



OPEN

## Insights into the molecular properties underlying antibacterial activity of prenylated (iso) flavonoids against MRSA

Sylvia Kalli<sup>1</sup>, Carla Araya-Cloutier<sup>1</sup>, Jos Hageman<sup>2</sup> & Jean-Paul Vincken<sup>1✉</sup>

High resistance towards traditional antibiotics has urged the development of new, natural therapeutics against methicillin-resistant *Staphylococcus aureus* (MRSA). Prenylated (iso)flavonoids, present mainly in the Fabaceae, can serve as promising candidates. Herein, the anti-MRSA properties of 23 prenylated (iso)flavonoids were assessed in-vitro. The di-prenylated (iso)flavonoids, glabrol (flavanone) and 6,8-diprenyl genistein (isoflavone), together with the mono-prenylated, 4'-*O*-methyl glabridin (isoflavan), were the most active anti-MRSA compounds (Minimum Inhibitory Concentrations (MIC)  $\leq 10 \mu\text{g/mL}$ ,  $30 \mu\text{M}$ ). The in-house activity data was complemented with literature data to yield an extended, curated dataset of 67 molecules for the development of robust in-silico prediction models. A QSAR model having a good fit ( $R^2_{\text{adj}}$  0.61), low average prediction errors and a good predictive power ( $Q^2$ ) for the training (4% and  $Q^2_{\text{LOO}}$  0.57, respectively) and the test set (5% and  $Q^2_{\text{test}}$  0.75, respectively) was obtained. Furthermore, the model predicted well the activity of an external validation set (on average 5% prediction errors), as well as the level of activity (low, moderate, high) of prenylated (iso)flavonoids against other Gram-positive bacteria. For the first time, the importance of formal charge, besides hydrophobic volume and hydrogen-bonding, in the anti-MRSA activity was highlighted, thereby suggesting potentially different modes of action of the different prenylated (iso)flavonoids.

*Staphylococcus aureus* (SA) and its oxacillin-resistant form (methicillin-resistant, MRSA) is one of the leading causes of healthcare-associated infections worldwide<sup>1</sup>. Furthermore, MRSA has also emerged as a major cause of community-associated and livestock infections<sup>2</sup>. In the United States, even though hospital-associated MRSA infections decreased by 5.4% from 2013 to 2016, community-associated infections increased by 1.6%<sup>3</sup>. In Europe, MRSA is a public health concern for Southern and Eastern European countries, in particular<sup>4</sup>. The World Health Organization has referred to MRSA as a high priority pathogen for the development of new therapeutics<sup>5</sup>. Over the last 30 years, only a few antibiotics have been approved as anti-MRSA agents, but MRSA has already developed resistance towards them<sup>6</sup>.

Novel chemical scaffolds with different modes of action to currently used antibiotics are constantly being investigated<sup>7</sup>. Prenylated flavonoids and isoflavonoids, collectively termed as (iso)flavonoids, as well as stilbenoids, have shown promising antibacterial activity against clinically-relevant pathogens, including MRSA<sup>8,9</sup>. Prenylated phenolic compounds are a class of secondary defence metabolites produced by species of the Fabaceae family upon (a)biotic stress<sup>10</sup>. Attachment of the prenyl moiety (3,3-dimethylallyl) to a phenolic skeleton is known to increase its antibacterial potency, due to the increased hydrophobicity conferred to the molecule<sup>11</sup>.

Several efforts have been made in correlating essential structural features of prenylated (iso)flavonoids with their anti-MRSA activity. Clear structure–activity relationships (SARs) remain difficult to be established due to the extensive structural diversity of prenylated (iso)flavonoids. This chemical diversity stems mainly from the different (iso)flavonoid subclasses (21 known to date). (Iso)flavones, (iso)flavanones, isoflavans, pterocarpanes, pterocarpene, 3-aryl coumarins, and 2-arylbenzofurans are some of the main (iso)flavonoid subclasses encountered in Fabaceae<sup>10,12</sup>. Chemical diversity increases further by the presence of substituents other than the prenyl group (hydroxyl, methoxyl), as well as the different configuration that the prenyl group can have (chain, pyran or furan). Therefore, SARs that apply to one subclass of prenylated (iso)flavonoids might not apply to another.

<sup>1</sup>Laboratory of Food Chemistry, Wageningen University & Research, Wageningen, The Netherlands. <sup>2</sup>Biometris, Applied Statistics, Wageningen University & Research, Wageningen, The Netherlands. ✉email: jean-paul.vincken@wur.nl

Methylation is one of the most common decorations of prenylated (iso)flavonoids. Methylation at the C7, for example, slightly reduces the anti-MRSA activity of prenylated isoflavones<sup>13,14</sup>, has no effect on the activity of prenylated isoflavans, while it was shown detrimental for prenylated coumarins<sup>13</sup>. Chain prenylation seems more favourable than ring prenylation in prenylated isoflavones with respect to anti-MRSA activity<sup>14–16</sup>. In contrast, the ring-prenylated pterocarpan, phaseollin (MIC 78  $\mu\text{M}$ ) is active, whereas its chain-prenylated analogue, phaseollidin has lower anti-MRSA activity (MIC 154  $\mu\text{M}$ )<sup>16</sup>. Prenylation at C6 of isoflavones<sup>14</sup> is better for anti-MRSA activity than C8-prenylation<sup>13</sup>, whereas the opposite seems to be preferred for flavanones<sup>17</sup>, consistent with the results for other G<sup>+</sup> bacteria<sup>18</sup>.

Since prenylation-dependent hydrophobicity is not the only determinant of antibacterial activity, more systematic SARs are necessary to rationally portray how the overall molecular characteristics influence anti-MRSA activity. In silico tools, such as QSAR and pharmacophore modelling<sup>19</sup> can aid the elucidation of these molecular characteristics. Properties, such as shape (flexibility and globularity) and surface (hydrophilic/hydrophobic regions) of prenylated (iso)flavonoids, have been shown to contribute to their antibacterial activity against the Gram-positive, *Listeria monocytogenes*<sup>18</sup>. The limitation of this QSAR analysis was the small dataset (30 molecules from 6 subclasses) used<sup>18</sup>, which did not allow splitting, thereby compromising the robustness of the models<sup>20</sup>. A (Q)SAR study on a larger dataset (67 (iso)flavonoids tested against SA or MRSA), which allowed splitting, has been performed by Sadgrove et al. (2020)<sup>21</sup>. However, this study considered molecules only from *Erythrina* species (Fabaceae) including non-prenylated, prenylated, geranylated (iso)flavonoids and a chromen-4-one derivative. By using regression QSAR modelling, the importance of hydrogen-bonding, hydrophobicity, primary oxygens and charge distribution was highlighted in this study<sup>21</sup>. Nevertheless, the importance of these molecular properties with respect to the mechanism of action of these compounds was not addressed.

In this study, a multiple linear regression model was developed from an extensive, curated dataset of 67 prenylated (iso)flavonoids from 9 subclasses tested only against MRSA by using a QSAR approach. The model was externally validated with an additional set of 10 prenylated (iso)flavonoids and further assessed for its capacity to predict the level of activity of prenylated (iso)flavonoids against other Gram-positive bacteria (including SA, *Staphylococcus epidermidis*, *Bacillus subtilis*, *Enterococcus faecalis*, and *Streptococcus mutans*). In addition, a 3D pharmacophore model was developed to visualize the structural requirements for anti-MRSA activity. Both in-silico approaches were based on molecules from both in-house and literature activity data. For the in-house data, the activity (MIC values) of 23 prenylated compounds belonging to 6 subclasses was determined.

## Materials and methods

**Materials.** Prenylated isoflavonoids (glabrene, 3'-hydroxy-4'-O-methyl-glabridin, 4'-O-methyl-glabridin, hispaglabridin A, hispaglabridin B, glyceofuran, glyceollidin II, glyceollin I, glyceollin II, glyceollin III, glyceollin IV, glyceollin V, dehydroglyceollidin II, dehydroglyceollin I, dehydroglyceollin II, dehydroglyceollin III, dehydroglyceollin IV) and one prenylated flavone (glabrol) were previously purified and chemically characterized<sup>22,23</sup>. Wightone, lupiwightone, luteone, 2,3-dehydrokivitone, licoisoflavone A, neobavaisoflavone, iso-osajin and 6,8-diprenygenistein were purchased from Plantech UK (Reading, UK). Isowightone and anhydrotuberoin were purchased from ChemFaces (Wuhan, Hubei, China). 6-Prenylnaringenin and ampicillin were purchased from Sigma Aldrich (St. Louis, MO, USA). Psoralidin and 8-prenylnaringenin were purchased from Sanbio B.V. (Uden, The Netherlands). Glabridin was purchased from Wako (Osaka, Japan). Bacto brain heart infusion (BHI) broth was purchased from BD (Franklin Lakes, NJ, USA), tryptone soya broth (TSB) and bacteriological agar from Oxoid Ltd (Basingstoke, UK), and peptone physiological salt solution (PPS) from Tritium Microbiologie (Eindhoven, The Netherlands). Ethanol absolute (EtOH) was purchased from Biosolve (Valkenswaard, The Netherlands).

**Methods.** *Antibacterial susceptibility assay.* Different prenylated (iso)flavonoids were tested for their antibacterial activity against MRSA 18HN (strain kindly provided by RIVM, Bilthoven, The Netherlands). Bacteria were streaked from a –80 °C glycerol stock onto a BHI agar plate and incubated 24 h at 37 °C. Next, one colony was transferred to BHI broth (10 mL) and further incubated for 18 h at 37 °C. These overnight cultures were diluted with TSB (final inoculum concentration  $3.8 \pm 0.4 \log \text{CFU/mL}$ ). Stock solutions of the different prenylated compounds were prepared in aqueous EtOH (70% v/v) or DMSO and subsequently diluted with TSB. A series of concentrations were tested ranging from 3.1 to 100 or 150  $\mu\text{g/mL}$  (2.1% v/v solvent max.) of prenylated (iso) flavonoids. No solubility issues were associated with prenylated (iso)flavonoids in hydrophilic growth media (TSB) at any of the concentrations tested. Equal volumes (100  $\mu\text{L}$ ) of bacteria and prenylated compound solutions in TSB were mixed into a 96-well plate. The 96-well plate was incubated in a SpectraMax M2e (Molecular Devices, Sunnyvale, CA, USA) at 37 °C with constant linear shaking. The optical density (OD) at 600 nm was measured every 5 min for 24 h.

A positive control (vancomycin, 2  $\mu\text{g/mL}$ ), negative controls (TSB suspension of bacteria with and without 2.5% (v/v) max. solvent) and blanks (compounds and TSB medium without bacteria) were used for optical comparison and sterility control. Growth inhibition was assessed by measuring the time to detection (TTD), i.e. the time to reach a change in OD of 0.05 units<sup>24</sup>. When no change in OD (i.e.  $\Delta\text{OD} < 0.05$ ) was observed after the 24 h of incubation, cell viability was confirmed by plate counting, as described by Araya-Cloutier et al. (2018)<sup>18</sup>. The minimum inhibitory concentration (MIC) and the minimum bactericidal concentration (MBC) were defined as described elsewhere<sup>18</sup>. Prenylated compounds were tested in two independent biological reproductions, each performed in duplicate.

*Inactivation kinetics.* An overnight MRSA culture was diluted to  $4.4 \pm 0.5 \log (\text{CFU/mL})$  and mixed with prenylated phenolics at their MIC and 2xMIC. Samples were incubated in duplicate at 37 °C/125 rpm. At different

time points (0, 2, 4, 6, 24 h), 100  $\mu\text{L}$  of culture medium were taken and decimally diluted in PPS. Dilutions were spread on BHI agar plates and incubated overnight at 37 °C, after which colonies were counted. Vancomycin and ampicillin at 2  $\mu\text{g}/\text{mL}$  were used as control antibiotics.

**QSAR modeling. Dataset construction.** An outline of the different steps followed for the QSAR modelling is demonstrated in Fig. S1. First, 106 prenylated (iso)flavonoids together with their anti-MRSA activity were mined from 19 different studies (2000–2017, Table S1). These studies were chosen under the premise that agar or broth (micro)dilution assays were used to determine the anti-MRSA activity. The activity data of prenylated (iso)flavonoids from 16 studies was combined with the activity data of the prenylated (iso)flavonoids generated in-house, to obtain a dataset of 94 molecules with reported MIC values. The activity data from the 3 other, randomly chosen, studies<sup>25–27</sup> was used to externally validate the model<sup>28</sup>. The choice for the construction of this independent external validation set was based on whole studies, instead of various molecules from different studies to minimize extreme experimental variability. Compounds for which no specific MIC value was obtained (i.e. MIC was higher than the highest tested concentration), were excluded from the QSAR modelling study. When more than one MRSA strain was tested and a range of MIC values was given, or more than one MIC value per compound was available, the highest MIC value was used to consider the worst-case scenario. Activity data [MIC values ( $\mu\text{M}$ )] was converted to pMIC [ $-\log \text{MIC (M)}$ ] (to improve the normality of the data distribution).

**Dataset curation.** A defined dataset of 76 prenylated (iso)flavonoids with established MICs was intended to enter the training/test phase. To ensure uniform chemical diversity upon the splitting of the modelling set, two selection criteria were applied for the compounds used for the QSAR modelling. First, (iso)flavonoids with common chain prenylation (i.e. 3,3-dimethylallyl) and ring prenylation (2''-isopropenylfuran and/or 2,2-dimethylpyran) were exclusively considered (**criterion 1**). Second, for each subgroup of mono- and di-prenylated compounds (in each subclass) two representative compounds had to be present, otherwise the subgroup was excluded (**criterion 2**). Chemical structures were extracted from literature and inputted to the modelling software (Molecular Operating Environment, MOE, v.2019.08, Chemical Computing Group, Montreal, QC, Canada) using the canonical SMILES codes from PubChem. If not available, the chemical structure was drawn manually using the PubChem sketcher to obtain the SMILES code before importing to MOE. Structures were energy minimized by using the molecular orbital package (MOPAC) PM3 forcefield at a root mean square (RMS) gradient of 0.01 kcal/mol/Å<sup>2</sup>. Then, a conformational search was performed by using LowModeMD (RSM gradient 0.1 kcal/mol/Å<sup>2</sup>, other settings were left on default) was performed for all compounds in the database.

Optimized chemical structures were used to calculate different molecular descriptors available in MOE. After eliminating descriptors that were identical for all molecules or redundant descriptors (inter-correlation  $R_{\text{pearson}} > 0.99$ ), a total number of 120 descriptors were finally incorporated into the database (list of descriptors in Table S2). Statistical modelling was performed using R (R Core Team, 2021<sup>29</sup>) and a combination of in-house created scripts and specific libraries mentioned where applicable. After the two selection criteria were applied, the modelling set of 70 complying compounds was split into a training set (80%, for model development) and test set (20%, for model selection based on model's predictive power on the test set)<sup>28</sup> using the Kennard–Stone algorithm from the R-package 'prospectr' (using principal components and retaining 95% explained variance)<sup>30</sup>. This procedure selects the best chemically representative subset as a training set, avoiding overoptimistic splitting which might occur during random splitting<sup>31</sup>.

**Model development and validation.** A genetic algorithm (GA) (R-package 'GA', version 3.2) was used to select a small subset of descriptors that best predict the antibacterial activity (pMIC). pMIC activity was modelled using multiple linear regression (MLR) using the selected descriptors. Model accuracy for a given set of predictors was determined by a leave-one-out cross validation (LOOCV) procedure and calculated as  $Q^2_{\text{LOO}}$ . Every GA run was repeated 12 times with different starting seeds. Combinations of predictors with a variance inflation factor (VIF)  $> 5$  (calculated with R-package 'car'), indicating a strong inter-correlation, were effectively removed from the GA population by penalizing the fit during the GA run. GA parameters were optimised using a full factorial experimental design and were found to be population size = 150, cross-over rate = 0.6, mutation rate = 0.3. The maximum number of iterations was set to 200 and elitism set to 8. The number of predictors to be selected during a GA run was fixed and varied between 2 and 7.

The applicability domain (AD) of the models was calculated by means of the William's plot. If the first GA run indicates the presence of outliers with high leverage then all outliers (including high residual outliers) are removed from the dataset and the GA is repeated until no high leverage outliers are detected. High leverage molecules are poorly fitting, highly impactful molecules that force the GA to accommodate them in the model, having great impact on its performance<sup>32</sup>. In our study, after the first GA-run, two molecules, glyceollin I (**87**) and glyceollin III (**89**) (Table S1) were flagged as high leverage molecules. Erysubin F (**55**) was highlighted as a high residual outlier. Erysubin F is reported to have an outstandingly low activity (MIC 100  $\mu\text{g}/\text{mL}$ ) for a double prenylated isoflavone<sup>33</sup>. All three molecules were removed from the dataset, the GA-MLR procedure was repeated once more and a new AD (AD') was defined.

The final dataset used for the development and selection of the best QSAR model comprised 67 prenylated (iso)flavonoids from 9 different (iso)flavonoid subclasses: 2-arylbenzofurans, 3-arylcoumarins, flavanones, isoflavans, isoflavanones, isoflavenes, isoflavones, pterocarpanes and pterocarpenes.

Several internal and external statistical parameters which are typically used to assess the statistical performance of the models were calculated. Some of the internal ones include: the significance of the models and the descriptors individually ( $p$  value  $< 0.05$ ), the coefficient of determination,  $R^2$ , and the adjusted  $R^2$  ( $R^2_{\text{adj}}$ ) which corrects for the difference in the number of descriptors, and the leave-one-out cross-validation coefficient of

determination ( $Q^2_{\text{LOO}}$ ) which is used to assess the model's internal predictivity. According to Tropsha et al. (2010), the  $R^2$  and  $Q^2_{\text{LOO}}$  should be  $> 0.6$  and  $> 0.5$ , respectively<sup>34</sup>. Furthermore, the maximum variance inflation factor ( $\text{VIF}_{\text{max}}$ ), which indicates potential inter-correlation of the descriptors in the generated models, should preferably be  $< 5$ <sup>35</sup>. The accuracy of the prediction of the test set was assessed by calculating the  $Q^2_{\text{F3}}$  for the test set ( $Q^2_{\text{test}}$ ) according to Consonni et al. (2009), a calculation based on the number of training objects<sup>36</sup>. For both sets, the % of prediction error was calculated with the following formula:  $(\text{pMIC}_{\text{observed}} - \text{pMIC}_{\text{predicted}})/\text{pMIC}_{\text{observed}} * 100$ . Last, the best QSAR model was further evaluated with the external validation set; an independent set of 10 compounds with established MICs verified to belong to the applicability domain of the selected model (AD'). The aim of this external validation set was to simulate the scenario where new prenylated (iso)flavonoids with unknown activity are predicted by the model. Based on the small number of datapoints and the narrow range of activities, model's prediction accuracy on the external validation set was assessed by determining the prediction error percentage and the mean absolute errors (MAE). An acceptable prediction error percentage was considered when this was less than 10% which corresponds to maximum one twofold dilution difference between predicted and experimental values<sup>37,38</sup>. MAE was calculated using the following formula:  $1/n * \sum |\text{pMIC}_{\text{observed}} - \text{pMIC}_{\text{predicted}}|$  with  $n$  being the number of external validation objects. If MAE is up to 10% of the training set range, then it is considered as a good prediction whereas if it is more than 15% it is considered a bad one, according to Roy et al. (2016)<sup>39</sup>.

**Prediction of activity of prenylated (iso)flavonoids against other Gram-positive bacteria.** The best QSAR model for MRSA was also used to predict the level of activity (low, moderate, high) of prenylated (iso)flavonoids tested against other Gram-positive bacteria. Thus, the antibacterial activity of 71 prenylated (iso)flavonoids tested against other Gram-positive bacteria was mined from literature (studies from 1988 to 2020). The compounds were gathered under the premise that they fall within the applicability domain of our QSAR-MRSA model. Furthermore, only compounds from the same subclasses as the ones used to construct the QSAR-MRSA were used, to avoid extreme extrapolation of the model. When the compounds were tested against different Gram-positive bacteria or when more than one MIC was available per compound, then the highest MIC was taken into account to consider the worst-case scenario.

The quality of prediction was assessed by verifying that the compounds were still classified as active ( $\text{MIC} \leq 25 \mu\text{g/mL}$ ), moderately active ( $25 < \text{MIC} < 100 \mu\text{g/mL}$ ) or inactive ( $\text{MIC} \geq 100 \mu\text{g/mL}$ ) after the prediction. Based on the nature of the MIC determination method (agar or broth dilution assays), misclassified compounds for which their predicted activity differed from their experimentally measured activity by one two-fold dilution (acceptable experimental error in dilution assays<sup>37,38</sup>), were considered as correctly predicted.

**Pharmacophore elucidation.** A pharmacophore model was built using the pharmacophore elucidation query module of MOE. Since no definite target site of these compounds is known, a ligand-based pharmacophore methodology was employed. This is performed by aligning the different active ligands and determining the essential common chemical queries necessary for activity. All 77 compounds (curated training, curated test and external validation set) were used to extract the pharmacophore features. An activity threshold of  $\text{pMIC} \geq 4.1$  ( $\text{MIC} \leq 25 \mu\text{g/mL}$ ) was set to distinguish the active from the moderate/inactive antibacterials. The quality of the model was assessed based on its capacity to discriminate the active from the moderately active or inactive molecules. The overall accuracy refers to the percentage of correct predictions, including both actives and inactive molecules. The positive accuracy refers to the proportion of correctly predicted actives, indicating the sensitivity of the model. The negative accuracy is calculated based on the percentage of correctly predicted inactives showing the specificity of the model.

## Results

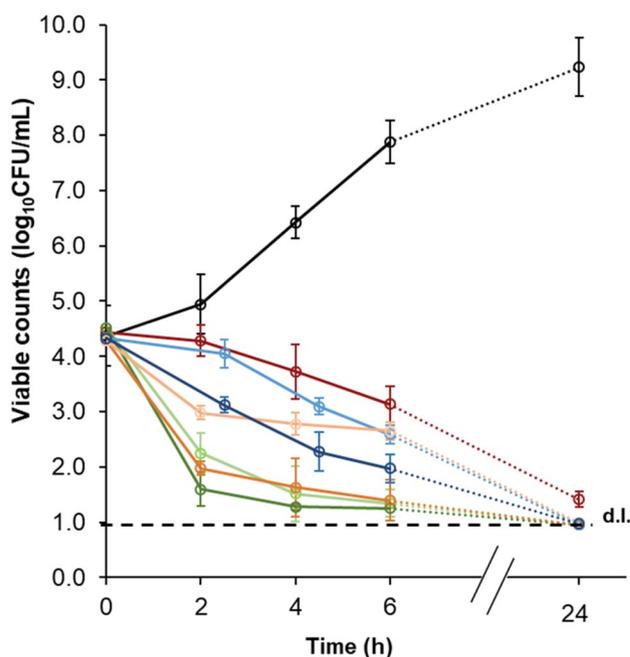
**Experimental anti-MRSA activity of prenylated (iso)flavonoids.** Table 1 shows the antibacterial potency of prenylated (iso)flavonoids against MRSA 18HN. The classification of the compounds with respect to their anti-MRSA activity was based on literature of antimicrobial phytochemicals<sup>40</sup>. Compounds with MIC values  $\leq 25 \mu\text{g/mL}$  were considered active, with most active being the ones having a MIC of  $\leq 10 \mu\text{g/mL}$ . Compounds with MIC values between 25 and  $100 \mu\text{g/mL}$  were considered moderately active, whereas those with MIC values  $\geq 100 \mu\text{g/mL}$  were classified as inactive. The most active compounds experimentally tested in this study were the isoflavone, 6,8-diprenylgenistein (**50**) and the di-prenylated flavanone, glabrol (**13**) which showed a MIC of  $9 \mu\text{g/mL}$ , corresponding to  $23 \mu\text{M}$  and  $24 \mu\text{M}$ , respectively. These compounds share double chain prenylation in their backbones (Fig. S2). The most active mono-prenylated compound was the ring-prenylated isoflavan, 4'-O-methylglabridin (**17**) with a MIC of  $10 \mu\text{g/mL}$  ( $30 \mu\text{M}$ ). Inactive prenylated (iso)flavonoids with an established MIC tested were glyceollins I–III (**87–89**) ( $> 100 \mu\text{g/mL}$ ,  $> 296 \mu\text{M}$ ). The mono-prenylated compounds (**51**, **63**, **68**, **76**, **85**, **91**, **101**) and the di-prenylated hispaglabridin B (**26**) had negligible antibacterial activity against MRSA (TTD similar to that of the control at  $100 \mu\text{g/mL}$ , therefore  $\text{MIC} \gg 100 \mu\text{g/mL}$ ) (Table S1).

Figure 1 shows the inactivation kinetics of MRSA 18HN in the presence of the top three antibacterial prenylated (iso)flavonoids tested in this study. The di-prenylated (iso)flavonoids, 6,8-diprenylgenistein (**50**) and glabrol (**13**) inactivated MRSA by  $> 99\%$  in the first 2 h of contact at  $2 \times$  their MIC, whereas the mono-prenylated isoflavan, 4'-O-methylglabridin (**17**) inactivated MRSA after 6 h also at  $2 \times$  its MIC. The control antibiotic vancomycin at its MIC decreased the initial inoculum by  $> 99\%$  only after 24 h. The slow action of vancomycin against MRSA is in line with what has been shown before for MRSA and *Streptococcus pneumoniae*<sup>41,42</sup>.

**Extended prenylated (iso)flavonoid dataset for QSAR modelling of anti-MRSA activity.** A QSAR study was performed to pinpoint overall molecular properties contributing to the anti-MRSA activity. To

Subclass	Name	MIC $\mu\text{g/mL}$ [ $\mu\text{M}$ ]	MBC $\mu\text{g/mL}$ [ $\mu\text{M}$ ]
Flavanones	6-Prenylnaringenin (12)	38 [110]	44 [129]
	Glabrol (13)	9 [24]	19 [48]
	Sophoraflavanone B (14)	22 [64]	25 [73]
Isoflavans	3'-OH-4'-O-Methylglabridin (16)	16 [44]	26 [73]
	4'-O-Methylglabridin (17)	10 [30]	23 [66]
	Glabridin (22)	13 [39]	19 [58]
	Hispaglabridin A (25)	44 [111]	44 [111]
Isoflavones	6,8-Diprenylgenistein (50)	9 [23]	16 [38]
	Isowighteone (64)	22 [65]	34 [102]
	Licoisoflavone A (65)	25 [71]	50 [141]
	Luteone (69)	25 [71]	44 [123]
	Neobavaisoflavone (70)	38 [116]	50 [155]
	Wighteone (74)	16 [46]	22 [65]
Isoflavene	Glabrene (49)	25 [78]	44 [136]
Pterocarpans	Glyceollidin II (86)	44 [129]	44 [129]
	Glyceollin I (87)	100 [296]	n.a
	Glyceollin II (88)	150 [443]	n.a
	Glyceollin III (89)	100 [296]	n.a
	Glyceollin IV (90)	44 [123]	75 [212]
Pterocarpenes	Dehydroglyceollidin II (98)	22 [68]	22 [68]
	Dehydroglyceollin I (99)	16 [49]	22 [68]
	Dehydroglyceollin II (100)	19 [59]	19 [59]
	Dehydroglyceollin IV (102)	44 [130]	50 [149]

**Table 1.** Minimum inhibitory concentration (MIC) and minimum bactericidal concentration (MBC) of prenylated (iso)flavonoids tested in this study against MRSA 18HN.



**Figure 1.** MRSA 18HN inactivation kinetics in the presence of the three most antibacterial prenylated (iso) flavonoids tested in this study, at their MIC (light shade) and 2xMIC (dark shade). Control (black), vancomycin (2  $\mu\text{g/mL}$ , red), 4'-O-methylglabridin (17) (blue), glabrol (13) (orange) and 6,8-diprenylgenistein (50) (green).

n	Equation	R <sup>2</sup>	R <sup>2</sup> <sub>adj</sub>	Q <sup>2</sup> <sub>LOO</sub>	Q <sup>2</sup> <sub>test</sub>
2	$y = 2.787(0.331) - 0.159(0.048)*h\_emd\_C + 0.006(0.001)*vsurf\_D4$	0.56	0.54	0.51	0.73
3	$y = 1.885(0.305) - 0.701(0.183)*h\_pavgQ + 0.006(0.001)*vsurf\_D4 - 0.054(0.018)*vsurf\_IW7$	0.60	0.57	0.53	0.78
4	$y = 2.551(0.343) - 0.178(0.045)*h\_emd\_C + 0.006(0.001)*vsurf\_D4 - 0.050(0.017)*vsurf\_IW7 + 0.047(0.016)*E\_vdw$	0.64	0.61	0.57	0.75
5	$y = 6.240(0.270) + 0.030(0.006)*PEOE\_VSA\_PPOS - 2.521(0.275)*vsurf\_CW3 - 0.061(0.020)*vsurf\_IW7 - 0.839(0.238)*h\_pavgQ + 0.011(0.006)*PEOE\_VSA + 2$	0.68	0.64	0.60	0.64
6	$y = 6.207(0.262) + 0.027(0.006)*PEOE\_VSA\_PPOS - 2.558(0.267)*vsurf\_CW3 - 0.063(0.019)*vsurf\_IW7 - 0.888(0.214)*h\_pavgQ + 0.011(0.005)*PEOE\_VSA + 2 + 0.025(0.012)*vsurf\_DD12$	0.69	0.65	0.61	0.65
7	$y = 3.530(1.470) + 0.030(0.006)*PEOE\_VSA\_PPOS - 2.690(0.270)*vsurf\_CW3 - 0.077(0.020)*vsurf\_IW7 - 0.878(0.208)*h\_pavgQ + 0.017(0.006)*PEOE\_VSA + 2 + 0.030(0.012)*vsurf\_DD12 - 0.317(0.171)*PM3\_IP$	0.71	0.67	0.60	0.67

**Table 2.** Best GA-MLR models for predicting the antibacterial activity of prenylated (iso)flavonoids against MRSA. n: number of descriptors; R<sup>2</sup>: coefficient of determination (>0.6)<sup>31</sup>; R<sup>2</sup><sub>adj</sub>: adjusted R<sup>2</sup>; Q<sup>2</sup><sub>LOO</sub>: leave-one-out cross-validated coefficient of determination (>0.5)<sup>31</sup>; Q<sup>2</sup><sub>test</sub>: correlation coefficient for the test set. All models were significant (*p* values < 1.0e<sup>-7</sup>) and no multicollinearity of descriptors was observed (VIF<sub>max</sub> < 3.2)<sup>32</sup>. The F statistics for all models were > 17 with the largest *p* value being 2.1e<sup>-10</sup>. Values in parentheses refer to the standard error of the regression coefficients of the models.

construct robust and reliable QSAR models, the experimental data shown above was combined with literature data to enlarge and diversify our collection of prenylated (iso)flavonoids. In total, 76 prenylated (iso)flavonoids with established MICs from 10 subclasses were collected (Table S1). This diverse set of compounds was curated (see criteria 1 and 2 in **Dataset construction**), which resulted in 70 molecules in the modelling set (Fig. S1). During the training phase, genetic algorithm (GA) flagged glyceollin I (**87**), glyceollin III (**89**) and erysubin F (**55**) as outliers. Thus, QSAR modelling was performed based on a curated modelling set 67 of molecules from 9 subclasses. The best model was further externally validated with a set of 10 molecules from three studies independent from those used in the modelling set<sup>28</sup>.

Similar to what was observed for our in-house data, more than 90% of the di-prenylated (iso)flavonoids compiled from literature, showed a high anti-MRSA activity (MIC ≤ 25 µg/mL, 8–64 µM), whereas half of the mono-prenylated ones were active. Structures of representative active anti-MRSA agents per isoflavonoid subclass derived from the extended dataset can be found in Fig. S3.

The most powerful anti-MRSA di-prenylated (iso)flavonoids were isolupalbigenin, erybraedin A and eryvarin W (**62**, **78**, and **106**, Table S1) with MIC = 3 µg/mL, 8 µM. These molecules have one prenyl-group on the A-ring and one on the B-ring (Fig. S3). The most active mono-prenylated isoflavonoid was orientanol B (**93**) with MIC = 6 µg/mL, 18 µM (Table S1). This molecule showed comparable activity to the in-house tested mono-prenylated isoflavan, 4'-O-methyl-glabridin (**17**) with MIC = 10 µg/mL, 30 µM. Both compounds are prenylated on the A-ring and possess the same type and amount of substituents (one hydroxy-group and one methoxy-group), yet in different positions over the backbone (Fig. S3).

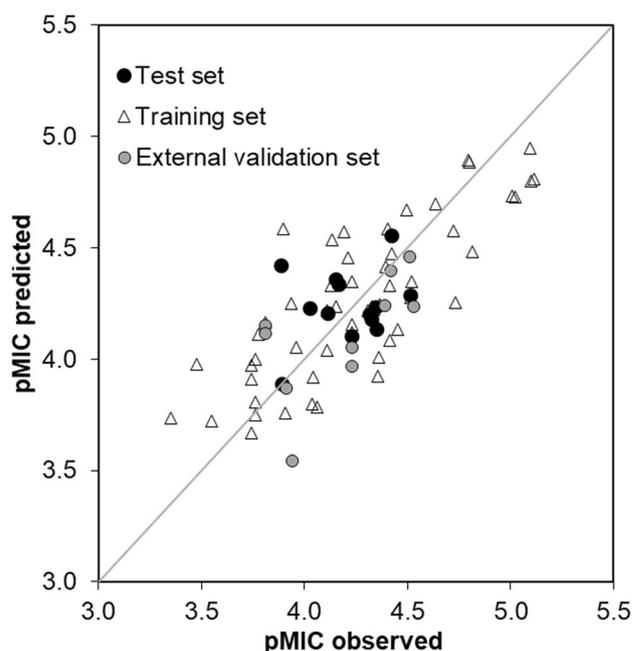
**QSAR model development, selection and external validation.** The best GA-MLR-models obtained per number of descriptors and their statistical performance are listed in Table 2.

Models with four or more descriptors complied with the thresholds of all validation parameters of the QSAR models, whereas models with less descriptors (2–3) gave relatively poor adjusted R<sup>2</sup>s (R<sup>2</sup><sub>adj</sub>). Usually, a higher number of descriptors results in over-fitted/more complex, less interpretable models. Thus, we chose the four-descriptor model as the best, since it balances well statistical validity, predictivity of the training (Q<sup>2</sup><sub>LOO</sub> 0.57 and average prediction error of 4%, Table S1) and the test set (Q<sup>2</sup><sub>test</sub> 0.75 and average prediction error of 5%, Table S1) (Fig. 2) and model's interpretability. The model was also externally validated by predicting an independent set of 10 compounds (**1**, **19**, **28–30**, **36**, **37**, **44**, **46**, **47**) coming from three studies<sup>25–27</sup> and low average prediction errors (5%) were obtained (Table S1). The mean average error (MAE) of the external validation set was calculated to be 11.5% being marginally above the 10.0% threshold for good predictions reported by Roy et al. (2016)<sup>39</sup>. The applicability domain of the best 4-descriptor model can be found in Fig. S4.

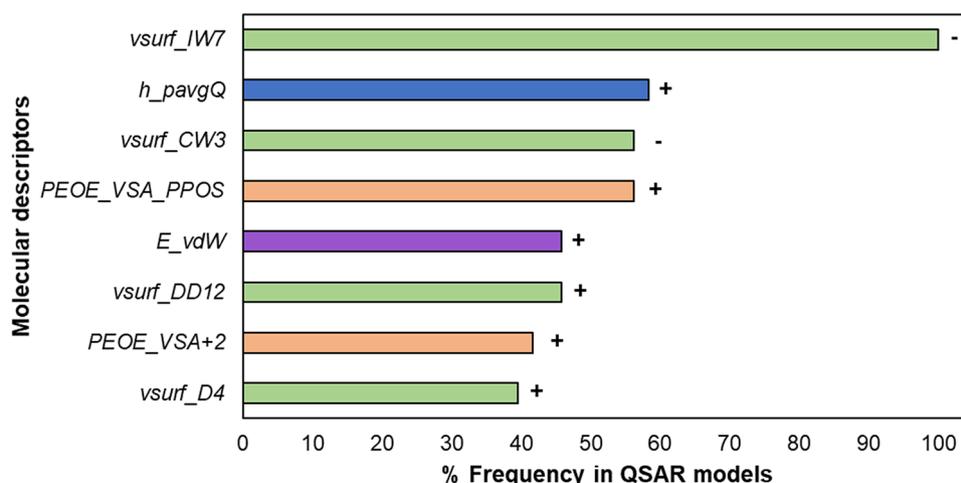
In the chosen four-descriptor model, the hydrophobic volume at the 4th energetic level, i.e. – 0.8 kcal/mol (*vsurf\_D4*) was the most significant descriptor (*p* value 2e<sup>-10</sup>), followed by the sum of hydrogen bond donor strengths of carbon atoms (*h\_emd\_C*, *p* value 2e<sup>-4</sup>). In addition, the unbalance between the centre of mass of a molecule and the position of the hydrophilic regions around it (*vsurf\_IW7*) and the van der Waals interaction energy (*E\_vdw*) were the least significantly correlated descriptors (*p* value 6e<sup>-3</sup>). Definitions of the descriptors can be found in Table S2.

**Predicting the level of antibacterial activity of prenylated (iso)flavonoids against other Gram-positive bacteria.** The QSAR model developed for MRSA was additionally assessed for its capacity to predict the level of activity (low, moderate, high) of prenylated (iso)flavonoids tested against other Gram-positive bacteria (Table S3).

Thus, activity data (MIC values) of 71 prenylated (iso)flavonoids tested against SA, *Staphylococcus epidermis*, *Bacillus subtilis*, *Enterococcus faecalis* and *Streptococcus mutans* was collected (Table S3 and Fig. S5). The model predicted correctly the level of activity of 73% of the compounds (Fig. S6). This shows that the model is robust enough to be used as a tool for quick screening for potent antibacterial prenylated (iso)flavonoids. Half of the incorrectly predicted molecules were flavanones (Table S3 and Fig. S5), possibly due to the limited structural variation of flavanones in the training and test set (Table S1) used to develop the model.



**Figure 2.** Correlation of the experimental and predicted anti-MRSA activity (pMIC, M) of prenylated (iso) flavonoids by using our best 4-descriptor QSAR model for the training, the test and the external validation sets. The applicability domain of the QSAR model can be found in Fig. S4.



**Figure 3.** Descriptors most frequently selected by the GA to predict the anti-MRSA activity. The positive (+) or negative (-) correlation of each descriptor with anti-MRSA activity is shown next to the bars. Colour coding is used to indicate the different descriptor families. Green indicates the surface area, volume and shape (*vsurf*) descriptors, blue indicates the Hückel-theory (*h*) descriptor, orange represents the partial charge descriptors (*PEOE\_VSA*) and purple is used to indicate the energy descriptor. Definitions of the descriptors can be found in Table S2.

**Interpretation of most frequently used descriptors in statistically compliant models.** To obtain a deeper insight into the molecular properties essential for activity, we analysed the most frequently used descriptors, i.e. present in more than 40% of the statistically compliant models ( $R^2_{adj} > 0.6$  and  $Q^2_{LOO} > 0.5$ )<sup>34</sup> (Fig. 3).

*Hydrophobic volume and balanced hydrophilic fraction favour anti-MRSA activity.* The most frequently used descriptor family included the surface area, volume and shape descriptors, *vsurf* (green colour in Fig. 3). These descriptors were calculated on multiple energetic levels, as volume and shape are influenced by the energy of the molecules<sup>43</sup>.

The descriptor *vsurf\_IW7* was present in all statistically compliant models. It represents the hydrophilic integrity moment at  $-5.0$  kcal/mol, an energy level known to be representative for polar and H-bond donor and acceptor regions<sup>43</sup>. It shows the distribution of hydrophilic moieties over the molecules by measuring the unbalance between the hydrophilic moieties and the centre of molecular mass. Its negative correlation to the anti-MRSA activity suggests that the hydrophilic regions close to the centre of mass or balanced at opposite directions in the molecule favour the activity<sup>44</sup>.

The descriptor *vsurf\_CW3* quantifies the ratio between the hydrophilic surface and the total molecular surface<sup>43</sup> and was one of the most strongly correlated descriptors ( $4.7e^{-13}$ – $4.1e^{-09}$ , Table S4). Its negative contribution to the anti-MRSA activity indicates that a small hydrophilic fraction favours anti-MRSA activity. All molecules with *vsurf\_CW3* of  $<1.043 \text{ \AA}^3$  possessed high activity (except dehydroglyceollin IV (**102**) and hispaglabridin A (**25**)).

The *vsurf\_DD12* descriptor quantifies the contact distance of lowest and the 2nd lowest hydrophobic energy points of a molecule. Its positive, although weak contribution to activity (*p* value,  $4.7e^{-02}$ – $1.5e^{-01}$ , Table S4) implies that the hydrophobically interacting groups or regions should be far away from each other and not localized closely together in the molecule<sup>45</sup>.

Last, the *vsurf\_D4* descriptor is a measure of the hydrophobic volume at an energy level of  $-0.8$  kcal/mol (known to account for polarizability and dispersion forces)<sup>43</sup>. This descriptor was positively correlated to activity indicating that larger hydrophobic surfaces are preferred.

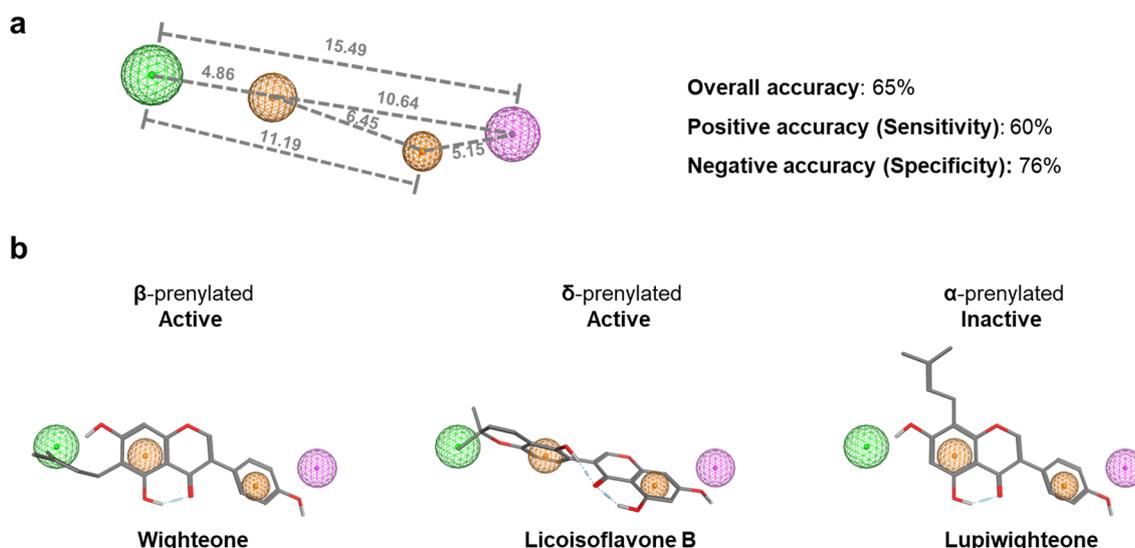
*(Partial) charges favour anti-MRSA activity of prenylated (iso)flavonoids.* Formal positive charge. The second most frequently used, positively correlated descriptor in the QSAR models was the Hückel theory descriptor, *h\_pavgQ* (blue colour in Fig. 3). Hückel theory descriptors take into account local resonance and electron withdrawing effects<sup>46</sup>. This descriptor refers to the average total (formal) charge sum across protonation states at pH 7<sup>47</sup>, calculated based on the relative concentration of various protonation states of the molecule<sup>48</sup>. Isoflavones, (iso)flavanones and 3-aryl coumarins had generally higher values for this descriptor. Possibly, the presence of the carbonyl group on the C-ring of these subclasses enhances the acidity of hydroxyl protons (at C7 and C5) due to stabilization of the negative charge by resonance. In contrast, pterocarpanes, pterocarpenes and isoflavans, lacking this carbonyl group, had low *h\_pavgQ* values. Examples of the correlation of *h\_pavgQ* with the activity can be found in Fig. S7.

*Partial positive charge.* The 2D partial charge descriptors, *PEOE\_VSA\_PPOS* and *PEOE\_VSA + 2* were frequently used, positively correlated variables in the QSAR models (orange in Fig. 3). *PEOE\_VSA* descriptors are calculated based on the Partial Equalization of Orbital Electronegativities method and capture electrostatic interactions<sup>49</sup>. *PEOE\_VSA\_PPOS* (*p* value  $3.6e^{-06}$  to  $1.4e^{-05}$ , Table S4) refers to the total positive polar van der Waals surface area (in  $\text{\AA}^2$ ), where the partial charge is greater than  $+0.200e$ . In fact, this descriptor depends on the number of oxygen atoms within the molecule<sup>50</sup>. Isoflavones, isoflavanones and 3-aryl coumarins were the most oxygenated subclasses in the dataset and showed large, partially positive surface areas within the molecule (*PEOE\_VSA\_PPOS*  $>35 \text{ \AA}^2$ ). *PEOE\_VSA + 2* (*p* value  $4.2e^{-03}$  to  $1.6e^{-01}$ , Table S4) refers to the sum of van der Waals (vdw) surface area for each atom (in  $\text{\AA}^2$ ) of which the partial charge is between  $+0.100e$  and (including)  $+0.149e$ . Based on the level of significance of these two *PEOE* descriptors, small partial positive charges are less determinant of anti-MRSA activity compared to larger partial positive charges.

*Van der Waals surface energy.* The energy descriptor *E\_vdw* (purple colour in Fig. 3) was also used as a positively correlated explanatory variable. This descriptor represents the van der Waals (vdW) component of the potential energy of the molecules<sup>51</sup>. Pterocarpanes, pterocarpenes and 2-arylbenzofurans had clearly lower vdW interaction energies (*E\_vdW*  $<11.1$  kcal/mol) than the rest of the subclasses. Pterocarpanes and pterocarpenes are more rigid molecules than the other (iso)flavonoid subclasses due to the presence of an additional D-ring<sup>18</sup>. This higher rigidity should facilitate vdW interactions, due to lower entropic penalties on the rigid molecules compared to more flexible ones<sup>52</sup>. 2-Arylbenzofurans, despite the lack of the extra D-ring, contain a furan C-ring, which contributes to increased molecular planarity compared to a pyran<sup>53</sup> (Fig. S8). Thus, the more rigid or planar a molecule is, the lower the vdW interaction energy and the lower the activity. This implies that other interactions, most likely electrostatic, are more important for activity.

**Pharmacophore modelling.** To understand further the effect of prenylation and of other substituents on the antibacterial activity against MRSA, a ligand-based pharmacophore model was developed. The analysis of the fit of the compounds in the 3D-pharmacophore model revealed four structural features mapping the active molecules; two aromatic ring features (orange) corresponding to the A- and B-ring of (iso)flavonoids, one hydrophobic feature representing the prenyl group (green) and a hydrogen donor projection feature referring to the position of a potential hydrogen bond partner (pink) (Fig. 4a). The pharmacophore model constructed for MRSA had a 65% overall accuracy (sum of molecules predicted correctly), a 60% positive accuracy (active molecules predicted correctly) and a 76% negative accuracy (inactive molecules predicted correctly). These accuracies are 20% higher than the accuracies derived when using a previously constructed pharmacophore model on *L. monocytogenes*<sup>18</sup>. The hits in pharmacophore search and the quality of fitting of each molecule to the model for MRSA and *L. monocytogenes* are shown in Tables S5 and S6, respectively.

The model predicted correctly all the active prenylated isoflavones, the most abundant subclass of the dataset, 70% of the active isoflavans, 50% of the active pterocarpanes and 43% of the active pterocarpenes. Figure 4b illustrates examples of isoflavones prenylated at different positions with respect to their fitting quality into the pharmacophore model. Prenylated isoflavones on the  $\beta$ -position (A-ring) and  $\delta$ -position (B-ring) of the



**Figure 4.** Ligand-based pharmacophore model for active anti-MRSA ( $MIC \leq 25 \mu\text{g/mL}$ ) prenylated (iso)flavonoids. The colour of the spheres represents the following features: orange spheres represent aromatic rings, the green and the pink spheres represent hydrophobic features and hydrogen bond donor projections (i.e. features that denote the presence of possible hydrogen bond partners), respectively. Numbers represent the distance between the features in Ångström. The donor/acceptor projections have a radius of 1.2 Å, the aromatic features have a radius of 0.8 Å and 1.0 Å, left and right respectively and the hydrophobic feature has a radius of 1.2 Å. Percentages correspond to the different prediction accuracies (overall, positive and negative) of the model (a). The fitting of molecules in the pharmacophore with respect to their potency are also illustrated (b).

isoflavonoid backbone map all the essential pharmacophore features. In contrast,  $\alpha$ -prenylation failed to map the hydrophobic moiety (A-ring).

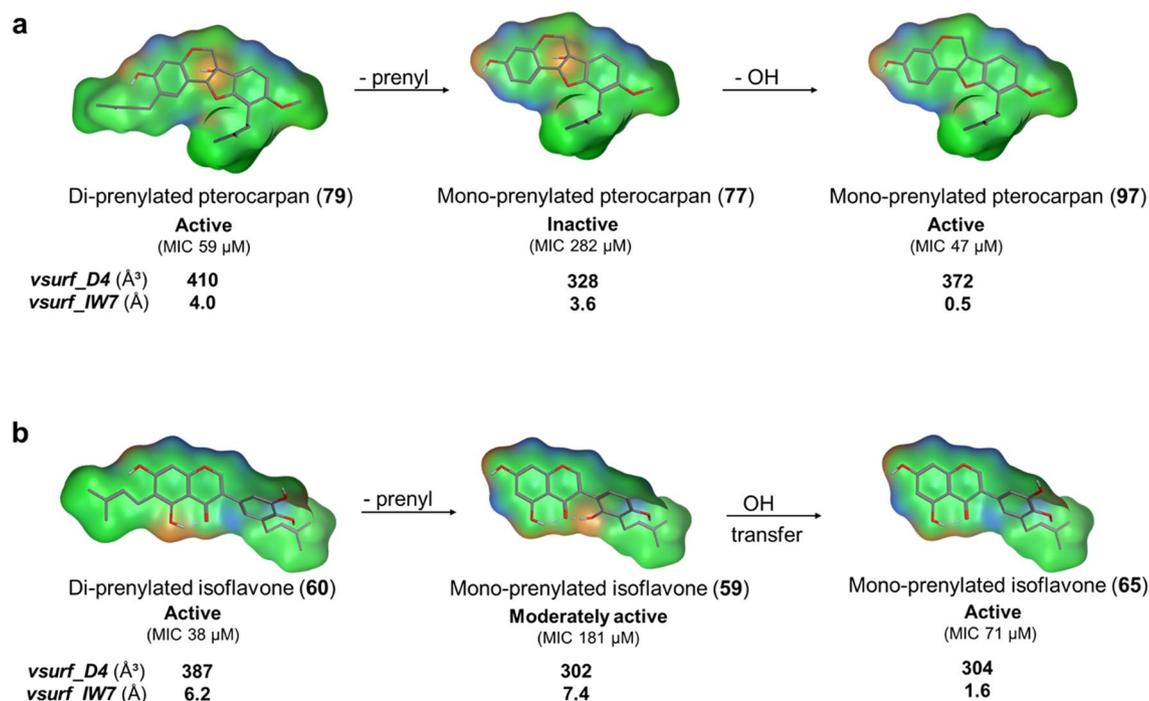
However, the model failed to predict the active di-prenylated isoflavanones (only the di-prenylated members of this subclass were active), while it correctly predicted analogue molecules such as di-prenylated isoflavones and di-prenylated isoflavans. The simultaneous absence of a double bond in the C-ring and the presence of a carbonyl group in isoflavanones, contrary to isoflavones and isoflavans, respectively, seems to significantly affect the orientation of the B-ring (Fig. S8). This different B-ring orientation could be the reason for the lack of fit of isoflavanones in the pharmacophore. Five molecules (8, 25, 29, 72 and 77, Table S1) were found to be false positives during the pharmacophore search (Fig. S9); These molecules are typical examples of activity cliffs, i.e. structurally similar molecules with large difference in potency<sup>54</sup>.

## Discussion

**Hydrophobicity and electrostatic interactions are the main determinants for the anti-MRSA activity of prenylated (iso)flavonoids.** Traditionally, the antimicrobial potency of prenylated (iso)flavonoids was attributed to their increased hydrophobicity mainly due to the presence of the prenyl group<sup>11,55</sup>. Generally, the amphiphilic cytoplasmic membrane is hypothesized as the first target of potential antibacterials<sup>56</sup>. Based on the ability of other antibiotics (of similar or larger size than prenylated (iso)flavonoids) to penetrate the bacterial cell wall, it is expected that prenylated (iso)flavonoids can also cross the open structure of the cell wall<sup>57</sup> and access the membrane<sup>58</sup>. Nonetheless, possible interactions of prenylated (iso)flavonoids with the cell wall cannot be completely excluded.

In this study, continuous hydrophobicity with balanced hydrophilic groups in the form of the descriptors *vsurf\_IW7*, *vsurf\_D4* and *vsurf\_CW3*, were the most strongly correlated properties to anti-MRSA activity, in accordance with previous results against the Gram-positive *L. monocytogenes*<sup>18</sup>. The contribution of the hydrophobic volume (*vsurf\_D4*) and hydrophilic integrity moment (*vsurf\_IW7*) to the anti-MRSA activity is illustrated in Fig. 5. Di-prenylated (iso)flavonoids are characterized by an extensive hydrophobic volume, as in molecules (79) and (60). In most cases, extensive hydrophobic volume accounts for the increased antimicrobial potency of the molecules<sup>18</sup>. Decreasing their hydrophobic volume by removal of one prenyl-group, as in molecules (77) and (59) indeed decreased the anti-MRSA activity. However, further removal of the C6a hydroxyl-group (molecule (77) to (97)) (Fig. 5a) restored the hydrophobic volume to a certain extent, reduced the hydrophilic moments in the molecule (*vsurf\_IW7*), leading to higher anti-MRSA activity.

Interestingly, transferring the OH group from C5' to C6' (molecule (59) to (65)) (Fig. 5b) restored the balance between the hydrophilic moments in the molecule, enhancing the anti-MRSA potency, even without influencing the hydrophobic volume. In this respect, isolupalbigenin (62) together with erybraedin A (78) and eryvarin W (106) were highlighted as the most active (di-prenylated) compounds of the dataset ("Extended prenylated (iso)flavonoid dataset for QSAR modelling of anti-MRSA activity" section). Isolupalbigenin (62) has relatively unfavorable distribution of hydrophilic groups (*vsurf\_IW7* = 4.5 Å) compared to the other two molecules (Fig. S3), both having *vsurf\_IW7* = 0.0 Å). By changing the position of the OH group from the C5 of the A-ring to the C6'



**Figure 5.** Surface maps of a series of active, moderately active and inactive prenylated pterocarpan (a) and isoflavones (b). Isoflavonoids from the same subclass (i.e. a or b) are analogues, differing only in the presence, absence or relocation of one functional group (indicated by the arrows). These structural differences influence their hydrophobic volume ( $vsurf\_D4$ ,  $\text{\AA}^3$ ) and the distribution of the hydrophilic regions ( $vsurf\_IW7$ ,  $\text{\AA}$ ) and therefore the anti-MRSA activity. Mildly polar regions are coloured in blue, H-bonding in orange and hydrophobic regions in green.

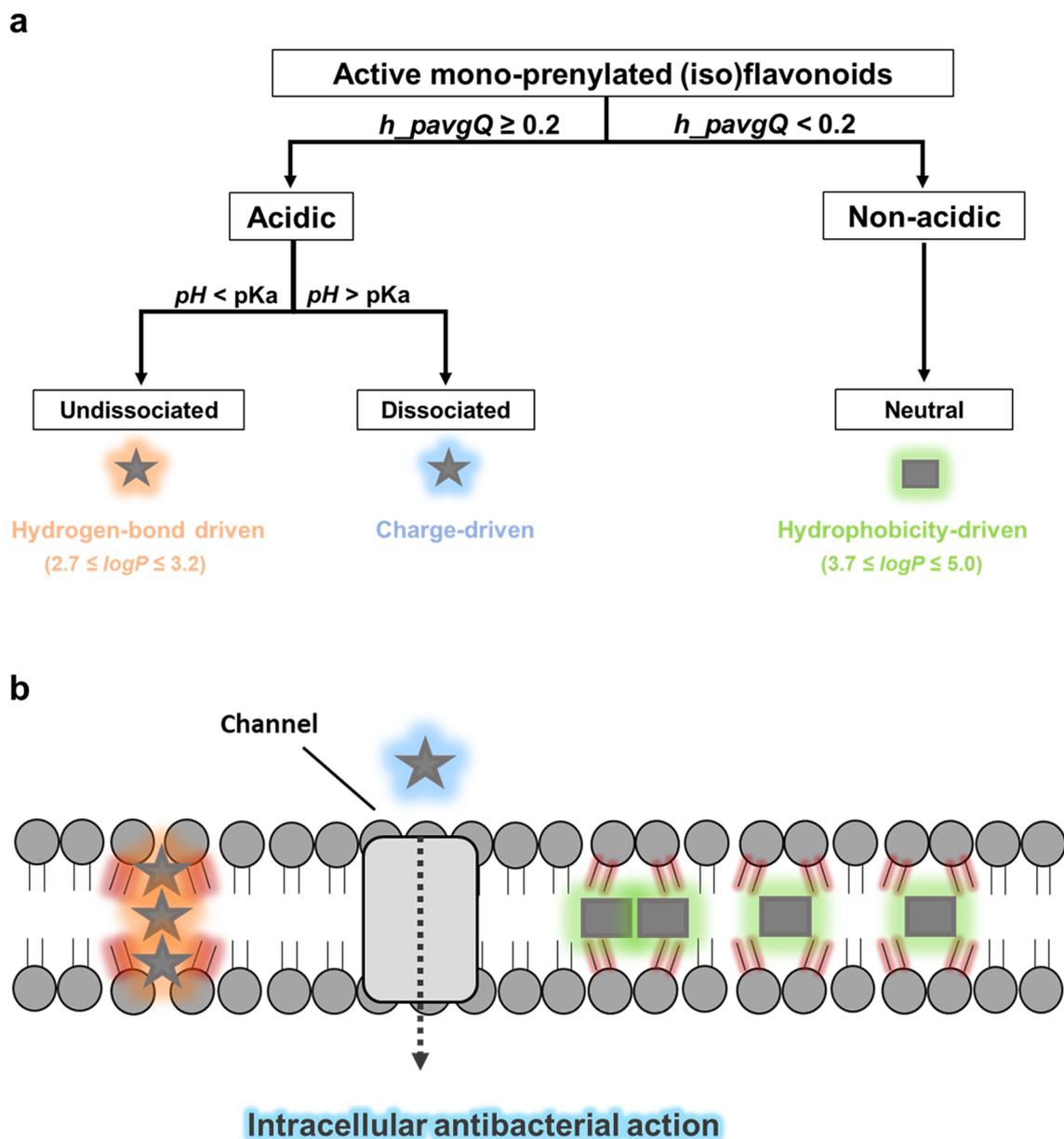
of the B-ring (Fig. S3), the new molecule has optimal distribution of the hydrophilic groups ( $vsurf\_IW7=0.0$   $\text{\AA}$ ) and was predicted to have the highest activity from all molecules of the dataset.

From the above it becomes also evident that prenylated (iso)flavonoids scoring low in hydrophobicity can still be active anti-MRSA agents (e.g. molecule 65, Fig. 5b). This observation agrees with the findings of our pharmacophore model, where the importance of single, and not necessarily double, prenylation together with a hydrogen bond acceptor were highlighted. Charge was highlighted as the second most important molecular property for the antibacterial activity of prenylated (iso)flavonoids. Large positive partial charges ( $PEOE\_VSA\_PPOS$  and  $PEOE\_VSA + 2$ ) and formal negative charges ( $h\_pavgQ$ ) suggest the importance of electrostatic interactions<sup>49</sup>. Sadgrove et al. (2020) also associated the importance of electrostatic interactions such as hydrogen-bonding with anti-(MR)SA activity<sup>21</sup>.

**New insights in the possible interaction of active anti-MRSA agents with the cytoplasmic membrane.** Based on the main molecular properties highlighted by the QSAR study and the pharmacophore model, it is hypothesized that different active mono-prenylated anti-MRSA (iso)flavonoids might be differently taken up by the cell or interact with the membrane. Three ways of interaction of prenylated (iso)flavonoids with the cytoplasmic membrane are proposed (Fig. 6).

*Acidic prenylated isoflavonoids might interact with both the membrane and with intracellular targets.* First, prenylated (iso)flavonoids with high potential to be charged at pH 7 are hypothesized to interact differently from the ones that remain neutral across the pH range. The average total formal charge sum across protonation states at pH 7<sup>47</sup> is represented by the descriptor,  $h\_pavgQ$  (Table S2). Twenty percent of the active mono-prenylated (iso)flavonoids had an absolute  $h\_pavgQ$  value higher than 0.2. These molecules were considered acidic and both their dissociated and undissociated form were taken into account (Fig. 6a, left side). The subclass of isoflavones is the predominant representative subclass (83%) of acidic, active mono-prenylated (iso)flavonoids. All, but one, molecules of this category in our dataset are characterized by the presence of three or more free hydroxyl groups. Prenylated (iso)flavones have been shown experimentally to permeabilize the membrane of MRSA and other Gram-positive bacteria<sup>18,59,60</sup>. Yet, the exact mechanism of membrane permeabilization has not been thoroughly investigated.

Recently, Li et al. (2018) showed that a very active (MIC 1  $\mu\text{g/mL}$ , 2  $\mu\text{M}$ ) di-prenylated xanthone ( $\alpha$ -mangostin) forms membrane-spanning, intermolecular aggregates primarily through hydrogen bonding<sup>58</sup>. This was shown through molecular dynamic simulation experiments. Aggregate formation ultimately led to membrane destabilization and subsequent water translocation across the membrane, without pore formation<sup>58</sup>. This is in line with the membrane permeabilization of Gram-positive bacteria observed by prenylated isoflavones<sup>18,60</sup>.



**Figure 6.** Proposed classification of active ( $MIC \leq 25 \mu\text{g/mL}$ ) prenylated (iso)flavonoids based on their potential to be negatively charged at pH 7. Molecules with high potential to be dissociated at pH 7 ( $h_{pavgQ} \geq 0.2$ ), i.e. acidic, were considered for their dissociated and undissociated forms separately. Molecules that remain fully undissociated at pH 7 ( $h_{pavgQ} < 0.2$ ), i.e. non-acidic, were considered neutral (a). Hypothesized interactions of active anti-MRSA prenylated (iso)flavonoids with the cytoplasmic membrane; possible interactions of prenylated (iso)flavonoids with the cell wall are not considered in this study (b). Stars represent active acidic mono-prenylated (iso)flavonoids, whereas rectangular shapes represent active, neutral mono-prenylated (iso)flavonoids. Different glows around the shapes indicate hypothesized, predominant driving forces that mediate cellular uptake or interaction with the membrane; blue denotes a negative charge, orange indicates strong hydrogen bonding and green glow represents hydrophobicity-driven interaction with the lipid bilayer. The intensity of red shading around phospholipids shows the level of compromise of membrane integrity.

Interestingly, it was specifically highlighted that the presence of three hydroxyl groups in alpha-mangostin are crucial for transmembrane aggregate formation, while their absence would make alpha-mangostin soluble in the lipid bilayer<sup>58</sup>. The formation of intermolecular hydrogen bonds possibly facilitates the unfavourable presence of polar groups into the hydrophobic interior of the membrane, stabilizing the cluster. Alpha-mangostin is structurally similar to the active isoflavones in our dataset, e.g. a mono-prenylated (wighteone, 74) and a di-prenylated one (lupalbigenin, 67), although the xanthone is more planar<sup>61</sup> (Fig. S10). The hydrogen bond strengths of mangostin and the two analogue isoflavones were quantified ( $h_{ema}$  and  $h_{emd}$  descriptors, Table S2, Molecular Operating Environment, MOE). The sum of hydrogen bond donor (HBD) and acceptor (HBA) strengths for these molecules were comparable (HBD: 10.2 kcal/mol for wighteone, 10.9 kcal/mol for lupalbigenin and

11.4 kcal/mol for mangostin; HBA: 4.6 kcal/mol for the two prenylated isoflavones and 5.6 kcal/mol for the xanthone). Since aggregate formation by the di-prenylated xanthone was shown to be primarily hydrogen-bond driven, possibly this mechanism of action is also employed by prenylated isoflavones given the similar degree of hydrogen-bond capacity.

Xanthone transmembrane aggregates were formed after a high enough concentration of the molecule reaches the lipid tail region of the membrane. Although wightone (74) and lupalbigenin (67) have similar hydrogen-bond capacities, the presence of only one prenyl group in the former should increase the free energy barrier of membrane penetration compared to the latter<sup>41</sup>. Possibly, higher concentrations are needed for mono-prenylated isoflavones to form these aggregates than for their di-prenylated counterparts, something which is also reflected in the higher MICs observed for the former.

The mechanism of membrane activity proposed by Li et al. (2018) considered only the undissociated form of the molecule<sup>58</sup>. A negative charge in the dissociated acidic isoflavones would be repelled by the polar (phosphate) head groups of the membrane and would not allow membrane penetration<sup>58,62</sup>. Instead, the dissociated acidic isoflavones might be transported to the cytosol through transmembrane carrier proteins or by active transport<sup>63,64</sup> (Fig. 6b, blue star) similar to what has been shown for other anionic antibiotics, such as carbenicillin and quinolones<sup>65</sup>. The proposed membrane transportation of these molecules implies that the membrane is not the only target for their anti-MRSA activity and their mode of action is further complemented by activity inside the cytosol. Recently, a di-prenylated flavone was shown to inhibit the biosynthesis of phosphatidic acids in the cytosol, the repair mechanism of bacterial membranes<sup>59</sup> after membrane disruption. Noticeable inhibition (>99%) of MRSA was observed within 2 h of exposure to this prenylated flavonoid<sup>59</sup>, similar to what was found in Fig. 1.

It is therefore proposed that the dissociated and undissociated forms of antibacterial acidic (iso)flavonoids might employ different modes of action (Fig. 6b, orange and blue stars). Internalization of antibacterials in different ways depending on their protonation state has been shown before for Gram-negatives<sup>66</sup>.

*Non-acidic prenylated isoflavonoids might be internalized by the membrane via diffusion.* Hydrogen-bonding, formal charge and hydrophobicity were the main molecular properties determining anti-MRSA activity. It was hypothesized that acidic prenylated isoflavonoids may interact with both the membrane and with intracellular targets, depending on their protonation state. Nevertheless, 80% of the active mono-prenylated isoflavonoids had an absolute  $h\_pavq$  value of less than 0.2, meaning that these molecules remain neutral at the pH of the medium. This category of non-acidic prenylated (iso)flavonoids (Fig. 6a, right side) comprised compounds with less than 3 hydroxyl groups and higher  $logP$  values (ranging from 3.7 to 5.0) than the acidic prenylated isoflavonoids ( $logP \leq 3.2$  for the undissociated species). It is therefore hypothesized that hydrophobicity becomes the main driver for interaction of these molecules with the membrane<sup>58</sup>. Nevertheless, the presence of a few (<3) hydroxyl groups in these molecules, may also trigger aggregate formation, similarly to the acidic, undissociated (iso)flavonoids. Individually or as aggregates, these molecules might be soluble within the lipid bilayer<sup>58</sup>, due to their higher hydrophobicity, ultimately disturbing its integrity<sup>65</sup> (Fig. 6b, green rectangular). Contrary to prenylated isoflavones (the acidic group in Fig. 6), a few examples from the non-acidic subclasses were shown not to permeabilize the cytoplasmic membrane, but have been hypothesized to still disrupt its integrity by other means<sup>18</sup>.

Interestingly, neutral prenylated isoflavonoids can be planar (pterocarpenes) or non-planar (pterocarpan and isoflavans), while acidic isoflavonoids included in this study are all non-planar. Notably, the two most active mono-prenylated isoflavonoids studied were the pterocarpan Orientanol B (93) (six times more active than its pterocarpene analogue, Dehydroglyceollin IV (102), Table S1 and Fig. S2)) and the isoflavan 4'-O-methylglabridin (17). Both molecules adopt non-planar conformations (Fig. S8). Planarity has been associated to the extent of insertion in the cytoplasmic membrane and the level of disruption of membrane integrity<sup>66-68</sup>. It has been hypothesized that non-planar molecules disrupt membrane integrity more effectively due to better interaction acyl chains<sup>58,67,68</sup>, possibly leading to higher antimicrobial activity. In contrast, planar, molecules may interact less efficiently with the acyl chains, despite their rapid diffusion into the membrane.

Overall, the *in silico* models developed in the study showed for the first time that active prenylated (iso)flavonoids can have fundamentally different molecular properties (such as acidic or highly hydrophobic), suggesting potentially different modes of action. Prenylated (iso)flavonoids might employ different mechanisms for cell uptake or interaction with the cytoplasmic membrane. Information on the contribution of these molecular properties to the anti-MRSA activity can aid the design and development of novel antibacterial agents against Gram-positive bacteria. The best QSAR model developed in the study can be used as a screening tool to predict the (level of) activity of new antibacterial agents against Gram-positive bacteria.

Received: 26 February 2021; Accepted: 9 June 2021

Published online: 09 July 2021

## References

1. Monegro, A. F. & Regunath, H. Hospital acquired infections. in *StatPearls* (StatPearls Publishing, 2018).
2. Antoci, E., Pinzone, M., Nunnari, G., Stefani, S. & Cacopardo, B. Prevalence and molecular characteristics of methicillin-resistant *Staphylococcus aureus* (MRSA) among subjects working on bovine dairy farms. *Le Infezioni in Medicina: Rivista Periodica di Epidemiologia, Epidemiologia, Diagnostica, Clinica e Terapia delle Patologie Infettive* **21**, 125–129 (2013).
3. Control, C. f. D. & Prevention. *Active bacterial core surveillance report, emerging infections program network, methicillin resistant Staphylococcus aureus*, <http://www.cdc.gov/abcs/reports-findings/survreports/mrsa14.html> (2014).
4. ECDC. Antimicrobial resistance surveillance in Europe 2014. Annual Report of the European Antimicrobial Resistance Surveillance Network (EARS-Net). (Stockholm, 2015).

5. Organization, W. H. 2019 antibacterial agents in clinical development: An analysis of the antibacterial clinical development pipeline (2019).
6. Lehman, S. M. *et al.* Design and preclinical development of a phage product for the treatment of antibiotic-resistant *Staphylococcus aureus* infections. *Viruses* **11**, 88 (2019).
7. Vestergaard, M., Frees, D. & Ingmer, H. Antibiotic resistance and the MRSA problem. *Gram-Positive Pathogens* **7**, 747–765 (2019).
8. Araya-Cloutier, C., Vincken, J.-P., van Ederen, R., den Besten, H. & Gruppen, H. Rapid membrane permeabilization of *Listeria monocytogenes* and *Escherichia coli* induced by antibacterial prenylated phenolic compounds from legumes. *Food Chem.* **240**, 147–155 (2018).
9. de Bruijn, W. J. *et al.* Antibacterial prenylated stilbenoids from peanut (*Arachis hypogaea*). *Phytochem. Lett.* **28**, 13–18 (2018).
10. Veitch, N. C. Isoflavonoids of the Leguminosae. *Nat. Prod. Rep.* **30**, 988–1027 (2013).
11. Botta, B. *et al.* Prenylated isoflavonoids: Botanical distribution, structures, biological activities and biotechnological studies. An update (1995–2006). *Curr. Med. Chem.* **16**, 3414–3468 (2009).
12. Veitch, N. C. Flavonoid chemistry of the leguminosae. *Recent Adv. Polyphenol Res.* **2**, 23–58 (2010).
13. Hatano, T. *et al.* Phenolic constituents of licorice. VIII. Structures of glicophenone and glicoisoflavone, and effects of licorice phenolics on methicillin-resistant *Staphylococcus aureus*. *Chem. Pharm. Bull.* **48**, 1286–1292 (2000).
14. Sato, M., Tanaka, H., Tani, N., Nagayama, M. & Yamaguchi, R. Different antibacterial actions of isoflavones isolated from *Erythrina poeppigiana* against methicillin-resistant *Staphylococcus aureus*. *Lett. Appl. Microbiol.* **43**, 243–248 (2006).
15. Wang, S. Y. *et al.* Flavonoids from *Sophora moorcroftiana* and their synergistic antibacterial effects on MRSA. *Phytother. Res.* **28**, 1071–1076 (2014).
16. Tanaka, H. *et al.* Antibacterial activity of isoflavonoids isolated from *Erythrina variegata* against methicillin-resistant *Staphylococcus aureus*. *Lett. Appl. Microbiol.* **35**, 494–498 (2002).
17. Zhou, B. & Wan, C.-X. Phenolic constituents from the aerial parts of *Glycyrrhiza inflata* and their antibacterial activities. *J. Asian Nat. Prod. Res.* **17**, 256–261 (2015).
18. Araya-Cloutier, C. *et al.* QSAR-based molecular signatures of prenylated (iso) flavonoids underlying antimicrobial potency against and membrane-disruption in Gram positive and Gram negative bacteria. *Sci. Rep.* **8**, 9267 (2018).
19. Nandi, S., Ahmed, S. & Saxena, A. Combinatorial design and virtual screening of potent anti-tubercular fluoroquinolone and isothiazoloquinolone compounds utilizing QSAR and pharmacophore modelling. *SAR QSAR Environ. Res.* **29**, 151–170 (2018).
20. Gramatica, P. Principles of QSAR models validation: Internal and external. *QSAR Comb. Sci.* **26**, 694–701 (2007).
21. Sadgrove, N. J., Oliveira, T. B., Khumalo, G. P., Vuuren, S. F. V. & van Wyk, B.-E. Antimicrobial isoflavones and derivatives from *Erythrina* (Fabaceae): Structure activity perspective (sar & qsar) on experimental and mined values against *Staphylococcus Aureus*. *Antibiotics* **9**, 223 (2020).
22. van de Schans, M. G. M. *et al.* Involvement of a hydrophobic pocket and helix 11 in determining the modes of action of prenylated flavonoids and isoflavonoids in the human estrogen receptor. *ChemBioChem* **16**, 2668–2677 (2015).
23. van de Schans, M. G. M. *et al.* Glyceollins and dehydroglyceollins isolated from soybean act as SERMs and ER subtype-selective phytoestrogens. *J. Steroid Biochem. Mol. Biol.* **156**, 53–63 (2016).
24. Aryani, D., Den Besten, H., Hazeleger, W. & Zwietering, M. Quantifying strain variability in modeling growth of *Listeria monocytogenes*. *Int. J. Food Microbiol.* **208**, 19–29 (2015).
25. Tanaka, H. *et al.* Three new isoflavanones from *Erythrina costaricensis*. *Nat. Prod. Res.* **23**, 1089–1094 (2009).
26. Tanaka, H. *et al.* Antibacterial constituents from the roots of *Erythrina herbacea* against methicillin-resistant *Staphylococcus aureus*. *Planta Med.* **76**, 916–919 (2010).
27. Tanaka, H. *et al.* Two new isoflavanones from the roots of *Erythrina variegata*. *Nat. Prod. Commun.* **10**, 1934578X1501000330 (2015).
28. Faulon, J.-L. & Bender, A. *Handbook of Chemoinformatics Algorithms* (CRC Press, 2010).
29. Team, R. C. R. A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria (2021). <https://www.R-project.org/>.
30. Kennard, R. W. & Stone, L. A. Computer aided design of experiments. *Technometrics* **11**, 137–140 (1969).
31. de Bruijn, W. J., Hageman, J. A., Araya-Cloutier, C., Gruppen, H. & Vincken, J.-P. QSAR of 1, 4-benzoxazin-3-one antimicrobials and their drug design perspectives. *Bioorg. Med. Chem.* **26**, 6105–6114 (2018).
32. Eriksson, L. *et al.* Methods for reliability and uncertainty assessment and for applicability evaluations of classification- and regression-based QSARs. *Environ. Health Perspect.* **111**, 1361–1375 (2003).
33. Tanaka, H. *et al.* Antibacterial properties of a new isoflavonoid from *Erythrina poeppigiana* against methicillin-resistant *Staphylococcus aureus*. *Phytomedicine* **11**, 331–337 (2004).
34. Tropsha, A. Best practices for QSAR model development, validation, and exploitation. *Mol. Inf.* **29**, 476–488 (2010).
35. Dolatabadi, M., Nekoei, M. & Banaei, A. Prediction of antibacterial activity of pleuromutilin derivatives by genetic algorithm-multiple linear regression (GA-MLR). *Monatsh. Chem. Chem. Mon.* **141**, 577–588 (2010).
36. Consonni, V., Ballabio, D. & Todeschini, R. Comments on the definition of the Q(2) parameter for QSAR validation. *J. Chem. Inf. Model.* **49**, 1669–1678 (2009).
37. Keepers, T. R., Gomez, M., Biek, D., Critchley, I. & Krause, K. M. Effect of *in vitro* testing parameters on ceftazidime-avibactam minimum inhibitory concentrations. *Int. Scholar. Res. Not.* **2015**, 1–6 (2015).
38. Brennan-Krohn, T., Smith, K. P. & Kirby, J. E. The poisoned well: Enhancing the predictive value of antimicrobial susceptibility testing in the era of multidrug resistance. *J. Clin. Microbiol.* **55**, 2304–2308 (2017).
39. Roy, K., Das, R. N., Ambure, P. & Aher, R. B. Be aware of error measures. Further studies on validation of predictive QSAR models. *Chemom. Intell. Lab. Syst.* **152**, 18–33 (2016).
40. Gibbons, S. Anti-staphylococcal plant natural products. *Nat. Prod. Rep.* **21**, 263–277 (2004).
41. Koh, J.-J. *et al.* Rapid bactericidal action of alpha-mangostin against MRSA as an outcome of membrane targeting. *Biochim. Biophys. Acta Biomembr.* **1828**, 834–844 (2013).
42. Coyle, E. A. & Rybak, M. J. Activity of coytavancin (LY333328), an investigational glycopeptide, compared to that of vancomycin against multidrug-resistant *Streptococcus pneumoniae* in an *in vitro* pharmacodynamic model. *Antimicrob. Agents Chemother.* **45**, 706–709 (2001).
43. Cruciani, G., Crivori, P., Carruti, P.-A. & Testa, B. Molecular fields in quantitative structure-permeation relationships: The VolSurf approach. *J. Mol. Struct. (Theochem)* **503**, 17–30 (2000).
44. Moorthy, N. H. N., Cerqueira, N. S., Ramos, M. J. & Fernandes, P. A. QSAR analysis of 2-benzoxazolyl hydrazone derivatives for anticancer activity and its possible target prediction. *Med. Chem. Res.* **21**, 133–144 (2012).
45. Darden, J. C. *et al.* Mechanism-based QSAR modeling of skin sensitization. *Chem. Res. Toxicol.* **28**, 1975–1986 (2015).
46. Labute, P., Kossner, M., Ajamian, A., Santavy, M. & Lin, A. Pharmacophore annotation using extended Hückel theory. *J. Cheminform.* **6**, 1–1 (2014).
47. Allen, C. H., Mervin, L. H., Mahmoud, S. Y. & Bender, A. Leveraging heterogeneous data from GHS toxicity annotations, molecular and protein target descriptors and Tox21 assay readouts to predict and rationalise acute toxicity. *J. Cheminform.* **11**, 36 (2019).
48. Ghafourian, T. *et al.* QSPR models for the prediction of apparent volume of distribution. *Int. J. Pharm.* **319**, 82–97 (2006).
49. Labute, P. Derivation and applications of molecular descriptors based on approximate surface area. in *Chemoinformatics* 261–278 (Springer, 2004).

50. Potta, T. *et al.* Discovery of antibiotics-derived polymers for gene delivery using combinatorial synthesis and cheminformatics modeling. *Biomaterials* **35**, 1977–1988 (2014).
51. Bhardwaj, R. M. *Control and Prediction of Solid-state of Pharmaceuticals: Experimental and Computational Approaches* (Springer, 2016).
52. King, E. M., Gebbie, M. A. & Melosh, N. A. Impact of rigidity on molecular self-assembly. *Langmuir* **35**, 16062–16069 (2019).
53. Klymchenko, A. S., Pivovarenko, V. G. & Demchenko, A. P. Perturbation of planarity as the possible mechanism of solvent-dependent variations of fluorescence quantum yield in 2-aryl-3-hydroxychromones. *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* **59**, 787–792 (2003).
54. Bajorath, J. *Representation and Identification of Activity Cliffs* (Taylor & Francis, 2017).
55. Yazaki, K., Sasaki, K. & Tsurumaru, Y. Prenylation of aromatic compounds, a key diversification of plant secondary metabolites. *Phytochemistry* **70**, 1739–1745 (2009).
56. Epan, R. M., Walker, C., Epan, R. F. & Magarvey, N. A. Molecular mechanisms of membrane targeting antibiotics. *Biochim. Biophys. Acta Biomembr.* **1858**, 980–987 (2016).
57. Lambert, P. Cellular impermeability and uptake of biocides and antibiotics in Gram-positive bacteria and mycobacteria. *J. Appl. Microbiol.* **92**, 46S–54S (2002).
58. Li, J., Beuerman, R. W. & Verma, C. S. Molecular insights into the membrane affinities of model hydrophobes. *ACS Omega* **3**, 2498–2507 (2018).
59. Pang, D. *et al.* Destruction of the cell membrane and inhibition of cell phosphatidic acid biosynthesis in *Staphylococcus aureus*: An explanation for the antibacterial mechanism of morusin. *Food Funct.* **10**, 6438–6446 (2019).
60. Wu, S.-C. *et al.* Natural flavones from *Morus alba* against methicillin-resistant *Staphylococcus aureus* via targeting the proton motive force and membrane permeability. *J. Agric. Food Chem.* **67**, 10222–10234 (2019).
61. Negi, J., Bisht, V., Singh, P., Rawat, M. & Joshi, G. Naturally occurring xanthenes: Chemistry and biology. *J. Appl. Chem.* **2013**, 1–9 (2013).
62. Cramariuc, O. *et al.* Mechanism for translocation of fluoroquinolones across lipid membranes. *Biochim. Biophys. Acta Biomembr.* **1818**, 2563–2571 (2012).
63. Tommasi, R., Brown, D. G., Walkup, G. K., Manchester, J. I. & Miller, A. A. ESKAPEing the labyrinth of antibacterial discovery. *Nat. Rev. Drug Discov.* **14**, 529–542 (2015).
64. Franklin, T. & Snow, G. Penetrating the defences: How antimicrobial drugs reach their targets. in *Biochemistry and Molecular Biology of Antimicrobial Drug Action* 107–118 (Springer, 1998).
65. Santos, R. S., Figueiredo, C., Azevedo, N. F., Braeckmans, K. & De Smedt, S. C. Nanomaterials and molecular transporters to overcome the bacterial envelope barrier: Towards advanced delivery of antibiotics. *Adv. Drug Deliv. Rev.* **136**, 28–48 (2018).
66. Li, X.-Z., Plésiat, P. & Nikaido, H. The challenge of efflux-mediated antibiotic resistance in Gram-negative bacteria. *Clin. Microbiol. Rev.* **28**, 337–418 (2015).
67. Wesołowska, O., Gąsiorowska, J., Petrus, J., Czarnik-Matusewicz, B. & Michalak, K. Interaction of prenylated chalcones and flavanones from common hop with phosphatidylcholine model membranes. *Biochim. Biophys. Acta Biomembr.* **1838**, 173–184 (2014).
68. van Dijk, C., Driessen, A. J. & Recourt, K. The uncoupling efficiency and affinity of flavonoids for vesicles. *Biochem. Pharmacol.* **60**, 1593–1600 (2000).

## Acknowledgements

Authors are grateful to John Chapman for critically reading the manuscript.

## Author contributions

S.K., C.A.-C. and J.-P.V. conceptualized the study; S.K. and C.A.-C. designed and performed the experimental work; S.K., C.A.-C. and J.H. performed the data analysis (molecular modelling and statistical analysis); S.K. drafted the manuscript. All authors reviewed the manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-021-92964-9>.

**Correspondence** and requests for materials should be addressed to J.-P.V.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021, corrected publication 2022