

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property
Organization
International Bureau

(43) International Publication Date
17 August 2023 (17.08.2023)



(10) International Publication Number
WO 2023/154943 A1

(51) International Patent Classification:

G01N 33/68 (2006.01) G01N 30/72 (2006.01)
G06N 20/00 (2019.01) G01N 30/86 (2006.01)

(21) International Application Number:

PCT/US2023/062542

(22) International Filing Date:

14 February 2023 (14.02.2023)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

63/310,101	14 February 2022 (14.02.2022)	US
63/311,932	18 February 2022 (18.02.2022)	US
63/482,683	01 February 2023 (01.02.2023)	US

(71) Applicant: VENN BIOSCIENCES CORPORATION

[US/US]; 2 TOWER PLACE, 5TH FLOOR, South San Francisco, 94080 (US).

(72) Inventors: LIANG, Zhewei; 30 Kenton Avenue, San Carlos, California 94070 (US). LIU, Mingqi; 359 Northam Ave, San Carlos, California 94070 (US).

(74) Agent: SISTRUNK, Melissa L.; Norton Rose Fulbright US LLP, 555 California Street, Suite 3300, San Francisco, CA 94104 (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, CV, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, MG, MK, MN, MW, MX, MY, MZ, NA, NG,

(54) Title: DE NOVO GLYCOPEPTIDE SEQUENCING

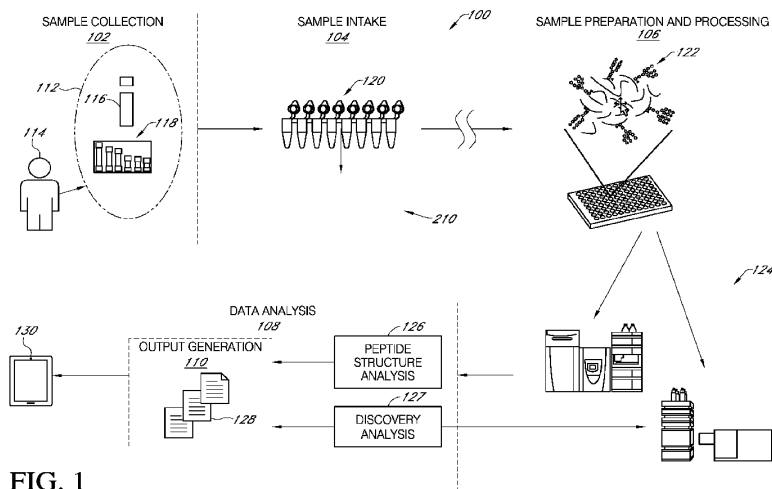


FIG. 1

(57) Abstract: A system and method of training a machine learning model to predict glycopeptide fragmentation patterns and retention times. Spectral data for a plurality of fragments of a glycopeptide structure is received. Glycan fragment composition data is generated using the spectral data. The glycan fragment composition data identifies a plurality of composition codes and a plurality of total intensities for a plurality of glycan fragments identified from the plurality of fragments using the spectral data. A linear glycan sequence is created using the glycan fragment composition data. A training input is formed for a machine learning model using the linear glycan sequence. The machine learning model is trained using the training input to predict a fragmentation pattern and a retention time for the glycopeptide structure.

[Continued on next page]

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, CV, 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, ME, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Declarations under Rule 4.17:

- *as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))*
- *as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii))*
- *of inventorship (Rule 4.17(iv))*

Published:

- *with international search report (Art. 21(3))*