Identification of Sensitive Enzymes in the Photosynthetic Carbon Metabolism

Citation: Umeton, Renato, Giovanni Stracquadanio, Alessio Papini, Jole Costanza, Pietro Lio, and Giuseppe Nicosia. "Identification of Sensitive Enzymes in the Photosynthetic Carbon Metabolism." Advances in Systems Biology (November 18, 2011): 441–459.

As Published: http://dx.doi.org/10.1007/978-1-4419-7210-1_26

Publisher: Springer-Verlag

Persistent URL: http://hdl.handle.net/1721.1/100961

Version: Author's final manuscript: final author's manuscript post peer review, without publisher's formatting or copy editing

Chapter 26
Identification of Sensitive Enzymes in the Photosynthetic Carbon Metabolism

Renato Umeton, Giovanni Stracquadanio, Alessio Papini, Jole Costanza, Pietro Liò, and Giuseppe Nicosia

Abstract Understanding and optimizing the CO_2 fixation process would allow human beings to address better current energy and biotechnology issues. We focused on modeling the C_3 photosynthetic Carbon metabolism pathway with the aim of identifying the minimal set of enzymes whose biotechnological alteration could allow a functional re-engineering of the pathway. To achieve this result we merged in a single powerful pipe-line Sensitivity Analysis (SA), Single- (SO) and Multi-Objective Optimization (MO), and Robustness Analysis (RA). By using our recently developed multipurpose optimization algorithms (PAO and PMO2) here we extend our work exploring a large combinatorial solution space and most importantly, here we present an important reduction of the problem search space. From the initial number of 23 enzymes we have identified 11 enzymes whose targeting

R. Umeton (✉)
Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139, USA
e-mail: umeton@mit.edu

G. Stracquadanio
The Johns Hopkins University, 217 Clark Hall, Baltimore, MD 21218, USA
e-mail: stracquadanio@jhu.edu

A. Papini
University of Florence, Via La Pira, 4, Firenze, FI I-50121, Italy
e-mail: alpapini@unifi.it

J. Costanza • G. Nicosia
University of Catania, Viale A. Doria 6, Catania, CT 95125, Italy
e-mail: costanza@dmi.unict.it; nicosia@dmi.unict.it

P. Liò
University of Cambridge, William Gates Bldg, 15 J J Thomson Avenue, Cambridge CB3 0FD, UK
e-mail: pl219@cam.ac.uk

I.I. Goryanin and A.B. Goryachev (eds.), *Advances in Systems Biology*, Advances in Experimental Medicine and Biology 736,

in the C_3 photosynthetic Carbon metabolism would provide about 90% of the overall functional optimization. Both in terms of maximal CO_2 Uptake and minimal Nitrogen consumption, these 11 sensitive enzymes are confirmed to play a key role. Finally we present a RA to confirm our findings.

1 Introduction

The Calvin Cycle (C_3 cycle) is a biochemical pathway of plants capable of fixing atmospheric inorganic CO_2 into an organic compound. This biochemical pathway is hence the basis of primary (plants) productivity. The modeling of this pathway (and allied pathway in Carbon metabolism plants) aims at optimization with respect to some specific functional targets of interest for possible biotechnological applications. Photosynthesis models showed that modifying enzyme concentration would allow the increase of the C_3-cycle efficiency, while maintaining the total amount of Nitrogen constant in the plant [1–3]. Where the biochemical pathway of the Calvin cycle is concerned, a seminal work by Zhu et al. [1], based on the Farquhar model [4], showed, with the help of an evolutionary algorithm, how enzyme concentration rearrangements could be capable of increasing the total amount of CO_2 Uptake by a factor of 76% with respect to the results obtained with the initial concentrations characteristic of the natural leaf. CO_2 Uptake rate at the natural enzyme concentration was calculated in the latter work as 15.5 μ mol m^{-2} s^{-1} in normal "air". This value is within the range of typical CO_2 Uptake rates calculated in the field for C_3 leaves [5] and can then be considered a good approximation. More recently, new efficient models showed that strategies modifying enzyme concentrations may lead to an increase in CO_2 amount of 135% with respect to the initial natural value [2, 3]. In particular, Stracquadanio et al. [2] used also the concepts of robustness and sensitivity for assessing the evaluation of confidence limits in the results obtained by perturbing the new identified solutions; this simulates typical "in-vitro" implementation variables (refer to Sensitivity and Robustness in Sect. 2). An important question regarding this re-optimization is: *why did the evolution process and not optimize enzyme concentration in order to maximize* CO_2 *fixation*? One hypothesis to answer this non-trivial question is that the Calvin cycle pathway evolved during a time in which CO_2 atmospheric concentration was much lower compared to current values. Additionally, some of the enzymes (such those belonging to the photorespiration pathway) whose reduction in concentration would result in a theoretical increase in photosynthesis efficiency, might be strongly linked to other biological functions (e.g., photosystem protection and Nitrogen assumption [6, 7]).

In this paper, we improve and advance the work started in [2] by defining and exploiting an investigation pipe-line that composes Sensitivity Analysis (SA), Single- and Multi-objective optimization, and Robustness Analysis (RA) to move toward the study of the artificial photosynthesis. More in detail, these techniques are composed into the following pipe-line: beginning with a (1) system of ODEs

(refer to coming paragraph) to have a computational model of C_3 photosynthetic pathway, (2) we exploit SA to identify which are the sensitive tuning gears of the system, then (3) we exploit SO and MO optimization to re-optimize the pathway in a functional fashion; (4) we adopt RA to have a quantitative prediction about the stability of the lately found optimizations. Each step is iterated a number of times, until a reasonable solution stability and then experimental feasibility is achieved. Then (5) we compare the newly optimized solutions with natural values to assess solution key changes and to possibly read new insights in the pathway mechanisms. The paper is structured as follows. Section 2 details all of the methods adopted in our design workflow. Section 3 presents the results obtained exploiting these methods. In Sect. 4 conclusions and future directions are presented.

2 Methods

As mentioned above, the computational simulation of the C_3 Carbon metabolism requires the definition of a set of ODEs; in our research work, it is considered the model proposed by [1]. The model takes into account rate equations for each discrete step in photosynthetic metabolism, check-point equations for conserved quantities (e.g., total leaf Nitrogen) and a set of ODEs to describe each pathway mechanism: from initial concentration of nutrients of a cell of a leaf, toward enzyme-mediated reactions, and having a consequent CO_2 uptake. The model assumes that the total protein–nitrogen in the enzymes is $1\,\mathrm{g\,m^{-2}}$; the mass nitrogen in each enzyme, in a $1\,\mathrm{m^2}$ leaf area, is computed on the basis of the number of active sites, catalytic rate per active site, molecular mass of each enzyme, and the ratios between V_m of different enzymes. Mole of each protein is then calculated based on the molecular mass and the mass of each protein, i.e., the total concentration of the adenylate nucleotides ([CA]) in the chloroplast stroma (i.e., the sum of [ATP] and [ADP]) is assumed to remain constant. The V_m for each enzyme is then calculated based on the amount of each enzyme and the volume of the compartment that it occupies in $1\,\mathrm{m^2}$ leaf area. The total concentration of the adenylate nucleotides ([CA]) in the chloroplast stroma, the sum of [ATP] and [ADP], are assumed to remain constant. Similarly, the sum of [NADPH] and [NADP] in the chloroplast stroma ([CN]) are assumed constant. The export of PGA, GAP, or DHAP from the chloroplast to the cytosol is associated with a counterimport of the phosphate, mediated by a phosphate translocator. Consequently, the total concentration of phosphate in the stroma ([CP]) is assumed constant. Finally, a set of ODEs encodes the rates of changes in concentration for each metabolite; the latter is represented by the difference between the rates of those reactions generating the metabolite and the rate of the reactions consuming it. It is clear that the volume of the chloroplast stroma can be different from the cytosol one in a typical higher plant cell; in this scenario, it has been assumed a 1:1 ratio in the computation of concentrations within these two compartments.

2.1 Pathway Sensitivity Analysis

Morris method [8] has been adopted for the evaluation of sensitive components in the set of ODEs mentioned. The main idea behind SA is the identification of crucial pathway gears, whose tuning results in a major system response. More in detail, the SA here adopted belongs to the class of the *one-factor-a-time* (OAT) methods [9], aiming at the evaluation of pathway sensitivity by means of a series of stimuli in a way such that only one input is perturbed while all of the others are kept at their nominal value. Considering for the moment our pathway as a black-box with certain inputs and certain outputs, each SA step-variation, computed for each input is calculated as: $u_i = (P(x_1, x_2, \ldots, x_i + \Delta x_i, \ldots, x_k) - P(x_1, x_2, \ldots, x_i, \ldots, x_k))/\Delta x_i$, where P is the result computed from the pathway model with input x; in particular $x_1, x_2, \ldots, x_i + \Delta x_i, \ldots, x_k$ is the perturbed input vector and $x_1, x_2, \ldots, x_i, \ldots, x_k$ is the nominal input vector. For each factor, a group of outcomes u_i is collected and, as metrics for sensitivity, the mean ${\mu_i}^*$ and the standard deviation σ_i are computed. Highly linear behaviors should be expected from those inputs with a high value of ${\mu_i}^*$. A completely different behavior, such as highly non-linear or counterintuitive responses should be expected from those inputs with high σ_i values. For each enzyme (i.e., input) we use the five concentrations under consideration (refer to Table 26.1 in Sect. 3) as nominal values, computing 20 different factor levels, altered for 10 times each one. As bounds of this SA we adopt $\pm 100\%$ of the nominal value, for each input enzyme concentration. The result of this analysis, highlighted how there are *eleven* enzymes that have to be considered extremely sensitive when compared to all of the others [2]. These enzymes are: *Rubisco, PGA kinase, GAP dehydrogenase, FBP aldolase, FBPase, SBP aldolase, SBPase, Phosphoribulose kinase, ADPGPP, Phosphoglycolate phosphatase, and GDC*. These enzymes showed indeed high values of σ_i (i.e., $1 < \sigma_i < 15$), when compared to all of the others (i.e., $10^{-4} < \sigma_i < 1$).

The definition of a set of linked ODEs gives a mathematical description of the chemical process and, successively, the Morris analysis gives useful insights on linear and non-linear contribution of enzymes to the Carbon metabolism. However, it is important to validate these results by taking into account the interaction map defined by the pathway; it is plausible to assume that sensitive enzymes should be hubs of the photosynthesis pathway. This information has been obtained using Rosvall community detection method [10]; the interaction map we gained confirms these assumptions. Figure 26.1 shows how Rubisco and GAP dehydrogenase are the most strongly regulated enzymes of the Calvin Cycle. Both enzymes are light regulated. Transketolase is another key enzyme, since it uses as substrates Fructose-6-P (otherwise destined to exit from the cycle toward the starch biosynthetic pathway) and 3-P-Glyceraldehyde, that is produced by the enzyme GAP dehydrogenase itself. These enzymes correspond to the main nodes of the Calvin Cycle leading to the other biosynthetic pathways. Phosphoglycolate phosphatase is the first enzyme of the photorespiration pattern linked to the Oxygenase activity of Rubisco and Glycine

Table 26.1 Concentrations of the enzymes, individual robustness, CO_2 uptake rate (at $c_i = 270\,\mu\,\mathrm{mol\,mol}^{-1}$, reflecting current CO_2 atmospheric concentration), global and local robustness values. The first enzyme value column reports touchstone concentrations used in our simulations: the initial/natural leaf. The second value column reports the result of the optimization in which only the eleven sensitive enzymes are altered, while all of the others are kept at their nominal values. Third column reports the best-known leaf design, in terms of CO_2 uptake and robustness. The fourth column reports the result of a simulation, where the enzymes Cytosolic FBP aldolase, Cytosolic FBPase, UDP-Glc pyrophosphorylase have been maintained to their initial values. Last column reports the most efficient known point in terms of CO_2, but corresponds to a highly instable solution

Enzyme name	Initial concentration mg N m^{-1} (the natural leaf)	Optimal concentration of 11 sensitive enzymes mg N m^{-1}	Optimal and robust concentration mg N m^{-1}	Optimal and robust concentration mg N m^{-1} (3 fixed enzymes)	Optimal but *not* robust concentration mg N m^{-1}
Rubisco	517.00 (100)	784.27 (**84**.5)	860.226 (100)	856.44 (100)	861.93 (**39**)
PGA kinase	12.20 (100)	4.66 (100)	3.989 (100)	3.63 (100)	3.98 (**0**)
GAP DH	68.80 (100)	69.03 (**81**.5)	64.483 (100)	65.08 (100)	63.55 (**17**)
FBP aldolase	6.42 (100)	10.40 (100)	9.050 (100)	10.86 (100)	9.29 (**30**.**5**)
FBPase	25.50 (100)	29.44 (100)	26.889 (100)	32.24 (100)	27.03 (**0**)
Transketolase	34.90 (100)	*34*.90 (100)	8.247 (100)	16.93 (100)	16.98 (*100*)
SBP aldolase	6.21 (100)	5.55 (100)	6.661 (100)	5.75 (100)	5.94 (**0**)
SBPase	1.29 (100)	4.70 (100)	4.397 (100)	4.43 (100)	4.31 (**1**)
PRK	7.64 (100)	7.04 (100)	7.007 (100)	6.38 (100)	7.99 (**22**.**5**)
ADPGPP	0.49 (100)	2.12 (100)	0.721 (100)	5.09 (100)	1.22 (**0**)
PGCA Pase	85.20 (100)	0.95 (100)	0.325 (100)	0.20 (100)	0.00 (**0**)
Glycerate kinase	6.36 (100)	*6*.36 (100)	0.005 (100)	0.00 (100)	0.00 (*100*)
Glycolate oxidase	4.77 (100)	*4*.77 (100)	0.019 (100)	0.16 (100)	0.00 (*100*)
GSAT	17.30 (100)	*17*.30 (100)	0.027 (100)	0.00 (100)	0.00 (*100*)
Glycer. dehyd.	2.64 (100)	*2*.64 (100)	0.003 (100)	0.00 (100)	0.00 (*100*)
GGAT	21.80 (100)	*21*.80 (100)	0.00005 (100)	0.00 (100)	0.00 (*100*)
GDC	179.00 (100)	0.02 (100)	0.00003 (100)	0.00 (100)	0.00 (*100*)
Cyt. FBP ald.	0.57 (100)	*0*.57 (100)	2.127 (100)	*0*.57 (100)	2.03 (**0**.5)

(continued)

Table 26.1 (continued)

Enzyme name	Initial concentration mg N m^{-1} (the natural leaf)	Optimal concentration of 11 sensitive enzymes mg N m^{-1}	Optimal and robust concentration mg N m^{-1}	Optimal and robust concentration mg N m^{-1} (3 fixed enzymes)	Optimal but *not* robust concentration mg N m^{-1}
Cyt. FBPase	2.24 (100)	*2*.24 (100)	5.554 (100)	*2*.24 (100)	5.27 (**30**.5)
UDPGPP	0.07 (100)	*0*.07 (100)	0.531 (100)	*0*.07 (100)	0.50 (**0**)
SPS	0.20 (100)	*0*.20 (100)	0.034 (100)	0.01 (100)	0.03 (**30**.5)
SPP	0.13 (100)	*0*.13 (100)	0.031 (100)	0.01 (100)	0.03 (**0**)
F26BPase	0.02 (100)	*0*.02 (100)	0.00 (100)	0.00 (100)	0.00 (*100*)
CO_2 uptake $\frac{\mu \text{ mol}}{\text{m}^2 \text{ s}}$	15.486	22.420	20.626	22.156	
(Local R. %, Global R. %)	(<u>100</u>, 81.80)	(81.5, 78.3)	(<u>100</u>, 97.2)	(<u>100</u>, 92.6)	(0, 39.18)

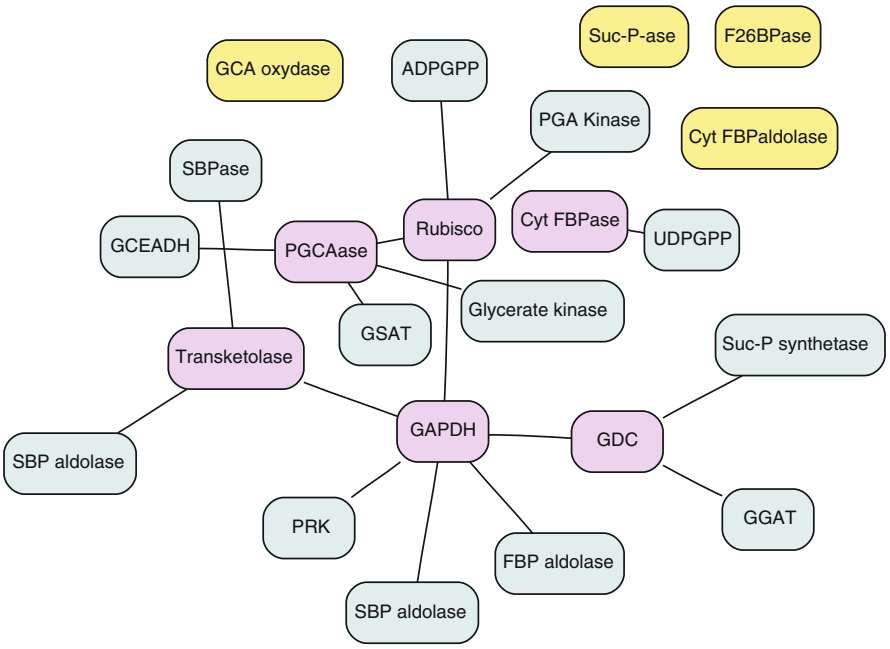

Fig. 26.1 Relationships within the Pathway. Enzyme interactions in the C_3 photosynthetic Carbon metabolism

decarboxylase the enzyme responsible for the loss of the CO_2 fixed by Rubisco. Figure 26.1 presents also 10 enzyme clusters in the C_3 photosynthetic Carbon metabolism pathway:

(1) Phosphoglycolate phosphatase, Glycerate kinase, Ser glyoxylate aminotransferase, Glycerate dehydrogenase; (2) Transketolase, SBPase, SBP aldolase, Enzyme 9; (3) GAP dehydrogenase, GAP dehydrogenase, FBPase, FBP aldolase, Phosphoribulose kinase; (4) Rubisco, PGA Kinase, Enzyme11, ADPGPP; (5) GDC, Glu glyoxylate aminotransferase, Suc-P synthetase; (6) Cytosolic FBPase, UDP-Glc pyrophosphorylase; (7) F26BPase; (8) Suc-P phosphatase; (9) Cytosolic FBP aldolase; (10) Glycolate oxidase.

2.2 Optimization of Sensitive Enzymes: Single- and Multi-Objective Optimization

Once these eleven enzymes have been identified as *sensitive*, their optimization has been then evaluated and compared to the optimization of the complete pathway system. This means inspecting a search space in 11 dimensions (sensitive enzymes), instead of inspecting the complete pathway one, that has 23 dimensions. The aim of this is evidently the evaluation of the contribution of these *sensitive enzymes*

in the optimization of the photosynthetic metabolism considered. To achieve this aim we have employed Single- and Multi-objective optimization algorithms; these approaches have considered both the complete domain space ($x \in \mathbb{R}^{23}$) and the "sensitive domain", as cropped by those 11 most sensitive enzymes ($x \in \mathbb{R}^{11}$). Considering a fixed total amount of protein Nitrogen available to the leaf ($1\,\mathrm{g\,m^{-2}}$, i.e., ca. 20.833×10^4 $\mathrm{mg\,l^{-1}}$), we have used Parallel Optimization Algorithm (PAO) [2] to let a pool of solutions evolve in an archipelago fashion. Fixing at their natural value all but sensitive enzymes, PAO has evaluated a number of new enzyme concentration profiles and has associated a CO_2 Uptake to each one of them by computing the system of ODEs mentioned above. The key aspect of PAO is its ability to share *portions* of promising solutions among optimization cores. We adopted two optimization cores (islands): on one PAO runs A-CMA-ES [2] algorithm and on the other, it runs DE [11]; solution portions are exchanged every 200 generations with probability 1/2. On both islands the optimization aim is the same: to find all those sensitive enzyme concentration vectors $\hat{x} = [\mathrm{conc}_1, \mathrm{conc}_2, \ldots, \mathrm{conc}_{11}]$, such that, when $\hat{x}$ is composed with the other enzyme values kept at their nominal value, the resulting CO_2 Uptake function is maximized:

$$\max_{\hat{x} \in \mathbb{R}^{11}} \left(f_1 \ (\hat{x}, x_{\text{non-sensitive}}) \right). \tag{26.1}$$

Relaxing the constraint about the fixed total amount of protein Nitrogen, a new optimization has been performed. The focus here is again on those *sensitive enzymes* and on the comparative evaluation with respect to the rest of the pathway. Parallel Multi-objective Optimization (PMO2) [12] has been adopted to evaluate the contextual optimization of CO_2 Uptake and total Nitrogen needed. Gaining higher CO_2 Uptake rates employing less Nitrogen mean absorbing more CO_2, while consuming less "leaf-fuel", this means, a more efficient metabolism cycle. This means that additionally to the maximization of the CO_2 Uptake function, a new function is taken into account; it is the minimization:

$$\min_{\hat{x} \in \mathbb{R}^{11}} \left(f_2 \ (\hat{x}) \right) = \min_{\hat{x} \in \mathbb{R}^{11}} \left(\sum_{i=1}^{11} \frac{\hat{x}[i] \times \mathrm{WM}_i}{\mathrm{BK}_i} \right), \tag{26.2}$$

where BK_i is the catalytic number or turnover number, and WM_i the molecular weight of the ith enzyme, respectively. Hence, our search for $\hat{x}$ has to accomplish a contextual trade-off between maximal CO_2 Uptake rate and minimal Nitrogen employment. Note that the minimization of f_2 does not take into account those enzymes that are not sensitive: since they are fixed in our search for minima, the second objective is simply shifted by a constant quantity for all of the points evaluated; this obviously does not impact the optimization as all of the point efficiencies are translated as a whole. In PMO2, at the beginning, a random initial

population P_0 is generated and it is sorted based on the non-domination criterion; a fitness proportional to its non-domination level is assigned to each solution. Non-dominated sorting has been introduced in order to rank the population according to its domination level. For each solution, we compute the domination count n_p, that denotes the number of solutions dominated by p, and S_p, that is, the set of solutions dominated by p; obviously, all the solutions belonging to the first front have a domination count set to zero. For each solution with $n_p = 0$, we pick each element of S_p and reduce its domination count by one; if for any member of S_p the domination count goes to zero, it is put into a separate list Q. The process is iterated until each solution is assigned to a front. The algorithm proceeds using binary tournament selection, recombination and mutation to create a population of offspring Q of size N. At each generation g, a population $R_g = P_g \cup Q_g$ is built and, hence, it is sorted according to non-domination; it is important to note that since the parent population is put in R_g, elitism is assured. The selection procedure chooses the individual with $n_p = 0$ and, then, it picks individuals from other domination levels if there are not N non-dominated individuals; this set of dominated solutions is chosen according to a *crowding-comparison* operator.

2.3 Leaf Candidate Robustness

Once single-objective optimization algorithms have found the best solutions to the f_1 maximization problem, we adopt the RA to assess the intrinsic stability of the solution. The definition of robustness here adopted has to be considered as the ability of a system to survive random perturbations [13]. In order to evaluate the robustness of enzymes partitions, the *robustness condition*, ρ, and the *uptake yield*, Γ, have been defined [12]. Let $\bar{x} \in \mathbb{R}^{23}$ an enzyme partitioning and $f : \mathbb{R}^n \to \mathbb{R}$ a function computing the expected CO_2 uptake rate value of $\bar{x}$. Given an enzyme partition $\bar{x}_*$ obtained by perturbing $\bar{x}$, the robustness condition ρ is defined as follows:

$$\rho(\bar{x}, \bar{x}_*, f, \epsilon) = \begin{cases} 1 \text{ if } \mid f(\bar{x}) - f(\bar{x}_*) \mid \leq \epsilon \\ 0 \qquad\qquad \text{otherwise} \end{cases}, \tag{26.3}$$

where the robustness threshold ϵ denotes the maximum percentage of variation from the nominal CO_2 uptake value.

Given an ensemble $\mathcal{T}$ of perturbed enzymatic concentrations obtained by perturbing $\bar{x}$, the uptake yield Γ is defined as follows:

$$\Gamma(\bar{x}, f, \epsilon) = \frac{\sum_{\tau \in \mathcal{T}} \rho(\bar{x}, \tau, f, \epsilon)}{|\mathcal{T}|}. \tag{26.4}$$

The ensemble T has been generated using a Monte-Carlo algorithm; mutations occurring on all the enzymes (global RA) and one enzyme at time (local RA) have

been considered [13]. A maximum perturbation of 10% has been fixed for each enzyme concentration, and then it has been generated an ensemble of 5×10^3 trials for the global RA and 200 trials, for each enzyme, for the local RA. All the experiments assume $\epsilon = 5\%$ of the nominal uptake rate value.

3 Results

3.1 Sensitive Enzymes for the Uptake Objective

The aim of the present research work is to compare how the exclusive targeting of sensitive enzymes varies either when f_1 (26.1) is maximized or when f_1 is maximized while f_2 (26.2) is minimized. As mentioned above, we want to evaluate if it is really worth moving from the original search space ($x \in \mathbb{R}^{23}$) to the sensitive enzymes search space ($x \in \mathbb{R}^{11}$), without important losses in terms of functional pathway optimization.

In order to suggest correct and minimal biotechnological targets, we present in Table 26.1, four alternative leaf designs, unraveled by PAO algorithm: these solutions represent candidate enzyme concentration whose task is the increase of the CO_2 Uptake rate, while maintaining the actual amount of total Nitrogen contained in the enzymes.

Best solutions obtained on the optimization of *sensitive enzymes* both by PAO (max f_1) and PMO2 (max $f_1 \cap min\ f_2$) have been further inspected and compared to the natural leaf. Figure 26.2 shows how sensitive enzymes changed their concentration in order to maximize the CO_2 Uptake (i.e., the best point found by "PAO 11 Sens" and the end of its convergence, Fig. 26.3). Exclusive targeting of sensitive enzymes brought an optimal uptake rate of $33.317\,\mu\,\mathrm{mol\,m^{-2}s^{-1}}$, that is, only -9% than the most efficient known point; this confirms how these 11 enzymes perform about 91% of the whole photosynthetic optimization. It is also worth noting in Fig. 26.2 histogram how all of the increases and decreases in optimal enzyme concentration are within the range $\times 0.001 - \times 4.3$; this is a plausible biotechnological range, indeed, around a fivefold increase can be achieved by means of enzyme promoters. Afterward, RA has been employed to assess the stability of this solution as well: as reported in Table 26.1, this point has an overall local robustness that is not very high (81.5%) and a global robustness comparable to the natural one (78.3%).

To evaluate the effective contribution of these 11 sensitive enzymes, we have fed into PAO the optimization problem in which the variable enzyme set is extended from 11 to all but those 3 that seemed to play an important role looking at single histograms but did not play an important role according to the SA. These three enzymes are: Cytosolic FBP aldolase, Cytosolic FBPase, and UDPGP. The result of this optimization is reported in Fig. 26.4 (refer to Table 26.1 for single enzyme variation). This configuration registered a CO_2 Uptake rate of $36.197\,\mu\,\mathrm{mol\,m^{-2}\,s^{-1}}$,

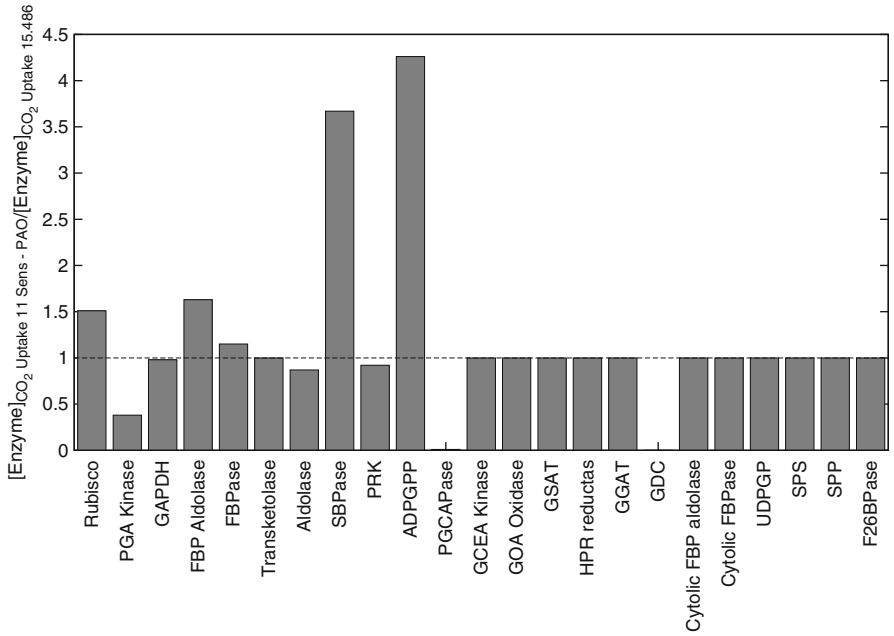

Fig. 26.2 The ratio of the enzyme concentrations optimized by the PAO algorithm ($33.317\,\mu\,\mathrm{mol\,m^{-2}\,s^{-1}}$) at a $c_i = 270\,\mu\,\mathrm{mol\,mol^{-1}}$ compared to the initial concentrations ($15.486\,\mu\,\mathrm{mol\,m^{-2}\,s^{-1}}$). Optimization of CO_2 uptake rate perturbing the 11 most sensitive enzymes only (*Rubisco, PGA kinase, GAP dehydrogenase, FBP aldolase, FBPase, SBP aldolase, SBPase, Phosphoribulose kinase, ADPGPP, Phosphoglycolate phosphatase, and GDC*). These enzymes are the most important enzymes in the studied model of the Carbon metabolism) while the remaining enzymes are maintained at their initial concentration

that is, ca. $+8\%$ when compared to the optimization of sensitive enzymes. RA has reported 100% and 92.6% for local and global robustness, respectively, proving the stability of this solution.

We have then decided to target four enzymes of the C_3 metabolic pathway: we all know biotechnological intervention is hard and error prone, then we want to minimize intervention points. This decision came from combining the information gained on different points of our research: (1) promising leaf engineering obtained with the alteration of the 11 most sensitive enzymes, (2) those three enzymes that could have played a crucial role do not seem to affect the Uptake objective, (3) out of those 11 enzymes pointed out by the SA only *six* of them present a change out of the range $0.2\times - 1.5\times$ (Rubisco, FBP aldolase, SBPase, ADPGPP, Phosphoglycolate phosphatase, and GDC). Because of the non-optimal local robustness showed (refer to Table 26.1, 84.5%), Rubisco has been filtered from the analysis to ensure a fair comparison and a more precise identification of robust targets for biotechnological intervention. These five enzymes have been sorted out into three simulations; we have designed each simulation such that there are two enzymes that showed a

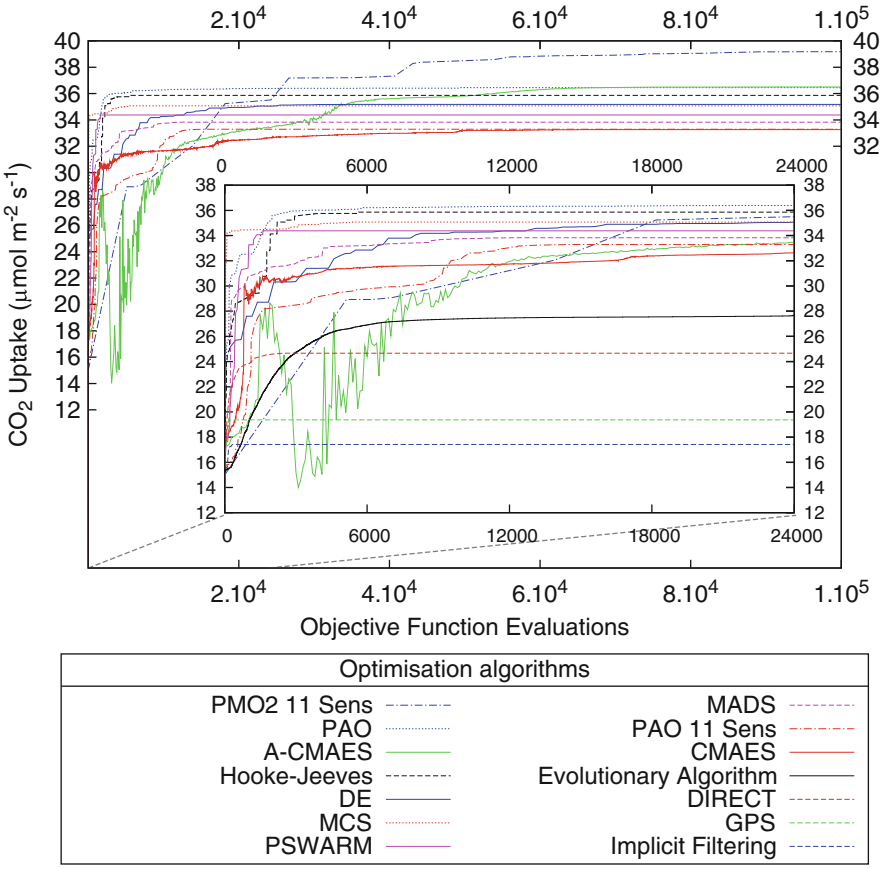

Fig. 26.3 Convergence process of fourteen derivative-free global optimization algorithms; on single objective, the PAO algorithm outperforms all of the other algorithms (from best to worst: Hooke–Jeeves [14], DE [11], MCS [15], PSWARM [16], MADS [17], CMAES [18], Evolutionary Algorithm in [1], DIRECT [19], GPS [20], and Implicit Filtering [21]) in the single-objective optimization of the full problem version (i.e., optimization of all of the enzymes) and then it has been adopted for the optimization of the reduced model which has optimized using the most sensitive enzymes (in the legend "PAO 11 Sens"). This optimization comparison has been performed to maximize light-saturated photosynthetic rate (CO_2 Uptake) at $c_i = 270\,\mu\,\mathrm{mol\,mol}^{-1}$, that is, the value characteristic of nowadays CO_2 atmospheric concentration. It is also reported the convergence of the PMO2 algorithm, in terms of non-dominated solutions, when the optimization enzyme set is restricted to the eleven sensitive ones ("PMO2 11 Sens" in the legend)

negative fold-change and other two with a positive one; this has been put into place to ensure a balance with respect to the total Nitrogen partitioning. Indeed, having a fixed amount of protein Nitrogen, it is likely that to allow ADPGPP to grow as in Fig. 26.2, we have to couple it with some of those enzymes that diminished their concentrations (i.e., Phosphoglycolate phosphatase, and GDC). Summarizing, we have further inspected our pathway through three more simulation configurations:

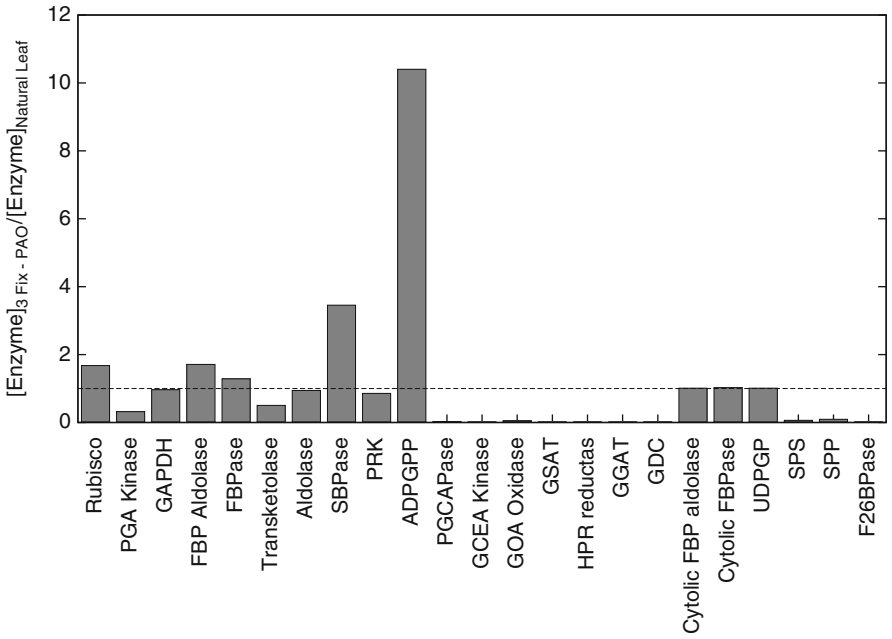

Fig. 26.4 Changes in the concentrations of Carbon metabolism enzymes with respect to their natural values when three metabolites are kept constant: Cytosolic FBP aldolase, Cytosolic FBPase, and UDPGP. The C_i has value 270 μ mol mol^{-1}, reflecting nowadays condition. This configuration obtains CO_2 Uptake rate of 36.197 μ mol m^{-2} s^{-1}, suggesting that even fixing these three enzymes the uptake performance can be very effective

Var4-1 (that tunes only FBP aldolase, SBPase, Phosphoglycolate phosphatase, and GDC), *Var4-2* (FBP aldolase, ADPGPP, Phosphoglycolate phosphatase, and GDC) and *Var4-3* (that targets only ADPGPP, SBPase, Phosphoglycolate phosphatase, and GDC). Table 26.2 presents these results from a quantitative point of view. It is of note how when the optimization is pushed selectively to the limit varying only 4 enzymes (i.e., *Var4* simulations), we observe enzyme concentrations that readjust their values within the range $0.001\times - 160\times$. Such a step variation has to be considered a strong signal, as targeting such a small set of enzymes we are unraveling how the metabolism can become functional (uptake maximization) without invalidating any other connected pathway.

3.2 Sensitive Enzymes on Uptake Maximization and Nitrogen Minimization

In order to compare the optimization of the whole system with the optimization of the sensitive enzymes, we have compared the Pareto frontiers obtained on

Table 26.2 Concentrations of the enzymes, individual robustness, CO_2 uptake rate (at $c_i = 270\,\mu$ mol mol^{-1}, reflecting current CO_2 atmospheric concentration), global and local robustness values. The first enzyme value column reports touchstone concentrations used in our simulations: the initial/natural leaf. Columns 3–5 present enzyme values obtained as result of Var4-1, Var4-2 and Var4-3 simulations and robustness values associated with each leaf engineering

Enzyme name	Initial concentration mg N m^{-1} (the natural leaf)	Optimal concentration of Var4-1 sensitive enzymes mg N m^{-1}	Optimal concentration of Var4-2 sensitive enzymes mg N m^{-1}	Optimal concentration of Var4-3 sensitive enzymes mg N m^{-1}
Rubisco	517.00 (100)	*517.00* **(99.5)**	*517.00* (100)	*517.00* **(98.5)**
PGA kinase	12.20 (100)	*12.20* (100)	*12.20* (100)	*12.20* (100)
GAP DH	68.80 (100)	*68.80* (100)	*68.80* (100)	*68.80* (100)
FBP aldolase	6.42 (100)	14.76 (100)	10.40 (100)	*6.42* (100)
FBPase	25.50 (100)	*25.50* (100)	*25.50* (100)	*25.50* (100)
Transketolase	34.90 (100)	*34.90* (100)	*34.90* (100)	*34.90* (100)
SBP aldolase	6.21 (100)	*6.21* (100)	*6.21* (100)	*6.21* (100)
SBPase	1.29 (100)	198.05 (100)	*1.29* **(70)**	91.13 (100)
PRK	7.64 (100)	*7.64* (100)	*7.64* (100)	*7.64* (100)
ADPGPP	0.49 (100)	*0.49* (100)	43.52 (100)	46.52 (100)
PGCA Pase	85.20 (100)	63.30 (100)	220.93 (100)	131.79 (100)
Glycerate kinase	6.36 (100)	*6.36* (100)	*6.36* (100)	*6.36* (100)
Glycolate oxidase	4.77 (100)	*4.77* (100)	*4.77* (100)	*4.77* (100)
GSAT	17.30 (100)	*17.30* (100)	*17.30* (100)	*17.30* (100)
Glycer. dehyd.	2.64 (100)	*2.64* (100)	*2.64* (100)	*2.64* (100)
GGAT	21.80 (100)	*21.80* (100)	*21.80* (100)	*21.80* (100)
GDC	179.00 (100)	0.02 (100)	0.49 (100)	22.19 (100)

Cyt. FBP ald.	0.57 (100)	*0.57* (100)	*0.57* (100)	*0.57* (100)
Cyt. FBPase	2.24 (100)	*2.24* (100)	*2.24* (100)	*2.24* (100)
UDPGPP	0.07 (100)	*0.07* (100)	*0.07* (100)	*0.07* (100)
SPS	0.20 (100)	*0.20* (100)	*0.20* (100)	*0.20* (100)
SPP	0.13 (100)	*0.13* (100)	*0.13* (100)	*0.13* (100)
F26BPase	0.02 (100)	*0.02* (100)	*0.02* (100)	*0.02* (100)
CO_2 uptake $\frac{\mu \text{mol}}{\text{m}^2\,\text{s}}$	15.486	22.420	20.626	22.156
(Local R. %, Global R. %)	(100, 81.80)	(99.5, 91.8)	(70, 69.4)	(98.5, 92.9)

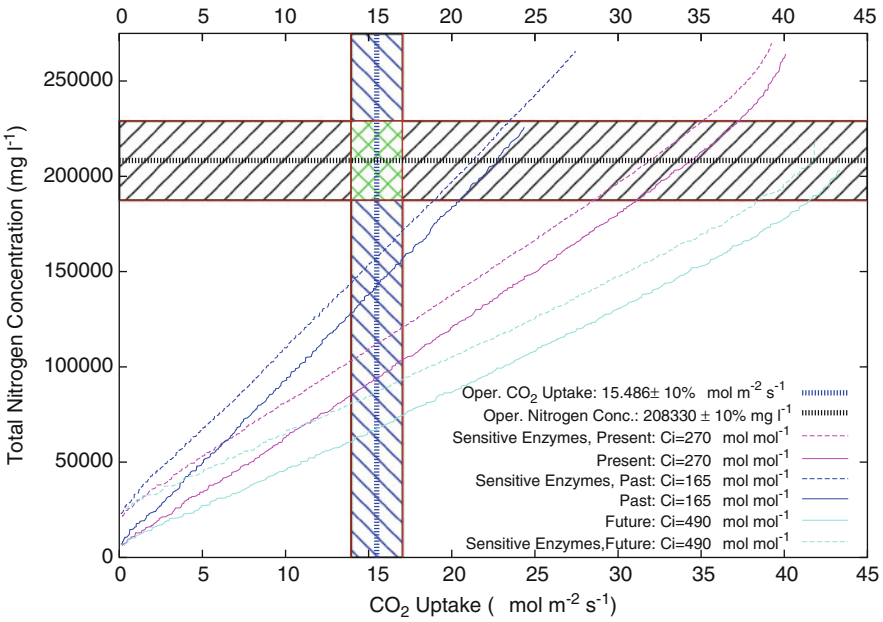

Fig. 26.5 CO_2 uptake and protein–nitrogen concentration trade-off. Maximizing the CO_2 Uptake while minimizing the total amount of protein–nitrogen concentration; the operative area of natural leaves is located in the green checked area. The label "Sensitive Enzymes" indicates the multi-objective optimization using the 11 most sensitive enzymes of the model, the three resulting Pareto Fronts have been dominated by the multi-objective optimization over all enzymes of the model. This trade-off search has been carried out for the three c_i concentration referring to the environmental condition in effect 25 million years ago, nowadays and in 2100 AC

both tasks by PMO2 in three conditions: Present, Past, and Future atmospheric conditions, i.e., $c_i \in \{270, 165, 490\}\ \mu$ mol mol^{-1}. Figure 26.5 presents these Pareto frontiers comparison: in this multi-objective optimization, as we saw in the single-objective one, the optimization of only the sensitive enzymes causes a minor loss in optimization performances. It is interesting, how reducing the search space to less than half of the dimensions (i.e., sensitive optimization), the performances are affected by a factor between 5% and 10%. It is also of note how this difference is consistently kept among all of the atmospheric conditions considered. Having a narrower search space means on one hand the achievement of sub-optimal solutions, but on the other hand, it means that during the biotechnological implementation we will have just half of the variables, compared to the original problem. Functional optimization, versus problem dimensionality, is an intrinsic trade-off that shows how we have to accept slightly lower efficiencies if we want the benefits of dealing with half of the unknowns.

Figure 26.6 shows the comparison between the natural leaf and the best non-dominated solution found by PMO2. This non-dominated solution (i.e., it belongs

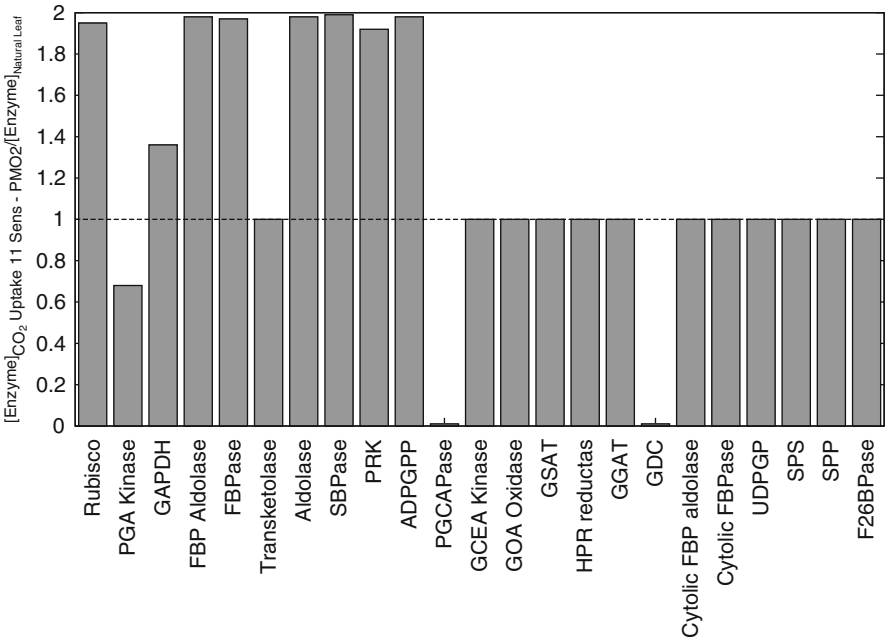

Fig. 26.6 Optimization of CO_2 uptake rate and Nitrogen consumption perturbing the 11 most sensitive enzymes only. The ratio of the enzyme concentrations optimized by the multi-objective optimization algorithm PMO2 ($39.242\,\mu\,\mathrm{mol\,m^{-2}\,s^{-1}}$) at a $c_i = 270\,\mu\,\mathrm{mol\,mol^{-1}}$ compared to the initial concentrations ($15.486\,\mu\,\mathrm{mol\,m^{-2}\,s^{-1}}$). The non-dominated solution here considered is the one with the highest CO_2 uptake rate, which shows a Nitrogen consumption of ca. $269658\,\mathrm{mg\,l^{-1}}$

to the Pareto front, and more in detail, it is the best point found by PMO2 11 Sens and the end of its convergence, Fig. 26.3) is the one with the overall maximal CO_2 Uptake rate ($39.242\,\mu\,\mathrm{mol\,m^{-2}\,s^{-1}}$). It is remarkable how, despite the tremendous increase in uptake rate of ca. 253%, and the relatively high increase in Nitrogen consumption (129%), all of the changes at the enzyme level are within the range $\times 0.01 - \times 2$. From a theoretical point of view, these changes are even easier to implement with the current chemical processing, when compared to the one reported in Fig. 26.2.

4 Conclusion

The statistician George Box said: "All models are wrong, but some are useful". Nowadays this sentence would reflect many things: the continuous improvement of developing new models in all scientific fields, the different level of

abstractions that a model could express and our difficulties in modeling multi-scale or compartmentalized dynamical systems. In conclusion, we are delighted to report that the modeling of C_3 carbon metabolism is a thriving field of research. It has two immediate and important benefits: the improved understanding of the process that shapes photosynthesis in plants and the possibility to test engineered solutions in silico using a mature single- and multi-objective optimization methodology. We believe that our quantitative findings of a small number of enzymes that concentrate the biotechnology potentialities and our methodological improvements could effectively represent a significant contribution to the community working in this area.

References

1. Zhu XG, de Sturler E, Long SP (2007) Optimizing the distribution of resources between enzymes of carbon metabolism can dramatically increase photosynthetic rate: A numerical simulation using an evolutionary algorithm. Plant Physiol 145:513–526
2. Stracquadanio G, Umeton R, Papini A, Liò P, Nicosia, G (2010) Analysis and optimization of C3 photosynthetic carbon metabolism. In: Rigoutsos I, Floudas CA (eds) Proc BIBE 2010, 10th IEEE Int Conf Bioinformatics and Bioengineering, May 31–June 3, 2010, Philadelphia, PA, USA, IEEE Computer Society, pp 44–51
3. Papini A, Nicosia G, Stracquadanio G, Lio P, Umeton R (2010) Key Enzymes for the optimization of CO_2 uptake and nitrogen consumption in the C3 photosynthetic carbon metabolism. J Biotechnol 150:525–526
4. Farquhar G, Caemmerer S, Berry J (1980) A biochemical model of photosynthetic CO_2 assimilation in leaves of C3 species. Planta 149(1):78–90
5. Wullschleger S (1993) Biochemical limitations to carbon assimilation in C3 plants: a retrospective analysis. J Exp Bot 44:907–920
6. Wingler A, Lea P, Quick W, Leegood R (2000) Photorespiration: metabolic pathways and their role in stress protection. Philos Trans Royal Soc London. Ser B: Biol Sci 355(1402):1517
7. Heber U, Bligny R, Streb P, Douce R (1996) Photorespiration is essential for the protection of the photosynthetic apparatus of C3 plants against photoinactivation under sunlight. Bot Acta 109:307–315
8. Morris M (1991) Factorial sampling plans for preliminary computational experiments. Technometrics 33(2):161–174
9. Saltelli A, Tarantola S, Campolongo F (2004) Sensitivity analysis in practice: a guide to assessing scientific models. John Wiley & Sons Inc.
10. Rosvall M, Bergstrom C (2007) An information-theoretic framework for resolving community structure in complex networks. Proc Natl Acad Sci 104(18):7327
11. Storn R, Price K (1997) Differential evolution – a simple and efficient heuristic for global optimization over continuous spaces. J Global Optim 11(4):341–359
12. Umeton R, Stracquadanio G, Sorathiya A, Papini A, Liò P, Nicosia G (2011) Design of robust metabolic pathways. In: Proc 48th design automation conference, DAC 2011, San Diego, CA, USA, June 5–9, 2011, ACM, pp 747–752
13. Stracquadanio G, Nicosia G (2011) Computational energy-based redesign of robust proteins. Comput Chem Eng 35(3):464–473
14. Hooke R, Jeeves TA (1961) "Direct search" solution of numerical and statistical problems. J ACM 8(2):212–229
15. Huyer W, Neumaier A (1999) Global optimization by multilevel coordinate search. J Global Optim 14(4):331–355

16. Vaz A, Vicente L (2007) A particle swarm pattern search method for bound constrained global optimization. J Global Optim 39(2):197–219
17. Audet C, Dennis JE (2007) Mesh adaptive direct search algorithms for constrained optimization. SIAM J Optim 17(1):188–217
18. Hansen N, Ostermeier A (2001) Completely derandomized self-adaptation in evolution strategies. Evol Comput 9(2):159–195
19. Jones DR, Perttunen CD, Stuckman BE (1993) Lipschitzian optimization without the Lipschitz constant. J Optim Theor Appl 79(1):157–181
20. Lewis R, Torczon V (1999) Pattern search algorithms for bound constrained minimization. SIAM J Optim 9(4):1082–1099
21. Gilmore P, Kelley CT (1995) An implicit filtering algorithm for optimization of functions with many local minima. SIAM J Optim 5(2):269–285