12th International Electronic Conference on Synthetic Organic Chemistry (ECSOC-12) 1-30 November 2008
http://www.usc.es/congresos/ecsoc/12/ECSOC12.htm & http://www.mdpi.org/ecsoc-12

[G0016]

MOOP-DESIRE-based Simultaneous Optimization of the Analgesic, Antiinflammatory, and Ulcerogenic Profiles of 3-(3-Methylphenyl)-2-Substituted Amino-3H-Quinazolin-4-ones

Maykel Cruz-Monteagudo,[1,2,3] Fernanda Borges,[1] M. Natália D. S. Cordeiro[4]

[1]*Department of Organic Chemistry, Faculty of Pharmacy, University of Porto, 4150-047 Porto, Portugal*
[2]*Applied Chemistry Research Center, Faculty of Chemistry and Pharmacy, Central University of ''Las Villas'', Santa Clara 54830, Cuba*
[3]*Chemical Bioactive Center, Central University of ''Las Villas'', Santa Clara 54830, Cuba*
[4]*REQUIMTE, Department of Chemistry, Faculty of Sciences, University of Porto, 4169-007 Porto, Portugal*

Abstract: Up to now, very few reports have been published concerning the application of multiobjective optimization (MOOP) techniques to quantitative structure–activity relationship (QSAR) studies. However, none reports the optimization of objectives related directly to the desired pharmaceutical profile of the drug. In this work, for the first time, it is proposed a MOOP method based on Derringer's desirability function that allows conducting global QSAR studies considering simultaneously the pharmacological, pharmacokinetic and toxicological profile of a set of molecule candidates. The usefulness of the method is demonstrated by applying it to the simultaneous optimization of the analgesic, antiinflammatory, and ulcerogenic properties of a library of fifteen 3-(3-methylphenyl)-2-substituted amino-3H-quinazolin-4-one compounds. The levels of the predictor variables producing concurrently the best possible compromise between these properties is found and used to design a set of new optimized drug candidates. Our results also suggest the relevant role of the bulkiness of alkyl substituents on the C-2 position of the quinazoline ring over the ulcerogenic properties for this family of compounds. Finally, and most importantly, the desirabilitybased MOOP method proposed is a valuable tool and shall aid in the future rational design of novel successful drugs.

Key words: chemoinformatics; drug discovery; global QSAR; multiobjective optimization; NSAIDs; overall desirability function; ulcerogenic index

Introduction

Developing a successful drug is a complex and lengthy process and failure at the development stage is due to multiple factors, such as lack of efficacy, poor bioavailability, and toxicity.[1] Improving the profile of a candidate drug requires finding the best compromise between various, often competing objectives. In fact, the ideal drug should have the highest therapeutic efficacy, the highest bioavailability and the lowest toxicity, which highlights the multiobjective nature of the drug discovery and development process. But even when a potent candidate has been identified, the pharmaceutical industry routinely tries to optimize the remaining objectives one at a time, which often results in expensive and time-consuming cycles of trial and error.[2] Roughly

75% of the total costs during the development of a drug are attributed to poor pharmacokinetics and/or toxicity.[3]

In the last years, the drug discovery/development process has been gaining in efficiency and rationality because of the continuous progress and application of chemoinformatics methods.[2] In particular, the quantitative structure–activity relationship (QSAR) paradigm has long been of interest in the drug-design process,[4] redirecting our thinking about structuring medicinal chemistry.[5] Yet, standard chemoinformatics approaches usually ignore multiple objectives and optimize each biological property sequentially.[6–17] Nevertheless, some efforts have been made recently toward unified approaches able of modeling multiple pharmacological, pharmacokinetic, or toxicological properties onto a single QSAR equation.[18–21]

Multiobjective optimization (MOOP) methods introduce a new philosophy for reaching optimality based on compromises among the various objectives. These methods aim at discovering the global optimal solution by optimizing several dependent properties simultaneously. The major benefit of MOOP methods is that local optima corresponding to one objective can be avoided by taking into account the whole spectra of objectives, leading thus to a more efficient overall process.[22]

Several applications of MOOP methods have appeared lately ranging from substructure mining to docking, including inverse quantitative structure property relationship (QSPR) and QSAR.[22, 24-33] Most of these MOOP applications have been based on the following approaches: weighted-sum-of-objective-functions (WSOF)[23] and pareto-based methods.[22] An excellent review on the subject has been most recently published by Nicolaou et al.[22]

Despite the availability of numerous optimization objectives, MOOP techniques have only recently been applied to the building of QSAR models. Actually, very few reports exist of the application of MOOP methods to QSAR. Nicolotti et al.[31] employed a variant of an evolutionary algorithm called multiobjective genetic programming that used pareto ranking to optimize the QSAR models. A number of conflicting objectives including model accuracy, number of terms, internal complexity, and interpretability of the descriptors used in the model were considered. On the other hand, Stockfisch[32] proposed a nonevolutionary multiobjective technique called the partially unified multiple property recursive partitioning method for building QSAR models. This method was successfully used to construct models to analyze selectivity relationships between cyclooxygenase 1 and 2 inhibitors.[33] Up to now, no QSAR study has nevertheless reported the simultaneous optimization of competing objectives directly related with the definitive pharmaceutical profile of drugs, such as therapeutic efficacy, bioavailability, and/or toxicity.

In the present work, we are proposing a MOOP method based on Derringer's desirability function[34] that allows running global QSAR studies jointly considering multiple properties of interest to the drug-design process. The method proposed is applied to a small set of 2-substituted amino-3H-quinazolin-4-one compounds with the aim of simultaneously optimizing their analgesic, antiinflammatory and ulcerogenic properties, as well as suggesting new improved drug candidates of this kind.

Materials and Methods

Data set. Our prediction models (PMs) were developed using a library of fifteen 3-(3-methylphenyl)-2-substituted amino-3H-quinazolin-4-one compounds published by Alagarsamy et al.[35] The analgesic activity (An) reported for these compounds (in %) was measured using the

tail-flick method in Wistar albino mice,[36] whereas the antiinflammatory activity (Aa) reported (in %) was evaluated using the carrageenan-induced paw oedema test in rats.[36] The ulcerogenic index (U) was determined by the method of Ganguly and Bhatnagar,[37] and the ulcers were induced in rats using the method described by Goyal et al.[38] All these assays[35] were performed by administering a maximum dose of 20 $mg \cdot kg^{-1}$.

Computational methods. The structures of all compounds were first drawn with the aid of ChemDraw software package,[39] and reasonable starting geometries obtained by resorting to the MM2 molcular mechanics force field.[40,41] Molecular structures were then fully optimized with the PM3 semiempirical Hamiltonian,[39] implemented in the MOPAC 6.0 program.[42] Here, it should be remarked that the final molecular structures pertain only to the compounds' global minimum energy conformations, and indeed, further molecular simulations and/or docking studies would be desirable to reach reliable conclusions about conformational requirements and ligand–receptor interactions. But the point of any QSAR model is to have a set of readily calculated descriptors, and such an approach would require much more extensive calculations.

Subsequently, the optimized structures were brought into the DRAGON software package[43] for computing a total of 120 atom-centered fragment (ACF) molecular descriptors.[44] ACF descriptors were chosen because their simple nature offers easy structural interpretation. To reduce noisy information that could lead to chance correlations, descriptors having constant or near constant values as well as highly pair-correlated ($|R|$ [0.95) were excluded. Thus, from an initial set of 120 ACF molecular descriptors only 12 remained for further variable selection. Table 1 summarizes and describes the ACF molecular descriptors used in this work.

Table 1. Symbols and description for the 12 acf descriptors remaining after variable reduction.

Symbol	Description	Symbol	Description
C-001	CH3R/CH4	C-038	Al—C(=X)—Al
C-002	CH2R2	C-039	Ar—C(=X)—R
C-024	R…CH..R	H-046	H attached to C0(sp^3) no X attached to next C
C-025	R..CR..R	H-052	H attached to C0(sp^3) with 1X attached to next C
C-026	R…CX..R	O-061	O…
C-037	Ar—CH=X	Cl-089	Cl attached to C1(sp^2)

The task of selecting the descriptors that will be more suitable to model the activity of interest is complicated, as there are no absolute criteria for ruling such selection. Approaches implementing genetic algorithms (GA) for solving optimization problems in ANN[45–47] and SVM[48] based QSAR have been recently reported. Herein, the GA optimization technique was applied for variable selection[49–52] by using the BuildQSAR software package.[53,54] The particular GA simulation conditions applied here were 10,000 generations, 300 model populations and 35% of mutation probabilities. Figure 1 depicts the ACF molecular descriptors selected by the GA method, which were finally applied to model the analgesic, antiinflammatory, and ulcerogenic properties of the present compounds.

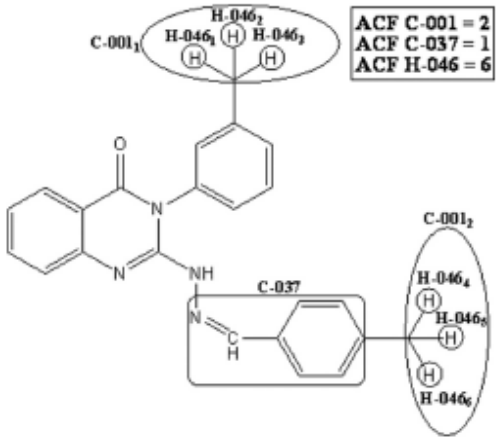


Figure 1. Atom-centered fragments (ACF) descriptors for compound AS14.

As to the modeling technique, we opted for a regression based approach; in this case, the regression coefficients and statistical parameters were obtained by multiple linear regression (MLR) analysis by means of the STATISTICA software package.[55] For each PM, the goodness of fit was assessed by examining the determination coefficient (R^2), the adjusted determination coefficient ($Adj.R^2$), the standard deviation (s), Fisher's statistics (F), as well as the ratio between the number of compounds (N), and the number of adjustable parameters (ρ) in the model, known as the q statistics. The predictive ability of the models was evaluated by means of internal cross-validation (CV), specifically by the leave-one-out (LOO) technique.[56] Quality of the new models (Q^2_{LOO}) gives then an estimated measure of the predictive ability of the full model.

We have also checked the validity of the preadopted parametric assumptions, another important aspect in the application of linear multivariate statistical-based approaches.[57] These include the linearity of the modeled property, homoscedasticity (or homogeneity of variance) as well as the normal distribution of the residuals and nonmulticollinearity between the descriptors.[58]

Finally, the applicability domain of the final PMs was identified by a leverage plot, that is to say, a plot of the standardized residuals.vs. leverages for each training compound.[56,59] The leverage (h_i) of a compound in the original variable space measures its influence on the model, and is calculated as follows:

$$h_i = \mathbf{t}_i (\mathbf{T}^T \mathbf{T})^{-1} \mathbf{t}_i^{\,T} \qquad (i = 1,..., N) \qquad \textbf{(1)}$$

where ti is the descriptor vector of that compound and T is the model matrix derived from the training set descriptor values. In addition, the warning leverage h^* is defined as

$$h^* = 3 \times p'/N \qquad \textbf{(2)}$$

Leverage values can be calculated for both training compounds and new compounds. A leverage higher than the warning leverage h^* means that the compound predicted response can be extrapolated from the model, and thus, the predicted value must be used with great care. On the other hand, a standardized residual value greater than two indicates that the value of the dependent variable for the compound is significantly separated from the remainder training data, and hence, such predictions must be considered with much caution too. In this work, only predicted data for new compounds belonging to the applicability domain of the training set were considered reliable.

***M*ulti*O*bjective *OP*timization based on the *D*esirability *E*stimation of *S*everal *I*nterrelated *RE*sponses (*MOOP-DESIRE*)**. Improving the profile of a molecule for the drug discovery and development process requires the simultaneous optimization of several different objectives. The ideal drug should have the highest therapeutic efficacy and bioavailability, as well as the lowest toxicity. Because of the conflicting relationship among the aforementioned properties, to discover such a drug is almost a chimera and, if possible, an extremely difficult, expensive and time-consuming task. However, finding the best compromise between such objectives is an accessible and more realistic target (see Figure 2).

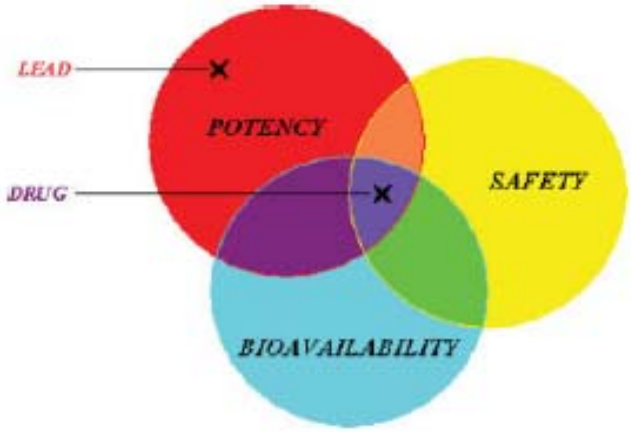


Figure 2. Graphic representation of the compromise between therapeutic efficacy (potency), bioavailability (ADME properties), and toxicity (safety) required to reach a successful drug.

In this work, we are proposing a MOOP technique based on the desirability estimation of several interrelated responses (MOOP-DESIRE) as a tool for performing global QSAR studies, taking into account both the pharmacological, pharmacokinetic and toxicological profiles of a set of candidates. MOOP-DESIRE methodology is intended to find the most desirable solution that optimizes a multiobjective problem by using the Derringer's desirability function,[56,59] specifically addressed to confer rationality to the drug development process.

The process of simultaneous optimization of multiple properties of a drug candidate can be described as follows. From now on, the terms ''response variable'' and ''independent variables'' should be understood as any property to be optimized, and any set of molecular descriptors used to model each property, respectively.

1. Prediction models set-up.

Each response variable (Y_i) is related to the n independent variables (X_n) by an unknown functional relationship, often (but not necessarily) approximated by a linear function. Each predicted response ($\hat{Y}_i$) is then estimated by a least-squares regression technique.

In some cases, the developed prediction model for some response may share the same independent variables of the other responses' prediction models, but with different coefficients. In this atypical case, attaining the best compromise among the responses turns out to be simpler. Actually, due to the multiplicity of factors involved in the "drugability" of a molecule, one should not expect that the same subset of independent variables can optimally explain both different types of biological properties (especially conflicting properties like potency and toxicity). However, in the latter case, there is still a way to maximize the desirability of both

biological properties, *i.e.* to set-up a global prediction model where the predicted values of each response are fitted to a linear function using the whole subset of independent variables employed in modeling the k original responses. Here, the independent variables used in computing the predicted values for the original responses will remain the same. Independent variables not used in computing the predicted values for the original responses will be set to zero.

2. Desirability functions selection and evaluation.

For each predicted response $\hat{Y}_i$, a desirability function d_i assigns values between 0 and 1 to the possible values of $\hat{Y}_i$. This transformed response d_i, can have many different shapes. Regardless of the shape, $d_i = 0$ represents a completely undesirable value of $\hat{Y}_i$, and $d_i = 1$ represents a completely desirable or ideal response value. The individual desirabilities are then combined using the geometric mean, which gives the overall desirability D:

$$D = (d_1 \times d_2 \times ... \times d_k)^{\frac{1}{k}} \quad (3)$$

with k denoting the number of responses.

This single value of D gives the overall assessment of the desirability of the combined response levels. Clearly, the range of D will fall in the interval [0, l] and will increase as the balance of the properties becomes more favorable. Notice that if for any response $d_i = 0$, then the overall desirability is zero. Thus, the desirability maximum will be at the levels of the independent variables that simultaneously produce the maximum desirability, given the original models used for predicting each original response.

Depending on whether a particular response is to be maximized, minimized, or assigned a target value, different desirability functions can be used. Here we used the desirability functions proposed by Derringer and Suich[34].

Let L_i, U_i and T_i be the lower, upper, and target values, respectively, that are desired for the response $\hat{Y}_i$, with $L_i \leq T_i \leq U_i$.

If a response is of the *target* best kind, then its individual desirability function is defined as:

$$d_i = \begin{cases} \left[\dfrac{\hat{Y}_i - L_i}{T_i - L_i}\right]^s & \text{if} \quad L_i \leq \hat{Y}_i \leq T_i \\ \left[\dfrac{\hat{Y}_i - U_i}{T_i - U_i}\right]^t & \text{if} \quad T_i < \hat{Y}_i \leq U_i \\ 0 & \text{if} \quad \hat{Y}_i < L_i \quad \text{or} \quad \hat{Y}_i > U_i \end{cases} \quad (4)$$

If a response is to be maximized instead, its individual desirability function is defined as:

$$d_i = \begin{cases} 0 & \text{if} \quad \hat{Y}_i \leq L_i \\ \left[\dfrac{\hat{Y}_i - L_i}{T_i - L_i}\right]^s & \text{if} \quad L_i < \hat{Y}_i < T_i \\ 1 & \text{if} \quad \hat{Y}_i \geq T_i = U_i \end{cases} \quad (5)$$

In this case, T_i is interpreted as a large enough value for the response, which can be U_i.

Finally, if one wants to minimize a response, one might use:

$$d_i = \begin{cases} 1 & \text{if} \quad \hat{Y}_i \le T_i = L_i \\ \left[\dfrac{\hat{Y}_i - U_i}{T_i - U_i}\right]^s & \text{if} \quad U_i < \hat{Y}_i < T_i \\ 0 & \text{if} \quad \hat{Y}_i \ge U_i \end{cases} \qquad (\mathbf{6})$$

Here, T_i denotes a small enough value for the response, which can be L_i. Moreover, the exponents s and t determine how important is to hit the target value T_i. For $s = t = 1$, the desirability function increases linearly towards T_i. Large values for s and t should be selected if it is very desirable that the value of $\hat{Y}_i$ be close to T_i or increase rapidly above L_i. On the other hand, small values of s and t should be chosen if almost any value of $\hat{Y}_i$ above L_i, and below U_i are acceptable or if having values of $\hat{Y}_i$ considerably above L_i are not of critical importance[34].

In this way, one may predict the overall desirability for each drug candidate determined by k responses, which in turn are at the same time determined by a specific set of independent variables. However, as the Derringer's desirability function is built using the estimated responses $\hat{Y}_i$, there is no way to know how reliable the predicted D value of each candidate is.

To overcome this shortcoming, we propose here a statistical parameter, the *overall desirability's determination coefficient* (R^2_D), which measures the effect of the set of independent variables X_n in reducing the uncertainty when predicting the D values.

If the response variable is estimated as a continuous function of the independent variables X_n, the individual desirabilities d_i are continuous functions of the estimated $\hat{Y}_i$'s (eqs. **4**, **5** and **6**), and the overall desirability D is a continuous function of the d_i's (eq. **3**), then D is also a continuous function of the X_n. Therefore, R^2_D can be computed in analogy with the so-called determination coefficient R^2. Specifically, R^2_D is computed by using the observed D_{Yi} (calculated from Y_i) and the predicted $D_{\hat{Y}i}$ (calculated from $\hat{Y}_i$) overall desirability values instead of using directly the measured (Y_i) and predicted ($\hat{Y}_i$) response values.

$$R^2_D = 1 - \frac{SSE}{SSTO} = 1 - \frac{\sum (D_{Y_i} - D_{\hat{Y}_i})^2}{\sum (D_{Y_i} - \overline{D}_{Y_i})^2} \qquad (\mathbf{7})$$

where D_{Yi} and $D_{\hat{Y}i}$ have been defined previously. $\overline{D}_{Y_i}$ is the mean value of D for the Y_i responses of each case included in the data set, *SSTO* is the total sum of squares, and *SSE* is the sum of squares due to error.

Similar to R^2, the *adjusted overall desirability's determination coefficient* (*Adj.* R^2_D) can be computed as shown below.

$$Adj.\ R^2_D = 1 - \frac{SSE}{SSTO} = 1 - \frac{\dfrac{\sum (D_{Y_i} - D_{\hat{Y}_i})^2}{N-2}}{\dfrac{\sum (D_{Y_i} - \overline{D}_{Y_i})^2}{N-1}} \qquad (\mathbf{8})$$

Like this, both R^2_D and *Adj.* R^2_D have the same properties of R^2 and *Adj.* R^2. Thus, both will fall in the range [0, 1] and the larger R^2_D /*Adj.* R^2_D is, the lower is the uncertainty in predicting D by using a specific set of independent variables X_n[60].

Since R^2_D and *Adj.* R^2_D measure the goodness of fit rather than the predictive ability of a certain PM, it is advisable to use an analogous of the leave one out cross validation determination coefficient (Q^2_{LOO}) to establish the reliability of the method in predicting D. For this, the *overall*

desirability's LOO–CV determination coefficient (Q^2_D) can be defined in an analogous way as R^2_D:

$$Q^2_D = 1 - \frac{SSE_{LOO-CV}}{SSTO} = 1 - \frac{\sum (D_{Y_i} - D_{\hat{Y}_i}(LOO-CV))^2}{\sum (D_{Y_i} - \overline{D}_{Y_i})^2} \quad (\mathbf{9})$$

where SSE_{LOO-CV} and $D_{\hat{Y}i}$ (*LOO-CV*) are the leave one out cross validation square sum of residuals and the predicted overall desirability by LOO-CV, respectively.

In this way, we can have a measure of how reliable will be the simultaneous optimization of the *k* responses over the independent variables domain.

3. Multi-objective optimization.

As seen before, the desirability function condenses a multivariate optimization problem into a univariate one. Thus, the overall desirability *D* can be maximized over the independent variables domain. For accomplish this, one can use the *Response/Desirability Profiler* option of any of the modules of regression or discriminant analysis implemented in STATISTICA[55]. The overall desirability *D* is optimized with the "*Use general function optimization*" option, that is, the simplex method of function optimization, [61-63] or the "*Optimum desirability at exact grid points*" option, which performs exhaustive searches for the optimum desirability at exact grid points. The first option is usually faster, but the default option is the later one, except when the number of predicted values that must be computed to perform the exhaustive grid search exceeds 200000, in which case the *Use general function optimization* option becomes the default.

The final result is to find the optimal levels (or an optimal range) of the independent variables that optimize simultaneously the *k* responses determining the final quality of the product. In this way, the best possible compromise between the *k* responses is found and consequently the highest overall desirability for the final compound is reached (*i.e.* the more enviable drug candidate).

Desirability Functions Specifications. Response/desirability profiling allows one to trace the response surface produced by fitting the observed response(s) using equation(s) based on the levels of the independent variables.[34] That is to say, one can inspect the predicted values for the response(s) at different combinations of levels of the independent variables, specify desirability function(s) for the response(s), and search for the levels of the independent variables that simultaneously produce the most desirable response or the best possible compromise among responses leading to the most desirable solution (candidate molecule).

In the present work, the optimization of the overall desirability was carried on by the Optimum desirability at exact grid points option of the general regression module of STATISTICA.[55] Three desirability functions, one for each response, were fitted. Specifically, the analgesic and anti-inflammatory activities ought to be maximized [eq. (3)]. For estimating their d_i's, the lower value L_i was set to 25%, and the upper value U_i, made equal to the target value T_i, was set to 100% for both responses. In contrast, the ulcerogenic index must be minimized where $L_i = T_i = 0$ and $U_i = 1.73$ [eq. (4)]. The value of $U_i = 1.73$ corresponds to the ulcerogenic index of aspirin (measured with the same protocol used for the training set[35]), a NSAID with a recognized ulcerogenic ability. Furthermore, the spline method[64,65] was used for fitting the desirability function and surface/contours maps, and the current level of each independent variable was set equal to its optimum value. As to the *s* and *t* parameters, these were fixed at 1.00 by assuming that the desirability functions increase linearly towards T_i on the three responses.

Results and Discussion

MOOP–DESIRE-based optimization. Following the strategy outlined previously, we began by seeking the best linear models relating each property to the ACF molecular descriptors. One should emphasize here that the reliability of the final results of the optimization process strongly depends on the quality of the initial set of PMs.

One MLR-based PM containing two ACF[44] variables previously selected by GA was developed for each property. The resulting best-fit models are given in Table 2 together with the statistical regression parameters, whereas the computed ACF molecular descriptors along with the measured and predicted values of the analgesic activity, antiinflammatory activity, and the ulcerogenic index for the 15 training compounds are shown in Table 3.

Table 2. Regression coefficients and statistical parameters for the MLR models.

Analgesic activity (*An*) model								
$An = 51.762(\pm 2.155) + 8.333(\pm 0.957) \cdot C - 001 - 6.929(\pm 1.534) \cdot C - 037$								
N	*R*	R^2	R^2 Adj.	Q^2	SPRESS	ρ	*F*	*p*
15	0.967	0.935	0.923	0.905	3.143	5.000	85.15699	0.000000
Antiinflammatory activity (*Aa*) model								
$Aa = 36.708(\pm 1.789) + 5.527(\pm 1.232) \cdot C - 001 + 1.475(\pm 0.430) \cdot H - 046$								
N	*R*	R^2	R^2 Adj.	Q^2	SPRESS	ρ	*F*	*p*
15	0.942	0.887	0.869	0.827	3.526	5.000	47.46719	0.000002
Ulcerogenic index (*U*) model								
$U = 0.718(\pm 0.044) - 0.056(\pm 0.020) \cdot C - 001 + 0.137(\pm 0.032) \cdot C - 037$								
N	*R*	R^2	R^2 Adj.	Q^2	SPRESS	ρ	*F*	*p*
15	0.896	0.803	0.771	0.713	0.065	5.000	24.56766	0.000057

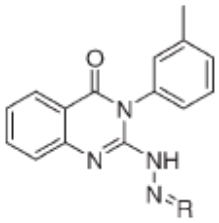
As can be noticed, the models are good in both statistical significance and predictive ability (see Table 2). Good overall quality of the models is revealed by the large *F* and small p values, satisfactory ρ values (ρ = 5), along with R^2 and $Adj.R^2$ (goodness of fit) values ranging from 0.803 to 0.935 and 0.771 to 0.923, respectively; as well as Q^2_{LOO} (predictivity) values between 0.713 and 0.905.

The next step is to find out if the basic assumptions of MLR analysis (linearity, normality, homoscedaticity and non multicollinearity,) are fulfilled. No violations of such assumptions were found that could compromise the reliability of the resulting predictions.

Another aspect deserving special attention is the applicability domain of the several PMs. The leverage values (*h*) and standardized residuals (Std. Res.) related to three PMs for the 15 training compounds are shown in Table 4, whereas Figure 3 shows the corresponding leverage plots. From these plots, the applicability domain is established inside a squared area within ±2 standard deviations and a leverage threshold h^* of 0.6 (Notice that each model was fitted using 15 training compounds and included 3 adjustable parameters: two ACF descriptors plus the intercept.). As seen in Figure 3, only one compound of the training set has a leverage greater than h^* for Aa, but shows standard deviation values within the limits, which implies that it should not be considered an outlier but instead as an influential compound.

Table 3. Computed ACF descriptors (C-001, C-037, and H-046), measured and predicted values for the analgesic (An) and antiinflammatory (Aa) activities, plus the ulcerogenic index (U) of the training set compounds.

3-(3-Methylphenyl)-2-substituted amino-3*H*-quinazoline-4-one



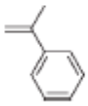
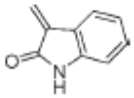
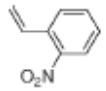
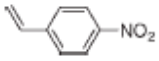
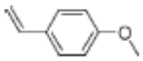
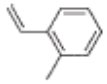
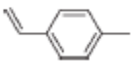
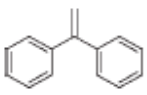
Compound	R	C-001	C-037	H-046	An_{meas} (%)	Aa_{meas} (%)	U_{meas} (%)	An_{pred} (%)	Aa_{pred} (%)	U_{pred} (%)
AS1		3	0	6	76	59	0.53	77	62	0.55
AS2		3	0	9	79	68	0.59	77	67	0.55
AS3		3	0	8	78	69	0.56	77	65	0.55
AS4		1	0	9	59	56	0.60	60	56	0.66
AS5		2	0	3	68	55	0.63	68	52	0.61
AS6		1	0	3	60	45	0.65	60	47	0.66
AS7		1	1	3	58	50	0.69	53	47	0.80
AS8		1	1	3	50	43	0.89	53	47	0.80
AS9		1	1	3	53	47	0.83	53	47	0.80
AS10		1	1	3	58	46	0.85	53	47	0.80
AS11		1	1	3	52	48	0.82	53	47	0.80
AS12		1	1	3	53	47	0.80	53	47	0.80
AS13		2	1	6	58	53	0.69	62	57	0.74
AS14		2	1	6	60	53	0.71	62	57	0.74
AS15		1	0	3	59	49	0.68	60	47	0.66

Table 4. Leverages (*h*) and standardized residuals (Std. Res.) for the analgesic (An) and antiinflammatory (Aa) activities, plus the ulcerogenic index (U) prediction models.

Compound	*h*(*An*)	Std. Res. (*An*)	*h*(*Aa*)	Std. Res. (*Aa*)	*h*(*U*)	Std. Res. (*U*)
AS1	0.276	−0.29	0.317	−1.11	0.276	−0.36
AS2	0.276	0.85	0.328	0.51	0.276	0.75
AS3	0.276	0.47	0.279	1.38	0.276	0.19
AS4	0.276	−0.42	0.776	0.17	0.276	−1.14
AS5	0.143	−0.16	0.226	0.99	0.143	0.45
AS6	0.276	−0.04	0.112	−0.58	0.276	−0.22
AS7	0.133	1.84	0.112	1.18	0.133	−2.02
AS8	0.133	−1.21	0.112	−1.29	0.133	1.68
AS9	0.133	−0.06	0.112	0.12	0.133	0.57
AS10	0.133	1.84	0.112	−0.23	0.133	0.94
AS11	0.133	−0.45	0.112	0.47	0.133	0.39
AS12	0.133	−0.06	0.112	0.12	0.133	0.02
AS13	0.200	−1.34	0.089	−1.27	0.200	−0.98
AS14	0.200	−0.57	0.089	−1.27	0.200	−0.61
AS15	0.276	−0.42	0.112	0.82	0.276	0.34

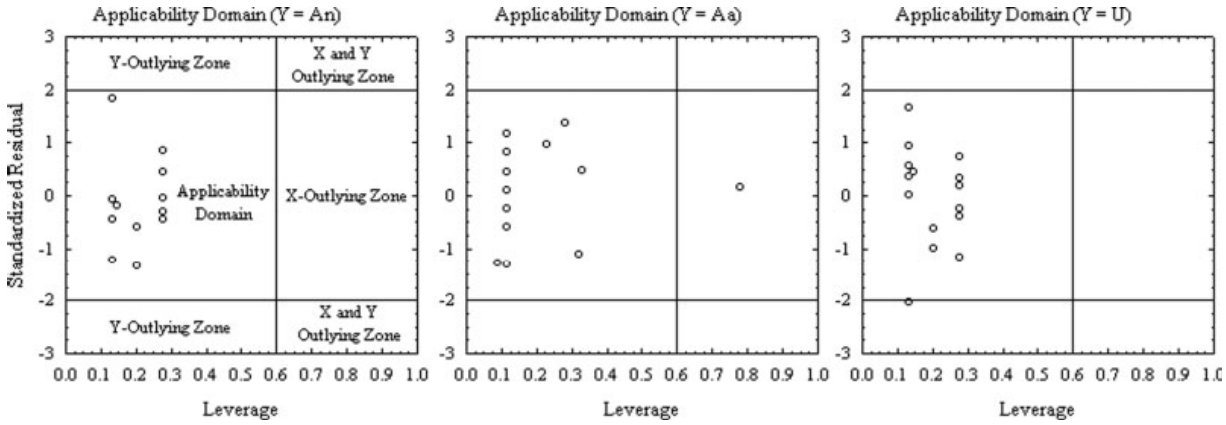


Figure 3. Leverage plots based on the three MLR models; i.e. plots of the standardized residuals.vs. leverage values for the training compounds, with a warning leverage of 0.6.

So far, we have demonstrated the satisfactory accuracy and the predictive ability of the developed PMs. We may now thus proceed with an adequate level of confidence to the simultaneous optimization of the analgesic, antiinflammatory and ulcerogenic properties for the set of compounds. Here it is important to remark that, since *D* is maximized directly over the independent variables domain, and at the same time, the predicted *D* values depend on the initial set of PMs, one should consider the applicability domain of each PM to determine the optimum level of each independent variable as well as for the selection of the optimal solution(s).

First, the predicted values for each property were used to fit a model containing all the independent variables (C-001, C-037, and H-046) applied in modeling the original properties (An, Aa and U). So, for the An and U properties, the original values of C-001 and C-037 were used (H-046 values were set to zero), and for Aa, the original values of C-001 and H-046 (C-037 values were set to zero). In so doing, one is able to discriminate opposite objectives like efficacy (analgesic and anti-inflammatory activities) and toxicity (ulcerogenic ability) with total or partial overlap of the descriptors set used to built the PMs (Notice that the An and U models both contain the C-001 and C-037 descriptors, and the An, Aa, and U models share a common descriptor, i.e. C-001; see Table 2.). Once the model has been set up, the desirability functions

for each property (d_i's) might be specified. In order to obtain candidate(s) with high analgesic and anti-inflammatory activities as well as low ulcerogenic index, An and Aa should be maximized [eq. (3)] and U minimized [eq. (4)]. In addition, the individual d_i values for the An, Aa, and U properties were determined by setting the L_i, U_i and T_i values as referred previously. Then, the three d_i's were combined into the single overall desirability D by means of eq. (1).

The expected and predicted desirability values attributable to each response plus the overall desirability for the training set are depicted in Table 5. In addition, the LOO-CV predicted values and the desirability values for each response, along with the overall desirability values are shown in Table 6.

Table 5. Expected and predicted values for the desirability due to the analgesic activity [d(An)], antiinflammatory activity [d(Aa)], ulcerogenic index [d(U)], and overall desirability [D(An-Aa-U)].

Compound	$d(An)$	$d(An)_{pred}$	$d(Aa)$	$d(Aa)_{pred}$	$d(U)$	$d(U)_{pred}$	$D(An\text{-}Aa\text{-}U)$	$D(An\text{-}Aa\text{-}U)_{pred}$
AS1	0.68	0.69	0.45	0.49	0.69	0.68	0.60	0.62
AS2	0.72	0.69	0.57	0.56	0.66	0.68	0.65	0.64
AS3	0.71	0.69	0.59	0.53	0.68	0.68	0.65	0.63
AS4	0.45	0.47	0.41	0.41	0.65	0.62	0.50	0.49
AS5	0.57	0.57	0.40	0.36	0.64	0.65	0.53	0.51
AS6	0.47	0.47	0.27	0.29	0.62	0.62	0.43	0.44
AS7	0.44	0.37	0.33	0.29	0.60	0.54	0.45	0.39
AS8	0.33	0.37	0.24	0.29	0.49	0.54	0.34	0.39
AS9	0.37	0.37	0.29	0.29	0.52	0.54	0.38	0.39
AS10	0.44	0.37	0.28	0.29	0.51	0.54	0.40	0.39
AS11	0.36	0.37	0.31	0.29	0.53	0.54	0.39	0.39
AS12	0.37	0.37	0.29	0.29	0.54	0.54	0.39	0.39
AS13	0.44	0.49	0.37	0.43	0.60	0.57	0.46	0.49
AS14	0.47	0.49	0.37	0.43	0.59	0.57	0.47	0.49
AS15	0.45	0.47	0.32	0.29	0.61	0.62	0.44	0.44
Overall desirability function [$D(An\text{-}Aa\text{-}U)$] statistics[a]				$R^2_{D(An\text{-}Aa\text{-}U)} = 0.934$			Adj.$R^2_{D(An\text{-}Aa\text{-}U)} = 0.929$	

[a]Statistical quality of the overall desirability function estimated by the overall desirability determination coefficient (R^2_D) and the adjusted determination coefficient (Adj. R^2_D).

Table 6. Leave-One-Out Cross-Validation (LOO-CV) Results.

Compound	An_{pred}	Aa_{pred}	U_{pred}	$d(An)_{pred}$	$d(Aa)_{pred}$	$d(U)_{pred}$	$D(An\text{-}Aa\text{-}U)_{pred}$
AS1	77	64	0.56	0.69	0.52	0.69	0.63
AS2	76	66	0.53	0.68	0.55	0.68	0.63
AS3	76	64	0.55	0.68	0.52	0.68	0.62
AS4	61	54	0.69	0.48	0.39	0.60	0.48
AS5	69	51	0.60	0.59	0.35	0.65	0.51
AS6	60	47	0.67	0.47	0.29	0.61	0.44
AS7	52	46	0.82	0.36	0.28	0.53	0.38
AS8	54	47	0.79	0.39	0.29	0.54	0.39
AS9	53	47	0.79	0.37	0.29	0.54	0.39
AS10	52	47	0.79	0.36	0.29	0.54	0.39
AS11	53	46	0.80	0.37	0.28	0.54	0.38
AS12	53	47	0.80	0.37	0.29	0.54	0.39
AS13	62	57	0.76	0.49	0.43	0.56	0.49
AS14	62	57	0.75	0.49	0.43	0.57	0.49
AS15	61	46	0.65	0.48	0.28	0.62	0.44
Overall desirability's LOO-CV determination coefficient				$Q^2_{D(An\text{-}Aa\text{-}U)} = 0.905$			

As can be seen, the overall desirability function exhibits good statistical quality as indicated by the R^2_D and $Adj.R^2_D$ values (~1). Moreover, the high Q^2_D value (0.905) provides an adequate level of reliability on the method in predicting D.

Finally, the optimization of the overall desirability was carried out to obtain the levels of the ACF descriptors that simultaneously produce the most desirable combination of all properties. Figure 4 shows the multiple response overall desirability, as well as the individual desirability functions determined by the respective pairs of predictor variables included on the three MLR models.

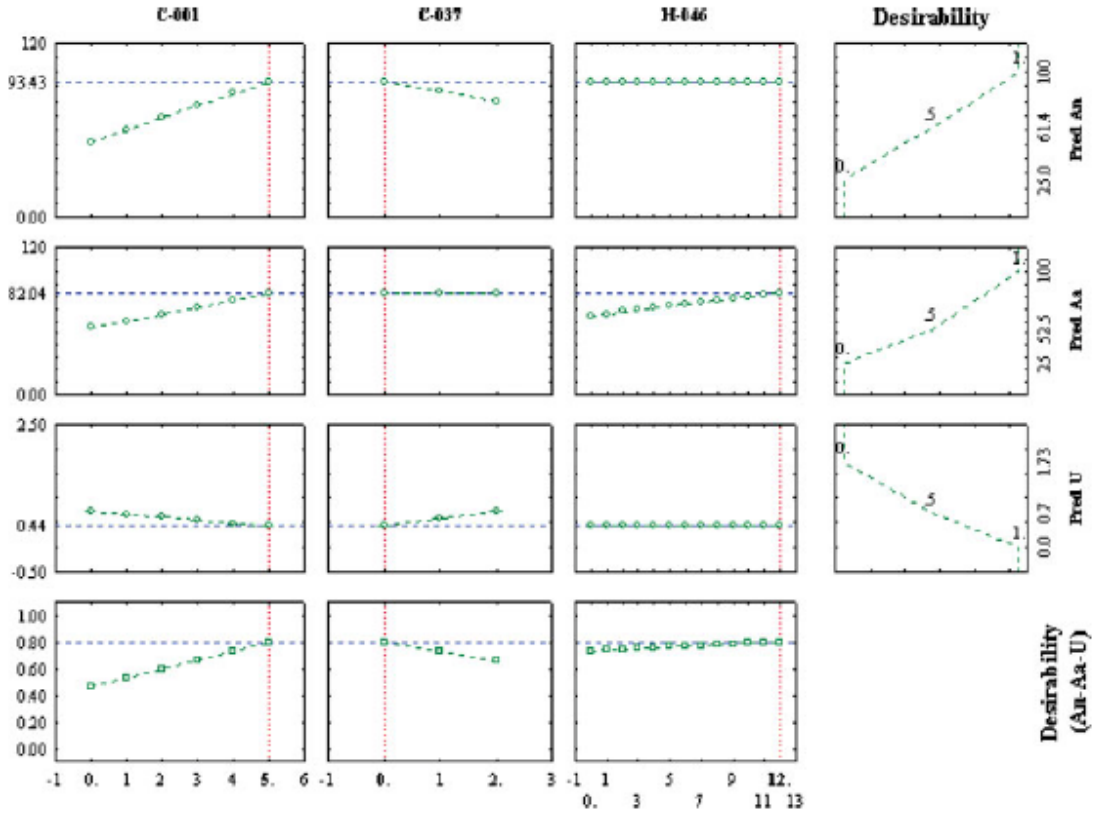


Figure 4. Multiple response desirability function due to the analgesic activity, anti-inflammatory activity and ulcerogenic index (*D*(An-Aa-U) (last row), along with the individual desirability functions coming from the pairs of predictor variables included on the three MLR models (first three rows).

By inspecting the form of each individual desirability function, it is possible to know the influence of a certain variable over each individual objective. In so doing, one can conclude that C-001 has a significant influence over the three properties, while H-046 has only a remarkable influence on the Aa activity. Here, one should note that the form of the An individual desirability function is similar to that obtained for the Aa activity (for these noncompeting objectives, both curves show a positive slope). However, opposite individual desirability function forms were obtained for competing objectives like Aa and U (i.e. the curve related to the ulcerogenic index has a negative slope).

Moreover, the data reveal that a 3-(3-methylphenyl)-2-substituted amino-3H-quinazolin-4-one optimized candidate must have analgesic and anti-inflammatory activities of 93.43% and 82.04%, respectively, plus an ulcerogenic index of 0.44. This represents an overall desirability of 0.8; that can be attained if the candidate has C-001, C-037 and H-046 values equal to 5, 0, and 12, respectively (see Fig. 4), being C-001 the most influencing variable. The significant slope of

the C-001 curve suggests that more attractive candidates could be designed if its values are greater than 5. However, due to the high influence of C-001 over the overall desirability, the optimal range for this variable should be close to 5. But one must also consider the applicability domain of the original PMs. In fact, the training set show C-001 values up to 3 and thus, if the new candidate has a C-001 value extremely far from 3, it might be out of the applicability domain of the original PMs. On the other hand, as the shape of the H-046 desirability function reveals no significant influence (slope near zero), the overall desirability could be increased by large departures from its optimum value (= 12). But again the applicability domain of the original PMs should be taken into account.

Design of new drug candidates. According to the previous results, the most important variable was found to be descriptor C-001 and the second one descriptor C-037. These two ACF descriptors represent, respectively, the number of methyl groups and heteroatoms attached to a sp2 carbon atom linked to the aromatic side ring in the drug candidates (see Figure 1). On the other hand, the less influencing ACF descriptor, H-046, represents the number of hydrogen atoms attached to a sp3 carbon no heteroatom attached to another carbon (see Figure 1).

This information allows one to guess the most important chemical modifications needed to improve the overall desirability of the present compounds. Considering the positive/negative influence of C-001/C-037 a different number *vs*. type of alkyl groups on the C-2 position of the quinazoline ring should be introduced. In fact, the introduction of branched alkyl substituents might lead to a positive role due to the bulkiness of the substituents.

So, a new set of nine compounds was designed in which several different alkyl substituents were linked to the C-2 position of the quinazoline ring. The chemical modifications and the predicted values of the expected pharmaceutical properties are shown in Table 7. The leverage values obtained for each new designed candidate were also considered to check whether or not each new candidate falls within the applicability domain of the original PMs (see Table 7).

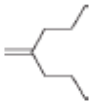
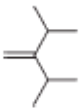
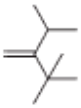
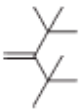
After a comprehensive data analysis, compound **ASNEW8** can be claimed to be the most desirable and reliable candidate designed in this study, displaying predicted percentages of analgesic and antiinflammatory activities of 93 and 82, respectively, plus a predicted ulcerogenic index of 0.44. Further, an excellent predicted overall desirability (0.8) is obtained. The data acquired allow us to propose also compounds **ASNEW4**, **ASNEW5**, **ASNEW6**, and **ASNEW9**, though having leverage values higher than h^*, i.e. out of the applicability domain of the original PMs. Interestingly, they possess the highest overall desirability and predictor variables values, significantly separated from those of the training compounds (see Table 8).

A noticeable profile improvement can be observed between the predicted properties displayed by compound **ASNEW8** and the most promising compound reported by Alagarsamy et al. (AS3).[35] Explicitly, **ASNEW8** displays analgesic and antiinflammatory activities 15 and 13% higher, respectively. At the same time, **ASNEW8** shows only the 78.6% of the ulcerogenic ability of AS3. On the other hand, if we compare the performance of **ASNEW8** with diclofenac (a known NSAIDs used as reference compound[35]), one can easily notice its enhanced predicted pharmaceutical properties. In effect, **ASNEW8** displays analgesic and antiinflammatory activities 31% and 22% higher than diclofenac, respectively. In addition, the ulcerogenic index is extensively reduced (**ASNEW8** has almost a quarter (3.75 times lower) of the ulcerogenic ability of diclofenac).

In summary, a remarkable simultaneous improvement on the analgesic and antiinflammatory activities plus ulcerogenic profile of the new designed candidates was obtained through

MOOPDESIRE-based methods combined with human expert interpretation and use of the results. The data suggest a positive role of the bulkiness of the alkyl substituents on the C-2 position of the quinazoline ring on the ulcerogenic properties. Anyhow, in the future, an experimental study of the analgesic, antiinflammatory and ulcerogenic properties of the designed candidates should be carried out to validate the process.

Table 7. Computed ACF descriptors (C-001, C-037, and H-046), predicted and leverage values for the analgesic (An) and antiinflammatory (Aa) activities, plus the ulcerogenic index (U) of the nine new designed compounds.

Compound	R	C-001	C-037	H-046	An_{pred} (%)	Aa_{pred} (%)	U_{pred} (%)	$h(An)$	$h(Aa)$	$h(U)$
ASNEW1		3	0	11	77	70	0.55	0.216	0.361	0.216
ASNEW2		3	0	13	77	72	0.55	0.216	0.496	0.216
ASNEW3		4	0	12	85	77	0.49	0.403	0.453	0.403
ASNEW4[a]		5	0	15	93	86	0.44	0.573	**0.614**	0.573
ASNEW5[a]		6	0	18	102	96	0.38	**0.695**	**0.724**	**0.695**
ASNEW6[a]		7	0	21	110	106	0.33	**0.777**	**0.796**	**0.777**
ASNEW7		4	0	9	85	72	0.49	0.403	0.401	0.403
ASNEW8		5	0	12	93	82	0.44	0.573	0.562	0.573
ASNEW9[a]		5	0	15	93	86	0.44	0.573	**0.614**	0.573

[a]Compounds out of the predictions model's applicability domain; leverage values greater than h^* are marked in bold.

Table 8. Predicted values for the desirability due to the analgesic activity [d(An)], antiinflammatory activity [d(Aa)], ulcerogenic index [d(U)], and overall desirability [D(An-Aa-U)] of the nine new designed compounds.

Compound	$d(An)_{pred}$	$d(Aa)_{pred}$	$d(U)_{pred}$	$D(An\text{-}Aa\text{-}U)_{pred}$
ASNEW1	0.69	0.60	0.73	0.67
ASNEW2	0.69	0.63	0.73	0.68
ASNEW3	0.80	0.69	0.65	0.71
ASNEW4[a]	0.91	0.81	0.75	0.82
ASNEW5[a]	1.00	0.95	0.78	0.90
ASNEW6[a]	1.00	1.00	0.81	0.93
ASNEW7	0.80	0.63	0.72	0.71
ASNEW8	0.91	0.76	0.75	0.80
ASNEW9[a]	0.91	0.81	0.75	0.82

[a]Compounds out of the predictions model's applicability domain.

Despite the limited size and homogeneity of our data set, this work offers the possibility of a deeper and case by case analysis of the results obtained by using the MOOP-DESIRE methodology. The use of small and homogeneous data set is more suitable for later stages of the drug development process once identified a lead rather than for early stages. Actually, the results of the optimization process can be used to perform specific structural modifications over the lead. For this, the use of clearly defined structural or physicochemical descriptors can lead to interpretable structure–desirability relationships which can be used to design new candidates with an improved pharmaceutical profile. The MOOP-DESIRE methodology can also be applied to handle larger and/or more diverse data sets, such as those frequently obtained in High-Throughput Screening processes, being there more appropriate for early stages of the drug development process. That is, molecules coming from large and heterogeneous data sets can be filtered and ranked according to a certain criterion rather than applying the results of the optimization process to design new candidates. To accomplish that, one can resort to the overall desirability of each molecule as a ranking criterion or to several distance measures between the optimal values of the descriptors determined by MOOP-DESIRE and the computed values of the descriptors. In this case, it is advisable to use descriptors leading to highly predictive structure–desirability relationships rather than interpretable descriptors in order to ensure the accuracy of the predictions and therefore, an accurate assessment of the molecule's overall desirability which will then be the ranking criterion.

Conclusions

In this work, a novel MOOP method sustained on the desirability estimation of several interrelated responses is proposed. The MOOP-DESIRE methodology based on Derringer's desirability function enables one to perform global QSAR studies, considering simultaneously the pharmacological, pharmacokinetic and toxicological profiles of a set of molecule candidates. The usefulness of the methodology was demonstrated by applying it to the simultaneous optimization of the analgesic, anti-inflammatory and ulcerogenic properties of a library of fifteen 3-(3-methylphenyl)-2-substituted amino-3H-quinazolin-4-one compounds. The best compromise between the mentioned properties was established and new drug candidates with the highest overall desirability then designed. In particular, one of the designed candidates (compound ASNEW8) is predicted to have 93% of analgesic activity, 82% of inflammatory inhibition and an

ulcerogenic index of 0.44, which represents an excellent overall desirability (50.8), being this accomplished by modifying the compounds' structure in such a way that pushed the values of the C-001, C-037, and H-046 predictor variables to 5, 0, and 12, respectively. Furthermore, it was observed that the presence of bulky alkyl substituents at the C-2 position of the quinazoline ring displayed a positive role on the ulcerogenic ability without a negative influence in the other properties. Yet, further experimental corroboration is still needed to validate the model. In conclusion, the desirability-based MOOP method herein proposed is regarded as a valuable tool and shall aid in the future rational design of novel successful drugs.

References

1. Ekins, S.; Boulanger, B.; Swaan, P. W.; Hupcey, M. A. J Comput Aided Mol Des 2002, 16, 381.
2. Jorgensen, W. L. Science 2004, 303, 1813.
3. Seifert, M. H. J.; Wolf, K.; Vitt, D. Drug Discov Today: Biosilico 2003, 1, 143.
4. Brown, N.; Lewis, R. A. Curr Opin Drug Discov Devel 2006, 9, 419.
5. Hansch, C. J Med Chem 1976, 19, 1.
6. Fukunaga, J. Y.; Hansch, C.; Steller, E. E. J Med Chem 1976, 19, 605.
7. Mayer, J. M.; van de Waterbeemd, H. Environ Health Perspect 1985, 61, 295.
8. Moriguchi, I.; Hirano, H.; Hirono, S. Environ Health Perspect 1996, 104, 1051.
9. Estrada, E. SAR QSAR Environ Res 2000, 11, 55.
10. Vilar, S.; Estrada, E.; Uriarte, E.; Santana, L.; Gutierrez, Y. J Chem Inf Model 2005, 45, 502.
11. Marrero-Ponce, Y.; Marrero, R. M.; Torrens, F.; Martinez, Y.; Bernal, M. G.; Zaldivar, V. R.; Castro, E. A.; Abalo, R. G. J Mol Model (Online) 2006, 12, 255.
12. Helguera, A. M.; Cabrera Perez, M. A.; Gonzalez, M. P. J Mol Model (Online) 2006, 12, 769.
13. Gonzalez-Diaz, H.; Cruz-Monteagudo, M.; Molina, R.; Tenorio, E.; Uriarte, E. Bioorg Med Chem 2005, 13, 1119.
14. Gonzalez-Diaz, H.; Cruz-Monteagudo, M.; Vina, D.; Santana, L.; Uriarte, E.; De Clercq, E. Bioorg Med Chem Lett 2005, 15, 1651.
15. Cruz-Monteagudo, M.; Gonzalez-Diaz, H.; Borges, F.; Gonzalez-Diaz, Y. Bull Math Biol 2006, 68, 1555.
16. Cruz-Monteagudo, M.; Cordeiro, M. N.; Borges, F. J Comput Chem 2008, 29, 533.
17. Cruz-Monteagudo, M.; Borges, F.; Perez Gonzalez, M.; Cordeiro, M. N. Bioorg Med Chem 2007, 15, 5322.
18. Cruz-Monteagudo, M.; Gonzalez-Diaz, H.; Aguero-Chapin, G.; Santana, L.; Borges, F.; Dominguez, E. R.; Podda, G.; Uriarte, E. J Comput Chem, 2007, 28, 1909.
19. Gonzalez-Diaz, H.; Aguero, G.; Cabrera, M. A.; Molina, R.; Santana, L.; Uriarte, E.; Delogu, G.; Castanedo, N. Bioorg Med Chem Lett 2005, 15, 551.
20. Prado-Prado, F. J.; Gonzalez-Diaz, H.; Santana, L.; Uriarte, E. Bioorg Med Chem 2007, 15, 897.
21. Gonzalez-Diaz, H.; Prado-Prado, F. J.; Santana, L.; Uriarte, E. Bioorg Med Chem 2006, 14, 5973.
22. Nicolaou, A. C.; Brown, N.; Pattichis, C. S. Curr Opin Drug Discov Devel 2007, 10, 316.
23. Yann, C.; Siarry, P., Eds. Multiobjective Optimization: Principles and Case Studies; Springer-Verlag: Berlin, Germany, 2004.
24. Jones, G.; Willett, P.; Glen, R. C. J Comput Aided Mol Des 1995, 9, 532.

25. Handschuh, S.; Wagener, M.; Gasteier, J. J Chem Inf Comput Sci 1998, 38, 220.
26. Shepphird, J. K.; Clarck, R. D. J Comput Aided Mol Des 2006, 20, 735.
27. Janson, S.; Merkle, D. In Hybrid Metaheuristics Second International Workshop, HM 2005; Blesa, M. J.; Blum, C.; Roli, A.; Sampels, M., Eds.; Springer-Verlag: Barcelona, Spain, 2005, p. 128.
28. Zoete, V.; Grosdidier, A.; Michielin, O. High Performance Computing for the Life Sciences Symposium, Lausanne, Switzerland, 2005.
29. Brown, N.; Mckay, B.; Gasteiger, J. J Comput Aided Mol Des 2006, 20, 333.
30. Lameijer, E. W.; Kok, J. N.; Back, T.; Ijerman, A. P. J Chem Inf Model 2006, 46, 545.
31. Nicolotti, O.; Gillet, V. J.; Fleming, P. J.; Green, D. V. J Med Chem 2002, 45, 5069.
32. Stockfisch, T. P. J Chem Inf Comput Sci 2003, 43, 1608.
33. Rao, S. N.; Stockfisch, T. P. J Chem Inf Comput Sci 2003, 43, 1614.
34. Derringer, G.; Suich, R. J Quality Technol 1980, 12, 214.
35. Alagarsamy, V.; Dhanabal, K.; Parthiban, P.; Anjana, G.; Deepa, G.; Murugesan, B.; Rajkumar, S.; Beevi, A. J. J Pharm Pharmacol 2007, 59, 669.
36. Arulmozhi, D. K.; Veeranjaneyulu, A.; Bodhankar, S. L.; Arora, S. K. J Pharm Pharmacol 2004, 56, 655.
37. Ganguly, A. K.; Bhatnagar, O. P. Can J Physiol Pharmacol 1973, 51, 748.
38. Goyal, R. K.; Chakrabarti, A.; Sanyal, A. K. Planta Med 1985, 29, 85.
39. ChemDrawn Ultra 9.0. CambridgeSoft. 2004.
40. Burkert, U.; Allinger, N. L. Molecular Mechanics; ACS: Washington, D.C., 1982.
41. Clark, T. Computational Chemistry; Wiley: N.Y., 1985.
42. Frank, J. MOPAC 2.0; Seiler Research Laboratory, US Air Force Academy, Colorado Springs: CO, 1993.
43. Todeschini, R.; Consonni, V.; Pavan, M. DRAGON 2.1; Milano Chemometrics: Milano, 2002.
44. Viswanadhan, V. N.; Ghose, A. K.; Revankar, G. R.; Robins, R. K. J Chem Inf Comput Sci 1989, 29, 163.
45. Caballero, J.; Fernandez, L.; Abreu, J. I.; Fernandez, M. J Chem Inf Model 2006, 46, 1255.
46. Fernandez, M.; Caballero, J. J Mol Graph Model 2006, 25, 410.
47. Fernandez, L.; Caballero, J.; Abreu, J. I.; Fernandez, M. Proteins 2007, 67, 834.
48. Caballero, J.; Fernandez, L.; Garriga, M.; Abreu, J. I.; Collina, S.; Fernandez, M. J Mol Graph Model 2007, 26, 166.
49. Leardi, R.; Boggia, R.; Terrile, M. J Chemom 1992, 6, 267.
50. Yasri, A.; Hartsough, D. J Chem Inf Comput Sci 2001, 41, 1218.
51. Hou, T. J.; Wang, J. M.; Liao, N.; Xu, X. J. J Chem Inf Comput Sci 1999, 39, 775.
52. Hasegawa, K.; Kimura, T.; Funatsu, K. J Chem Inf Comput Sci 1999, 39, 112.
53. Barbosa de Oliveira, D.; Gaudio, A. C.; BuildQSAR; University of Espı´rito Santo: Vito´ria ES, Brasil, 2000.
54. Barbosa de Oliveira, D.; Gaudio, A. C. Quant Struct-Act Relat 2000, 19, 599.
55. STATISTICA 6.0. Statsoft_Inc., 2001.
56. Eriksson, L.; Jaworska, J.; Worth, A. P.; Cronin, M. T.; McDowell, R. M.; Gramatica, P. Environ Health Perspect 2003, 111, 1361.
57. Stewart, J.; Gill, L. Econometrics; Prentice Hall: London, 1998.
58. Kutner, M. H.; Nachtsheim, C. J.; Neter, J.; Li, W., Eds. In Applied Linear Statistical Models; McGraw Hill: New York, 2005, p 278.

59. Atkinson, A. C. Plots, Transformations and Regression; Clarendon Press: Oxford 1985.
60. Kutner, M. H.; Nachtsheim, C. J.; Neter, J.; Li, W. Applied Linear Statistical Models; McGraw Hill: New York, 2005.
61. Nelder, J. A.; Mead, R. Comput J 1965, 7, 308.
62. Fletcher, R.; Reeves, C. M. Comput J 1964, 7, 149.
63. Hooke, R.; Jeeves, T. A. J Assoc Comp Machin 1961, 8, 212.
64. De Boor, C. A Practical Guide to Splines; Springer-Verlag: New York, 1978.
65. Gerald, C. F.; Wheatley, P. O. Applied Numerical Analysis; Addison Wesley Reading: MA, 1989.