Detection of potential enzyme targets by metabolic modelling and optimization. Application to a simple enzymopathy

Julio Vera-González1,2,3,*, Raul Curto4, Marta Cascante5 and Néstor V. Torres2,3

1Systems Biology and Bioinformatics Group, University of Rostock, Albert Einstein Str. 21, 18051, Germany, 2Grupo de Tecnología Bioquímica, Department of Biochemistry and Molecular Biology, University of La Laguna, Astrofísico Francisco Sanchez s/n, 38206, Spain, 3Instituto Canario de Investigación del Cáncer (ICIC), Canary Islands, Spain, 4University School of Experimental Sciences and Technology (EUCET), Universitat Internacional de Catalunya, Nova estacio s/n, 43500, Spain and 5Department of Biochemistry and Molecular Biology, Faculty of Chemistry and CERQT-Parc Científic de Barcelona (PCB), University of Barcelona, Martí i Franques 1, 08028, Spain

Associate Editor: Prof. Thomas Lengauer

ABSTRACT

Motivation: A very promising approach in drug discovery involves the integration of available biomedical data through mathematical modelling and data mining. We have developed a method called optimization program for drug discovery (OPDD) that allows new enzyme targets to be identified in enzymopathies through the integration of metabolic models and biomedical data in a mathematical optimization program. The method involves four steps: i) collection of the necessary information about the metabolic system and disease; ii) translation of the information into mathematical terms; iii) computation of the optimization programs prioritizing the solutions that propose the inhibition of a reduced number of enzymes; iv) application of additional biomedical criteria to select and classify the solutions. Each solution consists of a set of predicted values for metabolites, initial substrates and enzyme activities, which describe a biologically acceptable steady state of the system that shifts the pathologic state towards a healthy state.

Results: The OPDD was used to detect target enzymes in an enzymopathy, the human hyperuricemia. An existing S-system model and bibliographic information about the disease were used. The method detected six single-target enzyme solutions involving dietary modification, one of them coinciding with the conventional clinical treatment using allopurinol. The OPDD detected a large number of possible solutions involving two enzyme targets. All except one contained one of the previously detected six enzyme targets. The purpose of this work was not to obtain solutions for direct clinical implementation but to illustrate how increasing levels of biomedical information can be integrated together with mathematical models in drug discovery.

1 INTRODUCTION

One of the most potentially fruitful approaches to boost the development of new drugs is through the integration of available biomedical data by mathematical analysis and data mining. In this context, predictive biosimulation involves implementation of advanced mathematical models that can simulate the biological systems that make up the human body (Voit 2002). The models integrate available metabolic, genomic and proteomic information to describe and predict the behaviour of the simulated biomedical system.

The studied disease. In the classical hyperuricemia, a functional defect in the enzyme phosphoribosylpyrophosphate synthetase (PRPPS) –that controls the synthesis of purines- provokes an increase in its enzymatic activity and leads to an augmentation of degradative metabolic fluxes yielding uric acid (Scrivner et al. 1989). The uric acid is stored in the form of urate crystals in some tissues, especially in joints. The symptoms of the disease include acute episodes of arthritic pain and various kinds of nephropathy. Currently, treatment of this disease usually includes a symptomatic treatment for joint pain, a restricted diet that precludes consumption of food with high concentrations of purine precursors, and the prescription of allopurinol (Klinenberg et al. 1965). Allopurinol is a specific inhibitor of the enzyme xanthine oxidase that can lead to a drastic reduction in the concentrations of uric acid. Classical hyperuricemia is of particular interest for use in a case study because, in addition to having extensively described symptoms, its functional and metabolic basis is well understood.

Mathematical framework. S-system models are a special set of power-law models initially developed to study dynamic and steady-state properties of metabolic pathways (Savageau 1969a, 1969b, 1970; see Voit 2000 for a recent review). In S-system models, the metabolic network is represented by a mathematical model in ordinary differential equations in which a net (aggregated) input flux and a net (aggregated) output flux represent the dynamics of each time-dependent metabolite. Both aggregated fluxes are represented as power-law terms. The validity of the S-systems and other power-law models as an alternative approach to model complex biochemical systems has been proved in recent times for metabolic systems (Alvarez-Vasquez et al. 2005), genetic circuits (Atkinson et al. 2003) and cell signalling systems (Vera et al. 2007). The structural simplicity of S-system models generates a set of well-known analytical properties. By using a simple logarithmic transformation, the steady states of any S-system model can be evaluated using linear algebra, which permits rapid analysis of many important properties of the system. In addition, S-systems can explore optimization problems using linear programming, which allows biological optimization programs of high dimensionality to be analyzed with fast linear algorithms (Voit 1992, Torres et al. 1997). A complete section explaining the essential mathematical properties of S-systems and their use in biological optimization programs is available in the attached document, “Introduction to S-system models in optimization” (ISS1-ISS5). The optimization of metabolic systems is not a new topic, but in previous contributions (Voit 1992, Torres et al. 1997, Mendes et al. 1998) this approach was mainly used in metabolic engineering.

© The Author (2007). Published by Oxford University Press. All rights reserved. For Permissions, please email: journals.permissions@oxfordjournals.org
2 METHODS

The aim of the present work was to develop a mathematical approach to the rational identification of potential enzyme targets for the treatment of metabolic diseases, mainly single enzyme pathologies. The approach is based on dynamic mathematical modelling of the metabolic network responsible for the disease and the use of standard optimization routines.

The optimization program for drug discovery (OPDD) can be applied to those diseases that meet three general criteria. First, the metabolic disease should be caused by a structural or functional alteration (genetic or degenerative) that causes a significant alteration of the catalytic activity in one or more enzymes. Second, the enzyme deficiency should cause physiologically relevant changes (increases or decreases compared with the values in a healthy individual) in the concentration of one or more metabolic intermediates and/or fluxes. Finally, the changes should be responsible for the symptoms of the disease, either directly or by affecting other related metabolic processes. Thus, in the OPDD framework for the analysis is the metabolic network involved in the disease rather than a single biochemical process.

The usual steady state –values in metabolite concentrations and fluxes - of a metabolic network in a typical healthy individual will be referred as the healthy state (HS). In contrast, the pathologic state (PS) corresponds to that of an individual suffering from the disease. In this state, the system contains one or more enzymes with significantly altered catalytic activity that provokes changes in certain fluxes and metabolite concentrations that ultimately lead to the manifestation of the symptoms.

We can alter the dynamics of a metabolic network by using drugs to modulate enzyme activities. In this case, the aim is to alter the values of critical metabolite concentrations and fluxes in the network and shift them towards those encountered in the HS. Eventually we can also change the initial substrate concentrations, e.g. by dietary restriction. We hypothesise that the intervention proposed will restore metabolic levels existing in the healthy state, which will avoid enzyme overexpression due to pathological metabolic accumulations.

Once the metabolic network and variables associated with the PS are known, the OPDD aims to identify modifications of enzyme activities and substrate concentrations that shift the system to a steady state as close as possible to that of the HS, while also verifying all the additionally imposed physiological and biological constraints (see Figure AD1.1). Actually, the OPDD focuses on restoring the natural values for the metabolites and fluxes that are considered critical for the disease (key metabolites and fluxes), which are determined using the available biomedical information.

2.1 Mathematical structure of the OPDD method

The ideas proposed above can be translated into a mathematical framework for the identification of new target enzymes for therapeutic treatment of metabolic diseases. The starting point for any OPDD application is a reliable mathematical model in ordinary differential equations describing the biochemical system under consideration. The models must describe the main biochemical processes that make up the network. Given that the aim is to shift the PS towards the HS, we look for solutions in which the values of metabolite concentrations and fluxes that play a key role in the manifestation of the disease (the so-called key metabolites and fluxes) will be shifted towards those seen in the HS. This aim can be mathematically represented in the following objective function:

\[
\text{Min} \left[ \sum_{j=1}^{n_i} \lambda_j |X_j - X^{j^*}_i| + \sum_{i=1}^{n_i} \lambda_i |J_i - J^{i^*}_i| \right]
\]

where \( \lambda_j, \lambda_i \geq 0 \), and \( X_j \) and \( J_i \) are the key metabolites and fluxes, respectively. The values of \( \lambda_j \) and \( \lambda_i \) are determined by the relative importance of each key metabolite and flux in relation to the symptoms and they are assigned using the available biomedical information. In well studied diseases (e.g. the case study used in this work), this is easy because the symptoms of the disease are well described. In more complex diseases, interaction with biomedical researchers and techniques of inference will be necessary to elucidate the values of these coefficients. We model the functional origin of the disease by imposing a value to the activities of the damaged enzymes that represents the changes provoked by the disease:

\[
X_i = X^{def}_i
\]

In this equation, \( X_i \) represents the activity value for the deficient enzyme \( k \) and \( X^{def}_i \) is its characteristic value in the PS. The solution near to the HS will be a stable steady state of the considered metabolic network. This steady state condition translates in mathematical terms into the following set of equations:

\[
\frac{dX_i}{dt} = 0, \quad i = 1, \ldots, n
\]

Moreover, we have to assume additional conditions to guarantee physiologically acceptable solutions. These conditions can be modelled with the following set of equations:

\[
X^{LB}_i \leq X_i \leq X^{UB}_i, \quad i = 1 \ldots n_D
\]

where \( n_D \) is the number of metabolites in the network, and \( X^{LB}_i \) and \( X^{UB}_i \) establish the physiologically admissible lower and upper bounds for each metabolite. Similarly, we impose physiological limits for the independent variables (substrates as well as modifiable enzyme activities):

\[
X^{LB}_j \leq X_j \leq X^{UB}_j, \quad j = n_D + 1 \ldots n_D + n_I
\]

where \( n_I \) is the number of independent variables in the network, and \( X^{LB}_j \) and \( X^{UB}_j \) the bounds imposed. When information is available, similar constrains can be imposed to metabolic fluxes, in which we describe feasible physiological intervals for them. Taken together, these mathematical conditions make up a non-linear optimization program (see Scheme AD1.1).

2.2 General structure of the OPDD method

The outlined approach integrates modelling of the system and identification of target enzymes with the analysis of the available information and the choice of the best solutions. The method includes four sequential steps:

Step 1. Information gathering and construction of the disease model. Essential biological information about the system and the disease is collected in the first step. A reliable mathematical model based on differential equations that describe the metabolic system, including the relevant biochemical processes, has to be selected. A characterisation of the metabolic basis of the disease is also considered, including essential information about the genetic and/or
Detection of potential enzyme targets by metabolic modelling and optimization. Application to a simple enzymopathy

functional origin of the disease and the features defining its associated PS. In addition, analysis of the critical interactions between the specific network and general metabolism is important to avoid undesirable interactions between the designed treatments and other metabolic processes.

**Step 2. Optimization.** This step involves translation of the following into mathematical terms: i) definition of the pathologic state of the system, assigning an appropriate value for the activity of the deficient enzyme; ii) choice of the set of key metabolites and fluxes that deviate from the values seen in healthy individuals, leading to the manifestation of symptoms; iii) selection of the set of target enzymes that can be pharmacologically modified (mainly by inhibiting their activity); iv) establishment of a plausible physiologically admissible interval value for each metabolite and modifiable enzyme; and v) definition and implementation of the optimization objective that integrates the available information about the key metabolites and fluxes, their values in the HS and their importance in the development of symptoms.

**Step 3. Computation of solutions.** Each time the defined optimization program is carried out a feasible solution is obtained. Each solution consists of a set of predicted values for metabolites, initial substrates and enzyme activities that describes a biocompatible acceptable steady state of the system that shifts the PS towards the HS. Those solutions that modify the least number of enzymes are considered. The higher the number of enzymes to be modified, the greater the number of necessary drugs; this could increase the level of adverse effects of the potential treatment. This idea is applied in a two-step process: i) sequential generation of the solutions, starting with the simplest treatments (with a minimum number of modified enzymes) and increasing the complexity of the solutions; ii) mathematical analysis of the solutions in an effort to discard those that are not sufficiently stable and robust.

**Step 4. Selection and classification of solutions.** Additional selection criteria are necessary in order to select the most feasible solutions from within the generated set. A minimum acceptable value for the objective function can be established to allow selection of the solutions. In addition, biomedical criteria can be mathematically imposed. Examples include imposition of a certain biochemical condition on the solutions or limiting the range of changes in enzyme activities. In this way, a reduced, biologically acceptable set of good candidates is obtained for experimental selection using conventional protocols in pharmacological research. A complete explanation of the steps used in the method and a scheme are available in the additional information section (AD1). The information obtained allows us to determine which enzymes or set of enzymes in the network are candidates for inhibition and how much we have to alter their activities in order to get the best therapeutic results. The computational core of the OPDD is contained in Step 3. The computational requirements of this step will depend on the characteristics of the selected mathematical model. The method can be applied to different modelling approaches (Michaelis-Menten models, S-systems, General Mass Action (GMA) models, etc.), but its computational efficiency is maximized through the use of S-systems models: the sequentially generated optimization programs become linear in the logarithmic space of the variables and can then be solved using the Simplex algorithm (see ISS3 and ISS4). The OPDD results are highly intuitive and we can thus concentrate efforts and attention on the biological aspects of the problem.

**3 RESULTS**

**3.1 Application of the OPDD to human hyperuricemia**

In the following section we apply the OPDD method to the detection of single or combined target enzymes for the treatment of classical hyperuricemia in humans.

**Step 1. Modelling purine metabolism in humans.** In the present work we have used a modified version of an S-system model initially developed by the Cascante group (Curto et al. 1997, 1998a, b Voit 2000) which describes the purine metabolism in humans. This model includes the synthesis, recovery and degradation of purine nucleotides, but also takes into account the regulation of the enzyme activity for metabolites, either substrates or products. In Curto et al. (1998a) this S-System model was compared with a Complemented Michaelis-Menten model and a GMA model of the same system with similar level of detail; the results of such comparison were that the predictions of the three models were similar and consistent with clinical findings for physiological and moderately pathological perturbations in metabolites or enzymes. The original model was modified such that the activities of the enzymes involved in the network were converted into decision variables and the whole set of variables (including metabolites and enzyme activities) was normalised with respect the original basal state of the system defined in Curto et al (1997). This normalization is not mandatory but facilitates the visualization and analysis of the solutions. An extensive explanation of the model, a table with the complete name of the variables, a graphic scheme and the set of ordinary differential equations are included in AD2.

**Step 2. Optimization program.** Using the information available on the disease we can configure the proposed OPDD. Mathematical definition of the healthy state. The HS of the system, which in this case coincides with the original basal state of the system, can be defined by assigning a value equal to 1 to all of the variables in the model (metabolite concentrations and enzyme activities). In mathematical terms this translates to

\[ X_{1}^{HS} = 1 \quad i = 1, ..., 46 \]

Mathematical characterization of the pathologic state. In its most common presentation, hyperuricemia is caused by overactivity of the enzyme PRPPS \((X_{19}, \text{see Figure AD2.1})\). This increases the concentration of inosine monophosphate (IMP, \(X_{2}\)) causing an increase in the concentration of uric acid and leading to the symptoms of the disease. To model the PS, we doubled the value of the variable that represents the activity of PRPPS \((X_{19}, \text{with respect to the HS (Curto et al. 1998b):})\)

\[ X_{19}^{PS} = 2 \cdot X_{19}^{HS} \]

In this way, we model the overactivity of the enzyme. The concentration of uric acid \((X_{16})\) was considered the only key metabolite in the OPDD strategy because the deviation in uric acid concentration is considered the principal effect of hyperuricemia (Scriver et al. 1989, Curto et al. 1998b). Accordingly, the optimization objective for the OPDD program will be

\[ \text{Min} \quad \left| X_{16} - X_{16}^{HS} \right| \]
This means that we must look for solutions that push the value of the key metabolite $X_i$ as close as possible to the HS value.

Selection of target enzymes. Establishment of plausible, physiologically admissible metabolite and enzyme intervals. In a first approach to the optimization program, we will consider all of the enzymes (except $X_{37}$) as optimization targets. Thus, the set of target enzymes $\{S^{17}\}$ is:

$$S^{17} = \{X_{20}, X_{21}, \ldots, X_{46}\}$$

On the basis of previous studies (Vera et al. 2003) and the information available on the system (Curto et al. 1997, 1998a,b), lower and upper bounds were established for the values of the metabolites:

$$0.5 \cdot X^{HS}_i \leq X_i \leq 10 \cdot X^{HS}_i \quad i = 1, \ldots, 16$$

By using a lower bound of 0.5 we guarantee the existence of enough quantity to satisfy the need for metabolites of the cell in some basic processes, while a value of 10 times higher than the HS value does not seem critical in most of the metabolites. However, an improved version of the program would include specific bounds for each metabolite based on the updated knowledge of biomedical researchers.

In the model, two metabolites are considered as independent variables representing the network substrates: ribose-5-phosphate ($X_{37}$) and inorganic phosphate ($X_{18}$). $X_{37}$ can be partially regulated by controlling the diet. This coincides with the prescription of diet used by physicians to treat hyperuricemia (Sriver et al. 1989). Therefore, we considered two possible scenarios: i) No prescription of diet, in which the HS value is assigned to $X_{37}$ ($X_{37} = X^{HS}_{37}$); ii) Prescription of diet, in which an interval of feasible values for $X_{37}$ is defined at around 50% of the HS value to model the effects of diet on the uptake of this substrate ($0.5 \cdot X^{HS}_{37} \leq X_{37} \leq 1.5 \cdot X^{HS}_{37}$). The concentration of inorganic phosphate ($X_{18}$) affects several essential metabolic processes in the cell, and consequently, we keep it constant at the HS value ($X_{18} = X^{HS}_{18}$). In the following sections we discuss the results obtained with prescription of diet, which better illustrate the properties of the method. Results for treatment without prescription of diet are included in AD5.

In terms of the admissible interval of values for the enzyme activities, we only modelled treatments in which drugs act as inhibitors of the target enzymes. Therefore, an increase in the activity of enzyme targets with respect the HS value using drugs is not considered. Thus, by applying a criterion of simplicity and considering that we are only illustrating the use of the method, we chose a suitable interval for each enzyme activity ranging from a 10% of the HS value (strong inhibition) to the true HS value (no inhibition):

$$0.1 \cdot X^{HS}_i \leq X_i \leq 1.0 \cdot X^{HS}_i \quad i \in \text{TE} = \{20, \ldots, 46\}$$

In a more exhaustive case, the limits on the inhibition of each enzyme must be imposed separately using biomedical criteria.

Steady-state equations. We guaranteed that the computed solution represents a steady state for the system by introducing the set of equations that describes the steady-states of the system.

Section AD3 shows the optimization program as defined to date. In the current representation, the optimization program is a non-linear set of equations. However, if we apply the linearization strategy based on the use of an S-system model we obtain the linear optimization program used to generate the solutions (see Section AD4).

Step 3. Computation of solutions. The search process was formulated aiming in an effort to explore the simplest possible treatments to control pathological disequilibria. First, we analysed treatments that inhibit only one enzyme. In this case, we built and computed one optimization program for each possibility that included the following additional set of equations:

$$0.1 \cdot X^{HS}_i \leq X_i \leq X^{HS}_i \quad i \in \{20, \ldots, 46\}$$

where the activity of the target enzyme $X_i$ to vary while the others were fixed at the HS value. Afterwards, we explored solutions representing treatments involving the simultaneous inhibition of two different enzymes. In this case, for each possible combination of two modifiable enzymes it was necessary to compute one optimization program, which included the following additional equations:

$$0.1 \cdot X^{HS}_i \leq X_i \leq X^{HS}_i \quad 0.1 \cdot X^{HS}_j \leq X_j \leq X^{HS}_j \quad i, j \in \text{TE} = \{20, \ldots, 46\}$$

where the activity of the two targeted enzymes $X_i$ and $X_j$ could be modified, while the others remained fixed at the HS value. This involved a set of 351 optimization programs for treatments with prescription of diet and the same number in the case without prescription of diet. Although the application of this strategy to combinations of 3 or more enzymes is straightforward, we limited our exploration to the cases of one and two simultaneously modifiable target enzymes. We notice that, from a biological point of view, treatments that modulate a large number of enzymes are feasible but could increase the possibilities of adverse effects.

The computation of programs was carried out using the Matlab toolbox MetMAP (Vera and Torres 2003). The toolbox contains a set of functions that allows simple or sequential optimization using the Matlab Simplex routine included in the Matlab optimization toolbox. The calculations were computed using a Pentium IV 1.6 GHz processor with 512 Mb RAM. The computing time was in the range of milliseconds per optimization program. This means that the total computational time necessary to compute the whole set of $2^7 + 2^9 = 351$ programs was in the order of seconds, even using conventional Matlab m-files (not compiled code). All the computed solutions had the required dynamic stability (reflected in the properties of their eigenvalues) and robustness (considering the values of their logarithmic gains). With S-system models, the method scales very well with the complexity of the model (number of variables and equations) and the computational requirements of the simplex algorithm are not significant at the level of the complexity of the current and predictable mathematical models of metabolic systems.

Step 4. Choice of the most biomedically appropriate solutions. Additional biomedical criteria were defined to select the most viable and accurate solutions (Vera et al. 2003). Firstly, we imposed an upper limit on the concentration of uric acid ($X_{46}$) equal to 105% of the HS value:

$$X_{46} \leq 1.05 \cdot X^{HS}_{46}$$

The solutions that satisfy this inequality were called satisfactory solutions, and the use of this criterion is a procedure to select the most promising solutions. A second criterion was defined using the
Cartesian distance in the space of the metabolite concentrations from the considered solution \( X \) to the HS \( X^{HS} \):

\[
D(X, X^{HS}) = \sqrt{\sum_{i} (X_i - X_i^{HS})^2}
\]

We called this metabolic distance \( D(X, X^{HS}) \), which measures the deviation of the metabolite concentrations from the HS values in the considered solution. Solutions with a high value of \( D(X, X^{HS}) \) would cause metabolic imbalance (metabolites with concentrations far from the appropriate values represented by the HS) and therefore undesirable physiological side effects.

Finally, solutions with the highest values for the catalytic activity of the target enzymes \( X_{val} \) were chosen. We took into account the fact that higher enzyme activities imply the use of a lower dose of the specific (inhibitory) drug, and reduced drug doses minimise adverse side effects. Simultaneous application of these three criteria allowed identification of solutions with maximum reduction of uric acid levels, drug doses and metabolic imbalance.

### 3.2 Single target enzyme solutions

Six satisfactory solutions were obtained with prescription of diet: inhibition of amidophosphoribosyltransferase \( X_{36} \), AMP deaminase \( X_{36} \), RNases to AMP and GMP \( X_{36} \), 5’(3’)-Nucleotidase \( X_{36} \), guanine hydrolyase \( X_{36} \), and xanthine oxidase \( X_{36} \). Further details on the selected solutions can be found in Table AD2. Examination of the table shows that all the solutions have significant values of \( D(X, X^{HS}) \) and require substantial inhibition of the targeted enzyme (low values in \( X_{val} \)).

Another important observation relates to the solution involving inhibition of xanthine oxidase \( X_{36} \). This solution coincides with the current clinical treatment for hyperuricemia, which is based on the use of a combination of allopurinol and a diet low in purine precursors (see Table AD2). Moreover, our method predicts the well-known metabolic side effects of treatment with allopurinol, namely a strong increase in the concentration of xanthine and hypoxanthine \( X_{36} \). This represents an a posteriori pragmatic verification of our model predictions. In fact, the OPDD not only found the current clinical treatment of the disease but also several other possible target enzymes with potentially lower side effects.

In general terms, direct analysis of the regulatory effects of the solutions is complicated because they involve synergistic changes in several processes included in the network. However, in the case of single-target solutions, a tentative simple analysis is possible (see Figure AD5.1). In our case, the six satisfactory solutions include some intuitive solutions, such as direct inhibition of degradative fluxes that generate uric acid (xanthine oxidase -X_{44}- or guanine deaminase -X_{40}- and inhibition of the production of inosine monophosphate (amidophosphoribosyltransferase -X_{40}-), the so-called purine nucleus, which is the basic component of purine derivatives. Moreover, inhibition of the RNases \( X_{36} \) implies inhibition of the degradative fluxes for RNA, which in principle would appear not to be feasible from a biomedical point of view. Finally, two non-intuitive strategies have been detected based on inhibition of AMP deaminase \( X_{36} \) and 5’(3’)-Nucleotidase \( X_{36} \), which represent complex systemic responses of the system. The inhibition of the step directly before the production of uric acid would be a perfectly intuitive candidate to develop a drug for the treatment of hyperuricemia (case of xanthine oxidase). However, it is not easy to visualise how the inhibition of a unique enzyme six steps earlier in the pathway (case of AMP deaminase) could allow the equilibration of uric acid levels. Thus, the selected solutions combine highly intuitive biological solutions with unexpected ones.

### 3.3 Solutions with two target enzymes

In the studied case, the OPDD program yielded two sets of 351 solutions, with and without dietary restriction. Seventy-seven satisfactory solutions \( X_{36} \leq 1.05 \cdot X_{36}^{HS} \) were computed without dietary restriction and 142 remained with dietary restriction. We used three different additional processes for selection of the solutions. The aim was to make use of such a large number of promising solutions to establish different therapeutic strategies.

In the first process, we discarded all the satisfactory solutions that included any of the 6 target enzymes included in the previously analysed solutions (see section 3.2). The following equations were applied to both sets of 351 solutions generating the \( S_N \) solutions set:

\[ S_N = \{ X_{ij} | X_{ij} \neq S_i \lor X_{ij} \neq S_j \} \]

Where

\[ S_i = \{ X_{22}, X_{26}, X_{36}, X_{37}, X_{40}, X_{44} \} \]

and

\[ S_j = \{ X_{57}, X_{57}, X_{36} \} \]

In this way, alternative treatments were explored for patients in whom treatments based on the previous simpler solutions are ineffective. The only solution that satisfies these conditions was the treatment based on inhibition of GMP reductase \( X_{36} \) and the transmethylation pathway \( X_{36} \) with dietary restriction. However,
this treatment proved to be inadequate since it required strong inhibition of both enzyme targets and led to significant deviations from HS values in certain metabolites (see Table AD3.C).

In the second process, we imposed two additional restrictive conditions on both sets of satisfactory solutions using the previously defined quantities metabolic distance \( D(X, X^{\text{HS}}) \) and activity value (\( X_{\text{val}} \)):

1. The metabolic distance must have a value less than or equal to 1.2 times the minimum value of \( D(X, X^{\text{HS}}) \) in the sets of one target enzyme discussed in section 3.2:

\[
D(X, X^{\text{HS}}) \leq 1.2 \cdot \min \{D(X, X^{\text{HS}})\} \quad \Rightarrow D(X, X^{\text{HS}}) \leq 1.65
\]

Solutions with reduced metabolic distance define possible treatments with reduced metabolic imbalance and undesirable physiological side effects.

2. The enzyme activity in both modified enzymes must be at least 0.8 times the maximum computed value in the sets of one target enzyme:

\[
X_j \geq 0.8 \cdot \max_{X_j} \left[ X \right] \quad \Rightarrow X_j \geq 0.30 \quad \forall X_j \geq 0.30
\]

Selected solutions will have an enzyme activity higher than 0.3 for both modified enzymes. Solutions with high enzyme activity are chosen because they are supposed to require lower levels of drugs administered to obtain the desired results. Lower doses of drugs mean lower side effects and undesirable interactions and therefore administered to obtain the desired results. Lower doses of drugs chosen because they are supposed to require lower levels of drugs administered to obtain the desired results. Lower doses of drugs.

**Solutions with dietary restriction**

Selected solutions will have an enzyme activity higher than 0.3 for both modified enzymes. Solutions with high enzyme activity are chosen because they are supposed to require lower levels of drugs administered to obtain the desired results. Lower doses of drugs mean lower side effects and undesirable interactions and therefore administered to obtain the desired results. Lower doses of drugs.

Finally, a third approach to classify the solutions was developed. This approach was based on the ranking of satisfactory solutions in terms of a merit function that integrates both additional quality criteria. The merit function (\( MF \)) had the following formulation:

\[
MF[D(X, X^{\text{HS}})] = \min \left[ X \right] \quad \land \quad X_{\text{min}} = \max \left[ X \right]
\]

All of the magnitudes were calculated using the set of satisfactory solutions for the cases of two target enzymes with dietary restriction. We attempted to assign a similar importance to both criteria using the weights selected (1/2 and 1/2) and the proposed normalisation. In this way, a solution with the minimum metabolic distance and maximum values of activity in both inhibited enzymes will have \( MF \) equal to zero and will be ranked in the first position on the list. In contrast, a solution with the maximum metabolic distance and the minimum values for the enzyme activities will have \( MF \) equal to 1 and will be ranked in the last position. The two lists of solutions sorted according to the value of \( MF \) are included in the attached file (ranked_solutions.mat).

These lists suggest another strategy to the screening of solutions. In this case, the solutions should be tested according to the order assigned in the list, assuming that the lower the value of \( MF \), the higher the possibility of obtaining good experimental results.
An interesting result is obtained when we compare the list of solutions ranked according to $MF'$ with the previously selected solutions. To illustrate the effect of the use of different selection criteria, we analyse the cases of treatments with dietary restriction. In this case, the three first positions in the $MF$ ranking are occupied by solutions obtained in the previous selection ($MF\ (1):X_{25}-X_{36}$, $MF\ (2):X_{27}-X_{35}$, $MF\ (3):X_{32}-X_{36}$). However, the other three selected solutions ($MF\ (31):X_{27}-X_{38}$, $MF\ (39):X_{25}-X_{45}$, $MF\ (44):X_{25}-X_{49}$) do not rank in very high positions. On the other hand, positions ranked between 4-6 are occupied by previously undetected solutions ($MF\ (4):X_{27}-X_{46}$, $MF\ (5):X_{28}-X_{46}$, $MF\ (6):X_{27}-X_{49}$). These three solutions yield good results when we consider the merit function but violate the criteria considered in the previous case (the value for one enzyme activity of $X_{27}-X_{45}$ is too low, while the value for $D(X, X^{\text{val}})$ is too high in $X_{35}-X_{45}$ and $X_{32}-X_{45}$). This apparent contradiction is explained by the properties of multi-criterion selection processes. The second selection strategy is a goal-oriented strategy to select solutions, while the merit function is a reduction of a multi-criterion problem to a weighted sum. Both approaches are equally valid but the results might not be identical. This observation highlights the range of alternative solutions obtained when multiple criteria are considered (see Vera et al. 2003).

4 DISCUSSION

In this study, we analysed possible treatments involving simultaneous inhibition of one and two enzymes in the network, with and without prescription of diet. More complex treatments (inhibiting 3 or more target enzymes) could be interesting but the increasing risk of serious side effects must be considered. Some additional biomedical criteria were used to classify and select the solutions. In the simple solutions based on the inhibition of one enzyme, the OPDD lead us to three potential solutions without dietary restriction and six with dietary restriction. Analysis of these solutions in relation to the values of $D(X, X^{\text{val}})$ and $X_{\text{val}}$ facilitated further classification. The best solutions involved inhibition of adenosine phosphoribosyltransferase ($X_{23}$) (without dietary restriction) and AMP deaminase ($X_{30}$) or RNases ($X_{28}$) (with dietary restriction). This finding establishes a rational basis for experimental assays to verify the clinical potential of the proposed solutions. The method predicted a solution that coincided with the current clinical treatment of hyperuricemia (inhibition of xanthine oxidase -$X_{23}$) and the well-known adverse side effects associated with this therapy.

The method generated a large amount of solutions proposing the coordinated inhibition of two enzymes ($77$ without dietary restriction and $142$ with dietary restriction). All of these solutions derived from the previously detected single-enzyme solutions. The only exception was the treatment based on inhibition of GMP reductase ($X_{28}$) and the transmethylation pathway ($X_{30}$) with dietary restriction, which was inadequate in terms of additional biomedical criteria. This large number of solutions was reduced by imposing two additional quality criteria: limits for the values of $D(X, X^{\text{val}})$ and $X_{\text{val}}$. After this selection process the number of high-quality solutions was reduced to three without dietary restriction and six with dietary restriction (see Table A3). In addition, the sets of satisfactory solutions were sorted using a function, thereby providing another strategy through which to select and test solutions.

These different ways to classify the solutions have clearly different biomedical aims. When solutions that do not include the previously detected target enzymes are selected, potentially useful alternative therapies can be identified for use in patients who do not respond effectively to the other treatments. The strategy involving solutions belonging to set $S_5$ attempted to select the reduced set containing the most promising solutions based on additional biomedical criteria. Finally, the list generated using the merit function is fully operative when the aim is to perform a large-scale screening process in which a large number of solutions can be tested simultaneously or sequentially.

More complex objective functions can be considered in the OPDD, including for example, a trade-off between key metabolites and fluxes, metabolic distance and degree of inhibition. Several strategies are possible in this case. Although a weighted sum of sub-functions would represent the conventional approach, certain solutions could be lost as a consequence of the reduction of a multi-objective problem to a single-objective one. The fully multi-objective approach is affordable in the case of the S-systems models using the Simplex multi-objective algorithm but necessitates a complex selection process (Vera et al 2003).

The purpose of this article was not to obtain biomedically feasible solutions but to illustrate how increasing levels of biomedical information can be integrated together with mathematical models. Some additional requirements must be satisfied before any attempt to implement the solutions generated, and it is necessary at least to consider the following: a) The obtainment of an improved model integrating data from state-of-the-art experimental techniques; b) The participation of biomedical researchers in the decision-making process, specially defining the pathologic state, feasible physiological constrains and key metabolites and fluxes; and c) The experimental validation of the results in a biological model. In addition, we do not propose specific drugs to be used in the treatment of the enzymopathies like hyperuricemia, but inhibitory patterns to be satisfied by potential drugs. The necessary analysis of kinetics and detoxification, which is drug specific, is to be performed in further steps of the drug discovery proceeding.

Although the optimization of biochemical models have proved its utility in the analysis of basic properties in metabolic systems (Vera et al 2003), we consider the OPDD as a method for applied biomedicine in drug discovery instead of a tool for pure analysis like sensitivity analysis. The difference appears clear when we compare the aim of the present work with the recently published article by Vilaprinyo and collaborators (Vilaprinyo et al 2006). In that work, the aim was to clarify why some gene profiles are selected in response to heat shock and how these changes in gene expression provoked an efficient response of the system. The gene profiles preferred for the system in response to this stress were analysed according to potential physiological and functional criteria, and some of these criteria were selected as definition of the efficient response. The aim of that analysis was to detect priorities of the cell that define underlying design principles. In contrast, optimization has a mostly applied biotechnological purpose. In this vein, OPDD is to be used when the priorities and limitations of the system for a certain biomedical problem are known, and in this case the method tells us which enzymes and how much (in a qualitatively sense) must be inhibited in order to alleviate the effects of the studied disease. Consequently, the OPDD method is not to replace other successful methods in drug discovery but to be integrated in a more complex procedure together with basic experimental techniques, well established analytical methods (e.g. sensitivity analysis, Voit 2000 and Kell and Westerhoff 1986; QSAR,
Yang and Huang 2006; molecular docking, Jain 2006), and primarily experimental methods for drug screening (Carnero 2006).

The basic idea of the OPDD method (mathematical model + optimization to detect enzyme targets) can be used with any kind of metabolic model based on ODEs: Michaelis-Menten based models, power-law model (S-System or GMA) or other modelling formalisms. The computation time associated with the solution of the optimization programs included in the OPDD when one uses S-system models with an intermediate complexity (<100 variables) and simple treatments (1-2 simultaneous target enzymes) is in the order of seconds. Using other models in ODE and non-linear optimization algorithms, the total computational time will be in the range of $10^{-5}$ seconds and will imply increased hardware requirements. When we apply the method to the case of S-systems, the total computational time is not critical and efforts can be focused on refinement of the optimization program and the biological hypothesis formulated about the disease. In this case, the method allows an interactive improvement of the optimization program to be applied. Such improvements are not possible in other cases in which each round of optimization would take $10^{3}$ seconds. Moreover, when we apply the ideas included in the Indirect Optimization Method (IOM; Torres et al 1997), either the recasting of the model (Savageau and Voit 1987) or the expansion as an S-system model using a Taylor series approximation in the logarithmic space allows the use of the method in other kinds of models with its computational advantages (see ISS5-ISS6). On the other hand, in some cases the flux aggregation performed in S-system models can result in a loss of predictive capabilities of the model (De Atauri et al. 1999); in these cases, the use of the OPDD method with other non-aggregated ODEs mathematical models can be more adequate.

Finally, the proposed method was designed to analyse enzyme-pathways in which an enzyme deficiency has been unequivocally identified as the primary cause of the disease, and we currently discard the usability of the OPDD method to investigate complex multifactorial diseases (e.g. diabetes type II or obesity) in which several enzymes are simultaneously shifted in activity.

ACKNOWLEDGEMENTS

This work was supported by research grants from the Spanish Ministry of Science and Education (ref. BIO1999-0492-C02-02, BIO2005-08898-C02-02 and SAF2005-01627), by a research grant from the Government of the Canary Islands (ref. nº PI2000-071) and by Fundación La Caixa (ON03-70-0). J. Vera was the recipient of a research grant from the Spanish Ministry of Science and Education (ref. PN99-4382098).

REFERENCES


