
Communication

Modeling Drugs-PLGA Nanoparticles Interactions Using Gaussian Processes: Pharmaceutics Informatics Approach

Rania M. Hathout ^{1,*}, Orchid A. Mahmoud ^{2,3}, Dalia S. Ali ^{2,4}, Marina Mamdouh ², Abdelkader A. Metwally ^{1,5}

¹ Department of Pharmaceutics and Industrial Pharmacy, Faculty of Pharmacy, Ain Shams University, Cairo, Egypt; rania.hathout@pharma.asu.edu.eg;

² Drug Design Program, Faculty of Pharmacy, Ain Shams University, Cairo, Egypt; mari-na.tawfik@sce.aucegypt.edu

³ Department of Pharmacognosy, Faculty of Pharmacy, Ain Shams University, Cairo, Egypt; orchid.ahmed@pharma.asu.edu.eg

⁴ Department of Biotechnology, Organization of Research and Control of Biologics, Egyptian Drug Authority, Cairo, Egypt; dalia.samir.ali@pharma.asu.edu.eg

⁵ Department of Pharmaceutics, Faculty of Pharmacy, Health Sciences Center, Kuwait University, Kuwait; Abdelkader74@yahoo.com

* Correspondence: rania.hathout@pharma.asu.edu.eg; r_hathout@yahoo.com

Abstract: The objective of this study was to correlate the binding of drugs on a very popular nanoparticulate polymeric matrix; PLGA nanoparticles with their main constitutional, electronic and physico-chemical descriptors. Gaussian Processes (GPs) was the artificial intelligence machine learning method that was utilized to fulfil this task. The method could successfully model the results where optimum values of the investigated descriptors of the loaded drugs were deduced. A percentage bias of $12.68 \% \pm 2.1$ was obtained in predicting the binding energies of a group of test drugs. As a conclusion, GPs could successfully model the drugs-PLGA interactions associated with a good predicting power. The GPs-predicted binding energies (ΔG) can easily be projected to the drugs loading as was previously proven. Adopting the “Pharmaceutics Informatics” approach can save the pharmaceutical industry and the drug delivery scientists a lot of exerted resources, efforts and time.

Keywords: machine learning; Gaussian; nanoparticles; descriptors; docking; simulations; PLGA

1. Introduction

Poly-lactide poly-glycolide polymeric nanoparticles (PLGA) are nanoscale polymeric drug delivery carriers that were widely used as drug delivery systems specifically for lipophilic drugs. Accurate prediction of drug-loading in these nanoparticulate systems prior to wet-lab experimentations saves valuable time and resources.

In our previous studies, we have concluded that the integration of some computational methods that span chemoinformatics, molecular dynamics simulations, data mining, molecular docking and mathematical modeling together with machine learning techniques enables the successful prediction of important pharmaceutical objectives such as drug loading, as well as providing important information about the interactions that encompass different drugs with their carriers and subsequently the successful choice of the optimum drugs-carriers pairs [1-2].

Machine learning methods can be categorized into two main classes; the supervised and the unsupervised tools [3-4]. Supervised methods predict an outcome based on the

inputs, and depending on the type of the outcome they can carry-out classification or regression tasks. The supervised machine learning methods include: Artificial neural networks, support vector machines, partial least squares regression and the Gaussian processes.

In this current work, the previously introduced hypothesis of adopting computational approaches to predict drug loading in solid lipid nanoparticles was extended by combining Gaussian process modelling with chemoinformatics, molecular dynamics and molecular docking experiments using different scoring functions.

The all-atom molecular dynamics simulations (MDS) were used aiming for preparing the tripalmitin system, that was further used for the docking experiments. In MDS, each run consists of repetitive steps. The forces exerted on each atom in the simulated system are computed after the numerical solving of Newton's laws of motion. The aforementioned forces are computed utilizing a number of equations and some physical constants which together form the 'force field' used in the MDS. In each step, the positions and velocities of the atoms are updated. MDS are widely used in biophysical and chemical studies, however, its usage in drug formulation investigations is still less common [5-6].

Molecular docking is a different computational method where the affinity of a molecule towards a carrier (receptor) is evaluated by calculating the change in the free energy of binding (ΔG). Resembling the molecular dynamics simulations, this method highly contributed in speeding up the drug discovery process [7].

In order to obtain meaningful information about drug molecules, calculation of main chemoinformatics descriptors was carried-out. These descriptors are then used as inputs for modeling a specific outcome of interest using different machine learning techniques.

The Gaussian processes (GP) are a Gaussian distribution over functions, where the functions are a mean function and a covariance function. GP are non-parametric, which means that there is no specific form of the function that models the relationship between the input(s) and the outcome [8]. GP belong to the supervised learning set of methods, with the additional benefit of providing not only the predictions of outcomes, but also the probability (variance) associated with these predictions. Moreover, GP are less prone to over-fitting, which leads to poor predictions [9].

GP have been found to be of value in the drug delivery contexts and the relevant industries over the last decade. Furthermore, they were recorded to result into more robust models compared to other techniques such as the linear regression and the quantitative structure permeability relationships applied in studying percutaneous absorption of drugs through skin [10,11,12].

For GP to predict an outcome for a given unknown input x^* , the predicted mean μ^* is defined as $\mu^* = R^T(C + \sigma^2 I)^{-1}t$, where R is $N \times L$ cross-covariance matrix expressed as the kernel $k(x, x^*)$ with dimensions N (number of training points) and L (number of test points), C is the training points covariance matrix expressed as $k(x, x)$, I is the identity matrix of dimensions N , σ^2 is the variance of the random noise or random error and t is the training points outcome. The variance Σ^* of the predicted mean is calculated as $\Sigma^* = C^* - R^T(C + \sigma^2 I)^{-1}R$, with C^* being the covariance matrix of the test points expressed as $k(x^*, x^*)$.

To this end, Gaussian Processes were introduced in this study as a new artificial intelligence technique to model the drug-PLGA matrix binding in an attempt to further predict the loaded masses. Combining chemoinformatics tools together with machine learning methods poses a platform for a new approach in the drug delivery field that could be called "Pharmaceutics Informatics".

2. Results and Discussion

This study attempts to utilize the previously generated important data that can be collected from the literature and exploit them in a new generated correlation as a novel

proof-of-concept of the success of combining Bio/chemoinformatics and machine learning methods (as a type of artificial intelligence) in what we can name as “Pharmaceutics Informatics” in modeling and predicting the interactions between different drug-carrier pairs.

Particularly, this concept was previously utilized adopting Gaussian Processes (GPs) in few studies such as Moss et al., 2009 and Hathout et al., 2016 [13–14]. Figure 1 shows the Gaussian Processes generated 3D surface and contour plots.

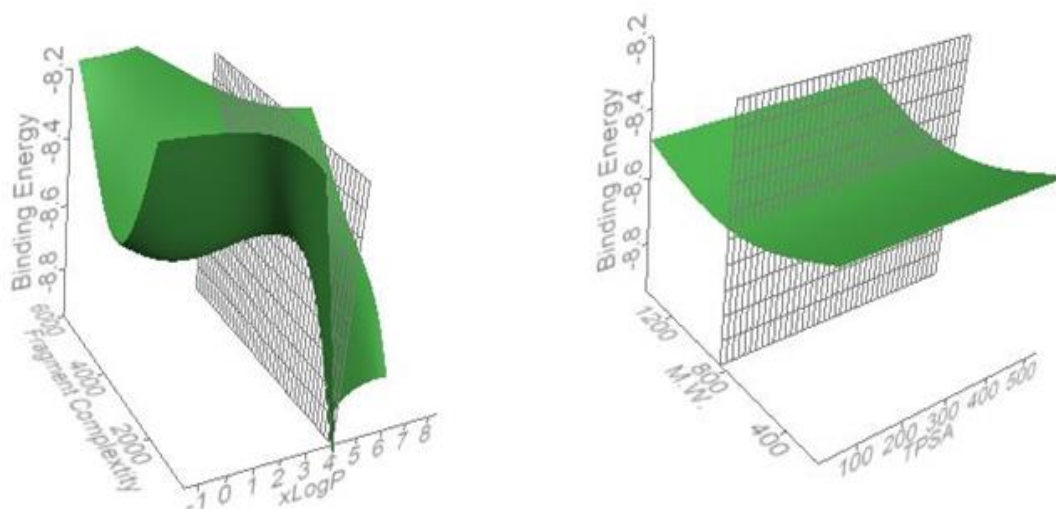


Figure 1. The 3D Gaussian Processes depicting the effect of the main descriptors (in pairs) on the binding energy: Left panel: Fragment Complexity versus xLogP and the right panel: Molecular weight versus TPSA (Total polar surface area). The grid indicates the least values of the dependent factor (the binding energy).

Figure 1 displays a successful modeling of the investigated drug-carrier pairs. It can be concluded from the results that the increase in an important physico-chemical descriptor such as xLogP, above a value of 4, reflecting the high lipophilicity of the drugs led to the lowest binding energy values (ΔG) indicating higher interactions and binding with the PLGA simulated nanoparticulate matrices. This is considered a logical inference since PLGA is a hydrophobic carrier that possesses high affinity to likewise molecules and drugs [15]. It can also be inferred from the same figure that decreasing the fragment complexity of the studied molecules leads to better binding of the PLGA. The fragment complexity of the molecule is a constitutional main descriptor that represents the number of bonds, number of non-hydrogen atoms and number of hetero-atoms of a molecule [1]. Decreasing the number of bonds and the aforementioned atoms could increase the flexibility of the molecules leading to better inclusion in the PLGA matrix and therefore better binding results. [16–17].

The right panel of Figure 1 displays the presence of an optimum value (almost 800 Daltons) for the better binding of the drug molecules with PLGA nanoparticulate matrix. Small molecular weight drugs may have high diffusion rates from the PLGA matrices leading to less binding and further low entrapment efficiencies. On the other hand, drugs possessing high molecular weights above 800 Daltons may lack the ability to be docked inside the pores and voids of the PLGA polymeric matrix [18–19]. Furthermore, it is shown in the figure that the total polar surface area has no obvious impact on the values of the binding energies. This can be ascribed to the hydrophobic nature of the carrier that led to the insignificant effect ($P < 0.05$) of the values of this electronic descriptor in the studied range (100–500 Angstrom squared). Gaussian Processes usually possess the ability of extracting the main features (descriptors) that affect a specific outcome [20].

In order to internally validate the Gaussian Processes generated model, the actual obtained binding energies (as previously reported in Hathout et al., 2015 [1]) were plotted against the predicted binding energy values (Figure 2). The figure displays the high correlation between the predicted and the actual ΔG values. Most of the obtained points were highly proximal to the 45° degrees line. This indicates the high similarity between the actual and the predicted values [21,22,23,24].

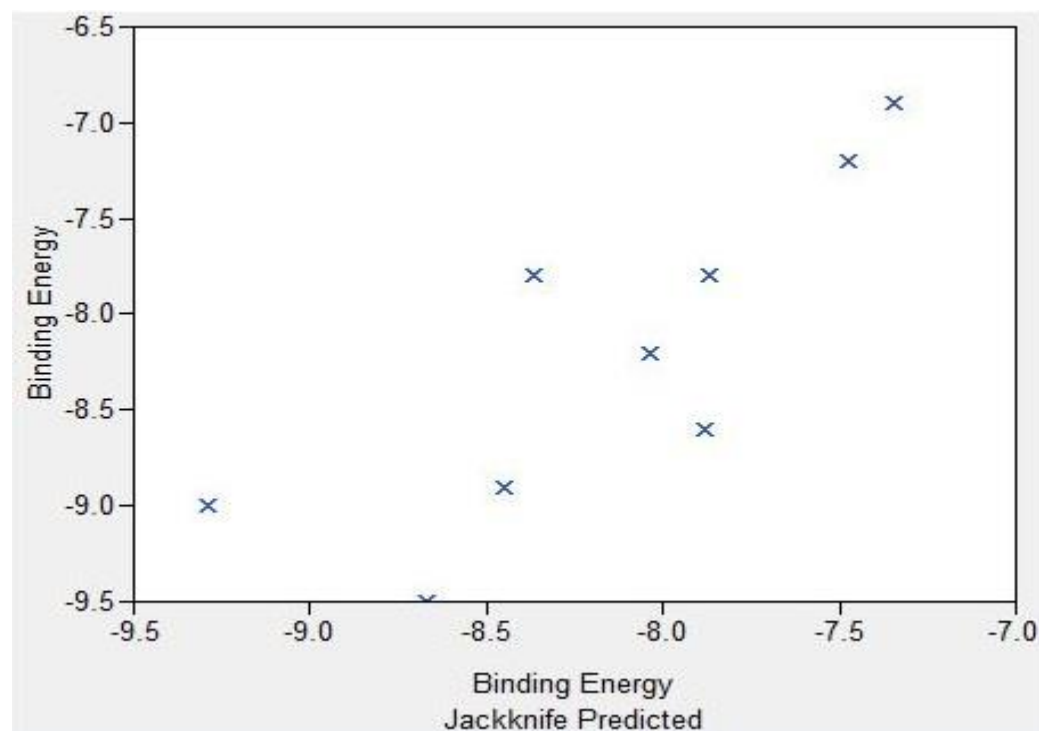


Figure 2. The predicted docked binding energies versus the actual counterparts of the studied drugs using Gaussian Processes.

In an attempt for the external validation of the Gaussian Processes, a group of four test drugs was fed in the GPs predictor profiler of JMP/SAS® software package; Aspirin, Prednisolone, Resveratrol and Testosterone (Figure 3). The percentage bias was calculated for these four additional molecules of the test group (Table 1). The scored % Bias was $12.68\% \pm 2$. This value was in a close agreement of that obtained using Artificial Neural Networks algorithms (11.83) that was calculated for the only two molecules that were involved in this study [1]. This acceptable predicting ability of Gaussian processes was previously reported in several studies [10-26].

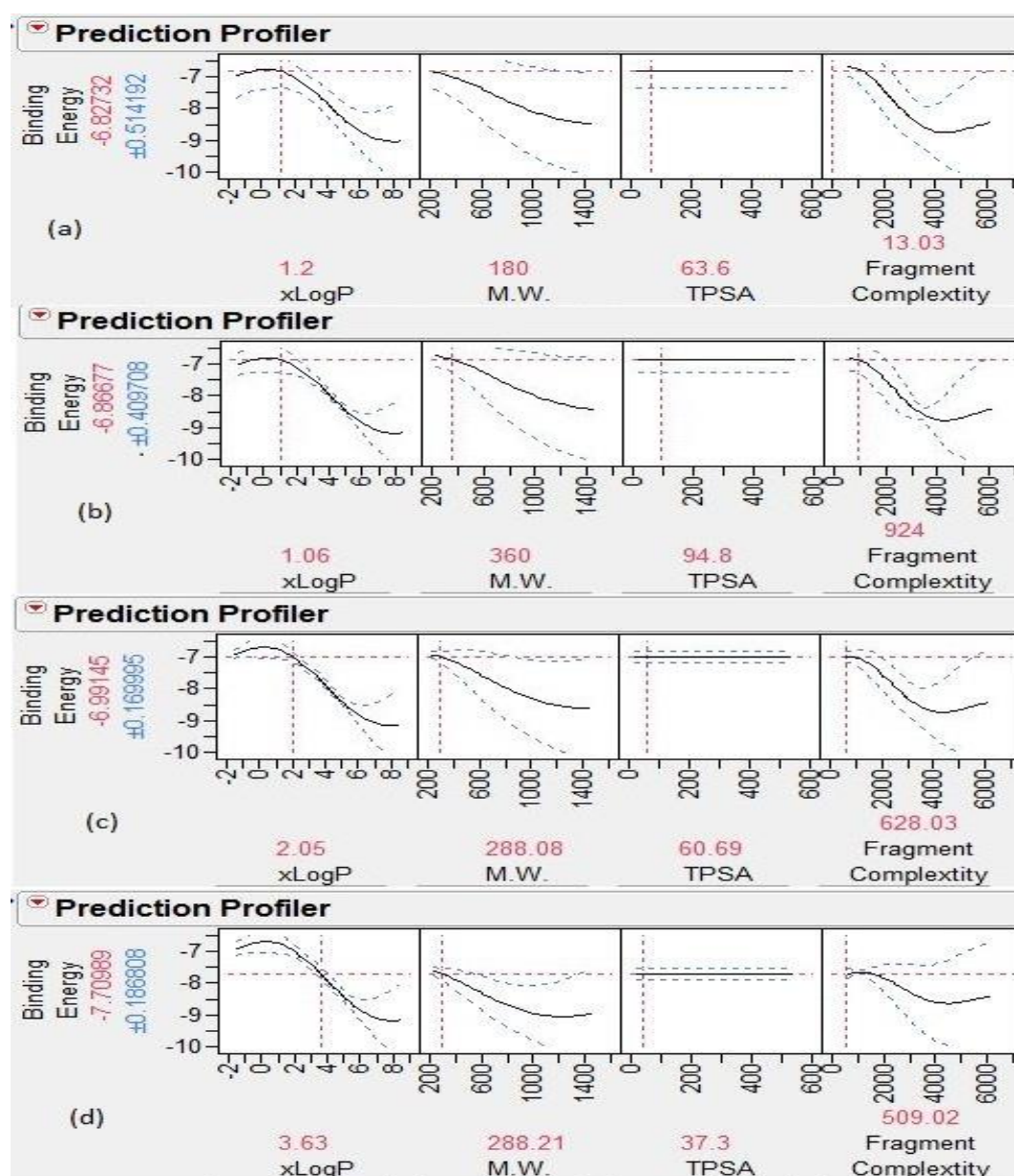


Figure 3. Calculating the binding energy of four validating drugs on PLGA nanoparticulate matrix (a) aspirin , (b) prednisolone , (c) resveratrol and (d) testosterone using the prediction profiler tool of SAS/JMP® software.

Table 1. Calculating the Bias percentages between the predicted and actual docking binding energies for the utilized validating drugs.

Drug	Actual ΔG (kcal/mol)	Predicted ΔG (kcal/mol)	Bias (%)
Aspirin	-5.9 ± 0.1	-6.8 ± 0.2	15.2
Prednisolone	-8 ± 0.2	-6.9 ± 0.0	13.7
Resveratrol	-7.8 ± 0.2	-7 ± 0.1	10.3
Testosterone	-8.7 ± 0.2	-7.7 ± 0.1	11.5

3. Methods

3.1. Data mining through the scientific literature databases.

The binding energies (ΔG) of the performed docking experiments of 10 drugs; Docetaxel,^[27] Paclitaxel ^[24], Triamcinolone acetonide ^[28], Indomethacin ^[29], Cyclosporin A ^[29], Ketoprofen ^[29], Cytarabine ^[30], Vancomycin ^[29], Phenobarbital ^[29], and Valproic acid ^[29] utilizing AutoDock vina[®] on the virtually simulated PLGA nanoparticulate matrix using molecular dynamics , were obtained from Metwally and Hathout, 2015 ^[1-2].

3.2. Computing the main descriptors of the studied drugs.

The main constitutional, electronic and topological descriptors were computed for the investigated drugs aiming to interpret the differences in docking scores that were observed and subsequently correlate them to their recorded entrapment efficiencies. The chosen descriptors were the molecular weight, xLogP, topological polar surface area and the fragment complexity. These main descriptors were computed using *Bioclipse*[®] version 2.6 (Bioclipse project, Uppsala University, Sweden) after incorporating the mol files of the drugs that were previously obtained using ChemDraw[®] Ultra version 10.

3.3. Modeling utilizing Gaussian Processes (GPs).

The modeling of the obtained binding energies (output) resulting from the docking experiments was performed using GPs modeling in JMP[®] 7.0 (SAS, Cary, NC). The x-factors (inputs) were the descriptors that were computed in section 2.2. The sole output was the binding energies (ΔG).

3.4. Validating the predicting ability of the developed GPs.

The predicting power of the utilized Gaussian Processes was evaluated after computing the predicted binding energies and contrasting these values with the actual obtained values. Furthermore, the molecular descriptors of two different tested drugs together with their docked binding energies on the PLGA matrices namely; Prednisolone and Testosterone were obtained from Metwally and Hathout, 2015 ^[1-2]. Two additional test drugs were added, namely; Aspirin and Resveratrol where their important constitutional, electronic and physic-chemical descriptors were calculated using *Bioclipse*[®] v.2.6 (Bioclipse project, Uppsala University, Sweden). The docking experiment was also performed using AutoDock Vina[®] and the virtually simulated PLGA matrix that was built using the same parameters that were used previously by Metwally and Hathout, 2015 ^[1]. Afterwards, the main descriptors of each test drug were fed into the prediction profiler of the constructed GPs in section 2.3. and consequently their corresponding binding energies were generated and contrasted to the actual docking results. Consequently, the percentage bias (% Bias) was computed between the actual and predicted values. The results were compared to the value that was previously reported utilizing ANN ^[1]. The percentage bias was calculated according to the following equation ^[25]:

$$\text{Bias (\%)} = \frac{|\text{Predicted} - \text{Actual}|}{\text{Actual}} \times 100 \quad \text{Eq. 1}$$

4. Conclusions

Gaussian Processes was successfully introduced in this study as an artificial intelligence technique that could correlate the important constitutional, electronic and physico-chemical descriptors such as the molecular weight, fragment complexity and total polar surface area and xLogP of several investigated drugs with their binding energies on a virtually simulated PLGA nanoparticulate matrix. It was concluded that molecules

possessing a molecular weight around 800 Daltons, decreased fragment complexity and xLogP above 4 would lead to the lowest (more negative) binding energy values and hence stronger binding and interaction with the PLGA matrix. Moreover, the obtained predicting power of the utilized GPs was highly acceptable ($12.38 \% \pm 2.1$). Projecting the binding energies results into masses of loaded drugs per 100 mg of the carrier matrix was previously attained using a simple mathematical relationship. Accordingly the loading masses can be further predicted. To this end, the introduced technique provides a new proof-of-concept of the success of what we can name as “Pharmaceutics Informatics” tools in providing clues for drug delivery scientists and formulators on the good selection of the drug-carrier pairs. Adopting this new technique would save the research and development sectors in the pharmaceutical industry and researches in universities and institutes huge amounts of resources, efforts and time that are usually spent in the wet-lab real experiments.

Author Contributions: Conceptualization, R.M.H.; methodology, R.M.H., O.M.A, D.S.A. and A.A.M.; validation, R.M.H. and A.A.M.; investigation, R.M.H. and A.A.M.; resources, R.M.H. and A.A.M.; data curation, R.M.H. and A.A.M.; writing—original draft preparation, R.M.H. and A.A.M.; writing—review and editing, R.M.H. and A.A.M.; visualization, R.M.H. and A.A.M.; All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

References:

1. Metwally, A. A. & Hathout, R. M. Computer-Assisted Drug Formulation Design: Novel Approach in Drug Delivery. *Mol. Pharmaceutics* (2015). **12**, 2800-2810
2. Metwally, A. A. & Hathout, R. M. Replacing microemulsion formulations experimental solubility studies with in-silico methods comprising molecular dynamics and docking experiments. *Chemical Engineering Research and Design* (2015). **104**, 453-456
3. Gad, H. A., El-Ahmady, S. H., bou-Shoer, M. I. & Al-Azizi, M. M. Application of chemometrics in authentication of herbal medicines: a review. *Phytochem. Anal.* (2013). **24**, 1-24
4. Hathout, R. M. Using principal component analysis in studying the transdermal delivery of a lipophilic drug from soft nano-colloidal carriers to develop a quantitative composition effect permeability relationship. *Pharm Dev. Technol.* (2014). **19**, 598-604
5. Alayoubi, A., Satyanarayanajois, S. D., Sylvester, P. W. & Nazzal, S. Molecular modelling and multisimplex optimization of tocotrienol-rich self emulsified drug delivery systems *Int. J Pharm.* (2012). **426**, 153-161
6. Uccello-Barretta, G. *et al.* Combining NMR and molecular modelling in a drug delivery context: investigation of the multi-mode inclusion of a new NPY-5 antagonist bromobenzenesulfonamide into beta-cyclodextrin *Bioorg. Med. Chem.* (2004). **12**, 447-458
7. Baskaran, C. & Ramachandran, M. Computational molecular docking studies on anticancer drugs. *Asian Pacific Journal of Tropical Disease* (2012) S734-S738
8. Shah, A. *et al.* Support vector regression to estimate the permeability enhancement of potential transdermal enhancers. *J Pharm Pharmacol.* (2016)
9. Brown, M. B. *et al.* An evaluation of the potential of linear and nonlinear skin permeation models for the prediction of experimentally measured percutaneous drug absorption. *J Pharm Pharmacol.* (2012). **64**, 566-577
10. Ashrafi, P., Moss, G. P., Wilkinson, S. C., Davey, N. & Sun, Y. The application of machine learning to the modelling of percutaneous absorption: an overview and guide. *SAR QSAR. Environ. Res.* (2015). **26**, 181-204
11. Lam, L. T. *et al.* The application of feature selection to the development of Gaussian process models for percutaneous absorption. *J Pharm Pharmacol.* (2010). **62**, 738-749
12. Moss, G. P. *et al.* The application of Gaussian processes in the prediction of percutaneous absorption. *J Pharm Pharmacol.* (2009). **61**, 1147-1153
13. Moss, G. P. *et al.* The application of Gaussian processes in the prediction of percutaneous absorption. *J Pharm Pharmacol.* (2009). **61**, 1147-1153
14. Hathout, R. M. & Metwally, A. A. Towards better modelling of drug-loading in solid lipid nanoparticles: Molecular dynamics, docking experiments and Gaussian Processes machine learning. *Eur. J Pharm Biopharm.* (2016). **108**, 262-268
15. Mehanny, M., Hathout, R. M., Geneidi, A. S. & Mansour, S. Studying the effect of physically-adsorbed coating polymers on the cytotoxic activity of optimized bisdemethoxycurcumin loaded-PLGA nanoparticles. *J. Biomed. Mater. Res.* (2017). **105**, 1433-1445

16. Hathout, R. M., Metwally, A. A., Woodman, T. J. & Hardy, J. G. Prediction of Drug Loading in the Gelatin Matrix Using Computational Methods. *ACS Omega* (2020). **5**, 1549-1556
17. Hathout, R. M., Abdelhamid, S. G., El-Housseiny, G. S. & Metwally, A. A. Comparing cefotaxime and ceftriaxone in combating meningitis through nose-to-brain delivery using bio/chemoinformatics tools. *Scientific Reports* (2020). **10**, 21250
18. Welling, S. H. *et al.* In silico modelling of permeation enhancement potency in Caco-2 monolayers based on molecular descriptors and random forest. *Eur. J Pharm Biopharm.* (2015). **94**, 152-159
19. Hathout, R. M., Abdelhamid, S. G. & Metwally, A. A. Chloroquine and hydroxychloroquine for combating COVID-19: Investigating efficacy and hypothesizing new formulations using Bio/chemoinformatics tools. *Inform. Med Unlocked.* (2020). **21**, 100446
20. Lam, L. T. *et al.* The application of feature selection to the development of Gaussian process models for percutaneous absorption. *J Pharm Pharmacol.* (2010). **62**, 738-749
21. Abdel-Hafez, S. M., Hathout, R. M. & Sammour, O. A. Towards better modeling of chitosan nanoparticles production: screening different factors and comparing two experimental designs. *Int. J Biol Macromol.* (2014). **64**, 334-340
22. Naguib, S. S., Hathout, R. M. & Mansour, S. Optimizing novel penetration enhancing hybridized vesicles for augmenting the in-vivo effect of an anti-glaucoma drug. *Drug Delivery* (2017). **24**, 99-108
23. Safwat, S., Ishak, R. A. H., Hathout, R. M. & Mortada, N. D. Nanostructured lipid carriers loaded with simvastatin: effect of PEG/glycerides on characterization, stability, cellular uptake efficiency and in vitro cytotoxicity. *Drug Dev. Ind. Pharm* (2017). **43**, 1112-1125
24. Safwat, S., Hathout, R. M., Ishak, R. A. & Mortada, N. D. Augmented simvastatin cytotoxicity using optimized lipid nanoparticles: a potential for breast cancer treatment. *J Liposome Res* (2017). **27**, 1-10
25. Abdel-Hafez, S. M., Hathout, R. M. & Sammour, O. A. Towards better modeling of chitosan nanoparticles production: Screening different factors and comparing two experimental designs. *International Journal of Biological Macromolecules* (2014). **64**, 334-340
26. Lam, L. T. *et al.* The application of feature selection to the development of Gaussian process models for percutaneous absorption. *J Pharm Pharmacol.* (2010). **62**, 738-749
27. Keum, C. G. Practical preparation procedures for docetaxel-loaded nanoparticles using polylactic acid-co-glycolic acid. (2011)
28. Sabzevari, A. *et al.* Polymeric triamcinolone acetonide nanoparticles as a new alternative in the treatment of uveitis: in vitro and in vivo studies. *Eur. J Pharm. Biopharm.* (2013). **84**, 63-71
29. Barichello, J. M., Morishita, M., Takayama, K. & Nagai, T. Encapsulation of Hydrophilic and Lipophilic Drugs in PLGA Nanoparticles by the Nanoprecipitation Method. *null* (1999). **25**, 471-476
30. Yadav, K. S. & Sawant, K. K. Modified nanoprecipitation method for preparation of cytarabine-loaded PLGA nanoparticles. *AAPS PharmSciTech* (2010). **11**, 1456-1465