 Cambridge Parkway, 8th Floor, Cambridge, MA 02142 (US).
- (74) **Agent: ALTIERI, Stephen, L. et al.;** Morgan, Lewis & Bockius LLP, 1111 Pennsylvania Avenue, Nw, Washington, DC 20004 (US).
- (81) **Designated States** (*unless otherwise indicated, for every kind of national protection available*): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, IT, JO, JP, KE, KG, KH, KN,

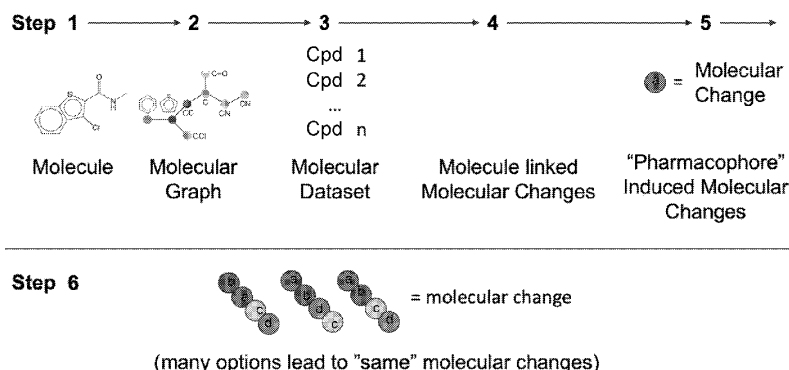
(54) **Title:** MOLECULE DESIGN

Figure 3

(57) **Abstract:** Systems and methods of discovering compounds with biological properties are provided. A first training dataset is obtained, including chemical structures and biological properties. Projections of compounds are obtained by projecting chemical structure information into a latent representation space using encoder weights. Compounds are classified by inputting projections into the classifier using classifier weights. The encoder and classifier are trained by comparing the classification of each compound to actual biological properties and updating the respective weights. A second training dataset is obtained including chemical structures. Projections of compounds are obtained by projecting chemical structure information into a latent representation space using encoder weights. Chemical structures are obtained by inputting projections into a decoder using decoder weights. The decoder is trained by comparing outputted and actual chemical structures and updating the respective weights. Candidate compounds not present in the first and second

[Continued on next page]

KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.

(84) Designated States (*unless otherwise indicated, for every kind of regional protection available*): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:

— with international search report (Art. 21(3))