# Machine learning uncovers natural product modulators of the 5-lipoxygenase pathway and facilitates elucidation of their biological mechanisms

Sigitas Mikutis[#*1], Stefanie Lawrinowitz[#2], Christian Kretzer[2], Lavinia Dunsmore[1], Laurynas Sketeris[1], Tiago Rodrigues[3], Oliver Werz[*2] & Gonçalo J. L. Bernardes[*1,4]

[1]Yusuf Hamied Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, UK

[2]Department of Pharmaceutical/Medicinal Chemistry, Institute of Pharmacy, Friedrich Schiller University Jena, Philosophenweg 14, 07743 Jena, Germany

[3]Instituto de Investigação do Medicamento (iMed), Faculdade de Farmácia, Universidade de Lisboa, Av. Prof. Gama Pinto, 1649-003 Lisbon, Portugal

[4]Instituto de Medicina Molecular João Lobo Antunes, Faculdade de Medicina, Universidade de Lisboa, Avenida Professor Egas Moniz, 1649-028, Lisboa, Portugal

[#]These authors contributed equally to this work.

Correspondence should be addressed to G.J.L.B. or S.M. or O.W.:

E-mail: gb453@cam.ac.uk
E-mail: siggitas@gmail.com
E-mail: oliver.werz@uni-jena.de

**Abstract**

Machine learning (ML) models have made inroads into chemical sciences, with optimization of chemical reactions and prediction of biologically active molecules being prime examples thereof. These models excel where physical experiments are expensive or time consuming, for example, due to large scales or the need for materials that are difficult to obtain. Studies of natural products suffer from these issues – this class of small molecules is known for its wealth of structural diversity and wide-ranging biological activities, but their investigation is hindered by poor synthetic accessibility and lack of scalability. To facilitate the evaluation of these molecules, we designed ML models that predict which natural products can interact with a particular target or a relevant pathway. Here, we focused on discovering natural products which are capable of modulating the 5-lipoxygenase (5-LO) pathway that plays key roles in lipid signalling and inflammation. These computational approaches led to the identification of 9 natural products which either directly inhibit the activity of the 5-LO enzyme or affect the cellular 5-LO pathway. Further investigation of one of these molecules, deltonin, led us to discover a new, cell type-selective mechanism of action. Our ML approach helped deorphanizing natural products as well as shedding light on their mechanisms, and can be broadly applied to other use cases in chemical biology.

## Introduction

The last decade has seen a rise in use of machine learning (ML) models for molecular applications. Some of their applications include prediction of reaction yields[1], optimization of molecular dynamics calculations[2] and inference of small molecule bioactivity[3]. The latter is a particularly challenging task as it needs to predict complex interactions between large biomolecules and their modulators. To highlight some examples, Zhavoronkov and colleagues used generative ML models that helped to design *de novo* inhibitors of DDR1 kinase[4], Stokes and colleagues reported a deep learning-based platform to screen for antibiotics in large datasets of molecules[5] whereas Svensson and colleagues utilised random forests and conformal predictions to assess cytotoxicity of chemical matter against 16 cell lines[6]. These examples highlight the potential of ML to streamline chemical biology and minimize experimentation.

One field of research where ML models could make a difference is the deorphanization of natural products – molecules that often exhibit complex biological activities and possess a wide variety of chemical motifs. Indeed, a number of studies report ML models able to predict properties of natural products including anti-microbial, anti-cancer and anti-inflammatory activities as well as abilities to modulate protein expression[7,8]. Natural products are well known for their polypharmacology[9], i.e., they can often interact with multiple biomolecules, which leads to complex bioactivities. The chemical space occupied by natural products is different to that of molecules produced in medicinal chemistry campaigns and functionalizing clinical candidates with natural products motifs can lead to more potent biomodulators[10]. Thus, the study of natural products can lead to discovery of new therapeutically-relevant chemical matter, uncover new mechanisms of action and, in many cases, directly lead to new pharmaceuticals. Indeed, 396 natural products or their derivatives (mostly semi-synthetically modified natural products) were approved by the FDA in 1981-2014, so they remain an important class of new therapeutics[11]. However, the chemical space covered by natural products is wide and for many it is difficult or even impossible to obtain amounts sufficient for comprehensive testing. As such, *in silico* approaches to predict properties of natural products are highly desirable and can help to streamline the use of scarce resources.

For this aim, we went on to develop ML models to discover natural product modulators of a particular biochemical process, and as a result could be suitable as chemical tool. Our approach utilizes publicly available small molecule-biomolecule interaction databases ChEMBL[12], data from which is used as the basis for predictive models, and PubChem[13], used to further refine the list of predicted modulators. We used the ML models to predict modulators of the 5-lipoxygenase (5-LO) pathway with roles in the inflammatory response[14,15,16]. 5-LO uses a non-heme iron to oxidise essential fatty acids, in particular (AA), with leukotrienes (LT) being one of the key products, thus 5-LO plays a central role in lipid signalling. A number of other proteins contribute to cellular 5-LO product formation forming a complex network of interactions. These proteins include the cytosolic phospholipase (cPL)A2 that release AA, 5-LO-activating protein (FLAP) that facilitates access of 5-LO to AA as well as LTA4 hydrolase and LTC4 synthase that generate the downstream LTB4 and LTC4, respectively.

Moreover, the activity of the 5-LO pathway is prone to complex regulation by $Ca^{2+}$, peroxides, phospholipids and phosphorylation events. 5-LO-derived products are of relevance for various inflammatory and allergic disorders, with one small molecule 5-LO inhibitor – zileuton – being approved against asthma[17]. 5-LO also plays an important role in acute myeloid leukemia, as recently disclosed by us[18,19]. Thus, it is conceivable that some natural products with anti-inflammatory properties could be modulating the activity of this pathway. We were especially interested in discovering natural products that would constitute novel 5-LO modulator chemical space. To accomplish this, we elected using physiochemical descriptors to build the ML models rather than molecular fingerprints or other structure-based methods, as they would make predictions based solely on the structural features.

The ML models we build led us to test 21 natural products against 5-LO. Due to the nature of the training data, we were able to discover new direct enzyme and cellular pathway inhibitors of 5-LO, highlighting the polypharmacology expressed by these molecules. A number of these natural products indeed modulated 5-LO product formation, which highlights the suitability of our approach to discover natural products modulating a particular pathway. We carried out a more in-depth characterization of deltonin and discovered that it modulates the 5-LO pathway in a cell type-dependent manner, with a mechanism that involves elevation of the intracellular $Ca^{2+}$ level and suppression of ROS modulation in human polymorphonuclear leukocytes. We believe that further study of this mechanism can provide new therapeutic insights, especially regarding its cell type selectivity.

## Methods

**Data preparation.** 5128 entries describing small molecule modulation of 5-LO and 4781 describing modulation of FLAP were downloaded from CHEMBL23 database an uploaded into KNIME analytics platform (v. 3.4.0)[20]. Only entries that report exact $IC_{50}$, $K_D$ or $K_i$ relationships at nanomolar concentrations were kept and duplicates were removed, leading to 2456 and 2245 entries for 5-LO and FLAP datasets. Molecules with reported activities below 1,000 nM were classified as strong modulators (label: 1) with the rest labelled as weak (label: 0). RDKit descriptors (117) were calculated for each molecule and were then normalised.

**ML models.** No code ML models were built with native nodes in the KNIME 3.4.0 environment, using the normalized descriptors of 2456 5-LO and 2245 FLAP modulators. A 10-fold cross-validation was performed to optimize hyperparameters and gauge the performance of three classifiers: gradient-boosted trees (GBT)[21], logistic regression (LR)[22] and naïve Bayesian (NB)[23] models, according to the True Negative Rate (TNR), Positive Predictive Value (PPV), the Matthew's Correlation Coefficient and Balanced Accuracy (BA)[24], defined as:

$$TNR = \frac{TN}{TN + FP} \quad (1)$$

$$PPV = \frac{TP}{TP + FP} \quad (2)$$

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \quad (3)$$

$$BA = \frac{1}{2}\left(\frac{TP}{TP + FN} + \frac{TN}{TN + FP}\right) \quad (4)$$

Where TP – number of true positives, TN – number of true negatives, FP – number of false positives and FN – number of false negatives. We use 1234 as random state. The relevant hyperparameters for the 5-LO models are as follows: GBT – tree depth = 4, number of trees = 500, learning rate = 0.05; LR – maximal number of epochs = 100, epsilon = 1.0E-05, step size = 0.1; NB – default probability = 0.0, maximum number of unique nominal values per attribute = 20. For FLAP models, the parameters are the same except GBT – learning rate = 0.15, NB – default probability = 0.005.

**Predicting 5-LO natural product inhibitors.** A dataset of 24,500 commercially available natural products (an overlap between DNP and Zinc commercially available databases) was uploaded into KNIME analytics platform. Molecules were pre-processed and the descriptors calculated and normalised as in the training set. The implemented GBT, LR and NB models were used to predict labels for the test set, leading to 1,245 molecules predicted as positives by all three models and further 2,254 predicted by any two models (Full metrics in Table S1) for the 5-LO dataset. For FLAP datased, only the 940 triple-positive predictions were considered. Triple-positive and part of random (500 out of 2,254 double-positive 5-LO set molecules) were further investigated in PubChem database for their anti-inflammatory, anti-leukemia and anti-arthritic activities. Eleven and ten randomly-chosen natural products from 5-LO and FLAP datasets, respectively, with known anti-inflammatory, anti-leukemic or anti-arthritic properties were selected for further testing. Molecules described as 5-LO modulators in ChEMBL database were not chosen for testing.

**Fragment analysis.** 2,456 5-LO modulators were split into strong and weak modulator datasets as described above. Molecules in both sets were broken down into fragments with path lengths 4-8. Fragments containing interrupted aromatic systems were discarded. Fragment frequencies (number of molecules containing a particular fragment in a set versus a total number of molecules in a set) in both sets were calculated. Ratio Frequency(active)/Frequency(inactive) was used to calculate relative representation of each fragment in the active and inactive datasets.

**Similarity search.** The natural products chosen for testing and all the 5-LO modulators in ChEMBL23 database were converted into RDKit (2048 bits, Min Path Length = 1, Max Path Length = 7) and AtomPair[25] (2048 bits, Min Path Length = 1, Max Path Length = 30) molecular fingerprints. Tanimoto similarity search was carried out between the fingerprints of chosen natural products and known 5-LO modulators in order to find the most similar molecules between the two sets.

**p$K_a$ predictions.** p$K_a$ values of resistomycin were predicted using MarvinView (ChemAxon/Infocom Marvin Extensions Feature 4.4.0.v211500) in KNIME Analytics Platform.

**Expression and purification of human recombinant 5-LO.** Human recombinant 5-LO was expressed in *E. coli* (BL21, DE3) transformed with pT3-5-LO plasmid at 30 °C overnight, as described before[26]. Cells were lysed using lysis buffer and homogenized by sonification (3 x 15 s, Branson Sonifier 250, Branson Ultrasonics Corporation). 5-LO was then purified from 40,000 x g supernatant (20 min, 4°C) using an ATP agarose column (Sigma-Aldrich), diluted with PBS (Dulbeco's formula, pH 7.4) buffer containing 1 mM EDTA, and immediately used for 5-LO activity assays.

**Blood cell isolation.** The process of isolation was performed as described by A. Boyum[27]. Briefly, peripheral blood was withdrawn from fasted healthy adult volunteers (University Hospital Jena, Germany) and centrifuged to obtain leukocyte concentrates. These were aliquoted and mixed with 2.5% dextran in PBS. After 45 min, leukocyte-rich supernatant was transferred to density centrifugation medium (Histopaque-1077; d = 1.077) and centrifuged (200 rpm, 10 min, room temperature, without brake, Heraeus Multifuge X3R Centrifuge, Thermo Fisher Scientific).

Peripheral blood mononuclear cells (PBMC) were concentrated on top of the density medium and separated from the other cells. Isolated PBMC were washed with PBS twice (1200 rpm, 5 min, 4 °C) and resuspended in 5 ml PBS.

For isolation of PMNL, contaminating erythrocytes of pelleted PMNL were removed by hypotonic lysis. Afterwards, PMNL were washed with PBS twice (1200 rpm, 5 min, 4 °C), resuspended in 5 ml PBS. A cell counting system (Vi-CellTMXR, Beckmann Coulter) was used to determine cell numbers and cell viability. For counting, the cell suspension was diluted (1:50) and trypan blue staining (0.4 % (v/v), sterile filtered) was used to determine cell viability.

**Determination of 5-LO product formation using isolated enzyme.** Isolated human 5-LO was diluted to optimal concentration (approx. 0.5 µg/ml) in 1 ml PBS containing 1 mM EDTA and treated with compounds or vehicle (0.1% (v/v) DMSO) for 10 min at 4 °C, and stimulated by addition of 2 mM $CaCl_2$ and 20 µM AA for 10 min at 37 °C. The reaction was stopped by adding 1 ml ice-cold methanol and samples were prepared for high performance liquid chromatography (HPLC) analysis.

**Determination of 5-LO product formation in PMNL.** Freshly isolated PMNL (5 x $10^6$) were suspended in 1 ml PBS-glucose (Dulbeco's PBS with 1% w/v glucose) buffer and 1 mM $CaCl_2$ was added. Cells were treated with compounds or vehicle (0.1% (v/v) DMSO) for 10 min at 37 °C, and stimulated by addition of 2.5 µM A23187 (and 20 µM exogenous AA as stated in the text) for 10 min at 37 °C. The reaction was stopped by adding 1 ml ice-cold methanol and samples were prepared for HPLC analysis.

**Solid phase extraction (SPE) and HPLC analysis.** After stopping the 5-LO reactions (see above), 530 µl PBS-HCl (500 µl Dulbeco's PBS mixed with 30 µl of 1N HCl aqueous solution) and 200 ng internal standard PGB1 were added to the samples. The samples from PMNL incubations were centrifuged (870 x g, 10 min, 4 °C) before SPE was performed using reverse phase C18 SPE cartridges Clean-Up (United Chemical Technologies). Briefly, columns were conditioned with methanol and $ddH_2O$, samples were added, washed with $ddH_2O$ and 25% (v/v) methanol, eluted with 300 µl

methanol, mixed with 120 µl $ddH_2O$; 100 µl were then subjected to analysis. Major 5-LO products ($LTB_4$, all-trans isomers of $LTB_4$, and 5-H(p)ETE) were analyzed by RP-HPLC using a Nova-Pak C18 Radial-Pak Column (4 µm, 5 × 100 mm, Waters) under isocratic conditions (73% methanol/27% water/0.007% trifluoroacetic acid) at a flow rate of 1.2 mL/min and detected at 235 nm (for 5-H(p)ETE) or 280 nm (for $LTB_4$ and its all-trans isomers).

**Lactate dehydrogenase (LDH) release assay.** The LDH assay was performed using the CytoTox 96® Non-Radioactive Cytotoxicity Assay kit. Briefly, 2 x $10^5$ PMNL or PBMC diluted in PBS-glucose buffer were seeded per well of a 96-well plate. Lysis control and 0.2% (v/v) triton X-100 were added to the cells and incubated for 45 min, compounds and vehicle (0.1% (v/v) DMSO) were added and incubated for 20 min at 37 °C. Stop solution was added, the plate was centrifuged (250 x g, 4 min, room temperature) and 50 µl of supernatant from each well was transferred. Afterwards, 50 µl of substrate mixture was added and incubated for 30 min at room temperature under exclusion of light. To finally stop the reaction, 50 µl of stop solution were added, the photometric measurement was done at 490 nm using a Multiscan Spectrum plate reader (Thermo Fisher Scientific). Cytotoxicity was calculated after background correction as:

$$lysis = \frac{A_{490}\ (compound)}{A_{490}\ (vehicle)} \times 100\%\ (5)$$

**Hemolysis assay.** Defibrinated OxoidTM sheep's blood (Thermo Fisher Scientific, Waltham, MA, USA) was diluted to a 5% (v/v) suspension in PBS. In a 96-well microtiter plate, 190 µl of blood suspension was added to 10 µl of compound in PBS containing 20 x stock. DMSO was diluted in PBS to a 20 x stock which after adding the blood gave a concentration of 0.1% (v/v, negative control). Addition of 1% (v/v) triton X-100 was used as a positive hemolysis control. Three replicates were performed for each compound concentration. The plate was then incubated at 37 °C for 2 or 24 h. Following the incubation, the plate was centrifuged (3,300 rpm, 5 min), 100 µl of supernatant were collected and transferred. To determine hemolysis, ultraviolet (UV) absorbance of free heme was measured immediately at 540 nm and 500 – 700 nm (spectrum) using a plate reader. Percentage of hemolysis was determined by:

$$hemolysis = \frac{A_{540}\ (compound) - A_{540}\ (PBS)}{A_{540}\ (1\%\ Triton\ X\text{-}100) - A_{540}\ (PBS)} \times 100\%\ (6)$$

**MTT assay.** Freshly isolated PBMC were diluted in RPMI supplemented with 10% fetal calf serum, 1% penicillin/streptomycin and 1 mM L-glutamine and 100 µl of the resulting cell suspension containing 2 x $10^5$ cells were seeded per well in a 96-well plate. A 333-fold stock of compound or 0.3% (v/v) DMSO was added and incubated for 1 h or 24 h at 37 °C. As a positive control 16.7% (v/v) ethanol or 0.05% (v/v) triton-X were used, a negative control contained media only. Finally, 20 µl of MTT solution (5 mg/ml in PBS, sterile-filtered) was added per well and incubated for 3 h at 37 °C. Cells were lysed by addition of 100 µl SDS lysis buffer shaking at 175 rpm (neolab

Multi shaker DOS-102, neoLab Migge) under exclusion of light overnight. The photometric measurement was performed at 400 – 600 nm (595 nm) using a Multiscan Spectrum plate reader (Thermo Fisher Scientific). Cell viability was calculated after background correction as:

$$cytotoxicity = \frac{A_{595}\ (compound)}{A_{595}\ (vehicle)} \times 100\%\ (7)$$

**Cultivation and preparation of human pathogenic *E. coli*.** Human pathogenic *E. coli* (O6:K2:H1 CFT073) was cultivated in 35 ml NB medium suspension in an Erlenmeyer flask shaking at 37 °C overnight. On the next day, 5 ml cell suspension were centrifuged (5000 x g, 5 min, 20 °C) and optical density OD600 was brought to 1.0 by diluting cells in PBS (pH 7.4) supplemented with 1 mM $CaCl_2$. $OD_{600}$ was determined using a Ultrospec10-Cell density meter (Amersham Biosciences).

**Differentiation and polarization of monocytes to M1 macrophages.** Freshly isolated PBMC were cultivated in 15 ml PBS (Dulbeco's formula, supplemented with 100 mg/L $CaCl_2$ and $MgCl_2$ hexahydrate) in a 75 $cm^3$ cell culture flask at 37 °C for 1 h in the WTC Binder incubator (WTC Binder GmbH) in a humid and $CO_2$ enriched (5% v/v) atmosphere. After attachment of monocytes, cells in suspension were removed and 15 ml RPMI supplemented with 10% fetal calf serum, 1% penicillin/streptomycin and 1 mM L-glutamine medium were added per flask. For macrophage differentiation, 20 ng/ml GM-CSF was supplemented and after 6 days, cells were polarized with 20 ng/ml interferon-γ and 100 ng/ml LPS for 48 h at 37 °C and 5% (v/v) $CO_2$.

**Stimulation of M1 macrophages and lipid mediator profiling by UPLC-MS/MS.** M1 macrophages (1 x $10^6$) were resuspended in 1 ml PBS (Dulbeco's formula, supplemented with 1 mM $CaCl_2$) and treated with test compounds or vehicle (0.1% (v/v) DMSO) for 20 min at 37 °C and 5% (v/v) $CO_2$. Afterwards, macrophages were stimulated by addition of *E. coli* (O6:K2:H1) at a ratio of 1:50 (M1:*E. coli* multiplicity of infection (MOI) of 50) for 90 min at 37 °C and 5% (v/v) $CO_2$. The reaction was stopped by transferring the supernatant to 2 ml ice-cold methanol and samples were prepared for ultraperformance liquid chromatography–tandem mass spectrometry (UPLC-MS/MS) analysis.

For stimulation with $Ca^{2+}$-ionophore A23187, medium was removed, and macrophages (1 x $10^6$) were treated by addition of compounds or vehicle (0.1% (v/v) DMSO) diluted in 1 ml PBS (Dulbeco's formula, supplemented with 1 mM $CaCl_2$) for 10 min at 37 °C and 5% (v/v) $CO_2$. Stimulation was performed by addition of 2.5 µM A23187 for 10 min at 37 °C and 5% (v/v) $CO_2$. The reaction was stopped by transferring the supernatant to 2 ml ice-cold methanol and samples were prepared for UPLC-MS/MS analysis.

To the samples, deuterated lipid mediators (d8-5S-HETE, d4-$LTB_4$, d5-$LXA_4$, d5-RvD2, d4-$PGE_2$; 200 nM each) and d8-AA (10 µM) were added as internal standards. Lipid mediators were extracted by solid phase extraction using Sep-Pak C18 6cc Vac Cartridges (500 mg; Waters), as previously described[28]. Briefly, samples were stored at -20 °C for at least 45 min to allow protein precipitation. After centrifugation (1200×g,

4 °C, 10 min), the supernatant was combined with acidified water (pH 3.5, 7 mL) and loaded onto pre-equilibrated solid phase cartridge columns. Washing steps with water and n-hexane (6 mL, each) followed. Lipid mediators were eluted with methyl formate (6 mL), brought to dryness using an evaporation system (TurboVap LV, Biotage, Uppsala, Sweden) and resuspended in methanol/water (50/50, v/v, 100 µL) for UPLC-MS/MS analysis.

For metabololipidomics analysis, lipid mediators were separated at 50°C on an Acquity UPLC BEH C18 column (130Å, 1.7 µm, 2.1 mm x 100 mm, Waters). The Acquity Ultraperformance LC system (Waters) was operated at a flow rate of 0.3 mL/min using a mobile phase consisting of methanol, water, and acetic acid (42:58:0.01, v/v/v), which was ramped to 86:14:0.01 (v/v/v) over 12.5 min followed by isocratic elution at 98:2:0.01 (v/v/v) for 3 min[28]. Eluted lipid mediators were detected by (scheduled) multiple reaction monitoring using a QTRAP 5500 mass spectrometer (Sciex, Framingham, MA), which was equipped with an electrospray ionization source that was operated in negative mode. Acquired mass spectra were processed using Analyst 1.6.2 (Sciex).

**Determination of the intracellular $Ca^{2+}$ concentration.** The determination of the intracellular $Ca^{2+}$ concentration was performed as previously described[29]. Briefly, 5 x $10^6$ PMNL or 1 x $10^6$ M1 macrophages were stained using 1 µM Fura-2-AM (30 min, 37 °C), washed with Krebs-HEPES buffer (centrifuged at 1000 rpm, 5 min, 4°C) and resuspended in 3 ml ice-cold Krebs-HEPES buffer plus BSA. 200 µl of cell suspension containing 5 x $10^5$ PMNL or 2.5 x $10^5$ M1 macrophages was seeded per well into a 96-well plate (black, clear bottom). Cells were incubated with 1 mM $CaCl_2$ for 10 min at 37 °C. To determine intracellular $Ca^{2+}$ concentrations, a microplate fluorometer NOVOstar (BMG Labtech Optima) was used to measure the $Ca^{2+}$-dependent fluorescence of the intracellular dye over time at 37 °C. Fluorescence at 340 nm ($Ca^{2+}$ chelating Fura-2-AM) and 380 nm (free Fura-2-AM) was measured every 1.18 s for 225 kinetic cycles (4.4 min total time). An automatic injecting system was used to perform the addition of the following solutions: 2 µl of 100-fold compound solution or vehicle (1% (v/v) DMSO) were injected after 11.8 s, 20 µl of 10% (v/v) triton-X in Krebs-HEPES buffer plus BSA were injected after 218.3 s, and 16.6 mM of EDTA was added after 253.7 s. Each treatment was performed in duplicates. Intracellular $Ca^{2+}$ was calculated from the ratio of fluorescence at 340/380 nm. Maximal $Ca^{2+}$ release was calculated as:

$$[Ca^{2+}]_{max}\ (n\text{-}fold\ of\ CTRL) = \frac{[Ca^{2+}]_{\max(cmpd)} - [Ca^{2+}]_{\mathrm{basal}(cmpd)}}{[Ca^{2+}]_{\max(vehicle)} - [Ca^{2+}]_{\mathrm{basal}(vehicle)}} \times 100\% \quad (8)$$

**Determination of intracellular ROS formation.** The detection of ROS was conducted using the peroxide-sensitive fluorescence dye DCFH-DA. PMNL were diluted to 5 x $10^6$ cells/ml in PBS-glucose buffer and 100 µl per well were seeded into a 96-well plate (black, clear bottom). Compound, vehicle (2% (v/v) DMSO) or DPI (positive control inhibitor of ROS formation) and 100 µl of ROS measuring solution (2 µg/ml DCFH-DA, 2 mM $CaCl_2$, in PBS-glucose buffer) were added and incubated for 10 min at 37 °C under exclusion of light. Afterwards, either PBS-glucose buffer or 1

μM PMA was added and ROS formation was measured immediately (t0) at 37 °C using microplate fluorometer NOVOstar (BMG Labtech Optima). Excitation occurred at 485 nm, where emission was measured at 520 nm every 6 s for 150 kinetic cycles (15 min total time). The detection level was reached after 600 s and thus, evaluation was performed after 480 s (t1) when the curve was still in the linear range. ROS formation was calculated from fluorescence change (t1 – t0) as percentage of vehicle control, values of cells treated with vehicle and stimulated with PMA were set as 100%. Calculation was performed as:

$$ROS\ formation\ (\%\ of\ CTRL) = \frac{F_{520\ (t1)\ cmpd} - F_{520\ (t0)\ cmpd}}{F_{520\ (t1)\ vehicle} - F_{520\ (t0)\ vehicle}} \times 100\% \quad (9)$$

**Statistical Analysis.** Data are presented as mean ± S.D. of n observations, where n represents the number of independent experiments performed at different time points. General statistical analyses were carried out using a two-sided Student's t-test at a confidence interval of 95%.

## Results and discussion

**Building ML models for discovery of 5-LO pathway modulators.** Suppression of cellular 5-LO product ($LTA_4$, 5-H(P)ETE) formation can be achieved by inhibition of $cPLA_2$, FLAP, 5-LO, and by modulation of cofactors such as $Ca^{2+}$, ROS/hydroperoxides, MAPK kinases, and phospholipids/glycerides[30] that collectively are referred here as "5-LO pathway". With the goal of developing a tool that would help identifying modulators of the 5-LO pathway we collected and curated ChEMBL 23 data pertaining to the targets of interest. Firstly, we collected data on 5-LO inhibitors. 5128 entries describing small molecules inhibiting the 5-LO enzyme were downloaded from the CHEMBL 23 database, which were then cleaned by removing duplicates and only keeping the relationships that declared exact $IC_{50}$, $K_D$ or $K_i$ values, trimming down the set to 2456 5-LO inhibitors (Fig. 1). For building the models, compounds with $IC_{50}$/$K_D$/$K_i$ values below 1 μM were classified as strong inhibitors and the rest defined as weak/inactive. This resulted in a moderately balanced label distribution, with 1358:1098 molecules in each bin, respectively. RDKit Descriptor Calculation Node from RDKit KNIME Integration (3.4.0) was used to generate 117 physicochemical descriptors for each of the modulators, which were then normalised and used to construct three predictive models, namely Gradient Boosted Trees, Naïve Bayesian and Logistic Regression (full description provided in the Methods section). We also tested the performance of other ML models, including k-nearest neighbour (k-NN), support vector machines (SVM) and random forests (RF) but elected not to include them in our model, for a number of different reasons – k-NN and SVM both were found to predict large number of positives in downstream applications (~25% of the natural product dataset), positive predictions made by k-NN were clustered in small chemical space whereas the performance of RF was found to be slightly worse than GBT and we opted to include only one decision tree-based model in our panel. We chose to use physiochemical descriptors rather than structural fingerprints so to minimise structural biases – as natural products cover a chemical space different to that of inhibitors

resulting from medicinal chemistry campaigns, biasing our models towards particular chemical moieties was not desirable. We built these models with the aim to screen large numbers of molecules, thus it was paramount to minimise the number of false positives. As such, the key metric we were optimising these models for was True Negative Rate (Definition in Methods section). To further minimise the false positive rate, only molecules predicted to be active by the plurality of models were considered as positive predictions, i.e., akin to a jury approach. Another metric we aimed to keep as high as possible was Positive Prediction Value, to ensure that the model can identify true positives while retaining good selectivity.

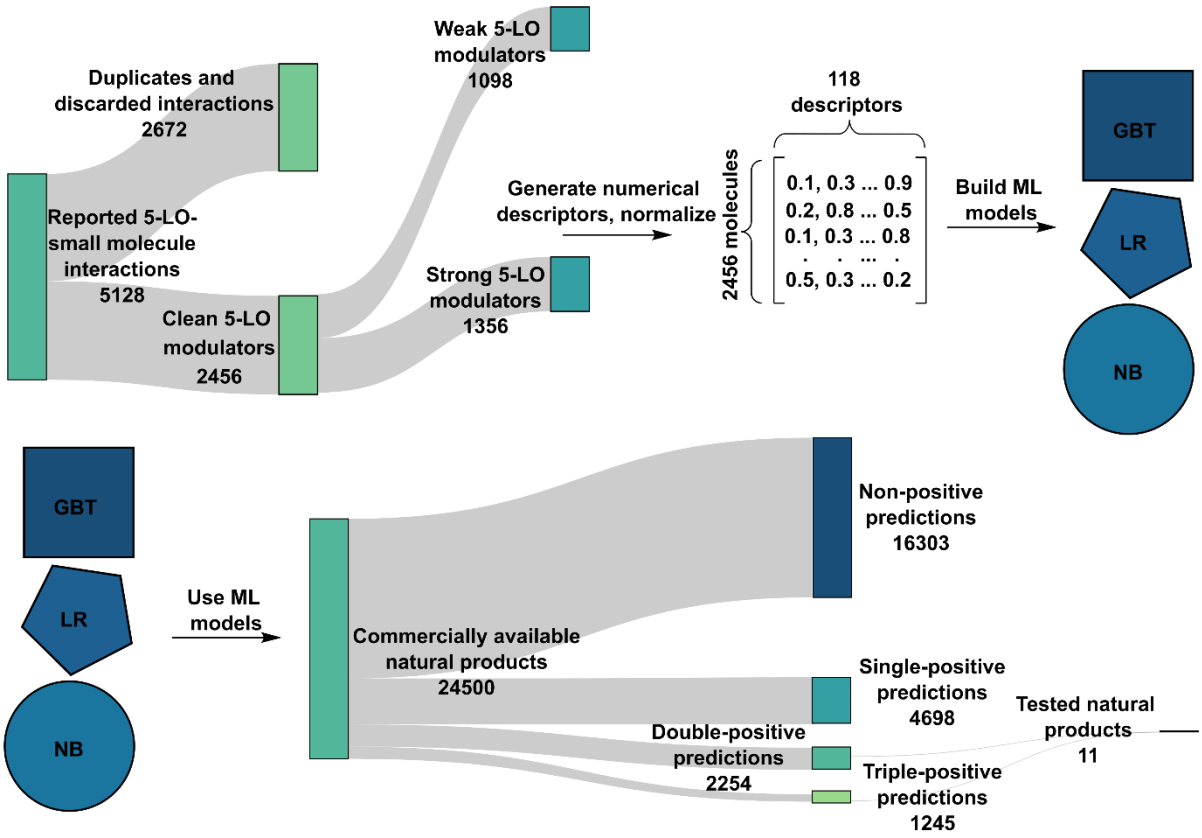



Figure 1 | **The workflow of predicting natural product 5-LO inhibitors.** Known small molecule – 5-LO interactions were cleaned up, the molecules were stratified into strong and weak/non-modulators (cut-off $IC_{50}$ = 1 μM). Numerical structure descriptors were generated for each molecule. Three predictive classification models were built (Gradient Boosted Trees, Logistic Regression and Naïve Bayesian) based on the descriptors. The models were then used to predict activities of 24,500 commercially available natural products and natural product derivatives. A subset of triple- and double-positive predictions were further investigated in PubChem database. 11 natural products with bioactivities that could be consistent with 5-LO inhibition were chosen for further testing. Similar approach was taken to predict 5-LO pathway inhibitors from dataset of FLAP activity modulators, as described in the text.

The three individual models and the consensus model (a model which considers a molecule positive only if the three individual models consider the molecule positive) were tuned and evaluated via 10-fold cross-validation. Briefly, the set of 2456 modulators was split 10 times into 9:1 training:validation subsets and used to train the models, with the resulting models then evaluated as described above. To scrutinise individual ML models, we plotted receiver operating characteristic (ROC) curves and calculated the area under curve (AUC) metric. For all the models the AUC value was found to be well over 0.5, which shows that they are providing meaningful, non-trivial predictions (Fig. S1a). Next, we calculated several different metrics for the individual and consensus models. We found that the consensus model was very good at differentiating true negatives from false positives, markedly better than any individual model (TNR = 0.89 for the consensus model, compared with 0.72, 0.52 and 0.71 for GBT, LR and NB, respectively; full metrics for all the models can be found in Table S1). The same holds true for distinguishing true positives from false positives (PPV = 0.82 for the consensus model, compared with 0.78, 0.71 and 0.68 for GBT, LR and NB, respectively).

As an additional control, we randomized the target label in the modulator dataset and used it to build new machine learning models, with expectation that these models will not result in meaningful predictions as label randomization breaks the link between biophysical information and modulatory activity. Indeed, the MCC for all three models was close to 0 (0.03, -0.01, -0.01 and 0.02 for GBT, LR, NB and consensus, respectively), implying random predictions. The non-randomized consensus model outperformed the randomized model according to all the metrics tested (TNR = 0.89 and 0.79, PPV = 0.82 and 0.58, MCC = 0.34 and 0.02 for the consensus prediction, and the consensus prediction with the shuffled label, respectively), showing that our models indeed provide meaningful outputs when trained on non-randomized data.

**5-LO modulator fragment analysis.** Before using the implemented ML models to select suitable natural products for further testing, we aimed at better understanding molecular features which might influence their ability to modulate the 5-LO pathway. For this aim we carried out a motif analysis on the known 5-LO inhibitors. Briefly, we counted the occurrence of all the fragments sized 5 to 9 atoms from the strong and weak modulator datasets. Using that we calculated the frequency of each fragment that corresponds to the fraction of molecules in a dataset containing a particular fragment (Fig. 2a, Table S2). In addition, we calculated ratio of frequencies for the strong versus weak datasets; this ratio shows which molecular fragments are overrepresented in the strong 5-LO inhibitor dataset thus might play a role in 5-LO modulation (Fig. 2b). Notably, we found that moieties known to affect ROS levels were overrepresented in the strong inhibitor dataset (e.g. paraquinones). This corroborates

findings presented in the literature – as 5-LO is an iron-dependent enzyme, redox cyclers can interfere with its oxidation state, thus reducing its activity[14].

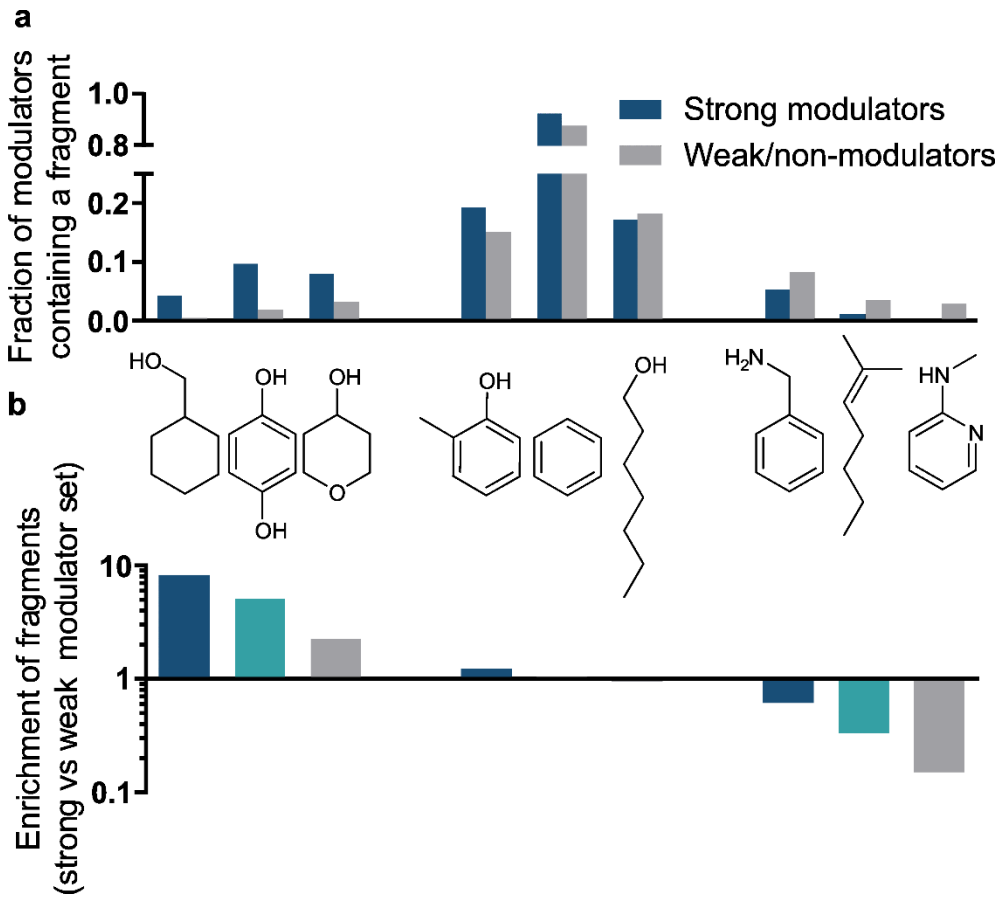



Figure 2 | **Examples of molecular fragments and their representations in the datasets of strong and weak 5-LO modulators.** (a) Frequency of selected fragments in strong and weak 5-LO modulator datasets. (b) Enrichment of selected fragments in the strong modulator dataset. Values above 1 correspond to fragment being more common in the strong modulator dataset compared to weak modulator dataset and values below 1 imply the contrary.

**Using a dataset of FLAP inhibitors to build models for discovery of 5-LO pathway modulators.** FLAP is a nuclear membrane-bound protein which assists 5-LO in product formation in intact cells by facilitating the access of 5-LO to its substrate AA[31]. Upon inspecting the dataset of FLAP inhibitors in ChEMBL, we have noticed a similar trend as for 5-LO – a sizeable part of reported FLAP inhibitors were determined via cell-based 5-LO assay, which screens for FLAP-dependent 5-LO product formation. Thus, we utilized our approach on this dataset with the aim of finding 5-LO pathway modulators.

To implement predictive ML models for FLAP we took an identical approach to the one with 5-LO. The FLAP inhibitor dataset was collected and cleaned the same way as the 5-LO dataset, leading to 2245 molecules being used for building FLAP models. Notably, the activity labels in this set were imbalanced, with 92% (2068) of the molecules being classified as strong inhibitors ($IC_{50}/K_D/K_i$ < 1 μM) whereas this is the

case for 55% of molecules in the 5-LO modulator dataset. As such, we used Balanced Accuracy (BA) as the metric to evaluate the resulting models as it is better suited for imbalanced datasets. LTA4 hydrolase, LTC4 synthase and (cPL)A2, other members of 5-LO pathway, have much lower numbers of reported inhibitors with less chemical diversity, thus we decided to focus solely on FLAP.

Using the FLAP modulator dataset, we have built GBT, LR and NB models. All had ROC AUC metrics above 0.5, thus their predictions were non-trivial (Fig. S1b). Again, the consensus model had superior metrics compared to individual models (BA = 0.67, 0.57, 0.54, 0.65 for consensus, GBT, LR and NB models, respectively, with full metrics in Table S1). As a control, we randomized the activity label while maintaining the class imbalance. As expected, these shuffled-data models could not predict the target class (MCC = 0.01, -0.01, 0.06, -0.01 for consensus, GBT, LR and NB models; MCC = 0.18 for consensus FLAP model).

**Screening of predicted 5-LO modulators.** With the models to predict 5-LO modulators and the motif analysis at hand, we went on to predict 5-LO pathway modulatory capabilities on the dataset of 24,500 natural products and their derivatives. 1,245 molecules were predicted as active by all the three ML models (Table S3). To further triage the list, we investigates known biological activities of these molecules in the PubChem database (v. 1.5), and chosen molecules with profiles which could arise from 5-LO inhibition. We looked out for properties that would be expected of 5-LO inhibitors, e.g., anti-inflammatory, anti-arthritic, anti-allergic properties as well as growth inhibition of cancer cell lines sensitive to 5-LO depletion. Based on these predictions, we have chosen 8 molecules from the pool of natural products predicted to be 5-LO modulators by all three models as well as 3 molecules predicted to be positive by two models (Fig. 3a, b).

As our aim was to use the ML models to predict and test novel 5-LO inhibitor space, we needed to be sure that the chosen molecules were structurally disparate from ones already tested against 5-LO and used to build the ML models. To investigate this, we carried out a similarity search between the molecules chosen for testing and 5-LO modulator dataset, so to find the most similar molecules between the two sets. Briefly, we converted the molecules into two different kinds of fingerprints and conducted Tanimoto similarity search. Molecules found to be the most similar to chosen natural products and corresponding Tanimoto coefficients reported in Table S4. Importantly, only one of our chosen molecules was found to have a similar scaffold to an already tested molecule – the flavone nobiletin. This demonstrates that the ML models used to predict 5-LO modulators are indeed capable of exploring previously untested chemical space.

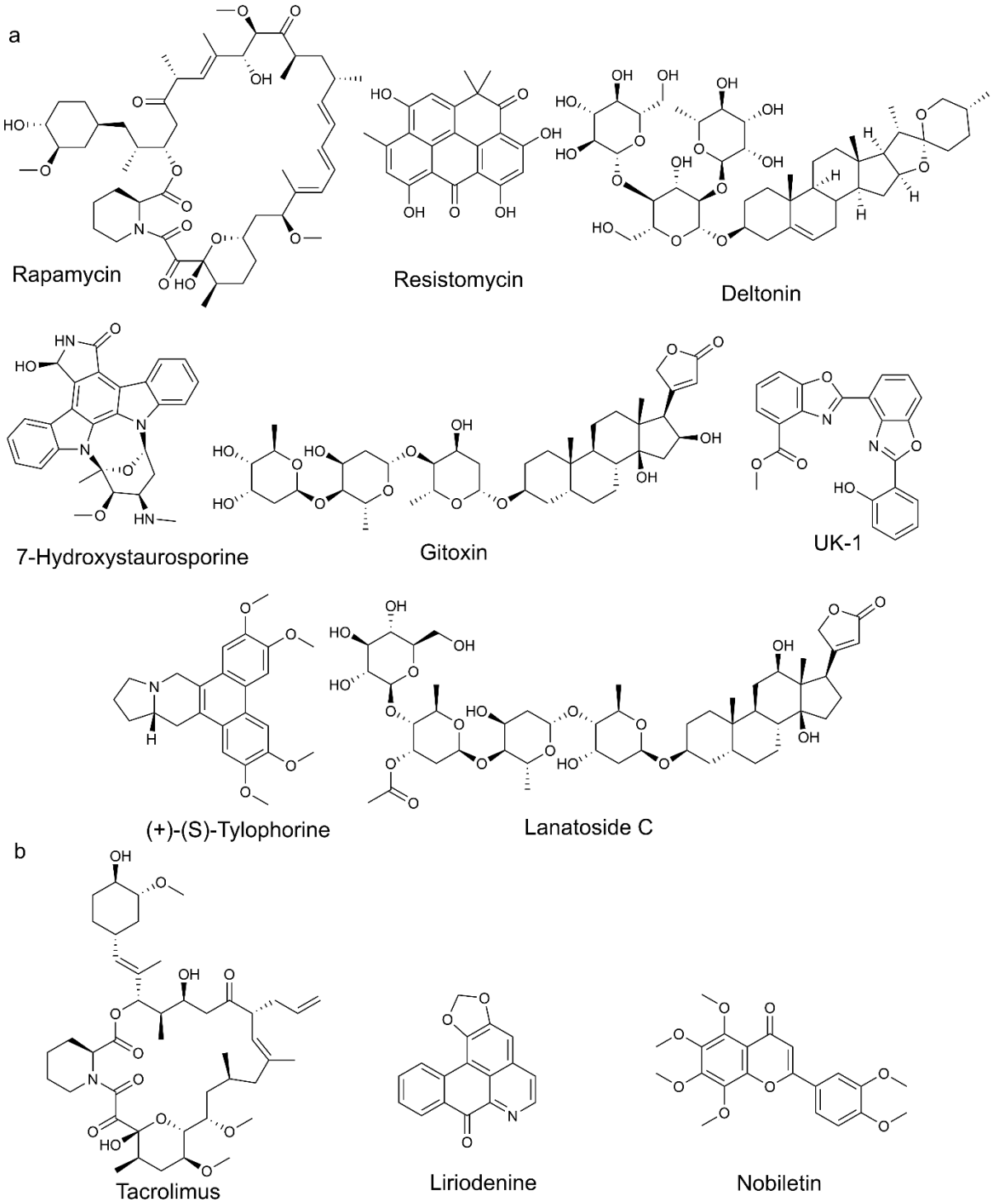



Figure 3 | **Natural products chosen for biological testing as 5-LO pathway modulators.** (a) 8 natural products predicted to be active against 5-LO by all 3 predictive AI models. (b) 3 natural products predicted to be active by 2 AI models.

For the initial assessment of activity of the 11 molecules, we carried out two assays – inhibition of isolated 5-LO, which reports whether a molecule is a direct 5-LO enzyme inhibitor, and inhibition of 5-LO in $Ca^{2+}$-ionophore A23187-activated

polymorphonuclear leukocytes (PMNL), which reports both direct 5-LO enzyme and 5-LO pathway modulators. As reference drug, the 5-LO inhibitor zileuton (3 μM) was used, as reported before,[32] and this drug blocked 5-LO activity in both test systems as expected (not shown). For the initial assessment we tested two concentrations for each molecule - 1 and 10 μM. Most of the molecules tested had some effect on 5-LO with at least two of them appearing to inhibit 5-LO directly with $IC_{50}$ values in the nanomolar range – resistomycin in the direct enzyme inhibition assay, deltonin in the cell-based PMNL assay (Fig. 4a, b). Importantly, neither of the molecules are similar to the ones previously tested against 5-LO, so constitute novel 5-LO modulator chemical space (Table S4). Resistomycin's inability to inhibit 5-LO product formation in PMNL could arise from its inability to cross cellular membranes due to the potential negative charge (its predicted p$K_a$ values are 6.66, 7.58, 8.11 and 8.73, suggesting negative charge under physiological conditions). Conversely, deltonin was only active in the cell-based PMNL assay, which suggests it being a 5-LO pathway modulator, thus we investigated it further.

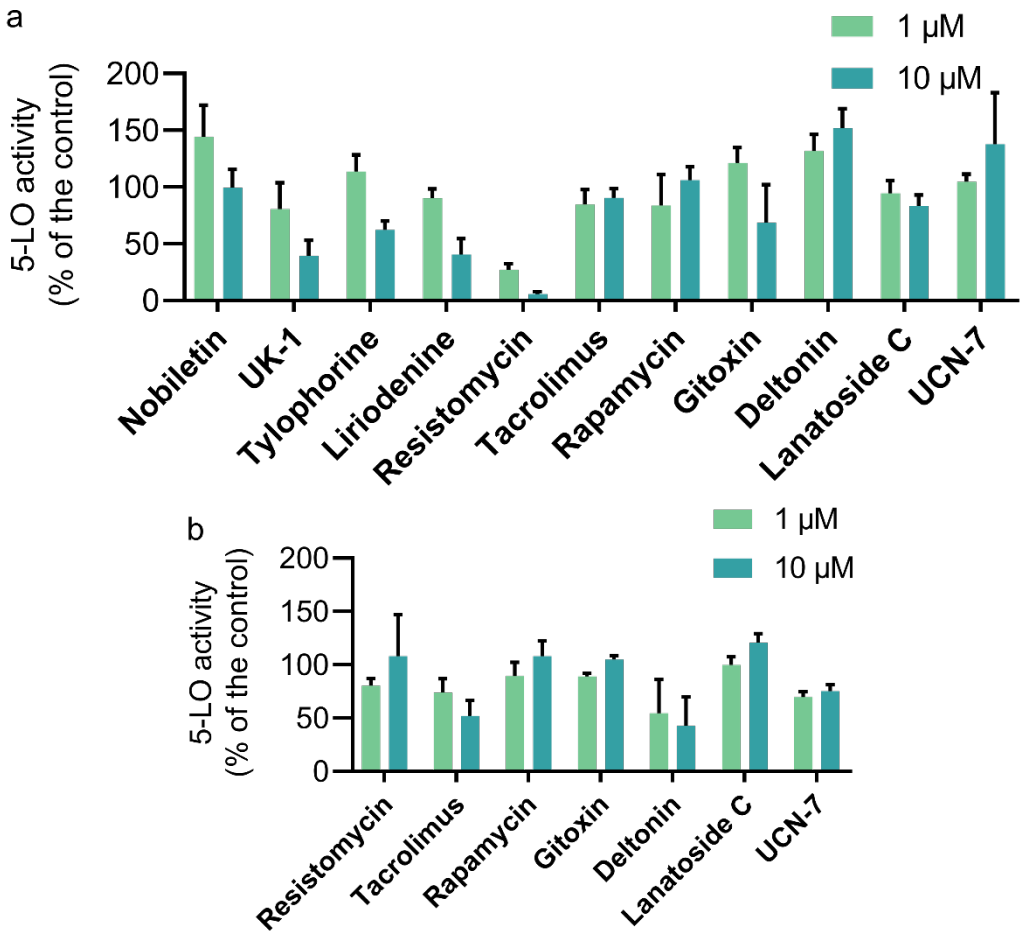



Figure 4 | **Screening results for 12 natural products as 5-LO pathway modulators.** (a) Ability of natural products to inhibit the activity of 5-LO directly in a cell-free assay, as inferred from suppressing 5-LO product ($LTB_4$ trans-isomers, and 5-HETE) formation, with natural product concentrations in μM. (b) Ability of natural products to inhibit 5-LO product ($LTB_4$, its trans-isomers, and 5-HETE) formation in the PMNL assay, with natural product concentrations in μM. Data are given as means + S.D., n = 3.

Similarly, we used the ML models made using the FLAP dataset to predict 5-LO pathway modulators. This resulted in 940 positive predictions by the consensus model - similar number to 1,245 for 5-LO (Table S5). We have chosen 10 molecules for further testing based on their known activities which we investigated in PubChem database (Fig. S2a). Chemical similarity search has again demonstrated that the molecules constitute novel chemical space compared to known 5-LO modulators, with pentacyclic triterpenes being the only overlapping scaffolds (Table S4). For initial screening, these compounds were tested either in an assay with isolated 5-LO or by using A23187-activated PMNL to study modulation of cellular 5-LO product formation; the FLAP antagonist MK886 (0.3 μM) was used, as reported before.[33] Intriguingly, three of the compounds – rhodomyrtone A, tipranavir and bevirimat - were found to be potent inhibitors of 5-LO product synthesis in the PMNL assay with reduced efficiency against the 5-LO enzyme under cell-free conditions, which is consistent with inhibition of the 5-LO/FLAP interaction (Fig. S2b, c). Another compound, manzamine A, was found to inhibit 5-LO directly but not in the intact PMNL assay (Fig. S2b, c).

We further investigated if these compounds may act as disruptors of the 5-LO – FLAP interaction. In the cellular context, FLAP is required for biosynthesis of 5-LO products from endogenously released substrate; however, it is not necessary when 5-LO converts exogenous AA under cell-free conditions[31]. Thus, if a molecule interferes with 5-LO-FLAP interactions, this inhibition can be partially overcome by providing cells with exogenous substrate. We carried out concentration-response experiments using the PMNL assay for the three FLAP dataset compounds active in both PMNL and direct 5-LO assays, either in absence or presence of exogenous AA (Fig. S2d). Interestingly, only one of the three compounds – bevirimat – partially lost its 5-LO inhibitory activity in the presence of AA, whereas the other two molecules became more potent. Moreover, bevirimat was found to be the strongest direct 5-LO inhibitor, so in principle it could interfere with both 5-LO catalytic activity and the 5-LO-FLAP interaction.

**Deltonin does not permeabilize cells akin to digitonin.** Deltonin is a saponin, thus we decided to compare its bioactivity with another molecule from the same family of natural products, digitonin. The alkaloid parts in these two molecules differ only by a double bond and a couple of hydroxyl groups, with the major difference between these molecules being the glycoside part (Fig. 5a). Digitonin permeabilizes cellular membranes[34], thus we explored the possibility of deltonin having similar features. Firstly, we tested the ability of these compounds to induce lysis of PMNL. For this aim, we have carried out an LDH assay, in which we treated PMNL with one of the two saponins for 20 min. This was followed by quantification of extracellular LDH which leaks from cells with compromised membranes, to evaluate the extent of cell lysis. We found that deltonin can have some effect on the integrity of cellular membranes at the highest concentration tested (10 μM), although this might be a result of its cytotoxicity. Digitonin, on the other hand, induced a complete lysis of the cells at the same concentration, in line with reports by others[34]. To gain further insight into deltonin's mode of action, we carried out a haemolysis assay. Briefly, we incubated deltonin with sheep blood for 24 hrs, followed by quantification of the released haemoglobin. Some

lysis was detected with 10 μM deltonin, potentially a result of its cytotoxicity, which echoes the results of LDH assay (Fig. S3). As discussed previously, deltonin can extensively inhibit the activity of 5-LO at lower concentrations, thus cell lysis cannot be the mechanism behind this phenomenon.

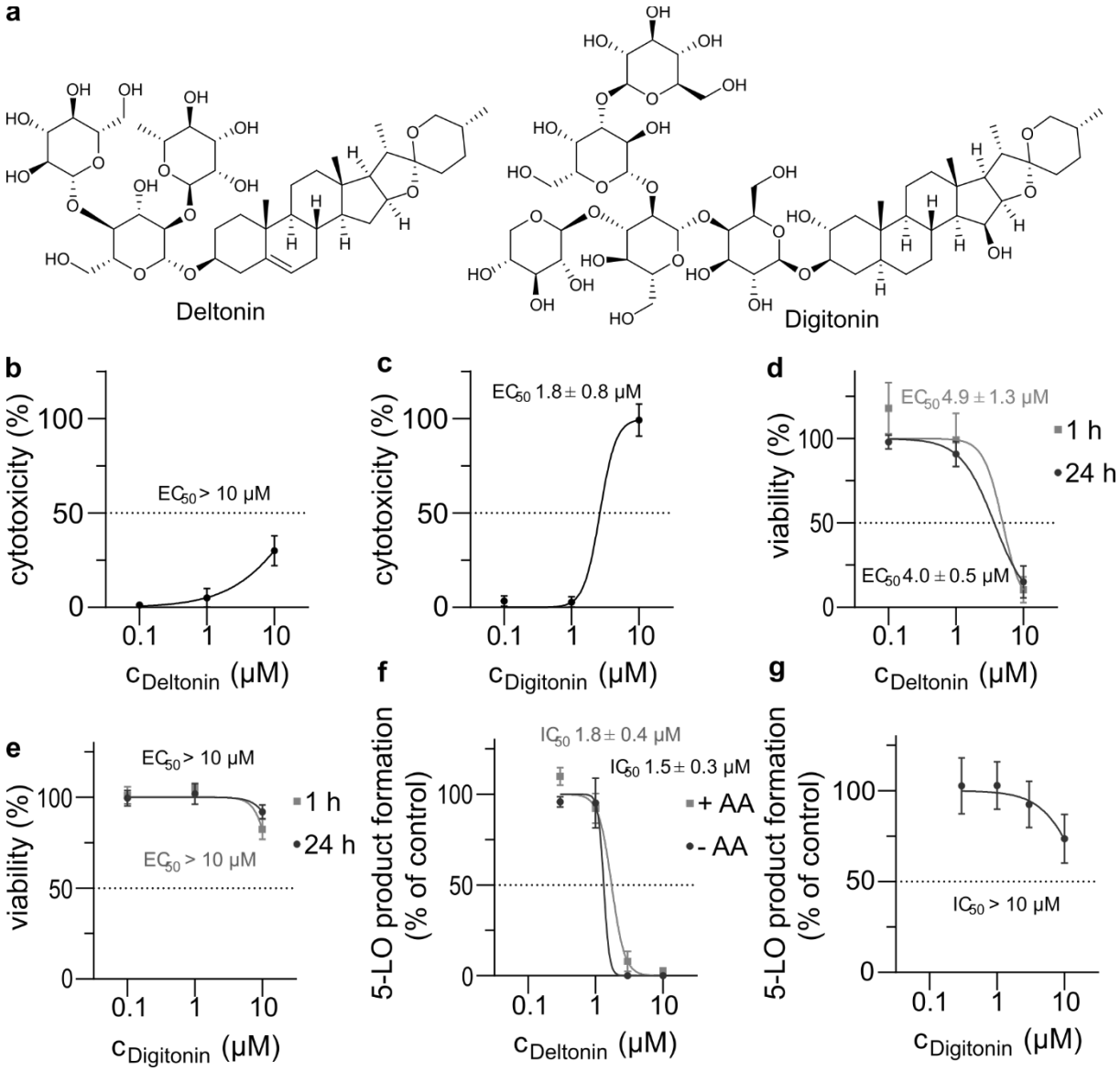



Figure 5 | **Deltonin and digitonin belong to the saponin family of natural products but act differentially on PMNL and PBMC.** (a) Structures of deltonin and digitonin. (b) LDH assays were performed to evaluate deltonin-induced cell lysis of PMNL. (c) LDH assays to evaluate digitonin-induced cell lysis of PMNL. (d) Results of an MTT assay to evaluate cytotoxicity of deltonin towards PBMC over 1 and 24 hrs. (e) Results of an MTT assay to evaluate cytotoxicity of digitonin towards PBMC over 1 and 24 hrs. (f) Concentration-dependent inhibition of 5-LO in a PMNL assay by deltonin, as inferred from 5-LO product ($LTB_4$, its trans-isomers, and 5-HETE) formation, with or without supplementation of exogenous arachidonic acid (AA). (g) Concentration-dependent inhibition of 5-LO in a PMNL assay by digitonin as inferred from 5-LO product ($LTB_4$, its trans-isomers, and 5-HETE) formation. Data are means ± S.D., n = 3.

Next, we investigated the short- and long-term cytotoxicity of deltonin and digitonin. For this aim, we treated PBMC (peripheral blood mononuclear cells) with one of the two saponins for 1 or 24 hrs (PMNL are too short-lived for this assay analysing effects at 24 hrs). We found that deltonin exhibits potent cytotoxicity, with estimated $EC_{50}$ value of 4.0 µM after 24 hrs treatment, in contrast with digitonin that was hardly active (Fig. 5d,e). Interestingly, little difference was observed between 1- or 24-hour treatment for both saponins, indicating a mechanism of action faster than the timeframes tested. Altogether, this corroborates that the observed deltonin-induced lysis is likely a result of its cytotoxicity whereas digitonin permeabilizes cells, thus the two compounds have different mechanisms of action despite similarities in structure.

**Deltonin inhibits 5-LO product formation in PMNL but not in M1 macrophages.** We went on to establish an $EC_{50}$ value against 5-LO for both compounds in the PMNL assay format as well as to investigate whether inhibition by deltonin could be reversed by carrying out the assay with exogenously added 5-LO substrate AA. Deltonin was found to inhibit 5-LO product formation with an $IC_{50}$ value of 1.5 µM, with exogenous AA (20 µM) having little influence on the inhibitory potency (Fig. 5f). This suggests that 5-LO inhibition is not a result of substrate depletion (due to $cPLA_2$ inhibition) or delivery inhibition (by blocking FLAP), as otherwise AA would reverse it. Digitonin was found to somewhat affect 5-LO product biosynthesis at higher concentrations, likely as result of compromising cell membranes (Fig. 5g).

The 5-LO pathway is also expressed in other innate immune cells, e.g., M1 macrophages. Thus, we investigated whether treatment of human monocyte-derived M1 macrophages with deltonin could affect formation of 5-LO products in these cells. To activate the M1 macrophages, we utilized either stimulation with A23187 or pathogenic *E. coli*. In both cases, treatment with deltonin had no significant effects (Fig 6a-d), in contrast to the 5-LO inhibitor zileuton that blocked 5-LO product formation in these cells under the same conditions.[28] Like in PMNL, also in M1 macrophages, digitonin failed to effectively inhibit 5-LO product formation. Additionally, we tested whether deltonin could affect the release of AA and other polyunsaturated fatty acids (PUFAs) and the activity of enzymes involved in the formation of additional inflammation-related lipid mediators in M1 macrophages, such as COX-1/2, 12-LO and 15-LO (Fig. S4a-h). Compared to 5-LO, we did not observe significant changes in the levels of various lipid mediators made by these enzymes. Altogether, this demonstrates that deltonin modulates 5-LO and lipid signalling in a cell type-selective

manner, being a potent 5-LO pathway modulator in PMNL, devoid of efficacy in M1 macrophages.

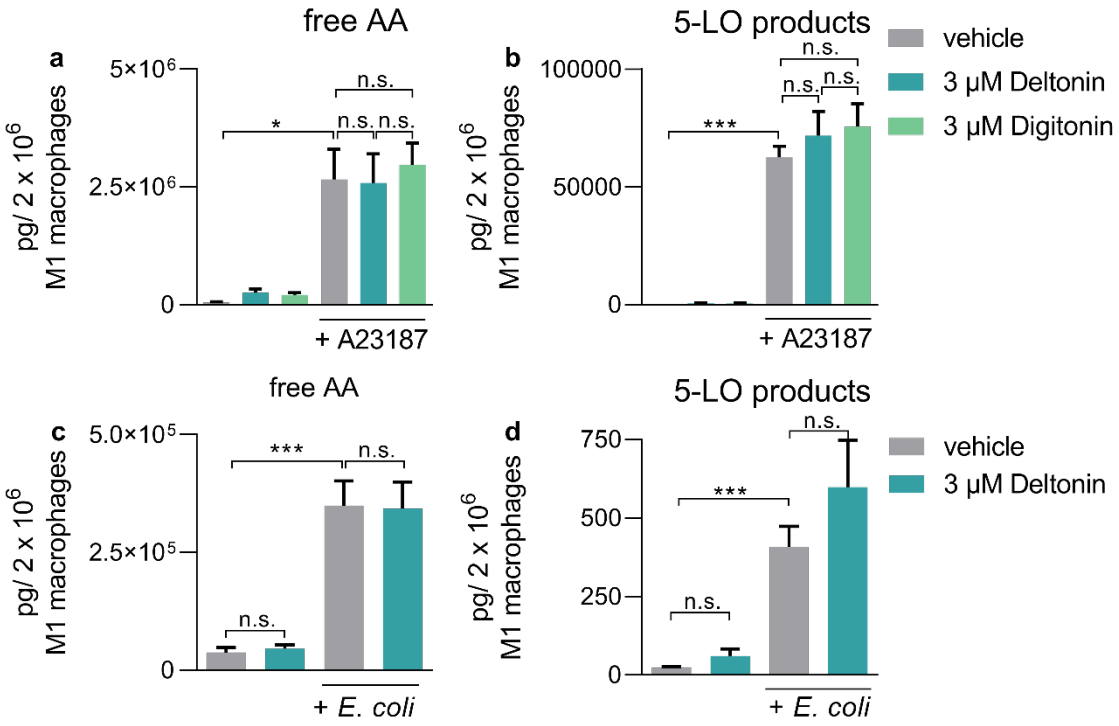



Figure 6 | **Treatment of activated M1 macrophages with deltonin or digitonin does not modulate 5-LO activity or substrate availability.** (a) Effects of deltonin or digitonin on AA release in A23187-stimulated M1 macrophages (n = 3). (b) Effects of deltonin or digitonin on the formation of 5-LO products ($LTB_4$, its trans-isomers, and 5-HETE) in A23187-stimulated M1 macrophages (n = 3). (c) Effects of deltonin or digitonin on the release of AA in *E. coli*-stimulated M1 macrophages (n = 4). (d) Treatment with deltonin or digitonin does not modulate the formation of 5-LO products in *E. coli*-stimulated M1 macrophages (n = 4). Data are means + S.D.; n.s. = not significant, * $p < 0.05$, *** $p < 0.001$.

**Deltonin elevates intracellular $Ca^{2+}$ levels in PMNL.** 5-LO and $cPLA_2$ activities are stimulated by $Ca^{2+}$, which is why $Ca^{2+}$ ionophores can be used to induce 5-LO product formation in intact cells[35,36]. On the other hand, exposure of 5-LO to $Ca^{2+}$ in the absence of AA (or other PUFAs as substrates) can inactivate the enzyme in a time-dependent manner due to complex redox-dependent mechanisms affecting the catalytic cycle of the active-site iron[37,38, 39]. For example, sphingosine-1-phosphate (S1P) induced $Ca^{2+}$ mobilization in PMNL and thereby led to irreversible inactivation of 5-LO seemingly involving ROS. Thus, we investigated if deltonin and digitonin could affect intracellular $Ca^{2+}$ levels in PMNL. For this aim, we utilised Fura-2-AM, a ratiometric $Ca^{2+}$ indicator in intact cells. As positive control, we used fMLP, a peptide known to elevate intracellular $Ca^{2+}$ and to activate macrophages, because unlike many

$Ca^{2+}$ ionophores (e.g. A23187), it does not interfere with fluorescent probes which we used in these assays[40].

We measured temporal intracellular $Ca^{2+}$ levels up to 200 s, after addition of vehicle, fMLP, deltonin or digitonin. fMLP caused rapid $Ca^{2+}$ influx, with the effect dissipating over 100 s and reaching the base level. Deltonin also led to a rapid increase of $Ca^{2+}$, but unlike fMLP, the increased $Ca^{2+}$ levels remained elevated throughout the course of the experiment (Fig. 7a,b). The effect of deltonin was concentration-dependent – the maximal elevation of $Ca^{2+}$ was similar to one achieved with fMLP (both at 1 µM) but higher levels were reached with deltonin at 5 and 10 µM. Digitonin also led to increase in intracellular $Ca^{2+}$ at higher concentrations but much less pronounced, even when 10 µM were used; the effect was smaller than with fMLP at 1 µM (Fig. 7a, b). This demonstrates that deltonin potently elevates $Ca^{2+}$ levels in PMNL, which might contribute to its 5-LO inhibitory activity.

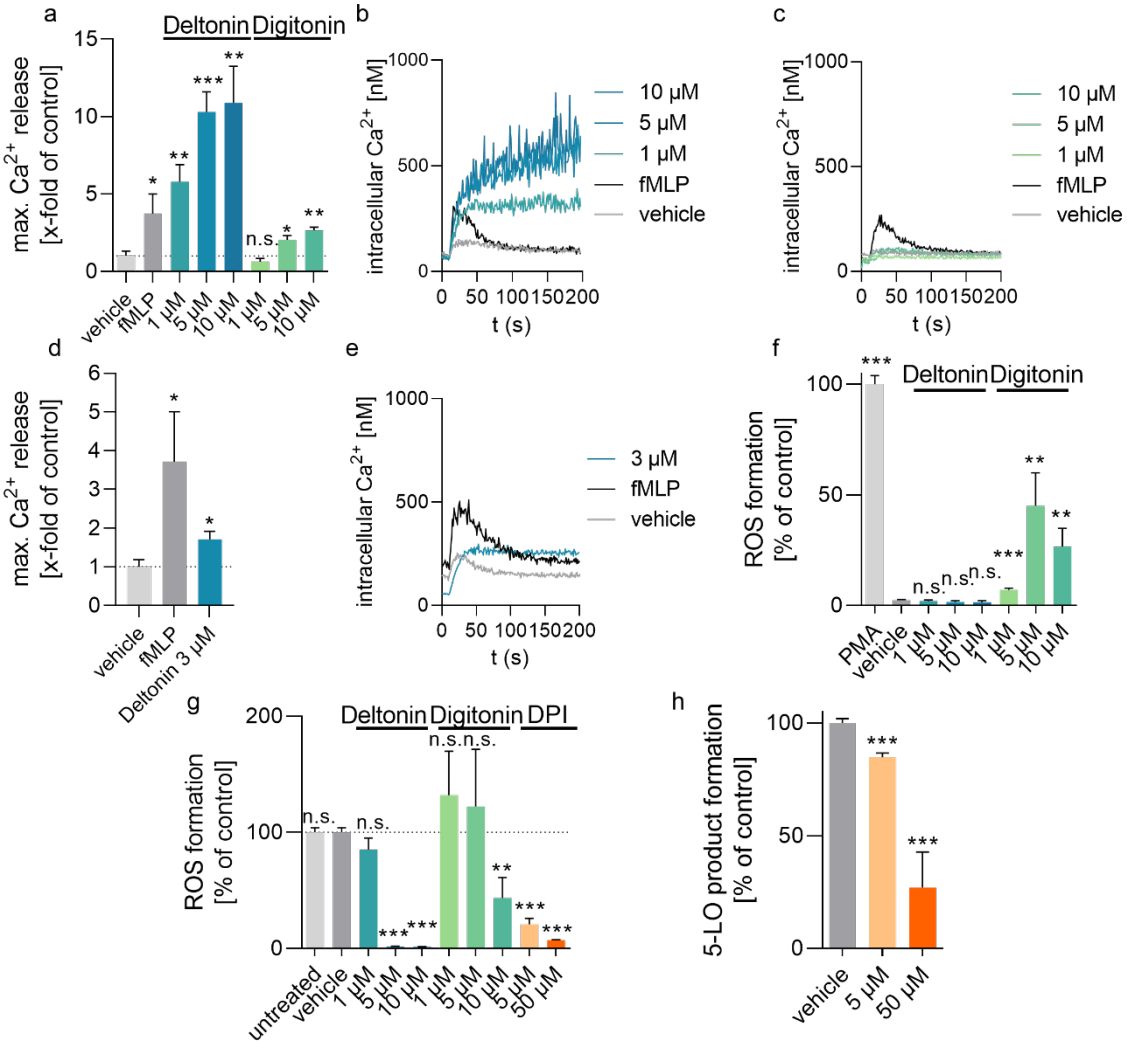



Figure 7 | **Deltonin potently modulates $Ca^{2+}$ influx and ROS levels in PMNL.** (a) Maximal calcium levels in PMNL upon treatment with fMLP (1 µM), deltonin or digitonin (n = 3). (b) Real-time measurement of $Ca^{2+}$ levels in PMNL upon treatment with fMLP or deltonin (n = 3, representative graph shown). (c) Real-time

measurement of $Ca^{2+}$ levels in PMNL upon treatment with fMLP or digitonin (n = 3, representative graph shown). (d) Maximal $Ca^{2+}$ levels in M1 macrophages upon treatment with fMLP or deltonin (n = 3). (e) Real-time measurement of $Ca^{2+}$ levels in M1 macrophages upon treatment with fMLP or deltonin (n = 3, representative graph shown). (f) Levels of ROS in PMNL stained with DCFH-DA (2 µg/ml) upon treatment with deltonin and digitonin (n = 3). (g) The effect of deltonin, digitonin and DPI on ROS formation on PMNL treated with PMA (1 µM) (n = 3). (h) Effect of DPI on 5-LO product ($LTB_4$ and its trans-isomers, 5-HETE) formation in PMNL (n = 3). Data are given as means + S.D.; n.s. = not significant, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$ when compared to the vehicle control.

If modulation of $Ca^{2+}$ levels indeed underlies deltonin-induced 5-LO inhibition in PMNL, $Ca^{2+}$ modulation by deltonin should be less pronounced in M1 macrophages as it does not inhibit 5-LO product formation in these cells. Thus, we have carried out real-time $Ca^{2+}$ measurements upon treating M1 macrophages with fMLP or deltonin. The known macrophage activator fMLP induced a strong elevation of intracellular $Ca^{2+}$, as expected (Fig. 7d, e). Deltonin treatment led to a modest increase in $Ca^{2+}$ levels, albeit a much smaller one when compared to its effect on PMNL or treatment of M1 macrophages with 1 µM fMLP under the same conditions. This further supports the hypothesis that deltonin-induced 5-LO inactivation occurs through $Ca^{2+}$ modulation as treatment with deltonin affects both processes more in PMNL than in M1 macrophages.

**Deltonin is a redox modulator.** The activity of 5-LO depends on the redox environment in the cell. 5-LO is a ferroprotein, with its catalytic cycle initiated by oxidation of $Fe^{2+}$ into $Fe^{3+}$, a process initiated by lipid hydroperoxides[14]. Thus, impaired cellular ROS levels can prevent 5-LO product formation. We measured the ability of deltonin and digitonin to affect ROS formation in PMNL by employing DCFH-DA, a redox state-sensitive fluorescent probe that gets localized inside cells upon treatment. To investigate the ability of tested molecules to induce ROS formation, we treated cells with PMA (phorbol 12-myristate 13-acetate) as positive control and compared its ability to elevate ROS with the two saponins. PMA is a known activator of NADPH oxidase, which leads to formation of ROS and subsequently lipid hydroperoxides[41]. In this experiment, deltonin failed to induce ROS formation whereas digitonin was active, although less pronounced than PMA (Fig. 7f). To test whether the compounds can prevent PMA-induced ROS formation, we pre-treated cells with one of the two compounds or the NADPH oxidase inhibitor DPI (diphenyleneiodonium) that suppresses ROS formation, followed by stimulation with PMA. Strikingly, deltonin completely abolished ROS formation at 5 and 10 µM with a milder effect at 1 µM, in line with its effects on 5-LO inhibition (Fig. 7g). 10 µM of digitonin also led to reduced ROS levels but to lesser extent than deltonin. DPI led to a dramatic reduction of ROS levels at the two tested concentrations, albeit it was not able to completely suppress ROS levels akin to deltonin. Interestingly, we found that in addition to reducing ROS levels, treatment with DPI also lead to reduced 5-LO product formation in PMNL which provides further evidence that impairment of the cellular redox tone can be linked to 5-LO product formation (Fig. 7h). Overall, this demonstrates that deltonin could be

inhibiting the 5-LO pathway in PMNL through reduction of cellular ROS levels, preventing activation of the 5-LO enzyme.

## Conclusions

We have built ML models to identify new modulators of the 5-LO pathway, using a database of known 5-LO or FLAP inhibitors as the training set. Interestingly, the two training sets contained both direct 5-LO enzyme and pathway modulators, and as such these models were able to detect both direct and indirect inhibitors. We applied these models on a dataset of natural products and their derivatives and, after further investigating their properties in PubChem database, we have chosen 12 and 10 natural products predicted to be active in the 5-LO and FLAP models, respectively. For both models, several tested molecules were found to be modulators of the 5-LO pathway: either direct 5-LO enzyme inhibitors or suppressors of 5-LO product formation in PMNL (5 molecules from the 5-LO set and 4 molecules from the FLAP set inhibited at least 50% of 5-LO activity at one or two of the tested concentrations). Overall, it demonstrates that the described approach is fit for the purpose of elucidating new natural product – 5-LO pathway interactions.

We went on to carry out an extensive mechanistic evaluation on one of the strongest discovered modulators, deltonin. The initial screen revealed this molecule to be a strong inhibitor in the cell-based PMNL assay, without direct inhibition of 5-LO. After testing several hypotheses, we found that deltonin affects both intracellular $Ca^{2+}$ levels (elevation) and the redox state (suppression) of cells, both of which are determinants for cellular 5-LO product formation. Intriguingly, we found that deltonin potently affects 5-LO in PMNL but not in M1 macrophages, thus this phenomenon is cell type-selective. Further studies are required to pinpoint more accurately which processes in the cell are affected by deltonin and to determine the basis of its cell type selectivity. Given the emerging emphasis on cell/tissue selectivity in therapeutics, understanding the mechanism through which deltonin achieves this selectivity can inform the design of novel medicines. We also tested whether natural products predicted to be active from a FLAP inhibitor dataset could suppress cellular 5-LO product formation. Intriguingly, we identified four 5-LO modulators from these predictions but only one of these molecules had an inhibitory profile that could be consistent with disruption of 5-LO – FLAP interactions. This is likely a result of the training set containing information about both direct and indirect FLAP modulators. Since FLAP has no enzymatic or any other measurable bioactivity under cell-free conditions, analysis of cellular 5-LO product formation from endogenous substrate is the only assessment of its functionality.

Together, this study provides a framework in how ML models may be used to elucidate targets and affected pathways of natural products or other small molecules as well as to investigate their biological mechanisms. The number of known natural products is exceedingly large for physical testing (e.g., SuperNatural III database reports 790,096 entries[42]), especially given their difficulty of isolation and limited availability. Natural products are rich in untapped chemical and biomechanistic diversity, hence new ways to investigate their functions can lead to a better understanding of interactions between

small molecules and biomolecules, with the potential to lead to new classes of therapeutics. As such, emerging *in silico* tools are well placed to probe the space of natural products and lead to new mechanistic and therapeutic insights.

## Data and code availability

- KNIME workflows and the relevant datasets are added as supplementary files.
- Any raw data required to reanalyze the reported findings is available from one of the lead authors upon request.


## Acknowledgements

We thank the Deutsche Forschungsgemeinschaft (DFG), Collaborative Research Center SFB 1278 “PolyTarget” (project number 316213987) and SFB1127 “ChemBioSys” (project number 239748522).


## Author contributions

Conceptualisation, S.M., T.R., O.W. and G.J.L.B.; methodology, S.M., S.L., C.K., L.D., T.R., O.W., G.J.L.B., investigation, S.M., S.L., C.K., L.D., L.S., visualisation, S.M., S.L., supervision, T.R., O.W., G.J.L.B.; funding acquisition, O.W., G.J.L.B., writing – original draft, S.M., S.L., writing – review and editing, all the authors.

## Supplemental Information

Data S1. A dataset of 5-LO modulators used to build ML models.

Data S2. A dataset of commercially-available natural products.

Data S3. A dataset of FLAP modulators used to build ML models.

Table S1. Metrics of ML models used in this study.

Table S2. List of molecular fragments in 5-LO modulators and their frequencies in strong and weak modulator subsets.

Table S3. 5-LO modulator-derived ML model predictions on the natural product dataset.

Table S4. Similarity search of tested natural products in the 5-LO modulator dataset. Numbers indicate Tanimoto similarity coefficient.

Table S5 FLAP modulator-derived ML model predictions on the natural product dataset.

Workflow S1. KNIME workflow used to build ML models.

Workflow S2. KNIME workflow used to carry out 5-LO fragment analysis.

## References


(1) Ahneman, D. T. Predicting reaction performance in C-N cross-coupling using machine learning. *Science* **2018**, *360* (6389), 613–613.

(2) Wang, Y. H.; Ribeiro, J. M. L.; Tiwary, P. Machine learning approaches for analyzing and enhancing molecular dynamics simulations. *Curr. Opin. Struct. Biol.* **2020**, *61*, 139–145.

(3) Conde, J.; Pumroy, R. A.; Baker, C.; Rodrigues, T.; Guerreiro, A.; Sousa, B. B.; Marques, M. C.; de Almeida, B. P.; Lee, S.; Leites, E. P.; et al. Allosteric Antagonist Modulation of TRPV2 by Piperlongumine Impairs Glioblastoma Progression. *ACS Cent. Sci.* **2021**, *7* (5), 868–881.

(4) Zhavoronkov, A.; Ivanenkov, Y. A.; Aliper, A.; Veselov, M. S.; Aladinskiy, V. A.; Aladinskaya, A. V.; Terentiev, V. A.; Polykovskiy, D. A.; Kuznetsov, M. D.; Asadulaev, A.; et al. Deep learning enables rapid identification of potent DDR1 kinase inhibitors. *Nat. Biotechnol.* **2019**, *37* (9), 1038–1040.

(5) Stokes, J. M.; Yang, K.; Swanson, K.; Jin, W. G.; Cubillos-Ruiz, A.; Donghia, N. M.; MacNair, C. R.; French, S.; Carfrae, L. A.; Bloom-Ackermann, Z.; et al. A Deep Learning Approach to Antibiotic Discovery. *Cell* **2020**, *181* (2), 475–483.

(6) Svensson, F.; Norinder, U.; Bender, A. Modelling compound cytotoxicity using conformal prediction and PubChem HTS data. *Toxicol. Res.* **2017**, *6* (1), 73–80.

(7) Zhang, R.; Li, X.; Zhang, X.; Qin, H.; Xiao, W. Machine learning approaches for elucidating the biological effects of natural products. *Nat. Prod. Rep.* **2021**, *38* (2), 346–361.

(8) Pang, X.; Fu, W.; Wang, J.; Kang, D.; Xu, L.; Zhao, Y.; Liu, A. L.; Du, G. H. Identification of Estrogen Receptor alpha Antagonists from Natural Products via In Vitro and In Silico Approaches. *Oxid. Med. Cell Longev.* **2018**, *2018*, 6040149.

(9) Rodrigues, T.; Reker, D.; Schneider, P.; Schneider, G. Counting on natural products for drug design. *Nat. Chem.* **2016**, *8* (6), 531–541.

(10) Young, R. J.; Flitsch, S. L.; Grigalunas, M.; Leeson, P. D.; Quinn, R. J.; Turner, N. J.; Waldmann, H. The Time and Place for Nature in Drug Discovery. *JACS Au* **2022**, *2* (11), 2400–2416.

(11) Newman, D. J.; Cragg, G. M. Natural Products as Sources of New Drugs from 1981 to 2014. *J. Nat. Prod.* **2016**, *79* (3), 629–661.

(12) Mendez, D.; Gaulton, A.; Bento, A. P.; Chambers, J.; De Veij, M.; Felix, E.; Magarinos, M. P.; Mosquera, J. F.; Mutowo, P.; Nowotka, M.; et al. ChEMBL: towards direct deposition of bioassay data. *Nucleic Acids Res.* **2019**, *47* (D1), D930–D940.

(13) Kim, S.; Chen, J.; Cheng, T.; Gindulyte, A.; He, J.; He, S.; Li, Q.; Shoemaker, B. A.; Thiessen, P. A.; Yu, B.; et al. PubChem in 2021: new data content and improved web interfaces. *Nucleic Acids Res.* **2021**, *49* (D1), D1388–D1395.

(14) Ford-Hutchinson, A. W.; Gresser, M.; Young, R. N. 5-Lipoxygenase. *Annu. Rev. Biochem.* **1994**, *63*, 383–417.

(15) Gilbert, N. C.; Newcomer, M. E.; Werz, O. Untangling the web of 5-lipoxygenase-derived products from a molecular and structural perspective: The

battle between pro- and anti-inflammatory lipid mediators. *Biochem. Pharmacol.* **2021**, *193*, 114759.

(16) Radmark, O.; Werz, O.; Steinhilber, D.; Samuelsson, B. 5-Lipoxygenase, a key enzyme for leukotriene biosynthesis in health and disease. *Biochim. Biophys. Acta Mol. Cell Biol. Lipids* **2015**, *1851* (4), 331–339.

(17) Werz, O.; Steinhilber, D. Therapeutic options for 5-lipoxygenase inhibitors. *Pharmacol. Ther.* **2006**, *112* (3), 701–718.

(18) Rodrigues, T.; Werner, M.; Roth, J.; da Cruz, E. H. G.; Marques, M. C.; Akkapeddi, P.; Lobo, S. A.; Koeberle, A.; Corzana, F.; da Silva, E. N.; et al. Machine intelligence decrypts -lapachone as an allosteric 5-lipoxygenase inhibitor. *Chem. Sci.* **2018**, *9* (34), 6899–6903.

(19) Dunsmore, L.; Navo, C. D.; Becher, J.; de Montes, E. G.; Guerreiro, A.; Hoyt, E.; Brown, L.; Zelenay, V.; Mikutis, S.; Cooper, J.; et al. Controlled masking and targeted release of redox-cycling ortho-quinones via a C-C bond-cleaving 1,6-elimination. *Nat. Chem.* **2022**, *14* (7), 754–765.

(20) Fillbrunn, A.; Dietz, C.; Pfeuffer, J.; Rahn, R.; Landrum, G. A.; Berthold, M. R. KNIME for reproducible cross-domain analysis of life science data. *J. Biotechnol.* **2017**, *261*, 149–156.

(21) Friedman, J. H. Greedy function approximation: A gradient boosting machine. *Ann. Stat.* **2001**, *29* (5), 1189–1232.

(22) Dreiseitl, S.; Ohno-Machado, L. Logistic regression and artificial neural network classification models: a methodology review. *J. Biomed. Inform.* **2002**, *35* (5-6), 352–359.

(23) Wickramasinghe, I.; Kalutarage, H. Naive Bayes: applications, variations and vulnerabilities: a review of literature with code snippets for implementation. *Soft Computing* **2021**, *25* (3), 2277–2293.

(24) Bender, A.; Schneider, N.; Segler, M.; Walters, W. P.; Engkvist, O.; Rodrigues, T. Evaluation guidelines for machine learning tools in the chemical sciences. *Nat. Rev. Chem.* **2022**, *6* (6), 428–442.

(25) Carhart, R. E.; Smith, D. H.; Venkataraghavan, R. Atom pairs as molecular features in structure-activity studies: definition and applications. *J. Chem. Inf. Comput. Sci.* **1985**, *25* (2), 64–73.

(26) Fischer, L.; Szellas, D.; Radmark, O.; Steinhilber, D.; Werz, O. Phosphorylation- and stimulus-dependent inhibition of cellular 5-lipoxygenase activity by nonredox-type inhibitors. *FASEB J.* **2003**, *17* (3), 949.

(27) Boyum, A. Isolation of mononuclear cells and granulocytes from human blood. Isolation of monuclear cells by one centrifugation, and of granulocytes by combining centrifugation and sedimentation at 1 g. *Scand. J. Clin. Lab. Invest. Suppl.* **1968**, *97*, 77.

(28) Werner, M.; Jordan, P. M.; Romp, E.; Czapka, A.; Rao, Z. G.; Kretzer, C.; Koeberle, A.; Garscha, U.; Pace, S.; Claesson, H. E.; et al. Targeting biosynthetic networks of the proinflammatory and proresolving lipid metabolome. *FASEB J.* **2019**, *33* (5), 6140–6153.

(29) Grynkiewicz, G.; Poenie, M.; Tsien, R. Y. A new generation of Ca2+ indicators with greatly improved fluorescence properties. *J. Biol. Chem.* **1985**, *260* (6), 3440–3450.

(30) Werz, O. Inhibition of 5-Lipoxygenase Product Synthesis by natural compounds of plant origin. *Planta Medica* **2007**, *73* (13), 1331–1357.

(31) Evans, J. F.; Ferguson, A. D.; Mosley, R. T.; Hutchinson, J. H. What's all the FLAP about?: 5-lipoxygenase-activating protein inhibitors for inflammatory diseases. *Trends Pharmacol. Sci.* **2008**, *29* (2), 72–78.

(32) Cerchia, C.; Küfner, L.; Werz, O.; Lavecchia, A. Identification of selective 5-LOX and FLAP inhibitors as novel anti-inflammatory agents by ligand-based virtual screening. *Eur. J. Med. Chem.* **2023**, *263*, 115932.

(33) Dahlke, P.; Peltner, L. K.; Jordan, P. M.; Werz, O. Differential impact of 5-lipoxygenase-activating protein antagonists on the biosynthesis of leukotrienes and of specialized pro-resolving mediators. *Front. Pharmacol.* **2023**, *14*, 1219160.

(34) Nishikawa, M.; Nojima, S.; Akiyama, T.; Sankawa, U.; Inoue, K. Interaction of digitonin and its analogs with membrane cholesterol. *J. Biochem.* **1984**, *96* (4), 1231–1239.

(35) Gijon, M. A.; Leslie, C. C. Regulation of arachidonic acid release and cytosolic phospholipase A2 activation. *J. Leukoc. Biol.* **1999**, *65* (3), 330–336.

(36) Rouzer, C. A.; Samuelsson, B. Reversible, calcium-dependent membrane association of human leukocyte 5-lipoxygenase. *Proc. Natl. Acad. Sci. U S A* **1987**, *84* (21), 7393–7397.

(37) Percival, M. D.; Denis, D.; Riendeau, D.; Gresser, M. J. Investigation of the mechanism of non-turnover-dependent inactivation of purified human 5-lipoxygenase. Inactivation by H2O2 and inhibition by metal ions. *Eur. J. Biochem.* **1992**, *210* (1), 109–117.

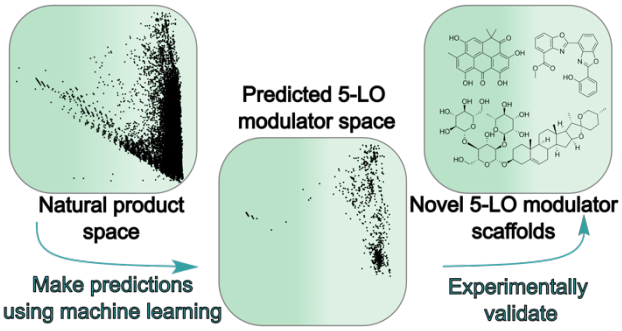
(38) Skorey, K. I.; Gresser, M. J. Calcium is not required for 5-lipoxygenase activity at high phosphatidyl choline vesicle concentrations. *Biochemistry* **1998**, *37* (22), 8027–8034.

(39) DeCarolis, E.; Denis, D.; Riendeau, D. Oxidative inactivation of human 5-lipoxygenase in phosphatidylcholine vesicles. *Eur. J. Biochem.* **1996**, *235* (1-2), 416–423.

(40) Liu, J. H.; Blanchard, D. K.; Wei, S.; Djeu, J. Y. Recombinant interleukin-8 induces changes in cytosolic Ca2+ in human neutrophils. *J. Infect. Dis.* **1992**, *166* (5), 1089–1096.

(41) Werz, O. 5-lipoxygenase: cellular biology and molecular pharmacology. *Curr. Drug Targets Inflamm. Allergy* **2002**, *1* (1), 23–44.

(42) Gallo, K.; Kemmler, E.; Goede, A.; Becker, F.; Dunkel, M.; Preissner, R.; Banerjee, P. SuperNatural 3.0-a database of natural products and natural product-based derivatives. *Nucleic Acids Res.* **2023**, *51* (D1), D654–D659.


Natural product space
Predicted 5-LO modulator space
Novel 5-LO modulator scaffolds
Make predictions using machine learning
Experimentally validate