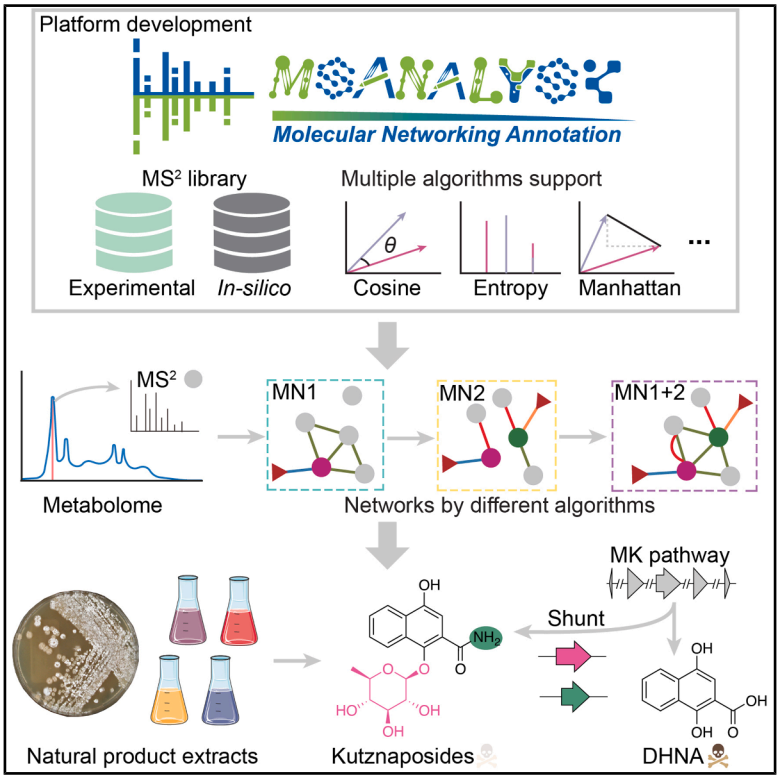
Resource

Cell Chemical Biology

Targeted discovery of aromatic glycosides with dual detoxification effects via a highly customized molecular networking platform

Graphical abstract



Authors

Wen-Chao Yu, Yan-Lei Yu, Bing-Cheng Dong, ..., Qihao Wu, Hong Wang, Bin Wei

Correspondence

qiw153@pitt.edu (Q.W.), hongw@zjut.edu.cn (H.W.), binwei@zjut.edu.cn (B.W.)

In brief

To overcome the limitations of single-algorithm annotation in mass spectrometry-based untargeted metabolomics, Yu et al. developed MSanalyst, which integrates multiple complementary MS^2 spectral algorithms. This approach enabled the discovery of kutznaposides, a class of aromatic glycosides, and uncovered their biosynthetic origin and antioxidant defense role.

Highlights

- MSanalyst enables the integration of multiple mass spectral similarity algorithms
- Complementary spectral metrics enhance the detection of hidden spectral connections
- MSanalyst facilitates the discovery of six aromatic glycosides, kutznaposides A–F
- Kutznaposides, from the menaquinone shunt pathway, mediate antioxidant defense

Yu et al., 2026, Cell Chemical Biology *33*, 1–13
January 15, 2026
https://doi.org/10.1016/j.chembiol.2025.12.001

CellPress

Please cite this article in press as: Yu et al., Targeted discovery of aromatic glycosides with dual detoxification effects via a highly customized molecular networking platform, Cell Chemical Biology (2025), https://doi.org/10.1016/j.chembiol.2025.12.001

Cell Chemical Biology

CellPress

Resource

Targeted discovery of aromatic glycosides with dual detoxification effects via a highly customized molecular networking platform

Wen-Chao Yu,[1,3,6] Yan-Lei Yu,[1,6] Bing-Cheng Dong,[1,6] Ze-Yu Wang,[2,3] Au-Qi Du,[1] Song-Wei Li,[1] Buddha Bahadur Basnet,[1,4] Xiao-Ze Bao,[1,3] Xuan-Rong Sun,[1] Xing-Nuo Li,[1] Qi Xuan,[2,3] Qihao Wu,[5,*] Hong Wang,[1,3,*] and Bin Wei[1,3,7,*]

[1]College of Pharmaceutical Science & Collaborative Innovation Center of Yangtze River Delta Region Green Pharmaceuticals, Zhejiang Key Laboratory of Green, Low-carbon, and Efficient Development of Marine Fishery Resources, Zhejiang University of Technology, Hangzhou 310014, China

[2]College of Information Engineering, Zhejiang University of Technology, Hangzhou 310023, China

[3]Binjiang Institute of Artificial Intelligence, Zhejiang University of Technology, Hangzhou 310056, China

[4]Central Department of Biotechnology, Tribhuvan University, Kathmandu 44600, Nepal

[5]Department of Pharmaceutical Sciences, University of Pittsburgh, Pittsburgh, PA 15261, USA

[6]These authors contributed equally

[7]Lead contact

*Correspondence: qiw153@pitt.edu (Q.W.), hongw@zjut.edu.cn (H.W.), binwei@zjut.edu.cn (B.W.)

https://doi.org/10.1016/j.chembiol.2025.12.001

SIGNIFICANCE Unlocking bioactive small molecules is essential for drug discovery, but discovering, prioritizing, and characterizing them remains difficult, as they are often found in complex biological extracts. Mass spectrometry-based untargeted metabolomics and MS^2 spectral similarity are critical strategies for establishing spectra-inferred structural relationships and discovering unknown metabolites. Current analytical platforms mainly rely on comparing MS^2 data using a single algorithm, which often leads to missed annotations of valuable metabolites. MSanalyst addresses this limitation by combining various mass spectral similarity algorithms, thereby enabling more comprehensive metabolite annotation. By applying MSanalyst to a well-studied strain, *Kutzneria viridogrisea*, we successfully discovered an undescribed class of aromatic glycosides, called kutznaposides. Integrated multi-omics and *in vitro* experiments further revealed that these molecules are produced through a previously uncharacterized menaquinone shunt pathway. This pathway serves as a crucial defense mechanism, enabling the host organism to eliminate reactive oxygen species and avoid self-toxicity. By demonstrating the strength of algorithmic integration, MSanalyst advances the systematic identification of hidden metabolites, metabolic pathways with essential biological functions, and potential biomedical applications.

SUMMARY

Natural products embedded within complex metabolomes are valuable sources of drug leads. Untargeted metabolomics using cosine-based MS^2 comparisons is widely used for discovering bioactive molecules. To improve annotation accuracy and resolution, alternative algorithms have been developed to complement cosine-based MS^2 comparison. Here, we present MSanalyst, a user-friendly platform that integrates 46 distinct mass spectral similarity algorithms. Benchmarking with microbial metabolite datasets and over three million MS^2 spectral pairs demonstrated that complementary algorithms substantially enhance the detection of metabolite-metabolite spectral connections. Applying MSanalyst to *Kutzneria viridogrisea* DSM 43850 led to the discovery of a class of aromatic glycosides, the kutznaposides. Biological assays and multi-omics analyses revealed that kutznaposides C–F arise from a previously unrecognized menaquinone shunt pathway, enabling the host to mitigate oxidative stress and avoid self-toxicity. Collectively, these findings highlight the potential of MSanalyst to uncover hidden metabolites, metabolic pathways, and their biological functions.

Please cite this article in press as: Yu et al., Targeted discovery of aromatic glycosides with dual detoxification effects via a highly customized molecular networking platform, Cell Chemical Biology (2025), https://doi.org/10.1016/j.chembiol.2025.12.001

INTRODUCTION

Natural products play crucial roles in diverse biological contexts, serving as potential drug candidates, disease biomarkers, signals for inter-microbial and host communication, and storage forms or inactive byproducts of primary metabolites within organisms.[1–3] However, natural products in complex matrices are often difficult to access due to the labor-intensive purification required for their discovery. Advanced dereplication or annotation strategies are crucial for the targeted discovery of undescribed secondary metabolites or metabolic pathways, as they help prevent the rediscovery of known compounds and minimize unnecessary experimental efforts. Liquid chromatography coupled with high-resolution mass spectrometry (LC-HRMS) is widely utilized in untargeted metabolomics to detect trace metabolites from small sample inputs, generating hundreds of thousands of informative tandem mass (MS^2) spectra. The Global Natural Products Social Molecular Networking (GNPS, http://gnps.ucsd.edu) platform has emerged as a central ecosystem for mass spectral analysis and knowledge dissemination.[4] GNPS provides diverse resources and tools, such as Mass Spectrometry Query Language (MassQL),[5] Mass Spectrometry Search Tool (MASST),[6,7] and Pan-Repository Reanalysis (PanREDU),[8] to facilitate the exploration and interpretation of complex molecular data. Among these, molecular networking (MN) has become a particularly powerful approach, leveraging MS^2 spectral comparison to construct networks of structurally related molecules.[9] Feature-Based MN (FBMN) further enhances MN by integrating feature detection and extraction from liquid chromatography-tandem mass spectrometry (LC-MS/MS) data, increasing precision and expanding analytical scope.[10] The recent GNPS 2.0 release features an extensive database of over 2 billion experimental MS^2 spectra, providing a rich source for accurate spectral comparisons. In parallel, several *in silico* annotation methods have been developed, including CSI-FingerID[11] and Competitive Fragmentation Modeling for metabolite IDentification (CFM-ID).[12] The *In Silico* MS^2 DataBase (ISDB) further broadens the MS^2 spectral search space,[13] providing an effective strategy for dereplicating natural products and annotating structural analogs of known compounds.

Within the GNPS ecosystem, MN has been demonstrated as a powerful approach for mapping chemical space from complex crude extracts through MS^2 spectral alignments. It organizes features sharing similar structural moieties into the same cluster according to their similarities using the modified cosine score.[9] Additional algorithms for MS^2 comparison and library searching continue to emerge, including Significant Interrelation of MS^2 Ions via Laplacian Embedding (SIMILE),[14] t-SNE,[15] spec2vec,[16] entropy,[17] and neutral loss.[18] These spectral similarity algorithms have been shown to outperform the classic cosine-based score by offering improved computational efficiency, enhancing receiver operating characteristics, and establishing more spectral correlations.[19] Therefore, integrating diverse mass spectral algorithms is essential to fully exploit the potential of MN in untargeted metabolomics.

Here, we present MSanalyst, a computational platform that enables seamless switching between and integration of multiple scoring metrics for MN. The stand-alone version is available at the GitHub repository https://github.com/WenchYu/MSanalyst, and the web-based version is accessible at https://msanalyst.net/. A total of 46 spectral similarity algorithms were benchmarked on a microbial natural product dataset and over three million spectral pairs derived from 2,571 small molecules. Of these, 29 algorithms provided complementary information to the classic modified cosine score for MS^2 comparison. MSanalyst incorporates these validated similarity metrics and assembles a comprehensive spectral library that includes both experimental and predicted spectra. By analyzing metabolomic samples of a widely studied actinobacterial strain (*Kutzneria viridogrisea* DSM 43850), we were able to identify the targeted natural products of this strain using MSanalyst. In addition, bioactivity and multi-omic analysis reveal that *K. viridogrisea* has evolved a shunt flux from the menaquinone pathway, which serves as a self-detoxification strategy against oxidative stress.

RESULTS

MSanalyst incorporates a comprehensive mass spectral library and various spectral similarity algorithms

In the MSanalyst workflow, raw data are preprocessed into molecular feature lists. These features serve as the basis for downstream analyses. Because key parameters such as mass tolerance and noise level vary across instruments and methods, preprocessed files are recommended as direct inputs. Alternatively, MSanalyst integrates UmetaFlow[20] and matchms[21] to preprocess commonly used mass data formats (e.g., mzML, mzXML, and mgf; Figure 1A).

MN is the core of MSanalyst, enabling annotation, clustering, and visualization of molecular features for prioritization. We enhanced this functionality by expanding the spectral library and incorporating multiple spectral similarity algorithms (Figure 1B). To address the relatively slow growth of experimental MS^2 spectra,[22] we employed CFM-ID to generate *in silico* spectra for 467,126 nonredundant structures, primarily sourced from the Comprehensive Marine Natural Products Database (CMNPD),[23] NPAtlas,[24] and the COlleCtion of Open Natural prodUcTs (COCONUT).[25] By integrating these with experimental MS^2 spectra from GNPS, MSanalyst provides comprehensive coverage of natural product chemical space. In total, 46 spectral similarity algorithms were incorporated to improve MS^2 comparison, many of which outperform the classic cosine-based method (Table S1).

In the main MN pipeline (Figure 1C), query features within the set mass tolerance undergo MS^2 comparison against the library. Features without a match above the defined spectral threshold are flagged as "unmatched," whereas matched features are classified as experimental or *in silico* depending on the reference spectrum. Since *in silico* spectra often diverge significantly from experimental ones (Figures S1 and S2), their use increases annotation efficiency but can introduce false positives (Figure S3). To mitigate this, *in silico* annotations are retained only when they meet a chemical dice similarity $\geq$ 0.75 to co-occurring experimental annotations within the same cluster.[26] Annotation results are then merged into the molecular network. Multiple similarity algorithms are supported in the library search and the subsequent self-clustering step, as analyses showed that different spectral similarity algorithms yield both overlapping and unique spectral correlations,

Please cite this article in press as: Yu et al., Targeted discovery of aromatic glycosides with dual detoxification effects via a highly customized molecular networking platform, Cell Chemical Biology (2025), https://doi.org/10.1016/j.chembiol.2025.12.001

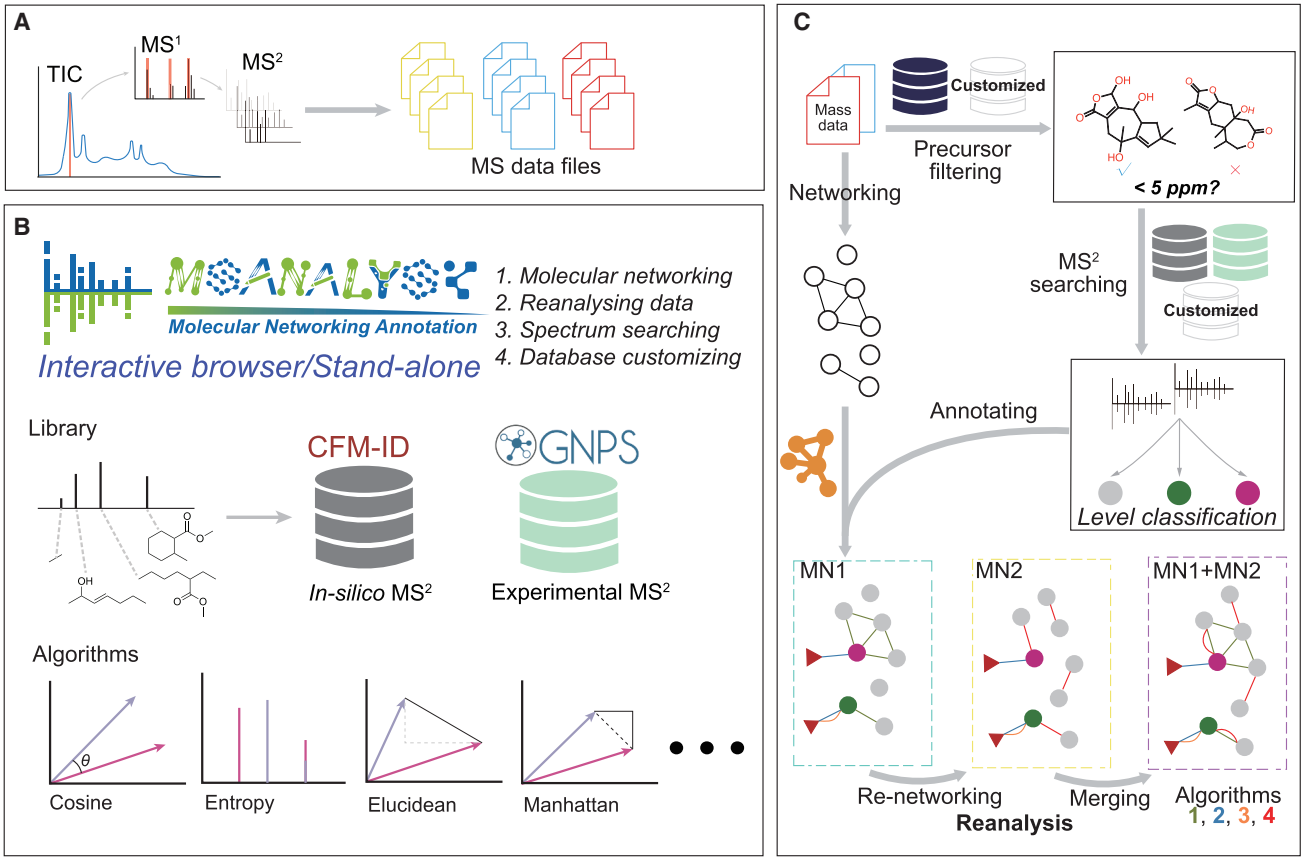


Figure 1. Overview of the MSanalyst worflow

(A) Unknown metabolomic features are acquired and preprocessed as input for MSanalyst.

(B) A comprehensive mass library, diverse spectral similarity algorithms, and additional functions are integrated.

(C) Features are combined within the MSanalyst workflow to enhance annotation and expand spectral connectivity.

complementing one another to strengthen network construction (Figure 2). Additional annotations are linked to their query features, clearly visualizing how different spectral algorithms contribute to the network.

To maximize accessibility, we provide both a web-based platform (https://msanalyst.net/) and a standalone version of MSanalyst, supported with instructional videos and documentation. Beyond the main pipeline, MSanalyst includes supplementary modules to improve functionality and user experience. Each task is assigned a unique identifier to facilitate data management and integration. The "reanalysis" module allows rapid refinement of results by adjusting thresholds and merging spectral connections missed by a single algorithm. The "spectrum searching" module enables rapid queries for metabolites of interest, while "database customizing" allows users to build private mass libraries for specific research needs. Together, these integrated features establish MSanalyst as a versatile and efficient platform for high-throughput, untargeted metabolomic data analysis.

Additional spectra-inferred structural relationships identified by diverse spectral algorithms

Modified cosine remains the most widely used similarity algorithm for establishing spectral correlations to infer structural similarity.[9,27] However, several algorithms have demonstrated greater accuracy and computational efficiency,[17] though differences in the correlations produced by these algorithms remain unexplored. To address this, we benchmarked 46 algorithms using 81 microbial natural product spectra to simulate real-world applications, alongside three million spectral pairs derived from 2,543 spectra for comparative analysis (Figure 2A). An initial library search at a widely accepted threshold of $\geq$0.7 showed that true positive (TP) annotations varied substantially across algorithms (1–34, Data S1), indicating that a universal threshold was unsuitable. Fragment counts also varied widely, from 4 to 275 (Figure 2B); thus, a fixed matched peak threshold could overlook some TP annotations (Figure S4). To optimize performance, we applied Youden's Index to identify thresholds maximizing the difference between the TP rate (TPR) and false positive rate (FPR).[28] This confirmed our hypothesis that optimal thresholds differ across algorithms, maintaining a false discovery rate (FDR) near 20% (Figure 2C). Among them, entropy achieved the highest performance with 36 TP annotations, while 39 algorithms contributed additional TP annotations beyond those detected by modified cosine (Data S1). Algorithms such as entropy, peak percentage, and ms_for_id outperformed modified cosine, largely because entropy and ms_for_id normalize intensity values, whereas peak percentage considers only matched fragments

Please cite this article in press as: Yu et al., Targeted discovery of aromatic glycosides with dual detoxification effects via a highly customized molecular networking platform, Cell Chemical Biology (2025), https://doi.org/10.1016/j.chembiol.2025.12.001

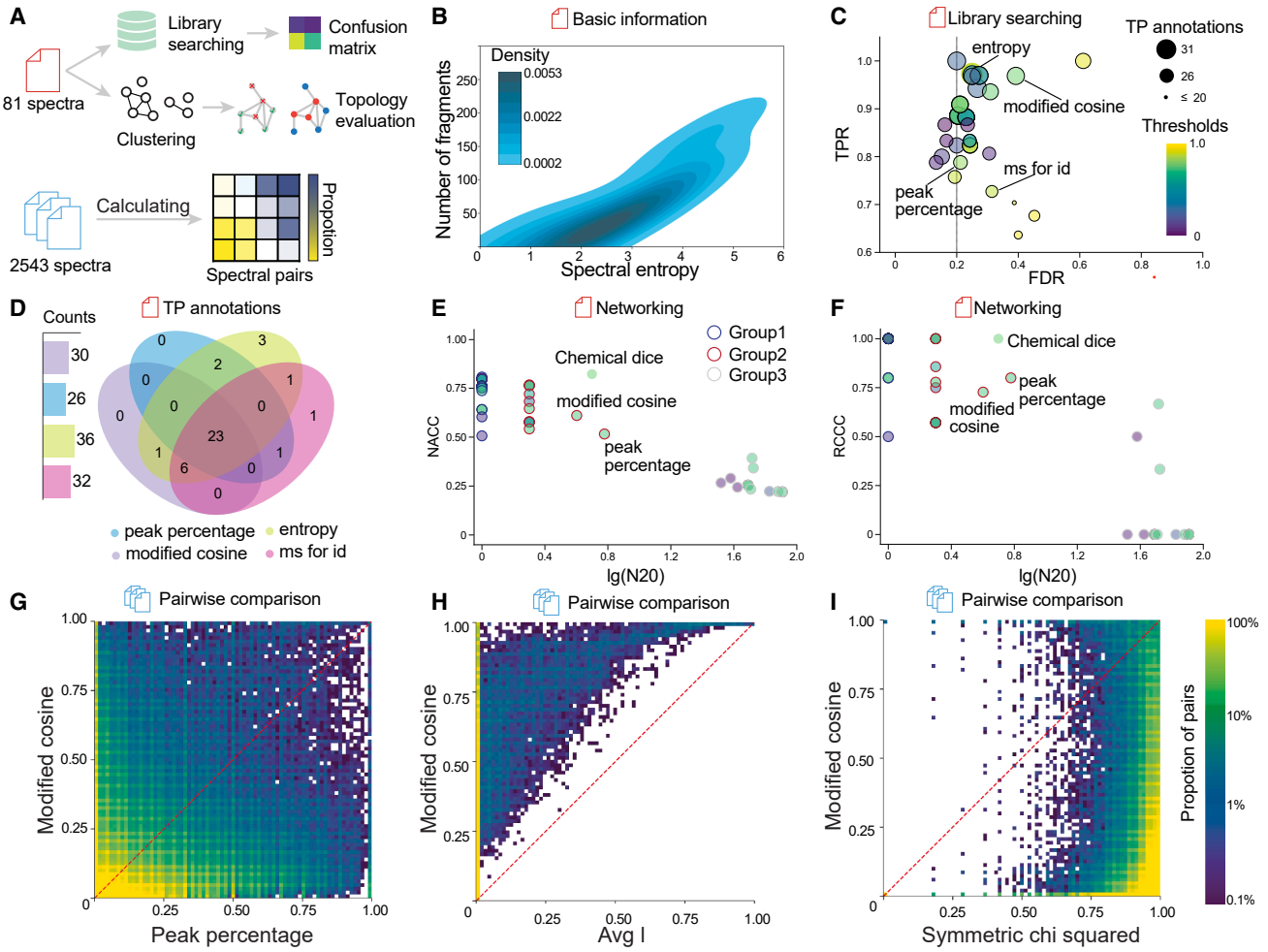


Figure 2. Comparative analysis of spectral similarity algorithms in MSanalyst

(A) Library searching and clustering were evaluated using an experimental dataset of microbial natural products (n = 81). The distribution of pairwise spectral similarity was calculated based on a curated dataset (n = 2543).

(B) Spectral entropy and peak numbers varied significantly in the dataset of microbial natural products.

(C) Different algorithms' library search results, filtered by optimal thresholds and top-scoring matches, can complement each other, (D) leading to more true positive (TP) annotations.

(E) Network accuracy and (F) the ratio of correctly classified clusters through N20 from networks generated by different spectral algorithms at their optimal thresholds; the chemical-dice-generated network represented the ideal situation.

(G) Peak percentage scored higher than modified cosine in 37% of spectral pairs sourced from the 2543 dataset.

(H) Avg_l was a subset of modified cosine, and (I) symmetric_chi_squared exhibited poor ability to reflect structural similarity.

and ignores intensity disparities (Figures S5 and S6). Confusion matrix analyses further showed that most algorithms experienced sharp TP declines only at very high thresholds, while FDRs decreased gradually with increasing stringency (Figures S7 and S8). In contrast, some algorithms maintained persistently high FDR values across thresholds (Figure S9). Importantly, the combined use of entropy and ms_for_id at their optimized thresholds captured all TP annotations identified by individual algorithms (Figure 2D). Thus, Youden's Index provides a practical criterion for algorithm-specific threshold optimization, with fine-tuning around these values to balance TP and FP annotations.

We performed network analysis to explore the algorithm's performance in identifying analogs. Based on three network indicators: N20, Network Accuracy of Classified Cluster (NACC) and Ratio of Correctly Classified Cluster (RCCC), algorithms were clustered into three groups (Figures 2E and 2F).[29] Group 1 algorithms generated a few small clusters with high accuracy and correct classification rates. Group 2 algorithms, including classic cosine-based methods, produced more and larger clusters, with modified cosine and peak percentage being the closest approximation to the ideal scenario (chemical dice similarity network). Group 3 algorithms produced large clusters with irrelevant structures, demonstrating low efficiency in correlating spectral and chemical similarity.

When scaled to three million spectral pairs, numerous algorithms scored higher than modified cosine in a significant fraction of comparisons, with peak percentage achieving up to 37% of spectral pairs (Figures 2G, S10, and S11). Some

Please cite this article in press as: Yu et al., Targeted discovery of aromatic glycosides with dual detoxification effects via a highly customized molecular networking platform, Cell Chemical Biology (2025), https://doi.org/10.1016/j.chembiol.2025.12.001

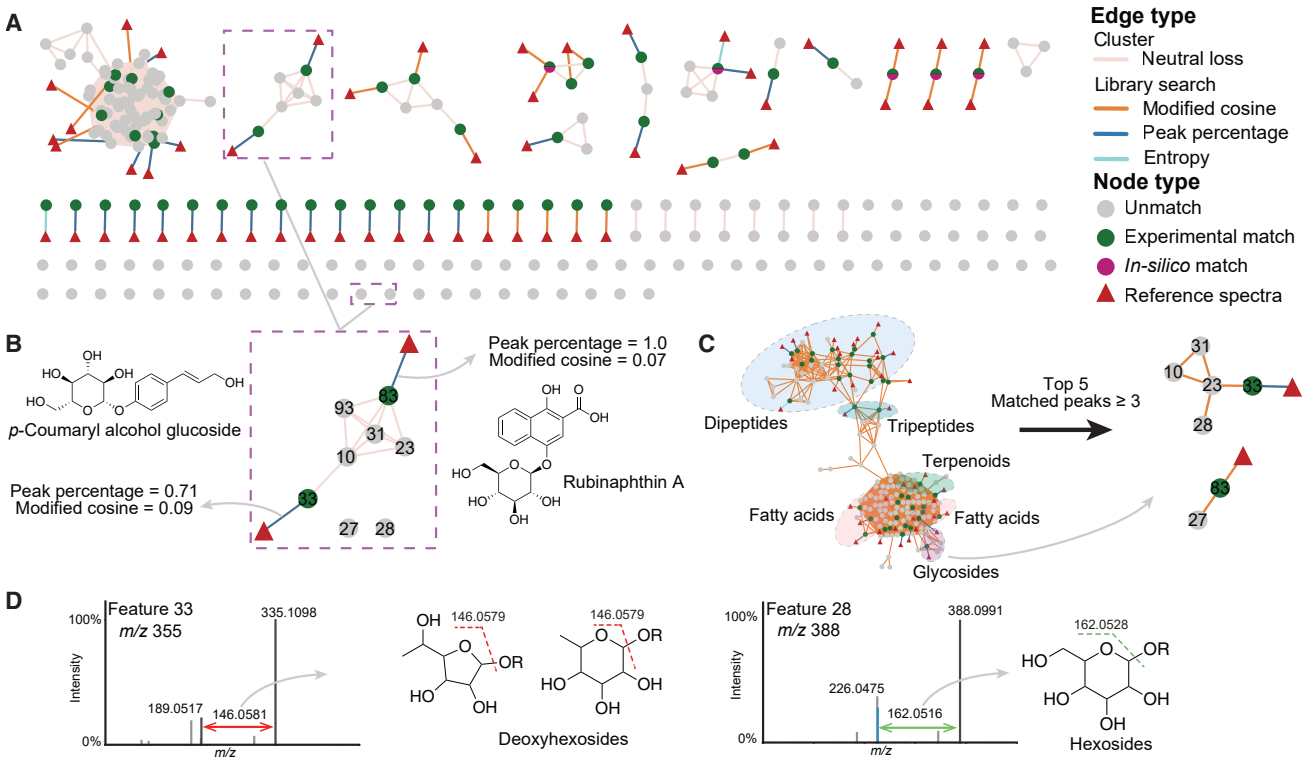


Figure 3. Integration of multiple algorithms enabled additional spectral connections that would have been overlooked by single approaches

(A) The MN was created in MSanalyst using modified cosine, peak percentage, and entropy for library search, and neutral loss for clustering, each applied at their optimal thresholds.

(B) Peak percentage identified additional glycoside annotations that would have been missed by any single algorithm. Feature IDs are shown on each node.

(C) The modified cosine produced a dense molecular network and utilized matched peaks and topK to limit the cluster, which helped distinguish the boundaries of molecular families.

(D) Manual inspection revealed that a neutral loss of 146 Da may originate from deoxyhexosides, whereas a neutral loss of 162 Da may originate from hexosides.

algorithms represented subsets of modified cosine (Figures 2H and S12), whereas others failed to capture structural similarity altogether (Figures 2I, S13, and S14). Collectively, these analyses help filter several low-accuracy and ineffective algorithms, thereby revealing 29 complementary algorithms—particularly peak percentage and entropy—which serve to enhance the establishment of spectral correlations (Data S1).

MSanalyst led to the discovery of kutznaposides A–F with antibacterial and antioxidant properties

MSanalyst integrates multiple algorithms to comprehensively explore the unknown metabolome, capturing features that may be missed by a single approach. It also provides opportunities to deepen our understanding of biosynthetic logic and biological functions. To showcase the power of MSanalyst, we targeted microbes that are rich in biosynthetic gene clusters (BGCs) encoding complex specialized metabolites with diverse activities, while their metabolomes have not yet been fully characterized. In our previous investigations searching for microbes with the potential to synthesize specialized metabolites with potent antimicrobial activities, the genus *Kutzneria* was found to harbor a diverse array of BGCs capable of encoding enzymes for synthesizing complex natural products.[30,31] However, although this genus is rich in BGCs, few corresponding natural products have been discovered, suggesting that many BGCs remain silent under laboratory conditions or that limited analytical tools have left them unexplored. Thus, to fully explore the metabolic potential of this genus and study its biological functions, we selected *Kutzneria viridogrisea* DSM 43850, which possesses the most abundant BGCs among commercially accessible *Kutzneria* strains, for further investigation. To maximize the activation of its BGCs, we cultured the strain using seven different media conditions and conducted LC-MS/MS analyses to thoroughly explore its chemical diversity.

Mass spectral data from crude extracts of *K. viridogrisea* DSM 43850 were processed using MSconvert and MZmine. After background subtraction, 214 molecular features were identified, and MSanalyst was used to profile its comprehensive metabolome. Algorithms that showed complementary performance with modified cosine were used for library searches. Together, modified cosine, peak percentage, and entropy identified 57 experimental and one *in silico* match, covering all annotated features detected by other algorithms (Figure 3A). Manual inspection led to the exclusion of synthetic drugs, such as fluocinolone acetonide and 9-DHAB III, and many common bacterial primary metabolites, including linear/cyclo dipeptides, imidazopyrimidines, and lipids (Figure S15). Additionally, two plant-derived glycosides stood out: feature 83 was annotated as rubinaphthin A[32] and feature 33 as p-coumaryl alcohol 4-*O*-glucoside[33] (Figure 3B). These

Please cite this article in press as: Yu et al., Targeted discovery of aromatic glycosides with dual detoxification effects via a highly customized molecular networking platform, Cell Chemical Biology (2025), https://doi.org/10.1016/j.chembiol.2025.12.001

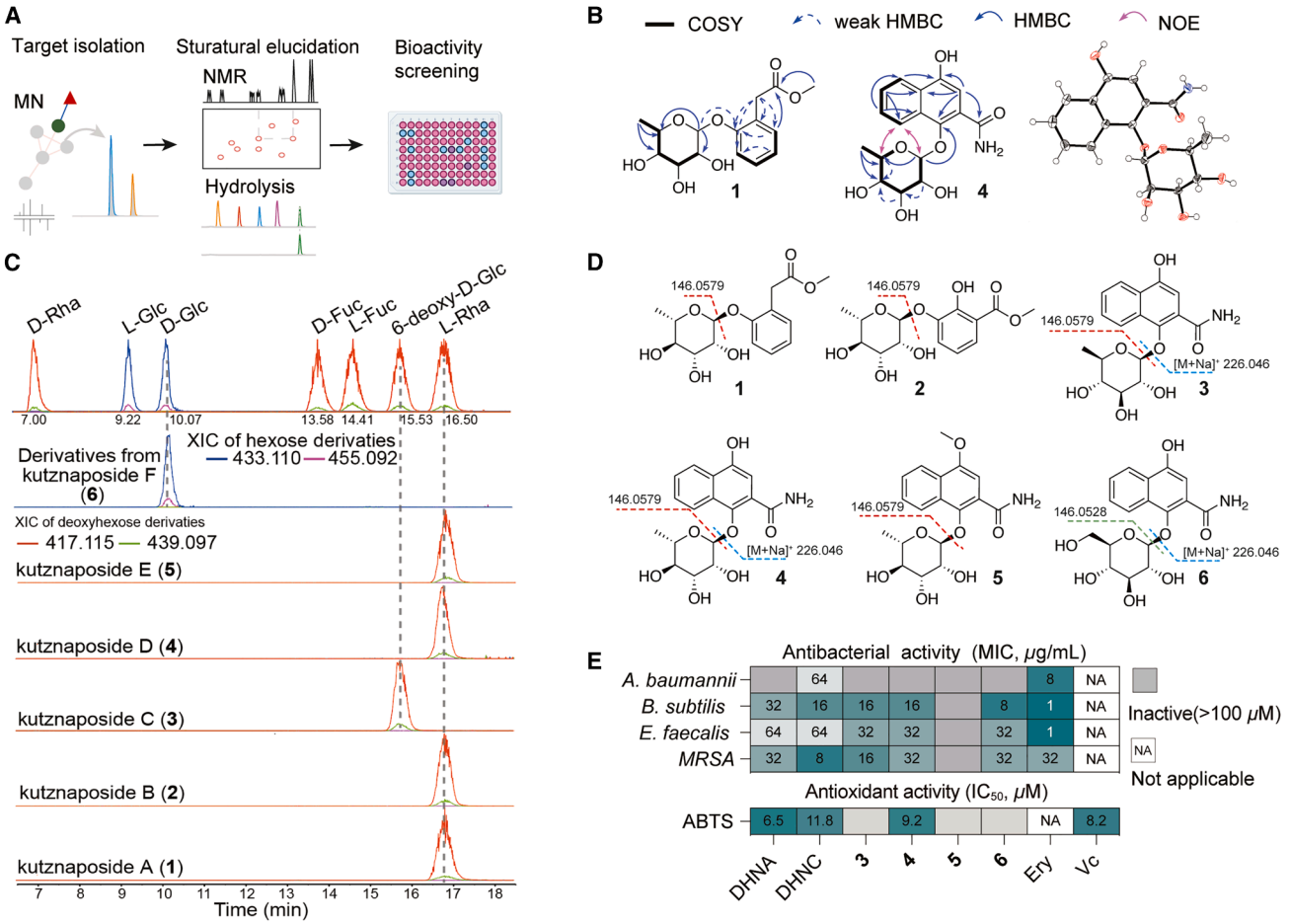


Figure 4. Structure elucidation of kutznaposides A–F (1–6) and their antimicrobial and antioxidant activities

(A) The cluster representing a class of phenolic glycosides was annotated by MSanalyst, followed by targted isolation, structural elucidation and bioactivity screening.

(B) NMR and crystallization were used to determine the scaffolds of the representative benzene and naphthalene glucosides.

(C) Extract ion chromatograms of monosaccharides (Hexose at m/z_{calcd} 433.110 of $[M + H]^+$ and m/z_{calcd} 455.092 of $[M+Na]^+$: L-glucose, 9.22 min; D-glucose, 10.07 min. Deoxyhexose at m/z_{calcd} 417.115 of $[M + H]^+$ and m/z_{calcd} 439.097 of $[M+Na]^+$: D-rhamnose, 7.00 min; D-fucose, 13.58 min; L-fucose, 14.41 min; 6-deoxy-D-glucose, 15.53 min; L-rhamnose, 16.5 min) and hydrolyzed samples derivatives.

(D) Absolute structures of compounds 1–6.

(E) The antibacterial and antioxidant activities of compounds 1–6, erythromycin (Ery), and vitamin C (Vc) served as the positive controls for antibacterial and antioxidant experiments, respectively.

glycosides scored highly with peak percentage (>0.7) but poorly with the modified cosine score (<0.1) against the reference spectrum (Figure S16), underscoring the risk of missing such compounds when relying solely on cosine scoring. Phenolic glycosides are rarely found in bacteria and may represent compounds of notable therapeutic and biological relevance.[34] Thus, we propagated these annotations through spectral similarity networks to search for additional analogs. At algorithm-specific thresholds, the two annotated glycoside features appeared as singletons in most networks, except those generated by neutral loss and modified cosine algorithms (Figure S17). The modified cosine-generated network linked the glycosides to their analogs but exhibits a crowded visualization architecture due to the allowance of precursor *m/z* shift matching. Additional parameters, such as topK and matched peaks, could help limit cluster sizes and improve separation between molecular families (Figure 3C). In contrast, the neutral loss-based network (Figure 3A) distinguished a well-defined glycoside cluster, reflecting consistent shared neutral losses.

From these analyses, we identified seven potential glycoside features. Specifically, feature 28 appeared to contain a hexoside moiety (neutral loss of 162.0528 Da), while the other molecular features contained a deoxyhexoside or deoxythiohexoside moiety (neutral loss of 146.0579 Da, Figure 3D).[35] Furthermore, five fragment ions resulting from sugar loss might suggest five different aglycones (Figure S18).

Integration of culture-medium metadata into the network revealed that these phenolic glycosides were detected exclusively in bacterial metabolomes cultivated in Gym Streptomyces Medium (GSM) (Figure S19). To validate the MS^2-predicted structures and assess their bioactivities, we scaled up fermentation in GSM medium, which yielded six previously undescribed aromatic glycosides, named kutznaposides A–F (Figure 4A).

Please cite this article in press as: Yu et al., Targeted discovery of aromatic glycosides with dual detoxification effects via a highly customized molecular networking platform, Cell Chemical Biology (2025), https://doi.org/10.1016/j.chembiol.2025.12.001

Their planar and absolute configurations were unambiguously established through comprehensive 1D/2D nuclear magnetic resonance (NMR) and mass spectrometric analyses (Figures 4B and S20–S26; Data S4. Correlation spectroscopy (COSY), heteronuclear single quantum coherence (HSQC), and Heteronuclear Multiple Bond Correlation (HMBC) spectra further demonstrated the presence of the sugar moiety in each compound (Figure 4C). Subsequent hydrolysis assays and the *J* value of every anomeric proton revealed that kutznaposides A, B, D, and E (1, 2, 4, and 5) contained the common α-L-rhamnose, which is replaced by β-deoxy-D-glucose and β-D-glucose in kutznaposide C (3) and kutznaposide F (6), respectively (Figure 4D). Substituted benzenes were characterized as the aglycons of 1 and 2, while 3, 4, 5, and 6 shared the same trisubstituted naphthalene backbone. Given that 2D NMR analysis could be misleading in some conjugated systems,[36] chemical synthesis (Figure S23) and crystallization approaches (Figure 4B, CCDC: 2373550) were performed to determine 1,4-dihydroxynaphthalene-2-carboxamide (DHNC) as the backbone of compounds 3, 4, and 6. In addition, a methoxyl signal was observed in the 1D NMR of 5, with 2D NMR further confirming its existence at the C-4 position. These combined methods not only verified the structures of kutznaposides but also provided sufficient DHNC for further bioactivity screening.

Hydroquinone derivatives, featuring a core benzene ring with two hydroxyl groups in the para position, exhibit robust antimicrobial activity and are strongly correlated with life-sustaining redox reactions.[37–39] Thus, we evaluated kutznaposides C–F, along with their precursors, against a panel of pathogenic microbes. Among them, 1,4-dihydroxy-2-naphthoic acid (DHNA), DHNC, and kutznaposides C (3), D (4), and F (6) exhibited different levels of inhibitory activity against *Acinetobacter baumannii*, *Bacillus subtillis*, *Enterococcus faecalis*, and methicillin-resistant *Staphylococcus aureus* (MRSA), with minimum inhibitory concentration (MIC) values ranging from 8 to 64 μg/mL (Figure 4E). In addition, the 2,2'-azino-bis(3-ethylbenzothiazoline-6-sulfonic acid (ABTS) assay indicated that DHNA, DHNC, and kutznaposide D (4) had comparable radical-scavenging capabilities (half maximal inhibtory concentration, IC_{50} values ranging from 6.5 to 9.2 μM) to that of the positive control, vitamin C (8.2 μM) (Figure 4E, inhibitory curve in Figure S27).

Kutznaposides C–F are bacterial menaquinone shunt products in *K. viridogrisea*

Kutznaposides C–F (4–6) represent a class of naphthol glycosides from microorganisms. They are synthesized through an uncommon shunt pathway that involves several unconventional enzymatic reactions, distinct from those typically found in the bacterial menaquinone pathway. The biosynthesis of kutznaposides starts with DHNA, an intermediate in the menaquinone (vitamin K2) pathway, which is synthesized from chorismate by seven *men*-enzymes encoded by *men*B to *men*F, *men*H, and *men*I.[40] Then, a hypothesized class II glutamine amidotransferase (GAT) might be involved in the transamination reaction, transferring a nitrogen group to the carboxylic acid group of DHNA.[41] Finally, uridine diphosphate glycosyltransferases (UGTs) and/or thymidine diphosphate glycosyltransferases (TGTs) are proposed to catalyze bacterial phenolic glycosylation.[42] These biochemical transformations of DHNA lead to bioactivity alterations and hint at possible biological functions (e.g., translation,[43] detoxification[44]) for the producer. Thus, we applied bioinformatic and proteomic analyses to decode these cryptic metabolic pathways. The protein annotation and hmmsearch results revealed that the putative homologs of *men*-encoded proteins, referred to as knps, were all present in strain DSM 43850 (Figure 5A and Data S2). As two kinds of glycosyltransferases (GTs), TGTs and UGTs, have only minor differences,[45] we used the reported bacterial GAT from *Streptomyces piomogeues* var. *Hangzhouwanensis*[46] and TGT ses60310 from *Saccharothrix espanaensis*[47] as reference sequences to identify candidate GATs and GTs in *K. viridogrisea*, respectively. After filtering the bioinformatically annotated results using proteomic data, we identified two GAT candidates and 16 potential GTs for further confirmation (Data S3). Based on the analysis of metabolomes, kutznaposides C–F were exclusively present in GSM medium, while DHNC was detected in lower quantities in R2A medium compared to GSM (Figure S28). Thus, we hypothesized that the related enzymes might exhibit differential expression in these two medium, providing an opportunity to understand how this shunt pathway was activated and to discover GAT and GT candidates. We then conducted proteomic profiling of *K. viridogrisea* enzymes under the growth conditions of GSM and R2A. Further investigation revealed the presence of two GATs in proteome analyses (Data S3), while only one GAT (GAT1) significantly increased under GSM cultural conditions. Notably, nine GTs showed a $\log_2$(fold change) > 1, and five of them (GTs 1–5) were exclusively expressed under GSM cultivation conditions. The remaining four GTs (GTs 6–9) showed a 2- to 6-fold increase in expression levels compared to those grown in R2A medium, suggesting their potential involvement in O-glycosylation (Figure 5B). Therefore, we performed heterologous expression of GTs 1–9 in *Escherichia coli* BL21. Among them, cells expressing GT5 and GT6 were able to convert DHNC into kutznaposides D and F, respectively (Figure 5C). Interestingly, these putative biosynthetic genes are dispersed across various genome loci and were not part of any BGCs predicted by antiSMASH (Figure 5D, Table S2). Moreover, neither GT5 nor GT6 glycosylated DHNA, indicating that glycosylation likely occurs after amidotransferase-mediated modification. Collectively, proteomic profiling and enzymatic assays support the biosynthetic pathway in which the production of kutznaposides is driven by *knp*-encoded machinery (KnpB1 to KnpF1, KnpH1, and KnpI1), followed by tailoring steps catalyzed by GAT and the identified GTs (Figure 5E). These findings indicate that kutznaposides are not readily identifiable using traditional bottom-up genome mining strategies, underscoring the powerful capability of MSanalyst to target valuable metabolites derived from noncanonical metabolic pathways directly.

The shunt flux from the menaquinone pathway functions as a self-detoxification strategy

Kutznaposides C–F are shunt products of the menaquinone pathway, whereas the original products, menaquinones, play a crucial role in microbial growth.[48,49] Therefore, GATs and GTs, which post-modify the key intermediate DHNA from the menaquinone pathway, are particularly intriguing. Moreover, since DHNA has been reported to exhibit self-harm effects,[39] and glycosylation/deglycosylation are known for their role in

Please cite this article in press as: Yu et al., Targeted discovery of aromatic glycosides with dual detoxification effects via a highly customized molecular networking platform, Cell Chemical Biology (2025), https://doi.org/10.1016/j.chembiol.2025.12.001

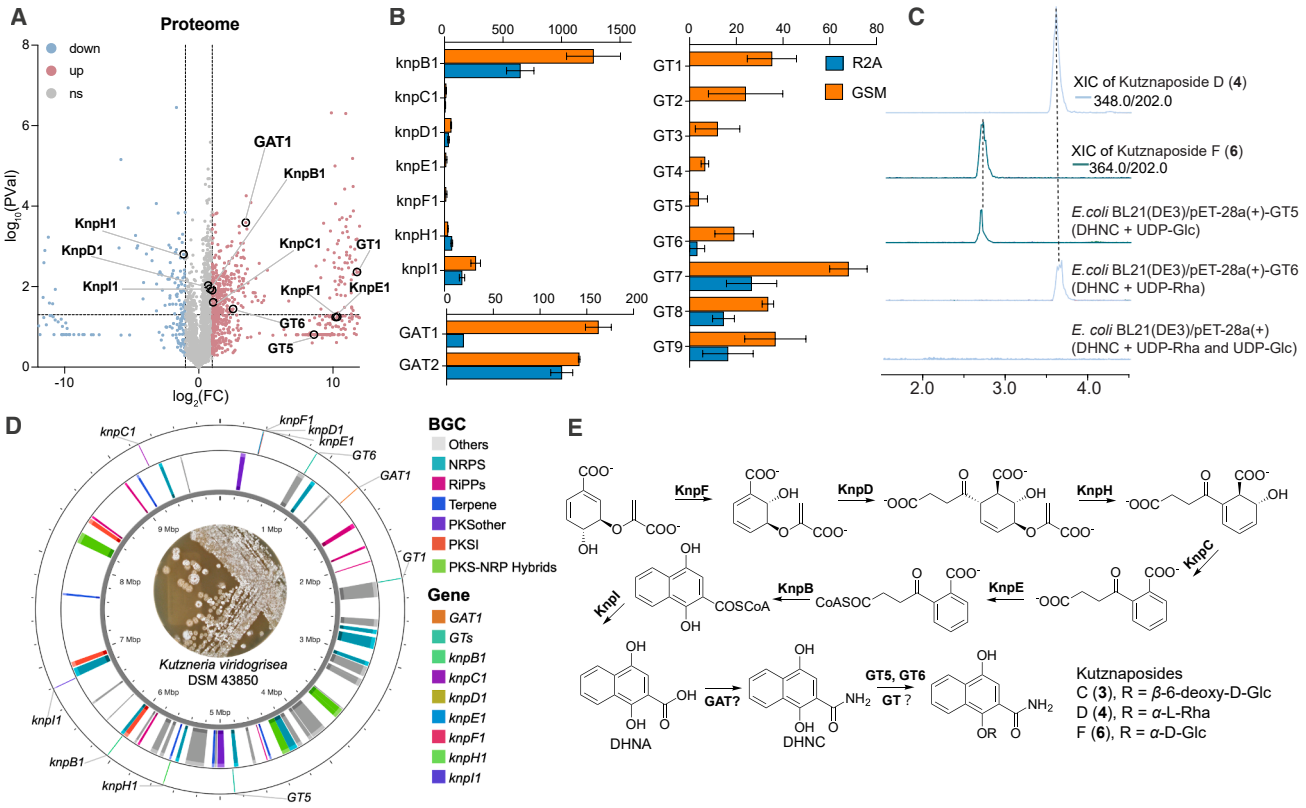


Figure 5. Proteomic investigation of key enzymes involved in the biosynthesis of kutznaposides C, D, and F

(A) Volcano plots and (B) protein expression in the proteome comparing GSM against R2A medium are depicted. Absolute expression is illustrated in a bar plot on the right (n = 3).

(C) Detection of glycosylated metabolites (4, the putative 4-*O*-glucoside; 6, the putative 4-*O*-rhamnoside) in the reaction mixtures of whole-cell catalysis using *E. coli* BL21(DE3) expressing GT5 and GT6, with DHNC and UDP-Glc/Rha.

(D) Whole genome assembly with the inner ring showing detected BGCs in different colors and the outer ring displaying distributed genes involved in the

(E) proposed biosynthesis of kutznaposides C, D, and F (3, 4, and 6).

self-resistance,[50] we propose that these two enzymes serve protective roles in enabling *K. viridogrisea* to avoid self-harm from DHNA. To test this hypothesis, a self-inhibitory growth assay was conducted using DHNA, DHNC, and kutznaposides C, D, and F. As expected, DHNA showed the most potent inhibitory effect against *K. viridogrisea*, followed by the downstream products DHNC and kutznaposides (Figure 6A). These findings suggest that this two-step enzymatic reaction represents a self-detoxification process, and the associated genes, GATs and GTs, are considered self-resistance genes.

To understand the mechanisms underlying the evolution and utilization of both the conventional and shunt menaquinone pathways in microorganisms,[51] as well as the activation factors of this discovered metabolic pathway, we conducted in-depth analyses of the proteomic data. The enrichment analysis revealed several Gene Ontology (GO) terms (Data S3) associated with detoxification observed in GSM medium, such as cellular response to xenobiotic stimulus (GO: 0071466) and xenobiotic metabolic process (GO: 0006805) (Figure 6B). Additionally, the top two enriched Kyoto Encyclopedia of Genes and Genomes (KEGG) metabolic pathways were alanine, aspartate, and glutamate metabolism (KEGG: map00250) and propanoate metabolism (KEGG: map00460) (Figure 6C). Almost all the genes were significantly upregulated (Data S3), suggesting a strong correlation with the activation of the shunt pathway.

We subsequently extracted these four sets of proteins from the overall protein-protein interaction (PPI) network of *K. viridogrisea* (Figure S30) to determine if specific enzymes interact with GAT1 (Figure 6D). The abundance of quinolinate synthase (EC: 2.5.1.72) and L-aspartate oxidase (EC: 1.4.3.16), which are involved in the first two steps of nicotinamide adenine dinucleotide (NAD^+) conversion from quinolinate, exhibited an indirect correlation with GAT1 (Figure 6D). The upregulation of these two enzymes could lead to the accumulation of reactive oxygen species (ROS)[52,53] (Figure 6E). During the oxidation of glyoxylate, the lack of catalase (EC: 1.11.1.6) and the upregulation of (*S*)-2-hydroxy-acid oxidase (EC: 1.1.3.15) could also result in the accumulation of ROS. This may be attributed to the more robust aerobic metabolism of *K. viridogrisea* in GSM compared to R2A, supported by the observation that all proteins involved in the TCA cycle exhibited increased expression levels (Figure 6E), and the biomass from GSM was approximately twice that of R2A (Figure S31). Thus, we hypothesize that *K. viridogrisea* biosynthesizes DHNA to scavenge excessive ROS, but this compound is toxic to the producer. To counteract this, the organism has evolved a mechanism to divert metabolic

Please cite this article in press as: Yu et al., Targeted discovery of aromatic glycosides with dual detoxification effects via a highly customized molecular networking platform, Cell Chemical Biology (2025), https://doi.org/10.1016/j.chembiol.2025.12.001

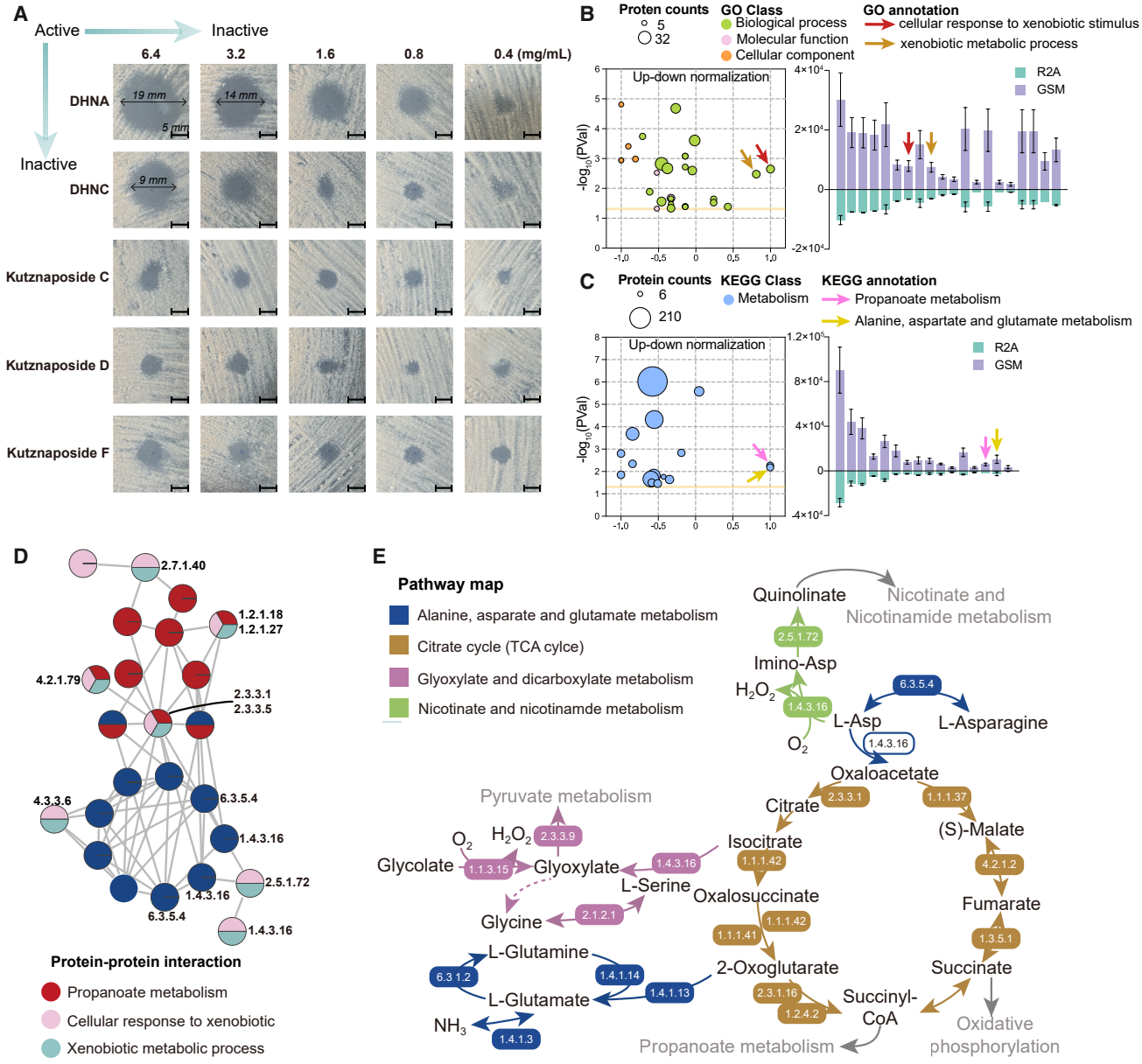


Figure 6. Multi-omic analyses reveal the menaquinone shunt pathway as an uncommon self-detoxification strategy

(A) Effects of *K. viridogrisea* on DHNA and its downstream metabolites (black bar, 5 mm) compared to DSMO control in Figure S29.

(B) Top 20 GO annotations and (C) top 17 KEGG pathways based on differentially expressed proteins detected in the proteome. In the bubble charts, the ordinate represents $-\log_{10}(p \text{ value})$, and the abscissa represents the proportion of up-down normalization (the difference between the number of up-regulated and down-regulated proteins among the total differential proteins). Yellow lines indicate the p value = 0.05 threshold, and different colors represent various ontologies. The left side shows bar plots of absolute protein expression within each ontology, summarized as the height of each bar. Arrows point to the top two highly up-regulated GO/KEGG annotated protein sets.

(D) Protein-protein interaction network of the most highly related gene sets analyzed in (B) and (C).

(E) Overview of the pathway map related to the enhanced aerobic metabolism of *K. viridogrisea* DSM 43850. All enzymes in color-filled boxes in the figure were significantly up-regulated ($Log_2FC > 1$).

flux toward the production of kutznaposides, which neutralize ROS and prevent self-toxicity. Therefore, we propose that this bacteria-originated menaquinone shunt pathway employs a series of unconventional biochemical transformations as a dual strategy for surviving in oxidative environments and self-detoxifying the antioxidants they produce.

DISCUSSION

MN is a powerful tool in LC-MS-based untargeted metabolome profiling, enabling clear annotation and propagation of structural information. However, its utility is constrained by the size of available spectral libraries and by reliance on a single spectral

Please cite this article in press as: Yu et al., Targeted discovery of aromatic glycosides with dual detoxification effects via a highly customized molecular networking platform, Cell Chemical Biology (2025), https://doi.org/10.1016/j.chembiol.2025.12.001

algorithm.[22] To address these limitations, we expanded the reference library with *in silico* spectra and integrated alternative algorithms. Although the inclusion of *in silico* libraries increases annotation coverage, FP annotations remain a major concern. Filtering *in silico* annotations using experimentally validated annotations within the same cluster significantly reduced the FDR. Evaluation across microbial natural product datasets and curated spectral pairs showed that algorithms such as entropy and peak percentage complement the classic modified cosine approach, improving annotation depth. The appropriate combination of alternative algorithms was able to fully explore the unidentified metabolome of *K. viridogrisea*. Multi-algorithm-based molecular networks revealed additional spectral relationships and annotations that remained hidden with a single similarity metric.

To maximize accessibility, we provide both web-based and stand-alone versions of MSanalyst, with detailed documentation and a tutorial video. All code and data are publicly available to ensure reproducibility. Because unknown metabolites display wide spectral diversity, fixed similarity and shared peak thresholds are not universally applicable. MSanalyst thus supports rapid re-analysis by adjusting thresholds or switching algorithms according to the metabolite of interest.

Library search evaluations demonstrated that strict spectral thresholds reduce the FDR but risk filtering out TPs, especially for sparsely fragmented spectra. Fine-tuning optimizes performance for each algorithm. Further, variations in calculation methods, such as allowing *m/z* shift matching, adjusting fragment weightings, or altering noise sensitivity, result in different similarity scores. Our evaluations indicate that combining multiple algorithms significantly improves both annotation efficiency and biological relevance. In cases where one algorithm scores higher than another, both TP and FP annotations may be present, which is one possible reason for the opposing conclusions about the accuracy of different algorithms.[17,54] Since the composition of the database could affect the overall accuracy, this reminds us that each algorithm is suited to different contexts. This emphasizes the necessity for multiple reliable algorithms to establish additional spectral connections. Taken together, these findings suggest that combining complementary metrics enhances the annotation of unknown metabolomic space and strengthens existing tools.

Application of MSanalyst enabled the discovery of a series of specialized glycoside metabolites, kutznaposides A–F, from *K. viridogrisea*. Although neutral-loss similarity is technically a subset of modified cosine, in this case it more clearly distinguished the glycoside family boundaries, whereas modified cosine alone produced a crowded, less interpretable cluster. Kutznaposides represent an undescribed class of bacterial-derived aromatic glycosides, with kutznaposides C–F identified as shunt products of the menaquinone pathway. Bioinformatic and multi-omic studies suggested that the enzymes involved in the biosynthesis of kutznaposides C–F are scattered across the genome and do not form a dedicated biosynthetic gene cluster. This indicates that these chemicals will be overlooked by genome mining approaches such as antiSMASH[55] in bottom-up natural product discovery. Meanwhile, MSanalyst optimizes the MN approach, providing a substantial complement to the traditional top-down natural product discovery strategy and significantly enhancing the ability to discover these cryptic metabolites associated with uncommon biological functions.

While the enzymes involved in the menaquinone pathway have been thoroughly studied, further investigation into how this pathway may be regulated is needed. This study investigates the underlying mechanism behind the pathway's shunt flux through proteomic analyses, coupled with *in vitro* bioactivity assays. We hypothesize that *K. viridogrisea* has evolved a series of enzymatic reactions to survive in an oxidative environment and prevent self-toxicity. As glycosyl modification is a common detoxification strategy, these self-resistance genes could lay the foundation for the discovery of other BGCs containing similar genes. Given the essential role of menaquinone in the pathogenesis of environmental and human pathogens,[51] our findings could facilitate the development of molecular probes for identifying potential drug targets.

Together, these findings highlight the power of MSanalyst in efficiently annotating unidentified metabolites and driving the discovery of natural products. By integrating and optimizing existing metabolomic approaches, we can gain deeper insights into data interpretation. We envision MSanalyst as an important metabolomics tool that, in synergy with multi-omics, bioactivity assays, and other advanced methods, will significantly propel the field of chemical biology forward.

Limitations of the study

Several limitations of our current methods and resources warrant consideration. First, although the MSanalyst platform integrates advanced machine-learning models (e.g., CFM-ID) for virtual MS/MS library generation, the systematic gap between predicted and experimental spectra still constrains its effectiveness for natural product annotation. Second, the annotation accuracy of different spectral similarity algorithms depends strongly on dataset composition. Spectral noise, low-quality fragment ions, and the lack of standardized benchmark datasets collectively limit our ability to rigorously evaluate these algorithms. Finally, the proposed self-protection mechanisms of *K. viridogrisea* remain preliminary and lack direct genetic or biochemical support, which will need to be addressed through future gene-knockout and *in vitro* reconstitution studies.

RESOURCE AVAILABILITY

Lead contact

Further queries and reagent requests may be directed to and will be fulfilled by the lead contact, Bin Wei (binwei@zjut.edu.cn).

Materials availability

All unique and stable reagents generated in this study are available from the lead contact with a completed materials transfer agreement.

Data and code availability

- The NMR and mass spectra for characterization of kutznaposides can be found in Data S4. Raw NMR data are available with specific IDs (NP-MRD: NP0351510, NP0351511, NP0351512, NP0351513, NP0351514, NP0351515). Crystallographic data for compound 4 (CCDC: 2373550) can be found in the Cambridge Crystallographic Data Center website (https://www.ccdc.cam.ac.uk/).
- Raw.mzML, Mzmine 2 preprocessed.mgf, and in-house microbial.mgf files used in this study were deposited in the MassIVE data repository and are publicly available (Massive: MSV000096272).

- The FBMN-generated network in Figure 3A can be consulted at https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=95c1c3dd1d0f4a15a0d414d55136862c. Data are also available on the website (https://msanalyst.net/i/manuals).
- The whole genome of *K. viridogrisea* DSM 43850 was uploaded to the National Center for Biotechnology Information with the accession number (NCBI: PRJNA1209424). Source data are provided with this paper.
- The stand-alone version of MSanalyst and relevant codes and data used in this study are available on GitHub at https://github.com/WenchYu/MSanalyst and from Zenodo (Record: 16356728).
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

ACKNOWLEDGMENTS

This work was financially supported by the National Key Research and Development Programs (nos. 2022YFC2804700 and 2022YFC2804104), the Fundamental Research Funds for the Provincial Universities of Zhejiang (no. RF-A2022013), the program of the National Natural Science Foundation of China (no. 42276137), the National 111 Project (no. D17012), the Zhejiang International Sci-Tech Cooperation Base for the Exploitation and Utilization of Nature Products and International (Regional) Cooperation and Exchange Program jointly initiated by the National Natural Science Foundation of China (No. W2412100). The authors would like to thank the Shiyanjia Lab (www.shiyanjia.com) for the XRD test. The authors would also like to acknowledge the valuable contributions from all three reviewers. Their suggestions have been extremely helpful in improving our manuscript.

AUTHOR CONTRIBUTIONS

B.W., H.W., Q.W., and W.-C.Y. conceived the study. W.-C.Y. developed the platform and carried out the mass spectrometric, bioinformatic, and statistical analyses. W.-C.Y., B.-C.D., and A.-Q.D. performed the chemical and biological experiments. Y.-L.Y., Z.-Y.W., Q.X., S.-W.L., B.B.B., X.-Z.B., X.-R.S., and X.-N.L. provided helpful discussions and technical assistance. W.-C.Y., Q.W., and B.W. wrote the manuscript with assistance from all authors. All authors approved the final version of the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
- METHOD DETAILS
  - Construction of MSanalyst
  - Comparative analysis of the spectral similarity algorithms
  - Metabolome analysis of *Kutzneria viridogrisea* DSM 43850 and target isolation
  - Structure elucidation of kutznaposides A-F
  - LC-MS-based enzymatic activity assay of glycosyltransferases
  - Protein preparation and mass spectrometry analysis
  - Genomic and proteomic analysis
  - Antimicrobial and ABTS decolorization assay
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.chembiol.2025.12.001.

Received: September 17, 2024
Revised: September 7, 2025
Accepted: December 3, 2025

REFERENCES

1. Atanasov, A.G., Zotchev, S.B., Dirsch, V.M., International Natural Product Sciences Taskforce, and Supuran, C.T. (2021). Natural products in drug discovery: advances and opportunities. Nat. Rev. Drug Discov. *20*, 200–216. https://doi.org/10.1038/s41573-020-00114-z.
2. Scott, T.A., and Piel, J. (2019). The hidden enzymology of bacterial natural product biosynthesis. Nat. Rev. Chem *3*, 404–425. https://doi.org/10.1038/s41570-019-0107-1.
3. van Bergeijk, D.A., Terlouw, B.R., Medema, M.H., and van Wezel, G.P. (2020). Ecology and genomics of Actinobacteria: new concepts for natural product discovery. Nat. Rev. Microbiol. *18*, 546–558. https://doi.org/10.1038/s41579-020-0379-y.
4. Wang, M., Carver, J.J., Phelan, V.V., Sanchez, L.M., Garg, N., Peng, Y., Nguyen, D.D., Watrous, J., Kapono, C.A., Luzzatto-Knaan, T., et al. (2016). Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. Nat. Biotechnol. *34*, 828–837. https://doi.org/10.1038/nbt.3597.
5. Damiani, T., Jarmusch, A.K., Aron, A.T., Petras, D., Phelan, V.V., Zhao, H.N., Bittremieux, W., Acharya, D.D., Ahmed, M.M.A., Bauermeister, A., et al. (2025). A universal language for finding mass spectrometry data patterns. Nat. Methods *22*, 1247–1254. https://doi.org/10.1038/s41592-025-02660-z.
6. Wang, M., Jarmusch, A.K., Vargas, F., Aksenov, A.A., Gauglitz, J.M., Weldon, K., Petras, D., da Silva, R., Quinn, R., Melnik, A.V., et al. (2020). Mass spectrometry searches using MASST. Nat. Biotechnol. *38*, 23–26. https://doi.org/10.1038/s41587-019-0375-9.
7. Gentry, E.C., Collins, S.L., Panitchpakdi, M., Belda-Ferre, P., Stewart, A.K., Carrillo Terrazas, M., Lu, H.-h., Zuffa, S., Yan, T., Avila-Pacheco, J., et al. (2024). Reverse metabolomics for the discovery of chemical structures from humans. Nature *626*, 419–426. https://doi.org/10.1038/s41586-023-06906-8.
8. El Abiead, Y., Strobel, M., Payne, T., Fahy, E., O'Donovan, C., Subramamiam, S., Vizcaíno, J.A., Yurekten, O., Deleray, V., Zuffa, S., et al. (2025). Enabling pan-repository reanalysis for big data science of public metabolomics data. Nat. Commun. *16*, 4838. https://doi.org/10.1038/s41467-025-60067-y.
9. Watrous, J., Roach, P., Alexandrov, T., Heath, B.S., Yang, J.Y., Kersten, R.D., van der Voort, M., Pogliano, K., Gross, H., Raaijmakers, J.M., et al. (2012). Mass spectral molecular networking of living microbial colonies. Proc. Natl. Acad. Sci. USA *109*, E1743–E1752. https://doi.org/10.1073/pnas.1203689109.
10. Nothias, L.-F., Petras, D., Schmid, R., Dührkop, K., Rainer, J., Sarvepalli, A., Protsyuk, I., Ernst, M., Tsugawa, H., Fleischauer, M., et al. (2020). Feature-based molecular networking in the GNPS analysis environment. Nat. Methods *17*, 905–908. https://doi.org/10.1038/s41592-020-0933-6.
11. Dührkop, K., Shen, H., Meusel, M., Rousu, J., and Böcker, S. (2015). Searching molecular structure databases with tandem mass spectra using CSI:FingerID. Anal. Chem. *112*, 12580–12585. https://doi.org/10.1073/pnas.1509788112.
12. Wang, F., Liigand, J., Tian, S., Arndt, D., Greiner, R., and Wishart, D.S. (2021). CFM-ID 4.0: More Accurate ESI-MS/MS Spectral Prediction and Compound Identification. Anal. Chem. *93*, 11692–11700. https://doi.org/10.1021/acs.analchem.1c01465.
13. Allard, P.-M., Péresse, T., Bisson, J., Gindro, K., Marcourt, L., Pham, V.C., Roussi, F., Litaudon, M., and Wolfender, J.-L. (2016). Integration of Molecular Networking and In-Silico MS/MS Fragmentation for Natural Products Dereplication. Anal. Chem. *88*, 3317–3323. https://doi.org/10.1021/acs.analchem.5b04804.

14. Treen, D.G.C., Wang, M., Xing, S., Louie, K.B., Huan, T., Dorrestein, P.C., Northen, T.R., and Bowen, B.P. (2022). SIMILE enables alignment of tandem mass spectra with statistical significance. Nat. Commun. *13*, 2510. https://doi.org/10.1038/s41467-022-30118-9.

15. Olivon, F., Elie, N., Grelier, G., Roussi, F., Litaudon, M., and Touboul, D. (2018). MetGem Software for the Generation of Molecular Networks Based on the t-SNE Algorithm. Anal. Chem. *90*, 13900–13908. https://doi.org/10.1021/acs.analchem.8b03099.

16. Huber, F., Ridder, L., Verhoeven, S., Spaaks, J.H., Diblen, F., Rogers, S., van der Hooft, J.J.J., and van der Hooft, J.J.J. (2021). Spec2Vec: Improved mass spectral similarity scoring through learning of structural relationships. PLoS Comput. Biol. *17*, e1008724. https://doi.org/10.1371/journal.pcbi.1008724.

17. Li, Y., Kind, T., Folz, J., Vaniya, A., Mehta, S.S., and Fiehn, O. (2021). Spectral entropy outperforms MS/MS dot product similarity for small-molecule compound identification. Nat. Methods *18*, 1524–1531. https://doi.org/10.1038/s41592-021-01331-z.

18. Aisporna, A., Benton, H.P., Chen, A., Derks, R.J.E., Galano, J.M., Giera, M., and Siuzdak, G. (2022). Neutral Loss Mass Spectral Data Enhances Molecular Similarity Analysis in METLIN. J. Am. Soc. Mass Spectrom. *33*, 530–534. https://doi.org/10.1021/jasms.1c00343.

19. Beniddir, M.A., Kang, K.B., Genta-Jouve, G., Huber, F., Rogers, S., and van der Hooft, J.J.J. (2021). Advances in decomposing complex metabolite mixtures using substructure- and network-based computational metabolomics approaches. Nat. Prod. Rep. *38*, 1967–1993. https://doi.org/10.1039/d1np00023c.

20. Kontou, E.E., Walter, A., Alka, O., Pfeuffer, J., Sachsenberg, T., Mohite, O.S., Nuhamunada, M., Kohlbacher, O., and Weber, T. (2023). UmetaFlow: an untargeted metabolomics workflow for high-throughput data processing and analysis. J. Cheminf. *15*, 52. https://doi.org/10.1186/s13321-023-00724-w.

21. de Jonge, N.F., Hecht, H., Strobel, M., Wang, M., van der Hooft, J.J.J., and Huber, F. (2024). Reproducible MS/MS library cleaning pipeline in matchms. J. Cheminf. *16*, 88. https://doi.org/10.1186/s13321-024-00878-1.

22. Bittremieux, W., Wang, M., and Dorrestein, P.C. (2022). The critical role that spectral libraries play in capturing the metabolomics community knowledge. Metabolomics *18*, 94. https://doi.org/10.1007/s11306-022-01947-y.

23. Lyu, C., Chen, T., Qiang, B., Liu, N., Wang, H., Zhang, L., and Liu, Z. (2021). CMNPD: a comprehensive marine natural products database towards facilitating drug discovery from the ocean. Nucleic Acids Res. *49*, D509–D515. https://doi.org/10.1093/nar/gkaa763.

24. van Santen, J.A., Jacob, G., Singh, A.L., Aniebok, V., Balunas, M.J., Bunsko, D., Neto, F.C., Castaño-Espriu, L., Chang, C., Clark, T.N., et al. (2019). The Natural Products Atlas: An Open Access Knowledge Base for Microbial Natural Products Discovery. ACS Cent. Sci. *5*, 1824–1833. https://doi.org/10.1021/acscentsci.9b00806.

25. Sorokina, M., Merseburger, P., Rajan, K., Yirik, M.A., and Steinbeck, C. (2021). COCONUT online: Collection of Open Natural Products database. J. Cheminf. *13*, 2. https://doi.org/10.1186/s13321-020-00478-9.

26. Morehouse, N.J., Clark, T.N., McMann, E.J., van Santen, J.A., Haeckl, F.P.J., Gray, C.A., and Linington, R.G. (2023). Annotation of natural product compound families using molecular networking topology and structural similarity fingerprinting. Nat. Commun. *14*, 308. https://doi.org/10.1038/s41467-022-35734-z.

27. Yang, J.Y., Sanchez, L.M., Rath, C.M., Liu, X., Boudreau, P.D., Bruns, N., Glukhov, E., Wodtke, A., de Felicio, R., Fenner, A., et al. (2013). Molecular Networking as a Dereplication Strategy. J. Nat. Prod. *76*, 1686–1699. https://doi.org/10.1021/np400413s.

28. Subtil, F., and Rabilloud, M. (2014). Estimating the optimal threshold for a diagnostic biomarker in case of complex biomarker distributions. BMC Med. Inform. Decis. Mak. *14*, 53. https://doi.org/10.1186/1472-6947-14-53.

29. Wang, X., Strobel, M., Aron, A.T., Phelan, V.V., Acharya, D.D., Brown, C.J., Clevenger, K., Hu, J., Kretsch, A., Mahood, E.H., et al. (2024). Network Topology Evaluation and Transitive Alignments for Molecular Networking. J. Am. Soc. Mass Spectrom. *35*, 2165–2175. https://doi.org/10.1021/jasms.4c00208.

30. Wei, B., Du, A.-Q., Ying, T.-T., Hu, G.-A., Zhou, Z.-Y., Yu, W.-C., He, J., Yu, Y.-L., Wang, H., and Xu, X.-W. (2023). Secondary Metabolic Potential of Kutzneria. J. Nat. Prod. *86*, 1120–1127. https://doi.org/10.1021/acs.jnatprod.3c00007.

31. Wei, B., Du, A.Q., Zhou, Z.Y., Lai, C., Yu, W.C., Yu, J.B., Yu, Y.L., Chen, J.W., Zhang, H.W., Xu, X.W., and Wang, H. (2021). An atlas of bacterial secondary metabolite biosynthesis gene clusters. Environ. Microbiol. *23*, 6981–6992. https://doi.org/10.1111/1462-2920.15761.

32. Liou, M.-J., Wu, P.-L., and Wu, T.-S. (2002). Constituents of the Roots of Rubia yunnanensis. Chem. Pharm. Bull. *50*, 276–279. https://doi.org/10.1248/cpb.50.276.

33. Bernards, M.A., Fleming, W.D., Llewellyn, D.B., Priefer, R., Yang, X., Sabatino, A., and Plourde, G.L. (1999). Biochemical characterization of the suberization-associated anionic peroxidase of potato. Plant Physiol. *121*, 135–146.

34. Johnson, J.B., Mani, J.S., Broszczak, D., Prasad, S.S., Ekanayake, C.P., Strappe, P., Valeris, P., and Naiker, M. (2021). Hitting the sweet spot: A systematic review of the bioactivity and health benefits of phenolic glycosides from medicinally used plants. Phytother Res. *35*, 3484–3508. https://doi.org/10.1002/ptr.7042.

35. Kersten, R.D., Ziemert, N., Gonzalez, D.J., Duggan, B.M., Nizet, V., Dorrestein, P.C., and Moore, B.S. (2013). Glycogenomics as a mass spectrometry-guided genome-mining method for microbial glycosylated molecules. Proc. Natl. Acad. Sci. USA *110*, E4407–E4416. https://doi.org/10.1073/pnas.1315492110.

36. Burns, D.C., and Reynolds, W.F. (2021). Minimizing the risk of deducing wrong natural product structures from NMR data. Magn. Reson. Chem. *59*, 500–533. https://doi.org/10.1002/mrc.4933.

37. Lewis, K. (2024). Streptonigrin kills bacteria by stealth. Proc. Natl. Acad. Sci. USA *121*, e2320942121. https://doi.org/10.1073/pnas.2320942121.

38. Mevers, E., Su, L., Pishchany, G., Baruch, M., Cornejo, J., Hobert, E., Dimise, E., Ajo-Franklin, C.M., and Clardy, J. (2019). An elusive electron shuttle from a facultative anaerobe. eLife *8*, e48054. https://doi.org/10.7554/eLife.48054.

39. Stanborough, T., Ho, N.A.T., Bulloch, E.M.M., Bashiri, G., Dawes, S.S., Akazong, E.W., Titterington, J., Allison, T.M., Jiao, W., and Johnston, J.M. (2023). Allosteric inhibition of Staphylococcus aureus MenD by 1,4-dihydroxy naphthoic acid: a feedback inhibition mechanism of the menaquinone biosynthesis pathway. Philos. Trans. R. Soc. Lond. B Biol. Sci. *378*, 20220035. https://doi.org/10.1098/rstb.2022.0035.

40. Johnston, J.M., and Bulloch, E.M. (2020). Advances in menaquinone biosynthesis: sublocalisation and allosteric regulation. Curr. Opin. Struct. Biol. *65*, 33–41. https://doi.org/10.1016/j.sbi.2020.05.005.

41. Massière, F., and Badet-Denisot, M.-A. (1998). The mechanism of glutamine-dependent amidotransferases. Cell. Mol. Life Sci. *54*, 205–222.

42. Williams, G.J., Goff, R.D., Zhang, C., and Thorson, J.S. (2008). Optimizing Glycosyltransferase Specificity via "Hot Spot" Saturation Mutagenesis Presents a Catalyst for Novobiocin Glycorandomization. Chem. Biol. *15*, 393–401. https://doi.org/10.1016/j.chembiol.2008.02.017.

43. Curnow, A.W., Hong, K.-w., Yuan, R., Kim, S.-i., Martins, O., Winkler, W., Henkin, T.M., and Söll, D. (1997). Glu-tRNAGln amidotransferase: a novel heterotrimeric enzyme required for correct decoding of glutamine codons during translation. Proc. Natl. Acad. Sci. USA *94*, 11819–11826.

44. Li, X., and Ju, J. (2024). Intracellularly driven chemical modifications of antimicrobial secondary metabolites: Potent mechanisms of self-resistance. Pharmaceutical Science Advances *2*, 100032. https://doi.org/10.1016/j.pscia.2023.100032.

45. Cho, K.W., Kim, T.-S., Le, T.T., Nguyen, H.T., Oh, Y.S., Pandey, R.P., and Sohng, J.K. (2019). Altering UDP-glucose Donor Substrate Specificity of Bacillus licheniformis Glycosyltransferase towards TDP-glucose.

J. Microbiol. Biotechnol. *29*, 268–273. https://doi.org/10.4014/jmb.1811.11009.

46. Liu, Q., Yao, F., Chooi, Y.H., Kang, Q., Xu, W., Li, Y., Shao, Y., Shi, Y., Deng, Z., Tang, Y., and You, D. (2012). Elucidation of Piericidin A1 Biosynthetic Locus Revealed a Thioesterase-Dependent Mechanism of α-Pyridone Ring Formation. Chem. Biol. *19*, 243–253. https://doi.org/10.1016/j.chembiol.2011.12.018.
47. Strobel, T., Schmidt, Y., Linnenbrink, A., Luzhetskyy, A., Luzhetska, M., Taguchi, T., Brötz, E., Paululat, T., Stasevych, M., Stanko, O., et al. (2013). Tracking Down Biotransformation to the Genetic Level: Identification of a Highly Flexible Glycosyltransferase from Saccharothrix espanaensis. Appl. Environ. Microbiol. *79*, 5224–5232. https://doi.org/10.1128/aem.01652-13.
48. Bentley, R., and Meganathan, R. (1982). Biosynthesis of vitamin K (menaquinone) in bacteria. Microbiol. Rev. *46*, 241–280. https://doi.org/10.1128/mr.46.3.241-280.1982.
49. Meganathan, R. (2001). Biosynthesis of menaquinone (vitamin K2) and ubiquinone (coenzyme Q): A perspective on enzymatic mechanisms. In Vitamins & Hormones, *61*, G. Litwack and T. Begley, eds. (Academic Press), pp. 173–218. https://doi.org/10.1016/S0083-6729(01)61006-9.
50. Zhang, Y., Bai, J., Zhang, L., Zhang, C., Liu, B., and Hu, Y. (2021). Self-Resistance in the Biosynthesis of Fungal Macrolides Involving Cycles of Extracellular Oxidative Activation and Intracellular Reductive Inactivation. Angew. Chem. Int. Ed. *60*, 6639–6645. https://doi.org/10.1002/anie.202015442.
51. Gatsios, A., Kim, C.S., York, A.G., Flavell, R.A., and Crawford, J.M. (2022). Cellular Stress-Induced Metabolites in Escherichia coli. J. Nat. Prod. *85*, 2626–2640. https://doi.org/10.1021/acs.jnatprod.2c00706.
52. Chen, H., Nwe, P.-K., Yang, Y., Rosen, C.E., Bielecka, A.A., Kuchroo, M., Cline, G.W., Kruse, A.C., Ring, A.M., Crawford, J.M., and Palm, N.W. (2019). A Forward Chemical Genetic Screen Reveals Gut Microbiota Metabolites That Modulate Host Physiology. Cell *177*, 1217–1231.e18. https://doi.org/10.1016/j.cell.2019.03.036.
53. Saunders, A.H., and Booker, S.J. (2008). Regulation of the activity of Escherichia coli quinolinate synthase by reversible disulfide-bond formation. Biochemistry *47*, 8467–8469.
54. Dlugas, H., Zhang, X., and Kim, S. (2025). Comparative analysis of continuous similarity measures for compound identification in mass spectrometry-based metabolomics. Chemometr. Intell. Lab. Syst. *263*, 105417. https://doi.org/10.1016/j.chemolab.2025.105417.
55. Blin, K., Shaw, S., Augustijn, H.E., Reitz, Z.L., Biermann, F., Alanjary, M., Fetter, A., Terlouw, B.R., Metcalf, W.W., Helfrich, E.J.N., et al. (2023). antiSMASH 7.0: new and improved predictions for detection, regulation, chemical structures and visualisation. Nucleic Acids Res. *51*, W46–W50. https://doi.org/10.1093/nar/gkad344.
56. Pluskal, T., Castillo, S., Villar-Briones, A., and Orešič, M. (2010). MZmine 2: modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. BMC Bioinf. *11*, 395–411.
57. Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. Bioinformatics *30*, 2068–2069. https://doi.org/10.1093/bioinformatics/btu153.
58. Stothard, P., and Wishart, D.S. (2005). Circular genome visualization and exploration using CGView. Bioinformatics *21*, 537–539. https://doi.org/10.1093/bioinformatics/bti054.
59. Zhang, Y., Yang, C., Wang, J., Wang, L., Zhao, Y., Sun, L., Sun, W., Zhu, Y., Li, J., and Wu, S. (2024). BioLadder: A bioinformatic platform primarily focused on proteomic data analysis. iMeta *3*, e215. https://doi.org/10.1002/imt2.215.
60. Cantalapiedra, C.P., Hernández-Plaza, A., Letunic, I., Bork, P., Huerta-Cepas, J., and Tamura, K. (2021). eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale. Mol. Biol. Evol. *38*, 5825–5829. https://doi.org/10.1093/molbev/msab293.
61. Chen, C., Chen, H., Zhang, Y., Thomas, H.R., Frank, M.H., He, Y., and Xia, R. (2020). TBtools: An Integrative Toolkit Developed for Interactive Analyses of Big Biological Data. Mol. Plant *13*, 1194–1202. https://doi.org/10.1016/j.molp.2020.06.009.
62. Szklarczyk, D., Kirsch, R., Koutrouli, M., Nastou, K., Mehryary, F., Hachilif, R., Gable, A.L., Fang, T., Doncheva, N.T., Pyysalo, S., et al. (2023). The STRING database in 2023: protein–protein association networks and functional enrichment analyses for any sequenced genome of interest. Nucleic Acids Res. *51*, D638–D646. https://doi.org/10.1093/nar/gkac1000.
63. Melnik, A.V., da Silva, R.R., Hyde, E.R., Aksenov, A.A., Vargas, F., Bouslimani, A., Protsyuk, I., Jarmusch, A.K., Tripathi, A., Alexandrov, T., et al. (2017). Coupling Targeted and Untargeted Mass Spectrometry for Metabolome-Microbiome-Wide Association Studies of Human Fecal Samples. Anal. Chem. *89*, 7549–7559. https://doi.org/10.1021/acs.analchem.7b01381.
64. Sud, M. (2016). MayaChemTools: An Open Source Package for Computational Drug Discovery. J. Chem. Inf. Model. *56*, 2292–2297. https://doi.org/10.1021/acs.jcim.6b00505.
65. Wei, B., Ying, T.-T., Lv, H.-W., Zhou, Z.-Y., Cai, H., Hu, G.-A., Liang, H.-M., Yu, W.-C., Yu, Y.-L., Fan, A.-L., et al. (2025). Global analysis of fungal biosynthetic gene clusters reveals the diversification of diketopiperazine biosynthesis. Bioresour. Technol. *422*, 132218. https://doi.org/10.1016/j.biortech.2025.132218.
66. Yang, M., Chen, T., Liu, Y.-X., and Huang, L. (2024). Visualizing set relationships: EVenn's comprehensive approach to Venn diagrams. iMeta *3*, e184. https://doi.org/10.1002/imt2.184.
67. Kim, H.W., Wang, M., Leber, C.A., Nothias, L.-F., Reher, R., Kang, K.B., van der Hooft, J.J.J., Dorrestein, P.C., Gerwick, W.H., and Cottrell, G.W. (2021). NPClassifier: A Deep Neural Network-Based Structural Classification Tool for Natural Products. J. Nat. Prod. *84*, 2795–2807. https://doi.org/10.1021/acs.jnatprod.1c00399.
68. Myers, O.D., Sumner, S.J., Li, S., Barnes, S., and Du, X. (2017). One Step Forward for Reducing False Positive and False Negative Compound Identifications from Mass Spectrometry Metabolomics Data: New Algorithms for Constructing Extracted Ion Chromatograms and Detecting Chromatographic Peaks. Anal. Chem. *89*, 8696–8703. https://doi.org/10.1021/acs.analchem.7b00947.
69. Wang, Y.-H., Avula, B., Fu, X., Wang, M., and Khan, I.A. (2012). Simultaneous Determination of the Absolute Configuration of Twelve Monosaccharide Enantiomers from Natural Products in a Single Injection by a UPLC-UV/MS Method. Planta Med. *78*, 834–837. https://doi.org/10.1055/s-0031-1298432.
70. Bubb, W.A. (2003). NMR spectroscopy in the study of carbohydrates: Characterizing the structural complexity. Concepts Magn. Reson. *19A*, 1–19. https://doi.org/10.1002/cmr.a.10080.
71. Chen, B., Wang, Y., Xie, F., Liu, H., and Dai, H. (2024). Identification of siderophores blocking infection of Pseudomonas aeruginosa from Kitasatospora sp. J. Antibiot. *77*, 4–12. https://doi.org/10.1038/s41429-023-00675-2.
72. Zhang, T., Zhong, H., Lin, L., Zhang, Z., Xue, K., He, F., Luo, Y., Wang, P., Zhao, Z., Cong, L., et al. (2024). Core microbiome-associated proteins associated with ulcerative colitis interact with cytokines for synergistic or antagonistic effects on gut bacteria. ISME J. *18*, wrae146.

Please cite this article in press as: Yu et al., Targeted discovery of aromatic glycosides with dual detoxification effects via a highly customized molecular networking platform, Cell Chemical Biology (2025), https://doi.org/10.1016/j.chembiol.2025.12.001

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Bacterial and virus strains		
Kutzneria viridogrisea	DSMZ	DSM 43850
E.coli BL21(DE3)/pET-28a(+)-GT5	This study	N/A
E.coli BL21(DE3)/pET-28a(+)-GT6	This study	N/A
Acinetobacter baumannii ATCC 19606	Nanjing Medical University	N/A
Bacillus subtitles ATCC 25923	Nanjing Medical University	N/A
Enterococcus faecalis FA2-2	Nanjing Medical University	N/A
Methicillin-resistant *Staphylococcus aureus* USA300.	Nanjing Medical University	N/A
Oligonucleotides		
GT5-F GTTTAACTTTAAGAAGGAGAATGGGGAAGCACTT CGCGTTTG	This study	N/A
GT5-R CTGTGATGATGATGATGATGCCGGGACAGGTACT CCTCCAGT	This study	N/A
GT6-F AACTTTAAGAAGGAGATATACCATGAGAGTCCTGG TCACCACCTC	This study	N/A
pET28a-GT5-F CATCATCATCATCATCACAGCA	This study	N/A
pET28a-GT5-R TCTCCTTCTTAAAGTTAAACAAAATTATTT	This study	N/A
pET28a-GT6-F CACCACCACCACCACCACTG	This study	N/A
pET28a-GT6-R GGTATATCTCCTTCTTAAAGTTAAACAAAATTA	This study	N/A
Chemicals, peptides, and recombinant proteins		
Kuznaposide A	This study	N/A
Kuznaposide B	This study	N/A
Kuznaposide C	This study	N/A
Kuznaposide D	This study	N/A
Kuznaposide E	This study	N/A
Kuznaposide F	This study	N/A
Glucose	Greagent	Cat# G24010A
Yeast extract	Adamas	Cat# 84106G
Malt extract	Hopebio	Cat# HB9178
R2A medium	Biofeng	Cat# M1318-01
Soluble·starch	Adamas	Cat# G8493
Tryptone	Adamas	Cat# M8084
NaCl	Adamas	Cat# BH0221
Mueller Hinton Broth medium	Millipore	Cat# 70192
Bacto Marine Broth DIFCO 2216 medium	Topbio	Cat# M0015B
Tryptic Soya Broth medium	Millipore	Cat# 1.05459
XAD16N	Supelco	Cat# XAD16
Methanol (AC)	Adamas	Cat# 75851AY
Methanol (LC/MS)	Supelco	Cat# 1.06035
Methanol-d4	Leyan	Cat# C400003

(*Continued on next page*)

Please cite this article in press as: Yu et al., Targeted discovery of aromatic glycosides with dual detoxification effects via a highly customized molecular networking platform, Cell Chemical Biology (2025), https://doi.org/10.1016/j.chembiol.2025.12.001

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Dimethyl sulfoxide	Solarbio	Cat# ID9010
Dimethyl sulfoxide-d6	Leyan	Cat# C400002
Dichloromethane	Adamas	Cat# 81014J-ZX
MCI gel	Supelco	Cat# 13630-U
Sephadex LH-20	Supelco	Cat# LH20100
Acetilenitrile (HPLC)	Supelco	Cat# 6.18030
1,4-Dihydroxy-2-naphthoate (DHNA)	Sigma-Aldrich	Cat# 281255
Phenyl 1,4-dihydroxy-2-naphthoate	Aladdin	Cat# P160594
Deoxygenated tetrahydrofuran	Sigma-Aldrich	Cat# 186562
Deoxygenated methanol	Aladdin	Cat# M116117-1L
Ammonium acetate	Aladdin	Cat# A471867
Ammonium hydroxide	Macklin	Cat# A765771
Hydrochloric acid	Sigma-Aldrich	Cat# 258148
D-glucose	Solarbio	Cat# ILA3069
L-glucose	Sigma-Aldrich	Cat# G5500
D-rhamnose	Biosynth	Cat# MR02762
L-rhamnose	Sigma-Aldrich	Cat# R3875
D-fucose	Solarbio	Cat# IF3390
L-fucose	Solarbio	Cat# F8410
6-deoxy-D-glucose	Yuanyebio	Cat# S31795
Trifluoroacetic acid	Aladdin	Cat# T103294
L-cysteine methyl ester	Sigma-Aldrich	Cat# 410209
Phenyl isothiocyanate	Aladdin	Cat# P110635
Formic acid	Sigma-Aldrich	Cat# 695076
Isopropanol	Sangon	Cat# A600918
Kanamycin	Sangon	Cat# A414917
Isopropyl β-D-Thiogalactopyranoside	Solarbio	Cat# I8070
UDP-glucose	Masterbio	Cat# S18074
UDP-rhamnose	Guidechem	Cat# 325325
TDP-glucose	Chemsrc	N/A
TDP-rhamnose	Chemsrc	N/A
Erythromycin	Chemsrc	Cat# A600192
2,2′-azino-bis(3-ethylbenzothiazoline-6-sulfonic acid (ABTS·+)	Solarbio	Cat# IA0010
K_2S_2O8	Aladdin	Cat# P433868
PBS	Hopebio	Cat# HBXBHP001
Vitamin C	Aladdin	Cat# A103536
Critical commercial assays		
Universal Genomic DNA Purification Mini Spin Kit	Beyotime	Cat# D0063
FastPure Plasmid Mini Kit	Vazyme	Cat# DC201
2 × Phanta Flash Master Mix	Vazyme	Cat# P520
FastPure Gel DNA Extraction Mini Kit	Vazyme	Cat# DC301
2 × ClonExpress Mix	Vazyme	Cat# C115
Software and algorithms		
Relevant code of MSanalyst	This study	https://github.com/WenchYu/MSanalyst
MSconvert	ProteoWizard	https://proteowizard.sourceforge.io/index.html
MZmine (version 2.5.1)	Pluskal et al.[56]	https://github.com/mzmine
CFM-ID (version 4.4.7)	Wang et al.[12]	https://hub.docker.com/r/wishartlab/cfmid
RDKit (version2024.9.6)	N/A	https://www.rdkit.org/

(*Continued on next page*)

Please cite this article in press as: Yu et al., Targeted discovery of aromatic glycosides with dual detoxification effects via a highly customized molecular networking platform, Cell Chemical Biology (2025), https://doi.org/10.1016/j.chembiol.2025.12.001

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Ms_entropy (version 1.3.4)	Li et al.[17]	https://github.com/YuanyueLi/MSEntropy
SpectralEntropy (version 1.0.2)	Li et al.[17]	https://github.com/YuanyueLi/SpectralEntropy
Match_ms (version 0.30.2)	de Jonge et al.[21]	https://github.com/matchms/matchms
FPsim2 (version 0.7.3)	N/A	https://github.com/chembl/FPSim2
Networkx (version 3.4.2)	N/A	https://github.com/networkx/networkx
Cyoscape (version 3.9.1)	N/A	https://cytoscape.org/
MestReNova (version 14.2.0)	mestrelab	https://mestrelab.com/download
Prokka (version 1.14.6)	Seemann et al.[57]	https://github.com/tseemann/prokka
CGview Server	Stothard et al.[58]	https://proksee.ca/
BioLadder	Zhang et al.[59]	https://www.bioladder.cn/web/
eggNOG-mapper (version 2.1.12)	Cantalapiedra et al.[60]	https://github.com/eggnogdb/eggnog-mapper
TBtools (version 2.092)	Chen et al.[61]	https://github.com/CJ-Chen/TBtools-II
STRING	Szklarczyk et al.[62]	https://string-db.org/
GraphPad Prism (version 10.4.2)	GraphPad	https://www.graphpad.com/features
Adobe illustrator 2025	Adobe	https://www.adobe.com/sg/products/illustrator.html
Deposited data		
Natural products atls (NPatls)	van Santen et al.[24]	https://www.npatlas.org/
The collection of open natural products (COCONUT)	Sorokina et al.[25]	https://coconut.naturalproducts.net/
Comprehensive marine natural products database (CMNPD)	Lyu et al.[23]	https://www.cmnpd.org/
The MSanalyst library	This study	https://zenodo.org/records/17088906
Metabolome of *K. viridogrisea and* MS^2 of pure isolates from *Aspergillus* sp. WHUF0304 (MSV000096272)	This study	https://massive.ucsd.edu/ProteoSAFe/dataset.jsp?task=2d7601057c30471e9375a20e89eda336
MS^2 used to generate dataset for Figure 2 (MSV000079758)	Melnik et al.[63]	https://gnps.ucsd.edu//ProteoSAFe/status.jsp?task=ac7d5ce7603345b99fd252bc56185aa5
Combined MS^2 dataset in Data S1	This study	https://zenodo.org/records/17088906
Raw NMR data of kutznaposides	This study	https://depositions.np-mrd.org/request-data/8987a2cc-f8c0-4ebd-ac16-48a54d9b109c
Others		
BEH C18 (2.1 × 50 mm, 1.7 μm) column	Waters	Cat# 186002350
XDB-C18 (9.4 * 250 mm, 5 μm)	Agilent	Cat# 990967-202
ZenoTOFTM 7600 mass spectrometer	SCIEX	N/A
One LC system	Evisep	N/A
Novaseq 6000	Illumina	N/A
Milli-Q Direct water purification system	Millipore	N/A
AVANCE III NMR	Bruker	N/A
Quadrupole Time-of-Flight Mass spectrometry X500B	SCIEX	N/A
Triple Quad 3500	SCIEX	N/A

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Strain *K. viridogrisea* JCM3282 (DSM 43850) was acquired from the Japan Collection of Microorganisms (JCM) Culture Collection Center (https://jcm.brc.riken.jp/en/). In small-scale fermentation, *K. viridogrisea* was cultivated in GSM, R2A, MHB, LB, 2216E, TSB, and M42 at 30°C with shaking at 180 rpm for 7 days. Media composition was listed in Table S3.

The growth inhibition assay of the tested compounds was performed against four bacterial pathogens: *Acinetobacter baumannii* ATCC 19606, *Bacillus subtilis* ATCC 25923, *Enterococcus faecalis* FA2-2, and Methicillin-resistant *Staphylococcus aureus* USA300. All strains are provided by Jiangsu Key Laboratory of Modern Pathogen Biology, Nanjing Medical University. These strains were cultured in LB medium at 37°C with shaking at 180 rpm.

Please cite this article in press as: Yu et al., Targeted discovery of aromatic glycosides with dual detoxification effects via a highly customized molecular networking platform, Cell Chemical Biology (2025), https://doi.org/10.1016/j.chembiol.2025.12.001

CellPress

METHOD DETAILS

Construction of MSanalyst

Generation of MS^2 library. The structures and relevant information for natural products were obtained from three databases: CMNPD (version 1.0, $n = 31560$),[23] NPAtlas (version 2024.09, $n = 32553$),[24] and COCONUT (version 1.0, $n = 406919$).[25] To ensure consistency and reliability, a molecular standardization step was applied to all Simplified Molecular Input Line Entry Systems (SMILES), which involved removing salts, neutralizing atoms, and eliminating duplicates and permanently charged compounds by mayachemtools (version 2025.06.22).[64] This resulted in a collection of 467,126 non-redundant natural products, and their theoretical adducts were calculated using RDKit (version 2024.9.6). These structures were then used to predict *In-silico* MS^2 spectra in protonated adducts under electrospray ionization (ESI) in positive mode at three different collision energies (Energy0: 10 V, Energy1: 20 V, and Energy2: 40 V) by the CFM-ID (Competitive Fragmentation Modeling, version 4.4.7).[12] Experimental MS^2 reference spectra were obtained from the GNPS library (ALL_GNPS_NO_PROPOGATED, downloaded on 2025.7.8).

Operation of spectral similarity algorithms. Before calculating spectral similarities, the spectra were preprocessed by removing peaks with *m/z* higher than the precursor *m/z*, centroiding peaks within 0.02 Da, and eliminating ions below 1% of maximum intensity. All the spectral algorithms were listed in Table S1. The cosine, modified cosine, and neutral loss were performed by matchms (version 0.21.1).[21] Different algorithms for handling signal intensity differences in spectral pairs can lead to varying similarity outcomes, with potential positive or negative impacts. Since signal differences are unavoidable in real mass data acquisition, we incorporate peak percentage to fully eliminate this effect and focus solely on the shared peaks between the compared spectra. The peak percentage was determined by dividing the matched peaks by the minimal fragments of the spectral pair. The remaining algorithms were performed by SpectralEntropy (version 1.0.2).[17]

Molecular networking pipeline. MSanalyt accepts several formats of spectral data (".json", ".mgf", ".msp", ".mzML", and ".mzXML"). An UmetaFlow workflow preprocessed raw data by detecting mass traces with a noise threshold and mass tolerance, then deconvoluted using width filtering with a default minimum of 1s and a maximum of 60s full-width-at-half-maximum for chromatographic peaks. Satellite isotopic traces were appended to a feature map. The feature map with the most features serves as the reference for aligning retention times. Raw spectral data are aligned based on feature map alignments. MS^2 spectra are mapped to aligned features and merged from multiple maps into a single consensus feature based on their shared retention time and *m/z* values. The consensus feature map is identical to the data from MZmine for submission to GNPS and can be used as input for MSanalyst. The molecular networking pipeline begins by screening query features against both the experimental and *in-silico* reference libraries to identify precursor-ion matches within a user-defined mass tolerance. Features without any MS^1 or MS^2 match are flagged as "unmatched". Spectral similarities are calculated between precursor-matched features and reference spectra using set algorithms; similarities above the set thresholds can be considered as annotations. *In-silico* annotations are retained only if they are similar to at least one experimental annotation within the same molecular family, with a chemical Dice similarity of 0.75 or higher (Morgen fingerprint, radius 2, 1024 nBits 1024).[26] Pairwise spectral similarities are calculated, and edges are created between nodes to form the molecular network when the calculated spectral similarities exceed the set threshold. Matched peaks and topK allow neighboring nodes of a single node to be used to limit the cluster size. After generating the network, annotations are added as additional nodes connected to their parent features, allowing for a clear visualization of how the spectral connection is established by specific algorithms.

MSanalyst is available as a web and stand-alone version (https://msanalyst.net/) with instructional videos and detailed documentation. Supplementary modules enhance the capabilities of the core pipeline by providing additional functionalities: Each analysis is assigned a unique task identifier to ensure reproducibility, facilitate data sharing, and support downstream integration. The "Reanalysis" module enables swift re-optimization of thresholds and the merging of networks produced with varying parameters or algorithms. The "Spectrum searching" feature allows for quick queries within libraries pertaining to individual metabolites of interest. The "Database customizing" module supports user-curated, project-specific spectral libraries and also promotes the expansion of reference data.

Comparative analysis of the spectral similarity algorithms

To better mimic real-world conditions, we used an experimental microbial natural product dataset to evaluate the results of library searching and network. This microbial natural product spectra dataset included 82 MS^2 spectra from protonated precursors from two parts: 42 spectra acquired in positive mode from the pure reference isolates of *Aspergillus* sp. WHUF0304 (upload to MSV000096272)[65] and 39 MS^2 spectra downloaded from MassIVE spectral library (MSV000079758, https://gnps.ucsd.edu//ProteoSAFe/status.jsp?task=ac7d5ce7603345b99fd252bc56185aa5).[63] Library search used the combined spectral dataset by all 46 spectral similarity algorithms against the GNPS library (ALL_GNPS_NO_PROPOGATED, downloaded on 2025.7.8) and the generated *in-silico* library with an allowed mass tolerance of 0.02 Da or 10 ppm. Due to the lack of prior knowledge of less-used algorithms, we determined the optimal thresholds for each algorithm within a similarity range of 0–1, using Youden's J statistic by sklearn (version 1.1.3) to maximize (true positive rate - false positive rate) (TPR - FPR), representing their optimal performance in library search. Since the number of fragments and spectral entropy varied greatly within the dataset, we compared the library searching results with and without matched peaks. The reference spectrum with the top 1 similarity to each query feature was included in the next confusion matrix evaluation. If the calculated similarity exceeded the spectral threshold, it was considered positive (P); if not, it was considered negative (N). Then, based on the structural similarity of the compounds corresponding to the two compared spectra, if the chemical

Please cite this article in press as: Yu et al., Targeted discovery of aromatic glycosides with dual detoxification effects via a highly customized molecular networking platform, Cell Chemical Biology (2025), https://doi.org/10.1016/j.chembiol.2025.12.001

dice (Morgen fingerprint, radius 2, 1024 nBits 1024) calculated by FPSim2 exceeded 0.75, the prediction was true (T); otherwise, it was false (F). Simply put, FDR = FP/(FP + TP), TPR = TP/(TP + FN), FPR = FP/(FP + TN), TP propotion = TP/81. The intersection of TP annotations obtained by modified cosine, peak percentage, entropy and ms_for_id was visualized by Evenn.[66]

Pairwise similarities of spectra from the same natural product dataset were calculated using 46 metrics with a 0.02 Da mass tolerance. The same threshold used for each algorithm in library searching was applied to create the molecular network. The networks were then evaluated using N20, network accuracy, and the ratio of correctly classified clusters (RCCC), and these indicators performed similarly to those in the literature, with minor modifications.[29] Chemical dice (Morgen fingerprint, radius 2, nBits 1024) and a cut-off. A value of 0.75 was employed to characterize the structural relationships between the structures. The RCCC was calculated under the NPclassifier[67] "superclass" level and retrieved from the API by SMILES ("https://npclassifier.gnps2.org/classify?smiles = "). The classification of whether a component is correct depends on whether its purity exceeds the threshold of "$\geq$0.7".

Comparative analysis of similarity distribution on a curated small molecule dataset. Dissimilar fragmentation behavior of the same molecule can be observed due to different mass acquisition methods. MS^2 spectra with protonated and positively charged precursors, obtained using the same mass spectrometry (qTOF), were filtered from the GNPS library (GNPS-LIBRARY, downloaded on 2025.7.8, https://external.gnps2.org/gnpslibrary) to eliminate the influence on MS^2 comparison, comprising the second dataset (n = 2571). All 46 spectral algorithms performed the general pairwise spectral calculation. All the mass spectra before comparison were preprocessed by removing peaks within a 0.1 m/z window higher than the precursor m/z, centroid peaks by merging peaks with 0.05 Da, and removing peaks with intensity below 1% of the base peak intensity.

Metabolome analysis of *Kutzneria viridogrisea* DSM 43850 and target isolation

The fermentation scale of *K. viridogrisea* DSM 43850 was increased to 40 L using GSM under identical conditions. The fermented culture underwent ultrasound treatment to disrupt the mycelium, followed by centrifugation to remove cells. The resulting supernatant was concentrated through evaporation. The concentrated supernatant was thoroughly extracted using macroporous resin XAD 16N (Dow Chemical Co., Ltd.), yielding a total of 19.3 g of crude extract. The targeted isolation of kutznaposides was conducted through MS-guided isolation, employing various chromatographic techniques.

The crude extracts of *K. viridogrisea* were analyzed using LC-qTOF-MS. Chromatographic separation was performed on an ExionLC high-performance LC system. LC-ESI-MS/MS analysis was performed by coupling the LC system to a hybrid quadrupole time-of-flight mass spectrometer (X500B, SCIEX AB, Germany), equipped with an electron spray ionization (ESI) source. HPLC-grade methanol was purchased from Merck, and HPLC-grade water was obtained by filtration using a Milli-Q Direct water purification system (Millipore). The LC conditions were as follows: column, Agilent poroshell 120 SB-C18 50 $\times$ 2.1 mm, 1.7 μm; mobile phase, (A) water and (B) methanol; flow rate, 0.3 mL/min; injection volume, 1 μL; gradient, 0–2 min (10% B), 2–18 min (10–100% B), 18–22 min (100% B), 22–25 min (10% B). The temperature of the column oven was set at 30°C. MS settings during the LC gradient were listed as follows. Method duration, 25min; total scan time: 0.495; estimated cycle, 3030; source name, TurbolonSpray; curtain gas, 35 psi, ion source gas1 35 psi, ion source gas2 35 psi; temperature 600°C; experiment mode, information-dependent acquisition; polarity, positive; spray voltage, 5500 V. TOF MS settings: TOF start mass, 200 Da; TOF stop mass 2000 Da; accumulation time, 0.2 s; Declustering potential, 80 V; collision energy, 10 V; IDA criteria. IDA criteria: Small molecule; Maximum candidate ions, 5; Intensity threshold exceeds, 100 cps; dynamic background subtraction.TOF MSMS settings: TOF start mass, 50 Da, TOF stop mass 2000 Da, Declustering potential, 80 V, accumulation time, 0.2 s, collision energy 30 V, CE spread 15 V.

The raw data files were converted to mzXML format with ProtoWizard and then preprocessed with mzMine 2.51.[56] The preprocessing included mass detection, ADAP chromatogram builder,[68] chromatogram deconvolution, isotopic grouper, and join aligner. Further details can be found in the For MS^1 feature detection, the noise level was set to 1E4, and for MS^2 feature detection, it was set to 10. The ADAP chromatogram builder was used with a minimum group size of 5 scans, a group intensity threshold of 1E4, a minimum highest intensity of 1E4, and an m/z tolerance of 0.02 Da (or 5 ppm). Local minimum search was applied to the chromatogram deconvolution step with the recommended parameters, including a chromatographic threshold of 85%, a search minimum in the retention time range of 0.05 min, a minimum relative height of 1%, a minimum absolute height of 1E4, a minimum ratio of peak top/edge of 1.7, and a peak duration range of 0–0.5 min. MS^2 scans were paired with an m/z tolerance range of 0.02 Da and an RT tolerance range of 0.5 min. The isotopic peaks grouper algorithm used an m/z tolerance of 0.02 Da (or 5 ppm) and an RT tolerance of 0.3 min. A join aligner step was performed with an m/z tolerance of 0.02 Da (or 5 ppm), a weight of 1 for both m/z and RT and an absolute RT tolerance of 0.5 min. Features with precursor m/z outside 300–1000 are deleted. Background features in the medium samples were removed, and the resulting feature list was exported using the 'Export for/Submit to GNPS' option, keeping only rows with MS^2 data in ".mgf" spectral data format and corresponding ".csv" quantification files. The exported files were analyzed by MSanalyst: a allowed mass tolerance of 0.02 Da or 10 ppm was used in the following MS^2 comparison. In library search, optimal similarity thresholds and matched peaks of "$\geq$3" were used as criteria for algorithms and selecting top-scoring matches as annotations. In networking, optimal thresholds without restriction of matched peaks were used for the algorithms. Further considering the top 5 maximum allowed edges connected to one node and matched peaks of "$\geq$3" were used to curate the modified cosine generated network. The resulting graphs are assembled with Networkx (version 3.1) and exported to Cytoscape (version 3.9.1) for visualization.

The isolation details were as follows: First, the extract was fractionated by MCI gel (Mitsubishi Chemical Co., Ltd.) using a gradient solvent system of decreasing polarity (Fr. 1 10% MeOH/H2O, 6.22 g; Fr. 2 20% MeOH/H2O, 1.18g; Fr. 3 30% MeOH/H2O, 2 g; Fr. 4

Please cite this article in press as: Yu et al., Targeted discovery of aromatic glycosides with dual detoxification effects via a highly customized molecular networking platform, Cell Chemical Biology (2025), https://doi.org/10.1016/j.chembiol.2025.12.001

40% MeOH/H2O, 2.41 g; Fr. 5 50% MeOH/H2O, 1.77 g; Fr. 6 60% MeOH/H2O, 1.3 g; Fr. 7 70% MeOH/H2O, 0.24 g; Fr. 8 80% MeOH/H2O, 1.09 g; Fr. 9 90% MeOH/H2O, 0.41g; Fr. 10 100% MeOH, 1.48 g). Fr. 6 and Fr. 8 were separately subjected to Sephadex LH-20 with 50% DCM/MeOH and collected every 40 min to enrich the targeted molecules. Fr. 6 and Fr. 8 were both divided into 28 subfractions. Subfractions 6–12 to 6–18 were combined, dried, and resuspended in methanol. A total of 240 mg of subfraction was subsequently purified by preparative HPLC using an Agilent Eclipse XDB-C18 (Agilent Tech., Inc., 9.4 * 250 mm, 5 μm), with isocratic elution (3 mL/min flow rate) at 35% MeOH/H2O for 20 min, producing 3 major fractions. Fr. 6-2 was further purified with isocratic elution (2 mL/min flow rate) at 20% ACN/H2O, yielding kutznaposide D (4) (12 mg, t_R = 24.5 min) as a faint yellow, amorphous solid. Similar HPLC conditions, but a different flow rate at 3 mL/min were used to isolate kutznaposide C (3) from Fr. 6-3 (8 mg, t_R = 10.1 min) as a faint yellow, amorphous solid. Kutznaposide F (6) (6.5 mg, t_R = 33 min) was enriched from Fr. 6-1 and further purified by semi-preparative HPLC using the same column and isocratic elution with 26% MeOH/H2O at the flow rate of 2 mL/min. Frs. 8–10 and 8–11 were combined (59.3 mg) and subsequently fractionated by RP HPLC (50% MeOH/H2O, 22.5 min, 3 mL/min) using an Agilent Eclipse XDB-C18 column (9.4 * 250 mm, 5 μm). As a result, fractions containing compound 1 were collected between 15 and 17 min. The same column further purified this fraction and isocratic elution with 20% ACN/H2O at a flow rate of 3 mL/min, yielding kutznaposide A (1) (1.3 mg, t_R = 33.5 min). The MeOH-soluble partition of Frs. 8–12 to 8–17 (136.8 mg) was fractionated by an Agilent Eclipse XDB-C18 column (9.4 * 250 mm, 5 μm). Fractions containing compounds kutznaposide B (2) and E (5) were subjected to HPLC purification (50% MeOH/H2O, 25 min, 3 mL/min). The fractions collected between 13 and 15 min were combined and further fractionated by HPLC (25% MeOH/H2O, 15 min, 2 mL/min), yielding 2 (3.6 mg, t_R = 11.9 min) and 5 (2.5 mg, t_R = 10.9 min).

Structure elucidation of kutznaposides A-F

All the spectra can be found in Figures S20–S26 and Data S4.

DHNC was synthesized (Figure S23) following the procedures included in the previously mentioned U.K. Patent No.9808746.3. 1g Pheynl 1,4-dihydroxy-2-naphthoate was dissolved in 5 mL deoxygenated tetrahydrofuran and 5 mL deoxygenated methanol was added. Stirred at room temperature under a nitrogen atmosphere, 0.5 g ammonium acetate was added to this solution, followed by 10 mL concentrated ammonium hydroxide. After stirring for three hours, the reaction solution was poured into 40 mL of 2 M ice-cold hydrochloric acid (HCl) and further adjusted to pH = 1.0 by concentrated HCl. The end product was collected by filtration, washed with water and dichloromethane, and air-dried. Yield: 125 mg (12.5%). MS and NMR data confirmed that the structure of the end product was right.

Hydrolysis analysis: To determine the stereochemistry of the sugar moieties in kutznaposides, monosaccharide analysis was conducted following the reported literature with minor modifications.[69] Stock solutions of various monosaccharides, including D/L-glucose, D/L-rhamnose, D/L-fucose, and 6-deoxy-D-glucose, were prepared in double-distilled water with a concentration of 2 mM. All tested compounds (100 μg each) were separately added to 1 mL of 4 M trifluoroacetic acid (TFA) and heated in an oven at 100°C for 6 h. After cooling to room temperature, the TFA was removed by nitrogen blowing. Next, L-cysteine methyl ester and phenyl isothiocyanate were added simultaneously, and the reaction mixture was incubated at 60°C for 1 h. The reaction mixture was diluted and centrifuged at 13,000 rpm for 10 min, and the supernatant was then subjected to LC-MS analysis. This entire process was repeated for the derivatization of monosaccharide standards. For LC-MS analysis, the mobile phases consisted of water with 0.05% formic acid (solvent A) and a mixture of methanol/isopropanol/acetonitrile (1:1:2, v/v/v) with 0.05% formic acid (solvent B). The elution (0.3 mL/min flow rate) was optimized as follows: 22% B from 0 to 15 min, increasing from 22% to 100% B from 15 to 18 min, and returning to 22% B from 22.01 to 25 min. A total of 1 μL of the sample or standard solution was injected into the LC-MS system for analysis. The samples were detected using information-dependent acquisition, with the ESI source operated in positive ionization mode. The source and gas temperature were maintained at 550°C and 600°C, respectively.

NMR spectra were collected on a Bruker AVANCE III NMR (both 500 MHz for ^{1}H and 125 MHz for ^{13}C NMR), and the resulting data were processed using MestReNova (v14.2.0) software. Efforts to grow crystals of kutznaposide D (4) from a 3:1 MeOH-H_2O system yielded crystals that confirmed the absolute configuration of 4 and DHNC as aglycon of compounds 3–6.

The elemental composition of kutznaposide A (1) was determined to be $C_{15}H_{20}O_7$ (*m/z* 335.1098 $[M+Na]^+$; $m/z_{calcd.}$ 335.1098). The ^{1}H, ^{13}C NMR and HSQC spectra displayed the three quaternary carbons (including one ester δ_C 168.6), four aromatic methines, five oxygenated methines (including one anomeric carbon δ_C 100.3 ppm, indicating the presence of one monosaccharide moiety), one-methylene, methoxy, and aliphatic methyl carbons. Thus, 6 degrees of unsaturation, deduced from the molecular formula, indicated one trisubstituted benzene, one rhamnose (Rha), and one ester of a carboxyl group at δ_C 168.9. In addition, the ^{1}H–^{1}H COSY and HMBC correlations confirm the structure of 1. To confirm the configurations of rhamnose, the acid hydrolysis of 1 gave L-rhamnose as a sugar component. The deoxy-glucopyranoside was assigned as L-rhamnose, and the *J*-value of the anomeric proton H-1' (*J* = 1.8) indicates its α-configuration.[70] Thus, the structure of 1 was 2-*O*-α-L-rhamnose-1-acetic acid methyl ester.

Kutznaposide B (2) ($[M+Na]^+$ m/z_{obs} 337.0889; m/z_{calcd} 337.0899) in the HRESIMS, identified as $C_{14}H_{18}O_8$, showed good agreement with the literature,[71] except for an additional 14 mass units. Moreover, the ^{1}H NMR and ^{13}C NMR and the HMBC correlation from the methoxy proton (δH 3.96) to a carbonyl group (δ_C 172.0), the *J*-value of the anomeric proton H-1' (*J* = 1.8) and acid hydrolysis indicated that compound 2 was characterized as 2-methoxy-4-*O*-α-L-rhamnose-1-benzoic acid.

Kutznaposide C (3) was obtained as a light yellow amorphous solid. Its molecular formula was determined to be $C_{17}H_{19}NO_7$ from the HRESIMS ion peaks at *m/z* 350.1248 ($[M + H]^+$ m/z_{calcd} 350.1234) and at *m/z* 372.1061 ($[M+Na]^+$ m/z_{calcd} 372.1054). Key fragments in the tandem mass, indicating a loss of an amide group (*m/z* 187.0394 $[M + H\text{-}NH_3]^+$; *m/z* 159.0440 $[M + H\text{-}NH_3\text{-}CO]^+$). The ^{1}H-NMR spectrum displayed the characteristic signal for five aromatic or olefinic protons and one aliphatic methine. The

Please cite this article in press as: Yu et al., Targeted discovery of aromatic glycosides with dual detoxification effects via a highly customized molecular networking platform, Cell Chemical Biology (2025), https://doi.org/10.1016/j.chembiol.2025.12.001

^{13}C-NMR spectrum showed seventeen carbon resonances, including one amide carbonyl (δ_C 171.77), ten aromatic or olefinic carbons (five of them are quaternary), and seven carbon signals, which were attached to oxygen. The planar structure of 3 was further determined by the ^{1}H–^{1}H COSY and HMBC experiments. Analysis of ^{1}H–^{1}H COSY discovered the proton connectivity for a deoxyhexoside moiety by clear correlations of H_3-1' (δ_H 4.87)/H-2' (δ_H 3.65)/H-3' (δ_H 3.4)/H-4' (δ_H 3.10)/H-5' (δ_H 3.10)/H-6' (δ_H 1.16). Since H-3 was presented as a single peak in ^{1}H-NMR, it indicated there are two different spin systems. Moreover, the presence of 1,2,4-trisubstituted naphthalene skeleton, bearing in two hydroxyl groups at C-1/C4 and one amide at C-2, was rapidly identified by the ^{1}H–^{1}H COSY connections from H-5 (δ_H 8.20)/H-6 (δ_H 7.53)/H-7 (δ_H 7.53)/H-8 (δ_H 8.51) and clear HMBC correlations from H-3 to C-1/C-4/C-9/C-10/C-11, H-5 to C-4/C-10, H-7 to C-8/C-9, H-8 to C-5/C-10 and, H-8 to C-1/C-9. The independent chemical synthesis and crystallization of kutznaposide D (4) (Figure 4B) unambiguously confirmed that 1,4-dihydroxynaphthalene-2-carboxamide served as the substrate of kutznaposides C-F. The sugar component was connected with the naphthalene as an oxyglycoside by the detailed interpretation of the well-resolved HMBC correlations from H-1′ to C-1. The absolute configuration of 3's sugar component was established by acid hydrolysis experiments, giving a 6-deoxy-D-glucose. The large *J*-value of anomeric proton H-1' (J = 7.9) indicated the anomer carbon of the deoxyhexoside was of β orientation. Finally, this aforementioned evidence resulted in the 3's absolute structure of 1-*O*-β-deoxy-D-glucose-4-hydroxy-2-naphthamide.

Kutznaposide D (4), with the same molecular formula of $C_{17}H_{19}NO_7$ as kutznaposide C (3) (HRESIMS m/z_{obs} 372.1057 $[M+Na]^+$), had a different retention time and was deduced to be an isomer. Slight differences with six were observed in the ^{1}H and ^{13}C NMR spectra of 3, especially in sugar signals (position 1′ to 5′). Acid hydrolysis deduced the glycosyl part was an L-rhamnose. The *J*-value of anomeric proton H-1' (J = 1.9) indicated the anomer was of α orientation. Detailed analysis of its 2D NMR spectra resulted in constructing a 1,2,4-trisubstituted naphthalene linking a rhamnose at C-1 position. Crystallization was successfully cultivated and helped unequivocally determine the absolute structure. Finally, kutznaposide D (4) was addressed as 1-*O*-α-L-rhamnose-4-hydroxy-2-naphthamide.

Robust interpretation of the 1D and 2D NMR and MS^2 data indicated that kutznaposide E (5), kutznaposide F (6) have the same 1,2,4-trisubstituted naphthalene backbone as 3, but different glycosyl moiety and minor post modifications. Compound 5 was determined to be $C_{18}H_{21}NO_7$ from its HRESIMS (m/z_{obs} 386.1204 $[M+Na]^+$, m/z_{calcd} 386.1212). The absolute stereochemistry of the sugar unit was determined as L-rhamnose after acid hydrolysis. The *J*-value of H-1' (J = 2.0) implied the anomer was of α orientation. Compared to the NMR spectra of 3, additional signals (δ_H 4.03, δ_C 56.3) were observed, accounting for the existence of the methoxy group. In subsequent HMBC analysis, key correlations between H-methoxy and C-4 resulted in methoxy at position C-4. The above spectral data led to the unequivocal determination of 1-*O*-α-L-rhamnose-4-methoxy-2-naphthamide. The molecular formula of compound 6 was deduced as $C_{17}H_{19}NO_8$ from its HRESIMS (m/z_{obs} 388.1011 $[M+Na]^+$, m/z_{calcd} 388.1003). Acid hydrolysis and spectral data provide definite evidence of D-glucose in β orientation and allow the assignment of its absolute structure as 1-*O*-β-D-glucose-4-hydroxy-2-naphthamide.

LC-MS-based enzymatic activity assay of glycosyltransferases

The coding regions of GT5 (1176 bp, GJFFDOON_04527) and GT6 (1116 bp, GJFFDOON_00995) were amplified from the cDNA extracted from *K. viridogrisea* DSM 43850 using Universal Genomic DNA Purification Mini Spin Kit (D0063, Beyotime). Linearized vector (about 5000 bp) was amplified from the pET28a(+) plasmid extracted from an overnight culture of *E.coli* BL21(DE3)/pET-28a(+) using FastPure Plasmid Mini Kit (DC201, Vazyme). The primers are listed in the key resources table, and their melting temperature (Tm) was calculated at https://crm.vazyme.com/cetool/en-us/tmcal.html. The PCR system (50 μL) contained 2 × Phanta Flash Master Mix (25 μL, P520, Vazyme), primers (2 μL, forward and reverse), template DNA (1 μL), and water (20 μL). The PCR reaction was carried out as follows: pre-denaturation at 98°C for 30 s, followed by 30 cycles (98°C for 10 s, annealing according to Tm for 5 s, extension at 72°C at 5 s/kb depending on the target length) and final extension at 72°C for 1 min. The PCR products were analyzed by gel electrophoresis (1%) and purified by FastPure Gel DNA Extraction Mini Kit (DC301, Vazyme). The linearized vector and purified gene product were mixed together at a 1:4 ratio and incubated with 2 × ClonExpress Mix (C115, Vazyme) at 50°C for 30 min. The one-step cloning system was cooled down and transferred into *E. coli* BL21 Chemically Competent Cell (C504, Vazyme). The single colony was picked up for sequencing after overnight incubation at 37°C in the LB solid plate (10 g/L tryptone, 5 g/L-1 yeast extract, 10 g/L NaCl, 2% agar and 30 μg/mL kanamycin).

Once the recombinant plasmid was successfully identified, the recombinant *E. coli* BL21 was cultured in 5 mL LB liquid medium with 30 μg/mL kanamycin at 200 rpm and 37°C. When OD600 reached 0.6–0.8, IPTG was added, reaching the final concentration of 0.1 mM, and the culture was continued to incubate at 180 rpm, 16°C for 16 h. The cells were harvested by centrifugation (4000 rpm, 4°C for 10 min) and resuspended in M9 medium for further whole cell catalysis69.

The whole cells were incubated with a sugar donor (40 μM, UDP-glucose/Rhamnose) and a sugar acceptor (40 μM, DHNC/DHNA) in a total volume of 200 μL at 30°C for 12 h. After incubation, the reaction mixture was quenched with 10 μL 1M HCl and freeze-dried. The dried mixture was dissolved in 100 μL methanol and centrifuged at 13000 rpm for 15 min. The supernatant was then used for LC-MS analysis.

The mixture was analyzed by a triple quad 3500 LC-MS/MS system (Sciex AB). Chromatographic separation was performed on a Waters BEH C18 (2.1 × 50 mm, 1.7 μm) column. The mobile phase consisted of water (containing 0.1% formic acid) in pump A and methanol in pump B. The flow rate was 0.3 mL/min with the gradient method (0–9 min 10–100% B, 9–12 min 100% B, 12.01–15 min 10% B). The source gas settings are listed as follows: Curtain Gas, 10 V; Collision Gas, 7 V; IonSpray Voltage, −4500.0 V, Temperature, 350.0°C; Ion Source Gas1, 35.0 V. The settings of multiple reaction monitoring pairs were as follows: 348/202 Da (Q1/Q3),

Please cite this article in press as: Yu et al., Targeted discovery of aromatic glycosides with dual detoxification effects via a highly customized molecular networking platform, Cell Chemical Biology (2025), https://doi.org/10.1016/j.chembiol.2025.12.001

100 msec dwell time, −104 V declustering potential and −17 collision energy for kutznaposide D; 364/202 Da (Q1/Q3), 100 msec dwell time, −91 V declustering potential and −18 V collision energy for kutznaposide F.

Protein preparation and mass spectrometry analysis

The cell pellets were collected from the 3-day cultures of *K. viridogrisea* in R2A and GSM by centrifugation at 12,000 rpm for 10 min (3 independent biological culture replicates). After freezing the samples with liquid nitrogen for 1 h, they were sequenced on an Illumina Novaseq 6000. The preparation of proteins and the mass spectrometry analysis workflow are the same as those reported in the literature.[72] In brief, the cell pellets were thoroughly washed and then lysed using an automatic frozen tissue grinder. After determining the protein concentration, trypsin and TFA were used for digestion, and the resulting peptides were loaded onto an Evotip for LC-MS/MS analysis. The tryptic-digested proteins were analyzed using reversed-phase C18 column chromatography on the Evosep One LC system (Evisep). MS and MS^2 data were collected in data-independent acquisition (DIA) mode with ZenoTOFTM 7600 mass spectrometer (SCIEX). Proteomic spectral libraries were constructed in data-dependent acquisition (DDA) mode. To capture as many theoretical peptide fragmentations as possible, the Zeno Sequential Window Acquisition (ZenoSWATH) method was used.

Genomic and proteomic analysis

The genome of *K. viridogrisea* DSM 43850 (GCF_046763725.1) was predicted with prokka (version 1.14.6)[57] and its graphical representation was produced on the CGview online server.[58] For proteomic data analysis, peptide identification was performed using the UniProtKB/Swiss-Prot Protein Knowledge Base (https://www.uniprot.org/) with AB SCIEX Protein Search software, followed by Spectronaut (version 12, Biognosys, Schlieren, Switzerland). Peptide quantification was carried out in Spectronaut using DDA data with exponential retention time by corrected retention time standards. Missing values in all three replicates were substituted with 0.01, while missing values in one or two out of the three replicates were substituted with 1/5 of the minimum observed values. T-tests and calculations of the false discovery rate (FDR) using the Benjamini-Hochberg (BH) method were conducted in BioLadder.[59] Both significant genes and proteins were defined as having $|Log_2FoldChange| > 1$ and P-value <0.05. GO and KEGG annotations of *K. viridogrisea* were generated by eggNOG-mapper (version 2.1.12),[60] and function and pathway enrichments were analyzed by TBtools (version 2.092).[61] Protein-protein interactions among differential proteins were recognized by STRING (version 11.5).[62] After uploading the annotated protein sequences of *K. viridogrisea* DSM 43850, differential proteins identified in proteome experiments were subsequently analyzed using the "Proteins by sequences" module with default settings (Network Type: full STRING network, Required score: medium confidence 0.4, FDR stringency: medium 5 percent, selection of the uploaded proteome). Volcano, bar, and bubble plots were performed using GraphPad Prism (version 10.2.0), and all the graphs were assembled and modified in Adobe Illustrator 2024.

Antimicrobial and ABTS decolorization assay

The MIC was determined using the standard microdilution method according to Clinical and Laboratory Standards Institute guidelines.

The inhibitory effect against *K. viridogrisea* was assessed on LB plates. No inhibition of kutznaposides C-F was detected. *K. viridogrisea* was cultured in LB medium at 30°C with shaking at 180 rpm for 2 days to recultivate from storage. The bacterial suspension was diluted 10-fold and spread onto LB agar plates. Two microliters of each compound were applied to the center of the plates with DMSO as the negative control. Inhibition zones of each compound were measured, and photographs were taken after 24 h of incubation at 30°C. In the antimicrobial assay, the strains were cultured in LB medium at 37°C with shaking at 180 rpm. After 16 h of cultivation, each strain was diluted 1:10,000 in medium and added to a 96-well microtiter plate. erythromycin served as a positive control. Positive control and kutznaposides A-F were dissolved in DMSO, giving stock solutions. The concentrations of the positive control and compounds tested were 32, 16, 8, 4, 2, 1, 0.5, and 0.25 μg/mL. All tests were performed in triplicate. Absorbance was measured at 600 nm after 18 h. The inhibition percentage was calculated using the formula: Inhibition (%) = [1-(A2-A1)/A2] × 100%, where A1 and A2 represented the absorbance values of bacterial cultures with and without samples, respectively.

The antioxidant assay was carried out using 2,2′-azino-bis(3-ethylbenzothiazoline-6-sulfonic acid ($ABTS^{\cdot+}$) scavenging assay, based on the previously reported literature with slight modifications.[68] ABTS and K_2S_2O8 were dissolved in water to produce ABTS stock solution, which was stored in the dark at 4°C–8°C for 12–16 h. The ABTS stock solution was then diluted with PBS (pH 7.0). Each tested compound was dissolved in methanol, with vitamin C serving as a positive control. The reaction consisted of 100 μL sample solution and 100 μL of diluted ABTS solution and was carried out at 28°C for 6 min, with absorbance measured at 734 nm. All tests were conducted in triplicate. The equation calculated radical scavenging activity of different solvent fractions: Inhibition (%) = [A2-(A1-A0)/A2] × 100%, where A1 and A0 represent the absorbance values of the sample with and without ABTS solution, respectively, and A2 indicates the absorbance of the ABTS solution with an equal volume of methanol.

QUANTIFICATION AND STATISTICAL ANALYSIS

Customized Python (version 3.10.18) scripts incorporating mainly sklearn (version 1.1.3) were used to calculate the TP, FP, TN, and FN of a confusion matrix as visualized in Figures 2C–2F. The scripts can be found at https://github.com/WenchYu/MSanalyst/tree/master/notebooks.

Please cite this article in press as: Yu et al., Targeted discovery of aromatic glycosides with dual detoxification effects via a highly customized molecular networking platform, Cell Chemical Biology (2025), https://doi.org/10.1016/j.chembiol.2025.12.001

The proteome of K. viridogrisea was analyzed in technical triplicate ($N = 3$) for both GSM and R2A medium cultures. The peptide identification was performed using the UniProtKB/Swiss-Prot Protein Knowledge Base (https://www.uniprot.org/) with AB SCIEX Protein Search software, followed by Spectronaut (version 12, Biognosys, Schlieren, Switzerland). Peptide quantification was carried out in Spectronaut using DDA data with corrected retention time standards. Missing values in all three replicates were substituted with 0.01, while missing values in one or two out of the three replicates were substituted with 1/5 of the minimum observed values. T-tests and false discovery rate (FDR) calculations using the Benjamini-Hochberg (BH) method were conducted in BioLadder.[70] Both significant genes and proteins were defined as having $|Log_2FoldChange| > 1$ and P-value <0.05 in Figures 5A and 5B.

GO function and KEGG pathway enrichments in Figures 6B and 6C were analyzed by TBtools (version 2.092).[72] Volcano, bar, and bubble plots were performed using GraphPad Prism (version 10.2.0), and all the graphs were assembled and modified in Adobe Illustrator 2024.