Active Machine Learning for Formulation of Precision Probiotics

Laura E. McCoubrey, Nidhi Seegobin, Moe Elbadawi, Yiling Hu, Mine Orlu, Simon Gaisford, Abdul W. Basit*.

University College London School of Pharmacy, London, United Kingdom.

* Corresponding author: A.W.B: a.basit@ucl.ac.uk

Highlights

- Active machine learning was applied to formulation and microbiome science for the first time.
- Leveraging a small dataset of 6 probiotic-excipient interactions, the effects of a further 111 pharmaceutical excipients on probiotic proliferation were predicted.
- Uncertainty sampling was used to obtain a final machine learning model certainty of 67.70%
- Experimental validation found that the effects of 3/4 tested excipients could be correctly predicted.
- Feature importance analysis via random forest and principal component analysis found the most influential chemical features determining excipients' effects on probiotic proliferation.

Keywords: artificial intelligence; in silico prediction; drug discovery and development; live biotherapeutic products; next generation probiotics; colonic delivery.

Abstract

It is becoming clear that the human gut microbiome is critical to health and well-being, with increasing evidence demonstrating that dysbiosis can promote disease. Increasingly, precision probiotics are being investigated as investigational drug products for restoration of healthy microbiome balance. To reach the distal gut alive where the density of microbiota is highest, oral probiotics should be protected from harsh conditions during transit through the stomach and small intestines. At present, few probiotic formulations are designed with this delivery strategy in mind. This study employs an emerging machine learning (ML) technique, known as active ML, to predict how excipients at pharmaceutically relevant concentrations affect the intestinal proliferation of a common probiotic, *Lactobacillus paracasei*. Starting with a labelled dataset of just 6 bacteria-excipient interactions, active ML was able to predict the effects of a further 111 excipients using uncertainty sampling. The average certainty of the final model was 67.70% and experimental validation demonstrated that 3/4 excipient-probiotic interactions could be correctly predicted. The model can be used to enable superior probiotic delivery to maximise proliferation in vivo and marks the first use of active ML in microbiome science.

1. Introduction

Research over the past decade has revealed the importance of the gut microbiome for human health (Proctor et al., 2019). The microbiome consists of trillions of microorganisms, including bacteria, fungi, and viruses, most of which inhabit the distal gastrointestinal (GI) tract (Martinez-Guryn et al., 2019). Many diseases have now been linked to gut dysbiosis, including metabolic syndrome, cancer, and neurological disorders, which represent the most common causes of death worldwide (Singer-Englar et al., 2019; World Health Organization, 2020). Dysbiosis can arise in response to many stimuli, with exposure to antimicrobials, a low fibre diet, and polypharmacy representing frequently investigated examples (Maier et al., 2021; So et al., 2018). The relationships between dysbiosis and disease are typically complex and condition-dependent, however in general terms dysbiosis increases risk of host disease because it involves alteration of the microbiome's symbiotic functions (Janssens et al., 2018). In recent years a new class of therapeutics has been founded, known as microbiome medicines, which are designed to prevent and treat disease through prevention and treatment of dysbiosis (Cammarota et al., 2020; Marcos-Fernández et al., 2021). Probiotics form a key division of microbiome medicines and are defined as live microorganisms that confer a health benefit when administered in sufficient quantities (Hill et al., 2014; Moens et al., 2019). The scientific community is increasingly investigating defined strains of probiotics as potential drug products with characterised mechanisms, specified indications, and the ability to demonstrate clinical benefit in human trials (Ghyselinck et al., 2021; Kim et al., 2018; Veiga et al., 2020; Yu et al., 2020). These therapeutics are known as precision or next generation probiotics (or live biotherapeutic products) and could revolutionise approaches to disease management as they enter the market in coming years (Aggarwal et al., 2020; Allegretti et al., 2019; O'Toole et al., 2017; Wilkinson et al., 2021).

Formulation of probiotics will play a key role in their translation to approved medicinal products. At present, probiotic colonisation of the GI tract is strain-dependent and mostly transient, often varying between individuals depending on pre-treatment microbiome composition . In part, exposure to harsh conditions in the upper GI tract including gastric acid and bile salts can contribute to low probiotic colonisation efficacy by reducing the number of viable microorganisms reaching the distal gut (Ding and Shah, 2007; Fredua-Agyeman and Gaisford, 2015; McConnell et al., 2008). This can be overcome by several strategies including using acid-tolerant probiotics and/or protecting probiotics during passage through the proximal GI tract. The latter strategy can employ coating technologies that release therapeutics specifically within the colon where microbiota density is highest (Dodoo et al., 2017; Liao et

al., 2020; Varum et al., 2020a; Varum et al., 2020b). However, delivery to the intended site of colonisation still does not guarantee that probiotic strains will assimilate with commensal species (Suez et al., 2019). Here, considered formulation design can be applied to maximise colonisation efficiency. Whilst this concept is in its relative infancy, advances have been made in recent years. For example, liquid formulations have been suggested to achieve superior probiotic viability compared to freeze-dried tablets (Fredua-Agyeman and Gaisford, 2015; Ghyselinck et al., 2021). Elsewhere, mucoadhesive microspheres have been shown to enhance gastric acid resistance and intestinal colonisation of probiotic species (Liu et al., 2020). The selection of functional excipients for probiotic formulations may therefore play an important role in their eventual therapeutic efficacy (Klayraung et al., 2009; Raddatz et al., 2020; Sreeja et al., 2016).

During development of a probiotic strain the number of available excipients for formulation is high, and the potential for an excipient to influence the viability and/or growth of a probiotic is significant. This introduces a barrier to the rapid and optimal formulation of probiotic products. In this work an advanced and emerging form of machine learning (ML), known as active ML, is applied to analyse and predict excipient effects on probiotic growth, hence enabling a strategic approach to probiotic delivery. Whilst a few applications of traditional ML techniques for advanced probiotic delivery exist, this is the first study to apply active ML to next generation probiotics and microbiome science as a whole (McCoubrey et al., 2021c; Westfall et al., 2021). Active ML is well suited to generating predictions from small datasets where collection of large data is not feasible or desired (McCoubrey et al., 2021a). During active ML, an initial ML model is developed on available labelled data. The model can then output predictions for unlabelled data, whilst indicating its certainty for each prediction. Users can experimentally test datapoints for which the model is most uncertain, and subsequently teach the model the new results, with the aim of improving overall model certainty (Reker, 2019). In this way, active ML is a symbiotic process between the user and ML workflow. Researchers with limited capacity to build the large datasets often required for traditional supervised ML techniques can use active ML to harness the potential of artificial intelligence with maximal resource efficiency.

Here, active ML was used to predict the effects of 111 pharmaceutical excipients on the growth of a common commercially available probiotic, *Lactobacillus paracasei*. Predictions were based on a small dataset of just 6 excipient-probiotic interactions, as this was considered as a feasible number of datapoints to collect during routine formulation development. *L. paracasei* is a well-research probiotic species, with documented benefits for a myriad of diseases, including atopic dermatitis, age-related

retinal cell loss, and colitis (Chen et al., 2019; Kim et al., 2020; Morita et al., 2018). The ML model developed in this study is designed to facilitate the selection of functional excipients for the optimisation of *L. paracasei* proliferation in vivo. Methods used can be easily translated for the development of any precision probiotic, and present active ML as a powerful tool for optimising the in vivo performance of pharmaceutical formulations.

2. Materials and Methods

2.1 Materials

Lactobacillus paracasei CASEI 431® was purchased from Wren Laboratories Ltd. (Hampshire, United Kingdom). MRS agar was obtained from Oxoid Ltd. (Basingstoke, United Kingdom). MRS broth, guar, D-mannitol, polysorbate (Tween®) 80, L-cysteine hydrochloride, dibutyl sebacate, acetic anhydride, sodium benzoate, phosphate buffered saline (PBS) tablets, and β-carotene were purchased from Sigma Aldrich (Dorset, United Kingdom). Sodium carbonate and magnesium sulphate were purchased from BDH (Dubai, United Arab Emirates). Aspartame and doxycycline hydrochloride were purchased from Fisher Scientific (Leicestershire, United Kingdom). Sucrose was purchased from Alfa Aesar (Lancashire, United Kingdom).

2.2 Methods

2.2.1 Analysis of probiotic growth

Isothermal microcalorimetry (IMC) was used to monitor microbial growth within media, as any heat produced by the metabolic activity of microorganisms can be used as a direct representation of their growth (Cabadaj et al., 2021; Said et al., 2014). *L. paracasei* was used as received in lyophilised powder form. Probiotic powder was emptied from capsules received from the supplier and suspended in 5.0 mL PBS, mixed by vortexing for 20 seconds. The enumeration of *L. paracasei* in this suspension was observed as 3.4×10^4 CFU/mL by plating on MRS agar and anaerobic incubation at 37 °C for 48 hours. Immediately after mixing, 2.8 µL of the *L. paracasei* PBS suspension was pipetted aseptically into 2.8 or 5 mL (depending on the calorimeter used) MRS broth inside glass vials pre-warmed to 37 °C (n=3). The inoculated vials were hermetically sealed and transferred to the calorimeter (2277 TAM; TA Instruments Ltd., UK). The calorimeter was operated at 37 °C to reflect in vivo conditions. Following a thermal acclimatisation period of 30 minutes, thermal data were recorded using a dedicated software package, Digitam 4.1 (set to record 1 data point every 10 seconds). *L. paracasei* growth was measured over 45 hours to allow visualisation of the bacterial growth curve.

2.2.2 Measuring excipient effects on *L. paracasei* growth

Each investigated excipient was incorporated into MRS broth at a concentration of 50 µM. This concentration was selected as drugs are typically present in the terminal ileum and colon, where microbiota populations are most dense, at concentrations of 20 µM (Maier et al., 2018). Because excipients are often present in higher concentrations than drugs in pharmaceutical formulations, the higher concentration of 50 µM was selected. This selected concentration was examined against each excipient's maximum potency permitted in single FDA-approved oral dosage forms, and confirmed as pharmaceutically relevant (U.S. Food and Drug Administration, 2020). The growth of *L. paracasei* in the presence of 50 µM excipient was measured via IMC as in Section 2.2.1 (n=3 for each excipient). Growth curves obtained in the presence of excipients were compared with growth curves obtained in pure microbial growth medium (n=3). The area under the curve (AUC), time taken to reach peak power (t_{max}), and the maximum power (P_{max}) were used as comparative features. A T-test, performed using the SciPy statistical software (version 1.5.2) for Python, was used to assess whether differences between microbial growth in the presence of excipients, compared with the absence of excipients, were significant ($p < 0.05$). Where an excipient significantly reduced AUC, increased t_{max}, or reduced P_{max}, then the excipient was determined as inhibiting microbial growth. Conversely, if an excipient significantly increased AUC, decreased t_{max}, or increased P_{max}, then the excipient was determined as promoting microbial growth. If excipients did not significantly alter AUC, t_{max}, or P_{max}, then they were deemed as neutral: neither inhibiting nor promoting microbial growth. The antibiotic doxycycline hydrochloride was used as a positive control for probiotic growth inhibition.

The initial dataset was built by measuring the effect of 6 excipients found in oral medicines (D-mannitol, polysorbate 80, aspartame, guar, β-carotene, and sucrose) on *L. paracasei* growth. The positive control, doxycycline hydrochloride, was included in this initial dataset. Excipients were chosen to cover a broad chemical and functional space: β-carotene (colourant); sucrose (natural sweetener); mannitol (diluent/plasticiser); aspartame (artificial sweetener); polysorbate 80 (surfactant); guar (binder) (Rowe et al., 2009).

2.2.3 Machine learning

2.2.3.1 Dataset curation

As outlined, the initial dataset consisted of the 6 excipients and doxycycline hydrochloride, labelled with their effects on the growth of *L. paracasei* (labels: neutral, promote, and prevent). The wider unlabelled dataset, known as the pool, consisted of a further 111 further pharmaceutical excipients whose effects

on bacterial growth were unknown. These 111 unlabelled excipients were selected based on their ability to dissolve in the aqueous bacterial growth medium and to cover a broad chemical and functional space. The labelled dataset and unlabelled pool can be found in the Supplementary Materials.

2.2.3.2 Feature selection and preprocessing

Each excipient in the labelled datasets and unlabelled pool were attached to 1613 chemical descriptors based on their simplified molecular-input line-entry system (SMILES) structure, which was obtained from the PubChem database (Weininger, 1988). Mordred, a molecular descriptor calculator, was used to transform excipients' SMILES annotations into the 1613 molecular features (Moriwaki et al., 2018). These molecular features included computational descriptors and more typically recognised chemical features, such as '4-membered aromatic ring count' and 'number of atoms'. All features were scaled prior to ML using Python's Standard Scaler to remove unit bias.

2.2.3.3 Data visualisation

Principal component analysis (PCA) was used for dimension reduction to allow visualisation of relationships between labelled excipients. The PCA tool with Python's scikit-learn package (version 0.23.2) was used with random state = 0. PCA is a commonly used unsupervised tool for dimension reduction in high-dimensional feature spaces and allowed feature relationships to be plotted on 2D graphs (PC1 plotted against PC2) (Taguchi et al., 2015). PCA plots were visually inspected to identify similarities between labelled excipients, i.e., to assess whether excipients with the same label clustered together in the same feature space.

2.2.3.4 Active machine learning

The ActiveLearner and uncertainty sampling packages from the modAL active learning framework for Python (version 0.4.1) were used to develop the ML pipeline in this study (Horvath, 2018). First, the initial labelled dataset was fitted to a Random Forest Classifier (random state = 0) from Python's scikit-learn package. This base model was chosen because it is relatively computationally light and facilitates the interpretation of nonlinear data relationships with protection against overfitting (Chandrashekar and Sahin, 2014; McCoubrey et al., 2021a). Using the modAL framework, predictions for how the 111 excipients in the unlabelled pool would affect the growth of *L. paracasei* were computed. Each prediction (options: neutral, promote, or prevent) was attached to an uncertainty score from 0 - 1.00, with higher values representing the most uncertain predictions. Uncertainty sampling was then implemented to improve model certainty. For each query, an unlabelled excipient with a prediction uncertainty above the average (mean) pool uncertainty was chosen for experimental testing (as in

Section 2.2.2). This represents an uncertainty-based sampling method, which is a common method used in active ML (Reker and Schneider, 2015). Once the true label was measured, this was incorporated into the labelled dataset and thus taught to the ML model. Based on this updated labelled dataset, predictions for the unlabelled excipients were updated along with their uncertainty scores. Further queries were then made using the same method, with the aim of actively increasing overall model certainty through targeted experimental validation.

2.2.3.5 Feature importance

Based on the final model obtained with active ML, the top 10 most important molecular features determining excipients' effects on *L. paracasei* growth were discovered. Two methods for assessing feature importance were used to improve insight and allow comparison between both techniques. In the first, a random forest classifier (random state = 0) within Python's scikit-learn package was applied to rank the importance of molecular features in ascending order of importance. In the second, PCA was used to fit and transform features and extract their factor loadings. Factor loadings describe the magnitude that individual molecular descriptors contribute towards a principal component (PC) and can be positive or negative depending on whether they are positively or negatively correlated with the PC (Esbensen and Geladi, 2009). The 10 features with the highest positive and negative correlations within PC1 were calculated, because PC1 accounts for the highest variance associated with features compared to any other PC.

2.2.4 Data analysis and statistics

A PC (running on Windows 10 64-bit, processor: Intel® Core i7 3770K (Santa Clara, CA, USA), RAM: 16GB DDR3, and graphics card: Asus Phoenix GTX 1660 OC Edition) was used for data analysis and active ML. Both the labelled and unlabelled datasets were saved on Microsoft® Excel® for Microsoft 365 64-bit. Statistical analysis and active ML were completed using Python (version 3.9.0) on Jupyter Notebook (version 6.0.3). PCA and random forest classification were developed using Python's scikit-learn package (version 0.23.2). As detailed in Section 2.2.2, a T-test was used to compare the AUC, t_{max}, and P_{max} values obtained from microbial growth curves, with statistical significance judged as $p < 0.05$. The T-test was run using SciPy statistical software for Python. Plots were constructed using the Matplotlib package in Python and OriginPro (version 2020b).

3. Results and Discussion

3.1 Initial excipient effects on bacterial growth

Table 1 shows the effects of the 6 excipients tested initially on the growth of the probiotic species, *L. paracasei.* 4 of the 6 excipients had no effect on bacterial growth whilst β-carotene was observed to promote growth and guar impaired growth (Supplementary Figure 1).

Table 1. Effects of excipients on the growth of *L. paracasei*. P_{max}: maximum power generated during bacterial growth; t_{max}: time taken for bacteria to reach peak power.

Excipient	Effect observed	P-value	Label
Aspartame	No change	-	Neutral
β-carotene	AUC increased, t_{max} decreased, P_{max} increased	0.014, 0.022, 0.007	Promote
Guar	AUC decreased, t_{max} increased	0.025, 0.023	Prevent
Mannitol	No change	-	Neutral
Polysorbate 80	No change	-	Neutral
Sucrose	No change	-	Neutral

β-carotene is used as a colourant for sugar-coated tablets and can produce formulations with pale yellow to dark orange appearances (Rowe et al., 2009). In the presence of β-carotene, the growth of *L. paracasei* was accelerated (t_{max} decreased, p = 0.022) and intensified (AUC increased, p = 0.014, P_{max} increased, p = 0.007). Past research has shown that the concentration of β-carotene in the diets of *drosophila melanogaster* is positively correlated with relative abundance of intestinal *Lactobacillaceae*, the bacterial family to which *L. paracasei* belongs (Garcia-Lozano et al., 2020). This highlights the in vivo applicability of the results. Guar, a galactomannan used as a binder, thickener, and controlled release agent, reduced the growth of *L. paracasei* (AUC decreased, p = 0.025, and t_{max} increased, p = 0.023). Though this reduction was not as severe as that recorded for the positive control antibiotic, which resulted in complete lack of probiotic growth (data not shown), it does demonstrate that pharmaceutically relevant concentrations of excipients can impair probiotic proliferation. A recent randomised controlled study in male athletes has shown guar intake to significantly alter gut

microbiome composition (Kapoor et al., 2020). However, the results found guar to promote the growth of *Lactobacilli*, which was not reflected in the behaviour of *L. paracasei* in this study. These differences could reflect in vitro vs. in vivo variability or could highlight heterogeneity between different strains of *Lactobacilli*.

3.2 Machine learning

3.2.1 Principal component analysis

Figure 1 demonstrates how the 6 excipient-probiotic interactions mapped onto a PCA plot, in which the location of the excipients relates to their chemical features. Each excipient in this study was assigned 1613 molecular descriptors, thus the PCA plots effectively reduce the large dimensions of the datasets to allow visualisation of how excipients' chemical structures influenced their effect on probiotic growth. Here weak patterns may be forming, for example the 3 excipients with no effect on probiotic growth lie closer in chemical space than those preventing growth. However, with the available data it is not possible to predict untested excipient effects on growth as more datapoints are needed to confirm any emerging relationships. .

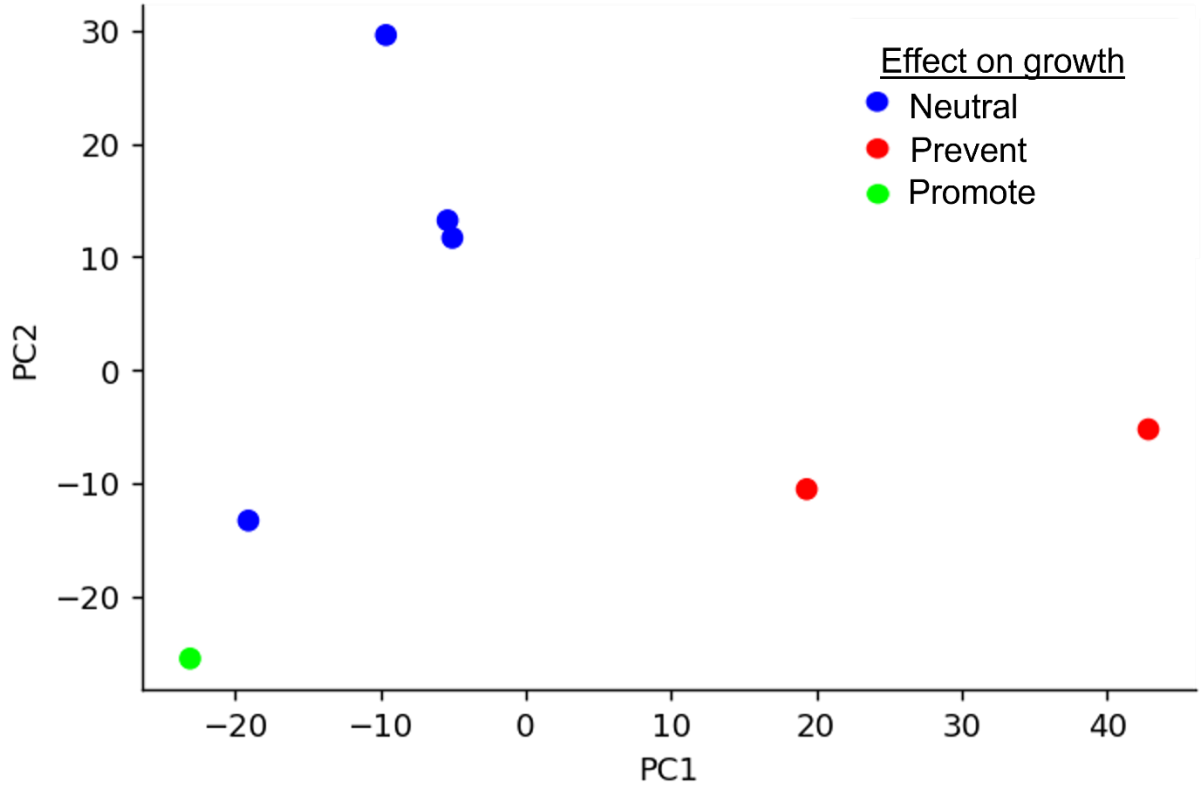


Figure 1. PCA plot (1st and 2nd principal components) depicting how the chemical structure of the 6 labelled excipients and doxycycline hydrochloride relate to their effect on the growth of *L. paracasei*.

Though measuring the effects of more excipients on *L. paracasei* growth was possible, each excipient test was time consuming, requiring 45 hours of measurement. The formulation phase of product development within pharmaceutical industry must be as resource economical as possible, to maintain project timelines and reduce final medicine cost. As such, it is not feasible to perform hundreds of tests per probiotic strain to identify optimal excipients for in vivo colonisation. This highlights the opportunity for an efficient means of predicting excipient-probiotic interactions based on small datasets. Traditional ML techniques, such as artificial neural networks and tree-based methods, typically require very large datasets containing hundreds to thousands of samples to reliably identify intra-data relationships and output accurate predictions (McCoubrey et al., 2021b). Here, active ML can be employed as a tool for leveraging small datasets for accurate in silico predictions (Elbadawi et al., 2021).

3.2.2 Active machine learning

Figure 2 demonstrates how the predictive certainty of the active ML model changed during the process of selective sampling. Based on the starting dataset of 6 excipient-probiotic interactions, the model could predict the effects of a further 111 untested excipients on *L. paracasei* growth with a certainty of 66.68% (± 13.34). The high certainty could be attributed to the comprehensive features used as inputs to the ML model. With the aim of improving model performance, excipients with below average prediction certainties were selected for experimental testing and subsequent teaching to the model. The first excipient selected for uncertainty sampling was sodium benzoate, a preservative and lubricant in tablet and capsule formulations (Rowe et al., 2009). The molecule was predicted to have no effect on *L. paracasei* growth with a model certainty of 63%. In reality, experimental testing showed sodium benzoate to impair probiotic growth by increasing t_{max} ($p = 0.036$). Logically, this result was expected due to the excipient's known use as an antimicrobial preservative and highlights the requirement of the ML algorithm for more training data. In response to the incorrect prediction the mean model certainty reduced to 60.34% (± 9.98).

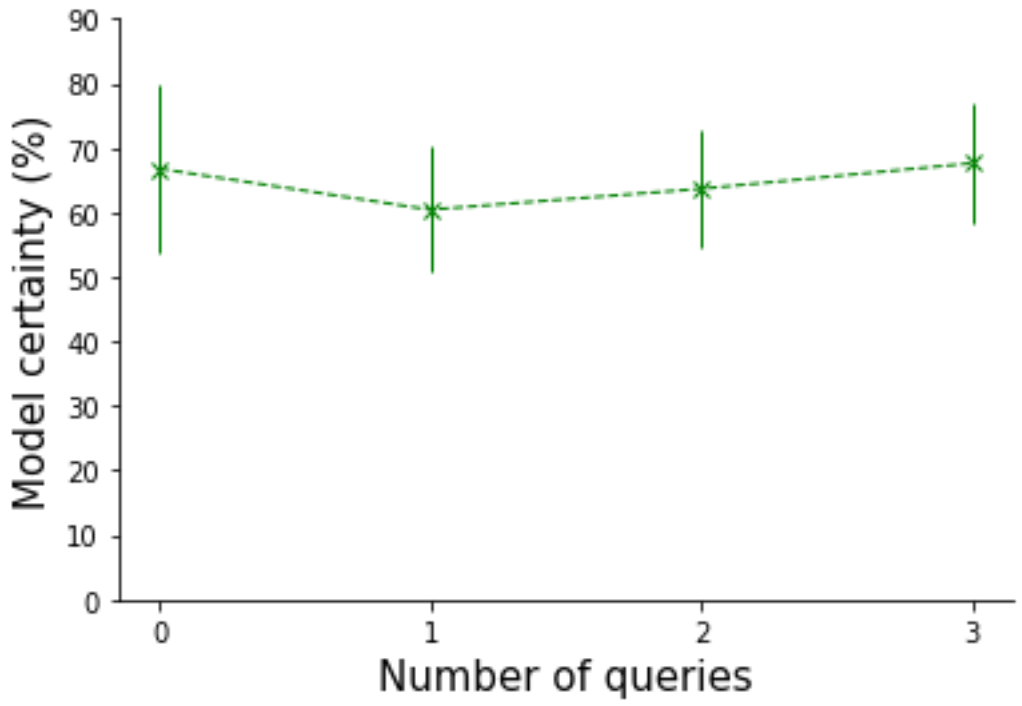


Figure 2. The relationship between average model certainty (for the prediction of the effects 111 unlabelled excipients on *L. paracasei* growth) and the number of queries (excipient-bacteria interactions experimentally tested and taught to the active machine learning model). Error bars represent standard deviation.

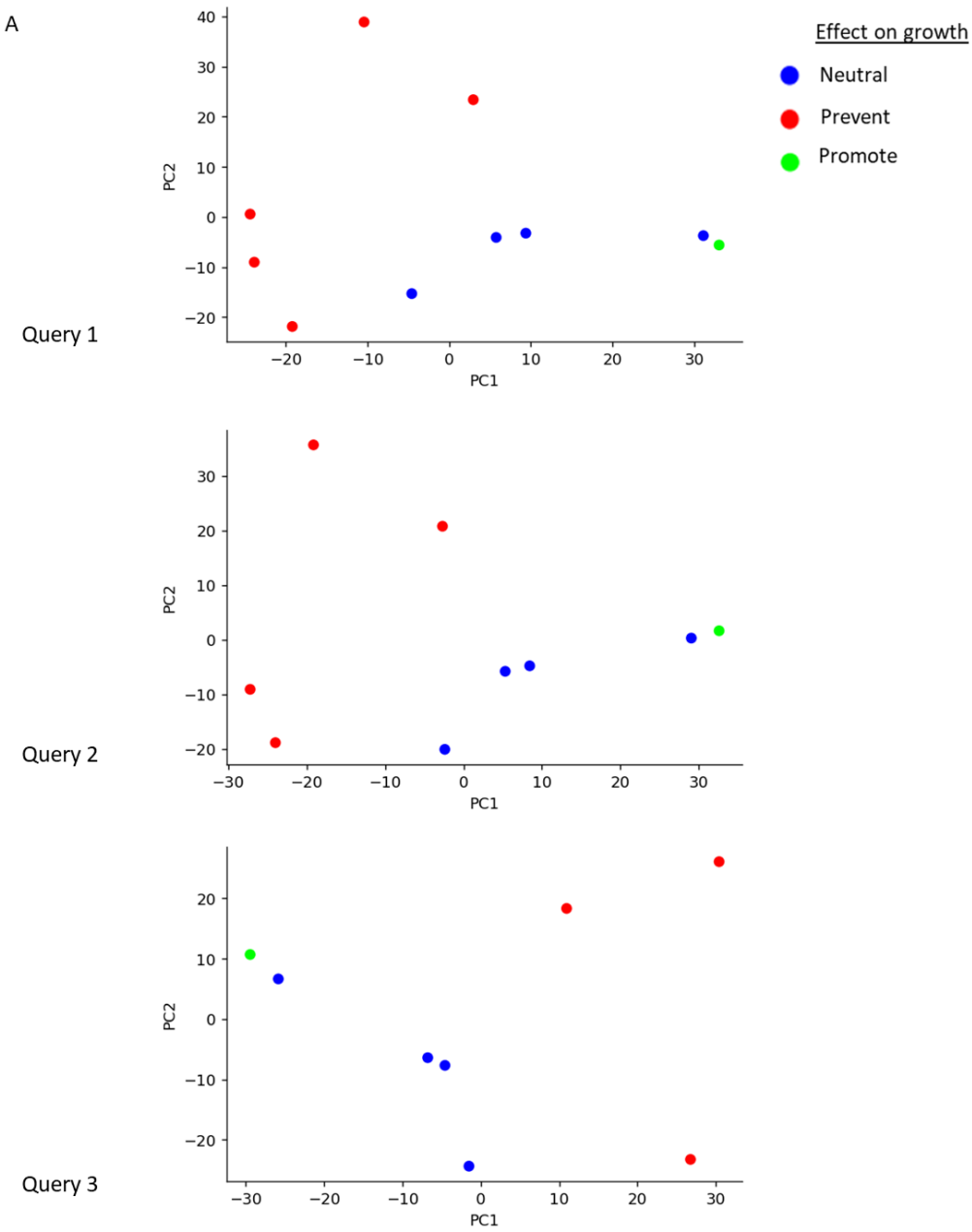
The second excipient selected for uncertainty sampling was the antioxidant cysteine hydrochloride, predicted to have no effect on probiotic growth with a certainty of 51%. In actuality the excipient significantly reduced the AUC ($p = 0.006$) of the *L. paracasei* growth curve, and thus was labelled as impairing bacterial growth. Though incorrect in its prediction, the mean certainty of the ML model increased to 63.64% (± 8.86) following this sample. This increase in mean certainty in response to an incorrect prediction is different to the first sample, where mean certainty decreased upon incorrect label assignment. The change likely demonstrates that the model was learning important relationships in the data and becoming more confident in predicting cases of impaired probiotic growth. The starting dataset contained just one excipient that impaired growth, thus more examples of this label gave the model more training data to learn from. Acetic anhydride was chosen as the final excipient to be tested through uncertainty sampling and was predicted to have a neutral effect on bacterial growth with a certainty of 53%. The molecule was subsequently observed to reduce both the AUC ($p = 0.022$) and t_{max}

of the *L. paracasei* growth curve, and so impaired growth. Supported by this additional data, the overall mean certainty of the model increased to 67.70% (± 9.30).

In principle, the more queries performed during uncertainty sampling, the more data the model has to learn from, and therefore the greater its performance will become. In this study, it was decided that 3 queries was acceptable as it is a relatively achievable number of experiments to carry out in practice when resources are finite. After the 3 queries average prediction certainty for the unlabelled excipients increased and standard deviation decreased, demonstrating how active ML can be used to improve models via targeted learning. Though the changes in model certainty were not statistically significant, the changes to mean certainties and the reduction in standard deviations represent a trend towards evolving confidence. It is likely that far more queries would be required to significantly increase the certainty, due to the small size of the dataset and the initially high certainty.

Figure 3A shows updated PCA plots with the excipient-probiotic interactions included from each query during uncertainty sampling. It highlights that the early excipient structure-activity relationships observed in Figure 1 were strengthened during active ML, i.e., excipients that impaired *L. paracasei* growth are more similar in chemical feature space than those with no effect. Figure 3B depicts a PCA plot including the pool of untested excipients labelled with the final model predictions. It demonstrates how the structure-activity relationships determined during excipient testing inform predictions for untested excipient. PCA is an unsupervised ML technique that can cluster groups without being supervised. Thus, in this study PCA found an inherent difference between the predictions derived from the active learning model. Hence, it is possible that the active model learned the key characteristics separating the two classes. There are noticeably no predictions for excipients promoting probiotic growth; this is likely because only one excipient resulting in *L. paracasei* growth promotion was supplied to the model during training. Based on the experimental data, it is probable that excipient-probiotic interactions resulting in growth promotion are rarer than those with no effect or inhibition of growth. Depending on resources available, researchers following the active ML process for drug development may wish to ensure equal representation of all label categories during model training to maximise accuracy. The final model predictions for the impact of the untested excipients on *L. paracasei* growth, with associated certainties, are provided in the Supplementary Materials.

A


Effect on growth
Neutral
Prevent
Promote
Query 1
Query 2
Query 3
PC1
PC2

B

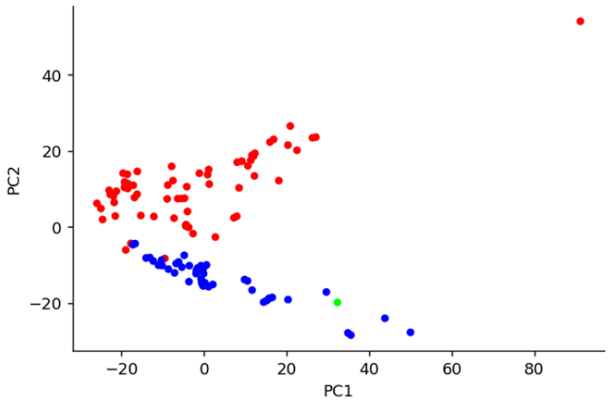

PC1
PC2

Figure 3. PCA plots (1st and 2nd principal components) constructed, A: after each query of the active machine learning process, depicting how excipients' chemical structures relate to their effect on the growth of *L. paracasei*, and B: incorporating the pool of untested excipients labelled with the final machine learning model predictions.

3.2.4 Model validation

Table 2 shows the accuracy of the final model's predictions for 4 randomly selected excipients used in oral pharmaceutical formulations. The results confirm that dibutyl sebacate (a plasticiser for film coatings); magnesium sulfate (present in solid dosage forms and used as a vaccine adjuvant); and sodium carbonate (a buffering/dispersing agent and oral dosage form diluent) impaired the growth of *L. paracasei*, and thus would not be ideal excipients for optimising probiotic proliferation in vivo. In comparison, glycerin (used as a solvent, sweetener, and viscosity-increasing agent) had no effect on bacterial growth therefore is more suitable for precision probiotic delivery (Rowe et al., 2009). Table 2 demonstrates that 3/4 of the validation experiments were predicted correctly following active ML. Before the uncertainty sampling process, the ML model could only predict 1/4 of the cases correctly (glycerin, with its neutral effect).

Table 2. Experimental validation results for untested excipient effects on *L. paracasei* growth, with associated machine learning model predictions and certainties.

Excipient	**Model prediction**	**Certainty**	**Experimental outcome**	**Model correct?**
Dibutyl sebacate	Neutral	73%	Prevent: AUC decreased ($p = 0.002$) and t_{max} increased ($p = 8.54 \times 10^{-5}$)	No
Magnesium sulfate	Prevent	75%	Prevent: AUC decreased ($p = 4.64 \times 10^{-5}$)	Yes
Glycerin	Neutral	68%	Neutral	Yes
Sodium carbonate	Prevent	76%	Prevent: AUC decreased ($p = 4.13 \times 10^{-6}$)	Yes

This study marks the first time that active ML has been applied to microbiome science and is among the first uses in drug discovery and development. The few existing pharmaceutical applications of active ML

have focused on drug discovery, and so this work highlights the benefits of applying active ML to drug development and delivery (Reker, 2019). Due to its high costs the drug development pipeline must be as streamlined as possible to allow novel therapeutics to reach the market with the longest active patent life. For this reason, active ML is an ideal technique for use in the pharmaceutical industry: it carries the predictive power of artificial intelligence whilst requiring minimal resources.

3.2.5 Feature importance

Figure 4 shows a random forest calculation of the top 10 molecular features determining excipient effects on the growth of *L. paracasei*. A molecular feature is a single quantifiable property of a compound's 2D or 3D chemical structure (Chandrashekar and Sahin, 2014). Notably, the most important features were computational, as opposed to more universally recognised descriptors, such as molecular weight or atom count. The utility of complex computational fingerprints has shown in other ML studies, demonstrating that detailed representation of compounds facilitates high performance (McCoubrey et al., 2021b; Reker et al., 2020; Schmidt et al., 2019) The most important feature determining excipient effects on probiotic growth was found to be the eccentric connectivity index (ECIndex), which is a fourth-generation topological descriptor commonly applied to model biological activities of compounds (Sharma et al., 1997). ECIndex has been described as 'the sum total of the product of eccentricity and degree of each vertex in a hydrogen-suppressed molecular graph having n total vertices', thus highlighting the descriptor's mathematical nature (Sharma et al., 1997). When the dataset was examined, labelled excipients with extremes of ECIndex values were observed to impair probiotic growth, whereas excipients with ECIndex values between 114 and 1953 only promoted or had no effect on growth (Table 3).

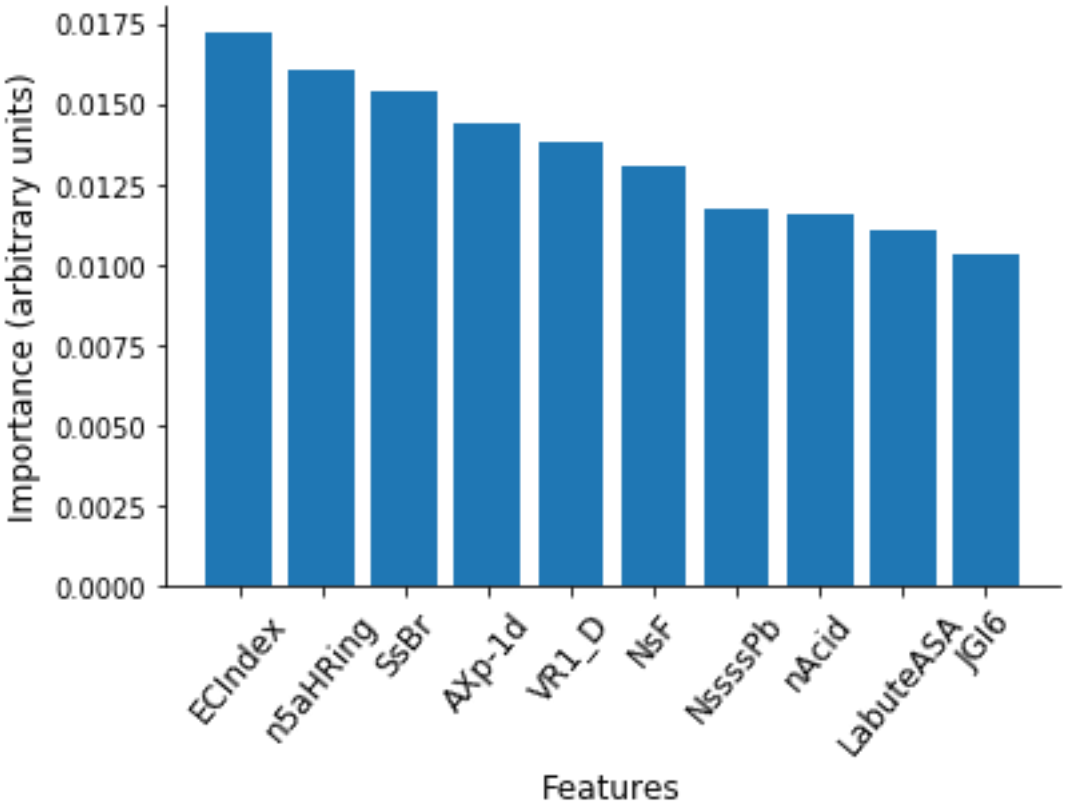


Figure 4. The top 10 most influential molecular features in determining excipients' effects on the growth of probiotic *L. paracasei*. When combined, importance of all 1613 features within the model were equal to 1.00.

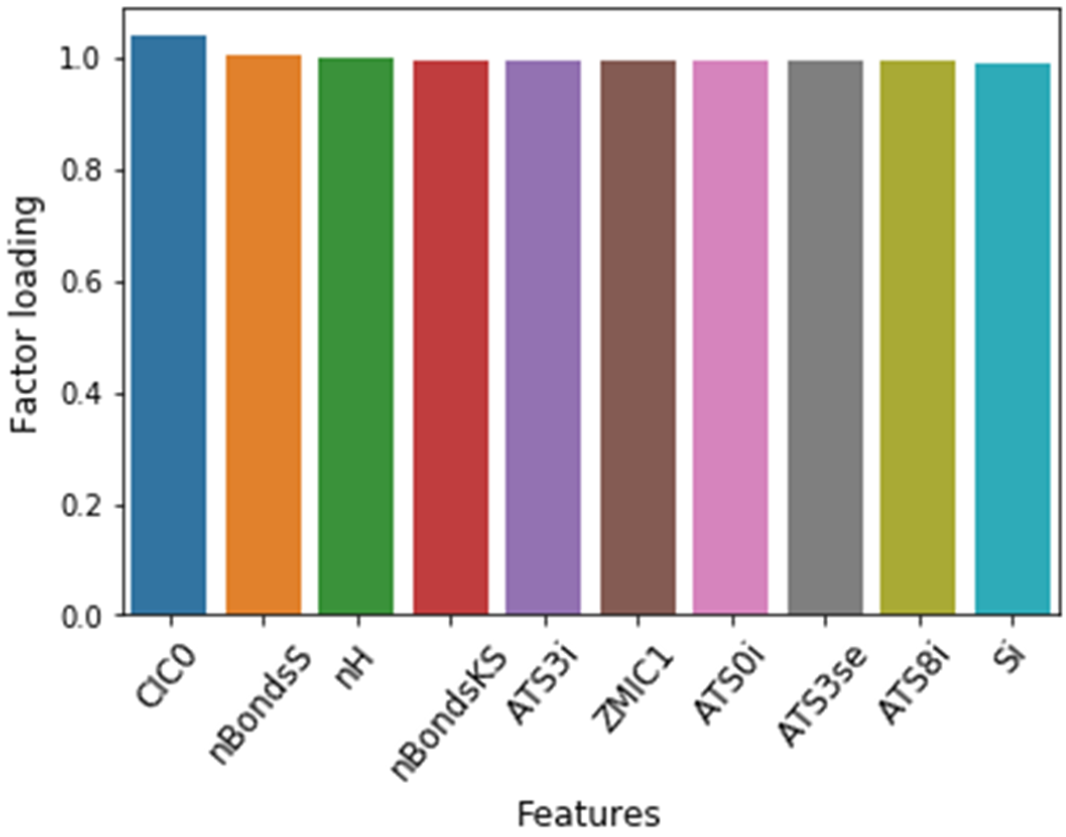
Table 3. The eccentric connectivity index (ECIndex) values of the 9 labelled excipients and antibiotic control, doxycycline hydrochloride, in correspondence to their effect on *L. paracasei* growth.

Excipient	**ECIndex**	**Effect on *L. paracasei* growth**
Doxycycline hydrochloride	-1.60×10^9	Prevent
Acetic anhydride	38	Prevent
Mannitol	114	Neutral
Aspartame	352	Neutral
Sucrose	367	Neutral
β-carotene	1642	Promote
Polysorbate 80	1953	Neutral
Cysteine hydrochloride	1.20×10^9	Prevent
Sodium benzoate	1.80×10^9	Prevent

Guar	2.11×10^9	Prevent

In comparison, Figure 5 shows the most influential features determined by PCA. Here, the features with the highest factor loadings within PC1 are depicted. PC1 accounted for more explained feature variance than any other PC (PC1 explained variance = 31.33%), thus its most determining features can be taken as those containing the most importance for the model's learning. Figure 5A demonstrates that 0-ordered complementary information content (CIC0) was the molecular feature with the highest positive correlation (+ 1.04), and Figure 5B shows that the averaged moreau-broto autocorrelation of lag 5 weighted by ionization potential (AATS5i) was the feature with the highest negative correlation (- 1.01). Whilst both features were almost equal in importance for the model (based on their similar loading magnitudes), their opposite correlations reveal that they are inversely related. For example, an excipient with a large CIC0 value is likely to have a low AATS5i value, and vice versa. CIC0 is a numerical indicator of molecular complexity with a long history of use in biological structure-activity studies; it is computationally calculated by partitioning a chemical graph into subsets (Roy et al., 1984). In the present dataset, excipients with CIC0 values < 3.23 always prevented *L. paracasei* growth. AATS5i is an autocorrelation descriptor that measures several topological features within a chemical structure (Ong et al., 2007). When examining the dataset, it was not visibly clear how AATS5i value affected excipient-probiotic interactions. This indicates that the descriptor's relationship to other features provided important information to the model. Interestingly, the features determined as the most influential by the random forest method were different to those calculated by the PCA method. This demonstrates the different learning styles of the techniques and supports a multi-method approach to feature importance to gain a broader appreciation of model learning.

A


Factor loading
1.0
0.8
0.6
0.4
0.2
0.0
ClC0
nBondsS
nH
nBondsKS
ATS3i
ZMIC1
ATS0i
ATS3se
ATS8i
Si
Features

B

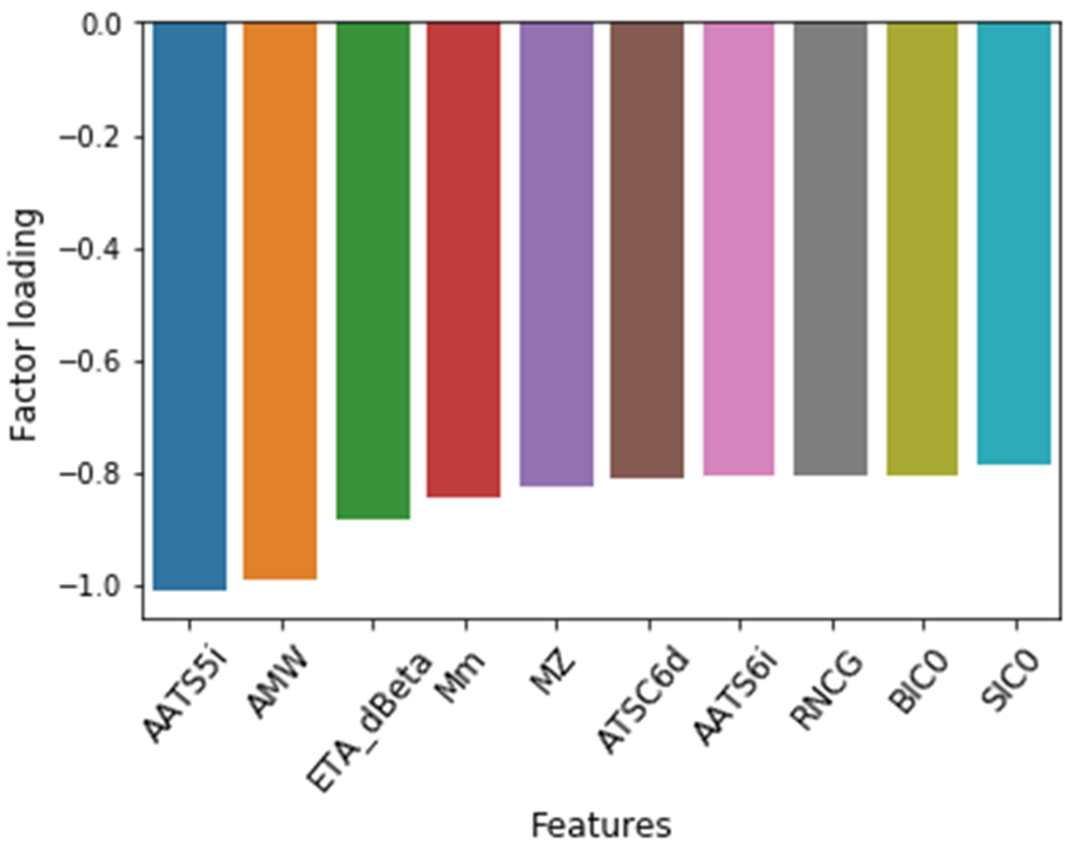

Factor loading
0.0
−0.2
−0.4
−0.6
−0.8
−1.0
AATS5i
AMW
ETA_dBeta
Mm
MZ
ATSC6d
AATS6i
RNCG
BIC0
SIC0
Features

Figure 5. Factor loadings of molecular features with A: the highest positive correlations, and B: the highest negative correlations, within principal component 1, after principal component analysis had been applied to measure the explained variance of excipients' features.

The key research question addressed by this study was whether there is a need to explore the effect of excipients on the growth of probiotics. There are many unanswered questions when it comes to the effects of the microbiota on the therapeutics, with even fewer questions regarding excipient-probiotic interactions. Recent work has revealed excipients can be active, whilst work by Kluneman et al. has revealed that metabolites such as glycoholic acid, a known excipient, can influence microbial cross-feeding and induce strain growth (Klünemann et al., 2021). Such recent work clearly highlights a need for a more thorough investigation of excipients' activity. Such a task could take decades to address, hence the need for ML.

From a small dataset it was possible to map the effects of 111 excipients on one probiotic strain, due to the ability of ML to comprehend over 1600 descriptors and provide the statistical likelihood for untested excipients. The subsequent in vitro study revealed that the current ML approach was able to accurately predict three out of the four untested excipients that were tested, while PCA highlighted an inherent pattern in the relationship between excipient activity (Figure 3). Without ML, the effects of excipients on probiotic proliferation could be highly complex to elucidate, as effects will likely be based on strain-specific sensitivity to excipient structure. Thus, ML correctly predicting 75% of the tested excipients provides a useful adjunct to the insight of formulation scientists.. In particular, the present study demonstrated the usefulness of active ML at early-stage of development when working with small data.

The model was used to address the standalone task of excipient-probiotic interactions but it is also envisaged to be combined with other ML models for a wider pipeline, such as a ML model for predicting whether drugs will impair bacterial growth (McCoubrey et al., 2021b). Merging the two models could help in understanding the combinatorial effect of individual materials in a formulation, leading to a more precise formulation development strategy for each individual patient. One current limitation of the use of ML for predicting excipient activity is the representation of excipients (i.e., featurisation). While the drug discovery community has successfully demonstrated that molecular descriptors provide adequate feature representation for small molecules, there is currently no initiative for polymeric excipients, whose polymeric properties, such as chain length, influence their properties. Currently, there is no standardised approach to representing such polymeric information for ML model development. Thus, future work will seek to find the optimal representation of polymeric excipients.

Conclusions

In this study active ML was applied to predict the effect of pharmaceutical excipients on the growth of a common probiotic, *L. paracasei*. An initial dataset based on 6 excipient-bacteria interactions was leveraged to predict the effects of a further 111 untested excipients on probiotic growth. The active learning process achieved an overall predictive certainty of 67.70% (± 9.30) for the final model and enabled the correct prediction of 3/4 excipient effects. These results were achieved after just 3 uncertainty sampling queries. Using the final model, the most important features determining excipient effects were calculated using random forest and PCA methods. Results show that excipients can exert significant effects on probiotic proliferation, and thus oral delivery of precision probiotics should be a considered process to optimise intestinal microbial colonisation and subsequent therapeutic benefits. This study is the first to apply active ML to microbiome science, and among the first within pharmaceutical sciences. Active ML is an ideal tool for harnessing the benefits of artificial intelligence when working with small datasets.

Funding: this work was funded by The Engineering and Physical Sciences Research Council [grant code EP/S023054/1.

Declarations of interest: the authors declare no competing interests associated with this work.

References

Aggarwal, N., Breedon, A.M.E., Davis, C.M., Hwang, I.Y., Chang, M.W., 2020. Engineering probiotics for therapeutic applications: recent examples and translational outlook. Curr Opin Biotechnol 65, 171-179.

Allegretti, J.R., Fischer, M., Sagi, S.V., Bohm, M.E., Fadda, H.M., Ranmal, S.R., Budree, S., Basit, A.W., Glettig, D.L., de la Serna, E.L., Gentile, A., Gerardin, Y., Timberlake, S., Sadovsky, R., Smith, M., Kassam, Z., 2019. Fecal Microbiota Transplantation Capsules with Targeted Colonic Versus Gastric Delivery in Recurrent Clostridium difficile Infection: A Comparative Cohort Analysis of High and Lose Dose. Digestive Diseases and Sciences 64, 1672-1678.

Cabadaj, M., Bashir, S., Haskins, D., Said, J., McCoubrey, L., Gaisford, S., Beezer, A., 2021. Kinetic analysis of microcalorimetric data derived from microbial growth: Basic theoretical, practical and industrial considerations. J Microbiol Methods, 106276.

Cammarota, G., Ianiro, G., Ahern, A., Carbone, C., Temko, A., Claesson, M.J., Gasbarrini, A., Tortora, G., 2020. Gut microbiome, big data and machine learning to promote precision medicine for cancer. Nature Reviews Gastroenterology & Hepatology 17, 635-648.

Chandrashekar, G., Sahin, F., 2014. A survey on feature selection methods. Computers & Electrical Engineering 40, 16-28.

Chen, C.L., Hsu, P.Y., Pan, T.M., 2019. Therapeutic effects of Lactobacillus paracasei subsp. paracasei NTU 101 powder on dextran sulfate sodium-induced colitis in mice. Journal of Food and Drug Analysis 27, 83-92.

Ding, W.K., Shah, N.P., 2007. Acid, bile, and heat tolerance of free and microencapsulated probiotic bacteria. J Food Sci 72, M446-450.

Dodoo, C.C., Wang, J., Basit, A.W., Stapleton, P., Gaisford, S., 2017. Targeted delivery of probiotics to enhance gastrointestinal stability and intestinal colonisation. International Journal of Pharmaceutics 530, 224-229.

Elbadawi, M., Gaisford, S., Basit, A.W., 2021. Advanced machine-learning techniques in drug discovery. Drug Discovery Today 26, 769-777.

Esbensen, K.H., Geladi, P., 2009. Principal Component Analysis: Concept, Geometrical Interpretation, Mathematical Background, Algorithms, History, Practice, Comprehensive Chemometrics, pp. 211-226.

Fredua-Agyeman, M., Gaisford, S., 2015. Comparative survival of commercial probiotic formulations: tests in biorelevant gastric fluids and real-time measurements using microcalorimetry. Benef Microbes 6, 141-151.

Garcia-Lozano, M., Haynes, J., Lopez-Ortiz, C., Natarajan, P., Peña-Garcia, Y., Nimmakayala, P., Stommel, J., Alaparthi, S.B., Sirbu, C., Balagurusamy, N., Reddy, U.K., 2020. Effect of pepper-containing diets on the diversity and composition of gut microbiome of drosophila melanogaster. International Journal of Molecular Sciences 21.

Ghyselinck, J., Verstrepen, L., Moens, F., Van Den Abbeele, P., Bruggeman, A., Said, J., Smith, B., Barker, L.A., Jordan, C., Leta, V., Chaudhuri, K.R., Basit, A.W., Gaisford, S., 2021. Influence of probiotic bacteria on gut microbiota composition and gut wall function in an in-vitro model in patients with Parkinson's disease. International Journal of Pharmaceutics: X.

Hill, C., Guarner, F., Reid, G., Gibson, G.R., Merenstein, D.J., Pot, B., Morelli, L., Canani, R.B., Flint, H.J., Salminen, S., Calder, P.C., Sanders, M.E., 2014. Expert consensus document. The International Scientific Association for Probiotics and Prebiotics consensus statement on the scope and appropriate use of the term probiotic. Nat Rev Gastroenterol Hepatol 11, 506-514.

Horvath, T.D.a.P., 2018. modAL: A modular active learning framework for Python.

Janssens, Y., Nielandt, J., Bronselaer, A., Debunne, N., Verbeke, F., Wynendaele, E., Van Immerseel, F., Vandewynckel, Y.P., De Tré, G., De Spiegeleer, B., 2018. Disbiome database: linking the microbiome to disease. BMC Microbiol 18, 50.

Kapoor, M.P., Koido, M., Kawaguchi, M., Timm, D., Ozeki, M., Yamada, M., Mitsuya, T., Okubo, T., 2020. Lifestyle related changes with partially hydrolyzed guar gum dietary fiber in healthy athlete individuals – A randomized, double-blind, crossover, placebo-controlled gut microbiome clinical study. Journal of Functional Foods 72, 104067.
Kim, N., Yun, M., Oh, Y.J., Choi, H.J., 2018. Mind-altering with the gut: Modulation of the gut-brain axis with probiotics. J Microbiol 56, 172-182.
Kim, W.K., Jang, Y.J., Han, D.H., Jeon, K., Lee, C., Han, H.S., Ko, G., 2020. Lactobacillus paracasei KBL382 administration attenuates atopic dermatitis by modulating immune response and gut microbiota. Gut Microbes 12, 1-14.
Klayraung, S., Viernstein, H., Okonogi, S., 2009. Development of tablets containing probiotics: Effects of formulation and processing parameters on bacterial viability. International Journal of Pharmaceutics 370, 54-60.
Klünemann, M., Andrejev, S., Blasche, S., Mateus, A., Phapale, P., Devendran, S., Vappiani, J., Simon, B., Scott, T.A., Kafkia, E., Konstantinidis, D., Zirngibl, K., Mastrorilli, E., Banzhaf, M., Mackmull, M.-T., Hövelmann, F., Nesme, L., Brochado, A.R., Maier, L., Bock, T., Periwal, V., Kumar, M., Kim, Y., Tramontano, M., Schultz, C., Beck, M., Hennig, J., Zimmermann, M., Sévin, D.C., Cabreiro, F., Savitski, M.M., Bork, P., Typas, A., Patil, K.R., 2021. Bioaccumulation of therapeutic drugs by human gut bacteria. Nature 597, 533-538.
Liao, N., Pang, B., Jin, H., Xu, X., Yan, L., Li, H., Shao, D., Shi, J., 2020. Potential of lactic acid bacteria derived polysaccharides for the delivery and controlled release of oral probiotics. J Control Release 323, 110-124.
Liu, Y., Liu, B., Li, D., Hu, Y., Zhao, L., Zhang, M., Ge, S., Pang, J., Li, Y., Wang, R., Wang, P., Huang, Y., Huang, J., Bai, J., Ren, F., Li, Y., 2020. Improved Gastric Acid Resistance and Adhesive Colonization of Probiotics by Mucoadhesive and Intestinal Targeted Konjac Glucomannan Microspheres. Advanced Functional Materials 30.
Maier, L., Goemans, C.V., Wirbel, J., Kuhn, M., Eberl, C., Pruteanu, M., Muller, P., Garcia-Santamarina, S., Cacace, E., Zhang, B., Gekeler, C., Banerjee, T., Anderson, E.E., Milanese, A., Lober, U., Forslund, S.K., Patil, K.R., Zimmermann, M., Stecher, B., Zeller, G., Bork, P., Typas, A., 2021. Unravelling the collateral damage of antibiotics on gut bacteria. Nature.
Maier, L., Pruteanu, M., Kuhn, M., Zeller, G., Telzerow, A., Anderson, E.E., Brochado, A.R., Fernandez, K.C., Dose, H., Mori, H., Patil, K.R., Bork, P., Typas, A., 2018. Extensive impact of non-antibiotic drugs on human gut bacteria. Nature 555, 623-628.
Marcos-Fernández, R., Ruiz, L., Blanco-Míguez, A., Margolles, A., Sánchez, B., 2021. Precision modification of the human gut microbiota targeting surface-associated proteins. Scientific Reports 11, 1270.
Martinez-Guryn, K., Leone, V., Chang, E.B., 2019. Regional Diversity of the Gastrointestinal Microbiome. Cell Host Microbe 26, 314-324.
McConnell, E.L., Fadda, H.M., Basit, A.W., 2008. Gut instincts: explorations in intestinal physiology and drug delivery. Int J Pharm 364, 213-226.
McCoubrey, L.E., Elbadawi, M., Orlu, M., Gaisford, S., Basit, A.W., 2021a. Harnessing machine learning for development of microbiome therapeutics. Gut Microbes 13, 1-20.
McCoubrey, L.E., Elbadawi, M., Orlu, M., Gaisford, S., Basit, A.W., 2021b. Machine Learning Uncovers Adverse Drug Effects on Intestinal Bacteria. Pharmaceutics 13.
McCoubrey, L.E., Gaisford, S., Orlu, M., Basit, A.W., 2021c. Predicting drug-microbiome interactions with machine learning. Biotechnol Adv, 107797.
Moens, F., Van den Abbeele, P., Basit, A.W., Dodoo, C., Chatterjee, R., Smith, B., Gaisford, S., 2019. A four-strain probiotic exerts positive immunomodulatory effects by enhancing colonic butyrate production in vitro. International Journal of Pharmaceutics 555, 1-10.

Morita, Y., Jounai, K., Sakamoto, A., Tomita, Y., Sugihara, Y., Suzuki, H., Ohshio, K., Otake, M., Fujiwara, D., Kanauchi, O., Maruyama, M., 2018. Long-term intake of Lactobacillus paracasei KW3110 prevents agerelated chronic inflammation and retinal cell loss in physiologically aged mice. Aging 10, 2723-2740.
Moriwaki, H., Tian, Y.-S., Kawashita, N., Takagi, T., 2018. Mordred: a molecular descriptor calculator. Journal of cheminformatics 10, 4-4.
O'Toole, P.W., Marchesi, J.R., Hill, C., 2017. Next-generation probiotics: the spectrum from probiotics to live biotherapeutics. Nature Microbiology 2, 17057.
Ong, S.A.K., Lin, H.H., Chen, Y.Z., Li, Z.R., Cao, Z., 2007. Efficacy of different protein descriptors in predicting protein functional families. BMC Bioinformatics 8, 300.
Proctor, L.M., Creasy, H.H., Fettweis, J.M., Lloyd-Price, J., Mahurkar, A., Zhou, W., Buck, G.A., Snyder, M.P., Strauss, J.F., Weinstock, G.M., White, O., Huttenhower, C., The Integrative, H.M.P.R.N.C., 2019. The Integrative Human Microbiome Project. Nature 569, 641-648.
Raddatz, G.C., Poletto, G., Deus, C., Codevilla, C.F., Cichoski, A.J., Jacob-Lopes, E., Muller, E.I., Flores, E.M.M., Esmerino, E.A., de Menezes, C.R., 2020. Use of prebiotic sources to increase probiotic viability in pectin microparticles obtained by emulsification/internal gelation followed by freeze-drying. Food Res Int 130, 108902.
Reker, D., 2019. Practical considerations for active machine learning in drug discovery. Drug Discovery Today: Technologies 32-33, 73-79.
Reker, D., Schneider, G., 2015. Active-learning strategies in computer-assisted drug discovery. Drug Discovery Today 20, 458-465.
Reker, D., Shi, Y., Kirtane, A.R., Hess, K., Zhong, G.J., Crane, E., Lin, C.H., Langer, R., Traverso, G., 2020. Machine Learning Uncovers Food- and Excipient-Drug Interactions. Cell Rep 30, 3710-3716 e3714.
Rowe, R.C., Sheskey, P.J., Quinn, M.E., 2009. Handbook of Pharmaceutical Excipients, Sixth ed. Pharmaceutical Press and the American Pharmacists Association.
Roy, A.B., Basak, S.C., Harriss, D.K., Magnuson, V.R., 1984. NEIGHBORHOOD COMPLEXITIES AND SYMMETRY OF CHEMICAL GRAPHS AND THEIR BIOLOGICAL APPLICATIONS, in: Avula, X.J.R., Kalman, R.E., Liapis, A.I., Rodin, E.Y. (Eds.), Mathematical Modelling in Science and Technology. Pergamon, pp. 745-750.
Said, J., Dodoo, C.C., Walker, M., Parsons, D., Stapleton, P., Beezer, A.E., Gaisford, S., 2014. An in vitro test of the efficacy of silver-containing wound dressings against Staphylococcus aureus and Pseudomonas aeruginosa in simulated wound fluid. Int J Pharm 462, 123-128.
Schmidt, J., Marques, M.R.G., Botti, S., Marques, M.A.L., 2019. Recent advances and applications of machine learning in solid-state materials science. npj Computational Materials 5, 83.
Sharma, V., Goswami, R., Madan, A.K., 1997. Eccentric Connectivity Index: A Novel Highly Discriminating Topological Descriptor for Structure−Property and Structure−Activity Studies. Journal of Chemical Information and Computer Sciences 37, 273-282.
Singer-Englar, T., Barlow, G., Mathur, R., 2019. Obesity, diabetes, and the gut microbiome: an updated review. Expert Rev Gastroenterol Hepatol 13, 3-15.
So, D., Whelan, K., Rossi, M., Morrison, M., Holtmann, G., Kelly, J., Shanahan, E., Staudacher, H., Campbell, K., 2018. Dietary fiber intervention on gut microbiota composition in healthy adults: A systematic review and meta-analysis. The American journal of clinical nutrition 107.
Sreeja, V., Prajapati, J., Thakkar, V., Gandhi, T., Darji, V., 2016. Effect of excipients on disintegration, viability and activity of fast disintegrating tablets containing probiotic and starter cultures. Current Trends in Biotechnology and Pharmacy 10, 108-117.
Suez, J., Zmora, N., Segal, E., Elinav, E., 2019. The pros, cons, and many unknowns of probiotics. Nature Medicine 25, 716-729.

Taguchi, Y.H., Iwadate, M., Umeyama, H., 2015. Principal component analysis-based unsupervised feature extraction applied to in silico drug discovery for posttraumatic stress disorder-mediated heart disease. BMC Bioinformatics 16.
U.S. Food and Drug Administration, F., 2020. Inactive Ingredient Search for Approved Drug Products, July 28, 2020 ed. Office of Pharmaceutical Quality, U.S. Food and Drug Administration, Maryland, U.S.A.
Varum, F., Cristina Freire, A., Bravo, R., Basit, A.W., 2020a. OPTICORE, an innovative and accurate colonic targeting technology. International Journal of Pharmaceutics 583, 119372.
Varum, F., Cristina Freire, A., Fadda, H.M., Bravo, R., Basit, A.W., 2020b. A dual pH and microbiota-triggered coating (Phloral(TM)) for fail-safe colonic drug release. International Journal of Pharmaceutics 583, 119379.
Veiga, P., Suez, J., Derrien, M., Elinav, E., 2020. Moving from probiotics to precision probiotics. Nat Microbiol 5, 878-880.
Weininger, D., 1988. SMILES, a Chemical Language and Information System: 1: Introduction to Methodology and Encoding Rules. Journal of Chemical Information and Computer Sciences 28, 31-36.
Westfall, S., Carracci, F., Estill, M., Zhao, D., Wu, Q.L., Shen, L., Simon, J., Pasinetti, G.M., 2021. Optimization of probiotic therapeutics using machine learning in an artificial human gastrointestinal tract. Sci Rep 11, 1067.
Wilkinson, J.E., Franzosa, E.A., Everett, C., Li, C., Hcmph researchers., Hcmph trainees., Hcmph investigators., Hu, F.B., Wirth, D.F., Song, M., Chan, A.T., Rimm, E., Garrett, W.S., Huttenhower, C., 2021. A framework for microbiome science in public health. Nat Med.
World Health Organization, W., 2020. The top 10 causes of death, in: WHO (Ed.), Online.
Yu, Y., Dunaway, S., Champer, J., Kim, J., Alikhan, A., 2020. Changing our microbiome: probiotics in dermatology. Br J Dermatol 182, 39-46.