Functional integration of a metabolic network model and expression data without arbitrary thresholding

Paul A. Jensen and Jason A. Papin*

Department of Biomedical Engineering, University of Virginia, Charlottesville, VA 22908, USA
Associate Editor: Trey Ideker

ABSTRACT

Motivation: Flux balance analysis (FBA) has been used extensively to analyze genome-scale, constraint-based models of metabolism in a variety of organisms. The predictive accuracy of such models has recently been improved through the integration of high-throughput expression profiles of metabolic genes and proteins. However, extensions of FBA often require that such data be discretized *a priori* into sets of genes or proteins that are either 'on' or 'off'. This procedure requires selecting relatively subjective expression thresholds, often requiring several iterations and refinements to capture the expression dynamics and retain model functionality.

Results: We present a method for mapping expression data from a set of environmental, genetic or temporal conditions onto a metabolic network model without the need for arbitrary expression thresholds. Metabolic Adjustment by Differential Expression (MADE) uses the statistical significance of changes in gene or protein expression to create a functional metabolic model that most accurately recapitulates the expression dynamics. MADE was used to generate a series of models that reflect the metabolic adjustments seen in the transition from fermentative- to glycerol-based respiration in *Saccharomyces cerevisiae*. The calculated gene states match 98.7% of possible changes in expression, and the resulting models capture functional characteristics of the metabolic shift.

Availability: MADE is implemented in Matlab and requires a mixed-integer linear program solver. Source code is freely available at http://www.bme.virginia.edu/csbl/downloads/.

Contact: papin@virginia.edu

Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 INTRODUCTION

The availability of sequenced genomes has enabled the assembly of high-quality, genome-scale metabolic reconstructions (Oberhardt *et al.*, 2009). Various constraint-based techniques such as flux balance analysis (FBA) have been developed to predict the steady-state flux distribution through a metabolic network after various environmental and genetic perturbations. This method assumes that an organism's metabolic objective (or objective function) is to maximize the flux through a subset of pathways [often corresponding to the production of biomass in unicellular organisms as an approximation of growth (Feist and Palsson, 2010)] and formulates a linear programming problem describing the stoichiometry of the metabolic network (Orth *et al.*, 2010).

With constraint-based techniques such as FBA, *in silico* reconstructions have successfully predicted relationships between the presence or absence of a gene and the metabolic capabilities of a microorganism, including gene essentially (Feist *et al.*, 2007), adaptive evolution (Applebee *et al.*, 2008; Lee and Palsson, 2010) and strain design for optimal production of a metabolic byproduct (Burgard *et al.*, 2003; Feist *et al.*, 2010; Kim and Reed, 2010). These predictions assume that the reactions in a reconstruction are either 'on' or 'off', as determined by a set of Boolean rules connecting reactions to the binary expression states of associated genes. This simplification attempts to overcome the paucity of kinetic details for metabolic gene/reaction associations and allows genes to be mapped to the hundreds or thousands of reactions in a genome-scale model.

Genome-wide transcriptional regulatory networks (TRNs) can be used to predict the likely 'on' or 'off' state of metabolic genes as a function of environmental factors. Unfortunately, complete TRNs are only available for a limited number of organisms [including *Escherichia coli* (Covert *et al.*, 2004) and *Saccharomyces cerevisiae* (Herrgård *et al.*, 2006)]. Other reconstructions must instead rely on experimental measurements (typically from gene expression microarrays) to determine the appropriate gene states, but this approach suffers from two complications. First, classifying reactions as 'on' or 'off' requires setting a threshold in the continuous expression levels of the associated genes; such thresholds cannot be accurately determined from measured mRNA levels alone. Recent evidence indicates that such a correlation between transcript levels and metabolic reaction activity does exist (Moxley *et al.*, 2009). However, determining a specific threshold above which microarray measurements imply physiologically appreciable activity in the corresponding metabolic reaction is an unresolved problem. Second, these thresholds are not independent, since functioning metabolic reconstructions require a minimum set of reactions to be active in each condition, regardless of any apparent change in the underlying gene expression data. Especially when considering decreases in expression levels, the choice to fit these data by inactivating the corresponding reaction must be balanced with the functional effects of removing a reaction from the metabolic network.

An early attempt at reconciling gene expression data with an FBA model was the GIMME algorithm (Becker and Palsson, 2008). Using a set of user-supplied thresholds for the transition of each gene from 'on' to 'off', GIMME iteratively reactivated 'off' reactions (by turning on genes below their threshold) until a functioning model was obtained. While GIMME did produce functioning FBA models, the method required the *a priori* determination of expression thresholds.

---

*To whom correspondence should be addressed.

A more recent approach (Shlomi *et al.*, 2008) used the expression thresholds in previously published databases to classify reactions as either highly, moderately or lowly active. After choosing a threshold for the minimum flux through a highly active reaction, a mixed-integer linear program (MILP) was formulated to calculate a steady-state flux distribution that maximized the number of highly (lowly) expressed reactions that carried at least (not more than) the threshold flux. All such reactions were assumed to be active (inactive). Although this approach was successfully applied to a reconstruction of the human metabolic network, the same technique would not necessarily guarantee that the set of active reactions could produce the necessary objective flux in microorganisms with a more defined objective function. Additionally, determining the state of each reaction does not uniquely determine the expression state of each gene in the reconstruction. (An active reaction, for example, could be activated if either of two genes, each encoding an isozyme, were expressed.)

We present a method for *de novo* determination of a functional, binary expression state for all genes in a metabolic reconstruction using experimentally derived expression data. Our algorithm, Metabolic Adjustment by Differential Expression (MADE), uses an optimization-based approach to create a sequence of binary expression states that reflect the most statistically significant changes in the series of gene expression measurements; it does not require the use of arbitrary expression thresholds for each gene. The optimization is constrained to produce only functioning models by simultaneously imposing a minimum value on the system's metabolic objective. MADE is formulated as a single MILP problem that can be solved on several freely available, in addition to commercial, linear programming packages.

2 APPROACH

MADE relies on expression data from two or more conditions to determine the best-fitting gene states. (Again, 'conditions' can be any set of environmental, genomic or temporal data.) To illustrate, consider the simple metabolic system in Figure 1a. The product of the gene *input* facilitates the transport of metabolite A into the system, after which it undergoes transformation into B, catalyzed by either of two isozymes—the products of genes *left* and *right*. Production of B is assumed to be the metabolic objective. This example system receives an external stimulus, and the transcriptional response is measured for each gene at 0, 2 and 4 h, as summarized in Figure 1b.

As indicated by the expression data, the expression levels of genes *input* and *left* decrease between 0 and 2 h; they remain at this lower level of expression until the 4 h time point. Expression of the enzyme Right appears to increase between each set of time points. The challenges of converting these data into a binary gene state are illustrated in Figure 1c. As a first approximation, having *input* and *left* 'on' at 0 h and 'off' at 2 and 4 h matches the overall expression pattern. However, the gene product of *input* is always necessary for a functioning model, so neither of the resulting 2 or 4 h gene states is feasible. Finding a binary representation for the gene *right* is also not straightforward. The continuous expression levels increase during each transition, but a binary approximation can only transition from 'off' to 'on' once. It is unclear whether *right* should be actively expressed beginning at 2 or 4 h.

Figure 1d demonstrates the results of the MADE approach. Consistent with the measured expression, *left* turns 'off' between 0 and 2 h, remaining in that state for the final time point. Although *input* expression decreases significantly, MADE leaves the gene 'on' for the entire time course because of its essentiality. It is assumed that while mRNA expression of *input* has decreased by 2 h, the functional abilities of the gene product must be preserved to facilitate transport of a necessary metabolite. Finally, MADE activates *right* transcription at 2 h rather than 4 h, using the greater statistical significance of the increase between 0 and 2 h as evidence for this approximation. The resulting model is functional at all time points and suggests that the metabolic system initially routes flux through the Left-catalyzed pathway and transitions to the Right-catalyzed pathway after 2 h.

3 METHODS

3.1 FBA

At steady state, the fluxes through each chemical reaction in a metabolic network satisfy the system of equations $Sv = 0$, where v is a vector of fluxes and S is a matrix representation of the network stoichiometry. The rows of S correspond to the metabolites in the system. Each column in S describes a single chemical reaction, with negative (positive) entries indicating the stoichiometric coefficients of the reactants (products).

The FBA problem is formulated as the following linear program that maximizes the flux through an objective reaction (v_{obj}).

$$\max v_{obj}$$

subject to

$$Sv = 0 \qquad (a)$$

$$lb \leq v \leq ub \qquad (b)$$

In addition to the aforementioned mass balance constraints (a), upper and lower bounds, *ub* and *lb*, are imposed on the fluxes in the network. These bounds enforce thermodynamic reversibility and mechanistic (i.e. maximum uptake and secretion rate) constraints for the reactions. The solution to the FBA problem is a unique maximum flux through the objective reaction, which is used as a measure of biological fitness.

Several methods have been used to incorporate gene–protein reaction (GPR) and TRN constraints on an FBA problem (Covert *et al.*, 2004; Gianchandani *et al.*, 2006; Lee *et al.*, 2008). The SR-FBA formalism converts the Boolean rules into a series of integer inequalities, allowing for simultaneous solution with the FBA problem as a single MILP (Shlomi *et al.*, 2007). More details on this conversion are available as Supplementary Methods.

3.2 MADE representation

Given a sequence of measured mean expression levels $\{e_1, e_2, \ldots, e_n\}$ for a gene, a sequence of differences $\{d_{1\to 2}, d_{2\to 3}, \ldots, d_{n-1\to n}\}$ can be defined as

$$d_{i\to i+1} = \begin{cases} +1 & e_i < e_{i+1} \\ 0 & e_i = e_{i+1} \\ -1 & e_i > e_{i+1} \end{cases} \qquad (1)$$

For example, the value of $d_{1\to 2}$ signifies that the mean expression of the gene significantly increases (+1), decreases (−1) or remains constant (0) between conditions 1 and 2. Statistical significance is determined by $p_{1\to 2}$, the *P*-value corresponding to the confidence of assigning the gene to one of the cases in Equation (1).

MADE finds a sequence of binary expression states $\{x_1, x_2, \ldots, x_n\}$, $x_i \in \{0, 1\}$ such that the differences between successive states $(x_{i+1} - x_i)$ most closely match the corresponding differences in the mean expression levels
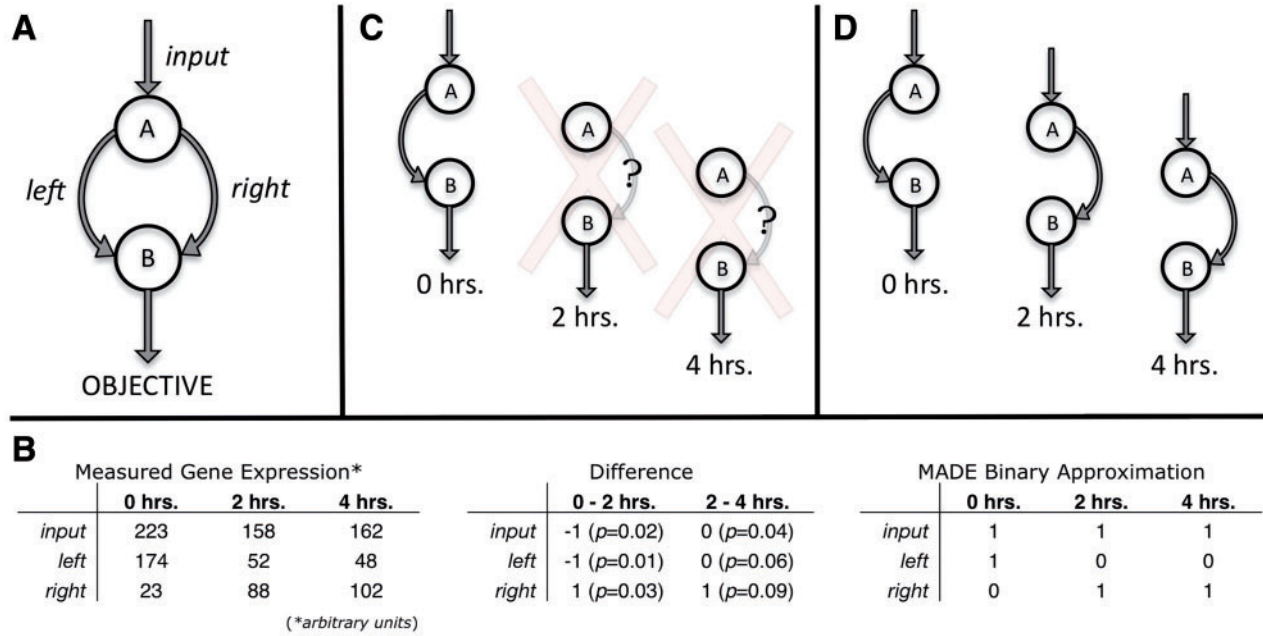
Fig. 1. Results of MADE on a prototypic system. (**A**) After transport facilitated by the gene product of *input*, metabolite A is converted to B by isozyme products of *left* or *right* before being removed through the objective reaction. (**B**) Measured expression data for each gene in the system, and the corresponding changes in expression (−1 and 1 indicate decreases and increases, respectively, and 0 represents no significant change). (**C**) Challenges with functional integration. Although *input* expression decreases, it must always remain 'on' for a functioning model. It is not clear from the expression data whether *right* should be first expressed at 2 or 4 h. (**D**) MADE finds the best functional model that reproduces the observed expression changes. The resulting model indicates that A is transformed to B by the product of *left* at early times. After 2 h, the flux is redirected through the reaction catalyzed by *right*'s product.

$d_{i \to i+1}$. Because the binary representation of gene expression assumed by FBA models cannot account for all possible expression patterns, MADE uses the statistical significance of the differences to create the most probable approximation. For example, consider a gene with mean expression levels (over four time points) of {10, 11, 65, 109}. If the calculated differences were {0, +1, +1} with *P*-values {0.02, 0.048, 0.0012}, an exact binary representation of this sequence is not possible, since the gene must increase over two consecutive intervals. MADE would choose the binary expression sequence {0, 0, 0, 1} over {0, 0, 1, 1}, since the increase between time points 3 and 4 is more statistically significant than the increase between time points 2 and 3.

More formally, the binary expression sequence returned by MADE is computed as the

$$\underset{x \in X}{\arg\min} \sum_{i=1}^{n-1} w(p_{i \to i+1}) |d_{i \to i+1} - (x_{i+1} - x_i)| \qquad (2)$$

where n is the number of conditions, $X \subset \{0, 1\}^n$ is the set of all gene expression states that allow functioning FBA models, and $w(p_{i \to i+1})$ is a weighting function that assigns larger weights to more significant *P*-values [for example, $w(p) = -\log p$]. The absolute value in the summand measures the discrepancy between the binary transitions and expression differences. Binary sequences that show an increase or decrease when the measured expression indicates no change (or vice versa) incur a 'penalty' of w; binary sequences that increase when expression data decrease (or vice versa) are penalized by $2w$.

3.3 Formulating the MILP problem

For each transition between conditions, the binary variables x representing the expression state of the metabolic genes are partitioned into three sets I, D and C, corresponding to increasing, decreasing or constant expression

between the two conditions. The optimization objective f for this transition is the weighted sum

$$
\begin{aligned}
f_{i \to i+1}(x) = & \sum_{x \in I} w(p_{x_{i \to i+1}})(x_{i+1} - x_i) \\
& + \sum_{x \in D} w(p_{x_{i \to i+1}})(x_i - x_{i+1}) \\
& - \sum_{x \in C} w(p_{x_{i \to i+1}}) \Delta_{x_i, x_{i+1}}
\end{aligned}
\qquad (3)
$$

where $\Delta_{x_i, x_{i+1}}$ is a binary variable that takes the value 0 when $x_i = x_{i+1}$ and 1 otherwise; this definition is enforced by adding the Boolean constraint $\Delta_{x_i, x_{i+1}} = (x_i \text{ XOR } x_{i+1})$.

The weights $w(p)$ were calculated as $(-\log p)$, but using other functions that mapped smaller *P*-values to a higher weight, e.g. $w(p) = 1 - p$, did not significantly change the results. The logarithmic transformation was chosen to minimize ill-conditioning for highly significant changes. Using a unit weighting function $w(p) = 1$ for all *P*-values would cause MADE to match the greatest number of transitions without regard for the significance of the changes. The *P*-values were determined by *t*-test on the normalized array data in accordance with standard microarray analysis procedures using the GeneSpring software package (Agilent Technologies, Palo Alto, CA, USA). The objective function reaches its theoretical maximum when all calculated expression states match the observed expression changes.

MADE maximizes the sum of the objective function values across all conditions subject to a regulated FBA formalism:

$$\max \sum_{i=1}^{n-1} f_{i \to i+1}(x)$$

subject to

$$Sv = 0 \qquad \text{(a)}$$

$$lb \leq v \leq ub \qquad \text{(b)}$$

$$v_{\text{obj}} \geq v_{\text{min}} \qquad \text{(c)}$$

$$N \begin{pmatrix} v \\ x \end{pmatrix} = b \qquad \text{(d)}$$

Constraints (a) and (b) represent the mass balance, stoichiometric and thermodynamic constraints of FBA. The metabolic objective (e.g. biomass production) is required to be above a minimum value in (c) to ensure a viable flux distribution. Minimum values of 0.1–0.3 of the optimal objective flux have been used to ensure viability in FBA-based bacterial strain design (Kim and Reed, 2010). The system of equations (d) contains an integer programming formalism for the GPR rules and a set of constraints that couple these rules to the metabolic fluxes v. These constraints were formulated using SR-FBA, as described above. A separate set of constraints (a–d) appear for each experimental condition.

MADE was implemented in Matlab r2009a (The Mathworks, Natick, MA, USA) using the genome-scale *S.cerevisiae* metabolic model iND750 (Duarte *et al.*, 2004). The final MILP problem was solved with the CPLEX solver (ILOG, Sunnyvale, CA, USA) on an Intel dual-core desktop computer running Linux. The MILP gap for all problems converged to less than 0.5% in under 500 s.

3.4 Flux variability analysis

Solutions to the FBA problem are not unique, since several flux distributions can yield the unique optimal objective value. As a variation on FBA, Flux Variability Analysis (FVA) captures the feasible range of reaction fluxes that satisfy an optimal objective value. The method calculates the minimum and maximum allowable fluxes through each reaction using the linear program

$$\max / \min v_i$$

subject to

$$Sv = 0 \qquad \text{(a)}$$

$$lb \leq v \leq ub \qquad \text{(b)}$$

$$v_{\text{obj}} = v_{\text{obj}}^* \qquad \text{(c)}$$

where (a) and (b) are the FBA constraints and v_{obj}^* is the optimal value of the objective flux. FVA was used in calculations described below.

4 RESULTS AND DISCUSSION

Expression data gathered from the glucose to glycerol shift in the yeast *S.cerevisiae* are ideal candidates for MADE application. Yeast treats glucose as its preferred carbon source, as evidenced by the repression of genes catabolizing all other carbon sources in glucose-rich media (Dickinson and Schweizer, 2004). After nearly all available glucose has been exhausted, yeast begin a drastic shift in the metabolic activity to initiate respiration of ethanol, the fermentation product of glucose. The effects of this 'diauxic shift' extend beyond central carbon metabolism, as the organism also slows production of amino acids and ribosomal-related proteins. The demand for these metabolic byproducts is decreased by the drastically slowed growth rate. Overall, nearly 1700 genes show differential expression in this process, accounting for over 25% of the yeast genome (DeRisi *et al.*, 1997). [For review of the diauxic shift, see Carlson (1999)]

While the metabolic effects of the diauxic shift are well studied, the transition from glucose- to glycerol-based growth is less understood. Roberts and Hudson used a series of microarrays to measure expression changes at 15, 30 and 60 min following a shift from glucose- to glycerol-rich media (Roberts and Hudson, 2006). The results indicated changes among expression levels in a number of gene ontologies and several individual genes. Integrating these data with a metabolic model would enable the analysis of pathway-level and functional shifts in the metabolic flux distribution.

We used MADE to integrate these data with iND750, a fully compartmentalized, genome-scale model of *S.cerevisiae* metabolism (Duarte *et al.*, 2004). A total of four environmental conditions were considered: growth to early log phase in YPD (glucose-rich media); and growth on YPG (glycerol-rich media) for 15, 30 and 60 min after the switch from YPD. Expression states were calculated for 743 metabolic genes in the model (a sample of these results is shown in Figure 2; complete results are shown in the Supplementary Material). Figure 3 shows the overall gene expression levels grouped by functional pathways. The expression levels for genes associated with each reaction in the pathway were averaged as a measure of overall subsystem-related transcriptional activity. The MADE-derived binary approximations for several subsystems follow the continuous expression patterns, although changes in measured expression often appear more subtle than switches between 'on' and 'off' states. Specifically, expression levels in the fatty acid and glutamate subsystems increase immediately after the glycerol transition, and membrane/ion transport activity decreases near the end of the transition period. All of these findings are consistent with published conclusions (Roberts and Hudson, 2006).

4.1 Flexibility of metabolic subsystems

The models returned by MADE also recapitulate several of the expected functional behaviors of the glucose/glycerol shift. We used FVA to characterize the metabolic flexibility in each subsystem. The heatmap in Figure 3 shows how the average range of admissible fluxes for each reaction in a subsystem changes during the shift. High values indicate that the flux through a subsystem can vary significantly while still achieving optimal growth. Lower values correspond to tight metabolic regulation.

As seen in Figure 3, the FVA results are similar for the iND750 model in the YPD and YPG environments before applying gene expression data. The MADE-calculated expression states change the metabolic capabilities of the model at each time point. Thus, a binary representation of gene expression enables multi-level resolution of metabolic functionality when examined on the pathway level.

Yeast grown with an ample glucose supply strongly repress catabolism of all other carbon sources (Dickinson and Schweizer, 2004). When glucose is removed, a genome-wide transcriptional adjustment reactivates pathways involved in transport and utilization of other nutrients. The MADE-generated models reflect this shift. Cells grown in YPD display the smallest range in flux variability, corresponding to a tightly regulated metabolic state that is highly optimized toward glucose utilization. After the shift to glycerol, the flux variability increases as the organism increases its metabolic versatility.

MADE requires all calculated expression states to yield functioning models, as defined by a minimum flux through the objective reaction. The models of the glucose/glycerol shift were required to produce at least 30% of their maximum objective flux. However, after integrating the expression data, the resulting models were capable of producing 44, 44, 60 and 69% of the maximum

Expression Data

MADE
Approximation
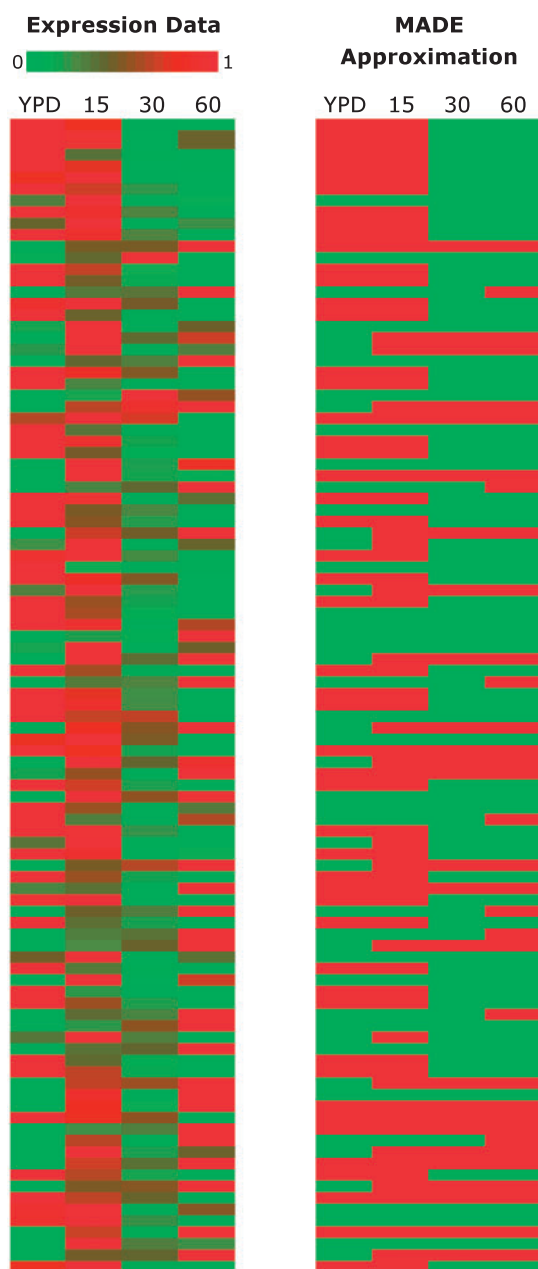
Fig. 2. MADE results for 100 randomly selected genes from the *S.cerevisiae* model. The left heatmap indicates the normalized expression levels across the four timepoints; the right heatmap displays the binary approximation calculated by MADE. Overall, 98.7% of the feasible transitions were correctly matched.

objective flux for the 0, 15, 30 and 60 min time points, respectively. While requiring a minimum level of metabolic functionality does change the gene states returned by MADE, the algorithm does not appear to be overly sensitive to the chosen minimum objective value.

The 30% objective flux threshold was chosen for this study to capture many of the short-term changes in expression following the transition in media. Previous work has shown that microbes behave suboptimally shortly after such transitions and evolve to

match the *in silico* optimal solution after hundreds of generations (Ibarra *et al.*, 2002). Future studies could use MADE to generate models that exhibit nearly optimal behavior by setting the objective flux threshold at a very high level (e.g. 90–99% of the optimal value).

4.2 Incorporating a transcriptional regulatory network

The differential expression initiated by the glucose to glycerol switch is due in part to transcriptional regulation. In order to refine the MADE-generated models, we coupled the iND750 model with iMH805, a transcriptional regulatory network previously assembled from primary literature (Herrgård *et al.*, 2006). The network consists of 436 Boolean rules capturing expression relationships among 82 nutrients, 55 transcription factors and 750 metabolic genes. The network was converted to a MILP system using the SR-FBA formalism, and the resulting equations were added to the GPR constraint set—no further modification of the MADE algorithm was necessary to calculate best fit expression states for each of the 55 transcription factors. The addition of the TRN changed the MADE prediction for 578 (77.8%) of the metabolic genes (see Supplementary Material). As expected, the network did help refine the MADE prediction: for example, *ENO2*, a phosphopruvate hydratase, is known to be expressed in glycerol-based media (SGD). Without the TRN, MADE predicted that only the related gene *ENO1* would be expressed. With the TRN, MADE correctly identified both *ENO1* and *ENO2* as being 'on'.

4.3 Accuracy of MADE

As previously mentioned, the binary variables representing gene expression in the FBA framework cannot describe all possible expression patterns. Using the prototypic model in Figure 1 as an example, the total number of comparisons is six (3 genes × 2 transitions). However, the number of feasible comparisons is only five, since the binary description of *right* expression cannot increase twice consecutively. Therefore, when measuring the ability of MADE's results to reproduce the expression change dynamics, it is useful to also calculate the percent of feasible transitions matched. The number of feasible matches can be calculated by running MADE without the minimum metabolic objective constraint. This allows MADE to fit gene expression states without regard for overall model functionality.

The fermentive/glycolytic shift data contained 2229 transitions, of which 1885 were feasible. MADE matched 98.7% of the feasible transitions (83.5% of all transitions). It is important to note that the 15.4% of transitions that are infeasible are a consequence of the binary representation of gene expression in FBA. Additionally, the assumption in MADE that the statistical significance of differential gene expression correlates with biological significance does not always hold. Also, some of the correctly matched transitions may not accurately reflect the true metabolic behavior. The transitions in gene expression data are not necessarily a completely accurate representation of the actual expression state due to errors in the experimental data.

5 CONCLUSION

Our MADE algorithm integrates gene or protein expression data and a metabolic model without *a priori* determination of activity thresholds. MADE can be applied to any set of two or more

Fig. 3. Average gene expression and flux variability by metabolic subsystem in *S.cerevisiae* after a transition from glucose- to glycerol-based respiration. Normalized expression levels were averaged across all genes associated with the subsystem. For comparison, the same approach was applied to the binary expression states calculated by MADE. The range of flux variability is the difference between the maximum and minimum allowable fluxes through a reaction while preserving an optimal objective value. These ranges were averaged for reactions in a subsystem and normalized: '1' indicates the largest average range for reactions in the subsystem; '0' is the smallest range. The 'Models Only' FVA results use the indicated media and no constraints on gene expression. The 'Models with Expression Data' show the changes to flux variability at each time point after the expression data were integrated with the model by MADE.

genetic, environmental or temporal conditions. The method attempts to match most closely the genes that exhibit the most statistically significant changes in expression levels. The resulting gene states always produce functioning models and match the direction of differential expression with high accuracy. Results from the glucose/glycerol shift in *S.cerevisiae* indicate that MADE-derived models also recapitulate the overall behavior of the actual biological system.

ACKNOWLEDGEMENTS

Conflict of Interest: none declared.

REFERENCES

Applebee,M.K. *et al.* (2008) Impact of individual mutations on increased fitness in adaptively evolved strains of escherichia coli. *J. Bacteriol.*, **190**, 5087–5094.

Becker,S.A. and Palsson,B.O. (2008) Context-specific metabolic networks are consistent with experiments. *PLoS Comput. Biol.*, **4**, e1000082.

Burgard,A.P. *et al.* (2003) Optknock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. *Biotechnol. Bioeng.*, **84**, 647–657.

Carlson,M. (1999) Glucose repression in yeast. *Curr. Opin. Microbiol.*, **2**, 202–207.

Covert,M.W. *et al.* (2004) Integrating high-throughput and computational data elucidates bacterial networks. *Nature*, **429**, 92–96.

DeRisi,J.L. *et al.* (1997) Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science*, **278**, 680–686.

Dickinson,J.R. and Schweizer,M. (2004) *The Metabolism and Molecular Physiology of Saccharomyces cerevisiae*, 2nd edn. CRC Press, Philadelphia, PA, USA.

Duarte,N.C. *et al.* (2004) Reconstruction and validation of saccharomyces cerevisiae ind750, a fully compartmentalized genome-scale metabolic model. *Genome Res.*, **14**, 1298–1309.

Feist,A.M. and Palsson,B.O. (2010) The biomass objective function. *Curr. Opin. Microbiol.*, **13**, 344–349.

Feist,A.M. *et al.* (2007) A genome-scale metabolic reconstruction for escherichia coli k-12 mg1655 that accounts for 1260 orfs and thermodynamic information. *Mol. Syst. Biol.*, **3**, 121.

Feist,A.M. *et al.* (2010) Model-driven evaluation of the production potential for growth-coupled products of escherichia coli. *Metab. Eng.*, **12**, 173–186.

Gianchandani,E.P. *et al.* (2006) Matrix formalism to describe functional states of transcriptional regulatory systems. *PLoS Comput. Biol.*, **2**, e101.

Herrgård,M.J. *et al.* (2006) Integrated analysis of regulatory and metabolic networks reveals novel regulatory mechanisms in Saccharomyces cerevisiae. *Genome Res.*, **16**, 627–635.

Ibarra,R.U. *et al.* (2002) Escherichia coli k-12 undergoes adaptive evolution to achieve in silico predicted optimal growth. *Nature*, **420**, 186–189.

Kim,J. and Reed,J.L. (2010) Optorf: optimal metabolic and regulatory perturbations for metabolic engineering of microbial strains. *BMC Syst. Biol.*, **4**, 53.

Lee,D.-H. and Palsson,B. Ø. (2010) Adaptive evolution of escherichia coli k-12 mg1655 during growth on a nonnative carbon source, l-1,2-propanediol. *Appl. Environ. Microbiol.*, **76**, 4158–4168.

Lee,J.M. *et al.* (2008) Dynamic analysis of integrated signaling, metabolic, and regulatory networks. *PLoS Comput. Biol.*, **4**, e1000086.

Moxley,J.F. *et al.* (2009) Linking high-resolution metabolic flux phenotypes and transcriptional regulation in yeast modulated by the global regulator gcn4p. *Proc. Natl Acad. Sci. USA*, **106**, 6477–6482.

Oberhardt,M.A. *et al.* (2009) Applications of genome-scale metabolic reconstructions. *Mol. Syst. Biol.*, **5**, 320.

Orth,J.D. *et al.* (2010) What is flux balance analysis? *Nat. Biotechnol.*, **28**, 245–248.

Roberts,G.G. and Hudson,A.P. (2006) Transcriptome profiling of Saccharomyces cerevisiae during a transition from fermentative to glycerol-based respiratory growth reveals extensive metabolic and structural remodeling. *Mol. Genet. Genomics*, **276**, 170–186.

SGD. The Saccharomyces genome database. Available at http://www.yeastgenome.org (last accessed date August 15, 2010).

Shlomi,T. *et al.* (2007) A genome-scale computational study of the interplay between transcriptional regulation and metabolism. *Mol. Syst. Biol.*, **3**, 101.

Shlomi,T. *et al.* (2008) Network-based prediction of human tissue-specific metabolism. *Nat. Biotechnol.*, **26**, 1003–1010.