



OPEN

## Using least angular regression to model the antibacterial potential of metronidazole complexes

Tahir Mehmood<sup>1</sup>✉, Mudassir Iqbal<sup>2</sup> & Bushra Rafique<sup>2</sup>

Imidazole has anti-inflammatory, antituberculous, antimicrobial, antimycotic, antiviral, and antitumor properties in the human body, to name a few. Metronidazole [1-(2-Hydroxyethyl)-2-methyl-5-nitroimidazole] is a widely used antiprotozoan and antibacterial medication. Using fourier transform infrared spectroscopy, the current study models the antibacterial activity of already synthesised Metronidazole (MTZ) complexes (MTZ – Benz, MTZ – Benz – Cu, MTZ – Benz – Cu – Cl<sub>2</sub>CHCOOH, MTZ, MTZ – Cu, MTZ – Cu – Cl<sub>2</sub>CHCOOH, MTZ – Benz – Ag, MTZ – Benz – Ag – Cl<sub>2</sub>CHCOOH, MTZ – Ag and MTZ – Ag – Cl<sub>2</sub>CHCOOH) against *E. coli*, *B. bronceptica*, *S. epidermidis*, *B. pumilus* and *S. aureus*. To characterise the Metronidazole complexes for antibacterial activity against 05 microbes, the least angular regression and least absolute shrinkage selection operators were used. Asymmetric Least Squares was used to correct the spectrum baseline. Least angular regression outperforms cross-validated root mean square error in the fitted models. Using Least angular regression, influential wavelengths that explain the variation in antibacterial activity of Metronidazole complexes were identified and mapped against functional groups.

The antibacterial agent metronidazole (MTZ) [1-(2-Hydroxyethyl)-2-methyl-5-nitroimidazole] is well known. Nitroimidazoles' biological action mechanism is based on nitro group reduction, which generates intermediate species that interact with DNA, oxidising it and liberating thymidine phosphate, resulting in a lesion characterised by double helix destabilisation and destruction<sup>1</sup>. The delivery of a specific and defined amount of drug to the human body via the gastrointestinal system can be accomplished in various dosage forms, such as oral tablet and suspension, among others. Various studies on the bioavailability of drugs synthesised by various manufacturers revealed that tablets and suspensions containing the same drug and drug content could not provide the same therapeutic response. The different parameters responsible for the dissolution profile and therapeutic effect variations are formulation, additives, physical form, manufacturing process, and so on. Metronidazole benzoate suspension can be used to treat infections caused by a variety of anaerobic bacteria, protozoa, and bacteroides. Metronidazole benzoate is frequently substituted for Metronidazole in paediatric oral preparations due to the blended flavour of the ester compared to the bitter taste of the free base<sup>2</sup>. Despite research into various ester structures, metronidazole benzoate is the only ester prodrug currently on the market. Metronidazole benzoate has demonstrated higher antimicrobial efficacy, for example, it is 20 times more effective against *Clostridium perfringens in vitro*<sup>3</sup>. However, because it is almost completely insoluble in liquids, its solubility in water bodies is insufficient for injections. Solubility is a major challenge for formulation scientists. To improve the solubility of poorly soluble products, physical and chemical modifications of the material, as well as other processes such as particle size reduction, crystal engineering, salt forming, stable dispersion, surfactant application, and complexation, are all used. In this analysis, copper complexes were used to solve the solubility problem. Metal chelates had a variety of biological functions, including anticancer, antidiabetic, antibacterial, antifungal, and antiviral properties. Metal chelates have been shown to have significantly greater antimicrobial activity than chelating agents<sup>4</sup>. Copper acetate monohydrate can react with monomeric ligands to form an adduct by replacing the water molecules. With the split of the [Cu<sub>2</sub>(acetate)<sub>4</sub>(H<sub>2</sub>O)<sub>2</sub>] dimer, it is even possible to lose an acetate group to allow the coordination of other ligands, or even shift the coordination of the acetate group to monodentate coordination, as with the imidazole ligand<sup>5</sup>. Antitumor, superoxide dismutase, and catecholase functions have been demonstrated for binuclear and mononuclear copper(II) carboxylate adducts with imidazole-type ligands<sup>6</sup>. These structural and electronic factors, in addition to the dependence of the structure of copper(II) carboxylates on the electronic and steric properties of added base, can influence the properties of copper(II) complexes as

<sup>1</sup>School of Natural Sciences, National University of Sciences and Technology (NUST), Islamabad, Pakistan. <sup>2</sup>Department of Chemistry, School of Natural Sciences, National University of Sciences and Technology (NUST), Islamabad, Pakistan. ✉email: tahime@gmail.com

metalloprotein models. Nitroimidazoles are used in the prevention of bacterial infections as chemotherapeutic agents and radiosensitizers<sup>7</sup>.

The Cu complexes of metronidazole form a dimer in the presence of acetate due to a Cu-Cu interaction and four bridging 2-acetate ligands. Complexation improves Metronidazole's anti-amoebic efficacy in vitro and in vivo samples, according to emphin vitro and in vivo samples<sup>8</sup>. The aliphatic carboxylic acid dichloroacetic acid contains only one carboxylate group (DCA). The carboxylate group in DCA can coordinate in either a monodentate or bidentate mode, according to research. Chelation can occur in bidentate or bridged systems. The researchers are interested in the synthesis, spectroscopic, computational, antimicrobial, and antidiabetic analyses of the reference drugs Metronidazole and Metronidazole benzoate with copper complexes<sup>9</sup>.

Fourier Transformation Infrared (FTIR) spectroscopy is used to evaluate the efficiency of Metronidazole complexes because it is simple, fast, and precise<sup>10</sup>. FTIR has been used in a variety of research projects. It can, for example, detect Metronidazole in environmental samples<sup>11</sup> and teratogenic effects of Metronidazole on the uterus of mice<sup>12</sup>. FTIR assist in capturing the phenotypic screening bioassay for *Escherichia coli* stress and antibiotic responses<sup>13</sup>. Metronidazole's chemical properties and antibacterial activity are characterised by FTIR using adhesive dopamine<sup>14</sup>.

When modelling Metronidazole antibacterial behaviour with FTIR spectral data, the sample size is much smaller than the wavelength. As a result, the data is highly dimensional (i.e. number of explanatory variables). In this case, the least absolute shrinkage and selection operator are used (Lasso)<sup>15</sup> and least angle regression (LAR)<sup>16</sup> are considered a viable option. The regression based algorithm are used for nondestructive pharmaceutical preparation discrimination<sup>17</sup>, for modeling the antibacterial activity of ionic liquids<sup>18,19</sup>, in a semi-solid matrix for multivariate analysis of nystatin and Metronidazole<sup>20</sup>, in tablet formulations for simultaneous quantitative assessment of paracetamol and tramadol<sup>21</sup>, for determining the active pharmaceutical ingredient in transdermal gel formulations on a quantitative basis<sup>22</sup>, for predicting effective dose as a cytotoxicity biomarker<sup>23</sup>, for antibiotic discovery using metabolic fingerprinting and a screening assay<sup>24</sup>, for determining the concentrations of active ingredients in semi-solid pharmaceutical formulations<sup>25</sup>, for orally disintegrating tablet template formulation and non-destructive methods of assessment<sup>26</sup>, for anti-*Helicobacter pylori* drug screening<sup>27</sup> and for the analysis of a binary amoxicillin-flucloxacillin mixture<sup>28</sup>.

The current study's key goal is to identify functional groups for characterization of already synthesized Metronidazole complexes in terms of antibacterial activity. For this least absolute shrinkage and selection operator (Lasso)<sup>15</sup> and least angle regression (LAR)<sup>16</sup> are considered. FTIR spectrum was used to model the Metronidazole's antibacterial activity, which was tested against bacterial strains *Escherichia coli* (*E. coli*), *Bordetella bronchiseptica* (*B. bronchiseptica*), *Staphylococcus epidermidis* (*S. epidermidis*), *Bacillus pumilus* (*B. pumilus*), and *Staphylococcus aureus* (*S. aureus*).

## Materials and methods

**Antibacterial activity of metronidazole.** The data set is taken from the study<sup>29</sup> where following Metronidazole complexes were synthesized (*MTZ – Benz*, *MTZ – Benz – Cu*, *MTZ – Benz – Cu – Cl<sub>2</sub>CHCOOH*, *MTZ – Cu*, *MTZ – Cu – Cl<sub>2</sub>CHCOOH*, *MTZ – Benz – Ag*, *MTZ – Benz – Ag – Cl<sub>2</sub>CHCOOH*, *MTZ – Ag* and *MTZ – Ag – Cl<sub>2</sub>CHCOOH*). The antibacterial activity of above synthesized complexes was measured against five bacteria i.e. *E. coli*, *B. bronchiseptica*, *S. epidermidis*, *B. pumilus* and *S. aureus*. For this Agar well diffusion process in accordance with the National Committee for Clinical Laboratory Practices' guidelines (M07-A8) was used. The biological effects against Gram-positive and Gram-negative bacteria were examined for the screening of synthesized metal complexes for antimicrobial activity. Gram-positive strains: *S. aureus* ATCC 6538, *B. pumilus* ATCC 14884 and *S. epidermidis* ATCC 12228. Gram negative bacteria: *B. bronchiseptica* ATCC 4617, *E. coli* ATCC 8739. Microorganism cultures were grown on sterilized nutrient agar as a substrate. Surface in which microbes were grown was washed with 0.9 % saline solution to detach and obtained fresh microbial culture for further use. More over sterilization was done using autoclave and the solution was simply poured to detach from the surface microbial culture. The test sample stock solution was made in ethanol at a concentration of 1 mg/ml. Metronidazole complexes samples. The Metronidazole complexes comprising of different concentration were poured in the agar well to determine the antimicrobial activity. Initially plates were prepared containing media with microbial culture. The agar was allowed to solidify then well are punched. Afterwards, Metronidazole complexes samples dilutions were introduced in the wells. The plates were then incubated for 24 hours at 37°C. The antimicrobial activities were assessed using a zone reader to measure the zone of antibacterial (mm) (Proto-COL3).

**Fourier transformation infrared spectroscopic experiment.** Attenuated Total Reflection-Fourier Transformation Infrared spectroscopy was used to conduct infrared spectral measurements. The infrared spectra of the ligands and complexes were measured in solid state as KBr pellets on a PerkinElmer Spectrum 100 FT-IR Spectrophotometer in the frequency range 4000–400  $cm^{-1}$ . It has a regular high linearity room temperature detector and a UATR Diamond ATR (Single Reflection). For each wavelength, 10 scans were collected, with 4  $cm^{-1}$  being the most recent resolution and 1  $cm/s$  scan speed, as used in<sup>10</sup>. Prior to each experiment, background spectra were obtained. The data matrix  $X_{10 \times 1676}$  is the outcome of a spectroscopic experiment with ionic liquid samples.

**Data acquisition and statistical software.** The data collection studies are being carried out at the National University of Sciences and Technology (NUST) in Islamabad, Pakistan's School of Natural Sciences (SNS). R is used for both computations and mathematical modelling. <https://www.R-project.org/>.

| Compound                                 | <i>E. coli</i> | <i>B. bronceptica</i> | <i>S. epidermidis</i> | <i>B. pumilus</i> | <i>S. aureus</i> |
|--|----------------|-----------------------|-----------------------|-------------------|------------------|
| MTZ – Benz                               | 22.09          | 19.37                 | 25.09                 | 25.93             | 18.29            |
| MTZ – Benz – Cu                          | 20.25          | 19.01                 | 22.90                 | 21.47             | 13.22            |
| MTZ – Benz – Cu – Cl <sub>2</sub> CHCOOH | 21.81          | 22.77                 | 23.72                 | 20.87             | 16.01            |
| MTZ                                      | 15.72          | 17.72                 | 14.32                 | 13.64             | 10.09            |
| MTZ – Cu                                 | 16.09          | 19.09                 | 15.51                 | 14.38             | 14.72            |
| MTZ – Cu – Cl <sub>2</sub> CHCOOH        | 16.90          | 19.90                 | 16.36                 | 16.98             | 14.77            |
| MTZ – Benz – Ag                          | 20             | 19                    | 21                    | 24                | 17               |
| MTZ – Benz – Ag – Cl <sub>2</sub> CHCOOH | 19             | 20                    | 22                    | 22                | 16               |
| MTZ – Ag                                 | 18             | 18                    | 23                    | 20                | 15               |
| MTZ – Ag – Cl <sub>2</sub> CHCOOH        | 17             | 17                    | 21                    | 23                | 14               |

**Table 1.** Antibacterial activity against the bacteria under consideration is presented.

**Baseline correction.** Baseline correction is part of the data preprocessing. FTIR spectra are supposed to have a straight line at 0, but this isn't always the case due to linear or nonlinear additions. For this, asymmetric least squares (ALS) is used. For weighting down predictor variables with substantial error, it employs the least square approach and smoothing by adding a 2<sup>nd</sup> derivative restriction<sup>18,30</sup>.  $S = \sum w_i(x_i - b_i)^2 + \lambda \sum (\Delta^2 b_i)$

where  $x_i$  is initial spectrum,  $b_i$  computed baseline,  $w_i$  is asymmetric residual weights,  $\Delta^2$  is the approximate baseline's second derivative. There are two parameters to tune in the algorithm: The smoothing parameter is  $\lambda$  and  $w$  is the weight. Lillard's proposed objective procedure is included in this case<sup>31</sup>.

**Predicting antibacterial activity.** We used the least absolute shrinkage and selection operator, as well as the least angular regression, to predict the antibacterial activity of synthesised Metronidazole complexes based on FTIR spectrum data. The specifics are provided below.

*Least absolute shrinkage and selection operator (Lasso).* Lasso<sup>15</sup> is potential modeling candidate. This model employs the coordinate descent algorithm with a penalty parameter known as *theta*. Several levels are considered for computational purposes, and the optimal choice of these penalty parameters is obtained through cross validation. The partial log-likelihood function defined by is used to penalise the multivariate regression coefficient  $\beta$ .

$$L(\beta) - \sum_{j=1}^p p_{\theta^*}(|\beta_j|) \quad (1)$$

where  $L(\beta)$  denotes the partial log-likelihood for  $n$  samples.  $p_{\theta}(|\cdot|)$  presents the penalty function.

In Lasso  $L_1$ -penalized model is used, which is defined as

$$p_{\theta}(|\beta|) = \theta|\beta| \quad (2)$$

LASSO do the variable selection by equating non significant variable's coefficients to zero. Here  $\theta_1$  is weight for  $L_1$  penalty Small  $\theta_1$  level will result large number of variables and vice versa.

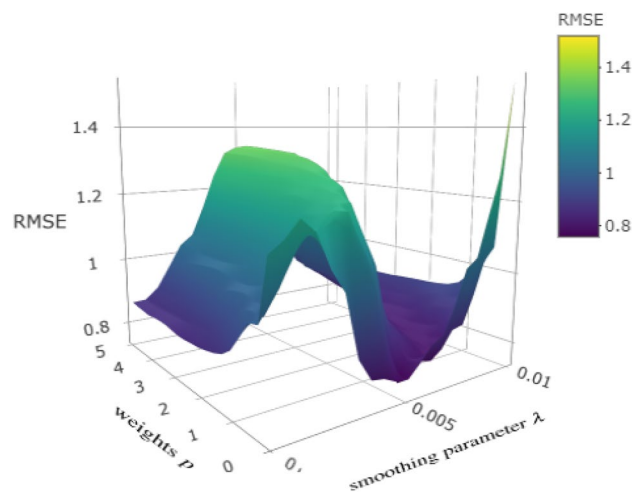
*Least angular regression (LAR).* The democratic equivalent of forward step-wise Lasso regression is LAR<sup>16</sup>. Just as much of an indicator as LAR requires is entered. The procedure is defined as follows:

1. Start with  $r = y, \beta_1, \beta_2, \dots, \beta_p = 0$ . Assume  $x_l$  standardized.
2. Find explanatory variable  $x_l$  most associated  $r$ .
3. Increase  $\beta_l$  in the path of  $\text{sign}(\text{corr}(r, x_l))$  so far some other explanatory variable  $x_m$  has the same degree of correlation with the new residual as  $x_l$ .
4. Move  $(\beta_l, \beta_m)$  in the course of the joint least squares  $(x_l, x_m)$  until some other explanatory variable  $x_\ell$  has the same degree of correlation as the new residual.
5. Continue in this manner until you've hit all of the predictors. Stop when  $\text{corr}(r, x_l) = 0 \forall l$ , i.e. OLS solution.

LAR is a state of art algorithm to solve  $L_1$  regularized linear regression or logistic regression. It offers unified explanation of forward selection, stage wise regression and Lasso.

## Results and discussion

Calculations are performed for both Metronidazole as well as complexes. These complexes' antibacterial efficacy is calculated for five bacteria: *E. coli*, *B. bronceptica*, *S. epidermidis*, *B. pumilus* and *S. aureus*. Table 1 shows the mean, limit, minimum, and standard deviation of antibacterial activity against the microbes studied. Synthesized Metronidazole Complexes tend to have the most activity against *S. epidermidis*, antibacterial behaviour against has shown the most difference *B. pumilus* and has lowest antibacterial activity against *B. bronceptica*.

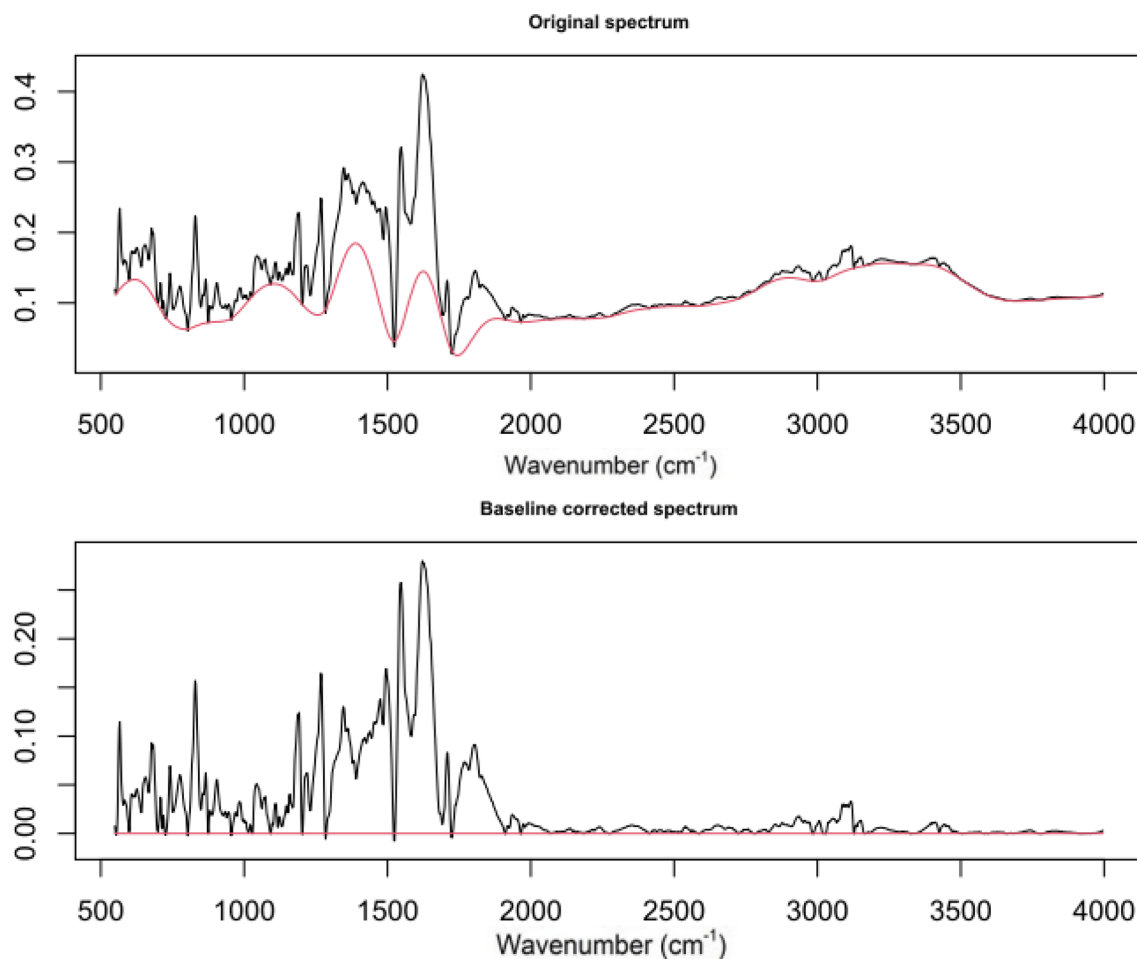


**Figure 1.** The model performance defined by RMSE of *S. aureus* with different levels of weights  $p$  on x-axis and smoothing parameter  $\lambda$  on y-axis is presented.

The ATR-FTIR spectroscopic analysis was carried out over the synthesized Metronidazole complexes samples. For identifying the influential spectrum peaks, FTIR spectrum baseline correction at zero is required. ALS is used for spectrum baseline correction. ALS has two parameters to tune, weight  $p$  and smoothing parameter  $\lambda$ . These parameters are optimized for each fitted antibacterial activity. For in-stance, the model performance defined by RMSE of *S. aureus* with different levels of weights and smoothing parameter is presented in Fig. 1. This indicates Metronidazole antibacterial activity against *S. aureus* is optimized with  $p = 0.010$  and  $\lambda = 4.100$ . The optimal tuned level of ALS parameters for all bacteria is presented in Table 2. By using the ALS optimal parameters for each microbe the spectrum baseline is corrected, which is further used in analysis. The comparison of original spectrum and baseline corrected spectrum obtained for *S. aureus* is presented in Fig. 2.

LASSO and LAR are used to model Metronidazole's antibacterial activity against the microbes in question. Regularized parameters  $\theta$  and fraction respectively determine the efficiency of LASSO and LAR. The change of these regularization parameters define the model prediction capability but also effects the number of wave-number selection. These parameters are optimized for prediction capability measured by RMSE in upper panel of Fig. 3 for *S. aureus*. For lasso, the optimal shrinkage parameter  $\theta$  is tuned to 1.5 and for LAR the optimal shrinkage parameter *fraction* is tuned to 0.02 which results the RMSE 2.14 and 1.91 respectively. For model fitting we have used bootstrap cross validation, at each fold the model RMSE was computed and is compared in lower panel of Fig. 3. The lower left panel indicates LAR has lower RMSE in most of the folds when compared to LASSO. Similar findings are found in lower right panel indicating LAR has better prediction capability of modeling antibacterial activity of Metronidazole against *S. aureus*. The distribution of ALS parameters which defines the baseline correction, the shrinkage parameters which defines the wave number selection and RMSE for LASSO and LAR is presented in Table 2. For prediction of antibacterial activity of Metronidazole for *E. coli*, LASSO uses  $\theta = 1.8$  to optimizes the RMSE= 2.38 and LAR uses *fraction* = 0.02 to optimizes RMSE= 2.32. For prediction of antibacterial activity of Metronidazole for *B. bronceptica*, LASSO uses  $\theta = 1.3$  to optimizes the RMSE= 1.97 and LAR uses *fraction* = 0.06 to optimizes RMSE= 1.91. For prediction of antibacterial activity of Metronidazole for *S. epidermidis*, LASSO uses  $\theta = 1.1$  to optimizes the RMSE= 2.33 and LAR uses *fraction* = 0.07 to optimizes RMSE= 2.01. For prediction of antibacterial activity of Metronidazole for *B. pumilus*, LASSO uses  $\theta = 1.6$  to optimizes the RMSE= 1.84 and LAR uses *fraction* = 0.01 to optimizes RMSE= 1.84. LAR is predicting the antibacterial activity of Metronidazole complexes as good as LASSO or in some cases it is better. Hence we have used LAR for extracting the influential wavenumbers. LAR computes the importance index for each variable that is wavenumber. Since LAR is defined by the shrinkage parameter *fraction*, see Table 2. We have extracted the wavenumbers by using the optimally tuned *fraction* in LAR for each microbial model. We have used repeated sampling based modeling that is bootstrap where the consistently selected LAR coefficients are used to identify the influential wave numbers. The influential wavenumbers are mapped for respective functional compound and group, and are presented for each microbes in Table 3.

For *E. coli* the influential wavenumbers are mapped to the functional compound  $C - H, CH_3, CH_2, CH, C = O$  and  $O - H$ , which respectively belong to Carboxylic Acids & Derivatives, Alkanes, and Carboxylic Acids & Derivatives functional group. For *B. bronceptica* the influential wavenumbers are mapped to the functional compound  $CH_2, CH_3, C = O$  and  $CH_3, CH_2, CH$  which respectively belong to Alkanes, Carboxylic Acids & Derivatives and Alkanes. For *S. epidermidis* the influential wavenumbers are mapped to the functional compound  $C \equiv C$  and  $O - H$ , which respectively belong to Alkenes and Carboxylic Acids & Derivatives functional group. For *B. pumilus* the influential wavenumbers are mapped to the functional compound  $O - H, O - C$  and  $CH_3, CH_2, CH$ , which respectively belong to Carboxylic Acids & Derivatives and Alkenes functional group. For *S. aureus* the influential wavenumbers are mapped to the functional compound  $O - H, = C - H, = CH_2$  and  $O - H$ , which respectively belong to Alcohols & Phenols, Alkenes, and Carboxylic Acids & Derivatives functional group.



**Figure 2.** A comparison of initial and baseline corrected spectral data for complexes is shown below. With the smoothing parameter, ALS corrects the baseline for *S. aureus*  $\lambda = 4.100$  and wights  $p = 0.001$ .

| Response              | ALS parameters |           | Shrinkage parameter |          | CVRMSE |      |
|-----------------------|----------------|-----------|---------------------|----------|--------|------|
|                       | P              | $\lambda$ | $\theta$            | fraction | LASSO  | LARS |
| <i>E. coli</i>        | 0.001          | 2.679     | 1.8                 | 0.02     | 2.38   | 2.32 |
| <i>B. bronceptica</i> | 0.001          | 0.616     | 1.3                 | 0.06     | 1.97   | 1.91 |
| <i>S. epidermidis</i> | 0.010          | 0.100     | 1.1                 | 0.07     | 2.33   | 2.01 |
| <i>B. pumilus</i>     | 0.001          | 4.226     | 1.6                 | 0.01     | 1.84   | 1.84 |
| <i>S. aureus</i>      | 0.010          | 4.100     | 1.5                 | 0.02     | 2.14   | 1.91 |

**Table 2.** For LASSO and LAR, the distribution of ALS parameters that define the baseline adjustment, shrinkage parameters that define the wave-number selection, and RMSE is shown.

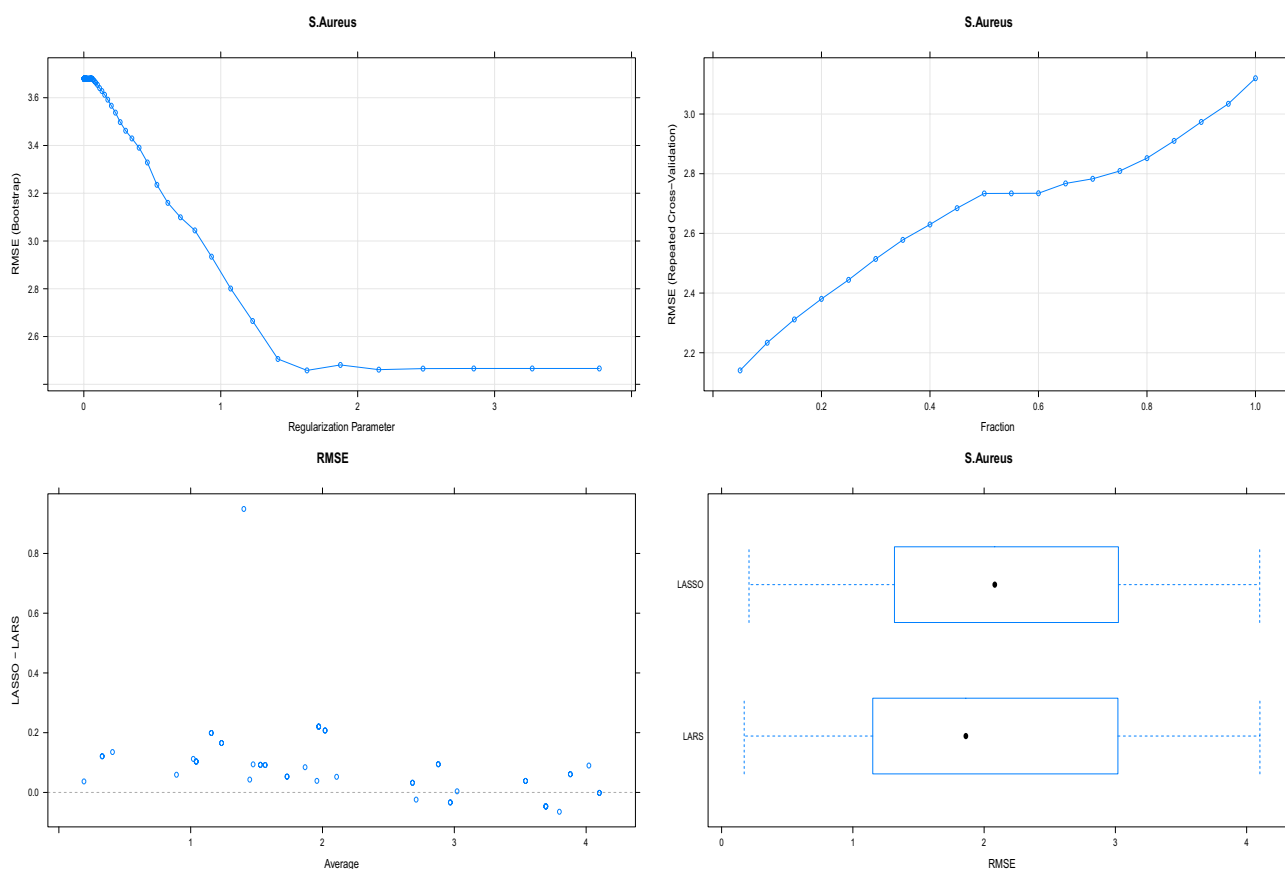
**Computations.** For computations, modeling and figures R software is used<sup>32</sup>. For baseline correction R package 'baseline'<sup>33</sup> and for LASSO and LAR model fitting R package 'caret'<sup>34</sup> is used.

## Conclusions

The aim of this study is to find functional groups for characterization of already synthesised Metronidazole complexes with regard to antibacterial activity. The least absolute shrinkage and selection operator (Lasso) and least angle regression (LAR) was used to model the antibacterial behaviour of Metronidazole complexes using the FTIR spectrum, where the antibacterial activity was tested against bacterial strains *E. coli*, *B. bronceptica*, *S. epidermidis*, *B. pumilus* and *S. aureus*. The practical compound's powerful group is also mapped. In addition, research is needed to analyses and examine existing findings for medical applications. Moreover in future work, we aimed to identify the cluster of influential wavenumbers which can better characterize the Metronidazole complexes.

| Response              | Influential wavenumbers ( $cm^{-1}$ ) | Functional compounds                   | Functional group                 |
|-----------------------|---------------------------------------|--|----------------------------------|
| <i>E. coli</i>        | 1802                                  | C = O                                  | Carboxylic acids and derivatives |
|                       | 2930                                  | CH <sub>3</sub> , CH <sub>2</sub> , CH | Alkanes                          |
|                       | 3156                                  | C – H                                  | Carboxylic acids and derivatives |
| <i>B. bronceptica</i> | 1475                                  | CH <sub>2</sub> , CH <sub>3</sub>      | Alkanes                          |
|                       | 135                                   | C = O                                  | Carboxylic Acids and Derivatives |
|                       | 2999                                  | CH <sub>3</sub> , CH <sub>2</sub> , CH | Alkanes                          |
| <i>S. epidermidis</i> | 2246                                  | C ≡ C                                  | Alkynes                          |
|                       | 2670                                  | O – H                                  | Carboxylic acids and derivatives |
| <i>B. pumilus</i>     | 1298                                  | O – C                                  | Carboxylic acids and derivatives |
|                       | 2545                                  | O – H                                  | Carboxylic acids and derivatives |
|                       | 3010                                  | CH <sub>3</sub> , CH <sub>2</sub> , CH | Alkanes                          |
| <i>S. aureus</i>      | 930                                   | = C – H, = CH <sub>2</sub>             | Alkanes                          |
|                       | 2845                                  | O – H                                  | Carboxylic acids and derivatives |
|                       | 3636                                  | O – H                                  | Alcohols and phenols             |

**Table 3.** By using the bootstrapping with LAR consistent influential wavenumbers for *B. bronceptica*, *E. coli*, *S. epidermidis*, *B. pumilus* and *S. aureus* are selected. The best-fitted model's powerful regression coefficients for each bacteria are presented. Furthermore, for each model, the influential compound assignment and respective group is indicated.



**Figure 3.** RMSE is used to assess the ability of a model to predict MTZ antibacterial activity against *S. aureus*. In the upper left and right panels, the RMSE behaviour of LASSO and LAR against their regularized parameter ( $\theta$ ) and fraction. We used bootstrap cross validation for model fitting, comparing the model RMSE in each fold.



Received: 29 April 2021; Accepted: 9 August 2021

Published online: 29 September 2021

## References

1. Timmerman, H., Mannhold, R. & Krogsgaard-Larsen, P. *Chemometric Methods in Molecular Design* Vol. 2 (Wiley, 2008).
2. Bempong, D. K., Manning, R. G., Mirza, T. & Bhattacharyya, L. A stability-indicating HPLC assay for metronidazole benzoate. *J. Pharm. Biomed. Anal.* **38**, 776–780 (2005).
3. Dubcy, S., Jain, V. & Preethi, G. Evaluation of lipophilicity, antimicrobial activity and mutagenicity of some novel ester prodrugs of metronidazole. *Indian J. Chem. Sect. B* **48**, 1571–1576 (2009).
4. Pandeya, K. B. *et al.* A critical review on traditional herbal drugs: An emerging alternative drug for diabetes. *Int. J. Organ. Chem.* **3**, 1–22 (2013).
5. Seguel, G. V., Rivas, B. L., Moreno, Y. & Peña, O. Poly (dl-aspartic acid) and copper (ii) acetate monohydrate interactions. *Transit. Metal Chem.* **32**, 1106–1109 (2007).
6. Singh, S., Bharti, N. & Mohapatra, P. P. Chemistry and biology of synthetic and naturally occurring antiamebic agents. *Chem. Rev.* **109**, 1900–1947 (2009).
7. Obaleye, J. A., Ajibola, A. A., Bernardus, V. B., Hosten, E. C. & Ozarowski, A. Synthesis, spectroscopic, structural and antimicrobial studies of a dimeric complex of copper (ii) with trichloroacetic acid and metronidazole. *Inorgan. Chim. Acta* **503**, 119404 (2020).
8. Bharti, N. *et al.* Synthesis, crystal structure, and enhancement of the efficacy of metronidazole against entamoeba histolytica by complexation with palladium (ii), platinum (ii), or copper (ii). *Helvetica Chim. Acta* **85**, 2704–2712 (2002).
9. Moncol, J. *et al.* Crystal structures and spectroscopic behavior of monomeric, dimeric and polymeric copper (ii) chloroacetate adducts with isonicotinamide, n-methylnicotinamide and n, n-diethylnicotinamide. *Inorgan. Chim. Acta* **360**, 3213–3225 (2007).
10. Kadiroğlu, P. FTIR spectroscopy for prediction of quality parameters and antimicrobial activity of commercial vinegars with chemometrics. *J. Sci. Food Agric.* **98**, 4121–4127 (2018).
11. Han, C., Chen, J., Wu, X., Huang, Y.-W. & Zhao, Y. Detection of metronidazole and ronidazole from environmental samples by surface enhanced Raman spectroscopy. *Talanta* **128**, 293–298 (2014).
12. Ashtarinezhad, A. *et al.* FTIR-microspectroscopy detection of metronidazole teratogenic effects on mice fetus. *Iran. J. Pharm. Res.* **13**, 101 (2014).
13. Ribeiro da Cunha, B., Fonseca, L. P. & Calado, C. R. A phenotypic screening bioassay for *Escherichia coli* stress and antibiotic responses based on Fourier-transform infrared (FTIR) spectroscopy and multivariate analysis. *J. Appl. Microbiol.* **127**, 1776–1789 (2019).
14. Fu, D., Lu, Y., Gao, S., Peng, Y. & Duan, H. Chemical property and antibacterial activity of metronidazole-decorated Ti through adhesive dopamine. *J. Wuhan Univ. Technol.-Mater. Sci. Ed.* **34**, 968–972 (2019).
15. Tibshirani, R. Regression shrinkage and selection via the lasso: A retrospective. *J. R. Stat. Soc. Ser. B* **73**, 273–282 (2011).
16. Efron, B. *et al.* Least angle regression. *Ann. Stat.* **32**, 407–499 (2004).
17. Chen, H., Lin, Z. & Tan, C. Nondestructive discrimination of pharmaceutical preparations using near-infrared spectroscopy and partial least-squares discriminant analysis. *Anal. Lett.* **51**, 564–574 (2018).
18. Mehmood, T., Iqbal, M. & Hassan, R. Prediction of antibacterial activity in ionic liquids through FTIR spectroscopy with selection of wavenumber by PLS. *Chemomet. Intell. Lab. Syst.* **206**, 104124 (2020).
19. Mehmood, T. & Iqbal, M. FTIR fingerprints discriminate ionic liquids' antibacterial activity. *Chemomet. Intell. Lab. Syst.* **208**, 104200 (2021).
20. Baratieri, S. C., Barbosa, J. M., Freitas, M. P. & Martins, J. A. Multivariate analysis of nystatin and metronidazole in a semi-solid matrix by means of diffuse reflectance NIR spectroscopy and PLS regression. *J. Pharm. Biomed. Anal.* **40**, 51–55 (2006).
21. Glavanović, S., Glavanović, M. & Tomišić, V. Simultaneous quantitative determination of paracetamol and tramadol in tablet formulation using UV spectrophotometry and chemometric methods. *Spectrochim. Acta Part A* **157**, 258–264 (2016).
22. Szabó, É., Gergely, S., Spaits, T., Simon, T. & Salgó, A. Near-infrared spectroscopy-based methods for quantitative determination of active pharmaceutical ingredient in transdermal gel formulations. *Spectrosc. Lett.* **52**, 599–611 (2019).
23. Zendehdel, R., Khodakarim, S. & Shirazi, F. H. Predicting of effective dose as biomarker for cytotoxicity using partial least square-Fourier transform infrared spectroscopy (PLS\FTIR). *Iran. J. Pharm. Res.* **14**, 1189 (2015).
24. Ribeiro da Cunha, B., Fonseca, L. P. & Calado, C. R. Metabolic fingerprinting with Fourier-transform infrared (FTIR) spectroscopy: Towards a high-throughput screening assay for antibiotic discovery and mechanism-of-action elucidation. *Metabolites* **10**, 145 (2020).
25. Schlegel, L. B., Schubert-Zsilavecz, M. & Abdel-Tawab, M. Quantification of active ingredients in semi-solid pharmaceutical formulations by near infrared spectroscopy. *J. Pharm. Biomed. Anal.* **142**, 178–189 (2017).
26. Rahman, Z., Zidan, A. S. & Khan, M. A. Risperidone solid dispersion for orally disintegrating tablet: Its formulation design and non-destructive methods of evaluation. *Int. J. Pharm.* **400**, 49–58 (2010).
27. Sampaio, P. S. & Calado, C. R. Potential of FTIR-spectroscopy for drugs screening against *Helicobacter pylori*. *Antibiotics* **9**, 897 (2020).
28. Attia, K. A., Nassar, M. W., El-Zeiny, M. B. & Serag, A. Effect of genetic algorithm as a variable selection method on different chemometric models applied for the analysis of binary mixture of amoxicillin and flucloxacillin: A comparative study. *Spectrochim. Acta Part A* **156**, 54–62 (2016).
29. Rafique, B. *et al.* Novel copper complexes of metronidazole and metronidazole benzoate: Synthesis, characterization, biological and computational studies. *J. Biomol. Struct. Dyn.* <https://doi.org/10.1080/07391102.2020.1871072> (2021).
30. Eilers, P. H. & Boelens, H. Baseline correction with asymmetric least squares smoothing. *Leiden university medical centre technical report* **1** (2005).
31. Liland, K., Almøy, T. & Mevik, B. Optimal choice of baseline correction for multivariate calibration of spectra. *Appl. Spectrosc.* **64**, 1007–10016 (2010).
32. R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. <https://www.R-project.org>.
33. Liland, K. H., Almøy, T. & Mevik, B.-H. Optimal choice of baseline correction for multivariate calibration of spectra. *Appl. Spectrosc.* **64**, 1007–1016 (2010).
34. Kuhn, M. *et al.* Building predictive models in R using the caret package. *J. Stat. Softw.* **28**, 1–26 (2008).

## Author contributions

All authors have equally contributed in writing the main manuscript text and M.I. conducted the analysis and prepared the results. 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/s41598-021-97897-x>.

**Correspondence** and requests for materials should be addressed to T.M.

**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