AEM
Journals.ASM.org

Evaluation of a Genome-Scale *In Silico* Metabolic Model for *Geobacter metallireducens* by Using Proteomic Data from a Field Biostimulation Experiment

Yilin Fang,[a] Michael J. Wilkins,[a] Steven B. Yabusaki,[a] Mary S. Lipton,[a] and Philip E. Long[b]

Pacific Northwest National Laboratory, Richland, Washington,[a] and Lawrence Berkeley National Laboratory, Berkeley, California[b]

Accurately predicting the interactions between microbial metabolism and the physical subsurface environment is necessary to enhance subsurface energy development, soil and groundwater cleanup, and carbon management. This study was an initial attempt to confirm the metabolic functional roles within an *in silico* model using environmental proteomic data collected during field experiments. Shotgun global proteomics data collected during a subsurface biostimulation experiment were used to validate a genome-scale metabolic model of *Geobacter metallireducens*—specifically, the ability of the metabolic model to predict metal reduction, biomass yield, and growth rate under dynamic field conditions. The constraint-based *in silico* model *of G. metallireducens* relates an annotated genome sequence to the physiological functions with 697 reactions controlled by 747 enzyme-coding genes. Proteomic analysis showed that 180 of the 637 *G. metallireducens* proteins detected during the 2008 experiment were associated with specific metabolic reactions in the *in silico* model. When the field-calibrated Fe(III) terminal electron acceptor process reaction in a reactive transport model for the field experiments was replaced with the genome-scale model, the model predicted that the largest metabolic fluxes through the *in silico* model reactions generally correspond to the highest abundances of proteins that catalyze those reactions. Central metabolism predicted by the model agrees well with protein abundance profiles inferred from proteomic analysis. Model discrepancies with the proteomic data, such as the relatively low abundances of proteins associated with amino acid transport and metabolism, revealed pathways or flux constraints in the *in silico* model that could be updated to more accurately predict metabolic processes that occur in the subsurface environment.

Microbial environments have been engineered for enhanced oil recovery (7, 23, 29, 33, 42) and remediation of soil and groundwater contaminated by chlorinated hydrocarbons, metals, and radionuclides (4, 24, 30, 40, 47, 51, 63) in the past 30 years and have recently been studied to manage carbon dioxide (46). Mathematical models of microbial processes have been used for experimental interpretation, design, and prediction in recent decades (3, 10, 20, 22, 39, 48, 49, 53, 56). These processes were very complex to model because there was no way to measure the detailed metabolisms inside the system. They were usually represented by specific terminal electron acceptor process (TEAP) reactions with fixed stoichiometry throughout the simulation and were traditionally modeled by Monod kinetics. These models are sufficient under nonlimiting conditions of nutrients. However, the kinetic parameters of these models, which were calibrated but not directly measurable, cannot reflect the sophisticated mechanisms developed by microorganisms to adapt to the changing environment through the regulation of metabolic pathways, such as their ability to respond to nutrient gradients and environmental stress (5, 6).

With the advent of high-throughput sequencing data for genomics and transcriptomics, a number of genome-scale *in silico* models have been built for various organisms to study cell metabolism (13, 15, 37, 52, 54, 60) using different mathematical models, such as detailed kinetic models (36) and constraint-based models (16, 44, 58). The genome-scale models make the integration of cellular dynamics and continuum scale model possible by exchanging information on the growth rate, substrate uptake rates, and by-product rates under changing environmental conditions between the models (17, 27, 50). Highly detailed, constraint-based genome-scale metabolic *in silico* models have been developed using genomic data to identify metabolic pathways (28, 31, 37, 38, 44, 45, 52, 55). Because the constraint-based flux balance approach uses constraints such as mass balance, reaction reversibility, and bounds of reaction fluxes, a feasible solution may be found, given an objective function. Among the nearly 20 constraint-based models (15) for bacterial species, 2 have been used to analyze the metabolism and physiology of the *Geobacter* species that are capable of U(VI) reduction (37, 52). A constraint-based metabolic model of *Geobacter sulfurreducens* (37) has been successfully incorporated into a continuum scale reactive transport model (17, 50). Although the initial goal of the *in silico* model development was to fundamentally account for intracellular and environmental exchange reactions to more accurately predict TEAP conversion rates, the advent of mass spectrometry-based protein identification technology now provides an experimental way to assess the detailed intracellular reactions. Further advancements in the measurement of subsurface *in situ* processes via environmental proteomics now offers the potential to track metabolic functions to specific bacterial species (9, 61).

Previous studies have suggested that protein abundances inferred from proteomic data can be assumed to be proportional to the metabolic flux through a specific reaction (14, 32, 67), making possible the assessment of genome-scale models. Colijin et al. (14)

Received 1 June 2012 **Accepted** 2 October 2012
Published ahead of print 5 October 2012
Address correspondence to Michael J. Wilkins, michael.wilkins@pnnl.gov.
Supplemental material for this article may be found at http://aem.asm.org/.

doi:10.1128/AEM.01795-12

used whole-cell measurement of gene expression data to model the maximum flux through individual metabolic reactions using a genome-scale model of *Mycobacterium tuberculosis*, the pathogen that causes tuberculosis. In that method, the level of mRNA was used as an approximate upper bound on the maximum available protein, and consequently an upper bound was provided on the potential flux through the corresponding metabolic reaction rates. Proteomic and metabolomics data have been used to infer metabolic flux to achieve higher prediction accuracy of a genome-scale metabolic network model (67). Lewis et al. (32) also found good correspondence between flux solutions of constraint-based modeling of *Escherichia coli* K-12 metabolism and the set of identified proteins and gene expression levels from transcriptomic and proteomic data.

In this study, the metabolic functions of the genome-scale metabolic model of *Geobacter* under dynamic field conditions was evaluated using shotgun global proteomic analyses of planktonic biomass in groundwater samples from the Integrated Field Research Challenge (IFRC) site near Rifle, CO (9, 61). At the Rifle IFRC site, acetate amendment was used to stimulate microbially mediated immobilization of uranium in the unconfined aquifer (2, 12, 59, 62). Field data have shown that during acetate biostimulation, biological U(VI) removal from groundwater occurs concomitantly with the enzymatic reduction of Fe(III) minerals. These geochemical changes are associated with significant increases in the abundance of members of the *Geobacteraceae* family, in some instances accounting for ~90% of the subsurface microbial community (2, 19, 21, 41, 59). Previous modeling efforts of microbial processes during bioremediation field experiments at the Rifle site used a traditional modeling approach based on a single reaction with fixed stoichiometry for each TEAP with various forms of Monod-type rate laws (18, 34, 35, 64). The mechanistic representation of biologically mediated TEAP reactions is still a principal challenge to modeling *in situ* bioremediation because microorganisms modify their metabolic activity in response to spatial and temporal variations in environmental conditions, as demonstrated by the temporal impact of geochemical conditions on the detected metaproteome in the Rifle aquifer during biostimulation (9).

Proteomic and quantitative PCR analyses of planktonic biomass samples were obtained from the 2008 experiment at the Rifle IFRC. Proteomics analyses of these samples have previously been shown to link active metabolic pathways to individual microbial species (9, 61) and provide insights on uranium bioremediation under engineered biostimulation in naturally complex field settings. The approach of the present study was to replace the single reaction Fe(III) TEAPs and rate laws in the reactive transport model with the genome-scale metabolic model of *G. metallireducens*. While this would allow a comparison of the new process model coupling with the previous modeling results, a more important outcome is the opportunity to use proteomic data from a groundwater sample to assess the metabolic processes predicted by the genome-scale model.

MATERIALS AND METHODS

Field experiment and biomass samples. In an experimental plot at the Rifle IFRC site that had previously been amended with acetate in 2007, acetate and bromide were amended to the subsurface in 2008 ("secondary stimulation"). This experimental plot has a line of 10 injection wells perpendicular to the flow direction at 1-m spacing, 3 upgradient monitoring wells, and three rows of 4 downgradient monitoring wells 2.5, 5.0, and 8.5 m from the injection gallery. Monitoring wells D01 through D04 are present in the first row, wells D05 through D08 are present in the second row, and wells D09 through D12 are present in the third row. The experiment began on 20 July 2008, and the acetate biostimulation period was 110 days. The initial target field concentrations of bromide and acetate were 1.3 mM and 5 mM, respectively. The mixing of amendment between injection wells was controlled by continuous cross-well mixing using peristaltic pumps to transfer liquid between the 10 injection gallery wells. The continuous injectate release was interrupted for 1 week after 2 weeks of injection and then resumed. During the course of the experiment, dissolved oxygen concentrations were generally below 16 μM. Further detailed descriptions of the experiment field plot, the monitoring well network, and the site geochemistry can be found in Williams et al. (62).

The metabolic fluxes calculated from the constraint-based model are mmol/h normalized to gram dry weight of biomass. Initial biomass concentration is therefore an important parameter controlling the magnitude of the *G. metallireducens* reaction fluxes. During the 2008 experiment, 1- to 2-liter water samples were filtered through Sterivex 0.20-μm-pore-size filters at each depth interval of the U02-D02-D06 well transect during the Fe reduction phase (day 11), the Fe-sulfate reduction transition phase (day 25), and the sulfate reduction phase (day 38) (11). Quantitative PCR (qPCR) for specific bacteria (including *Geobacter*) were based on the Microbial Insights series of CENSUS qPCR assays. The quantities of bacteria were reported as cell equivalents per ml of groundwater during each phase.

Planktonic biomass samples for proteomics analysis were recovered 5, 7, and 10 days from well D04 after the start of the *in situ* biostimulation (9, 61) to quantitatively evaluate proteomes during the phase of dominant Fe reduction and to study the impact of secondary biostimulation on the proteome of the microbial community at the start of 2008. Per sample, 200 liters of groundwater were prefiltered (1.2-μm pore size) at a rate of ~2 liters min^{-1}, and a Pall tangential flow filtration (0.2 μm) system (Pall Corp., Port Washington, NY) was used to concentrate the biomass. Biomass was preserved by passing the extracted groundwater through chilling baths that contained an ice-rock salt mixture to maintain a temperature of ~1°C. Biomass pellets were formed after biomass was concentrated (~200 ml) in the retentate vessel and centrifuged at 4,000 rpm for 40 min at 4°C. The pellet was resuspended in ~5 ml of groundwater and immediately frozen in an ethanol-dry ice mix.

Protein mixtures were extracted from biomass pellets using a barocycler. (For protocol details, see reference 9.) Tandem two-dimensional (2-D) liquid chromatography-mass spectrometry analysis was then performed on peptide mixtures, using an LTQ Orbitrap Velos ETD mass spectrometer fitted with a custom electrospray ionization interface. The generated mass spectra were then matched to predicted peptides from *G. metallireducens* using the database-searching algorithm SEQUEST and filtered using an MSGF value of 1 e−10. Spectral count data across 15 fractions were summed for each sample and then divided by each protein length to normalize the data. These values were subsequently log transformed.

Genome-scale metabolic models for dominant Rifle subsurface *Geobacter* species, such as *Geobacter bemidjiensis*, have not yet been constructed. Callister et al. (9) and Aklujkar et al. (1) indicated that core metabolic functions within *G. metallireducens* are relatively similar to those within *G. bemidjiensis* at the Rifle IRFC site, even though *G. metallireducens* has not been identified as a dominant member within the *Geobacter* community in the subsurface during biostimulation. Thus, the genome-scale *in silico* model for *G. metallireducens* from Sun et al. (52) was used in the present study. The proteomic analysis detected 637 *G. metallireducens* proteins. Of these, 180 were associated with the metabolic reactions in the *in silico* model.

Modeling approaches. (i) Constraint-based *in silico* modeling of *G. metallireducens*. One of the widely used constraint-based modeling approaches is flux balance analysis, which assumes the system is in pseudo

steady state. An advantage of this technique is that kinetic rates for each individual metabolic reaction are not needed (25). In addition to the elemental balances, the approach imposes physicochemical constraints, such as reaction reversibility, and biomass component demands (reaction to produce biomass composition) and growth-associated energy requirements. Feasible limits of intracellular metabolic flux distribution can then be calculated by optimizing a cellular objective function (e.g., maximization of growth) consistent with the existing biomass and extracellular reaction fluxes. In the absence of kinetic information, the constraint-based metabolic modeling approach can calculate the individual reaction fluxes to determine the metabolic capabilities and metabolic flux distributions of genome-scale biochemical networks (16, 44, 58).

Mathematically, the dynamic material balance of the system can be written as follows:

$$\frac{dC_i}{dt} = \sum_{j=1}^{n} S_{ij} v_j \rightarrow \frac{dC_i}{dt} = S \cdot v \; (i = 1,m)$$

where C_i is the concentration of the ith compound, which could be an intracellular metabolite, biomass component, substrate, or product; v_j is the reaction flux (rate) of the jth reaction in the system; $S(m \times n)$ with m compounds and n reactions is the stoichiometric matrix of a metabolic reaction network; and v is a flux (rate) vector of all of the reactions. Assuming intracellular reactions rapidly equilibrate to changes in the extracellular environment, based on metabolic dynamics that are relatively quick compared to the rates of biogeochemical transport in the subsurface (14, 58), a system of mass balance for all of the metabolites can be represented in a pseudo-steady state as $S \cdot v = 0$. The mass balance equation is subject to reaction reversibility and a rate capacity constraint for each reaction in the form of $v_{j,min} \leq v_j \leq v_{j,max}$, where $v_{j,min}$ and $v_{j,max}$ are the lower and upper bounds of the rate. Because the equation is a simple linear combination of fluxes, linear optimization tools are used to efficiently find the optimal metabolic flux distribution within the allowable range of cellular capabilities, given a specific objective function. This approach differs from biogeochemical reaction solvers in which detailed kinetics is required for each reaction and a set of nonlinear equations needs to be solved.

(ii) Coupling constraint-based and reactive transport models. Genome-scale constraint-based models can be incorporated into reactive transport models using a look-up table (50) or direct call of a linear programming tool, such as CPLEX (ILOG, Inc., Chicago, IL), glpk, etc. (17), depending on computation power and how dynamic the system is. Direct coupling with the reactive transport simulator HYDROGEOCHEM (66) was used in here. In this model, only the TEAP reaction of iron reduction was replaced with the *in silico* model of *G. metallireducens*. The *in silico* model is solved with linear programming during each Newton-Raphson iteration of the chemistry solver. At each grid point in the reactive transport simulation domain, substrate uptake rates are constrained by concentrations of geochemical species of interest using kinetic rules, such as Michaelis-Menten kinetics, derived from experimental measurement (37). Outside the microbial cell walls, the concentrations of these substrates change during the bioremediation experiment through transport and biogeochemical reactions, such as aqueous complexation, surface complexation, and mineral precipitation/dissolution, which affects the uptake rates by the microorganisms. These uptake rates are then passed to the metabolic model as upper or lower bounds for linear optimization of the metabolic network depending on whether the substrate is limiting or not. The model outputs are fluxes of biogeochemical compound per gram dry weight (gdw) of *G. metallireducens* biomass (mmol/gdw/h) that exchange with the extracellular environment. When multiplied by the local biomass, these fluxes become sources and sinks of species in the chemistry solver of the reactive transport model. Details of the coupling can be found in Scheibe et al. (50) and Fang et al. (17).

(iii) Conceptual model and parameters. For this assessment of the *in silico* metabolic model of *G. metallireducens*, flow was modeled as steady and one-dimensional (1-D), based on the consistency of the south-southwest groundwater flow direction and magnitude of the hydraulic head gradient during the 2008 field experiment. More detailed treatment using 3-D variably saturated flow and reactive transport modeling of the 2008 Rifle IFRC field experiment can be found in Yabusaki et al. (65). For the 1-D modeling approach, the porosity, Darcy velocity, and dispersivity were calibrated from the measurement of bromide data to be 0.15, 0.079 m/day, and 0.67 m, respectively. The 1-D model represents the injection gallery as an upgradient boundary with the downgradient domain discretized with 120 uniform 0.1-m-long grid cells. At the upgradient boundary, the injectate was assumed to be completely mixed over the saturated thickness and injection gallery width.

Previous studies (18, 65) have developed a comprehensive reaction network for the field biostimulation experiment that included biologically mediated terminal electron acceptor processes for solid-phase Fe(III), aqueous U(VI), and aqueous sulfate; nonelectrostatic aqueous and surface complexation for Fe(II) and U(VI); calcite, siderite, FeS, and S secondary minerals; sulfide-promoted dissolution of Fe(III) minerals, cation exchange, protonation/deprotonation of Fe(III) surfaces; and major ion chemistry. The TEAP reaction for solid-phase Fe(III) in the reaction network was replaced by the constraint-based *in silico* model for *G. metallireducens*. The *in silico* model associates 747 genes and 697 reactions, which can be found in the supplementary information of Sun et al. (52). In the model, a specific gene can be involved in several different metabolic reactions. A first-order decay model of *G. metallireducens* was incorporated in the simulation. The decay rate was set to 0.15 day^{-1}. All other reactions and parameters were the same as in Yabusaki et al. (65) without recalibration. The uptake rates for acetate, Fe(III), and ammonium needed to constrain the *in silico* model were calculated using Michaelis-Menten kinetics. The parameters of the maximum rate (V_{max}) and the substrate saturation constant (K_s) followed those in Fang et al. (17). The V_{max} parameters were set to be 1.0 mmol/gdw/h for acetate, 10.0 mmol/gdw/h for Fe(III), and 0.05 mmol/gdw/h for NH_4. The K_s parameters were set to 0.01 mM for acetate, 1 mM for Fe(III), and 0.1 mM for NH_4. Initial *G. metallireducens* cell concentrations were set to 32 cells ml^{-1} based on the qPCR data measured for *Geobacter* species from the groundwater sample taken at well D02 2 days before the 2008 experiment (11).

RESULTS

Comparisons of model prediction and observed breakthrough curves for bromide, Fe(II), and acetate in the first row of downgradient monitoring wells are presented in Fig. 1a to c. The temporal and spatial variability is due to the nonuniform release of injectate from the injection gallery and preferential flow pathways between the first row of downgradient monitoring wells and the injection gallery. Despite the variability, the 1-D model reproduced the general behavior of bromide, Fe(II), and acetate during the Fe reduction phase and Fe-sulfate transition phase (Fig. 1a to c). The model predicted the enrichment of *Geobacter* strains during the Fe reduction phase and the reduction of growth during the Fe-sulfate transition phase. This general response is consistent with the qPCR data (Fig. 1d).

Mass spectrometry measurements of proteins and peptides recovered at the three time points during Fe(III) reduction were searched against predicted peptides from the *G. metallireducens* genome. In total, 637 proteins were detected using this technique; of this total, 180 proteins were associated with specific metabolic reactions in the *in silico* model (see Table S1 in the supplemental material). The majority of proteins not associated with *in silico* model reactions were involved in cell maintenance and protein synthesis (e.g., ribosomal proteins). Figure 2 classifies the *in silico*-matching protein identifications based on Clusters of Orthologous Groups (COG) protein categories. Proteomic results confirmed the activity of the majority of *in silico* pathways that were

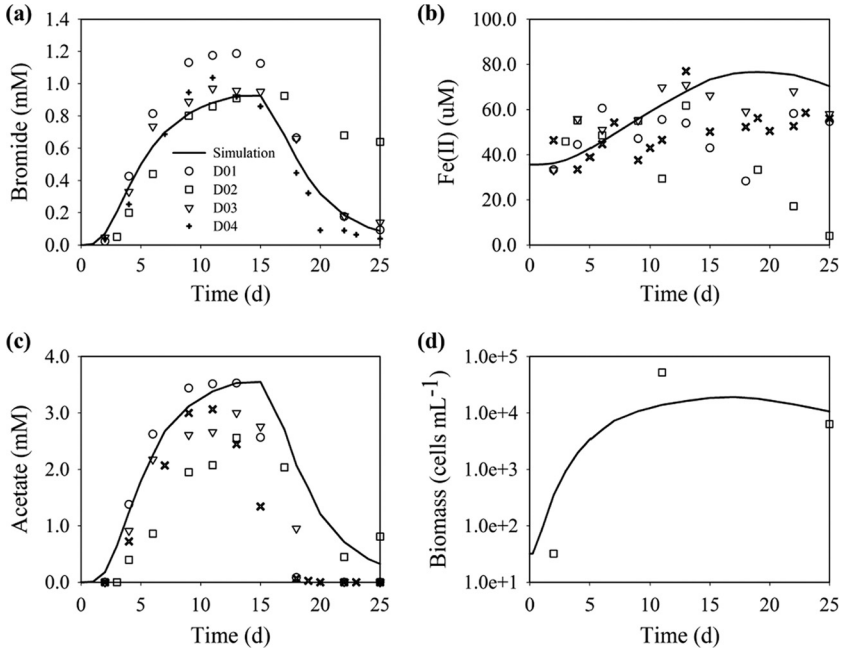

FIG 1 (a to c) Simulated (solid line) and observed (symbols) bromide breakthrough (a), Fe(II) breakthrough (b), and acetate breakthrough (c) at the first row of monitoring wells D01, D02, D03, and D04. (d) Comparison of *Geobacter* response from the simulation (solid line) and the qPCR data (open squares) at downgradient monitoring well D02 in Chandler et al. (11).

associated with "energy production and conversion" reactions (category C). In categories E (amino acid transport and metabolism), I (lipid transport and metabolism), and Q (secondary metabolites biosynthesis, transport, and catabolism), some discrepancies between the number of identified proteins and the number of reactions were observed. These results could be due to a number of reasons, i.e., (i) one metabolic reaction could associate with more than one enzyme and (ii) the category contains proteins that are either below the detection limit from proteomic analysis or were unidentified due to search database discrepancies. The discrepancy in category M (cell wall/membrane/envelope biosynthesis) also has two possible explanations: (i) these proteins associated with the *in silico* model also contribute to other synthesis-related processes that are not explicitly accounted for by the model and will require the development of additional genome-scale model capabilities (54) or (ii) some of the metabolic pathways were not accounted for by the *in silico* model (38, 52).

Following studies that have used proteomic data to constrain or infer metabolic flux rates (14, 67), we assumed that normalized protein abundances inferred from tandem mass spectrometry measurements (8, 9, 43) are proportional to the metabolic flux through a specific reaction. Within the data sets, there was good correspondence in general trends between *in silico* relative fluxes and protein abundances. Previous studies have indicated that up to 90% of carbon flux through *Geobacter* strains is directed through respiratory pathways rather than toward biosynthesis (57). Consistent with this observation, the highest *in silico* fluxes and protein abundances were observed for proteins associated with the activation of acetate (via an acetyl coenzyme A [acetyl-CoA] hydrolase/transferase enzyme) and the incorporation of acetyl-CoA into the tricarboxylic acid (TCA) cycle (via citrate synthase). In addition, malate dehydrogenase, which catalyzes the conversion of malate into oxaloacetate, is also one of the most abundant proteins within the data set and generated a high flux within the *in silico* model (Fig. 3). In some instances, differences exist between the *in silico* output and detected protein abundances; flux through the beta subunit of the 2-oxoglutarate ferredoxin oxidoreductases enzyme is seemingly underrepresented in the *in silico* model relative to protein abundances. Both the *in silico* fluxes and protein abundances remained at relatively stable levels over the course of the next 5 days, despite increasing acetate concentrations over this time period (Fig. 3).

Ammonium concentrations in well D04 were at least 118 μM during the first 10 days of biostimulation, which should not be limiting (41). Under these conditions, there was no *in silico* flux passing through nitrogen fixation pathways, which some *Geobacter* strains use for obtaining bioavailable nitrogen. However, this modeling result was in disagreement with experimental proteomic data, which indicated the expression of nitrogen-fixing nitrogenase enzymes by the *Geobacter* population. Nitrogen fixation is energetically expensive, so it is surprising that it occurs when ammonium should be available in excess. To investigate this behavior, the model was used to identify parameters that may reproduce the observations by altering the model parameters to ammonium limiting conditions. After the parameters were altered to account for limiting ammonium concentrations, the *in silico* model outputs predicted both flux through nitrogenase enzymes and evidence of increasing flux over time, presumably as any remaining ammonium is utilized and cells become more dependent on nitrogenase-fixed nitrogen for growth. Concurrent to this, the model predicts decreasing expression of an ammonium transporter as ammonium becomes scarce (Fig. 4). The environmental proteomics data supports these *in silico* fluxes, with some evidence for increasing nitrogenase activity across the three samples coupled to decreasing detection of an ammonium transporter

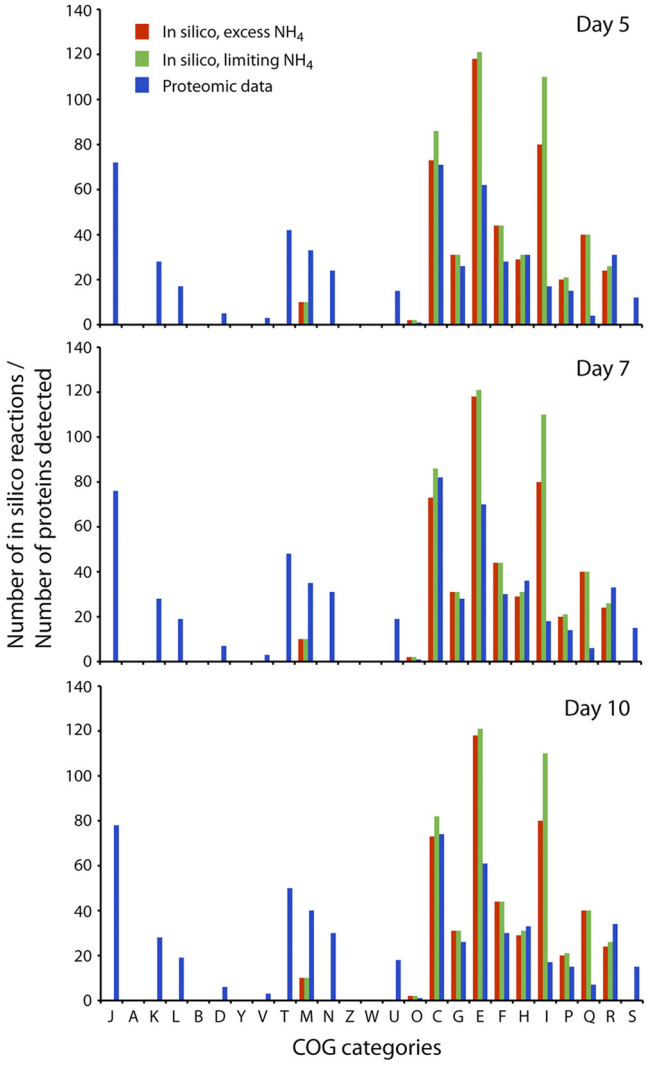

FIG 2 Distribution of number of *in silico* pathways and detected proteins into COG functional categories at days 5, 7, and 10. Columns: A, RNA processing and modification; B, chromatin structure and dynamics; C, energy production and conversion; D, cell cycle control, cell division, chromosome partitioning; E, amino acid transport and metabolism; F, nucleotide transport and metabolism; G, carbohydrate transport and metabolism; H, coenzyme transport and metabolism; I, lipid transport and metabolism; J, translation, ribosomal structure, and biogenesis; K, transcription; L, replication, recombination, and repair; M, cell wall/membrane/envelope biogenesis; N, cell motility; O, post-translational modification, protein turnover, chaperones; P, inorganic ion transport and metabolism; Q, secondary metabolites biosynthesis, transport, and catabolism; R, general function prediction only; S, function unknown; T, signal transduction mechanisms; U, intracellular trafficking, secretion, and vesicular transport; V, defense mechanisms; W, extracellular structures; Y, nuclear structure; Z, cytoskeleton.

(Fig. 4) (41). The development of potential ammonium-limiting microenvironments over the duration of biostimulation may require further investigations. Under ammonium-limiting conditions, more *in silico* pathways are active in categories C (energy production and conversion), E (amino acid transport and metabolism), and especially I (lipid transport and metabolism) (Fig. 2). These responses might be expected given that nitrogen fixation is an expensive process in terms of metabolic energy.

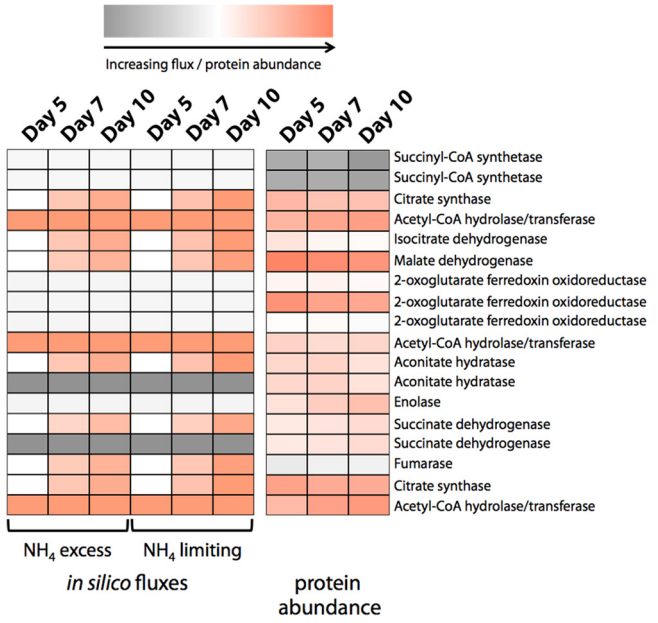

FIG 3 *In silico* flux and proteomic abundance data for TCA cycle reactions across the three samples.

DISCUSSION

This study was an initial attempt to assess the metabolic function roles within an *in silico* model by combining the model with environmental proteomic data collected from a field test site. The results showed that (i) the coupling of *in silico* model flux outputs with environmental proteomic data can be used for model validation and (ii) discrepancies between these two data types can be used to alter model parameters and potentially refine *in silico* models in the future. Proteomic data analysis showed that 180 of the 637 *G. metallireducens* proteins detected during the biostimulation experiment were associated with specific metabolic reactions in the genome-scale model of *G. metallireducens*. Although there are some discrepancies between predictions from the genome-scale *G. metallireducens* model and field proteomic observations, the most abundant proteins (e.g., acetyl-CoA transferase and citrate synthase) consistently match the highest metabolic reaction fluxes in the genome-scale model, with the result that the two data sets generally

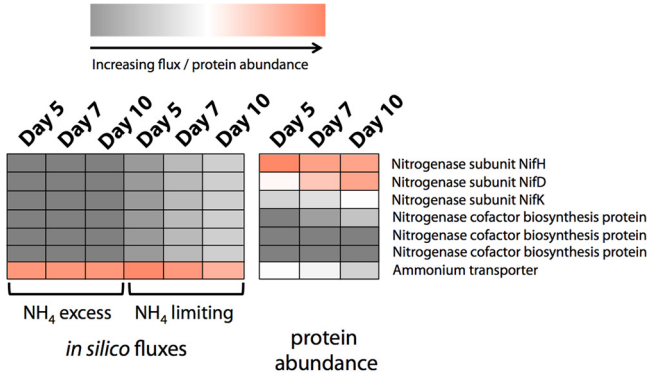

FIG 4 Comparison between *in silico* fluxes and proteomic data associated with nitrogen acquisition under both NH_4 excess and limiting conditions.

support each other. Given the metabolism of *Geobacter*-like strains that direct a large proportion of carbon flux through respiratory pathways, this result likely reflects the dominance of respiratory reactions in the *in silico* model, and associated high protein abundances. Conversely, for reactions exhibiting lower fluxes (e.g., biosynthesis pathways such as amino acid transport and metabolism), some discrepancies between the two data types may be at least partly associated with low protein abundances in experimental data and detection limits during tandem mass spectrometry analyses. Across a temporal scale, both *in silico* fluxes and protein abundances were relatively stable over the 5-day period of this experiment. Callister et al. (9) previously observed the inferred "stability" of microbial community members during secondary stimulation in the context of global proteome profiles, and suggested that these results were due to the rapid enrichment of dominant *Geobacter* strains to acetate amendment. By day 5 in this experiment, the *Geobacter* population may exist in a pseudo "steady-state" having already increased greatly in abundance by this time point. Although this response has been previously measured in the field (9), this is the first instance where *in silico* modeling data has supported these observations.

In some instances where discrepancies exist between the two different data sets, these observations have been used to alter the *in silico* model. For example, differences between predicted fluxes and measurements of enzymes associated with nitrogen utilization and fixation were used to infer processes occurring in the environment and subsequently alter certain parameters within the *in silico* model associated with ammonium concentrations. Given bulk geochemical measurements and the existing *in silico* model, the activity of nitrogen-fixing pathways would not have been anticipated without proteomic inferences. As these analyses become more complex, we anticipate that this feedback will become an important component of model validation when related to field systems. The identification of nitrogen-fixing activity offers an insight into another mechanism by which *Geobacter* strains are able to dominate these subsurface environments following acetate amendment (21a, 21b). While the efficient use of acetate by these strains has been previously documented (57, 61), the ability to actively obtain a potentially limiting element (N) may offer a competitive advantage over non-nitrogen-fixing microorganisms, as suggested by Zhuang et al. (68).

At a more general level, experimental-model discrepancies can be linked to a number of potential factors: (i) the genome-scale model constructed from *G. metallireducens* grown from pure culture is not representative of metabolism under field conditions, (ii) indigenous *Geobacter* strains may be distantly related to *G. metallireducens*, and (iii) the details of other major metabolic pathways were overlooked by the model. In addition, the genome-scale metabolic model is used to analyze the metabolic behavior of microorganisms as an individual cell and does not include mechanisms to handle external biological stresses (e.g., overpopulation or interaction with other organisms) or account for other microorganisms present in the aquifer that also utilize the electron donor and other key essential elements (26). Furthermore, protein extraction from biological membranes still poses technical challenges during global protein sample preparations and can sometimes result in missed identifications. Further development of the genome-scale model and the multiscale coupling should take into account these factors to better understand the relevant site processes, properties, and conditions and the metabolic functions of microorganisms in the context of engineered bioremediation.

The benefits of these developments may be significant. Given the ability to use high-throughput metagenomic sequencing data to reconstruct complete or near-complete genomes of indigenous environmental bacterial strains, and the availability of proteomic data to validate fluxes between reaction pathways, simplified "*in silico*" models may be constructed. Although not initially generating the high-resolution metabolic models that can be achieved using lab strains, this approach would allow modeling to be carried out using microorganisms that cannot currently be grown in laboratories and whose activity can only be inferred from metagenomic and proteomic data sets. The knowledge gained from this multiscale approach could be used to engineer electron donor (e.g., acetate), terminal electron acceptor (e.g., iron), and biogeochemical conditions that enhance the desired metabolic pathways of the target microorganism(s) to accomplish cost-effective uranium bioreduction and other desired processes.

ACKNOWLEDGMENTS

The work conducted at the Rifle Integrated Field Research Challenge (IFRC) site in Rifle, CO, is supported by the Subsurface Biogeochemical Research Program in the Climate and Environmental Sciences Division of the Office of Biological and Environmental Research at the U.S. Department of Energy Office of Science. Access to the site is in accordance with the city of Rifle and is managed under the DOE Long-Term Surveillance and Monitoring Program. Portions of this study were performed in the Environmental Molecular Sciences Laboratory, a DOE/BER national scientific user facility at the Pacific Northwest National Laboratory. The Pacific Northwest National Laboratory is operated by Battelle for the U.S. Department of Energy under contract DE-AC05-76RL01830.

REFERENCES

1. **Aklujkar M, et al.** 2010. The genome of *Geobacter bemidjiensis*, exemplar for the subsurface clade of *Geobacter* species that predominate in Fe(III)-reducing subsurface environments. BMC Genomics **11**:490. doi:10.1186/1471-2164-11-490.
2. **Anderson RT, et al.** 2003. Stimulating the *in situ* activity of *Geobacter* species to remove uranium from the groundwater of a uranium-contaminated aquifer. Appl. Environ. Microbiol. **69**:5884–5891.
3. **Aniszewski A.** 2011. Particular application of a mathematical transport model incorporating sub-surface reactive pollutants. Acta Geophys. **59**:110–123.
4. **Baek KH, et al.** 2007. Monitoring of microbial diversity and activity during bioremediation of crude OH-contaminated soil with different treatments. J. Microbiol. Biotechnol. **17**:67–73.
5. **Baker MD, Wolanin PM, Stock JB.** 2006. Signal transduction in bacterial chemotaxis. Bioessays **28**:9–22.
6. **Balaban NQ, Merrin J, Chait R, Kowalik L, Leibler S.** 2004. Bacterial persistence as a phenotypic switch. Science **305**:1622–1625.
7. **Brown LR.** 2010. Microbial enhanced oil recovery (MEOR). Curr. Opin. Microbiol. **13**:316–320.
8. **Callister SJ, et al.** 2006. Normalization approaches for removing systematic biases associated with mass spectrometry and label-free proteomics. J. Proteome Res. **5**:277–286.
9. **Callister SJ, et al.** 2010. Analysis of biostimulated microbial communities from two field experiments reveals temporal and spatial differences in proteome profiles. Environ. Sci. Technol. **44**:8897–8903.
10. **Chambon JC, Broholm MM, Binning PJ, Bjerg PL.** 2010. Modeling multi-component transport and enhanced anaerobic dechlorination processes in a single fracture-clay matrix system. J. Contam. Hydrol. **112**:77–90.
11. **Chandler DP, et al.** 2010. Monitoring microbial community structure and dynamics during in situ U(VI) bioremediation with a field-portable microarray analysis system. Environ. Sci. Technol. **44**:5516–5522.
12. **Chang YJ, et al.** 2005. Microbial incorporation of C-13-labeled acetate at

the field scale: detection of microbes responsible for reduction of U(VI). Environ. Sci. Technol. **39**:9039–9048.
13. **Chung BKS, et al.** 2010. Genome-scale metabolic reconstruction and *in silico* analysis of methylotrophic yeast *Pichia pastoris* for strain improvement. Microb. Cell Fact. **9**:50.
14. **Colijin C, et al.** 2009. Interpreting expression data with metabolic flux models: predicting *Mycobacterium tuberculosis* mycolic acid production. PLoS Comput. Biol. **5**:e1000489. doi:10.1371/journal.pcbi.1000489.
15. **Durot M, Bourguignon PY, Schachter V.** 2009. Genome-scale models of bacterial metabolism: reconstruction and applications. FEMS Microbiol. Rev. **33**:164–190.
16. **Edwards JS, Covert M, Palsson B.** 2002. Metabolic modeling of microbes: the flux-balance approach. Environ. Microbiol. **4**:133–140.
17. **Fang YL, et al.** 2011. Direct coupling of a genome-scale microbial *in silico* model and a groundwater reactive transport model. J. Contam. Hydrol. **122**:96–103.
18. **Fang YL, Yabusaki SB, Morrison SJ, Amonette JP, Long PE.** 2009. Multicomponent reactive transport modeling of uranium bioremediation field experiments. Geochim. Cosmochim. Acta **73**:6029–6051.
19. **Finneran KT, Anderson RT, Nevin KP, Lovley DR.** 2002. Potential for bioremediation of uranium-contaminated aquifers with microbial U(VI) reduction. Soil Sediment Contam. **11**:339–357.
20. **Greskowiak J, Prommer H, Massmann G, Nutzmann G.** 2006. Modeling seasonal redox dynamics and the corresponding fate of the pharmaceutical residue phenazone during artificial recharge of groundwater. Environ. Sci. Technol. **40**:6615–6621.
21. **Holmes DE, Finneran KT, O'Neill RA, Lovley DR.** 2002. Enrichment of members of the family *Geobacteraceae* associated with stimulation of dissimilatory metal reduction in uranium-contaminated aquifer sediments. Appl. Environ. Microbiol. **68**:2300–2306.
21a. **Holmes DE, et al.** 2008. Transcriptome of Geobacter uraniireducens growing in uranium-contaminated subsurface sediments. ISME J. **3**:216–230.
21b. **Holmes DE, Nevin KP, Lovley DR.** 2004. In situ expression of *nifD* in *Geobacteraceae* in subsurface sediments. Appl. Environ. Microbiol. **70**:7251–7259.
22. **Hunter KS, Wang YF, Van Cappellen P.** 1998. Kinetic modeling of microbially driven redox chemistry of subsurface environments: coupling transport, microbial metabolism and geochemistry. J. Hydrol. **209**:53–80.
23. **Jenneman GE, Knapp RM, McInerney MJ, Menzie DE, Revus DE.** 1984. Experimental studies of in situ microbial enhanced oil-recovery. SPE J. **24**:33–37.
24. **Kaszycki P, Petryszak P, Pawlike M, Koloczek H.** 2011. Ex situ bioremediation of soil polluted with oily waste: the use of specialized microbial consortia for process bioaugmentation. Ecol. Chem. Eng. S **18**:83–92.
25. **Kauffman KJ, Prakash P, Edwards JS.** 2003. Advances in flux balance analysis. Curr. Opin. Biotechnol. **14**:491–496.
26. **Kerkof LJ, Williams KH, Long PE, McGuinness LR.** 2011. Phase preference by active, acetate-utilizing bacteria at the Rifle, CO Integrated Field Research Challenge site. Environ. Sci. Technol. **45**:1250–1256.
27. **King EL, Tuncay K, Ortoleva P, Meile C.** 2009. *In silico Geobacter sulfurreducens* metabolism and its representation in reactive transport models. Appl. Environ. Microbiol. **75**:83–92.
28. **Kjeldsen KR, Nielsen J.** 2009. *In silico* genome-scale reconstruction and validation of the *Corynebacterium glutamicum* metabolic network. Biotechnol. Bioeng. **102**:583–597.
29. **Knabe S.** 1984. Overview of microbial enhanced oil recovery. Oil Gas J. **82**:59–60.
30. **Korotkevych O, et al.** 2011. Functional adaptation of microbial communities from jet fuel-contaminated soil under bioremediation treatment: simulation of pollutant rebound. FEMS Microbiol. Ecol. **78**:137–149.
31. **Lee J, Yun H, Feist AM, Palsson BO, Lee SY.** 2008. Genome-scale reconstruction and *in silico* analysis of the *Clostridium acetobutylicum* ATCC 824 metabolic network. Appl. Microbiol. Biotechnol. **80**:849–862.
32. **Lewis NE, et al.** 2010. Omic data from evolved E-coli are consistent with computed optimal growth from genome-scale models. Mol. Syst. Biol. **6**:390.
33. **Li JA, et al.** 2011. Interactions of microbial-enhanced oil recovery processes. Transport Porous Media **87**:77–104.
34. **Li L, Steefel CI, Kowalsky MB, Englert A, Hubbard SS.** 2010. Effects of physical and geochemical heterogeneities on mineral transformation and biomass accumulation during biostimulation experiments at Rifle, Colorado. J. Contam. Hydrol. **112**:45–63.
35. **Li L, Steefel CI, Williams KH, Wilkins MJ, Hubbard SS.** 2009. Mineral transformation and biomass accumulation associated with uranium bioremediation at Rifle, Colorado. Environ. Sci. Technol. **43**:5429–5435.
36. **Loew LM, Schaff JC.** 2001. The Virtual Cell: a software environment for computational cell biology. Trends Biotechnol. **19**:401–406.
37. **Mahadevan R, et al.** 2006. Characterization of metabolism in the Fe(III)-reducing organism *Geobacter sulfurreducens* by constraint-based modeling. Appl. Environ. Microbiol. **72**:1558–1568.
38. **Mahadevan R, et al.** 2008. Characterizing regulation of metabolism in *Geobacter sulfurreducens* through genome-wide expression data and sequence analysis. OMICS **12**:33–59.
39. **Mayer KU, Frind EO, Blowes DW.** 2002. Multicomponent reactive transport modeling in variably saturated porous media using a generalized formulation for kinetically controlled reactions. Water Resour. Res. **38**:1174.
40. **Molnaa BA, Grubbs RB.** 1989. Bioremediation of petroleum contaminated soils using a microbial consortia as inoculums, p 219–232. *In* Calabrese EJ, Kostecki PT (ed), Petroleum contaminated soils, vol 2. Lewis Publishers, Chelsea, MI.
41. **Mouser PJ, et al.** 2009. Influence of heterogeneous ammonium availability on bacterial community structure and the expression of nitrogen fixation and ammonium transporter genes during in situ bioremediation of uranium-contaminated groundwater. Environ. Sci. Technol. **43**:4386–4392.
42. **Nielsen SM, Shapiro AA, Michelsen ML, Stenby EH.** 2010. 1D simulations for microbial enhanced oil recovery with metabolite partitioning. Transport Porous Media **85**:785–802.
43. **Polpitiya AD, et al.** 2008. DAnTE: a statistical tool for quantitative analysis of -omics data. Bioinformatics **24**:1556–1558.
44. **Price ND, Papin JA, Schilling CH, Palsson BO.** 2003. Genome-scale microbial *in silico* models: the constraints-based approach. Trends Biotechnol. **21**:162–169.
45. **Reed JL, Vo TD, Schilling CH, Palsson BO.** 2003. An expanded genome-scale model of *Escherichia coli* K-12 (iJR904 GSM/GPR). Genome Biol. **4**:R54.
46. **Resat H, Bailey V, McCue LA, Konopka A.** 2011. Modeling microbial dynamics in heterogeneous environments: growth on soil carbon sources. Microb. Ecol. **63**:883–897. doi:10.1007/s00248-011-9965-x.
47. **Ros M, Rodriguez I, Garcia C, Hernandez T.** 2010. Microbial communities involved in the bioremediation of an aged recalcitrant hydrocarbon polluted soil by using organic amendments. Bioresour. Technol. **101**:6916–6923.
48. **Salvage KM, Yeh GT.** 1998. Modeling chemically and microbiologically reactive contaminant transport in porous media, p 263–270. *In* Burganos VN, et al (ed), Computational methods in water resources XII, vol 1. Computational methods in contamination and remediation of water resources. WIT Press, Wessex Institute of Technology, Southampton, United Kingdom.
49. **Schafer D, Schafer W, Kinzelbach W.** 1998. Simulation of reactive processes related to biodegradation in aquifers: 1. Structure of the three-dimensional reactive transport model. J. Contam. Hydrol. **31**:167–186.
50. **Scheibe TD, et al.** 2009. Coupling a genome-scale metabolic model with a reactive transport model to describe in situ uranium bioremediation. Microb. Biotechnol. **2**:274–286.
51. **Sinha S, et al.** 2009. Microbial transformation of xenobiotics for environmental bioremediation. Afr. J. Biotechnol. **8**:6016–6027.
52. **Sun J, et al.** 2009. Genome-scale constraint-based modeling of *Geobacter metallireducens*. BMC Syst. Biol. **3**:15. doi:10.1186/1752-0509-3-15.
53. **Sun Y, Petersen JN, Bear J, Clement TP, Hooker BS.** 1999. Modeling microbial transport and biodegradation in a dual-porosity system. Transport Porous Media **35**:49–65.
54. **Thiele I, Jamshidi N, Fleming RMT, Palsson BO.** 2009. Genome-scale reconstruction of *Escherichia coli*'s transcriptional and translational machinery: a knowledge base, its mathematical formulation, and its functional characterization. PLoS Comput. Biol. **5**:e1000312. doi:10.1371/journal.pcbi.1000312.
55. **Thiele I, Vo TD, Price ND, Palsson BO.** 2005. Expanded metabolic reconstruction of *Helicobacter pylori* (iIT341 GSM/GPR): an *in silico* genome-scale characterization of single- and double-deletion mutants. J. Bacteriol. **187**:5818–5830.
56. **Thullner M, Van Cappellen P, Regnier P.** 2005. Modeling the impact of microbial activity on redox dynamics in porous media. Geochim. Cosmochim. Acta **69**:5005–5019.
57. **Tang YJ, et al.** 2007. Flux analysis of central metabolic pathways in *Geobacter metallireducens* during reduction of soluble Fe(III)-nitrilotriacetic acid. Appl. Environ. Microbiol. **73**:3859–3864.

58. **Varma A, Palsson BO.** 1994. Metabolic flux balancing: basic concepts, scientific and practical use. Nat. Biotechnol. **12**:994–998.
59. **Vrionis HA, et al.** 2005. Microbiological and geochemical heterogeneity in an in situ uranium bioremediation field site. Appl. Environ. Microbiol. **71**:6308–6318.
60. **Widiastuti H, et al.** 2011. Genome-scale modeling and *in silico* analysis of ethanologenic bacteria *Zymomonas mobilis*. Biotechnol. Bioeng. **108**:655–665.
61. **Wilkins MJ, et al.** 2009. Proteogenomic monitoring of *Geobacter* physiology during stimulated uranium bioremediation. Appl. Environ. Microbiol. **75**:6591–6599.
62. **Williams KH, et al.** 2011. Acetate availability and its influence on sustainable bioremediation of uranium-contaminated groundwater. Geomicrobiol. J. **28**:519–539.
63. **Woo SH, Park JM.** 2004. Microbial degradation and enhanced bioremediation of polycyclic aromatic hydrocarbons. J. Ind. Eng. Chem. **10**:16–23.
64. **Yabusaki SB, et al.** 2007. Uranium removal from groundwater via in situ biostimulation: field-scale modeling of transport and biological processes. J. Contam. Hydrol. **93**:216–235.
65. **Yabusaki SB, et al.** 2011. Variably saturated flow and multicomponent biogeochemical reactive transport modeling of a uranium bioremediation field experiment. J. Contam. Hydrol. **126**:271–290.
66. **Yeh GT, et al.** 2010. Numerical modeling of coupled fluid flow and thermal and reactive biogeochemical transport in porous and fractured media. Comput. Geosci. **14**:147–170.
67. **Yizhak K, Benyamini T, Liebermeister W, Ruppin E, Shlomi T.** 2010. Integrating quantitative proteomics and metabolomics with a genome-scale metabolic network model. Bioinformatics **26**:i255–i260.
68. **Zhuang K, et al.** 2011. Genome-scale dynamic modeling of the competition between Rhodoferax and Geobacter in anoxic subsurface environments. ISME J. **5**:305–316.