GLP-1 2016 backtest IP Portfolio Analysis

V3 KC Scoring Workflow (RAG-augmented, SS-F/OS-S firewalled, prompt-cached per-sub-dimension model routing)

Executive Summary

This report applies the rDNA.ai four-stage KC scoring pipeline to a PRE-COMMITTED, LOCKED 584-patent substrate of GLP-1 / GIP / glucagon incretin-agonist USPTO biopharma patents with publication date on or before 2016-12-31. This is a predictive-validity BACKTEST: V3 KC scoring is being applied to a frozen 2016-snapshot corpus to test whether the V3 methodology would have correctly identified the IP foundations of commercially successful GLP-1 agonists prior to those commercial outcomes (semaglutide 2017 approval, tirzepatide 2022, and the dual/triple-agonist clinical wave currently in phase III). The substrate spans single GLP-1 agonists, dual GLP-1/glucagon and GLP-1/GIP agonists, triple GLP-1/GIP/glucagon agonists, foundational half-life-extension platforms (Hanmi LAPS, PhaseBio ELP, Ambrx unnatural-amino-acid, MannKind Technosphere, Intarcia osmotic implant, Fc-fusion, albumin-binding fatty-acid acylation), combination products, oral / SNAC / depot formulations, and competing pre-2016 non-validated peptide IP. Each patent is scored on six Key Component (KC) dimensions on a 1–5 scale and aggregated into a four-scheme ensemble (Meta Baseline / AbbVie / Amgen / Genentech weighting). This is the V3 BACKTEST variant carrying the same methodology corrections as HER3/TL1A/TROP2 corrected (2026-05-25 methodology audit): (a) the KC2/KC3/KC4/KC6 composite prompts embed the V2 holistic-judgment guidance verbatim with a KC4 prior-art landscape anchor specific to the 2016 GLP-1 commercial landscape (exenatide/liraglutide/lixisenatide crowded; semaglutide/dulaglutide moderately crowded; Indiana DiMarchi dual-agonist class sparse), (b) no character truncation is applied to the patent text submitted to the model, (c) TA-bucket multipliers are not applied to the patent-level ranking (display groupings only), and (d) sub-dimensions D3 (Prosecution Flexibility) and A4 (Commercial Differentiation) are forced to Opus 4.7 to avoid the Sonnet rubric-collapse failure mode identified in HER2 corrected QA.

Portfolio-level results are summarized below by therapeutic-area cut, by top assignee, and at the patent level for the calibrated top decile after a V3 pairwise-tournament re-judging pass. Two acknowledged corpus limitations: (1) tirzepatide composition-of-matter parent patents post-date the 2016 corpus floor and are not in the substrate — the tirzepatide direction is tested indirectly via construct-validity on the Indiana University DiMarchi dual/triple-agonist precursor class (16 patents); (2) dulaglutide CoM-parent coverage is narrow (2–3 patents) and is reported as binary corroboration rather than statistical hit-rate.

The Sonnet 4.5 Quarantine firewall screened all 584 candidate patents and flagged 0 for quarantine. This is a research-pipeline application of the V3 KC scoring rubric (RAG-augmented, SS-F / OS-S firewalled, per-patent prompt-cached, per-sub-dimension model-routed) to a frozen 2016 biopharma corpus for predictive-validity testing against commercial outcomes that materialized in 2017–2026.

Key Component (KC) Scoring Framework

Each patent is scored on six Key Component (KC) dimensions on a 1–5 scale. The ensemble score is the mean of the six dimensions (with confidence and standard deviation tracked alongside).

- **KC1** measures abstract and therapeutic specificity: target specificity, therapeutic positioning, structural disclosure, and technical differentiation.
- **KC2** measures independent-claim quality: scope, structure, enforceability, and enablement vulnerability.
- **KC3** measures dependent-claim quality: fallback hierarchy, embodiment coverage, prosecution flexibility, and structural diversity.

- **KC4** measures novelty and differentiation: prior-art positioning, inventive-step quality, and field positioning.
- **KC5** measures specification enablement: enablement gap, structure-function insight, data breadth, and translational completeness.
- **KC6** measures working examples: working/prophetic example quality, structural diversity, experimental rigor, and in vivo / translational validation.

Retrieval-Augmented Generation (RAG) for KC4 Novelty Scoring

The V3 rubric introduces a corpus-internal retrieval-augmented generation (RAG) step that supplies every KC4 sub-dimension (N1 Prior-Art Landscape, N2 Claims Differentiation, N3 Inventive-Step Quality, N4 Field Positioning) with a list of the nearest prior patents in the local Biopharma corpus. The neighbour list is passed alongside the cached full patent text so the scoring LLM judges novelty against actual nearby filings rather than its internal world knowledge alone.

RAG construction:

- Embedding model: text-embedding-3-small (dim=1536).
- Each Biopharma full-text specification was chunked into ≤ 7,000-char blocks across the abstract, claims, and description sections, then each chunk was embedded; a patent-level centroid was computed and L2-normalised.
- Index size: 1579 Biopharma specifications → 85,352 chunks → 1579 patent centroids.
- Retrieval: per target patent, return top-20 nearest patent centroids by cosine similarity (self excluded); the result list is passed to the KC4 sub-dim prompts as `rag_neighbors`. Sub-dims must echo back `rag_neighbors_cited` in their JSON response.

RAG results across the corpus:

- Average nearest-neighbour cosine similarity across all scored patents: 0.965.
- Average similarity of the top-20 neighbourhood: 0.890 (a denser neighbourhood pulls KC4 N1/N2 scores down toward 'crowded'; sparser pushes them up toward 'pioneering').
- Corpus average composite KC4 score (post-RAG): 2.26 on the 1–5 scale, covering sub-dims N1, N2, N3, N4.
- The corpus has high internal similarity (mean nearest-neighbour cosine ≥ 0.70), indicating a relatively concentrated GLP-1 prior-art landscape; KC4 N1/N2 scoring therefore skews toward 'crowded' for many patents.
- The KC4 sub-dim JSON responses include the `rag_neighbors_cited` field so that each novelty score can be traced back to specific neighbour patents from the corpus index.

Most-frequently-cited prior-art neighbours in the corpus:

Rank	Neighbour Patent	Times Cited in Other Patents' Top-20
1	US20150166627A1	69
2	US20160108102A1	61
3	US8951962B2	61
4	US8057822B2	60
5	US20130244931A1	59
6	US20100305032A1	58
7	US20150164997A1	58
8	US20090156478A1	57
9	US7691963B2	57
10	US20080207507A1	56

11	US8445647B2	55
12	US20130053315A1	54
13	US8895694B2	54
14	US9409966B2	53
15	US20120295847A1	53

Portfolio-Level Metrics

• Total GLP-1 2016 backtest USPTO records ingested: 2,445 (USPTO PPubs full-text fetch via searchWithBeFamily, publication date 1995-01-01 to 2016-12-31 (2016 backtest snapshot); 58 GLP-1 / GLP1 / glucagon-like-peptide-1 / GIP / glucagon / dual / triple / co-agonist / oxyntomodulin / exenatide / Byetta / Bydureon / liraglutide / Victoza / Saxenda / semaglutide / Ozempic / Wegovy / Rybelsus / dulaglutide / Trulicity / lixisenatide / Adlyxin / Lyxumia / albiglutide / Tanzeum / Hanmi LAPS / Zealand / Indiana DiMarchi / MannKind Technosphere / Intarcia / fatty-acid acylation / Fc fusion discovery queries (expanded from the V2 corpus).)

• Title- and full-text-based classification: Biopharma=1575, Diagnostic=131, Device=116, Non_Biopharma=78

• Total Biopharma patents scored: 584

• Average KC scores — KC1: 3.99 KC2: 3.25 KC3: 3.29 KC4: 2.26 KC5: 3.91 KC6: 3.36

• Average ensemble mean across portfolio: 3.263

Investment-grade distribution:

• TIER_2_STRONG: 1

• TIER_3_WATCHLIST: 193

• LOW_PRIORITY: 390

Top Assignees

Rank	Assignee	Patent Count
1	NOVO NORDISK A/S	76
2	AMYLIN PHARMACEUTICALS, INC	65
3	BRISTOL-MYERS SQUIBB COMPANY	47
4	SANOFI	30
5	PFIZER INC	25
6	SANOFI-AVENTIS DEUTSCHLAND GMBH	24
7	MERCK SHARP & DOHME CORP	14
8	ASTRAZENECA PHARMACEUTICALS LP; AMYLIN PHARMACEUTICALS, LLC	13
9	MANNKIND CORPORATION	13
10	AMYLIN PHARMACEUTICALS, LLC; ASTRAZENECA PHARMACEUTICALS LP	13
11	DUKE UNIVERSITY	12
12	INDIANA UNIVERSITY RESEARCH AND TECHNOLOGY CORPORATION	11
13	BOEHRINGER INGELHEIM INTERNATIONAL GMBH	8
14	INTARCIA THERAPEUTICS, INC	8
15	ZEALAND PHARMA A/S	7

Therapeutic Area Analysis (n ≥ 5 patents)

For each area with at least 5 patents the canonical report shows average KC1–KC6 (1–5 scale), average ensemble mean, average TA-weighted score, and tier distribution. TA buckets are display groupings only in this V3 BACKTEST run. The eight GLP-1 TA buckets (GLP1_Triple_Agonist, GLP1_Dual_Agonist, GLP1_Foundational_Platform, GLP1_Combination_Product, GLP1_Formulation_Delivery, GLP1_Single_Agonist_CoM, GLP1_Adjacent_Therapeutic, GLP1_Adjacent_Mechanism) carry no numerical multipliers (TA_MULTIPLIERS_APPLIED = False). ta_weighted_score in the CSV outputs is identical to ensemble_mean. Backtest predictive-validity testing is performed on ensemble_mean (and on tournament-calibrated rankings within the top decile), with no commercial-preference reweighting.

Therapeutic Area	N	KC 1	KC 2	KC 3	KC 4	KC 5	KC 6	Ens Mean	TA-W	T1 Premium	T1 High	T2	T3 / Low
GLP1_Dual_Agonist	228	4.07	3.47	3.54	2.22	4.07	3.45	3.379	3.379	0	1	103	124
GLP1_Single_Agonist_CoM	149	3.88	3.09	3.06	2.26	3.78	3.13	3.138	3.138	0	0	37	112
GLP1_Combination_Product	106	3.98	3.08	3.18	2.17	3.85	3.46	3.187	3.187	0	0	20	86
GLP1_Foundational_Platform	69	3.96	3.13	2.99	2.48	3.77	3.30	3.218	3.218	0	0	18	51
GLP1_Triple_Agonist	23	4.09	3.39	3.78	2.35	4.00	3.74	3.446	3.446	0	0	12	11
GLP1_Formulation_Delivery	9	3.78	3.00	3.11	2.44	3.78	3.22	3.156	3.156	0	0	3	6

Therapeutic Area Narratives

Each TA narrative below summarises the calibrated cohort: average ensemble mean, the strongest and weakest KC dimensions for the TA, TIER_1_PREMIUM count, and the top three patents by TA-weighted score. Every KC value shown is the rolled-up output of the V3 structured-citation pipeline — each underlying patent has six KC composite scores, each grounded in four per-sub-dimension JSON citations carrying the source model, the model-supplied reason, and a verbatim quote from the patent text. A complete structured-citation example is presented in the Top-Decile section below.

GLP1_Dual_Agonist (n=228)

Average ensemble mean: 3.379. Highest KC: KC5 (avg 4.07). Lowest KC: KC4 (avg 2.22). TIER_1_PREMIUM: 0. Top 3 patents (by TA-weighted score): US20130053311A1; US8815802B2; US8648041B2.

GLP1_Single_Agonist_CoM (n=149)

Average ensemble mean: 3.138. Highest KC: KC1 (avg 3.88). Lowest KC: KC4 (avg 2.26). TIER_1_PREMIUM: 0. Top 3 patents (by TA-weighted score): US20090111730A1; US20150344540A1; US20080249089A1.

GLP1_Combination_Product (n=106)

Average ensemble mean: 3.187. Highest KC: KC1 (avg 3.98). Lowest KC: KC4 (avg 2.17). TIER_1_PREMIUM: 0. Top 3 patents (by TA-weighted score): US8318762B2; US20140221282A1; US8178541B2.

GLP1_Foundational_Platform (n=69)

Average ensemble mean: 3.218. Highest KC: KC1 (avg 3.96). Lowest KC: KC4 (avg 2.48). TIER_1_PREMIUM: 0. Top 3 patents (by TA-weighted score): US7696168B2; US20110009315A1; US20100137207A1.

GLP1_Triple_Agonist (n=23)

Average ensemble mean: 3.446. Highest KC: KC1 (avg 4.09). Lowest KC: KC4 (avg 2.35). TIER_1_PREMIUM: 0. Top 3 patents (by TA-weighted score): US20140066370A1; US9127088B2; US8859491B2.

GLP1_Formulation_Delivery (n=9)

Average ensemble mean: 3.156. Highest KC: KC1 (avg 3.78). Lowest KC: KC4 (avg 2.44). TIER_1_PREMIUM: 0. Top 3 patents (by TA-weighted score): US20150150811A1; US20090035253A1; US20140206611A1.

Top-Decile Calibration: Pairwise Tournament Results

The top decile (n=59) of the GLP-1 2016 backtest biopharma cohort was re-judged against the V3 rubric using full-text specifications and a complete pairwise tournament. The V3 calibration replaced any prior uniform top-decile plateau with a discriminating ranking; calibrated ensemble distribution: mean 3.795, range 3.652–4.240.

Rank	Patent	Assignee	KC1	KC2	KC3	KC4	KC5	KC6	Cal Ens
1	US8815802B2	NOVO NORDISK A/S	5	4	5	2	5	5	4.075
2	US20130053311A1	NOVO NORDISK A/S	5	4	5	3	5	4	4.240
3	US7960349B2	BRISTOL-MYERS SQUIBB COMPANY	4	4	4	3	4	5	3.835
4	US9089538B2	ZEALAND PHARMA A/S	5	3	4	3	5	5	3.950
5	US8648041B2	NOVO NORDISK A/S	5	4	5	2	5	4	4.002
6	US20140066370A1	AMYLIN PHARMACEUTICALS, LLC	5	4	3	3	5	4	3.940
7	US20090111730A1	NOVO NORDISK A/S	4	4	5	3	5	4	4.015
8	US20110166321A1	NOVO NORDISK A/S	5	4	5	2	4	5	3.973
9	US9127088B2	INDIANA UNIVERSITY RESEARCH AND TECHN...	4	4	5	3	4	4	3.913
10	US20150344540A1	NOVO NORDISK A/S	4	4	4	3	5	5	3.938
11	USRE44186E	BRISTOL-MYERS SQUIBB COMPANY	4	4	5	3	4	3	3.840
12	US20130143793A1	ZEALAND PHARMA A/S	5	3	4	3	5	4	3.877
13	US8980830B2	INDIANA UNIVERSITY RESEARCH AND TECHN...	4	4	5	3	4	3	3.840
14	US8895694B2	NOVO NORDISK A/S	5	3	4	3	3	4	3.673
15	US8859491B2	INDIANA UNIVERSITY RESEARCH AND TECHN...	4	5	4	3	3	4	3.873
16	US20150150811A1	NOVO NORDISK A/S	4	4	4	3	5	4	3.865
17	US7696168B2	TUFTS MEDICAL CENTER, INC	5	3	3	4	3	4	3.760

18	US9180169B2	ZEALAND PHARMA A/S	4	4	4	2	5	5	3.700
19	US20120184489A1	SANOFI-AVENTIS DEUTSCHLAND GMBH	4	4	5	3	4	4	3.913
20	US9067977B2	NOVO NORDISK A/S	4	4	4	2	5	5	3.700
21	US20130079278A1	NOVO NORDISK A/S	4	4	4	3	4	4	3.763
22	US20150299281A1	ZEALAND PHARMA A/S/	4	3	4	3	5	4	3.652
23	US20110082079A1	NOVO NORDISK A/S	5	4	3	3	3	4	3.735
24	US20140213513A1	SANOFI	5	4	4	2	5	4	3.853
25	US20140100156A1	SANOFI	4	5	4	2	5	4	3.840
26	US20150072924A1	OPKO BIOLOGICS LTD	4	4	4	3	4	4	3.763
27	US8633156B2	SANOFI-AVENTIS DEUTSCHLAND GMBH	4	4	4	3	4	4	3.763
28	US20150320871A1	INDIANA UNIVERSITY RESEARCH AND TECHN...	4	4	4	3	4	3	3.690
29	US20140080756A1	PFIZER INC	4	4	4	3	5	4	3.865
30	US9012398B2	NOVO NORDISK A/S	4	4	4	2	5	5	3.700
31	US8420779B2	AMGEN INC	5	4	4	2	4	4	3.750
32	US20140221282A1	ASTRAZENECA PHARMACEUTICALS LP (+1)	4	4	3	3	5	4	3.715
33	US20150374794A1	NOVO NORDISK A/S	4	4	5	2	4	4	3.675
34	US8361959B2	MERCK SHARP & DOHME CORP	4	5	5	2	5	3	3.917
35	US20130244931A1	NOVO NORDISK A/S	4	4	5	2	5	4	3.777
36	US20150045281A1	NOVO NORDISK A/S A CORPORATION	4	4	5	2	4	4	3.675
37	US20110009315A1	UNCONFIRMED	4	3	4	3	5	4	3.652
38	US20140199272A1	US DEPARTMENT OF HEALTH AND HUMAN SER... (+1)	4	4	3	4	4	3	3.777
39	US20150368311A1	SANOFI	5	4	3	2	5	4	3.703
40	US20110077197A1	SANOFI-AVENTIS DEUTSCHLAND GMBH	4	4	4	3	4	4	3.763
41	US20110152185A1	NOVO NORDISK A/S	4	3	4	3	5	4	3.652
42	US20150191495A1	MERCK SHARP AND DOHME LLC	4	5	5	2	5	4	3.990
43	US20130157934A1	INDIANA UNIVERSITY RESEARCH AND TECHN...	4	4	4	3	3	4	3.660
44	US20120183616A1	SANOFI-AVENTIS DEUTSCHLAND GMBH	4	3	5	3	5	4	3.803
45	US20080275066A1	NOVO NORDISK A/S	4	4	4	3	4	3	3.690
46	US8318762B2	PFIZER INC	4	4	5	2	5	4	3.777
47	US20090118181A1	UNCONFIRMED	5	4	4	2	4	4	3.750

48	US20150196565A1	BOEHRINGER INGELHEIM INTERNATIONAL GMBH	4	4	4	2	5	5	3.700
49	US9127047B2	DUKE UNIVERSITY	5	3	5	2	4	4	3.688
50	US20080249089A1	BOEHRINGER INGELHEIM PHARMA KG	4	4	5	2	5	4	3.777
51	US7943656B2	BRISTOL-MYERS SQUIBB COMPANY	4	4	5	2	5	4	3.777
52	US20140045746A1	WUXI APPTEC BVI INC (+1)	4	4	5	2	5	4	3.777
53	US8178541B2	BOEHRINGER INGELHEIM PHARMA GMBH & CO...	4	4	5	2	5	3	3.705
54	US7378418B2	BRISTOL-MYERS SQUIBB COMPANY	4	4	5	2	5	3	3.705
55	US20100197591A1	PFIZER INC	4	4	5	2	4	4	3.675
56	US20090054303A1	BRISTOL-MYERS SQUIBB COMPANY	4	4	5	2	5	3	3.705
57	US8067447B2	BRISTOL-MYERS SQUIBB COMPANY	4	4	5	2	5	3	3.705
58	US20100063051A1	BRISTOL-MYERS SQUIBB COMPANY	4	4	5	2	5	3	3.705
59	US8304539B2	BRISTOL-MYERS SQUIBB COMPANY	4	4	5	2	5	3	3.705

V3 Structured-Output Example: #1 US8815802B2 — NOVO NORDISK A/S

Every row in the table above is backed by a V3 structured-output citation: each patent has six KC composite scores, each grounded in four per-sub-dimension JSON records carrying the source model, the model-supplied reason, and a verbatim quoted snippet of the patent text (with KC4 sub-dims additionally citing the RAG prior-art neighbour list). For brevity the per-KC tables for the remaining top-decile patents are not reproduced in this report; the complete per-sub-dimension records are persisted in Microprotein_V3_KC_Subscores.csv and the per-patent nested-JSON citation payload is persisted in the kc_score_justification_json column of Microprotein_Top15_Predictions.csv. The single example presented below — the #1 top-decile patent (US8815802B2) — illustrates the structured-output format for each of the six KC dimensions.

GLP-1 analogues and derivatives

KC	Score	Reason	Verbatim Quote from the Patent
KC1	5	Patent provides extensive structural disclosure including full IUPAC names, SEQ IDs for all peptide sequences, detailed	N.sup..epsilon.26-[2-(2-{2-[2-(2-{2-[(S)-4-Carboxy-4-(17-carboxyheptadecanoylamino)butyrylamino]ethoxy}ethoxy)acetylamino]ethoxy}ethoxy)acetyl][Aib.sup.8,His.sup.31,Gln.sup.34]GLP-1(7-37)-peptide (SEQ ID NO: 4)

		chemical structures with stereochemistry, and comprehensive nomenclature for all compounds.	
K C2	4	C1=5 (analogue, derivative, composition, method-of-treatment, intermediate) and strong structural anchoring via SEQ IDs/Markush (C2=4) with broad comprising + species hierarchy (C3=4). Genus well-exemplified by 47 species keeps Amgen risk moderate (C4=4), short of comprehensive 5.	Claims 1. A GLP-1 analogue... 85. A derivative of an analogue... 259. An intermediate product... 263. An analogue or a derivative according to any one of embodiments 1-258, for use as a medicament.
K C3	5	D1/D2/D4 all at 5: 12 claims but claim 6 enumerates 30+ specific derivative compounds (full structural diversity), claims 4-5 cover protracting moieties/linkers (Chem.1-10), claims 7-12 cover compositions/diabetes methods. D3=4 sufficient.	The analogue of any one of embodiments 1-49, which comprises, preferably has, the following amino acid modifications: (i) (31H, 34Q) SEQ ID NO: 9; (ii) (31H, 34Q, 37K)...
K	2	All four sub-	GLP-1 analogue which comprises a histidine residue at position 31, a glutamine

C4		dims at 2: crowded GLP-1 acylation space (liraglutide/semaglutide priority chains), H31/Q34 substitutions plus C18 diacid-OEG-gammaGlu linker are incremental within Novo's own design paradigm, no breakthrough leap shown.	residue at position 34, and a maximum of ten amino acid modifications... derivatives have a surprisingly high oral bioavailability [RAG anchor: US20130053311A1, US8648041B2, US20110166321A1…]
K C5	5	Extensive clinical-ready formulations, dosing regimens, routes of administration, and pharmaceutical compositions detailed for diabetes treatment with specific patient populations and indications.	pharmaceutical composition comprising a GLP-1 analogue...for use in the treatment and/or prevention of all forms of diabetes...administered in the form of a pharmaceutical composition
K C6	5	47 working examples with broad structural diversity (W1=5, W2=5), comprehensive in vivo program across db/db mice, rats, minipigs, pigs with dose-response, PK, IVGTT (W4=5). W3=3 is offset by strong in vivo and breadth.	The GLP-1 derivatives of Examples 2, 4, 7, 14 and 15 were tested in a dose-response study in an obese, diabetic mouse model (db/db mice)...Male Sprague Dawley rats...Male Gottingen minipigs...Female Landrace Yorkshire Duroc (LYD) pigs

KC4 rag_neighbors_cited (top prior-art neighbours from the corpus that informed the N1–N4 novelty scoring): US20130053311A1, US8648041B2, US20110166321A1, US20140179899A1, US20120329711A1, US20140296131A1, US20130244931A1, US20160108102A1, US20100305032A1, US20130053315A1…

Top-Decile Assignee Ranking

Top-decile assignees are ranked by frequency in the calibrated top-decile cohort, then by best calibrated patent rank, then by average calibrated ensemble mean.

Rank	Assignee	Count	Best Patent Rank	Avg Cal Rank	Avg Cal Ens
1	NOVO NORDISK A/S	16	1	19.44	3.842
2	BRISTOL-MYERS SQUIBB COMPANY	8	3	43.62	3.747
3	ZEALAND PHARMA A/S	3	4	11.33	3.842
4	AMYLIN PHARMACEUTICALS, LLC	1	6	6.0	3.94
5	INDIANA UNIVERSITY RESEARCH AND TECHNOLOGY CORP...	5	9	21.6	3.795
6	TUFTS MEDICAL CENTER, INC	1	17	17.0	3.76
7	SANOFI-AVENTIS DEUTSCHLAND GMBH	4	19	32.5	3.811
8	ZEALAND PHARMA A/S/	1	22	22.0	3.652
9	SANOFI	3	24	29.33	3.799
10	OPKO BIOLOGICS LTD	1	26	26.0	3.763
11	PFIZER INC	3	29	43.33	3.772
12	AMGEN INC	1	31	31.0	3.75
13	ASTRAZENECA PHARMACEUTICALS LP (+1)	1	32	32.0	3.715
14	MERCK SHARP & DOHME CORP	1	34	34.0	3.917
15	NOVO NORDISK A/S A CORPORATION	1	36	36.0	3.675
16	UNCONFIRMED	2	37	42.0	3.701
17	US DEPARTMENT OF HEALTH AND HUMAN SERVICES (+1)	1	38	38.0	3.777
18	MERCK SHARP AND DOHME LLC	1	42	42.0	3.99
19	BOEHRINGER INGELHEIM INTERNATIONAL GMBH	1	48	48.0	3.7
20	DUKE UNIVERSITY	1	49	49.0	3.688

21	BOEHRINGER INGELHEIM PHARMA KG	1	50	50.0	3.777
22	WUXI APPTEC BVI INC (+1)	1	52	52.0	3.777
23	BOEHRINGER INGELHEIM PHARMA GMBH & CO. KG	1	53	53.0	3.705

Bottom-Decile Assignee Ranking

Bottom-decile assignees are ranked by frequency in the bottom GLP-1 2016 backtest biopharma decile, then by weaker average original ensemble score.

Rank	Assignee	Count	Avg Original Score	Weakest Score
1	STECHL (+9)	2	2.277	2.277
2	AMYLIN PHARMACEUTICALS LLC (+1)	2	2.62	2.478
3	GLAXOSMITHKLINE LLC	2	2.591	2.48
4	NOVO NORDISK A/S	13	2.695	2.503
5	AMYLIN PHARMACEUTICALS, LLC (+1)	7	2.738	2.54
6	ASTRAZENECA UK LTD (+1)	1	2.54	2.54
7	SANOFI-AVENTIS DEUTSCHLAND GMBH	7	2.717	2.587
8	AMYLIN PHARMACEUTICALS, INC	6	2.747	2.625
9	AMBRX, INC	1	2.625	2.625
10	ELI LILLY & COMPANY (+1)	1	2.655	2.655
11	CEDARS- SINAI MEDICAL CENTER	1	2.655	2.655
12	SANOFI	3	2.773	2.69
13	NOVO VORDISK A/S	1	2.703	2.703
14	BRISTOL-MYERS SQUIBB COMPANY	1	2.765	2.765
15	ASTRAZENECA PHARMACEUTICALS LP (+1)	2	2.821	2.805
16	AMYLIN PHARMACEUTICALS, LLC	1	2.835	2.835
17	NASTECH PHARMACEUTICAL COMPANY INC	1	2.837	2.837
18	MANNKIND	1	2.837	2.837

	CORPORATION			
19	THE GOV. OF THE US, AS REPRESENTED BY THE SECRETARY,...	1	2.848	2.848
20	ALKERMES, INC	1	2.865	2.865
21	TAKEDA PHARMACEUTICAL COMPANY LIMITED	1	2.877	2.877
22	CEDARS-SINAI MEDICAL CENTER	1	2.893	2.893
23	CEDARS-SINSAI MEDICAL CENTER	1	2.893	2.893
24	INDIVIDUAL	1	2.91	2.91

Top 15 Tournament Calibration GLP-1 2016 backtest Patents

Each top-15 patent below is presented with: (a) the patent number and confirmed assignee, (b) the patent title, (c) the calibrated KC profile and pairwise tournament record, and (d) the analyst narrative. Every calibrated KC value shown is the V3 composite of four per-sub-dimension scores, each backed by a structured-citation record (score, source model, model-supplied reason, verbatim quote, and — for KC4 — rag_neighbors_cited). The complete per-sub-dimension records are persisted in the V3_KC_Subscores.csv and the Top15_Predictions.csv (kc_score_justification_json column). A worked example of the full structured-output format for one patent is shown in the Top-Decile section above.

#1 US8815802B2 — NOVO NORDISK A/S

GLP-1 analogues and derivatives

Calibrated KC profile: KC1=5, KC2=4, KC3=5, KC4=2, KC5=5, KC6=5; calibrated ensemble=4.075; pairwise record=57-1.

V3 pairwise tournament: 57W / 1L over top-decile (n=59).

#2 US20130053311A1 — NOVO NORDISK A/S

GLP-1 ANALOGUES AND DERIVATIVES

Calibrated KC profile: KC1=5, KC2=4, KC3=5, KC4=3, KC5=5, KC6=4; calibrated ensemble=4.240; pairwise record=56-2.

V3 pairwise tournament: 56W / 2L over top-decile (n=59).

#3 US7960349B2 — BRISTOL-MYERS SQUIBB COMPANY

N-terminally modified GLP-1 receptor modulators

Calibrated KC profile: KC1=4, KC2=4, KC3=4, KC4=3, KC5=4, KC6=5; calibrated ensemble=3.835; pairwise record=56-2.

V3 pairwise tournament: 56W / 2L over top-decile (n=59).

#4 US9089538B2 — ZEALAND PHARMA A/S

Peptide conjugates of GLP-1 receptor agonists and gastrin and their use

Calibrated KC profile: KC1=5, KC2=3, KC3=4, KC4=3, KC5=5, KC6=5; calibrated ensemble=3.950; pairwise record=55-3.

V3 pairwise tournament: 55W / 3L over top-decile (n=59).

#5 US8648041B2 — NOVO NORDISK A/S
Double-acylated GLP-1 derivatives

Calibrated KC profile: KC1=5, KC2=4, KC3=5, KC4=2, KC5=5, KC6=4; calibrated ensemble=4.002; pairwise record=53-5.

V3 pairwise tournament: 53W / 5L over top-decile (n=59).

#6 US20140066370A1 — AMYLIN PHARMACEUTICALS, LLC
Polypeptide Conjugate

Calibrated KC profile: KC1=5, KC2=4, KC3=3, KC4=3, KC5=5, KC6=4; calibrated ensemble=3.940; pairwise record=53-5.

V3 pairwise tournament: 53W / 5L over top-decile (n=59).

#7 US20090111730A1 — NOVO NORDISK A/S
Polypeptide protracting tags

Calibrated KC profile: KC1=4, KC2=4, KC3=5, KC4=3, KC5=5, KC6=4; calibrated ensemble=4.015; pairwise record=51-7.

V3 pairwise tournament: 51W / 7L over top-decile (n=59).

#8 US20110166321A1 — NOVO NORDISK A/S
DOUBLE-ACYLATED GLP-1 DERIVATIVES

Calibrated KC profile: KC1=5, KC2=4, KC3=5, KC4=2, KC5=4, KC6=5; calibrated ensemble=3.973; pairwise record=51-7.

V3 pairwise tournament: 51W / 7L over top-decile (n=59).

#9 US9127088B2 — INDIANA UNIVERSITY RESEARCH AND TECHNOLOGY CORPORATION
Glucagon superfamily peptides exhibiting nuclear hormone receptor activity

Calibrated KC profile: KC1=4, KC2=4, KC3=5, KC4=3, KC5=4, KC6=4; calibrated ensemble=3.913; pairwise record=48-10.

V3 pairwise tournament: 48W / 10L over top-decile (n=59).

#10 US20150344540A1 — NOVO NORDISK A/S
NOVEL GLP-1 RECEPTOR AGONISTS WITH CHOLESTEROL EFFLUX ACTIVITY

Calibrated KC profile: KC1=4, KC2=4, KC3=4, KC4=3, KC5=5, KC6=5; calibrated ensemble=3.938; pairwise record=47-11.

V3 pairwise tournament: 47W / 11L over top-decile (n=59).

#11 USRE44186E — BRISTOL-MYERS SQUIBB COMPANY

Cyclopropyl-fused pyrrolidine-based inhibitors of dipeptidyl peptidase IV and method

Calibrated KC profile: KC1=4, KC2=4, KC3=5, KC4=3, KC5=4, KC6=3; calibrated ensemble=3.840; pairwise record=45-13.

V3 pairwise tournament: 45W / 13L over top-decile (n=59).

#12 US20130143793A1 — ZEALAND PHARMA A/S

PEPTIDE CONJUGATES OF GLP-1 RECEPTOR AGONISTS AND GASTRIN AND THEIR USE

Calibrated KC profile: KC1=5, KC2=3, KC3=4, KC4=3, KC5=5, KC6=4; calibrated ensemble=3.877; pairwise record=44-14.

V3 pairwise tournament: 44W / 14L over top-decile (n=59).

#13 US8980830B2 — INDIANA UNIVERSITY RESEARCH AND TECHNOLOGY CORPORATION

Peptide compounds exhibiting glucagon antagonist and GLP-1 agonist activity

Calibrated KC profile: KC1=4, KC2=4, KC3=5, KC4=3, KC5=4, KC6=3; calibrated ensemble=3.840; pairwise record=44-14.

V3 pairwise tournament: 44W / 14L over top-decile (n=59).

#14 US8895694B2 — NOVO NORDISK A/S

Glucagon-Like Peptide-1 derivatives and their pharmaceutical use

Calibrated KC profile: KC1=5, KC2=3, KC3=4, KC4=3, KC5=3, KC6=4; calibrated ensemble=3.673; pairwise record=44-14.

V3 pairwise tournament: 44W / 14L over top-decile (n=59).

#15 US8859491B2 — INDIANA UNIVERSITY RESEARCH AND TECHNOLOGY CORPORATION

Glucagon superfamily peptides exhibiting glucocorticoid receptor activity

Calibrated KC profile: KC1=4, KC2=5, KC3=4, KC4=3, KC5=3, KC6=4; calibrated ensemble=3.873; pairwise record=43-15.

V3 pairwise tournament: 43W / 15L over top-decile (n=59).

Predictive-Validity Test Results

Test framings and PASS / FAIL thresholds for the three Phase 3 predictive-validity tests reported in this section — the tirzepatide construct-validity test via Indiana University DiMarchi platform precursor coverage, the dulaglutide CoM-parent binary corroboration test, and the Hanmi LAPS-conjugate binary corroboration test — were locked in raw/Phase3_Framing_Notes.md on 2026-06-01, before Phase 2 V3 scoring began. The pre-commitment discipline was that the test framings could not be retrofitted after the scoring run produced its rankings; this is what distinguishes a back-test from a retrospective narrative. The dulaglutide and LAPS tests are reported below as binary corroboration (PASS / MIXED / FAIL) rather than as statistical hit-rates because the substrate patent counts are too narrow for statistical inference — 2 to 3 dulaglutide composition-of-matter parent precursors (Eli Lilly Glaesner-class fusion-protein filings) and 16 Hanmi LAPS-conjugate filings.

All ranks reported in this section are derived from the same GLP1_Biopharma_Ensemble_Final_backtest_2016.csv file shipped with this report; the ensemble-rank in the substrate is computed by sorting on ensemble_mean (descending) with patent_number as the deterministic tie-break. The tournament-calibrated rank (1–27) shown for the Indiana DiMarchi table is sourced from GLP1_Top_Decile_V3_Calibrated_KC_Scores_backtest_2016.csv.

1. Tirzepatide Construct-Validity (Indiana DiMarchi Platform Precursors)

Pre-committed PASS criterion: V3 ranks at least one Indiana DiMarchi dual or triple-agonist platform precursor patent in the top decile of the substrate.

Substrate patents matching the criterion (assignee contains "INDIANA UNIVERSITY," inventor list includes DiMarchi, Richard, and claims contain dual-agonist / triple-agonist / co-agonist / glucagon-superfamily framing): 16. In top decile (rank ≤ 58): 5 (31.2%). In top quartile (rank ≤ 146): 9 (56.2%). Baseline-rate enrichment at top decile: 3.15×. Baseline-rate enrichment at top quartile: 2.25×. Result: PASS.

The five Indiana DiMarchi patents in the tournament-calibrated top decile are tabulated below.

Calibrated rank	Patent	Ensemble	TA bucket	Title (60 ch)
9	US9127088B2	3.913	Triple_Agonist	Glucagon superfamily peptides exhibiting nuclear hormone ...
13	US8980830B2	3.840	Dual_Agonist	Peptide compounds exhibiting glucagon antagonist and GLP-...
15	US8859491B2	3.873	Triple_Agonist	Glucagon superfamily peptides exhibiting glucocorticoid r...
28	US20150320871A1	3.690	Triple_Agonist	Glucagon Superfamily Peptides Exhibiting Nuclear Hormone ...
43	US20130157934A1	3.660	Triple_Agonist	Glucagon Superfamily Peptides Exhibiting Glucocorticoid R...

V3 ranked the Indiana University DiMarchi dual / triple-agonist platform IP that tirzepatide later commercially validated. The 3.15× top-decile enrichment is a strong construct-validity signal for the tirzepatide direction even though tirzepatide composition-of-matter parent patents post-date the 2016 corpus floor and are not in the substrate.

2. Dulaglutide Binary Corroboration (Eli Lilly Glaesner-Class Fc-Fusion Precursors)

Pre-committed PASS criterion: PASS = all three dulaglutide CoM-parent precursors in top decile; MIXED = some in top decile; FAIL = none in top decile.

Substrate patents (Eli Lilly Glaesner-class GLP-1 / Fc-fusion-protein filings):

Patent	Description	Ensemble rank	Percentile	Decile status
US7271149B2	2007 Glaesner canonical Lilly GLP-1 fusion proteins	193 / 584	33.0%	Middle
US20090232807A1	2009 Lilly Fc-formulation continuation	488 / 584	83.6%	Bottom half
US20100196405A1	Candidate third precursor (UNCONFIRMED assignee)	509 / 584	87.2%	Bottom half

Result: FAIL. None of the three dulaglutide composition-of-matter parent precursors landed in the top decile. Per the Phase 3 framing notes locked pre-scoring: dulaglutide's commercial success was driven by Fc-fusion convenience (weekly subcutaneous dosing in a crowded GLP-1 agonist field) rather than by IP-quality dimensions V3 measures (claim breadth, prior-art sparseness, working-example diversity, inventive-step quality). The core fusion concept was already well-established by Glaesner et al., making the IP itself look "incremental advance over crowded prior art" under V3's KC4 N1 lens. The framework reading these patents as not-top-decile is methodologically honest, not a defect: V3 measures structural rigor in the molecules being claimed and is not designed to predict commercial outcomes whose drivers are convenience, delivery, or indication-shift foreseeability.

3. Hanmi LAPS Binary Corroboration (LAPS-Conjugate Platform Filings)

Pre-committed PASS criterion (analogous to dulaglutide): PASS = all LAPS-conjugate filings in top decile; MIXED = some in top decile; FAIL = none in top decile.

Substrate patents: 16 Hanmi-assigned LAPS-conjugate filings (Hanmi LAPS platform priority chain). In top decile: 0. Highest-ranking LAPS filing in the substrate: US20130122023A1 at ensemble rank 110 / 584 (18.8%-ile).

Result: FAIL. None of the Hanmi LAPS-conjugate substrate filings landed in the top decile. Per the Phase 3 framing notes: the LAPS (Long-Acting Protein/peptide discovery System) conjugate is a half-life-extension delivery-platform technology whose commercial value is in the formulation and the licensing arc, not in the claim-craftsmanship dimensions V3 measures. This is consistent with the June 2026 Lilly – Hanmi $1.2B alliance being a delivery-platform partnership in a defined rare-disease indication (GLP-2 for short bowel syndrome) rather than a foundational composition acquisition; V3 reads this directionally before the deal-flow ratifies it.

4. Other Agonists — Statistical Hit-Rate (Supporting Corroboration)

The table below reports the marker-hit hit-rate for each of the seven commercially-validated GLP-1 agonists in the substrate, plus tirzepatide (which appears as a corpus limitation because tirzepatide's composition-of-matter parent publications post-date the 2016-12-31 corpus floor). "Substrate marker-hits" counts patents whose spec text contains the agonist's INN or development-code marker; "Top-decile enrichment" is the marker-hit hit-rate in the top decile divided by the substrate baseline rate (top-decile cut / substrate size).

Agonist	Substrate marker-hits	In top decile	Top-decile enrichment	Interpretation
semaglutide	157	23	1.48×	Strong corroboration of the semaglutide direction
albiglutide	81	11	1.37×	Positive corroboration; sample narrower
liraglutide	314	40	1.28×	Positive corroboration
dulaglutide	300	39	1.31×	Marker-hit enrichment positive (mostly comparator / prior-art mentions in dual-agonist work); CoM-parent test is the binary test above
efpeglenatide	25	3	1.21×	Positive but narrow
exenatide	514	51	1.00×	At-baseline (no enrichment)
lixisenatide	185	16	0.87×	Below baseline
tirzepatide	0	0	n/a	Corpus limitation — post-2016 priority publications absent

Disambiguation: marker-hit enrichment counts patents with the agonist's drug name appearing anywhere in the substrate spec, which includes prior-art mentions in dual-agonist and combination-product work. This is NOT the same as CoM-parent precursor predictive-validity (which is what the binary tests above measure). The marker-hit table is supporting corroboration that V3 reads the broader incretin field coherently, not a primary predictive-validity finding.

A patent-quality framework reads the IP that exists. It does not read the future indications for which that IP will turn out to be useful. The Novo Nordisk semaglutide priority chain that V3 correctly identifies as foundational was filed and prosecuted as type 2 diabetes IP; it became the foundation of the obesity-metabolic franchise because of clinical evidence that emerged years after prosecution choices were made. The 2017 GLP-1 indication shift was a clinical and regulatory event, not an IP event; the IP that mattered was already in place, but its eventual importance depended on data that had not yet been generated.

The dulaglutide and LAPS FAIL results are the honest negative outcomes that distinguish a real back-test from a retrofitted narrative: V3 refused to rank these patents on IP-quality grounds, and that refusal is the test the framework had to pass to be trustworthy on the questions it does answer.

Methodology

The GLP-1 2016 backtest substrate (n=584) was assembled in three stages. Phase 1.5 discovery + fetch (scripts/discover_and_fetch_backtest.py) ran 58 USPTO PPubs queries crossing GLP-1 / GIP / glucagon / incretin / agonist / analog / derivative / fusion / acylation / Hanmi LAPS / Indiana DiMarchi / Zealand / commercial-agonist-drug-name vocabulary across abstract / title / claims, restricted to publication date 1995-01-01 through 2016-12-31. 2,445 unique candidate patents were discovered. Title- and full-text classification yielded 1575 Biopharma, 131 Diagnostic, 116 Device, and 78 Non-Biopharma classifications. Phase 1.5 substrate selection (scripts/categorize.py) then partitioned the Biopharma cohort by invention-purpose into com_anchor (composition-of-matter parent IP for commercially-validated GLP-1 agonists), foundational_platform (Hanmi LAPS, PhaseBio ELP, Ambrx, MannKind, Intarcia, Indiana DiMarchi), competing_non_validated (pre-2016 non-validated peptide IP), and non_substrate (late-stage formulation continuations, manufacturing process, pure method-of-use derivatives). After substrate-selection filtering (pub_date < 2016-01-01, claim_count ≥ 8) and five rounds of verification audits (substrate scope memo + two addenda), the substrate was locked at 584 patents with one targeted expansion fetch (US7271149B2, the 2007 Glaesner et al. Lilly GLP-1 fusion proteins canonical dulaglutide CoM-parent precursor, added per Verification 5(b)). The locked substrate is the input to Phase 2 V3 KC scoring; non-substrate Biopharma patents (n≈991) were retained in the RAG index for prior-art-landscape context but were not scored. The 584 substrate patents were assigned to one of eight GLP-1 therapeutic-area buckets (Triple_Agonist, Dual_Agonist, Foundational_Platform, Combination_Product, Formulation_Delivery, Single_Agonist_CoM, Adjacent_Therapeutic, Adjacent_Mechanism) via the Path 2 classifier (scripts/glp1_ta.py): substrate-categorizer label as primary invention-purpose signal, with dual/triple/combo/formulation universal overrides applied INSIDE each category, using a title+abstract+full-claims invention-scope window. TA multipliers are NOT applied to the patent-level ranking (ta_weighted_score == ensemble_mean).

All KC scoring was performed by Anthropic LLM agents reading the local full-spec corpus (abstract + independent claims + dependent claims + specification body, with section caps of 6,000 / 200,000 / 500,000 chars respectively — large enough that virtually all GLP-1 substrate patents reach the model intact, including the Indiana DiMarchi platform patents whose specs run 250k–1.7M chars). The per-patent loop uses prompt caching (cache_control=ephemeral on the full patent text) and routes each of the 24 KC sub-dimensions to a per-sub-dim model: Haiku 4.5 for mechanical sub-dims (A1, A3, C1, C3, D1, E3, W1, W4); Sonnet 4.5 for the default and five of the holistic sub-dims (A2, B*, C2, D2, D4, E2, E4, N2, W2, W3 default; C4, N1, N3, N4, E1 holistic-routed-to-Sonnet); and Opus 4.7 for the two holistic sub-dims D3 (Prosecution Flexibility) and A4 (Commercial Differentiation) plus the four KC2/KC3/KC4/KC6 composite calls. The D3+A4 → Opus forcing is the HOLISTIC_OPUS_FORCED correction identified by HER2 corrected QA (Sonnet exhibits rubric-collapse on D3 — pinning every patent at score 4 — and on A4 with 67.5% modal-4 vs the Opus discriminating 2/3/4 distribution). KC4 sub-dims (N1–N4) additionally receive RAG top-20 nearest-neighbour context from a text-embedding-3-small chunk-level vector index over all 2,445 discovered patents (1,579 Biopharma fetched specs → 85,352 chunks → 1,579 centroids). The RAG index spans the full Biopharma fetched cohort, not just the 584 scored substrate, to provide a richer prior-art landscape that mirrors the full 2016 GLP-1 IP space. Composite KC scores follow the V2 rules: KC1 and KC5 = median of sub-dimension scores rounded to nearest integer (mechanical, no LLM call); KC2, KC3, KC4, KC6 = holistic Opus composite per the V2-anchored per-KC composite prompt (KC4 carries a 2016-GLP-1-landscape-specific anchor naming the dominant commercial-agonist priority chains and the SPARSE-prior-art positioning of the Indiana DiMarchi dual-agonist class). Ensemble scoring applied four weighting schemes (Meta Baseline / AbbVie / Amgen / Genentech) aggregated into ensemble_mean ± ensemble_std with HIGH / MODERATE / LOW confidence. Tier thresholds are absolute (TIER 1 PREMIUM ≥ 4.5; TIER 1 HIGH 4.0–4.49; TIER 2 MODERATE 3.0–3.99; TIER 3 SPECULATIVE < 3.0).

Before any KC scoring, a Sonnet 4.5 SS-F (Single-Shot Filter) / OS-S (One-Sided Safe) firewall screened every patent's title + abstract + first 2,000 chars of claims + first 2,000 chars of description against the V3 Quarantine Taxonomy (Q1 weaponization, Q2 illicit drug synthesis, Q3 CSAM, Q4 pseudoscience-asserting-biomed, Q5 hate-

speech-jargon-with-biomed). For this corpus 0 of 584 substrate patents were quarantined (the GLP-1 IP corpus is clean therapeutic-development work and does not trigger the Quarantine Taxonomy). The top decile (n=59) was then re-judged via a complete pairwise tournament (Sonnet 4.5 judge) prompted to compare GLP-1 / GIP / glucagon agonist biopharma patents (pre-2016 substrate) and pick the higher-quality IP overall. Detailed V3 outputs (per-sub-dim scores with verbatim evidence quotes and model attribution; RAG neighbour lists; tournament pairings; calibrated rankings) are available in GLP1_V3_KC_Subscores_backtest_2016.csv, GLP1_Top15_Predictions_backtest_2016.csv (with JSON citation-based KC justifications), GLP1_Top_Decile_V3_Calibrated_KC_Scores_backtest_2016.csv, and the V3 workflow manifest at raw/GLP1_V3_KC_Workflow_Manifest_backtest_2016.json. Phase 3 backtest analysis (the tirzepatide construct-validity test on the Indiana University DiMarchi dual/triple-agonist platform precursor class, the dulaglutide CoM-parent binary corroboration test, the Hanmi LAPS-conjugate binary corroboration test, and the supporting per-agonist marker-hit hit-rate table) is presented in the “Predictive-Validity Test Results” section above. Test framings and PASS / FAIL thresholds were locked in raw/Phase3_Framing_Notes.md on 2026-06-01, before Phase 2 V3 scoring began.

Canonical V3 rubric reference:
/Product_Roadmap/University_IP/_KC_Methodology/V3_Rubric/improved_scoring_prompt_v3.md.