## Supplementary figures

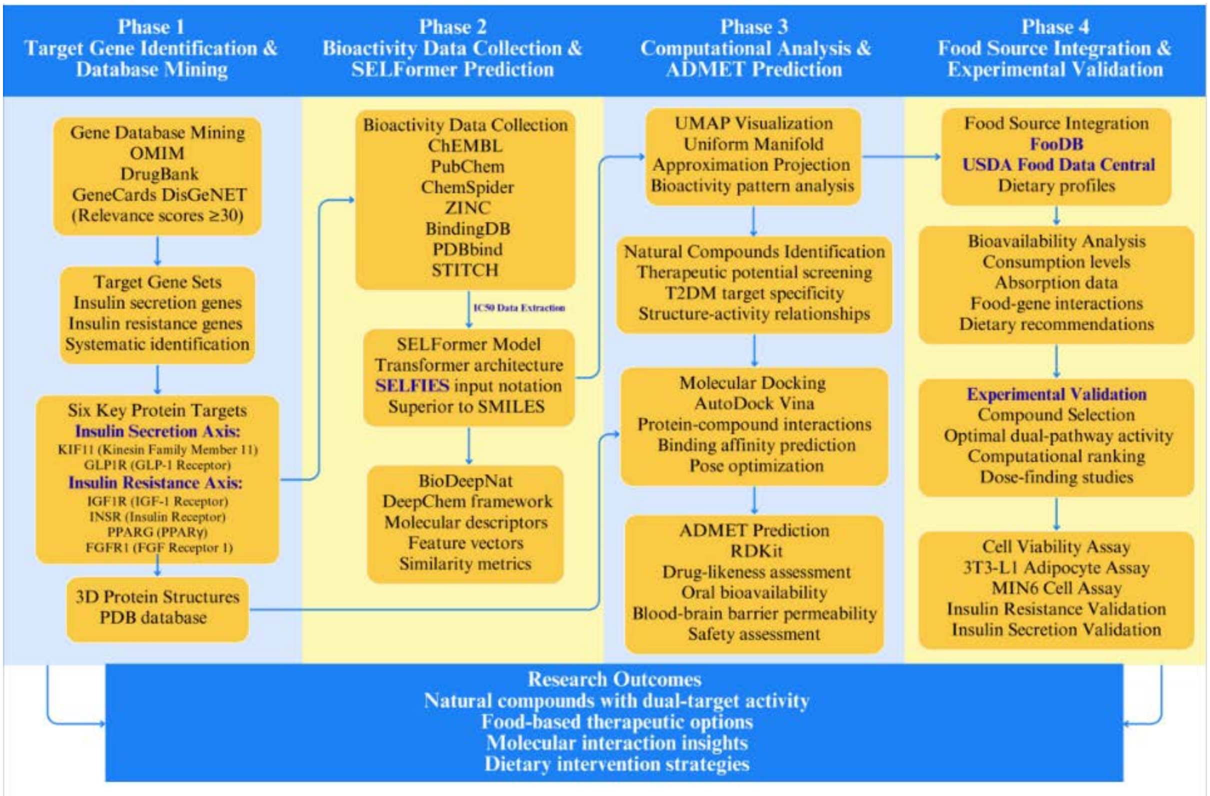

**Figure S1.** Computational Framework for Discovery of Natural Compounds with Therapeutic Potential in T2DM Acting through Modulation of Insulin Secretion and Resistance Pathways. The workflow encompassed four major phases: (1) Target identification and data collection, including gene selection from curated databases and retrieval of $IC_{50}$ data; (2) Machine learning prediction, featuring fine-tuning using SELFormer model and prediction of natural compound activity; (3) Natural compound similarity analysis and deep learning classification; (4) Computational analysis, including UMAP visualization, clustering, SAR analysis, and molecular docking studies; Experimental validation through cell-based assays for insulin secretion and resistance, culminating in food source identification for practical dietary applications; and *in vitro* validation.

**Abbreviations:** T2DM**,** Type 2 diabetes mellitus**;** $IC_{50}$, Half-maximal Inhibitory Concentration; UMAP, Uniform Manifold Approximation and Projection; SAR, structure activity relationship.

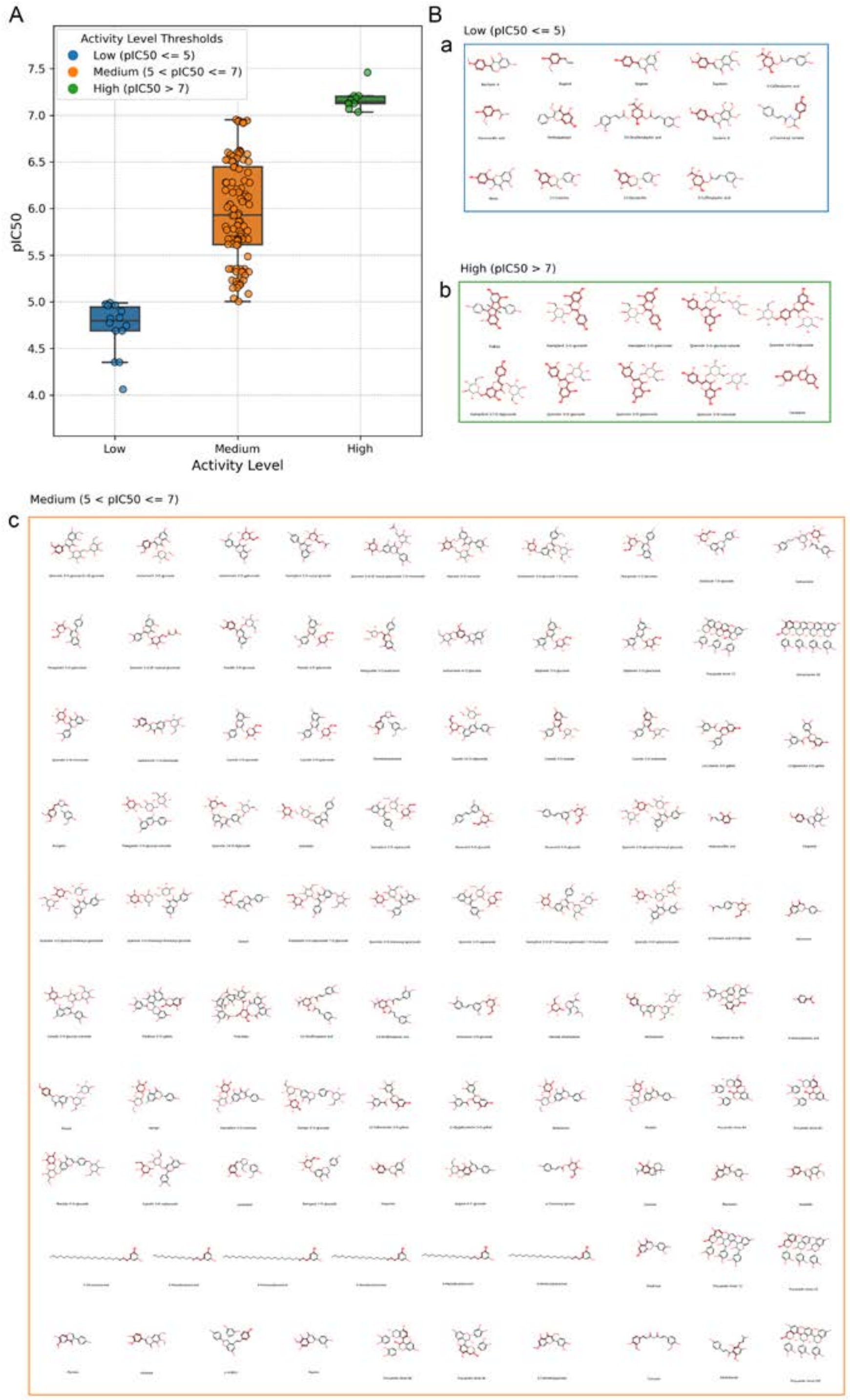

**Figure S2.** Natural Compound Activity Distribution

(A) Histogram plot showing pIC$_{50}$ values of natural compounds with a threshold line (pIC$_{50}$ = 7) separating different active compounds. (B) Chemical structures of representative natural compounds of (a) low-activity, (b) high-activity, and (c) medium-activity.

**Abbreviations:** pIC$_{50,}$ negative logarithm of the IC50.

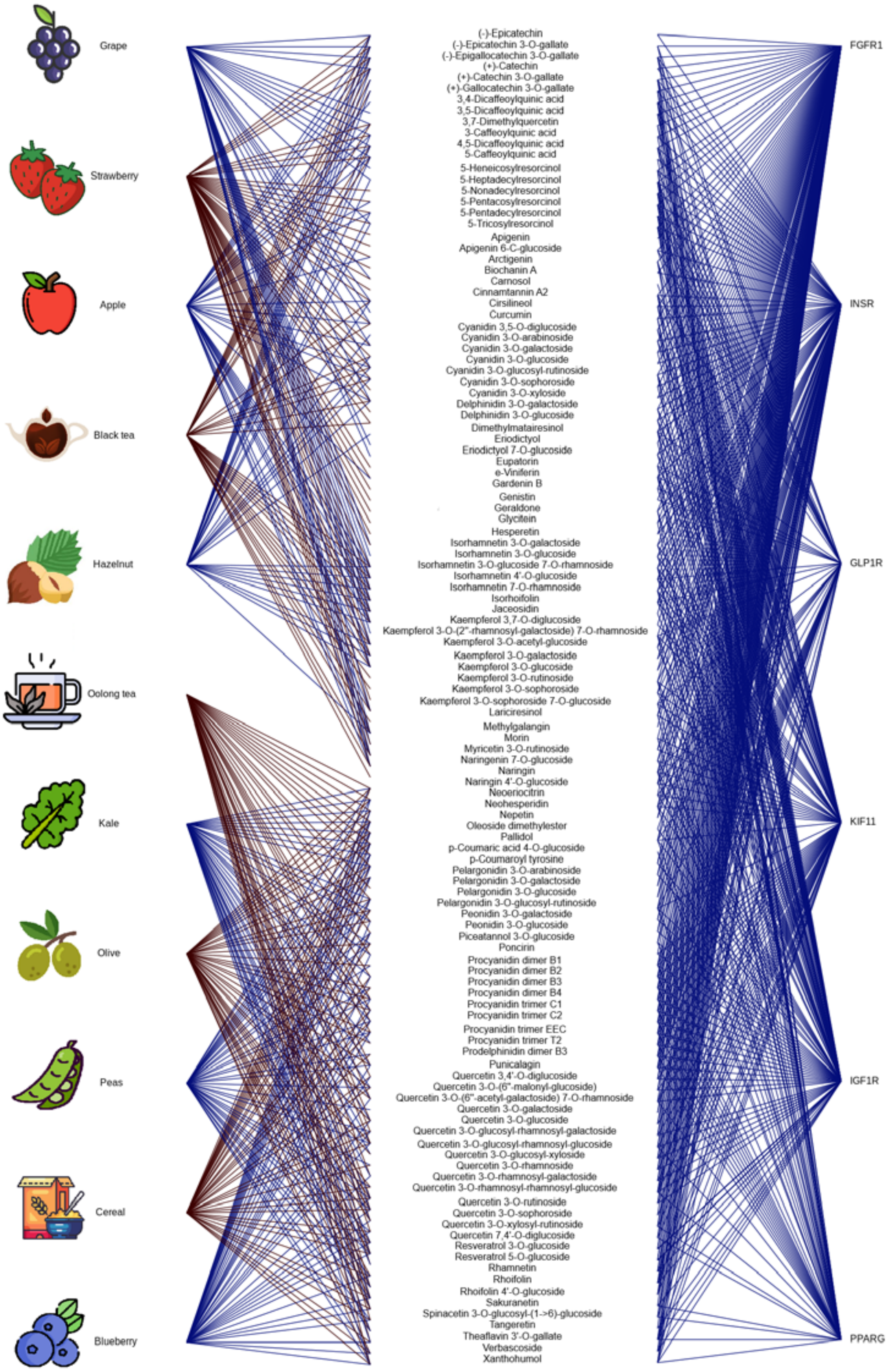
Grape
Strawberry
Apple
Black tea
Hazelnut
Oolong tea
Kale
Olive
Peas
Cereal
Blueberry
(-)-Epicatechin
(-)-Epicatechin 3-O-gallate
(-)-Epigallocatechin 3-O-gallate
(+)-Catechin
(+)-Catechin 3-O-gallate
(+)-Gallocatechin 3-O-gallate
3,4-Dicaffeoylquinic acid
3,5-Dicaffeoylquinic acid
3,7-Dimethylquercetin
3-Caffeoylquinic acid
4,5-Dicaffeoylquinic acid
5-Caffeoylquinic acid
5-Heneicosylresorcinol
5-Heptadecylresorcinol
5-Nonadecylresorcinol
5-Pentacosylresorcinol
5-Pentadecylresorcinol
5-Tricosylresorcinol
Apigenin
Apigenin 6-C-glucoside
Arctigenin
Biochanin A
Carnosol
Cinnamtannin A2
Cirsilineol
Curcumin
Cyanidin 3,5-O-diglucoside
Cyanidin 3-O-arabinoside
Cyanidin 3-O-galactoside
Cyanidin 3-O-glucoside
Cyanidin 3-O-glucosyl-rutinoside
Cyanidin 3-O-sophoroside
Cyanidin 3-O-xyloside
Delphinidin 3-O-galactoside
Delphinidin 3-O-glucoside
Dimethylmatairesinol
Eriodictyol
Eriodictyol 7-O-glucoside
Eupatorin
e-Viniferin
Gardenin B
Genistin
Geraldone
Glycitein
Hesperetin
Isorhamnetin 3-O-galactoside
Isorhamnetin 3-O-glucoside
Isorhamnetin 3-O-glucoside 7-O-rhamnoside
Isorhamnetin 4'-O-glucoside
Isorhamnetin 7-O-rhamnoside
Isorhoifolin
Jaceosidin
Kaempferol 3,7-O-diglucoside
Kaempferol 3-O-(2''-rhamnosyl-galactoside) 7-O-rhamnoside
Kaempferol 3-O-acetyl-glucoside
Kaempferol 3-O-galactoside
Kaempferol 3-O-glucoside
Kaempferol 3-O-rutinoside
Kaempferol 3-O-sophoroside
Kaempferol 3-O-sophoroside 7-O-glucoside
Lariciresinol
Methylgalangin
Morin
Myricetin 3-O-rutinoside
Naringenin 7-O-glucoside
Naringin
Naringin 4'-O-glucoside
Neoeriocitrin
Neohesperidin
Nepetin
Oleoside dimethylester
Pallidol
p-Coumaric acid 4-O-glucoside
p-Coumaroyl tyrosine
Pelargonidin 3-O-arabinoside
Pelargonidin 3-O-galactoside
Pelargonidin 3-O-glucoside
Pelargonidin 3-O-glucosyl-rutinoside
Peonidin 3-O-galactoside
Peonidin 3-O-glucoside
Piceatannol 3-O-glucoside
Poncirin
Procyanidin dimer B1
Procyanidin dimer B2
Procyanidin dimer B3
Procyanidin dimer B4
Procyanidin trimer C1
Procyanidin trimer C2
Procyanidin trimer EEC
Procyanidin trimer T2
Prodelphinidin dimer B3
Punicalagin
Quercetin 3,4'-O-diglucoside
Quercetin 3-O-(6''-malonyl-glucoside)
Quercetin 3-O-(6''-acetyl-galactoside) 7-O-rhamnoside
Quercetin 3-O-galactoside
Quercetin 3-O-glucoside
Quercetin 3-O-glucosyl-rhamnosyl-galactoside
Quercetin 3-O-glucosyl-rhamnosyl-glucoside
Quercetin 3-O-glucosyl-xyloside
Quercetin 3-O-rhamnoside
Quercetin 3-O-rhamnosyl-galactoside
Quercetin 3-O-rhamnosyl-rhamnosyl-glucoside
Quercetin 3-O-rutinoside
Quercetin 3-O-sophoroside
Quercetin 3-O-xylosyl-rutinoside
Quercetin 7,4'-O-diglucoside
Resveratrol 3-O-glucoside
Resveratrol 5-O-glucoside
Rhamnetin
Rhoifolin
Rhoifolin 4'-O-glucoside
Sakuranetin
Spinacetin 3-O-glucosyl-(1->6)-glucoside
Tangeretin
Theaflavin 3'-O-gallate
Verbascoside
Xanthohumol
FGFR1
INSR
GLP1R
KIF11
IGF1R
PPARG

**Figure S3.** Food Source Compound Profiles. Detailed visualization of natural compound profiles across various food sources mentioned in the network analysis. The figure shows the relative abundance and types of bioactive compounds found in strawberries, grapes, tea varieties, and other dietary sources, organized to highlight the richest sources of compounds with therapeutic potential in type 2 diabetes mellitus (T2DM).

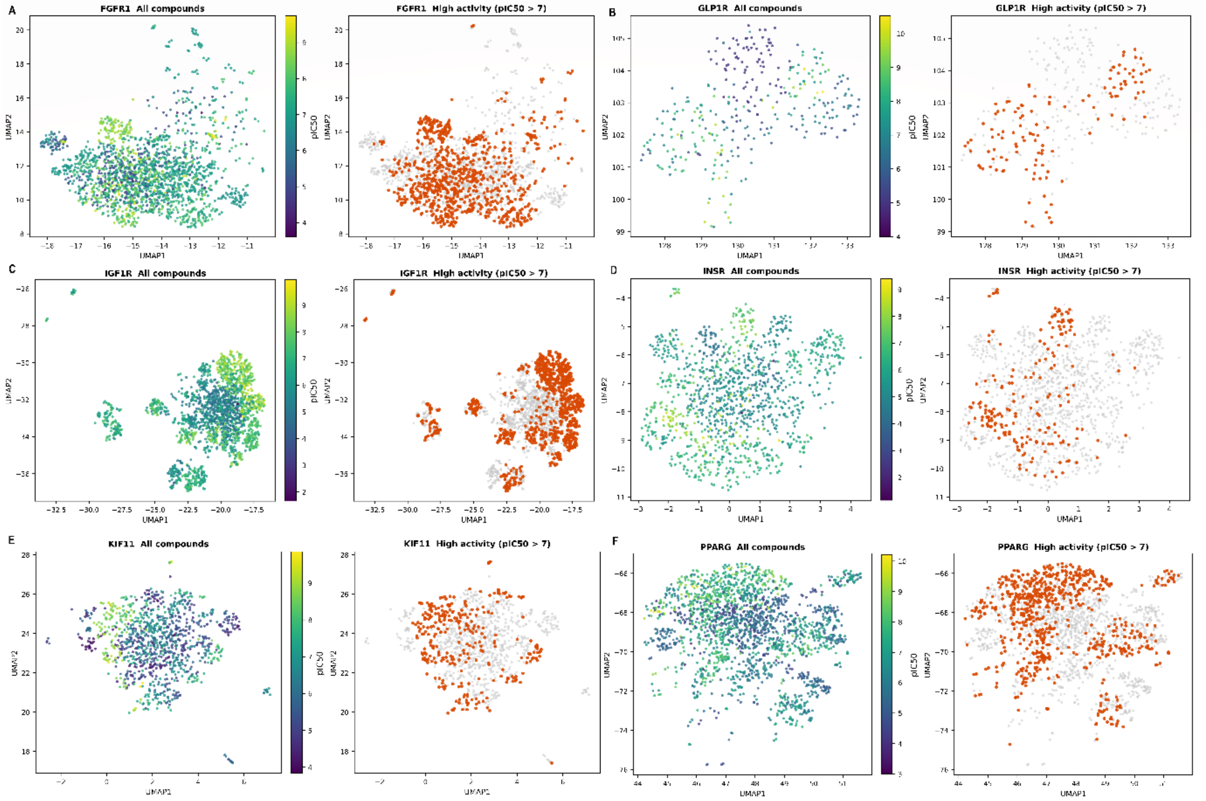

Figure S4. UMAP Visualization Comparing the Full Compound Space and High-Activity Subsets Across Six T2DM-Related Target Proteins.

UMAP projections illustrating the chemical space distribution of all screened compounds (left panels) versus the high-activity subset ($pIC_{50} > 7$, right panels) for each of the six T2DM-related target proteins: (A) FGFR1, (B) GLP1R, (C) IGF1R, (D) INSR, (E) KIF11, and (F) PPARG. In the left panels, all compounds are displayed and colored continuously by predicted $pIC_{50}$ values (viridis scale; purple = low activity, yellow = high activity), revealing the overall bioactivity landscape of the screened chemical space. In the right panels, the full compound set is shown in gray as background reference, with the high-activity subset ($pIC_{50} > 7$) highlighted in red.

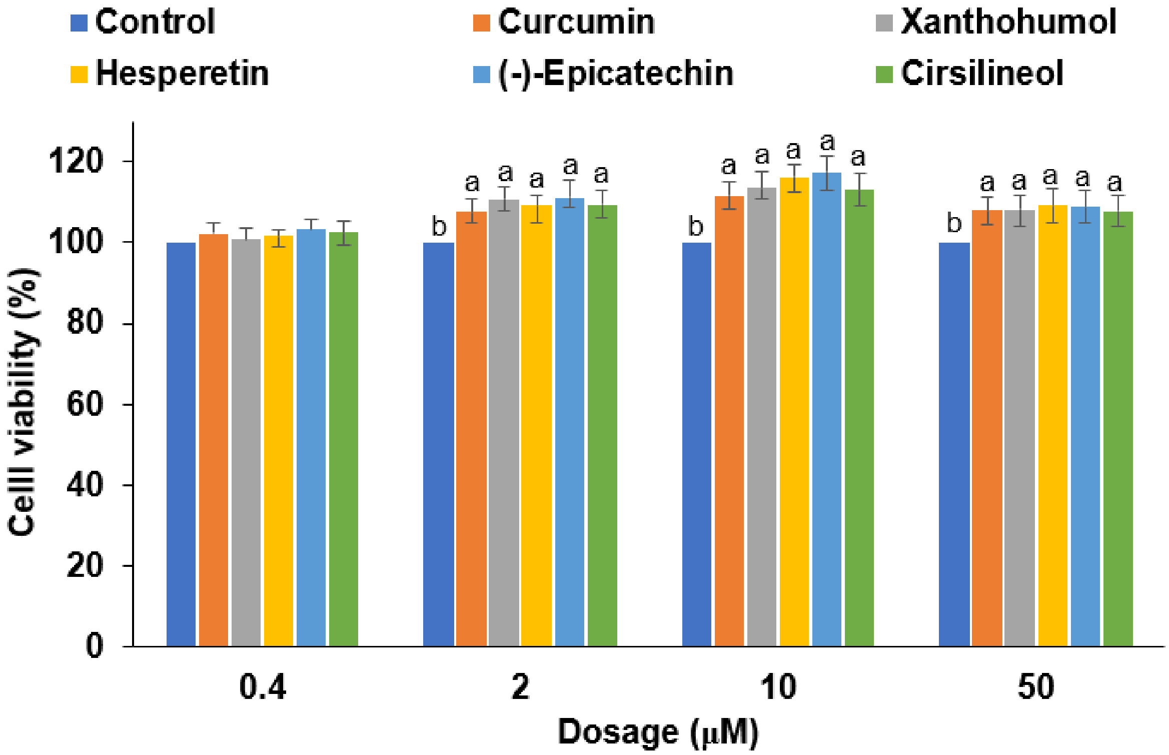

**Figure S5. Cell viability assessment of selected natural compounds in 3T3-L1 adipocytes.** Differentiated 3T3-L1 adipocytes were treated with selected natural compounds at concentrations of 0.4, 2, 10, and 50 µM for 24 hours. Cell viability was assessed using the MTT (3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide) assay. Results are expressed as percentage viability relative to vehicle control (0.1% DMSO, set as 100%). Values represent means ± SD from three independent experiments performed in triplicate. Statistical significance was determined by one-way ANOVA followed by Tukey's post-hoc test for multiple comparisons. Different letters (a, b, c, etc.) indicate statistically significant differences between groups at $p < 0.05$.

DMSO, Dimethyl sulfoxide

**Supplementary Table S1.** Individual SELFormer model performance metrics for each T2DM target protein

| Target Gene | Pathway | MSE | RMSE | MAE |
|---|---|---|---|---|
| GLP1R | Insulin Secretion | 0.497 | 0.705 | 0.311 |
| KIF11 | Insulin Secretion | 0.853 | 0.924 | 0.237 |
| INSR | Insulin Resistance | 0.991 | 0.996 | 0.603 |
| PPARG | Insulin Resistance | 0.884 | 0.94 | 0.614 |
| FGFR1 | Insulin Resistance | 0.88 | 0.938 | 0.235 |
| IGF1R | Insulin Resistance | 0.993 | 0.996 | 0.55 |

**Supplementary Table S2.** Quantitative comparison of autoencoder variants using Taylor diagram statistics relative to the experimental reference.

| Model | Correlation (r) | Standard Deviation (σ) | Centered RMSE (cRMSE) |
|---|---:|---:|---:|
| SimpleAE | 0.996 | 0.924 | 0.113 |
| ResAE | 0.995 | 0.875 | 0.158 |
| DeepAE | 0.986 | 0.912 | 0.180 |
| VAE | 0.991 | 0.737 | 0.288 |

**Supplementary Table S3.** Key amino acid residues involved in the interactions between compounds and target proteins.

| Target Protein | Compound | Residues involved in hydrogen bonding | | Residues involved in hydrophobic interactions | | | | | | Docking energy, ΔG (kcal mol−1) |
|---|---|---|---|---|---|---|---|---|---|---|
| | | Conventional Hydrogen Bond | Carbon Hydrogen Bond | Alkyl | Pi-Alkyl | Pi-Sigma | Pi-Pi T-shaped | Pi-Donor Hydrogen Bond | Pi-Sulfur | |
| FGFR1 | Curcumin | GLU571,ALA564 | No | No | VAL492,LYS514,LEU484 | VAL561 | No | No | No | -9.54 |
| | Xanthohumol | No | GLU562 | No | LEU630,VAL561,ALA640,ILE545 | LEU484 | PHE489 | No | No | -8.664 |
| GLP1R | Xanthohumol | GLN221,TRP214 | LEU217 | No | ARG299,VAL36,TRP39,TYR88 | LEU218 | TRP33 | No | No | -8.984 |
| IGF1R | Cirsilineol | ARG104,SER187 | MET184，CYS185 | No | LEU129 | No | No | No | No | -6.766 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Xanthohumol | SER210,THR49,TRP244,ARG77,THR23 | TYR225，CYS221 | VAL219,VAL50,ALA220 | No | No | TRP244 | No | No | -7.632 |
| KIF11 | Xanthohumol | ASN287 | GLY296,LEU293,THR300 | ILE332,ALA356,ALA353,VAL21,ALA334 | No | TYR104，LEU292 | No | No | No | -9.393 |
| INSR | Xanthohumol | LEU1002,ARG1000,SER1090 | ALA1080 | MET1139,LEU1002 | No | No | No | No | No | -7.727 |
| | Cirsilineol | ASP1150,LEU1002 | MET1079,ALA1080 | No | MET1139,ALA1028,VAL1010,LEU1002 | No | No | No | No | -7.713 |
| PPARG | Xanthohumol | GLU343,SER342 | No | No | CYS285,PHE287 | ARG288 | TYR327 | HIS449 | MET364 | -8.425 |