J Comput Aided Mol Des (2013) 27:79–90
DOI 10.1007/s10822-012-9627-1

# Combined use of pharmacophoric models together with drug metabolism and genotoxicity "in silico" studies in the hit finding process

**Ma José Jerez · Miguel Jerez · Coral González-García · Sara Ballester · Ana Castro**

Received: 8 November 2012 / Accepted: 15 December 2012 / Published online: 8 January 2013


**Abstract** In this study we propose a virtual screening strategy based on the generation of a pharmacophore hypothesis, followed by an in silico evaluation of some ADME-TOX properties with the aim to apply it to the hit finding process and, specifically, to characterize new chemical entities with potential to control inflammatory processes mediated by T lymphocytes such as multiple sclerosis, systemic lupus erithematosus or rheumatoid arthritis. As a result, three compounds with completely novel scaffolds were selected as final hits for future hit-to-lead optimization due to their anti-inflammatory profile. The biological results showed that the selected compounds increased the intracellular cAMP levels and inhibited cell proliferation in T lymphocytes. Moreover, two of these compounds were able to increase the production of IL-4, an immunoregulatory cytokine involved in the selective deviation of T helper (Th) immune response Th type 2 (Th2), which has been proved to have anti-inflammatory properties in several animal models for autoimmune pathologies as multiple sclerosis or rheumatoid arthritis. Thus our pharmacological strategy has shown to be useful to find molecules with biological activity to control immune responses involved in many inflammatory disorders. Such promising data suggested that this in silico strategy might be useful as hit finding process for future drug development.

**Keywords** Virtual screening · Pharmacophore · ADME in silico · Toxicity · PDE7 · T lymphocytes

## Introduction

Virtual screening still continues to be a complementary method to high-throughput screening of large chemical databases, and it has been widely used in the drug discovery process, being one of the methods applied as a source of lead compounds [1].

On the other hand, toxicity is a major impediment to reach a successful drug discovery and development, and in silico methods form a valuable part of the battery of techniques to act as alternative to traditional toxicity testing [2].

Moreover, determining ADME properties by using in silico methods reduces the risk during the selection process of drug candidates in the early drug discovery process [3]. In this specific context, cytochromes P450 (CYP) are crucial targets not only because metabolic transformations are frequently related to the incidence of toxic effects that may result from the emergence of reactive species or by induction of metabolic pathways, but also because CYP inhibition is pivotal for drug-drug interaction, an important issue in the discovery and development process of new drugs.

In the present work, we propose a combined strategy of virtual screening to identify new potential hit compounds, merging the use of developing pharmacophore modeling with an in silico evaluation of some ADME-TOX

**Electronic supplementary material** The online version of this article (doi:10.1007/s10822-012-9627-1) contains supplementary material, which is available to authorized users.

M. J. Jerez · A. Castro (✉)
Instituto de Química Médica-CSIC, Juan de la Cierva 3, 28006 Madrid, Spain
e-mail: acastro@iqm.csic.es

M. Jerez
Facultad de Ciencias Económicas y Empresariales, Universidad Complutense, Campus de Somosaguas, 28223 Madrid, Spain

C. González-García · S. Ballester
Unidad de Regulación Génica, CNM, Instituto de Salud Carlos III, Madrid, Spain

properties, using a representative set of phosphodiesterase 7 (PDE7) inhibitors as a case study.

PDE7 is one of at least eleven isoenzymes of mammalian cyclic nucleotide phosphodiesterase that have been identified [4]. Specifically, their relevance is based on the fact that the inhibition of the activity or the expression of PDE7 was found to block the activation of T-cells as well as the proliferation and function of preactivated T-cells and cytotoxic T-lymphocytes [5, 6]. Although its role in these cells is not very well defined yet and investigations on PDE7A knockout mice show that it is not essential for the lymphocyte activation [7], data in which they based their conclusions are limited to the PDE7A isoform, so the PDE7B isoform could make up for PDE7 functions in PDE7A knockout mice. In addition, there are studies that demonstrate that PDE7 is able to regulate T lymphocytes functions, including cytokine production, proliferation and expression of activation markers [8].

Therefore, currently there is an increasing interest in the development of new selective inhibitors of this enzyme that can have a great therapeutic importance for the treatment of T lymphocytes-mediated autoimmune and inflammatory pathologies [9, 10].

At first place, Catalyst was used to develop a 3D-QSAR model to carry out a database screening. The pharmacophore thus developed imparts information about important features for PDE7 inhibitory activity and geometry, and it was used to mine 3D-virtual databases of drug-like molecules. Once the contours generated from QSAR studies highlighted the structural features required for PDE7 inhibition, the reliability of the postulated pharmacophore model was validated by the biological evaluation of new PDE7 inhibitors.

Subsequent, based on the results of the Ames test available in the GENE-TOX database, we developed an early prediction of the mutagenicity of the compounds by means of a non-linear statistical model.

After that, the programs Volsurf and Metasite were employed for the evaluation of some relevant ADME properties, including BBB penetration, thermodynamic stability in solution, Caco2 permeability, human serum albumin binding and hERG potassium channels blockers (through Volsurf) and the determination of cytochrome responsible for their metabolism (using Metasite).

Finally, the biological profile of the hits compounds was studied by means of three parameters related to the control of inflammation: effect on intracellular cAMP levels, influence on T cell proliferation, and consequences on the production of the Th2 immunoregulatory cytokine IL-4.

## Methods

### Training and evaluation set for the pharmacophore model

31 molecules [**1–15** training set (Fig. 1) and **16–31** evaluation set (Fig. 2)] were selected as the training and

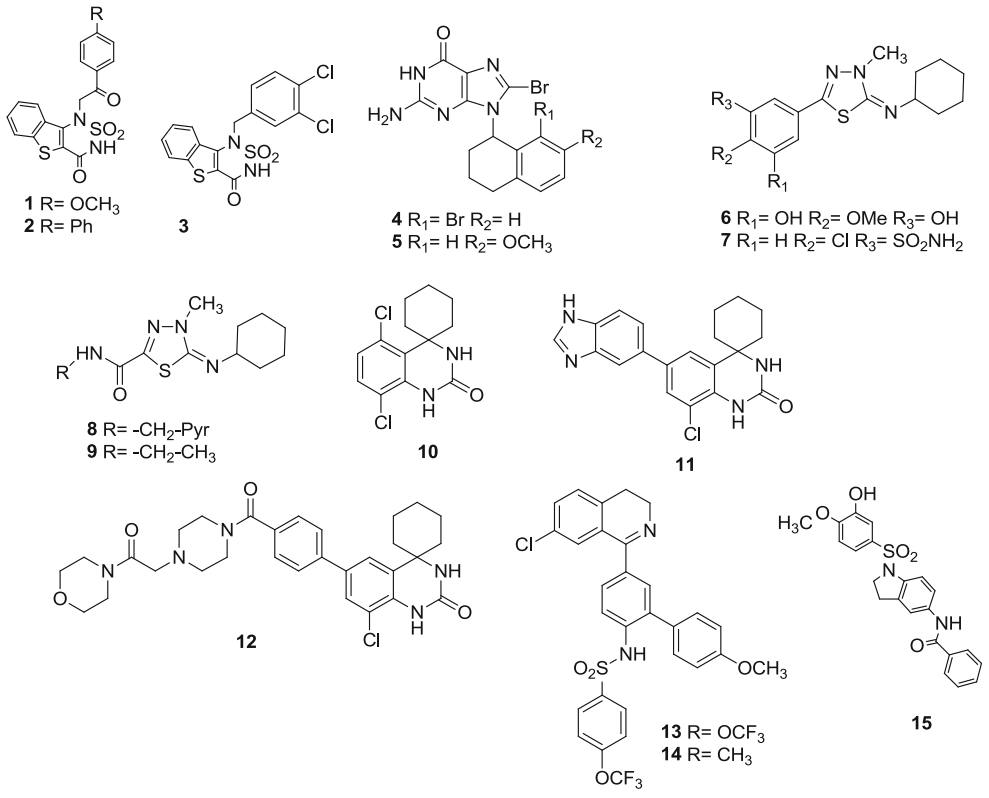


**Fig. 1** Compounds of the training set

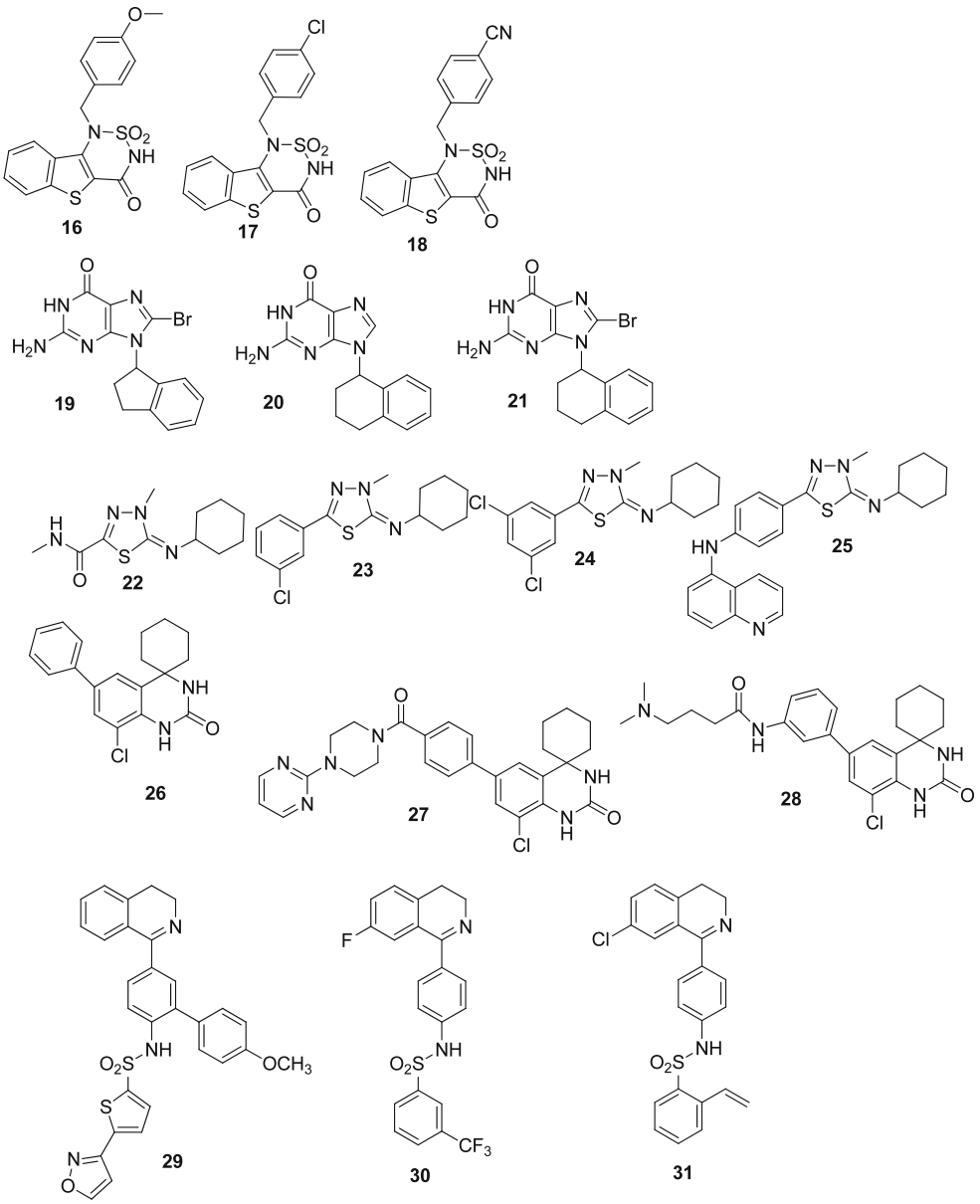


**Fig. 2** Compounds of the evaluation set

evaluation set, respectively, representing some of the most interesting compounds of known PDE7 inhibitors. The selected families belong to six structurally diverse chemical classes including: benzothiadiazine [11, 12], guanine [13], iminothiadiazol [14–16], spirotricycles [17–19], 3,4-dihidroisoquinolines [20, 21] and benzene-sulphonamides [22], assuming in all the cases a similar binding mode with the enzyme. These compounds are characterized by significant in vitro biological tests and show activity range within 5 orders of magnitude (Table 1).

All structures, considered in all the cases as neutral forms, were generated using 2D/3D editor sketcher in Catalyst software package [23] and their energy were minimized to the closest local minimum using the CHARMM-like force field implemented in the program.

To build conformational models of up to 256 conformers for each molecule, the "best conformer generation" option and a 20 kcal/mol energy cutoff were chosen.

All molecules with their associated conformations were regrouped together with their biological activities data (Table 1).

## Pharmacophore generation

Based on the chemical features of 15 compounds in the training set, 10 pharmacophore models were generated using the pharmacophoric analysis of the Catalyst/HypoGen program with a default uncertainty value of 3.0. This uncertainty value represents a ratio range of uncertainty in

**Table 1** $IC_{50}$ values of the training set compounds

| Chemical family | Code | IC50 (μM) | Number of conformers |
|---|---|---|---|
| Benzothiadiazines | **1** | 11 | 108 |
| | **2** | 25 | 35 |
| | **3** | 8 | 35 |
| Guanines | **4** | 1.31 | 7 |
| | **5** | 4.22 | 31 |
| Iminothiadiazoles | **6** | 0.061 | 27 |
| | **7** | 0.044 | 37 |
| | **8** | 0.062 | 46 |
| | **9** | 0.07 | 57 |
| Spirocycles | **10** | 0.014 | 4 |
| | **11** | 0.012 | 20 |
| | **12** | 0.008 | 100 |
| Isoquinolines | **13** | 0.032 | 200 |
| | **14** | 0.12 | 122 |
| Sulphonamides | **15** | 0.036 | 75 |

the activity value based on the expected statistical irregularities of biological data collection.

In the hypothesis generation and on the basis of the atom types in the molecules of the training set, the following chemical features were selected: H-bond acceptor (HBA), H-bond donor (HBD), hydrophobic (HY), aromatic (RA) and positive ionizable (PI), including all pharmacophoric features of molecules in both training and evaluation sets.

## Pharmacophore model quality examination

In order to evaluate the reliability and accuracy of the generated 3D pharmacophore models, prediction capability, cost analysis, and Cat-Scramble validation studies were performed.

### *Prediction capability of the generated pharmacophore model*

The pharmacophore model was also used to predict the PDE7 inhibitory activities of other compounds beyond of those in the training set. For such a purpose, 16 PDE7 inhibitors in the evaluation set (Fig. 2) were estimated by the best pharmacophore model with the highest correlation coefficient (r) and the lowest total cost.

### *Cost analysis*

The quality of HypoGen-generated pharmacophore models can be described in terms of fixed cost, null cost and total cost, all of which are well-defined by Debnath. The fixed cost represents the simplest model that fits the data. The null cost represents the cost of a hypothesis with no features which estimates every activity to be the average activity. If a returned cost (total cost) differs from the null cost by 40–60 bits, it is highly probable that the hypothesis has 75–90 % chance of representing the true correlation of the data. If the difference becomes more than 60 bits, the probability of representing a true correlation of the data is even higher, more than 90 %.

### *Cat-Scramble validation*

The Cat-Scramble validation procedure is based on the Fischer's randomization test. This is to check whether there is a strong correlation between chemical structures and bioactivities of ligands. It is done by randomizing the activity data associated with the training set of compounds and generating pharmacophore hypotheses using the common chemical features and parameters to develop the original pharmacophore hypothesis. 19 random spreadsheets were generated for a 95 % confidence level. If the randomized data set has similar or better cost values, RMSD and correlation in the generation of a pharmacophore model, the original hypothesis is considered to have been generated by chance.

Finally database searching was performed through a 3D query using the highest scoring pharmacophore model accomplished by means of Catalyst.

## Codification of structures for toxicity in silico studies

Initial molecules were selected from GENE-TOX database (Genetic Toxicology) [24], which contains peer-reviewed mutagenicity test data from the Environmental Protection Agency (EPA) for over 3,000 compounds, using the following three restrictive criteria: (a) positive or negative Ames test done with the same experimental method on strain TA98 of *Salmonella*; (b) organic molecules; and (c) therapeutic or industrial use described in the Merck Index.

In this way, 132 molecules (see Figure S1 in Supplementary Information) with 8–111 atoms and 64–808 Da of molecular weight were selected (Table S1 in Supplementary Information). From these molecules, 57 were toxic and 75 non toxic.

Codification of the molecules was carried out using CODES program. This program, which has never been used for virtual screening purposes, has been previously used for QSAR studies, as for example the prediction of the nematocide action of pteridine derivatives [25], of new drugs for potassium channels opening [26], and the oral absorption and BBB penetration of several drugs [27]. In such way, it has also been used to carry out QSPR studies, such as the determination of pharmacokinetic properties as for example the mean life of antihistaminic drugs, classifying them in their corresponding therapeutic categories [28].

**Table 2** Series of data using PCA criteria

| Series | Descriptors | Explanatory variables | Number of compounds |
|---|---|---|---|
| M5 | >5 | 1 | 128 |
| M8 | >8 | 2 | 119 |
| M11 | >11 | 3 | 111 |
| M15 | >14 | 4 | 91 |

Since the output of CODES is a matrix where each column represents one of the atoms of the molecule different than hydrogen, it is necessary to reduce those matrixes to the same dimensionality. This reduction was carried out by PCA using Tsar program [29]. Therefore five series of data were created to group the molecules using 1, 2 and 3 principal components (PC) (able to explain 100 % of the variance in 99 % of the cases) as follows: with 5 or more descriptors (1 explanatory variable, 128 compounds, M5), with 8 or more (2 variables, 119 compounds, M8), with 11 or more (3 variables, 111 compounds, M11), with 14 or more (4 variables, 101 compounds, M14) and with more than 14 (4 variables, 91 compounds, M15) as described in Table 2.

## Mathematical description of the models

We used these series to estimate the following non linear statistical model.

Let yi be a variable which may take only two values: yi = 1, meaning that the i-th molecule is toxic, and yi = 0, meaning that it is non-toxic. Assume also that there is a k × 1 vector of real continuous variables, xi, that summarize the physical properties of the i-th molecule. Under these conditions, the conditional probability that the i-th molecule is non-toxic and toxic can be written, respectively, as:

$$\mathrm{P}(\mathrm{yi} = 0|\mathrm{xi}.\beta) = \mathrm{F}(-\mathrm{xiT}\beta)$$

$$\mathrm{P}(\mathrm{yi} = 0|\mathrm{xi}.\beta) = 1 - F(-\mathrm{xiT}\beta)$$

where:

P () is the probability of the event defined within the parentheses.

F () is a cumulative distribution function (CDF), i.e., any continuous, strictly increasing function that yields real values from zero to one.

β is a k × 1 vector of real parameters defining a linear combination of xi that can be interpreted as a toxicity indicator.

In general the values in β are unknown, so one must compute an estimate, þ, that optimizes the model forecasting power. Given a sample of n values of yi and xi, a common criterion consists of maximizing the likelihood function:

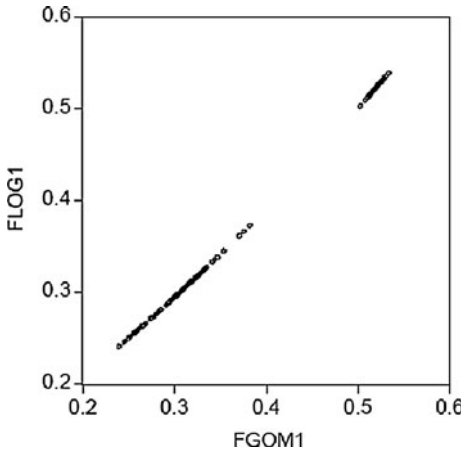


**Fig. 3** Prediction of the probability functions Logit and Gompit, with r = 0.9998

$$\ell(\beta|y_i, x_i) = \sum_{i=1}^{n} y_i \,\mathrm{In}\left[1 - F(-x_i^T\beta)\right] + (1 - y_i)\mathrm{In}\left[F(-x_i^T\beta)\right]$$

In this work, we considered two alternative specifications for F (): the CDF of a logistic distribution:

$$F(-x_i^T\beta) = \frac{\exp(x_i^T\beta)}{1 + \exp(x_i^T\beta)}$$

characteristic of the Logit model, and the Type-I extreme value distribution:

$$F(x_i^T\beta) = 1 - \exp\left[-\exp(x_i^T\beta)\right]$$

that defines a model known as Gompit.

These statistical models defined above provide estimates of the probability that a molecule is toxic, conditional to its physical properties. To transform these probabilities into binary forecasts one must define a "cut-off value", i.e., a probability threshold c such that, if $P(y_\mathrm{i} = 1|x_i,\beta) > c$, then the forecast for the i-th molecule is "toxic" and, if $P(y_\mathrm{i} = 1|x_i,\beta) \leq c$, then the forecast is "non-toxic".

This cut-off was established as the percentage of toxic molecules of the series: 39/91 = 0.4. Under these conditions, both models guessed correctly 60 % of the internal M15 series data. Gompit and Logit models performance measures are identical (Fig. 3); in fact, a one-to-one comparison of the binary forecasts provided by both specifications reveal that they are the same. Therefore, this exercise does not provide compelling evidence to favor any of these specifications.

## Calculation of Volsurf descriptors

For each compound, a 3D structure was generated using the program VolSurf and Molecular Interaction Fields (MIFs) were calculated by the program GRID. Grid spacing was set to 0.5 Å and the following probes were used: $H_2O$ (the

water molecule), O (sp2 carbonyl oxygen atom), N1 (neutral NH group, e.g., amide) and DRY (the hydrophobic probe). From the MIFs, VolSurf + derived series of 94 descriptors (independent of molecular alignment) that referred to some relevant ADME properties [30].

### Metasite analysis

MetaSite software was employed in this computational analysis. Two-dimensional structures of the compounds (generated in a simplified molecular input line entry system notation) were submitted to the MetaSite program, which is a fully automated procedure. The MetaSite version used in this analysis contains the homology three-dimensional models of P450 1A2, P450 2C9, P450 2C19, P450 2D6, and P4503A4. MetaSite analyses to predict sites of metabolism involve the calculation of two sets of descriptors: one set for P450 enzyme(s) and one set for substrate. The resulting analysis provides a "chemical fingerprint" of the enzyme and the substrate [31]. Chemical fingerprints are distance based descriptors that are calculated from molecular interaction forces computed by GRID [32]. MetaSite considers the chemical reactivity of the compound by taking into consideration the activation energy in the hydrogen abstraction step that ultimately generates metabolite(s).

### Cellular assays

The cellular anti-inflammatory properties of the selected compounds were evaluated in D10.G4.1 cells, an extensively characterized T cell line specific for conalbumin [33]. Cells were maintained in Click's medium [34] supplemented with 10 % heat-inactivated FCS and stimulated every 2 weeks at $5 \times 10^4$ cells/ml with mitomycin C(MMC)-treated feeder cells and antigen as previously described [35]. The biological effects mediated by the compounds selected were analyzed by three different assays indicative of anti-inflammatory features: power to increase intracellular cAMP levels, ability to inhibit T cell proliferation and effect on the production of the cytokine IL-4.

(a) *cAMP measurements*. D10.G4.1 cells were stimulated for 1 h with ionomycin and PMA in 96-well plates ($5x10^4$ cells per well), in the presence or absence of the compounds assayed, vehicle (DMSO) as negative control, or the PDE4 inhibitor Rolipram as a positive control. Intracellular cAMP was determined by a competitive enzyme immunoassay (EIA) system (Amersham, GE Healthcare).

(b) *Proliferation assays*. Proliferation was measured after incubating $10^4$ D10.G4.1 cells, at 37 °C and 5 % $CO_2$ for 72 h on 96-well plates, coated with 10 μg/ml of anti-CD3 (YCD3-1) [36]. Each sample was assayed in triplicate. Cell growth was measured by colorimetric assay as formerly described [37].

(c) *Cytokine quantification*. IL-4 determinations were performed by ELISA as previously described [38]. Briefly, anti-IL-4 11B11 (ATCC HB188) and biotinylated anti-IL-4 BVD6-24G2 were used respectively as capture and detecting antibodies. Absorbance was measured at 492 nm. Reference standard curves were set up in each assay with serial dilutions of recombinant murine IL-4 (Genzyme diagnostics). Each sample was assayed in triplicate.

## Results and discussion: Pharmacophore model and database searching

### Pharmacophore generation using HypoGen module

To assure the flexibility of the study, a conformational analysis was carried out for each inhibitor from the optimal geometry of minimum energy previously calculated. Although this step does not ensure the finding of the biologically active conformation, it increases the probability of this conformation to be in the set of conformers.

After that, HypoGen-generated pharmacophore models displayed a set of 10 pharmacophore hypotheses generated from the training set of 15 PDE7 inhibitors using Catalyst/HypoGen. Among these 10 hypotheses, Hypo_test is characterized by the highest cost difference of 84.10, and the highest correlation coefficient of 0.960. Noticeably, the cost range between Hypo_test and the fixed cost is 52.6, while that between the null hypothesis and Hypo_test is 84.10, showing that Hypo_test has probability higher than 90 % of correlating the data.

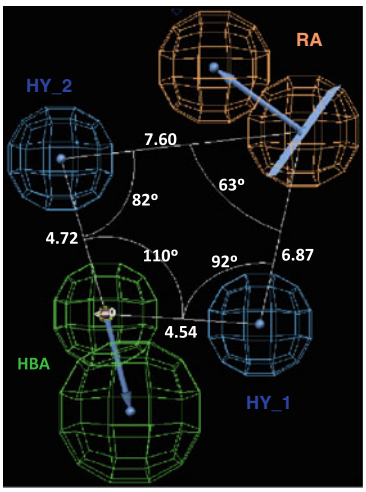


**Fig. 4** Topological features of Hypo_test hypothesis showing distances (Å) and angles

**Table 3** Observed versus predicted activities ($IC_{50}$) of the training set of molecules

| Molecule | Observed $IC_{50}$ (μM) | Predicted $IC_{50}$ (μM) |
|---|---|---|
| **1** | 11 | 8.2 |
| **2** | 25 | 12 |
| **3** | 8 | 8.2 |
| **4** | 1.31 | 1.3 |
| **5** | 4.22 | 0.38 |
| **6** | 0.061 | 0.094 |
| **7** | 0.044 | 0.11 |
| **8** | 0.062 | 0.13 |
| **9** | 0.07 | 0.085 |
| **10** | 0.014 | 0.013 |
| **11** | 0.012 | 0.014 |
| **12** | 0.008 | 0.0078 |
| **13** | 0.032 | 0.023 |
| **14** | 0.12 | 0.12 |
| **15** | 0.036 | 0.056 |

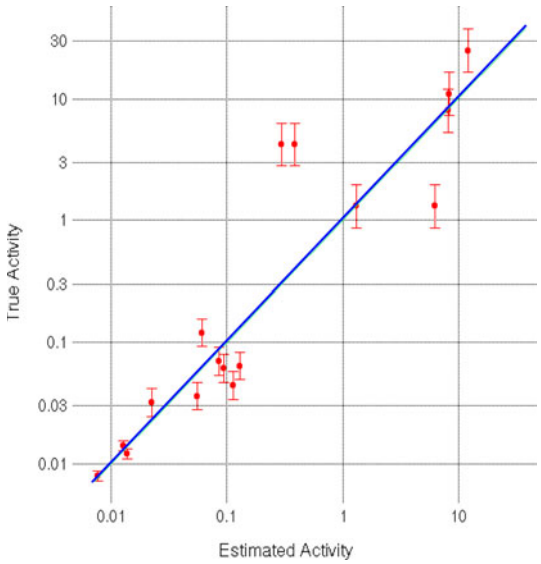


**Fig. 5** Correlation between observed and recalculated activities of the training set

Figure 4 shows the topological features of Hypo_test, which consists of four pharmacophore features, including one H-bond acceptor (HBA, colored by green), one aromatic ring (RA, colored by pink), and two hydrophobic groups (HY, colored by blue).

The recalculated results of PDE7 inhibitory activities (IC50 values) of the 15 training set molecules are listed in Table 3 and the plot of recalculated versus experimental values is recorded in Fig. 5. Although in some cases (see compound **5**) the correlation between experimental and newly predicted values is not optimal, the correlation coefficient of r = 0.96, indicates a good predictive power of this hypothesis.

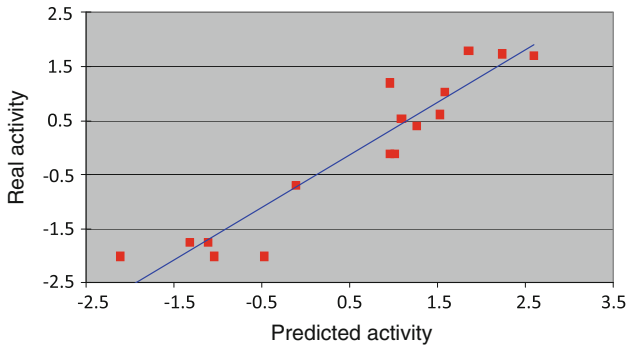


**Fig. 6** Correlation between observed and recalculated activities of the evaluation set

## Prediction capability of the generated pharmacophore model

Evaluation set prediction. To test whether Hypo_test can also predict the activities of compounds beyond the training set, we applied a total of 16 compounds representing diverse activity classes and not used in the training set. As shown in Fig. 6, there is a good line correlation between the experimental and predicted PDE7 activities of the test set compounds, showing a good correlation coefficient of 0.95. Thus, Hypo_test has a good prediction capability for both training set and test set compounds.

## Cat-Scramble validation

With the aid of the Cat-Scramble program, the experimental activities of compounds in the training set were scrambled randomly, and the resulting data set was used for HypoGen generation. All parameters were adopted and used in the initial HypoGen calculation. This procedure was reiterated 19 times and 190 hypotheses were analyzed, none of which showed better correlation values than the generated hypothesis (Hypo_test). On Table S2 in Supplementary Information, the activity variations and correlations can be observed. In variation number 12, the closest correlation to the generated hypothesis was obtained (r = 0.79). However, in this case the cost of the hypothesis was 248.3 bits (higher than our null hypothesis) and moreover it was a 3 hydrophobic point model. All of this indicates a probability of 90 % of true correlation among data and suggests that the HypoGen model is optimal.

## In silico screening

The representative common pharmacophore hypothesis (Hypo_test) was used as a search query to retrieve compounds with novel structural scaffolds and desired chemical features. The four-point pharmacophore hypothesis was directly used as a query for virtual screening, looking for

additional candidates for biological testing among the molecules that contain the pharmacophore in their lower energy conformation.

The "fast flexible search" option in Catalyst was first selected to screen the following 3D-databases: Maybridge (55.273 compounds), Chemical Diversity (374.316 compounds), Specs (255.040 compounds) and NCI (238.819 compounds). This yielded a collection of about 10.000 compounds that shared, in some conformation, the same 3D representations of the functional groups. Based on this search, iterative searched were performed restricting tolerance in angles and distances.

In the first search we only restrict the distances with a tolerance of 0.1 Å because the tolerance of the correlation sphere was 1.6 Å. In this way we found 10,793 compounds that match this restriction. On top of this search, the second run limited the opening of the angles to 10°, which is ±5° over the original value. In this case we found 90 compounds, what constitutes a high number to carry out an analysis of all of them. That is why in the third run we limited the tolerance of the angles to 5° (±2.5°) and the correlation sphere to 1 Å. In this last search 56 compounds were found.

Additionally, a subset of these structures was selected by removing all the molecules that did not satisfy the well-known Lipinski rules, describing properties of drug-like compounds (see Figure S2-S3 in Supplementary Information). Therefore, Lipinski indexes were calculated for each one of the 56 compounds and those without a number equal to 4 were discarded.

Thus, through three computational techniques, a cascade selection of candidates took place (Fig. 7), which allowed the identification of 32 compounds, a 0.0035 % of the initial compounds.

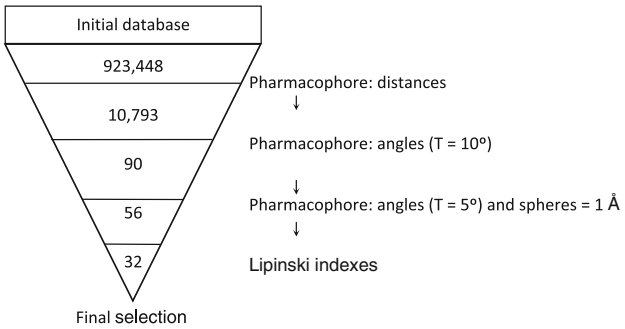


**Fig. 7** Summary of the virtual screening cascade during the hit finding process

**Table 4** Statistical a priori analysis of all the data series to select the prediction of the toxicity model

| Series | | McFadden $R^2$ (%) | | Hannan-Quinn criterion | | *p* Values (%) | |
|---|---|---|---|---|---|---|---|
| | | Logit | Gompit | Logit | Gompit | Logit | Gompit |
| **M5** | 3 PC | 1.83 | 1.81 | 1.436 | 1.436 | 36.30 | 36.88 |
| | 2 PC | 1.82 | 1.80 | 1.411 | 1.411 | 20.53 | 20.76 |
| | 1 PC | 1.63 | 1.60 | 1.389 | 1.389 | 9.21 | 9.46 |
| **M8** | 3 PC | 3.25 | 3.14 | 1.422 | 1.423 | 15.32 | 16.59 |
| | 2 PC | 3.25 | 3.14 | 1.395 | 1.397 | 7.19 | 7.88 |
| | 1 PC | 1.80 | 1.77 | 1.389 | 1.389 | 8.76 | 9.06 |
| **M11** | 3 PC | 3.52 | 3.45 | 1.432 | 1.433 | 14.86 | 15.55 |
| | 2 PC | 3.31 | 3.22 | 1.406 | 1.408 | 8.09 | 8.68 |
| | 1 PC | 2.54 | 2.48 | 1.389 | 1.390 | 4.95 | 5.23 |
| **M14** | 3 PC | 4.21 | 4.12 | 1.441 | 1.442 | 15.58 | 16.30 |
| | 2 PC | 4.01 | 3.88 | 1.410 | 1.412 | 8.26 | 8.96 |
| | 1 PC | 3.31 | 3.22 | 1.387 | 1.388 | 4.25 | 4.55 |
| **M15** | 3 PC | 4.01 | 3.93 | 1.424 | 1.426 | 13.84 | 14.57 |
| | 2 PC | 3.29 | 3.29 | 1.404 | 1.404 | 10.44 | 11.29 |
| | 1 PC | 2.23 | 2.18 | 1.388 | 1.389 | 4.01 | 4.38 |

## In silico studies of toxicity

### Statistical analysis

The five data series shown in Table 2, were analyzed to measure their a priori forecasting capacity in the estimation of two nonlinear models named Logit and Gompit (Table 4). The criteria employed are detailed below:

McFadden R2: It is similar to the statistical R2, but used for non linear models. Higher values indicate a better fit. The values obtained clearly indicate that the "PC variables" provide larger values and, therefore, better fits [39].

Hannan-Quinn criterion: It takes into account the fitting (the more, the better) and the number of variables (the less, the better). Higher values indicate higher forecasting power out of the sample [40]. According to this statistic, the best specifications are those that use a single principal component as explanatory variable.

*p* values of a significance statistic: They contrast the null hypothesis that all the β parameters are zero, meaning that there is no relationship between the physical properties of a molecule and its toxicity, against the alternative that $\beta$ is non-zero. The *p* values corresponding to the best model are 4.01 and 4.38 %, meaning that the risk of error in rejecting the null is roughly 4 %.

In all cases, the forecasting capacity of the series increase as follows: M5 < M8 < M11 < M14 < M15 (Table 4). Therefore, the introduction of small size molecules reduces the forecasting power but increases the range of molecules that the model would be able to predict. Moreover, among all series, the first component is the most significant variable for the prediction of toxicity.

To test the model robustness, we chose for evaluation a set of molecules which chemical structure could be compromised by toxic effects due to the presence, in the case of

non toxic compounds, or the absence in the case of toxic ones, of structural alerts (Figure S4-S5 in Supplementary Information) [41]. Table S3 in Supplementary Information shows the predictions of models for the evaluation set. Percentage of correct guesses was 66.7 %.

Although considering smaller molecules will reduce the forecasting power of the series, it increases the number of the molecules the model can predict. Therefore, we decided to estimate a new model with series M5, therefore able to predict the theoretical mutagenicity of any molecule with more than 5 atoms different than Hydrogen. In this case, the cut off was established in 54/128 = 0.42, obtaining a 55 % of good hits with both models. Table S4 in Supplementary Information shows the predictions of both models for the evaluation set, with a percentage of correct guesses of 62.5 % and able to predict any size molecules.

As result, using the initial M15 model, because all molecules had more than 14 atoms different than Hydrogen, the mutagenicity of molecules obtained from 3D database search from the designed pahrmacophoric model was predicted, showing that only the compound coded 5989–3949 was predicted as toxic. Therefore, the rest of them do still show appropriate drug-like properties (Table S5 in Supplementary Information).

## ADME in silico studies

### Prediction of ADME properties by Volsurf

The importance of favorable pharmacokinetic characteristics has been widely recognized in drug design [42]. The Volsurf approach was used in ADME modeling of drug-like compounds [43]. Compounds selected for the previous in silico screening were submitted to Volsurf to predict some ADME properties such as BBB penetration, thermodynamic stability in solution, Caco2 permeability, human serum albumin binding and hERG potassium channels blockers. Molecular descriptors calculated by Volsurf were applied to model pharmacokinetic parameters.

The pharmacokinetic features of selected compounds are shown in Table S6 (Supplementary Information).

From BBB penetration results, none of the compounds was discarded since any of them could be consider as a lead compound both in central or peripheral nervous system diseases. Concerning thermodynamic stability in solution, Caco2 permeability and human serum albumin binding, which is the main biological carrier of drugs, all the results showed suitable profiles.

For the prediction of blockage of hERG potassium channels, 6 compounds were discarded because of they were able to block these potassium channels, and consequently they could produce cardiac arrhythmias.

Finally, 12 of the selected compounds (Fig. 8) were purchased from commercial sources, and the biological evaluation confirmed their PDE7 inhibition for compounds 4571-0010, 2275-1061 and K784-3786 in the micromolar range [44].

### Identifications of the responsible cytochromes of the new prototypes

Since cytochrome inhibition and induction are the key mechanisms in drug-drug interactions, the molecules were introduced in Volsurf to predict in silico which cytochrome would be responsible for the metabolism of each of the compounds. As consequence, the elaborated model was used to predict that cytochromes CYP3A4 and CYP2C9 would be responsible of the metabolism of the 12 selected compounds (Figure S6 in Supplementary Information).

## Biological evaluation

To check their anti-inflammatory profile, we analyzed three parameters in the T cell line D10.G4.1.

### Intracellular cAMP levels

cAMP is an important second messenger whose levels are regulated by synthesis, mediated by the action of the enzyme Adenyl cyclase, and degradation by phosphodiesterases [45]. To validate the hypothesis that the found compounds are able to inhibit PDE activity, first of all we determined their ability to increase the intracellular cAMP levels. As shown in Fig. 9, the three selected compounds were able to significantly increase the cAMP levels in T lymphocytes, although none of them could reach levels as high as the PDE4 inhibitor Rolipram used as a positive control, what is not surprising since PDE4 is the major PDE in T lymphocytes. cAMP is involved in a wide range of cellular functions, including inhibition of the progression phase of T cell activation in response to mitogen [46]. Sustained increase of cAMP mediates a variety of immunoregulatory signals which leads to suppression of immune reactions and inflammation [47]. Thus, the ability of these three selected compounds to increase intracellular cAMP levels could constitute an anti-inflammatory potential.

### Proliferation assays

Deregulation of immune response leads to inflammation. As T cell activity is involved in both cell and humoral responses [48, 49], inhibition of T lymphocyte proliferation is considered as a suitable feature of anti-inflammatory drugs. To address whether these compounds had anti-proliferative activity, we next carried out proliferation assays

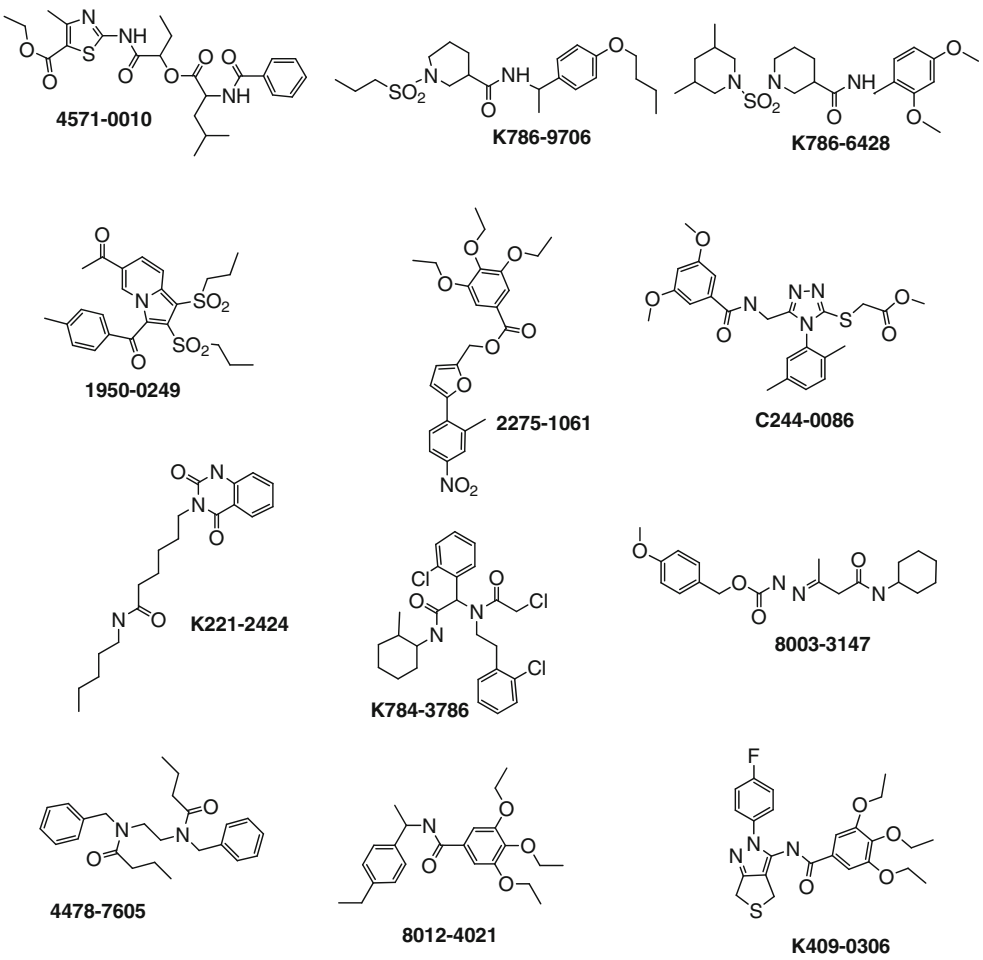


**Fig. 8** Hit compounds

with T lymphocytes activated through the T cell receptor by anti-CD3 antibody in the presence of increasing doses of the selected compounds. Results demonstrated an anti-proliferative effect in a dose-dependent way for the three compounds assayed (Fig. 10). Thus, these three compounds could control an inadequate immune response mediated by T lymphocytes.

Production of the anti-inflammatory cytokine IL-4

T helper (Th) cells produce and induce other cell types to produce cytokines with great influence on inflammation. Different cytokines favor or restrain different inflammatory processes. Th2 cells are IL-4 and IL-5 producers, which are involved in allergic inflammation and asthma [50, 51]. Meanwhile cytokines produced by Th1 (IFNγ) and Th17 (IL-17, IL-22) subsets are related to allograft rejection [52, 53], and to autoimmune disorders as multiple sclerosis [54], rheumatoid arthritis [55, 56], celiac disease [57], diabetes mellitus [58, 59] or psoriasis [60]. Th phenotype definition is thoroughly regulated by the cell environment. Th1, Th2 or Th17 phenotypes are directed by IL-12, IL-4

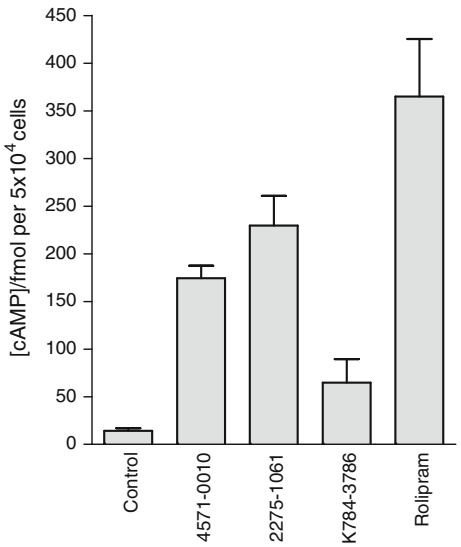


**Fig. 9** Intracellular cAMP levels in D10.G4.1 cells treated during 1 h at 37 °C with the indicated compound (100 μM). Values were obtained as fmol of cAMP in $5 \times 10^4$ cells by competitive enzyme immunoassay (EIA) system. Each sample was assayed in triplicate. Error bars represent standard deviation

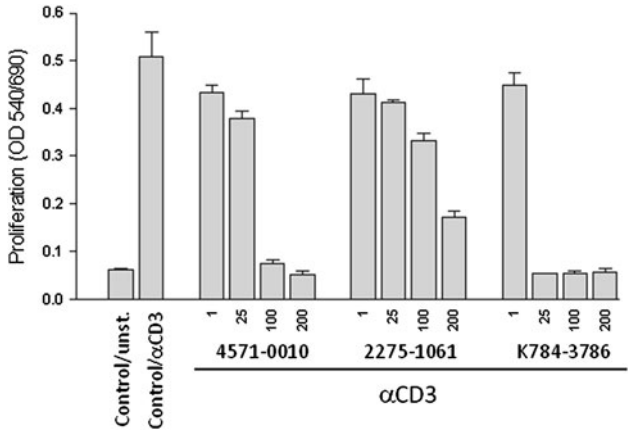


**Fig. 10** Proliferation of D10.G4.1 cells stimulated by plate-bound anti-CD3 in the presence of the indicated compound. Concentrations for each compound are indicated in μM. After 72 h of culture, proliferation was determined by colorimetric assay. Results are presented as absorbance (O.D.). Each sample was assayed in triplicate. Controls unstimulated and anti-CD3-stimulated are showed on the left. *Error bars* represent standard deviations

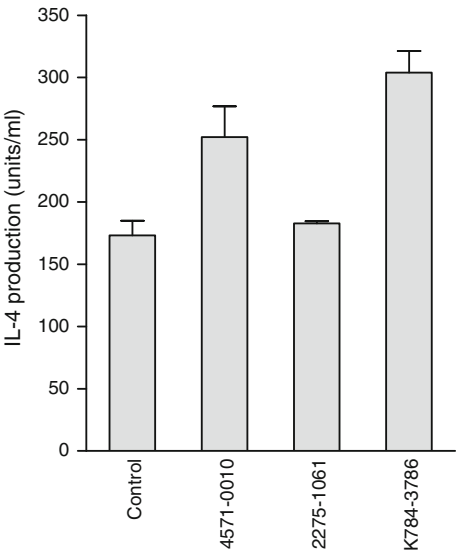


**Fig. 11** IL-4 production by D10.G4.1 cells stimulated by plate-bound anti-CD3 in the presence of the indicated compound (100 μM). Levels of IL-4 in the supernatants of cultures after 48 h were measured by ELISA. Each sample was assayed in triplicate and IL-4 values were normalized to proliferation for each culture. Error bars represent standard deviations

and IL-6 + TGFβ, respectively. So, immunotherapy searching for Th1- or Th17-mediated diseases is very often based in strategies for deviation of Th phenotype in favor of the IL-4 producer subtype Th2 [61–65].

Thus, we wanted to evaluate the effect of hit compounds on the levels of IL-4 secreted to the medium by T lymphocytes stimulated by anti-CD3. As shown in Fig. 11, two of them led to an increase of the amount of this cytokine, suggesting that they could be suitable agents for Th2 phenotype promotion.

## Conclusions

Many immune responses are regulated by cAMP as second messenger in cell signal transduction. Thus, agents able to increase intracellular cAMP levels have immunosuppressive properties by inhibition of T cell proliferation and/or modulation of cytokine production. Biological assays also showed ability to decrease T cell proliferation, most probably due to the increase of intracellular cAMP levels mediated by those molecules. Therefore, the iterative strategy and the combined use of some computational techniques reported herein have allowed the discovery of some molecules with potential to control inflammatory processes mediated by T lymphocytes, such as multiple sclerosis, diabetes mellitus, psoriasis, systemic lupus erythematosus or rheumatoid arthritis. In addition, T lymphocytes treated with two of the drugs found here secreted increased levels of IL-4, a cytokine with well known anti-inflammatory properties in autoimmune disorders in which Th1 and Th17 are involved.

This virtual screening strategy has taken into account not only biological activity criteria but also some ADME-TOX properties, making it a useful tool for the determination of new hit compounds. On one hand, toxicity risk was evaluated taking into account the predictive value of statistical models based on Ames test data. On the other hand, the study of which cytochrome could show more influence over those compounds was also analyzed.

In addition to these studies, it is also important to consider the common chemical modifications for a further lead optimization processes, such as the substitution of ester function to avoid the rapid hydrolysis by esterases or the elimination of chloroacetyl amide as reactive specie.

Finally, it is important to take into account that no predictive model is infallible. From this premise, the main advantage of this strategy is the fact that together with other criteria such as their chemical accessibility, pharmacokinetic properties should reinforce an appropriate selection of candidate compounds for preclinical development.

**Acknowledgments** The authors acknowledge computational resources provided by Center for Scientific and Academic Services of Catalonia (CESCA), the technical assistance of Beatriz Bravo and Raquel Gómez, and the support of the FIS project PS09/00604 and Ministerio de Economia y Competitividad (project: CTQ2010-19690). Miguel Jerez gratefully acknowledges financial support from Ministerio de Economía y Competitividad through the Grant ECO2011-23972.

## References


1. Oprea TI, Matter H (2004) Curr Opin in Chem Biol 8:349
2. Gleeson MP, Modi S, Bender A, Robinson RL, Kirchmair J, Promkatkaew M, Hannongbua S, Glen RC (2012) Curr Pharm Des 18(9):1266
3. Ekins S, Waller CL, Swaan PW, Cruciani G, Wrighton SA, Wikel JH (2000) J Pharmacol Toxicol Methods 44(1):251
4. Lugnier C (2006) Pharmacol Ther 109:366
5. Giembycz MA, Smith SJ (2006) Curr Pharm Des 12:3207
6. Li L, Yee C, Beavo JA (1999) Science 283:848
7. Yang G, McIntyre KW, Towsend RM, Shen HH, Pitts WJ, Dodd JH, Nadler SG, McKinnon M, Watson AJ (2003) J Immunol 171:6414
8. Nakata A, Ogawa K, Sasaki T, Koyama N, Wada K, Kotera J, Kikkawa H, Omori K, Kaminuma O (2002) Clin Exp Immunol 128:460
9. Castro A, Jerez MJ, Gil C, Martinez A (2005) Med Res Rev 25:229
10. Gil C, Campillo NE, Perez DI, Martinez A (2008) Expert Opin Ther Pat 18:1127
11. Martínez A, Castro A, Gil C, Miralpeix M, Segarra V, Doménech T, Beleta J, Palacios JM, Ryder H, Miro X, Bonet C, Casacuberta JM, Azorin F, Pina B, Puigdoménech P (2000) J Med Chem 43:683
12. Castro A, Abasolo MI, Gil C, Segarra V, Martínez A (2000) Eur J Med Chem 36:333
13. Barnes MJ, Cooper N, Davenport RJ, Dyke HJ, Galleway FP, Galvin FCA, Gowers L, Haughan AF, Lowe C, Meissner JWG, Montana JG, Morgan T, Picken CL, Watson RJ (2001) Bioorg Med Chem Lett 11:1081
14. EP 1 193 261 A1 2000
15. Vergne F, Bernardelli P, Lorthiois E, Pham N, Proust E, Oliveira C, Mafroud AK, Royer F, Wrigglesworth R, Schellhaas JK, Barvian MR, Moreau F, Idrissi M, Tertre A, Bertin B, Coupe M, Bernaa P, Soulardc P (2004) Bioorg Med Chem Lett 14:4607
16. Vergne F, Bernardelli P, Lorthiois E, Pham N, Proust E, Oliveira C, Mafroud AK, Ducrot P, Wrigglesworth R, Berlioz-Seux F, Coleon F, Chevalier E, Moreau F, Idrissi M, Tertre A, Descours A, Bernaa P, Lib M (2004) Bioorg Med Chem Lett 14:4615
17. US 2002/0198198 A1 2002
18. Lorthiois E, Bernardelli P, Vergne F, Oliveira C, Mafroud AK, Proust E, Heuze L, Moreau F, Idrissi M, Tertre A, Bertin B, Coupe M, Wrigglesworth R, Descours A, Soulardb P, Berna P (2004) Bioorg Med Chem Lett 14:4623
19. Bernardelli P, Lorthiois E, Vergne F, Oliveira C, Mafroud AK, Proust E, Pham N, Ducrot P, Moreau F, Idrissi M, Tertre A, Bertin B, Coupe M, Chevalier E, Descours A, Berlioz-Seux F, Berna P, Li M (2004) Bioorg Medi Chem Lett 14:4627
20. Bykgulden Lomberg Chemische Farbrik GMBH. WO 02/40449 A1 2002
21. Bykgulden Lomberg Chemische Farbrik GMBH. WO 02/40450 A1 2002
22. SmithKline Beecham Corp. US2002/0156064 2002
23. Catalyst 4.6, Accelerys Inc.: San Diego, CA, 2004
24. http://toxnet.nlm.nih.gov/cgi-bin/sis/htmlgen?GENETOX
25. Ochoa C, Rodríguez J, Rodríguez M, Chana A, Stud M, Alonso-Villalobos P, Martínez-Grueiro MM (1998) Med Chem Res 7:530
26. Martinez A, Castro A, Stud M, Rodriguez J, Cardelus I, Llenas J, Fernandez A, Palacios JM (1999) Med Chem Res 8:171
27. Dorronsoro I, Chana A, Abasolo MI, Castro A, Gil C, Stud M, Martinez A (2004) QSAR 23:89
28. Quiñones C, Caceres J, Stud M, Martinez A (2000) QSAR 19:448
29. Version 3.3.; Oxford Molecular, Ltd
30. Cruciani G, Crivori P, Carrupt PA, Testa B (2000) J Mol Struc THEOCHEM 503:17
31. Cruciani G, Carosati E, De Boeck B, Ethirajulu K, Mackie C, Howe T, Vianello R (2005) J Med Chem 48:6970
32. Kastenholz MA, Pastor M, Cruciani G, Haaksma EEJ, Fox T (2000) J Med Chem 43:3033
33. Kaye J, Porcelli S, Tite J, Jones B, Janeway CA Jr (1983) J Exp Med 158:836
34. Click RE, Benck L, Alter BJ (1972) Cell Immunol 3:156
35. Dorado B, Jerez MJ, Flores N, Martín-Saavedra FM, Durán C, Ballester S (2002) J Immunol 169:3030
36. Portoles P, Rojo J, Golby A, Bonneville M, Gromkowski S, Greenbaum L, Janeway CA Jr, Murphy DB, Bottomly K (1989) J Immunol 142:4169
37. Mosman T (1983) J Immunol Methods 65:55
38. Martín-Saavedra FM, Flores N, Dorado B, Eguiluz C, Bravo B, García-Merino A, Ballester S (2007) Mol Immunol 44:3597
39. McFadden D, Zarembka P (ed) (1974). Academic Press, Nueva York, pp 105–142
40. Hannan EJ, Quinn BG (1979) J R Stat Soc Ser B Stat Methodol 41:190
41. Snyder RD, Pearl GS, Mandakas G, Choy WN, Goodsaid F, Rosenblum IY (2004) Environ Mol Mutagen 43:143
42. Crivori P, Poggesi I (2006) Eur J Med Chem 41:795
43. Cruciani G, Pastor M, Guba, W (2000) Eur J Pharm Sci 11
44. Gil C, Castro A, Jerez MJ, Ke H, Wang H, Ballester S, Gonzalez-Garcia C, Martinez A (2008) Drugs Future 33(Suppl A):228
45. Mosenden R, Tasken K (2011) Cell Signal 23:1009
46. Lingk DS, Chan MA (1990) J Immunol 145:449
47. Brudvik KW, Tasken K (2012) Br J Pharmacol 166:411
48. Castellino F, Germain RN (2006) Annu Rev Immunol 24:519
49. Davis MM, Krogsgaard M, Huse M, Huppa J, Lillemeier BF, Li QJ (2007) Annu Rev Immunol 25:681
50. Romagnani S (2004) J Allergy Clin Immunol 113:395
51. Antoniu SA (2009) BioDrugs 23:241
52. Brandacher G, Margreiter R, Fuchs D (2007) Curr Drug Metab 8:273
53. Heidt S, Segundo DS, Chadha R, Wood KJ (2010) Curr Opin Organ Transplant 15:456
54. Pierson E, Simmons SB, Castelli L, Goverman JM (2012) Immunol Rev 248:205
55. Toh ML, Miossec P (2007) Curr Opin Rheumatol 19:284
56. Annunziato F, Cosmi L, Liotta F, Maggi E, Romagnani S (2009) Nat Rev Rheumatol 5:325
57. Lerner A (2010) Autoimmun Rev 9:144
58. Raz I, Eldor R, Naparstek Y (2005) Trends Biotechnol 23:128
59. Haskins K, Cooke A (2011) Curr Opin Immunol 23:739
60. Farkas A, Kemeny L (2012) Int Immunopharmacol 13:215
61. Slavin AJ, Tarner IH, Nakajima A, Urbanek-Ruiz I, McBride J, Contag CH, Fathman CG (2002) Autoimmun Rev 1:213
62. Ghoreschi K, Rocken M (2004) Curr Drug Targets Inflamm Allergy 3:193
63. McGeachy MJ, Anderton SM (2005) Cytokine 32:81
64. Pierson E, Simmons SB, Castelli L, Goverman JM (2012) Immunol Rev 248:205
65. Oreja-Guevara C, Ramos-Cejudo J, Stark Aroeira L, Chamorro B, Diez-Tejedor E (2012) BMC Neurol 12:95