IFSCC 2025 full paper (IFSCC2025-1302)

"Discovering Innovative Natural Compounds for Cosmetic Applications with Advanced AI-Driven Modeling"

Coralie Ebert [1], Bat-Chen Tamim-Yecheskel [1], Shani Zev [1], Hilla Ben-Hamo Arbel [1], and Marie-Jose EPAULE CHAUVIN [2,*]

[1] MeNow Ltd., Ramat-Gan, Israel; [2] Ornaturm Cosmetologie, Murs-Erigne, France

Abstract

This study presents an AI-based methodology for discovering safe and effective bioactive natural compounds for cosmetic applications. This methodology enables the identification of novel natural compounds with high efficacy and safety profile targeting biological pathways critical to skin health. It combines a structure-based virtual screening with a Bayesian network-based tissue model. To ensure safety, additional predictive models are used to assess carcinogenic, allergenic, and hormonal toxicity, prioritizing candidates with favorable profiles.

This methodology was applied to identify novel heparanase inhibitors, a pivotal mechanism for preserving the extracellular matrix (ECM). Heparanase degrades heparan sulfate proteoglycans—molecules essential for cell signaling, adhesion, and tissue repair. Using this methodology, an analysis of over 500,000 natural compounds was performed and led to the discovery of three novel inhibitor classes: sulfated flavone glucuronides, sulfated phenyl glycosides and glycoside esters, all showing high predicted efficacy and safety profiles.

Notably, four of those molecules were found in plants already used in cosmetics, such as Salvadora persica and Sterculia urens, supporting their translational potential.

These findings demonstrate the power of AI to accelerate the discovery of innovative, multifunctional ingredients for safe and effective cosmetic and biomedical formulations.

Keywords

Heparanase inhibition, AI-driven natural compound discovery, Bayesian tissue modeling, Bioactive cosmetic ingredients, Extracellular matrix preservation

1. Introduction

The development of innovative cosmetic ingredients increasingly relies on the discovery of bioactive natural compounds that are both effective and safe. Traditional approaches to ingredient discovery, often based on empirical screening and serendipitous findings, are time-consuming and limited by the sheer diversity of natural chemical space. In parallel, growing regulatory pressures and consumer demand for transparency and safety have intensified the need for scientifically validated, multifunctional ingredients [1].

Artificial Intelligence (AI) offers transformative opportunities for cosmetic science by enabling systematic exploration of vast molecular libraries, predictive modeling of biological activity,

and early de-risking through in silico safety assessment. In particular, the integration of machine learning models, systems biology simulations, and cheminformatics tools can dramatically accelerate the identification of promising bioactive compounds tailored for specific cosmetic claims such as anti-aging, skin barrier reinforcement, pigmentation modulation, and environmental protection [2].

In this study, we describe an advanced AI-driven platform designed to address these challenges. The platform combines multi-fingerprint QSAR screening, in silico toxicological filtering, chemical innovation assessment through SMARTS and molecular fingerprints, and a Bayesian virtual skin model trained with transcriptomic data. This integrative strategy allows not only the identification of active molecules against critical skin targets but also the estimation of their holistic effect on skin health.

To illustrate the potential of this methodology, we applied it to the identification of novel heparanase inhibitors. Heparanase, an endo-β-D-glucuronidase, plays a central role in the degradation of the extracellular matrix (ECM), contributing to skin aging, impaired barrier function, and inflammatory processes. By inhibiting heparanase, it is possible to preserve ECM integrity and promote skin resilience and rejuvenation — key claims in modern cosmetic formulations [3].

This work highlights the capacity of AI-based strategies to revolutionize natural ingredient discovery, bridging scientific rigor, innovation, and regulatory compliance in the development of next-generation cosmetic actives.

2. Materials and Methods

2.1. Construction of the Natural Compounds Database

A comprehensive database containing over 500,000 natural compounds was compiled from multiple high-quality sources, including plant extracts, marine metabolites, and microbial products. Compounds were standardized through cheminformatics preprocessing: removal of duplicates, normalization of chemical structures, and calculation of key physicochemical descriptors (molecular weight, logP, hydrogen bond donors/acceptors, topological polar surface area, etc.). This ensured consistency and high data integrity for downstream analyses.

2.2. Predictive Modeling for Activity Assessment

Quantitative structure-activity relationship (QSAR) models were developed to predict bioactivity on a range of 1,694 molecular targets relevant to cosmetic applications. Training datasets consisted of known actives and inactives for each target, curated from public databases and literature.

For QSAR modeling, molecular features were generated using four RDKit fingerprints: Morgan, Count, AVALON, and MACCS. These combined features served as inputs for an XGBoost classifier implemented via scikit-learn, using logistic regression with logloss as the evaluation metric. Model training involved k-fold cross-validation for robustness, followed by training on the full dataset. Hyperparameters (number of estimators [100, 200, 300], learning rate [0.01, 0.05, 0.1], max_depth [3, 6, 9]) were optimized using random search cross-validation. To address data imbalance, SMOTE oversampling was applied. Final model performances were assessed on a validation set and via 10-fold cross-validation.

2.3. In Silico Toxicological Filtering

Toxicity predictions were performed using established computational models. Each compound was screened for: Carcinogenicity, Mutagenicity, and Reproductive toxicity (CMR), as well as skin sensitization risk, respiratory sensitization potential and environmental toxicity, including aquatic toxicity and bioaccumulation risk.

2.4. Chemical Clustering and Innovation Scoring

Surviving compounds were clustered using a proprietary approach combining SMARTS-based substructure matching and molecular fingerprint analysis. This method enabled a fine-grained identification of chemical families based on both structural motifs and overall molecular similarity. Within each group, an innovation score was assigned by comparing structural novelty against known cosmetic ingredients and documented bioactives having the target activity.

2.5. Bayesian Virtual Skin Model Simulation

A Bayesian network model representing skin biology was constructed by seeding scientific literature and training with transcriptomic data to mimic biological pathways present in the skin. For each compound, multi-target prediction scores were integrated into the network to estimate the global effect on cosmetic claims of interest (e.g. skin barrier integrity, anti-aging, pigmentation evenness). Compounds predicted to positively modulate key biological pathways were prioritized.

2.6. Sourcing and Regulatory Compliance Analysis

Top candidate compounds were traced back to their natural sources using taxonomy databases and literature reports.

Special emphasis was placed on sourcing from plants listed in the Inventory of Existing Cosmetic Ingredients in China (IECIC). Compounds derivable from IECIC-compliant plants were preferentially considered for lead optimization, ensuring regulatory alignment for international markets.

3. Results and Discussion

3.1. Development of an Innovative AI-Driven Pipeline for Natural Compound Discovery

We developed an integrated AI-driven pipeline designed to accelerate the discovery of innovative natural cosmetic actives. The process starts with a database of over 500,000 natural compounds, followed by predictive modeling across 1,694 biological targets using QSAR models based on combined molecular fingerprints. Compounds predicted as active are filtered through in silico toxicological assessments, ensuring safety profiles suitable for cosmetic use Figure 1, top panel).

Surviving candidates are clustered using a proprietary SMARTS- and fingerprint-based method to identify novel chemical families. A Bayesian virtual skin model, trained on literature and transcriptomic data, integrates predicted bioactivities to prioritize compounds likely to enhance cosmetic claims. Finally, sourcing analyses ensure alignment with regulatory standards, preferentially for the Chinese cosmetic market (Figure 1, lower panel).

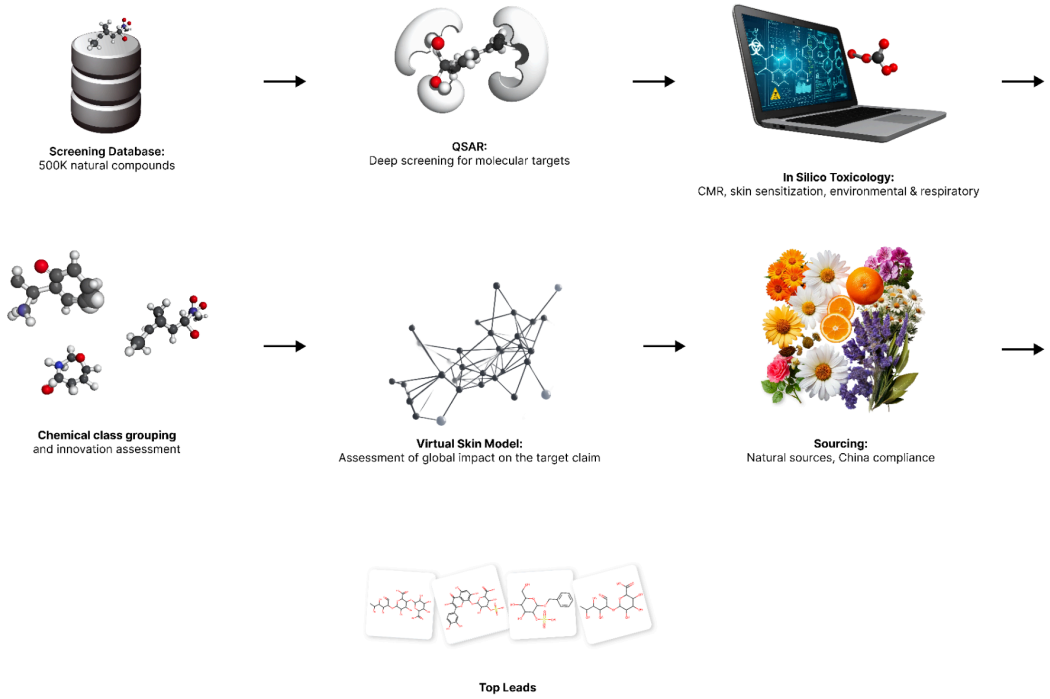

Figure 1 - Development of an Innovative AI-Driven Pipeline for Natural Compound Discovery. An illustration of the suggested pipeline workflow including the following steps: data preparation of over 500K natural compounds, QSAR screening for molecular targets, toxicology assessment, chemical class grouping and innovation assessment, global impact on target claims by virtual skin model, and sourcing identification with preference for china compliance.

By combining high-throughput predictive modeling, rigorous safety filtering, and holistic efficacy evaluation, this pipeline enables the rapid identification of new molecular targets and active ingredients for cosmetics. This approach has the potential to significantly reduce development timelines, expand the landscape of cosmetic innovation, and open avenues for creating next-generation products that are both highly effective and regulatory compliant.

As a practical application of this pipeline, we focused on the identification of new heparanase inhibitors, targeting extracellular matrix (ECM) preservation. Heparanase plays a critical role in ECM degradation and skin aging processes. In the next subsection, we present how our AI-driven strategy enabled the rapid discovery of promising natural inhibitors with potential cosmetic benefits related to ECM integrity and skin health.

3.2. Identification of Novel Heparanase Inhibitors for ECM Preservation

By applying our AI-driven discovery pipeline, we identified three distinct chemical groups of natural compounds exhibiting high predicted activity for heparanase inhibition, excellent safety profiles, and strong potential for extracellular matrix (ECM) preservation: Sulfated flavone glucuronides; Sulfated phenyl glycosides; Glycoside esters.

These chemical groups, characterized and illustrated in Figure 2, were defined based on core chemical structures and potential radical variations. Notably, these molecules have not

previously been reported as heparanase inhibitors, underscoring the innovative value of our approach. Interestingly, one sulfated flavone glucuronide derivative was previously mentioned in the literature for its hyaluronidase inhibition, suggesting a broader relevance for ECM stabilization [4].

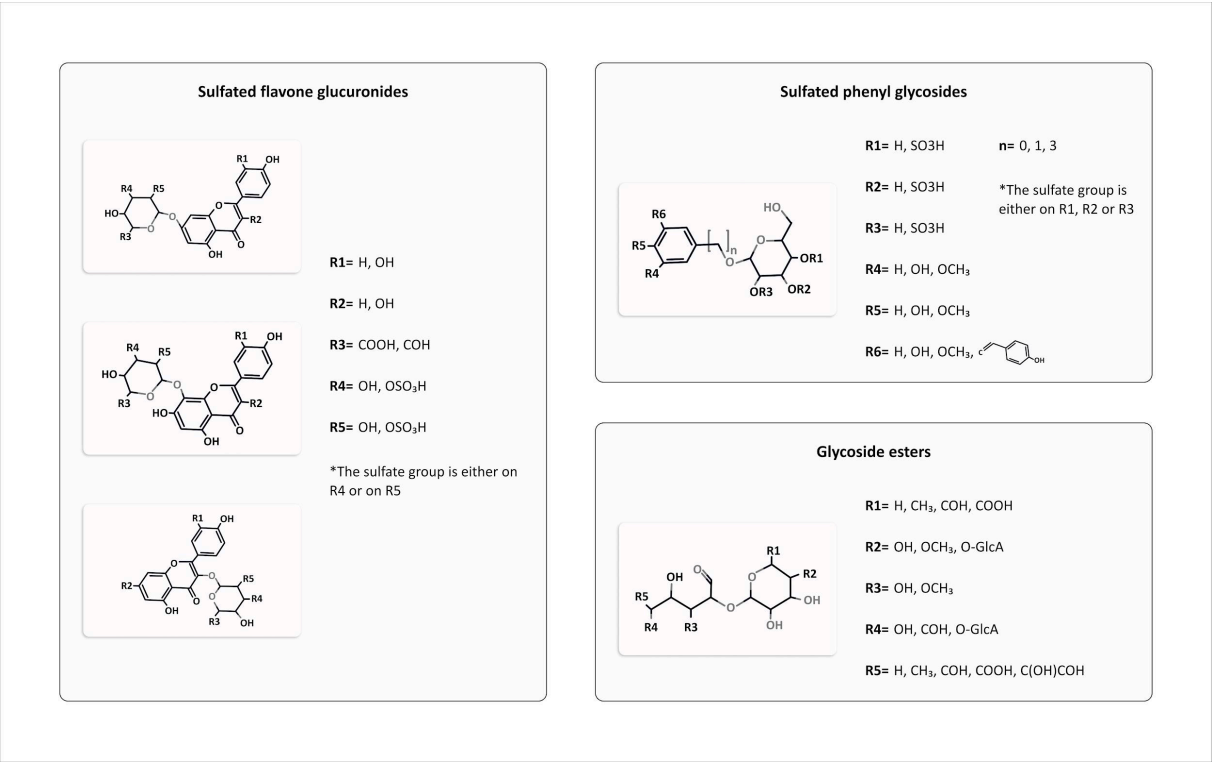

Figure 2 - Identification of three novel chemical groups with heparanase inhibition activity. An illustration of the three newly identified heparanase inhibitors chemical groups: Sulfated flavone glucuronides; Sulfated phenyl glycosides; Glycoside esters. The core chemical structures and potential radical variations are presented.

Sixteen existing natural compounds belonging to these three chemical groups were identified. Their safety and efficacy profiles were compared to those of known heparanase inhibitors, including Trachyspic acid, RK-682, 2-(1,3-Benzoxazol-5-yl)acetic acid, Trimellitimide, Suramin, Pixatimod (PG545), and Phosphomannopentaose sulfate (PI-88). The natural candidates demonstrated either comparable or superior profiles in predicted activity and safety (Figure 3), further highlighting their potential for ECM preservation and cosmetic applications.

This discovery demonstrates the capability of the described platform to unveil new bioactive chemical spaces for cosmetic ingredient innovation, providing promising leads for the development of novel skin health-promoting actives.

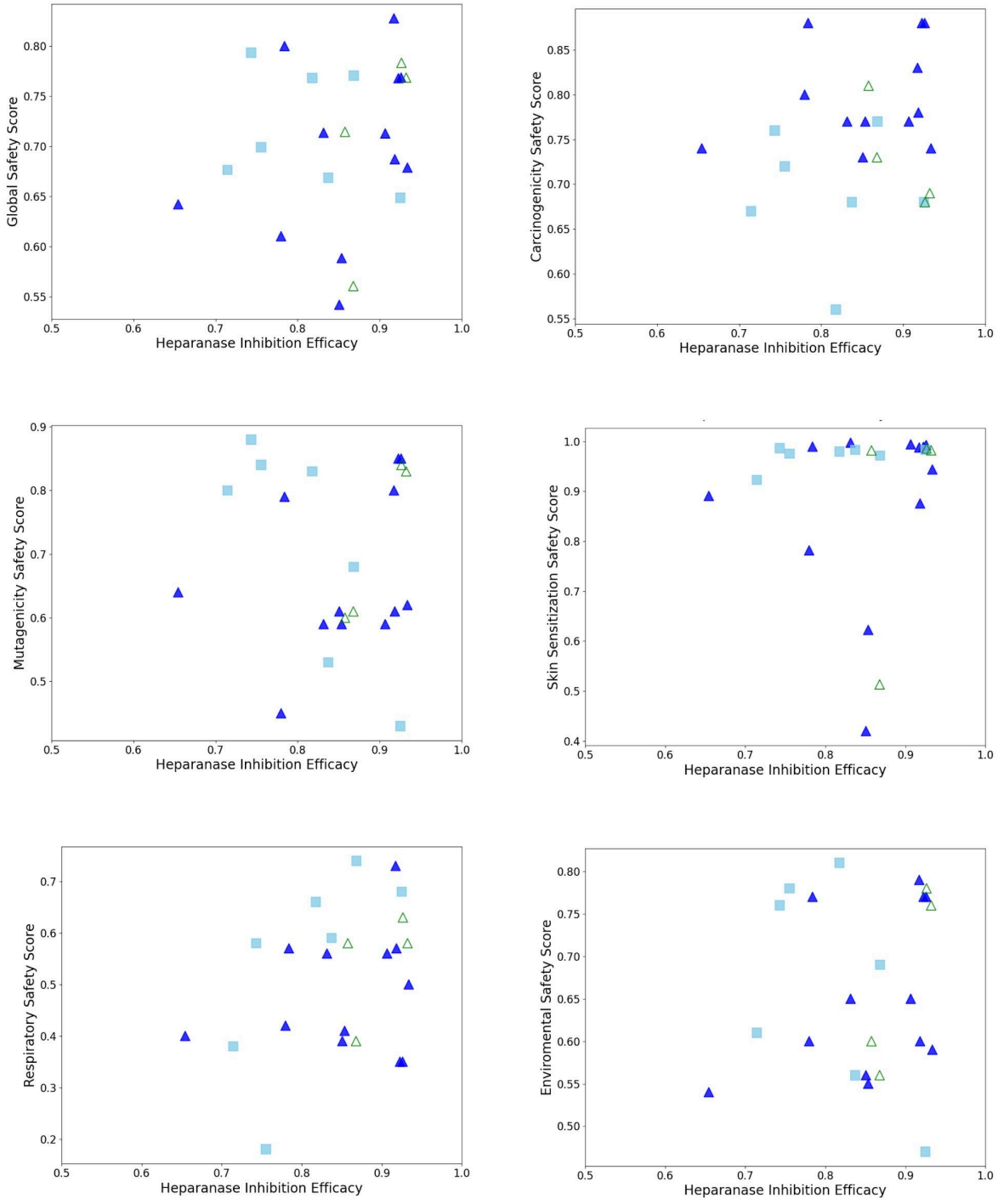

Figure 3 - The newly identified natural compounds demonstrated comparable or superior activity and safety profiles prediction. Sixteen newly identified heparanase inhibitors from three chemical groups (triangles, Solid blue and Green outlined) and seven known inhibitors (squares) were tested for their safety profile utilizing in silico computational techniques. A representation of safety score vs. heparanase inhibition predicted efficacy is presented as follows: Global safety (top left); Carcinogenicity safety (top right); mutagenicity safety (middle left); skin sensitization safety (middle right); respiratory safety (bottom left); and environmental safety (bottom right). Green outlined triangles represent the four selected lead compounds (see section three in results and discussion).

3.3. Selection of Lead Compounds with China-Compliant preferences

Among the sixteen promising natural compounds, four were identified as leading candidates combining high predicted efficacy and excellent safety profiles, together with a reliable sourcing. Three compounds with sourcing from China-compliant plants: one sulfated flavone glucuronide sourced from the flowers of Malva sylvestris; and two glycoside esters sourced from the fruits of Sterculia urens. Another interesting candidate is a sulfated phenyl glycoside sourced from the seeds of Salvadora persica, a plate with an INCI (International Nomenclature of Cosmetic Ingredients) registration.

These molecules, along with their chemical structures and sourcing details, are presented in Figure 4. The identification of these leads highlights the ability of the platform not only to discover innovative actives but also to align discoveries with key regulatory requirements, facilitating global market applications.

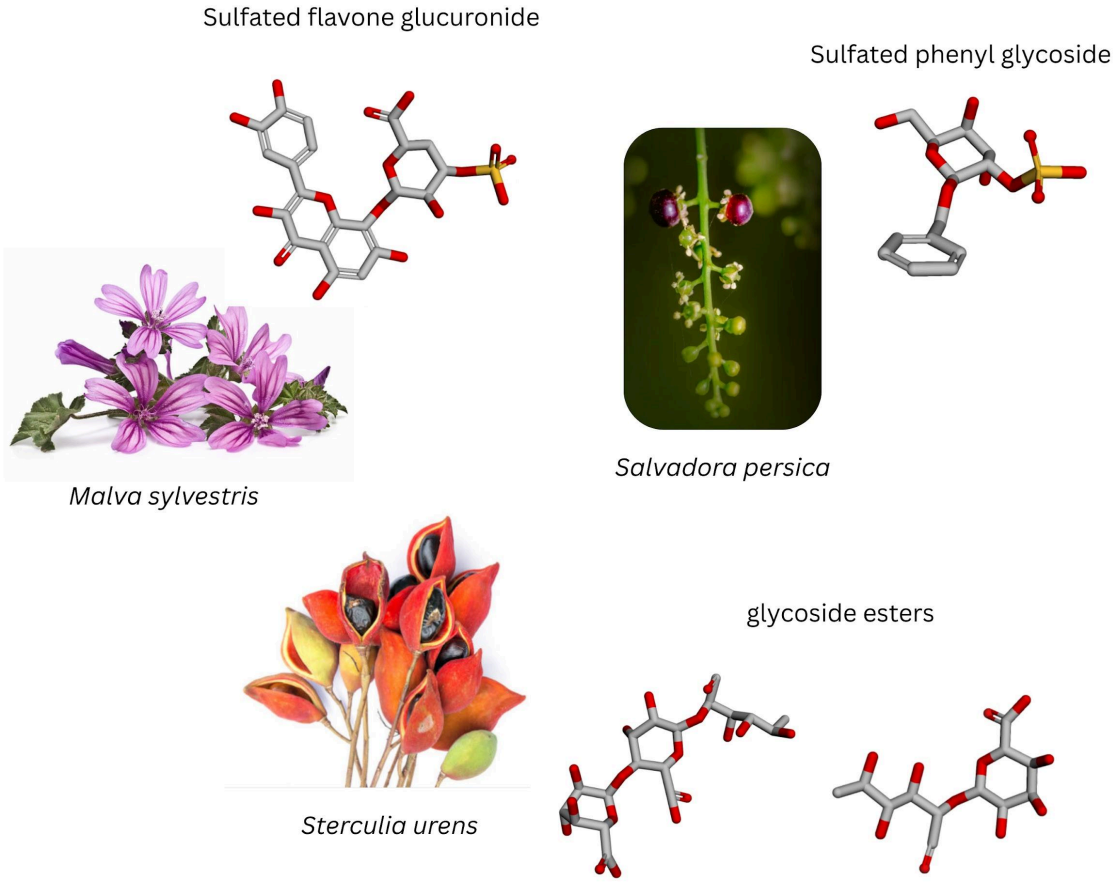

Figure 4 - Selected lead compounds with china-compliant and INCI registration. Illustration of the four leading compounds, combining high predicted efficacy and excellent safety profiles, together with a reliable sourcing. A sulfated flavone glucuronide sourced from the flowers of Malva sylvestris (top right); a sulfated phenyl glycoside sourced from the seeds of Salvadora persica (top left); and two glycoside esters sourced from the fruits of Sterculia urens (bottom).

4. Conclusion

The AI-driven discovery platform described in this work represents a significant advancement for cosmetic ingredient innovation. By integrating predictive modeling, toxicological assessment, chemical clustering, systems biology simulations, and regulatory compliance screening, the platform enables rapid identification of novel, safe, and effective natural actives.

Its application to the search for heparanase inhibitors demonstrated the capacity to uncover new bioactive chemical families with high potential for ECM preservation and skin health benefits. The identification of China-compliant and INCI registered sources leads further validates the platform's readiness for real-world cosmetic development.

This approach not only accelerates the ingredient discovery process but also expands the frontier of innovation, offering the cosmetic industry a powerful tool to create next-generation products that meet both efficacy and safety expectations in a competitive global market.

Acknowledgements

This work was supported by funding from the Israel Innovation Authority (IIA). The authors gratefully acknowledge the IIA's support, which made the development of this research possible.

References

1. Atanasov, Atanas G., et al. "Discovery and resupply of pharmacologically active plant-derived natural products: A review." *Biotechnology advances* 33.8 (2015): 1582-1614.

2. Moshawih, Said, et al. "Synergy between machine learning and natural products cheminformatics: Application to the lead discovery of anthraquinone derivatives." *Chemical Biology & Drug Design* 100.2 (2022): 185-217.
3. Masola, Valentina, et al. "Heparanase: A multitasking protein involved in extracellular matrix (ECM) remodeling and intracellular events." *Cells* 7.12 (2018): 236.
4. Sendker, Jandirk, et al. "Phytochemical characterization of low molecular weight constituents from marshmallow roots (Althaea officinalis) and inhibiting effects of the aqueous extract on human hyaluronidase-1." *Journal of Natural Products* 80.2 (2017): 290-297.