Preprints of the 12th IFAC Symposium on Computer Applications in Biotechnology
The International Federation of Automatic Control
16-18, 2013, December. Mumbai, India

A Compressed Sensing Based Basis-pursuit Formulation of the Room Algorithm

Tirthankar Sengupta*,. Shivi Jain*. Mani Bhushan ***

*Department of Chemical Engineering, Indian Institute of Technology Bombay, Mumbai 400076, India
(e-mail: tirthankar.s@iitb.ac.in ; shivi.jain@iitb.ac.in ; mbhushan@iitb.ac.in).
** On Leave From: Department of Chemical Engineering, Heritage Institute of Technology, Kolkata 700107, India.

Abstract: Regulatory on off minimization (ROOM) is a popular metabolic modeling strategy for obtaining the fluxes of various metabolic reactions in a mutant. It is based on minimization of the number of flux changes with respect to the wild-type. The ROOM approach involves solving an integer programming problem. In ROOM, the number of integer decision variables is equal to the number of reactions in the metabolic network under consideration. Typically, metabolic networks of interest are genome scale implying that the number of reactions in the network and hence the number of integer decision variables is large. The ROOM approach thus has inherent difficulties associated with large scale integer programming problems. In the current work, motivated by the emerging area of compressed sensing, we propose a reformulation, known as basis-pursuit, of the ROOM algorithm. The proposed formulation is an L1 norm minimization problem and is thus convex in nature. The proposed approach is used to obtain the flux profiles for various mutants of the *Synechocystis* species strain PCC 6803. The results are compared with the existing ROOM approach. It is observed that the proposed algorithm performs better in most cases. Use of compressed sensing based formulation creates exciting possibilities of efficiently reformulating various other metabolic network analysis problems.

Keywords: ROOM, Compressed Sensing, *Synechocystis*, LP, MILP.

1. INTRODUCTION

Metabolic models are quantitative frameworks for studying the metabolism of wild type and mutant organisms. Several metabolic modeling techniques exist in literature. Flux Balance Analysis (FBA) is a simple and yet a very important metabolic modeling strategy for wild type organisms. It is based on the assumption that physicochemical principles such as the laws of conservation of mass and energy hold true for metabolic networks (Kauffman et al., 2003). Further, in FBA, it is assumed that organisms have evolved over millennia to maximize growth (Varma and Palsson, 1993). However, this assumption of maximal growth may not hold true for mutant organisms. Studying the metabolism of mutant organisms is important since, in such organisms fluxes can get redirected towards some desired metabolite due to the genetic modifications. For example, a gene deletion will cause the flux of the reaction catalyzed by the enzyme coded by the gene to become zero, thereby potentially perturbing the fluxes in the entire metabolic network.

Several metabolic modeling strategies have been proposed which specifically apply to mutant organisms. Regulatory On Off Minimization (ROOM) (Shlomi et al., 2005) is one such popularly used strategy. Several studies based on ROOM exist in literature. Results of ROOM simulations performed on *E. coli* gene deletion mutants have been found to be in close agreement with experimental data on such mutants (Chen et al., 2010). The ROOM algorithm has been included as a part of the open source software package OptFlux which consists of computer programs of several metabolic modeling techniques (Rocha et al., 2010).

The basic premise in the ROOM approach is that the aim of the mutant is to minimize the number of significant flux changes compared to the wild type organism. This hypothesis is reasonable since there is evidence to suggest that the final metabolic steady state after a mutation is quite close to that in the wild type with major flux changes occurring in only a few reactions (Shlomi et al., 2005). Identification of the flux profiles of a mutant using ROOM thus involves solving an optimization problem where the number of significant flux changes compared to the wild type fluxes are minimized in the mutant, subject to the fluxes satisfying the conservation laws. This then results in an integer programming optimization problem. The number of integer decision variables in this problem is the same as number of fluxes in the metabolic network and can easily vary from a few hundreds to a few tens of thousands. The resulting optimization problem can then be a large scale integer programming problem and will have the inherent difficulties associated with solving large scale integer programming problems.

In this work, we propose an alternate formulation of the ROOM algorithm that alleviates the problems associated with the traditional formulation. Our proposed formulation is motivated by compressed sensing theory. Compressed sensing is a new paradigm in measurement and signal processing. It refers to a sampling technique in which only a

few measurements suffice for the reconstruction of the original signal. The number of required samples are much fewer than that required by Nyquist/Shannon sampling theory (Baraniuk, 2007). The key requirement in this reconstruction is that the underlying signal should be sparse. It is shown in literature that the unknown underlying signal can then be almost always obtained using a convex program (Candes, 2006). This convex program involves minimization of one norm of the sparse signal and can even be recast as a linear program known as basis pursuit (Baraniuk, 2007). In recent past, compressed sensing has been applied to a wide variety of areas such as imaging (Romberg, 2008), electron tomography (Thomas, 2012) and microarrays (Dai et al. 2009). Motivated by compressed sensing, our proposed approach for the alternate ROOM formulation is based on the reasonable premise that the number of fluxes in a mutant significantly different from the wild type fluxes is a small fraction of the total number of fluxes. This assumption of relatively few significant flux changes, however, may not always be true such as for lethal mutants. For example, for *S. cerevisiae*, the ROOM algorithm correctly predicted experimentally determined lethal genes as lethal only 29% of the time (Shlomi et al., 2005). For scenarios where the assumption of relatively few significant flux changes holds true, the ROOM approach can be used. For such cases, the flux change vector is expected to be sparse and we propose to obtain it using the known equality constraints on fluxes as a proxy for measurements. We thus formulate the ROOM algorithm as a linear programming problem.

We demonstrate the applicability of our proposed approach by obtaining flux profiles for mutants of the unicellular cyanobacterium *Synechocystis* species strain PCC 6803 in the context of ethanol production. This organism is selected since it is a model organism and can synthesize useful organic molecules such as ethanol by virtue of photoautotrophic metabolism. In particular, we predict the flux profiles corresponding to several mutants obtained by deleting single or a pair of genes. We compare our results to the standard ROOM approach and show that in several cases we obtain profiles which have lesser number of significant flux changes compared to the standard approach.

2. RELEVANT TECHNIQUES

2.1 The Flux Balance Analysis (FBA) Model

In this work, we consider the FBA model as described in *Formulation I* (Montagud et al., 2010). In equation (1), Z represents the biomass flux and v represents the flux vector containing all the fluxes. The biomass flux is maximized subject to mass balance and thermodynamic constraints. The term S is the stoichiometric matrix. The $(i,j)^{th}$ entry of matrix S is the signed stoichiometric coefficient of the i^{th} internal metabolite in the j^{th} reaction. It is assumed in FBA that all internal metabolites are at steady state. Hence the linear equality mass balance constraints $S.v = 0$ are enforced. The other constraints in equation (1) arise from thermodynamics. Irreversible reactions (corresponding to indices of the set I_{irr}) are constrained to the positive real line whereas reversible reactions (I_{rev}) are left unconstrained. Special bounds are put on constrained (I_{const}) and uptake reactions (I_{uptake}). The set I is thus the set of indices of all reactions in the metabolic network. The mathematical problem posed by equation (1) is thus a linear programming optimization problem with v being the vector of decision variables.

Formulation I: FBA model

$$\left.\begin{aligned} &\text{Maximize } Z = v_{biomass} \\ &\text{subject to,} \\ &S.v = 0 \\ &0 \le v_j < \infty, \quad \forall j \in I_{irr} \\ &-\infty < v_j < \infty, \ \forall j \in I_{rev} \\ &v_{j,\min} < v_j < v_{j,\max}, \forall j \in I_{const} \\ &v_{j,\min} < v_j < v_{j,\max}, \forall j \in I_{uptake} \\ &\text{where, } I_{irr} \cup I_{rev} \cup I_{const} \cup I_{uptake} = I \end{aligned}\right\} \quad (1)$$

2.2 Geometric FBA

The linear programming problem posed in *Formulation I* can have a unique optimal solution or infinite number of optimal solutions. The latter case is referred to as multiple optima. To address the issue of multiple optima, we consider a FBA algorithm that always obtains an unique solution (Smallbone and Simeonidis, 2009). We refer to this algorithm as "geometric FBA". The algorithm uses geometric arguments to obtain a solution devoid of internal cycles (Smallbone and Simeonidis, 2009). It is important to note that internal cycles here refer to cycles in which there is no net reaction. Thus redox cycles or futile cycles are not internal cycles as per this definition.

2.3 Regulatory on off minimization (ROOM)

Regulatory on off minimization (ROOM) is a modeling strategy for mutants which does away with the assumption of maximal growth. The ROOM algorithm is described by *Formulation II* (Shlomi et al., 2005).

In *Formulation II*, the superscript M denotes that the metabolic network of a mutant is under consideration. Also, N and N' refer to the number of reactions in the metabolic network of the wild type and the mutant respectively. Thus the number of gene deletions is $(N - N')$. The variables y_j are binary variables that can take values of either zero or one. The vector w^{MT} represents the wild type flux distribution. From equation (4) in *Formulation II,* it is observed that if the mutant flux value v_j^M lies outside the bounds w_j^u and w_j^l, the integer variable y_j must necessarily assume a value of one. On the contrary, if v_j^M lies within the bounds w_j^u and w_j^l, then y_j can assume a value of zero or one. Since the sum of the values of y_j is minimized, y_j is

forced to assume a value of zero in the latter case. δ and ε in equation (4) are user specified parameters. Their choice determines the bounds w_j^u and w_j^l that are used for labelling a flux change as significant or not significant. In this work, the values of δ and ε were taken to be 0.03 and 0.001 respectively (Shlomi et al., 2005).

Formulation II: ROOM formulation

$$\text{Minimize } \sum_{j=1}^{N'} y_j \tag{2}$$

subject to,

$$\left.\begin{aligned} & S^M . v^M = 0 \\ & 0 \le v_j^M < \infty, \forall j \in I_{irr}^M \\ & -\infty < v_j^M < \infty, \ \forall j \in I_{rev}^M \\ & v_{j,\min}^M < v_j^M < v_{j,\max}^M, \ \forall j \in I_{const}^M \\ & v_{j,\min}^M < v_j^M < v_{j,\max}^M, \ \forall j \in I_{uptake}^M \end{aligned}\right\} \tag{3}$$

$$\left.\begin{aligned} & v_j^M - y_j (v_{j,\max}^M - w_j^u) \le w_j^u \\ & v_j^M - y_j (v_{j,\min}^M - w_j^l) \ge w_j^l \\ & w_j^u = w_j^{MT} + \delta \mid w_j^{MT} \mid + \varepsilon \\ & w_j^l = w_j^{MT} - \delta \mid w_j^{MT} \mid - \varepsilon \\ & y_j \in \{0,1\} \end{aligned}\right\} \quad j = 1,2,..,N' \tag{4}$$

2.4 Compressed sensing theory

We now briefly discuss the theory of compressed sensing. More details can be obtained from literature (for example, Baraniuk, 2007). Consider a real valued signal x with P elements which in some basis ψ can be represented as:

$$x = \psi z \tag{5}$$

In equation (5), ψ is a PXP basis matrix and z is the $PX1$ vector of weighting coefficients. The signal x is said to be K-sparse if only K of the coefficients in vector z are non-zero (Baraniuk, 2007). Thus the information in the signal x can be stored in the non-zero elements of z which are less than the number of elements in x. Hence the signal x is called a compressible signal. The essential idea of compressed sensing is to make $Q \approx K$ measurements and then reconstruct the entire P dimensional K-sparse signal x or equivalently its representation z in basis ψ using the available Q measurements. These Q measurements can be represented as linear combinations of the P dimensional signal x as:

$$y = \phi x = \phi \psi z = \theta z \tag{6}$$

where ϕ is a QXP matrix. In Equation (6), y is a $QX1$ vector which stores the $Q << P$ measurements, resulting in an underdetermined system of equations. In equation (6), it is important to ensure that for a unique z there is a unique y. This is ensured if the matrix θ satisfies the restricted isometry property (RIP) (Baraniuk, 2007). The RIP is satisfied when all the possible subsets of P columns of the matrix θ are nearly orthogonal (Candes, 2006). A random QXP matrix is found to satisfy the RIP with high probability when $Q \ge cK \log(P/K)$ where c is a small constant (Baraniuk, 2007). After finding a suitable matrix θ, the next task is to develop a technique to actually recover the P dimensional signal z from just the Q measurements. The following formulation recovers the signal z:

$$\text{Minimize } \|z\|_0 \text{ such that } y = \theta z \tag{7}$$

where $\|z\|_0$ represents the number of nonzero elements in vector z. The above optimization problem is both numerically unstable and NP-complete (Baraniuk, 2007). Therefore, getting a solution in a reasonable amount of time is not possible with the above problem formulation. An alternative optimization based on L_1 norm minimization is able to recover K-sparse signals and closely approximate compressible signals with high probability (Baraniuk et al., 2007). The optimization based on L_1 norm, known as basis-pursuit problem, is as follows (Baraniuk, 2007):

$$\left.\begin{aligned} & \text{Minimize } \|z\|_1 = \sum_{i=1}^{P} |z_i| \\ & \text{such that } y = \theta z \end{aligned}\right\} \tag{8}$$

The quantity $\|z\|_1$ is the L_1 norm of vector z which is a convex function of z (Ganguli and Sompolinsky, 2012). Several efficient algorithms for solving the above convex optimization problem are available in literature (Ganguli and Sompolinsky, 2012). The problem posed in equation (8) can also be cast as a linear programming problem (Chvatal, 1983).

2.5 The proposed alternate ROOM formulation

Motivated by compressed sensing, we now propose an alternative formulation of the ROOM algorithm similar to equation (7). In particular, the vector γ defined as $(v^M - w^{MT})$ can be thought of as a sparse signal as mutant fluxes tend to remain close to the corresponding wild type values while satisfying the constraint $S^M . v^M = 0$. The proposed alternate ROOM formulation is described by *Formulation IIIa*.

Formulation IIIa: Alternate ROOM formulation

$$\text{Minimize } \|\gamma\|_0 \tag{9}$$

subject to,

$$\left.\begin{aligned} & S^M.\gamma = -S^M.w^{MT} \\ & -w_j^{MT} \le \gamma_j < \infty \,, \forall j \in I_{irr}^M \\ & -\infty < \gamma_j < \infty \,, \quad \forall j \in I_{rev}^M \\ & v_{j,\min}^M - w_j^{MT} < \gamma_j < v_{j,\max}^M - w_j^{MT} \,, \quad \forall j \in I_{const}^M \\ & v_{j,\min}^M - w_j^{MT} < \gamma_j < v_{j,\max}^M - w_j^{MT} \,, \quad \forall j \in I_{uptake}^M \end{aligned}\right\} \tag{10}$$

The vector γ obtained by solving *Formulation IIIa* can be used to obtain the vector v^M (the mutant flux distribution) by the relation $v^M = \gamma + w^{MT}$, since the vector w^{MT} is known from FBA.

Formulation IIIa will thus enable obtaining the flux profile v^M which has the minimum deviation from the wild type flux distribution w^{MT}, given the mass balance and thermodynamic constraints posed by equation (10) which the metabolic network of the mutant must obey. However, the optimization problem posed in *Formulation IIIa* is NP-complete (Baraniuk, 2007). Hence this problem is numerically intractable. The stoichiometric matrix S^M in equation (10) plays the role of the measurement matrix θ in equation (7) whereas $(-S^M.w^{MT})$ assumes the role of measurement y in equation (7). To ensure that *Formulation IIIa* can be solved by reformulating as a basis-pursuit optimization problem, the measurement matrix is required to satisfy RIP (section 2.4). In equation (10), this stoichiometric matrix is not necessarily random in nature. We still expect that a basis-pursuit reformulation of *Formulation IIIa* will lead to the same solution as in equation (7) due to the following important result encoded as a phase transition in compressed sensing. This is briefly discussed next.

Phase Transition in Compressed Sensing: Recently it has been reported (Ganguli and Sompolinsky, 2012) that the ability to extract the sparse solution in compressed sensing is insensitive to the details of the measurement matrix θ and the unknown signal *z* in the limit when Q and P are large. It instead depends only on the degree of sampling α and the signal sparsity *f*. The degree of sampling is defined as the number of measurements divided by the total number of elements in the unknown signal. The signal sparsity is defined as the number of non-zero elements in the signal divided by the total number of elements in the signal. In particular, a phase transition diagram in compressed sensing in the $\alpha - f$ plane has been reported (Figure 3 in Ganguli and Sompolinsky, 2012). If for a given *f*, the value of α is greater than the phase transition value $\alpha_c(f)$, then L_1 minimization typically yields perfect signal reconstruction (Ganguli and Sompolinsky, 2012).

The number of rows in stoichiometric matrix in metabolic modeling will typically be comparable (though less) to the number of columns. Thus the degree of sampling α in *Formulation IIIa* is quite close to 1. Further the signal sparsity would be quite low as it is expected that only a few fluxes in a mutant will be different from the wild type fluxes. Thus we expect α to be above the critical value obtained from the phase transition diagram. This will be verified in the case study to be presented later. The optimization problem in equations (9) and (10) can then be reformulated as a basis-pursuit problem as:

Formulation IIIb:L1 reformulation of IIIa

$$\left.\begin{aligned} & \text{Minimize } \|\gamma\|_1 \\ & \text{subject to} \\ & \text{equation (10)} \end{aligned}\right\} \tag{11}$$

3. CASE STUDY

We now apply the above mentioned techniques to study the metabolism of *Synechocystis* species strain PCC 6803 (Montagud et al., 2011). The most recent wild type model as communicated to us by the authors was used (Montagud A, personal communication, December 14, 2012). This model consists of 923 internal metabolites, 62 external metabolites and 983 reactions.

Fu (2009) created a *Synechocystis* 6803 mutant by inserting the genes *pdc* and *adh* into the genome of the *Synechocystis* 6803 wild type. This resulted in a mutant with an additional ethanol producing pathway and this mutant will henceforth be referred to as the "*Synechocystis* mutant". The number of reactions in the *Synechocystis* mutant is thus 985. The flux profile in the *Synechocystis* mutant was obtained from geometric FBA.

Sengupta et al. (2013) performed exhaustive single and double gene deletion simulations on the *Synechocystis* mutant using the ROOM algorithm. For these, the wild type flux distribution was the flux profile of the *Synechocystis* mutant as predicted by geometric FBA. For the single gene deletion simulations, the Pareto front with the biomass flux and the ethanol flux as the two objectives to be maximized was obtained and analyzed. The Pareto front represents an optimal trade off between the two fluxes. A similar analysis was done for the double gene deletion simulations (Sengupta et al., 2013). All simulations were performed in the autotrophic metabolic mode. In each simulation, glucose uptake was constrained to zero. In addition, carbon dioxide and bicarbonate uptake were each bounded to 1.7 mmol/gDW/hr (Montagud et al. 2010).

In this work, we compare the existing ROOM algorithm with the proposed L1 reformulation by performing gene deletion simulations on the *Synechocystis* mutant using both the techniques in the photoautotrophic metabolic mode. The L1

reformulation (*Formulation IIIb*) is solved by converting to a linear programming formulation (Chvatal, 1983). The criterion of deciding whether a flux change is significant or not in the proposed L1 reformulation remained the same as in the original ROOM algorithm. Thus, the flux change for the j^{th} reaction was considered significant if the flux value predicted by the proposed L1 reformulation lied outside the bounds w_j^u and w_j^l with δ and ε being 0.03 and 0.001 respectively. This is in accordance with equation (4). All simulations were performed using the CPLEX solver in TOMLAB® version 5.9 optimization suite operating in MATLAB® version 7.4 running on Ubuntu 10.04 LTS operating system on an Intel Quad Core PC with 3 GB of RAM.

3.1 Single gene deletions

Systematic exhaustive single gene deletion simulations were performed using both the optimization techniques. Amongst all such simulations, the number of common deletions for which simulations from both strategies converged was 778. In these, the maximal number of significant flux changes as predicted by the L1 reformulation was 306. Also, for single gene deletions, the number of reactions in the metabolic network is 984 (one less than 985). Hence the maximum value of the signal sparsity f for all the single gene deletion cases is (306/984) =0.311. Further, the value of the degree of sampling α is (923/984) = 0.938 since the number of internal metabolites for the *Synechocystis* mutant is 923. For f =0.311, the phase transition value $\alpha_c \approx 0.7$ (Figure 3 in Ganguli and Sompolinsky, 2012). Thus the value of α is well above the critical value of 0.7 for all the cases considered in this work. Hence the model posed by the L1 reformulation is expected to lead to solutions that minimize the number of significant flux changes for all mutants.

For the 778 cases, the number of significant flux change values as obtained from the proposed L1 reformulation algorithm were subtracted from the corresponding values as obtained from the orginal ROOM algorithm and the resultant values were plotted in Figure 1. From Figure 1, it is observed that barring a few exceptions, the original ROOM algorithm has predicted a higher number of significant flux change values compared to the proposed approach.

The number of significant flux changes in various pathways have been shown for two single gene deletion mutants in Table 1. For mutant A, the biomass flux as predicted by both the strategies is the same as that in the wild type. Hence it is expected that for A, the flux distribution should be very close to that in the wild type thereby leading to only a few significant flux changes. From Table 1 we note, that for mutant A, the proposed reformulation predicts much fewer significant flux changes as compared to ROOM and thus gives more meaningful results. For mutant B, the biomass flux predicted by both the strategies is zero and a large number of significant flux changes are thus expected. From Table 1, we observe that this is indeed the case and the predictions from the two strategies are comparable.

For the sake of comparison, we also report the total time taken for running the 778 simulations. For the ROOM approach, the total simulation time was 396 seconds whereas for the L1 reformulation approach, the same was 310 seconds. Thus the reformulation required less running time.

Table 1. Significant flux changes for two mutants

Pathway	Fluxes	A		B	
		ROOM	L1	ROOM	L1
Glycolysis	16	4	2	6	3
TCA cycle	8	0	0	6	6
PP Pathway	14	2	1	3	2
Amino Acid	146	6	0	75	78
Nucleic Acid	102	13	0	45	35
Fatty Acid	73	10	0	61	61
All	985	53	5	317	284

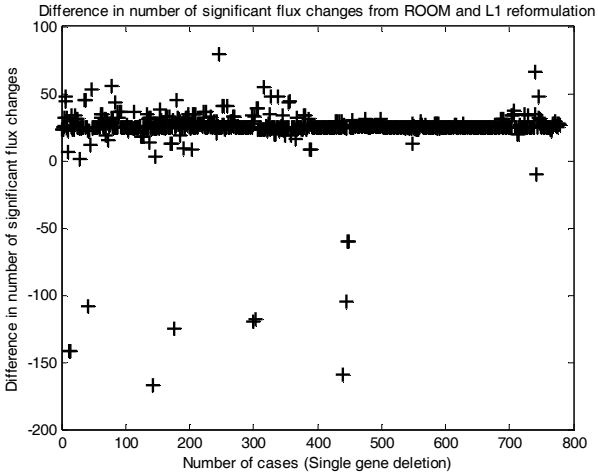

Fig. 1. Figure showing the difference in the number of significant flux changes from single gene deletion simulations on the *Synechocystis* mutant (ROOM minus L1).

3.2 Double gene deletions

For the double gene deletion simulations, the total number of possible systematic deletion pairs is high. Thus, we considered 554 gene deletion pairs which formed the Pareto front for the double gene deletion ROOM simulations (Sengupta et al., 2013).

For the 554 cases, the number of significant flux change values as obtained from the proposed L1 reformulation algorithm were subtracted from the corresponding values as obtained from the orginal ROOM algorithm and the resultant values were plotted in Figure 2. From this figure, it is observed that the original ROOM algorithm again predicts a higher number of significant flux changes when compared to the proposed approach in most cases.

4. CONCLUSION

ROOM is a popular metabolic modeling strategy for obtaining flux distributions in mutant organisms. In this

work, we have proposed a reformulation of the original ROOM algorithm. Motivated by compressed sensing, the proposed approach results in a L1 norm minimization problem which is convex in nature as opposed to the existing integer programming formulation. The utility of the proposed approach was demonstrated by applying it to analyze flux profiles for mutants obtained by single and double gene deletions on a mutant strain of the unicellular cyanobacterium *Synechocystis* species strain PCC 6803. In most of the cases, fewer number of significant flux changes were obtained as compared to the traditional ROOM implementation. Future work can be directed towards inspecting the flux values obtained from the two strategies and comparing the predicted values with actual in-vivo flux estimates obtained from metabolic flux analysis using 13C isotopic labelling. We are currently investigating the application of the proposed compressed sensing based approach for other type of metabolic modeling applications involving minimization of significant (non-zero) fluxes.

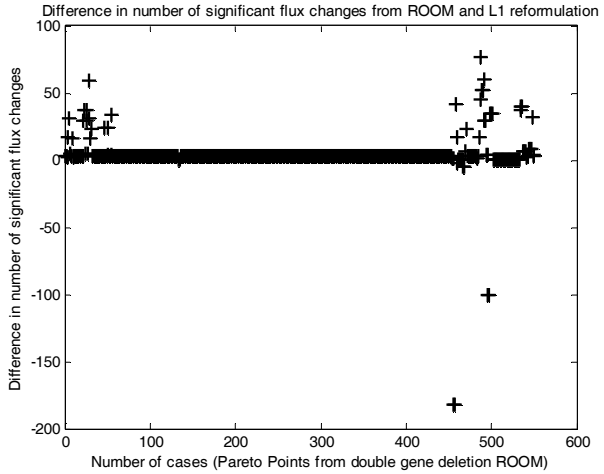

Fig. 2. Figure showing the difference in the number of significant flux changes from double gene deletion simulations on the *Synechocystis* mutant (ROOM minus L1).

Acknowledgement: We acknowledge Arnau Montagud for providing the most recent *Synechocystis* species strain PCC 6803 wild type metabolic model to us.

REFERENCES

Baraniuk, R. (2007). Compressive sensing. *IEEE Signal Processing Mag*, 118–120.

Candes, E. (2006). Proceedings of the International Congress of Mathematicians, Madrid, Spain, 3, 1433-1452.

Chen, Q., Wang, Z. and Wei D. (2010). Progress in the applications of flux analysis of metabolic networks. *Chinese Science Bulletin*, 55 (22), 2315–2322.

Chvatal V. (1983). *Linear Programming*. W.H. Freeman and Company, New York.

Dai, W., Sheikh, M. A., Milenkovic, O. and Baraniuk, R. G. (2009). Compressive sensing DNA microarrays. *EURASIP Journal on Bioinformatics and Systems Biology,* article id 162824, doi: 10.1155/2009/162824.

Fu, P. (2009). Genome-scale modeling of *Synechocystis* sp. PCC 6803 and prediction of pathway insertion. *J. Chem. Technol. Biotechnol.,* 84, 473-483.

Ganguli, S. and Sompolinsky, H. (2012). Compressed sensing, sparsity, and dimensionality in neuronal information processing and data analysis. *Annu. Rev. Neurosci.,* 35, 485-508.

Kauffman, K.J., Prakash, P. and Edwards, J.S. (2003). Advances in flux balance analysis. *Current Opinion in Biotechnology*, 14, 491-496.

Montagud, A., Navarro, E., Cordoba, P.F., Urchueguia, J.F. and Patil, K.R. (2010). Reconstruction and analysis of genome-scale metabolic model of a photosynthetic bacterium. *BMC Systems Biology*, 4, 156.

Montagud, A., Zelezniak, A., Navarro, E., Fernández de Córdoba, P., Urchueguía, J.F. and Patil, K.R. (2011). Flux coupling and transcriptional regulation within the metabolic network of the photosynthetic bacterium Synechocystis sp. PCC6803. *Biotechnology Journal*, 6, 330-342.

Rocha, I., Maia, P., Evangelista, P., Vilaça, P., Soares, S., Pinto, J.P., Nielsen, J., Patil, K.R., Ferreira, E.C. and Rocha, M. (2010). OptFlux: an open-source software platform for in silico metabolic engineering. *BMC Systems Biology*, 4, 45.

Romberg, J. (2008). Imaging via compressive sensing. *IEEE Signal Processing Magazine,* March 2008, 14-20.

Sengupta, T., Bhushan, M. and Wangikar, P.P. (2013). Metabolic modeling for multi-objective optimization of ethanol production in a *Synechocystis* mutant. *Submitted to Photosynthesis Research*.

Shlomi, T., Berkman, O. and Ruppin, E. (2005). Regulatory on/off minimization of metabolic flux changes after genetic perturbations. *PNAS*, 102 (21), 7695–7700.

Smallbone, K. and Simeonidis, E. (2009). Flux balance analysis: a geometric perspective. *J. Theor. Biol.*, 258 (2), 311–315.

Thomas, J. M., Leary, R., Midgley, P. A. and Holland D. J. (2012). A new approach to the investigation of nanoparticles: Electron tomography with compressed sensing. *Journal of Colloid and Interface Sciences,* 392, 7–14.

Varma, A. and Palsson, B.O. (1993). Metabolic capabilities of Escherichia coli. II. Optimal growth patterns. *J. Theor. Biol.*, 165, 503-522.