NUCLEAR ENGINEERING AND TECHNOLOGY
ISSN / eISSN 1738-5733
Publisher KOREAN NUCLEAR SOC, NUTOPIA BLDG,
342-1 JANGDAE-DONG, DAEJEON, SOUTH KOREA,
Volume 26-Issue 1-(2026)
DOI:https://doi.org/10.65281/731830

OPEN ACCESS

A HYBRID QSAR AND GROUP-INTERACTION CONTRIBUTION APPROACH TO PREDICT THE PERMEABILITY AND SOLUBILITY PROPERTIES OF SFT-MTX COMPLEXES

Check for updates

Saliha Gacem[*1,2,4], Khadra Mokadem[3], Regadia Aissaoui [1,2,4]

[1] Univ. Ziane Achour of Djelfa, Lab. Exploration and Valorization of Steppic Ecosystems, Djelfa, Algeria

[2] Department of Chemistry, Faculty of Exact Science and Computer Science, University Ziane Achour of Djelfa, B.P.3117, 17000, Djelfa, Algeria.

[3] Faculty of Mathematics & Matter Sciences, University of Kasdi Merbah, B.P.511, 30000, Ouargla, Algeria.

[4] Center of Research in Agro-Pastoralism, Djelfa - Algeria

[*]Corresponding author: saliha.gacem@univ-djelfa.dz ; salihagacem91@yahoo.com

Received: 21/11/2025 ; Accepted: 23/04/2026

ABSTRACT

In this study, a hybrid approach combining QSAR and group-interaction contributions (GIC) is proposed to predict the permeability and solubility of SFT-MTX complexes. It was developed using a database containing 67 datasets related to a wide variety of SFTs. The entire dataset was randomly split into a training set of 57 data points and a validation set of 10. the QSAR-GIC technique demonstrated a strong relationship between molecular structure and the effectiveness of complexes, where the Linear and Non-Linear modeling yielded a (R^2_{Linear} = 97,02% , $\%AARD_{Linear}$ = 5,8573% and $R^2_{Non-Linear}$ = 99,97% , $\%AARD_{Non-Linear}$ = 2,0042%) for permeability; (R^2_{Linear} = 97,28% , $\%AARD_{Linear}$ = 3,3365% and $R^2_{Non-Linear}$ = 99,86% , $\%AARD_{Non-Linear}$ = 0,7048%) for solubility. It was concluded that the models are sufficiently accurate for reliable predictions, underscoring the importance of incorporating SFTs into MTX to enhance its effectiveness. This work validates surfactants' capacity to ameliorate MTX and enhance its therapeutic efficacy against breast cancer, using computational techniques and the QSAR-GIC approach, thereby paving the way for the development of more effective and safer anticancer therapies.

Keywords: SFT-MTX Complex; QSAR-GIC; Property Estimation; Solubility; Permeability;

INTRODUCTION

Drugs (e.g., mitoxantrone (MTX)) are characterized by their hydrophobic nature, which leads to lower bioavailability. This can be increased by adding adjuvant substances to the drug formulation, such as surfactants (SFT). They can interact effectively with drugs and protect them from damaging microbes thanks to their diverse properties, such as amphiphilic structure, adsorption and aggregation at the interfaces, micellization, and antimicrobial activity. Thanks to these properties, SFTs can be used in many applications, such as delivery systems, improved solubility and permeability, stabilizing drugs, and preventing their precipitation. [1,2]

Surfactants are well-known surface-active agents and have been widely used as solubility enhancers. The most useful surfactant property for improving solubility and other applications is the ability to reduce the surface tension of water and other media. Surfactants with different functionalities in the backbone can dissolve in both aqueous and non-aqueous environments and exhibit many desirable properties in solution. They have demonstrated their importance in drug manufacturing processes. They may be used to dissolve many drug species that are generally insoluble in the aqueous phase and to manufacture various anticancer drugs, such as Mitoxantrone. [3]

Mitoxantrone (MTX) is a strong chemotherapy antineoplastic and immunosuppressive drug that can help cure multiple sclerosis (MS) and many types of cancer, including Breast cancer, various leukemias (e.g., Acute lymphocytic leukemia (ALL), Acute nonlymphocytic leukemia (ANLL), Acute myelogenous leukemia (AML)…), advanced hormone-refractory prostate cancer, non-Hodgkin's lymphoma (NHL), and certain solid tumors.[4] MTX is classified as a polar, non-amphiphilic drug [5], which results in low solubility, decreased absorption, and limited permeability. MTX (Novantrone, dihydroxyanthracenedione) belongs to a new structural class of synthetic antineoplastic agents. These anthracenediones have basic side chains, producing free radicals [called anthracyclines] attached to a planar aromatic ring system [tricyclicchromophores, which is a tricyclic aromatic backbone, less complex, and lacks the sugar group], facilitating their intercalation into DNA, and leading to inhibition of replication and transcription processes. Anthracyclines may cause cardiotoxicity of varying risk, particularly in cumulative doses. In contrast, Anthracenediones, such as Mitoxantrone, exhibit strong anticancer activity and lower cardiotoxicity due to their low generation of free radicals. However, it may also cause other toxicities, such as myelosuppression, predominantly leukopenia and neutropenia, necessitating periodic hematologic monitoring. It shows immunosuppressive effects. [4,6]

MTX works against cancer mainly by intercalating into deoxyribonucleic acid (DNA) and forming hydrogen bonds with it, and can also interfere with ribonucleic acid (RNA) and is considered a potent inhibitor of topoisomerase II. [4,7] Clinically, MTX is usually given intravenously, with its pharmacological profile supporting multiple routes under study (e.g., intraperitoneal, intrapleural). Yet, intravenous injection remains the standard method approved in therapy, where continuous injection regimens achieve steady plasma levels and increased cellular uptake but require further investigation. MTX offers a favorable balance between effectiveness and safety, making it a valuable chemotherapy option for various types of cancer. [8] The interaction between mitoxantrone and surfactants is an important focus of research, attracting significant attention due to the clinical relevance of mitoxantrone as an anticancer drug, particularly for breast cancer, and surfactants' ability to address some of its limitations (e.g., reduced permeability and water solubility).

METHODOLOGY

Computational chemistry methods have been developed to study reactions and predict reactivity in synthetic chemistry. As a result, vast ranges of compound reactivity can be predicted computationally, enabling the design of drugs for Breast cancer, Acute lymphocytic leukemia (ALL), Acute myelogenous leukemia (AML), and others. Many theoretical physical parameters were computed using computational methods. Various regression methods were used to compare the predicted data with the computed values of Bending Energy, Permeability, and Solubility for complexes formed using two primary models: linear and nonlinear.[9]

An in-depth theoretical study of the interactions between MTX and a 65 SFT was conducted using molecular docking techniques. Then, the effect of SFT on MTX effectiveness was evaluated by estimating Log S (ESOL) and Consensus Log $P_{o/w}$ based on the SwissADME tools [10]. To predict and validate the Permeability and solubility properties of SFT-MTX Complexes formed, the QSAR-GIC method was applied to identify the contributions of overlapping groups using two models (linear and non-linear). [11,12]

Data sets

A database was compiled containing 67 SFT-MTX complexes, which are classified into four classes: anionic, cationic, zwitterionic, and non-ionic, originally generated from synthetic or natural substances. The chemical structures of MTX and selected SFT were retrieved from the PubChem database (https://pubchem.ncbi.nlm.nih.gov/). [13] Figure 1 presents the statistics of the distribution of surfactant groups across the entire database.

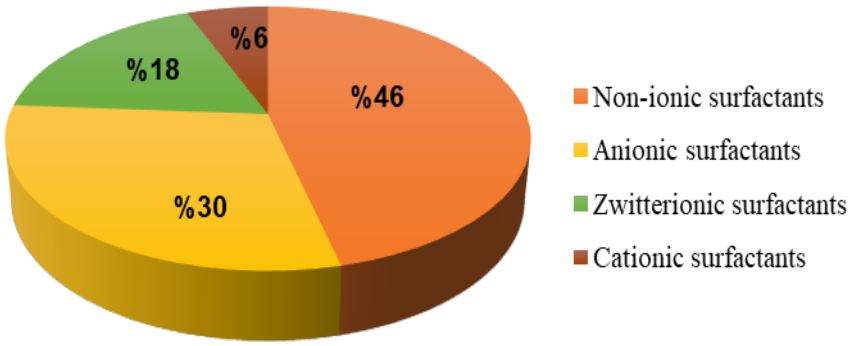

Figure 1. The database's relative surfactant groups class.

Method proposed in this work

All calculations were performed using AutoDock Tools 1.5.7 and AutoDock Vina. [14,15] The structure of the SFT-MTX complexes was cleaned and optimized using Avogadro [16] and BIOVIA Discovery Studio 2021 [17]. The Evaluation of complex properties was performed using the SwissADME Tools [10], which enables the calculation of various parameters, including Log S (ESOL) and Consensus Log $P_{o/w}$ values.

The suggested modelling technique has several benefits over other models, as it is based only on molecular structure and enables the prediction of generated complexes. Furthermore, it allows the identification of newly developed compounds not present in the database, thereby supporting experimental analysis while reducing time and cost. To validate the Permeability and solubility properties of SFT-MTX Complexes, a highly accurate predictive model was developed using the QSAR-GIC approach. [11,12]

To highlight the significance of incorporating SFTs with MTX, the proposed model relies on a three-level approximation: first-order contribution, second-order contribution, and third-order correction contribution, as shown in Figure 2. This was done according to the principles outlined in previous works by Marrero and Pardillo [12] and Constantinou and Gani [18]. The method suggested by these researchers is articulated around the following five points:

i) In the proposed QSAR-GIC modeling, the Property P (log S (ESOL) or Consensus Log P) was defined as the sum of the contribution values of the added SFT fragments and the property value of the drug Mitoxantrone.

ii) Conventional group contribution method: Log S (ESOL) and Consensus Log $P_{o/w}$ of an SFT_MTX complexes is regarded as a function of structurally-dependent parameters, determined as the sum of the number frequency of each simple group-interaction occurring in the complex, times

its contribution between bonding groups instead of the contribution of simple groups. This approach was proposed by Pardillo and Gonzalez-Rubio [19] and Marrero and Pardillo [12]. Various simple groups were selected to form a set of group interactions, allowing complex property prediction. The groups selected in this study are like those used by Marrero and Pardillo [12]. However, additional groups have been added to account for specific structural features of the complexes formed.

iii) As in the approach suggested by Constantinou and Gani [18], three levels of approximation are involved in the property prediction procedure. The first (basic) level uses contributions from first-order simple groups. A rather small set of second-order groups is used. These have the first-order groups as building blocks. This concept was adopted in this study to avoid the complexity of property estimation that would be engendered by incorporating a multi-order approach. As illustrated in Figure 3, the first order is associated with interaction contributions from simple groups (e.g.: CH3- , -CH2- and >CH- …) and interactions between a cation and anion (e.g.: interactions of single groups $> N^+ <$, Cl^-, Br^- , and interactions of second groups SO_4^{2-}, PO_4^{2-} CO_2^-,…). The second order uses binary interaction contributions between bonding groups (e.g.: xyl, A, and Anthra).

iv) The third-order is a correction term incorporating the equivalent number denoted as E, which equals 1 or 2. The number of Hydrogen bonds n_H ranges between 1 and 8, and the number of Hydrophobic bonds denoted as n_h, ranges between 1 and 7.

v) To achieve greater accuracy, the Non-Linear model was applied to two properties: Solubility (reported as Log S (ESOL), and Permeability (reported as Consensus Log Po/w).

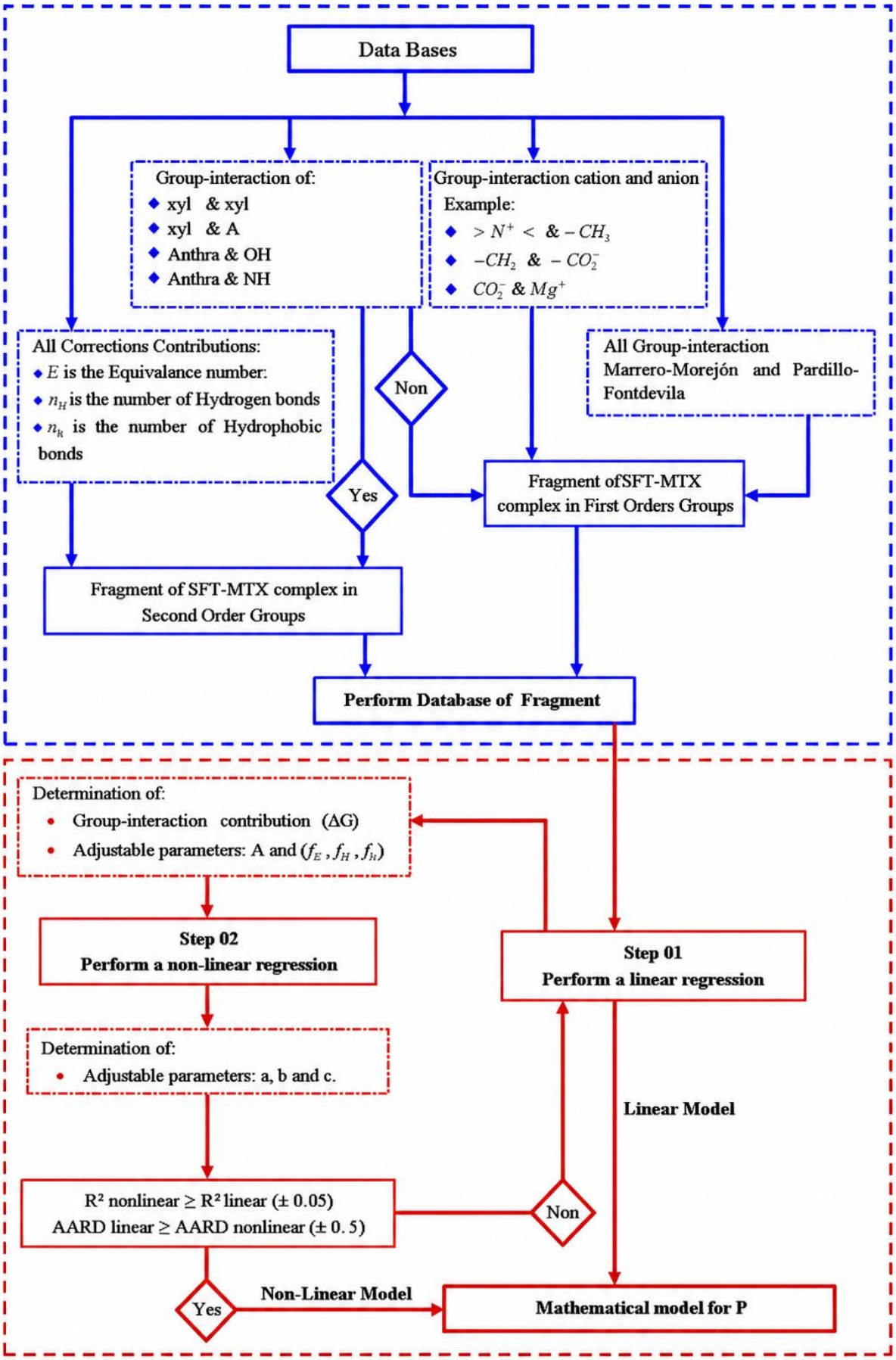

Figure 2. Technical study of the method of estimating properties

The permeability and solubility properties of SFT-MTX Complexes are greatly influenced by the number, nature, and functional groups of the surfactants interacting with MTX. These factors are among the most important in evaluating the effectiveness of surfactants in improving MTX properties. The performance and the effectiveness of complexes were evaluated by estimating two principal factors: Log S (ESOL) and Consensus Log $P_{o/w}$. To validate this relationship, the QSAR-GIC technique was employed. To link the contributions of interacting groups to the property (P), linear and non-linear models were applied as described by the following equations:

$$P = A + \sum_{j} n_j \Delta C_j \quad (1)$$

$$P_{\mathbf{Linear}} = P_{drug} + \left(A + \sum_{j} n_j \Delta C_j \right) \quad (2)$$

$$P_{\mathbf{Linear}} = P_{drug} + \left(A + \sum_{j} n_j \Delta C_j \right) + f_E \times E + f_H \times n_H + f_h \times n_h \quad (3)$$

$$P_{\mathbf{Non-Linear}} = g\ O\ P_{\mathbf{Linear}} \quad (4)$$

Where P is property study: "Solubility (Log S (ESOL), and Permeability (Consensus Log $P_{o/w}$)", A is constants, n_j represents the number of interacting groups (SFT-MTX) complexes, C_j indicates the contribution of each group, (f_E , f_H , f_h) were the equivalence factors, E is the Equivalence number, n_H is the number of Hydrogen bonds, and n_h is the number of Hydrophobic bonds.

The performance of the model was evaluated using three statistical indicators: the Percent Average Absolute Deviation (AAD), the Average Absolute Relative Deviation Percent (%AARD), and the Coefficient of Determination (R^2). [11,12]

$$AAD = \frac{1}{n}\sum_{i=1}^{n} |P_i - PredP_i| \quad (5)$$

$$\%AARD = \frac{100}{n}\sum_{i=1}^{n} \left| {PredP_i}/{P_i} - 1 \right| \quad (6)$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n} \left| {PredP_i}/{(P_i^{ext} - P_i)} \right|^2}{\sum_{i=1}^{n} \left| {PredP_i}/{P_i - average(P_i)} \right|^2} \quad (7)$$

RESULTS AND DISCUSSION

Overall, the combination of MTX with different types of surfactants has great potential to enhance therapeutic efficacy in many types of cancers, especially breast cancer. The docking and ADMET findings demonstrated that surfactants can increase the binding affinity of MTX to the target protein and improve the solubility and permeability of MTX, which could boost its therapeutic effectiveness. The prediction by ProTox3 and VEGA-QSAR tools to suggest that the complexes still have a good toxicity profile equivalent to the MTX drug can also support their potential safety for use in pharmaceuticals. This evaluation is only a summary of the general properties found for the most efficient complexes. More information is needed to assess these properties and move toward a synthesis of ideal anticancer therapeutic agents that are effective and safe. The QSAR-GIC approach have yielded accurate results that highlight a considerable relationship between interaction efficiency and the molecular structure of complexes. The study was based on a dataset consisting of 67 complexes, which were divided into two sets: 75% (57 complexes) for Training and 15% (10 complexes) for Validation, representing a balanced split for adequate model learning and reliable testing.

Linear model

Overall, the SFT compounds examined in this study demonstrate positive effects, with several of them being found to improve MTX performance and strengthen interactions between complexes and protein. In this study, we estimated three properties: Energie (E), Solubility (reported as Log S (ESOL), and Permeability (reported as Consensus Log $P_{o/w}$). These properties of the complex are significantly linked to the functional groups present in the surfactants and mitoxantrone drug. This relationship was modeled using the QSAR-GIC technique through a linear approach, according to the following equations:

$$\text{Log S (ESOL)}_{\mathbf{Linear}} = \text{Log S (ESOL)}_{\text{drug}} + \left(1{,}488 + \sum_{\text{j}} \text{n}_\text{j}\Delta\text{C}_\text{j}\right) + 1{,}671 \times E - 0{,}070 \times n_H + 0{,}068 \times n_h \qquad (8)$$

$$\text{Consensus Log P}_{\text{o/w}_{\mathbf{Linear}}} = \text{Consensus Log P}_{\text{o/w}_{\text{drug}}} + \left(-0{,}622 + \sum_{\text{j}} \text{n}_\text{j}\Delta\text{C}_\text{j}\right) - 1{,}544 \times E - 0{,}258 \times n_H + 0{,}372 \times n_h \qquad (9)$$

The results of the examined linear model revealed statistically significant coefficients estimated at all modeling stages. Table 1 presents the statistical parameters corresponding to each stage of the modeling process.

Table 1. Statistical Results of the Proposed Linear Models Using the QSAR-GIC Method for Property (P)**.**

		Linear Model		
		Solubility	**Permeability**	**No of data points**
R^2	Overall set	0,9728	0,9702	67
	Training set	0,9729	0,9702	57
	Validation set	0,9329	0,9277	10
%AARD	Overall set	3,3365	5,8573	67
	Training set	2,8356	5,0979	57
	Validation set	5,3110	16,6156	10
AAD	Overall set	0,1813	0,1423	67
	Training set	0,1527	0,1392	57
	Validation set	0,2662	0,3793	10

In analyzing the model in depth and comparing the results, several observations can come to light. Depending on the results of the two studied properties, the interacting group model exhibits a strong correlation coefficient between the Training and overall modeling stage, whereas a slight discrepancy is observed during the validation modeling stage. The absolute error values show a greater convergence between the Training and overall modeling stages. The AARD value was slightly high in the solubility property, while we recorded a value greater than expected for the AARD on the permeability property.

For the overall set, the coefficient of Determination (R^2) was 0,9728 for Solubility and 0,9702 for Permeability, reflecting excellent overall model performance across all data and confirming that the two models can explain more a 97% of the included data. These high values further support the overall efficiency of the two models. The Average Absolute Relative Deviation (%AARD) was 3,3365 for Solubility and 5,8573 for Permeability, which is a relatively low values that verify the model's stability and accuracy on the entire dataset. For Average Absolute Deviation (AAD) 0,1813 for Solubility and 0,1423 for Permeability, fall in between training and validation values, indicating that data are well-balanced and equally distributed.

R^2 in Training Set was 0,9729 for Solubility and 0,9702 for Permeability, indicating that the model provides variance explanations of 97%, indicating an excellent fit and good match to Training data. The percentage of %AARD and AAD were estimated for Solubility at (2,8356; 0,1527) and for Permeability (5,0979; 0,1392), which indicated small relative errors, lower dispersion, and good accuracy in Training data.

While in the Validation set, the Coefficient of Determination R^2 was estimated at 0,9329 for Solubility and 0,9277 for Permeability, which is a good value, indicating excellent predictive accuracy (more a 92%) and strong generalization on unseen data. The percentage of %AARD was 5,3110 for Solubility and 16,6156 for Permeability. Although the relative error for the Validation set was slightly larger than the %AARD values for the Training and Overall modeling sets, it indicates improved prediction accuracy on new data. The AAD value for the validation set was 0,2662 for Solubility and 0,3793 for Permeability, indicating lower dispersion and, therefore, greater stability and better accuracy on the tested data.

The linear model in the two properties studied performs well, especially validated by a high R^2 on validation data, confirming excellent predictive capability and good generalization. Small relative deviations (%AARD) support prediction reliability and its stable performance, while differences in AAD values between sets point to expected variability in the training and validation data. Generally, these models demonstrate accurate and consistent predictions with acceptable variability across all data points. Figure 4 shows a curve displaying the relationship between Property P (Solubility, which is reported as Log S (ESOL), and Permeability, which is reported as Consensus Log $P_{o/w}$) values and the Predicted Properties values (PredP) from the linear model, which was estimated by the QSAR-GIC technique.

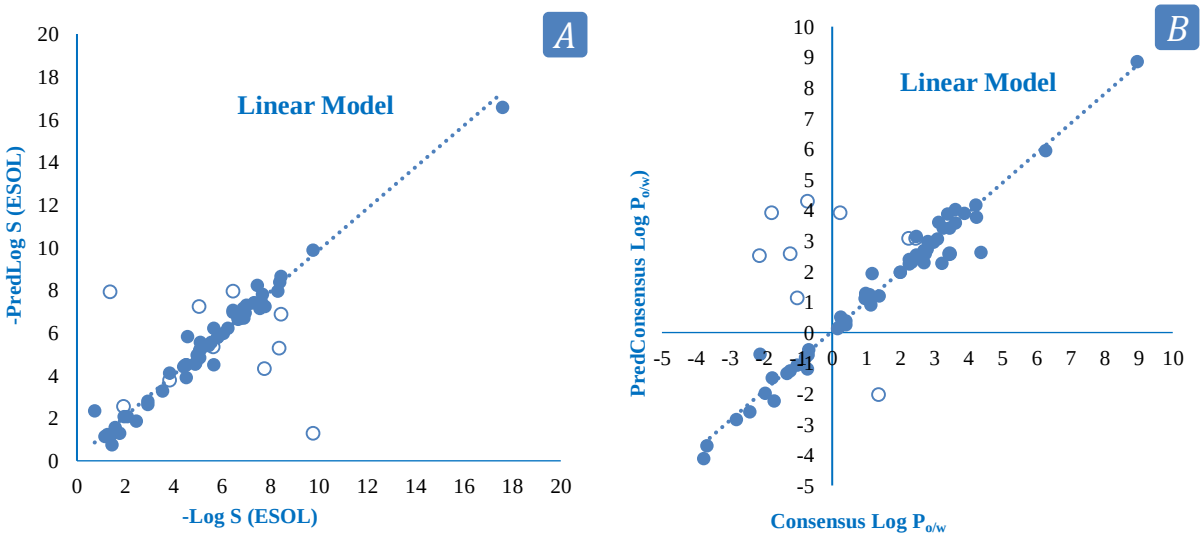

Figure 3. Comparison between Property P and -PredP in the Training set (●) and Validation set (○).

A: $-PredLog\ S\ (ESOL) = f(-Log\ S\ (ESOL))$; ***B***: $PredConsensus\ Log\ P_{o/w} = f(Consensus\ Log\ P_{o/w})$

Through an in-depth understanding of the curve, the modeled and measured points appear to have come together in the midrange area of the curve, clearly evidencing clustering of points around the linear regression line (Linear (PredP)). The fact that points are homogeneously spread near the linear

regression line indicates a robust linear relationship has developed, which means the predictions made from this model will be reliable and precise, or the data exhibits strong agreement and reliability between measured and modeled variables. Naturally, it is common to expect some level of variability in predictions in any modeling framework. The other small deviations of some points showing some level of scatter reflect minor deviations in predictions. A small amount of scatter from the linear regression line does not alter the overall findings related to the modeling, which are still strong. Overall, the curve demonstrates that the two models are linear and stable, capable of delivering precise predictions of Log S (ESOL) and Consensus Log Po/w values.

Table 2. Statistical Error Analysis and Percentage Distribution of %AAD in Linear Model.

	Linear Model	
	Solubility	**Permeability**
%AARD	3,3365	5,8573
AAD	0,1813	0,1423
AAD min	-1,607	-1,721
AAD max	1,149	1,427
AAD < 5	67	67
AAD >5	0	0

Visible in Table 2 are the statistical Error Analysis and Percent Distribution of Average Absolute Deviation in Linear Model Predictions. The Average Absolute Deviation (AAD) of the Linear model in the Overall set reached 0,1813 for Solubility and 0,1423 for Permeability, which a low values, indicating a very high predictive accuracy. The minimum absolute error (%AAD min) is -1,607 for Solubility and -1,721 for Permeability, meaning the model predicted some values with perfect accuracy. while the maximum absolute error is 1,149 for Solubility and 1,427 for Permeability, representing the higher deviation between predicted PredP and Property value (P), we observed that no sample had an absolute error greater than 5, as all samples have an absolute error (AAD) less than 5. This demonstrates that all predictions were highly accurate and confirms the model's accuracy and reliability.

For a more detailed interpretation of both models, Table 3 presents the relative error across the different categories contained in the studied database for Solubility, the relative error (%AARD) is low in the Cationic surfactant classes, followed closely by the Non-ionic, Zwitterionic, and Anionic surfactant classes, which have a a low relative error (%AARD < 5,144). In contrast, For Permeability. It is observed that the relative error (%AARD) is low in the Cationic surfactant classes, whereas it is slightly higher for the Anionic surfactant class and markedly higher for the Non-ionic and

Zwitterionic surfactant classes compared to the other groups. According to Table 3, it can be concluded that the best class, with the highest proportion of property value Property P (Solubility, reported as Log S (ESOL), Permeability, reported as Consensus Log $P_{o/w}$), corresponds to the Cationic surfactant group, which exhibited a low relative error (≈ 0%), followed by the Anionic and Zwitterionic surfactant group. In contrast, the Non-ionic surfactant group showed relatively lower efficiency, as the relative error was slightly higher (10,295%) in Permeability, reflecting decreased prediction precision or higher data variability.

Table 3. Statistical Results Proposed Using the QSAR-GIC Method for property (P) across the four classes of Surfactants interacting with Mitoxantrone in the Linear Model.

Class of Surfactants	No of Surfactants	% Surfactants	Solubility (%AARD)	Permeability (%AARD)
Non-ionic surfactants	31	46,27	2,686	10,295
Anionic surfactants	20	29,85	5,144	4,440
Zwitterionic surfactants	12	17,91	4,432	8,159
Cationic surfactants	4	5,97	4,441E-14	6,273E-13

Non-Linear model

During the first Phase of the study, the QSAR-GIC approach applied to estimate the properties of SFT-MTX complexes used a linear model. However, the AARD calculated from the validation set was higher than those calculated from the training and the overall sets, indicating the linear model was limited in estimating the complex relationships. To make results better, a Non-Linear Model was adopted to increase estimation and prediction accuracy and improve performance. The results revealed that the non-linear model estimated much better and was more reliable at making predictions. This proved that the non-linear model was more effective in treating the problems of the linear model.

The properties of SFT-MTX complexes are significantly linked to the functional groups present in the surfactants and mitoxantrone drug. This relationship was modeled using the QSAR-GIC technique through a Non-Linear approach, according to the following equations:

$$\text{Log S (ESOL)}_{\mathbf{Non-Linear}} = 0{,}0861721 + 1{,}05479 \times \left(\text{Log S (ESOL)}_{\text{drug}} + \left(1{,}488 + \sum_{\text{j}} \text{n}_{\text{j}} \Delta \text{C}_{\text{j}} \right) + 1{,}671 \times E - 0{,}070 \times n_H + 0{,}068 \times n_h \right) + 0{,}00575671$$
$$\times \left(\text{Log S (ESOL)}_{\text{drug}} + \left(1{,}488 + \sum_{\text{j}} \text{n}_{\text{j}} \Delta \text{C}_{\text{j}} \right) + 1{,}671 \times E - 0{,}070 \times n_H + 0{,}068 \times n_h \right)^2 \quad (10)$$

$$\text{Consensus Log P}_{o/w_{\textbf{Non-Linear}}} = 0{,}0588531 + 0{,}984701 \times \left(\text{Consensus Log P}_{o/w_{drug}} + \left(-0{,}622 + \sum_{j} n_j \Delta C_j \right) - 1{,}544 \times E - 0{,}258 \times n_H + 0{,}372 \times n_h \right) - 0{,}00478381 \times \left(\text{Consensus Log P}_{\frac{o}{w}_{drug}} + \left(-0{,}622 + \sum_{j} n_j \Delta C_j \right) - 1{,}544 \times E - 0{,}258 \times n_H + 0{,}372 \times n_h \right)^2 \quad (11)$$

The results of the examined Non-Linear model revealed statistically significant coefficients, which were estimated in all modeling stages. Table 4 presents the statistical parameters corresponding to each stage of the modeling process.

Table 4. Statistical Results of the Proposed Non-Linear Models Using the QSAR-GIC Method for Property (P).

		Non-Linear Model		
		Solubility	**Permeability**	**No of data points**
R^2	Overall set	0,9986	0,9997	67
	Training set	0,9985	0,9997	57
	Validation set	0,9999	1	10
%AARD	Overall set	0,7048	2,0042	67
	Training set	0,7048	2,1540	57
	Validation set	0,5356	1,4300	10
AAD	Overall set	0,0308	0,0499	67
	Training set	0,0308	0,0504	57
	Validation set	0,0208	0,0419	10

Through a thorough investigation of the model and the results from the analysis of the two properties, the interacting group model shows a strong correlation across the range of estimated statistical parameters (R^2, AARD, AAD) and across all three areas of modeling (Overall, Training, and Validation). In general, the two Non-Linear models for Solubility and Permeability presented accurate and consistent predictions across all of the data. The overall Non-Linear models for each property, Solubility and Permeability, performed excellently, indicated by the high values for the coefficient of determination ($R^2 > 0{,}9985$ for Solubility, $R^2 > 0{,}9997$ for Permeability) for all three modeling sets, thereby confirming excellent ability to predict and adequacy in generalization. Average Absolute Relative Deviation Percent (%AARD) results were low across all three areas of modeling (%AARD < 0,7048 for Solubility, %AARD < 2,1540 for Permeability), and Average Absolute Deviation (AAD) was low as well (AAD < 0,0308 for Solubility, AAD < 0,0504 for Permeability), which also confirms predictive reliability and stability of the models. Figure 5 displays a curve illustrating the relationship between property P (solubility, reported as Log S (ESOL), and

permeability, reported as Consensus Log Po/w) values and the predicted properties (Pred(PredP)) generated by the Non-linear model, estimated using the QSAR-GIC technique.

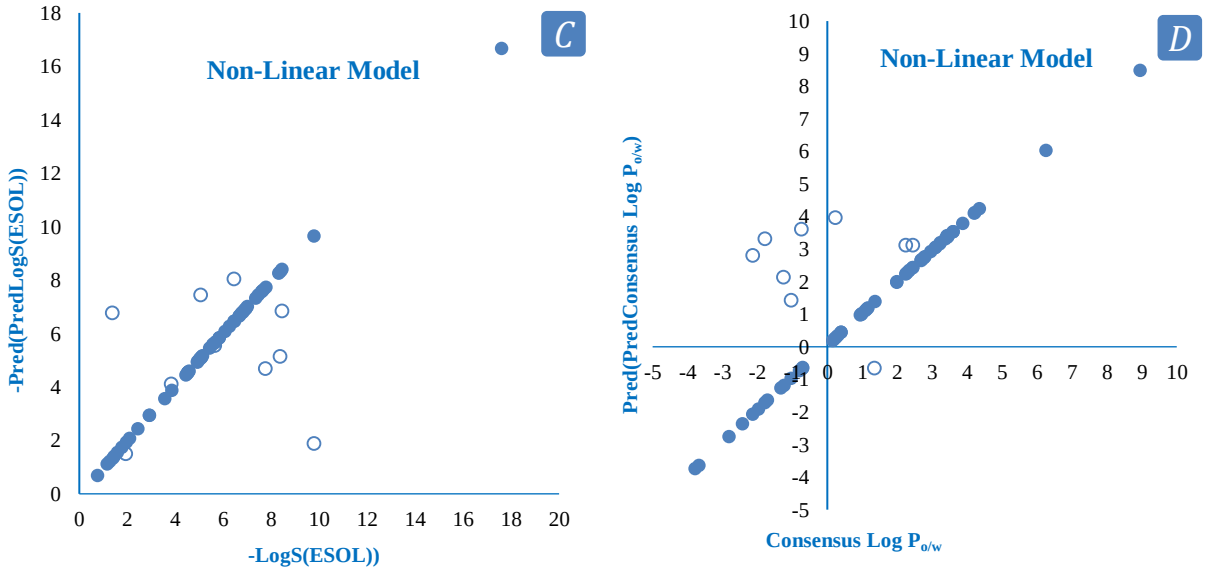

Figure 4. Comparison between Property P and -PredP in the Training set (●) and Validation set (○).

$\boldsymbol{C}$: $-Pred(PredLog\ S\ (ESOL)) = f(-Log\ S\ (ESOL))$; $\boldsymbol{D}$: $Pred(PredConsensus\ Log\ P_{o/w}) = f(Consensus\ Log\ P_{o/w})$

By thoroughly examining both curves, it is clear that modeled and measured points cluster around the central segment of the curve. with a clear clustering around the non-linear regression line (Non-Linear (Pred(PredP)). The consistency in the distribution of the points towards the line demonstrates a strong Non-Linear correlation, which not only confirms both models are accurate but are also highly reliable models, as shown by the trends and consistency of the data sets, especially the modeled and measured points alike. The slight scattering of some of the points in the curves indicates minor differences in prediction effect; however, these do not affect the model's overall quality. To summarize, both curves demonstrate that both models are stable and reliable in predicting Log S (ESOL) and Consensus Log Po/w values.

Table 5. Statistical Error Analysis and Percentage Distribution of %AARD in Non-Linear Model.

	Non-Linear Model	
	Solubility	**Permeability**
%AARD	0,7048	2,0042
AAD	0,0307	0,0499
AAD min	-0,044	-0,461
AAD max	0,902	0,071
AAD < 5	67	67
AAD >5	0	0

As indicated in Table 5, which presents the statistical error analysis and percentage distribution of Average Absolute Deviation in Non-Linear Model Predictions. The Average Absolute Deviation (AAD) of the Non-Linear model in the Overall set reached 0,0307 for Solubility and 0,0499 for Permeability, which are very low values indicating high predictive accuracy. The lowest absolute error (%AAD min) was recorded at -0.044 for Solubility and -0.461 for Permeability, signifying that the value predicted by the model is accurately close to the actual property value. Conversely, the highest absolute error (%AAD max) was recorded at 0.902 for Solubility and 0.071 for Permeability, demonstrating the highest deviation between the predicted property (Pred(PredP)) and the actual property value (P). It was observed that no sample had an absolute error exceeding 5, with all samples showing an absolute deviation (AAD) less than 5. This demonstrates that all predictions were highly accurate, confirming the reliability and precision of both models.

For a more precise estimation, the Non-Linear model was applied on both properties, Solubility and Permeability. Table 6 presents the relative error across the different categories contained in the studied database for Solubility. The relative error (%AARD) is very low in all classes of surfactants (%AARD < 0,870), en particulary the Cationic and the Non-ionic surfactants. For Permeability, the relative error (%AARD) of all classes of surfactant was improved, where its values are low (%AARD < 3,854). So, it can be concluded that the best class, with the highest proportion of property P, is the Non-ionic surfactant, followed closely by the Cationic surfactant, which exhibited a low relative error on two properties. These results reflect that the Non-Linear model of both properties (Solubility and Permeability) was predicted with high precision.

Table 6. Statistical Results Proposed Using the QSAR-GIC Method for property (P) across the four classes of Surfactants interacting with Mitoxantrone in the Non-Linear Model.

Class of Surfactants	No of Surfactants	% Surfactants	Solubility (%AARD)	Permeability (%AARD)
Non-ionic surfactants	31	46,27	0,396	2,014
Anionic surfactants	20	29,85	0,870	1,430
Zwitterionic surfactants	12	17,91	0,846	3,854
Cationic surfactants	4	5,97	0,512	1,448

CONCLUSION

In this study, a hybrid QSAR and Group Interaction Contribution (GIC) approach is applied to estimate and predict the major properties of SFT-MTX complexes generated. This novel study provided valuable results and important insights into the potential to develop and improve the

properties of mitoxantrone. Our study demonstrated the importance of incorporating surfactants with MTX, thereby improving its efficacy and properties, particularly permeability and water solubility.

Theoretical modeling using the QSAR-GIC approach demonstrated the strong relationship between the structure of SFT-MTX complexes and the amelioration of drug' properties; linear modeling provided an AARD% of 3.3365% (R^2 = 0.9728) for solubility and 5.8573% (R^2 = 0.9702) for permeability, and the Non-Linear model provided an AARD% of 0.7048% (R^2 = 0.9986) for solubility and 2.0042% (R^2 = 0.9997) for permeability. Therefore, these results confirm that the surfactants tested exhibit high performance and support the formation of SFT-MTX complexes, thereby improving drug permeability and solubility. This data provides strong, powerful numbers to inform and support future experimental studies.

ABBREVIATIONS

Abbreviations	Full form
SFT	Surfactant
MTX	Mitoxantrone
Log S	Aqueous Solubility
Log P	Partition Coefficient
QSAR	Quantitative Structure–Activity Relationship
GIC	Group Interaction Contributions
xyl	β-D-xylopyranose
Anthra	Anthracenediones
n_H	number of Hydrogen bonds
n_h	number of Hydrophobic bonds
AAD	Average Absolute Relative Deviation
AARD	Average Relative Deviation
R^2	Coefficient of Determination

Declarations

Ethical Approval

This manuscript does not involve any human or animal studies. Therefore, ethical approval, consent to participate, and consent to publish are not applicable.

Funding

No, this research did not receive any funding.

Competing Interests

The authors declare that they have no competing interests or any other interests that could be perceived as influencing the results and/or discussion reported in this paper.

Dual Publication

The authors also declare that the results, data, and figures included in this manuscript have not been published elsewhere and are not under consideration by any other publisher.

Availability of Data and Materials

All materials used in this work are owned by the authors. No permissions are required. This manuscript does not report data generation or analysis; therefore, data availability is not applicable.

REFERENCES

1. Vaishnavi, S.N.; Cardoza, C.; Ahmad Yasin, H.K.; Suraj, N.M.; Srushti, M.T.; Pritish, R.; Kartikeya, S.; Antriksh, G.; Purnima, D.A.; Bapu, R.T.; Jorddy, N.C.; Amit, P.P. Green surfactants (biosurfactants): a petroleum-free substitute for sustainability—comparison, applications, market, and future prospects. ACS Omega **2023**, 8, 11674–11699. https://doi.org/10.1021/acsomega.3c00591

2. Pokhrel, D.R.; Sah, M.K.; Gautam, B.; Basak, H.K.; Bhattarai, A.; Chatterjee, A. A recent overview of surfactant–drug interactions and their importance. RSC Adv. **2023**, 13, 18045–18063. https://doi.org/10.1039/D3RA02883F

3. Rafikul Islam , M.d.; Rehan Alam, M.d.; Munshi, R.I.; Javed, M.K.; Priyankar, S.; Ummey, R.; Abdul Goni, M.d. Estimation of phase separation and thermodynamic parameters of triton X-100 and chlorpheniramine maleate drug mixture: influence of lower chain alcohols. BCSE. **2025**, 39(11), 2109-2127. https://dx.doi.org/10.4314/bcse.v39i11.3

4. White, R.J.; Durr, F.E. Development of mitoxantrone. Invest. New Drugs **1985**, 3, 85–93. https://doi.org/10.1007/BF00174154

5. Azum, N.; Naqvi, A.Z.; Rub, M.A.; Asiri, A.M. Multi-technique approach towards amphiphilic drug–surfactant interaction: a physicochemical study. J. Mol. Liq. **2017**, 237, 264–272. https://doi.org/10.1016/j.molliq.2017.05.066

6. Eszter F.; Katalin U.; Gitta S. Advances and challenges in structural studies of bioactive peptide-Anthracycline Conjugates: A Mass Spectrometric Insight. Int. J. Mol. Sci. **2024**, 26, 4896. https://doi.org/10.3390/ijms26104896

7. Faulds, D.; Balfour, J.A.; Chrisp, P. Mitoxantrone: a review of its pharmacodynamic and pharmacokinetic properties and therapeutic potential in the chemotherapy of cancer. Drugs **1991**, 41, 400–449. https://doi.org/10.2165/00003495-199141030-00007

8. Hande, K.R. Pharmacokinetics and metabolism of mitoxantrone: a review. Cancer Chemother. Pharmacol. **1992**, 30, 1–7. https://doi.org/10.2165/00003088-199018050-00003

9. Abdulrahman, A.A.; Ammar, A.I. Theoretical study of some drugs using quantum chemical, molecular docking and statistical analysis for determination of reduction potential. BCSE. **2025**, 39(11), 2199-2207. https://dx.doi.org/10.4314/bcse.v39i11.9

10. Daina, A.; Michielin, O.; Zoete, V. SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. Sci. Rep. **2017**, 7, 42717. https://doi.org/10.1038/srep42717

11. Mokadem, K.; Korichi, M.; Tumba, K. A new group-interaction contribution method to predict the thermal decomposition temperature of ionic liquids. Chemometr. Intell. Lab. Syst. **2016**, 157, 189–195. https://doi.org/10.1016/j.chemolab.2016.08.001

12. Marrero-Morejón, J.; Pardillo-Fontdevila, E. Estimation of pure compound properties using group-interaction contributions. AIChE J. **1999**, 45, 615–621. https://doi.org/10.1002/aic.690450318

13. Bolton, E.E.; Wang, Y.; Thiessen, P.A.; Bryant, S.H. PubChem: integrated platform of small molecules and biological activities. In: Wheeler, R.A.; Spellmeyer, D.C. (Eds.), Annu. Rep. Comput. Chem. **2008**, 4, 217–241. Elsevier. https://doi.org/10.1016/S1574-1400(08)00012-1

14. Sarkar, A.; Concilio, S.; Sessa, L.; Marrafino, F.; Piotto, S. Advancements and novel approaches in modified AutoDock Vina algorithms for enhanced molecular docking. Results Chem. **2024**, 7, 101319. https://doi.org/10.1016/j.rechem.2024.101319

15. Che, X.; Liu, Q. An accurate and universal protein–small molecule batch docking solution using AutoDock Vina. Results Eng. **2023**, 19, 101335. https://doi.org/10.1016/j.rineng.2023.101335

16. Hanwell, M.D.; Curtis, D.E.; Lonie, D.C.; Vandermeersch, T.; Zurek, E.; Hutchison, G.R.; Avogadro: an advanced semantic chemical editor, visualization, and analysis platform. J. Cheminform. **2012**, 4, 17. https://doi.org/10.1186/1758-2946-4-17

17. Dassault Systèmes; BIOVIA Discovery Studio®: Overview datasheet. **2021**. https://www.3ds.com/products-services/biovia

18. Constantinou, L.; Gani, R. New group contribution method for estimating properties of pure compounds. AIChE J. **1994**, 40, 1697–1710.

19. Pardillo-Fontdevila, E.; Rubio, R.G. A group-interaction contribution approach: a new strategy for the estimation of physico-chemical properties of branched isomers. Chem. Eng. Commun. **1998**, 163, 245–254