backbone–backbone ineractions. *J. Phys. Chem. B*, 2006, **110**, 563–578.

30. Churchill, C. D. M., Navarro-Whyte, L., Rutledge, L. R. and Wetmore, S. D., Effects of the biological backbone on DNA–protein stacking interactions. *Phys. Chem. Chem. Phys.*, 2009, **11**, 10657–10670.
31. Churchill, C. D. M., Rutledge, L. R. and Wetmore, S. D., Effects of the biological backbone on stacking interactions at DNA-protein interfaces: the interplay between the backbone–π and π–π components. *Phys. Chem. Chem. Phys.*, 2010, **12**, 14515–14526.
32. Parker, T. M., Hohenstein, E. G., Parrish, R. M., Hud, N. V. and Sherrill, C. D., Quantum-mechanical analysis of the energetic contributions to π-stacking in nucleic acids versus rise, twist and slide. *J. Am. Chem. Soc.*, 2013, **135**, 1306–1316.
33. Wang, S. and Kool, E. C., Origins of the large differences in stability of DNA and RNA helixes: C-5 methyl and 2′-hydroxyl effects. *Biochemistry*, 1995, **34**, 4125–4132.
34. Šponer, J. *et al.*, The DNA and RNA sugar–phosphate backbone emerges as the key player. An overview of quantum-chemical, structural biology and simulation studies. *Phys. Chem. Chem. Phys.*, 2012, **14**, 15257–15277.
35. Pavelka, M. and Burda, J. V., Pt-bridges in various single-strand and double-helix DNA sequences. DFT and MP2 study of the cisplatin coordination with guanine, adenine, and cytosine. *J. Mol. Model.*, 2007, **13**, 367–379.
36. Sherrill, C. D., Computations of non-covalent π-interactions. In *Reviews in Computational Chemistry* (eds Lipkowitz, K. B. and Cundari, T. A.), John Wiley, New Jersey, 2009, vol. 26, pp. 1–38.
37. Tschumper, G. S., Reliable electronic structure computations for weak noncovalent interactions in clusters. In *Reviews in Computational Chemistry* (eds Lipkowitz, K. B. and Cundari, T. A.), John Wiley, New Jersey, 2009, vol. 26, pp. 39–90.
38. Grimme, S., Accurate description of van der Waals complexes by density functional theory including empirical corrections. *J. Comput. Chem.*, 2004, **25**, 1463–1473.
39. Grimme, S., Semiempirical hybrid density functional with perturbative second order correlation. *J. Chem. Phys.*, 2006, **124**, 034108.
40. Zhao, Y. and Truhlar, D. G., The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: two new functionals and systematic testing of four M06-class functionals and 12 other functionals. *Theor. Chem. Acc.*, 2008, **120**, 215–241.
41. Grimme, S., Density functional theory with London dispersion corrections. *WIREs Comput. Mol. Sci.*, 2011, **1**, 211–228.
42. Grimme, S., Semiempirical GGA-type density functional constructed with a long-range dispersion correction. *J. Comput. Chem.*, 2006, **27**, 1787–1799.
43. Grimme, S., Antony, J., Ehrlich, S. and Krieg, H., A consistent and accurate *ab initio* parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *J. Chem. Phys.*, 2010, **132**, 154104.
44. dftd3 program; http://toc.uni-muenster.de/DFTD3/
45. Goerigk, L., Kruse, H. and Grimme, S., Benchmarking density functional methods against the S66 and S66x8 datasets for noncovalent interactions. *Chem. Phys. Chem.*, 2011, **12**, 3421–3433.
46. Shui, X., Sines, C. C., McFail-Isom, L., VanDerveer, D. and Williams, L. D., Structure of the potassium form of CGCGAATTCGCG: DNA deformation by electrostatic collapse around inorganic cations. *Biochemistry*, 1998, **37**, 16877–16887.
47. Eichkorn, K., Treutler, O., Öhm, H., Häser, M. and Ahlrichs, R., Auxiliary basis sets to approximate Coulombic potentials. *Chem. Phys. Lett.*, 1995, **240**, 283–290.
48. Ahlrichs, R., Bär, M., Häser, M., Horn, H. and Kölmel, C., Electronic structure calculations on workstation computers: the program system turbomole. *Chem. Phys. Lett.*, 1989, **162**, 165–169.
49. Zheng, G., Lu, X.-J. and Olson, W. K., Web 3DNA – a web server for the analysis, reconstruction, and visualization of three-dimensional nucleic-acid structures. *Nucleic Acids Res.*, 2009, **37** (web server issue), W240-W246.
50. Biegler-König, F., Schönbohm, J., Derdau, R., Bayles, D. and Bader, R. F. W., AIM 2000, Version 1, Bielefeld, Germany, 2000.
51. Parthasarathi, R., Subramanian, V. and Sathyamurthy, N., Hydrogen bonding without borders: an atoms-in-molecules perspective. *J. Phys. Chem. A*, 2006, **110**, 3349–3351.

ACKNOWLEDGEMENTS. N.S. is an Honorary Professor at the Jawaharlal Nehru Centre for Advanced Scientific Research, Bengaluru. N.S. thanks the Department of Science and Technology, New Delhi for the J.C. Bose National Fellowship.

Received 19 June 2014; revised accepted 13 January 2015

In silico prediction of gene knockout candidates in *Escherichia coli* genome-scale model for enhanced succinic acid production from glycerol

Bashir Sajo Mienda[1,*], Mohd Shahir Shamsir[1] and Rosli Md Illias[2]

[1]Bioinformatics Research Group, Biosciences and Health Sciences Department, Faculty of Biosciences and Medical Engineering, and [2]Department of Bioprocess Engineering, Faculty of Chemical Engineering, Universiti Teknologi Malaysia, 81310 Skudai, Johor, Malaysia

The use of genome-scale models of *Escherichia coli* to guide future metabolic engineering strategies for increased succinic acid production has received renewed attention in recent years. Substrate selectivity such as glycerol is of particular interest, because it is currently generated as a by-product of biodiesel industry and therefore can serve as a solitary carbon source. However, study on the prediction of gene knockout candidates for enhanced succinate production from glycerol using Minimization of Metabolic Adjustment Algorithm with the OptFlux software platform remained underexplored. Here, we show that metabolic engineering interventions by gene knockout simulation of some pyruvate dissimilating pathway enzymes (lactate dehydrogenase A and pyruvate formate lyase A) using *E. coli* genome-scale model can reduce acetate flux and enhance succinic acid production under anaerobic conditions. The introduced genetic perturbations led to substantial improvement in succinate flux of about 597% on glycerol and 120% on glucose than that of the wild-type control strain BSKO. We hypothesize that the deletion of pyruvate formate

*For correspondence. (e-mail: bsmienda@gmail.com)

lyase A (*pflA*) in *E. coli* can led to no acetate production from glucose, lower acetate production from glycerol and increased succinic acid productivities on both substrates under anaerobic conditions. Our results demonstrate a predicted increase in succinate production (597% higher than the wild-type model) among others, from glycerol after deletion of *pflA*/*b0902* gene in *E. coli* genome-scale model. This would open up a novel platform for model-guided experimental inquiry and/or allow a comprehensive biological discovery on the metabolic processes of *pflA* in *E. coli* for succinate production when glycerol is the substrate.

Keywords: *Escherichia coli*, genome-scale model, gene knockout simulation, metabolic engineering, OptFlux software, succinic acid.

SUCCINIC acid has been established to be one among the fermentation products of anaerobic metabolism in addition to being an intermediate of the tricarboxylic acid cycle[1]. The invaluable uses of succinic acid as a precursor for various chemicals, an iron chelator and a supplement to many pharmaceuticals have been recognized[2]. Previously, succinic acid was mostly produced using a chemical approach from maleic anhydride[1]. Recently, much research attention has been devoted to the fermentative production of succinic acid, as several bacteria can now produce the acid as major fermentation product. However, large-scale production suffered some challenges because of complex physiology and metabolism of some obligate anaerobe producers and lack of adequate genetic knowledge for advanced systems metabolic engineering.

Escherichia coli has been established to produce mixed acid fermentation end-products with a small amount of succinic acid among others[3,4]. The mixed acid fermentation products of *E. coli* are more of acetic acid, formic acid, lactic acid and ethanol, rather than succinic acid[1]. Thus, it is deemed necessary to redirect metabolic fluxes from increasing succinic acid production by knocking out competing pathway genes interfering with reactions that facilitate succinic acid production.

Systems metabolic engineering of *E. coli* has received renewed attention in recent times because of the availability of genome-scale metabolic models[5–7]. These models were found to predict reasonably well accurate growth rates, metabolite excretion rates and growth phenotype on a number of substrates and genetic conditions that are consistent with established experimental observations[6,8,9]. Palsson and co-workers recently reported that prediction of microbial growth in relation to single and double gene knockouts using genome-scale models coupled with the integration of genomic and biochemical data would enable large-scale prediction of cellular functions[10]. This could open an opportunity for model-guide experimental inquiry and biological discovery[10]. These efforts have stimulated the development of several computational tools/software to study the *E. coli* system *in silico*. It was previously reported elsewhere[11] that synthetic microbiology and computational breakthrough could be synergistically combined to improve strain performance for increased ethanol production[11]. One of the notable examples of computational breakthroughs for metabolic engineering community with various applications is a software platform called OptFlux[12]. It has been developed and implemented with a genome scale-metabolic model of *E. coli* to predict the phenotype simulation of both wild-type and mutant strain using the method of flux balance analysis (FBA). We previously demonstrated the use of this approach for metabolic engineering interventions of *E. coli* for increased ethanol production from glucose[13], glycerol and xylose[14]. In a similar study, metabolic gene knockout of enhanced D-lactate production has been reported[15]. The software is characterized with plug-in architecture, where an algorithm called Minimization of Metabolic Adjustment (MOMA) was plugged in to simulate mutant whole-cell behaviour after genetic perturbation or gene knockout[12,16].

Few studies have reported the use of computational tools such as Opt Knock[17], for studying a number of gene deletions to increase the production of lactic acid using *E. coli* genome-scale model. Another study by Yim *et al.*[18] combined constrained based modelling, Opt Knock[17] and biosynthetic prediction algorithms[19] in guiding metabolic engineering interventions for the production of an important commodity chemical called 1,4 butanediol (BDO) in *E. coli*[18]. It achieved a production rate of 18 g l^{-1} of this highly reduced, non-natural chemical from renewable carbohydrate feedstock[18]. In addition, Lee *et al.*[20] have reported the use of MetaFluxNet for *in silico* comparative genome-scale metabolic interventions in combination with experimental validation for succinic acid production[1]. On the other hand, Zhang *et al.*[21] have reported re-engineering of *E. coli* for increased succinate production from mineral salts medium with inactivation of pyruvate formate lyase B (*pflB*). In a similar study by the same group, glycerol was chosen as the substrate for metabolic engineering to increase succinate production in *E. coli*[22]. Their metabolic engineering interventions focused on the inactivation of *pcK*, *ptsI* and *pflB* in *E. coli* using the glycerol substrate. This indicates the significance of substrate selectivity to increase succinate production using engineered *E. coli* strains. Until now, study on the *in silico* deletion of *pflA*/*b0902* using *E. coli* genome-scale model from glycerol for increased succinate production remained largely underexplored. We report here a predicted increase in succinate production (597% higher than the wild-type model) among others, from glycerol after deletion of *pflA*/*b0902* gene in *E. coli* genome-scale model using the OptFlux software platform.

The metabolic reconstruction of *E. coli* iAF1260 was used as a model[6] for all the wild-type and mutant strains described herein. The model was previously tested and validated against experimental data, and was shown to be

capable of predicting accurate growth rates, metabolite excretion rates and a growth phenotypes on a number of substrates and genetic conditions[6,8,23]. The substrates used in this study are glucose and glycerol, unless otherwise stated.

OptFlux software[12], as an open source platform (www.optflux.org) and a reference computational tool for metabolic engineering was used for FBA. MOMA was used as a simulation method for gene knockouts; and it was implemented using Java programing within the framework of OptFlux. All simulations of the mutant strains and wild-type models were performed using the OptFlux v3.06.

The chosen solitary carbon sources were glucose and glycerol, and the uptake rates for each carbon source were constrained to a maximum of 10 mmol g DW^{-1} h^{-1}. The oxygen uptake rate was considered to be 0.0 mmol g DW^{-1} h^{-1}, as the simulation condition was anaerobic for fermentative production of succinate. These values were chosen based on slightly close experimental observation of anaerobic growth of *E. coli*[9,24,25].

Gene knockout simulation was conducted under the OptFlux software platform using MOMA[16] as simulation method. FBA was used for simulation of the wild-type model, which predicts metabolic flux distributions at steady state using linear programing, whereas MOMA employs quadratic programing to identify the point in the flux space, which is closest to the wild-type point and consistent with the gene knockout constraint[16]. The wild-type model obtained from the Biomodels database[26], constructed by Feist *et al.*[6] was designated as BSK and the mutant model/strain with pyruvate formate lyase (*b0902*/*pflA*) single gene knockout was designated as BSK101 (Δ*flA*). The mutant model/strain with lactate dehydrogenase (*b1380*/*ldhA*) single gene knockout was designated as BSK102 (Δ*ldhA*). A mutant model with double knockout of both the *pflA* and *ldhA* was designated as BSK103 (Δ*pflA* Δ*ldhA*). The *in silico* gene knockout simulations were run to completion using MOMA, as previously described in their original documentation[16].

In this study, the *E. coli* genome-scale model constructed elsewhere[6] was metabolically engineered using the OptFlux software platform to convert glycerol to value-added succinate under anaerobic condition. Competing pathways genes such as lactate dehydrogenase (*ldhA)* and pyruvate formate lyase (*pflA*) were knocked out to see whether succinate flux can be increased. Two substrates (glucose and glycerol) were used in this *in silico* study. Deletion of *pflA* on glucose in strain BSK101 (Δ*flA*) led to an increase in succinate production of 6.9% higher than that of the wild-type strain BSKO, and maintaining a growth rate that 98.15% of the wild-type (Table 1 and Figure 1). The same mutant strain BSK101 (Δ*flA)* was simulated using glycerol as the substrate, showing a dramatic increase of succinate production with about 479% higher (fourfold) than that of the wild-type control strain BSKO (Table 2 and Figure 2). A growth rate of about 6.24% of the wild-type was seen when glycerol was the substrate (Table 2 and Figure 2). The slow growth rate might be attributed to the fact that the hydrolysate is not rich in xylose and mannose, which have been previously established to have effects on both cell growth and solvent production[27,28].

The slow growth rate of *E. coli* when glycerol is the substrate could be improved using co-substrates during fermentation such as glucose and glycerol as reported by Jin *et al.*[28]. In addition, other alternative strategies to overcome the anaerobic barrier entail the generation of enough biomass under aerobic conditions, and then switching into anaerobic conditions for succinate production[29]. This strategy was shown to be effective using a 'dual phase' fermentation system[30]. Another recent strategy reported[31] that an *E. coli* strain carrying mutations in the *rpoC* and *glpK* genes which was evolved from adaptive laboratory evolution in glycerol is capable of showing fast growth rate of about 89% of the wild-type[31]. This strategy could be useful in addressing the aforementioned

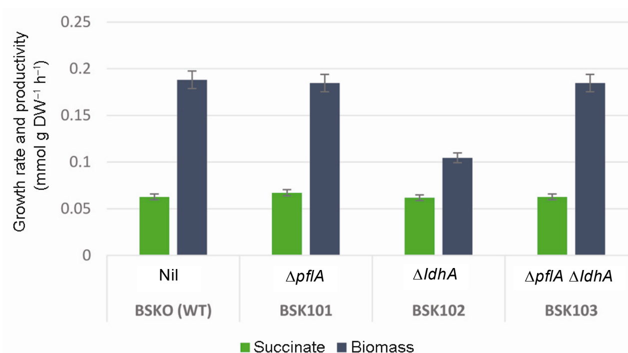

Figure 1. Maximum specific growth rate of *E. coli* BSK (WT) and mutants simulated using glucose as substrate with respect to succinate productivity. Error bars indicate standard deviation of the replicates.

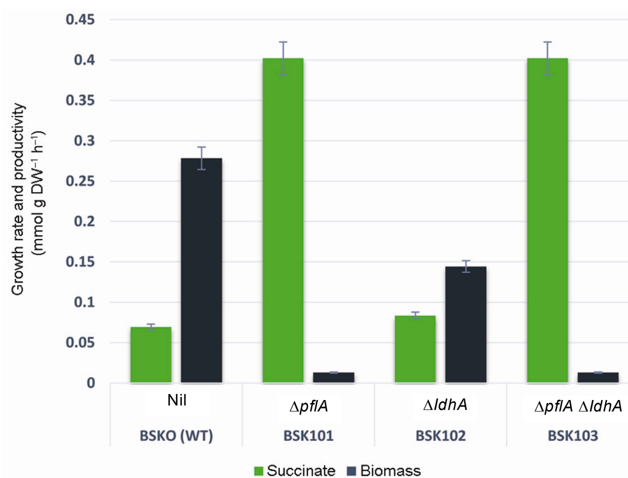

Figure 2. Maximum specific growth rates of *E. coli* BSK (WT) and mutants simulated using glycerol as substrate with respect to succinate productivity. Error bars indicate standard deviation of the replicates.

Table 1. *E. coli* strain design properties from glucose under the OptFlux software platform

Strain	Knockout genes	Succinate (mmol g DW^{-1} h^{-1})	% Succinate	Acetate (mmol g DW^{-1} h^{-1})	% Acetate	Biomass (h^{-1})	% Biomass
BSKO (WT)	Nil	0.06279	100.0	8.61685	100	0.18814504	100.0
BSK101	Δ*pflA*	0.06716	106.9	0.0	0.0	0.18466564	98.15
BSK102	Δ*ldhA*	0.06186	98.5	8.61453	99.97	0.10457363	55.58
BSK103	Δ*pflA* Δ*ldhA*	0.06279	106.9	8.61147	99.93	0.18466564	98.15

Maximum uptake rates for glycerol were set to be 10 mmol g DW^{-1} $h^{-1}$and the corresponding oxygen uptake rate was 0.0 mmol g DW^{-1} h^{-1} for anaerobic simulation.

Table 2. *E. coli* strain design properties from glycerol under the OptFlux software platform

Strain	Knockout genes	Succinate (mmol g DW^{-1} h^{-1})	% Succinate	Acetate (mmol g DW^{-1} h^{-1})	% Acetate	Biomass (h^{-1})	% Biomass
BSKO (WT)	Nil	0.06936	100.0	16.4721	100	0.2783578	100.0
BSK101	Δ*pflA*	0.40221	579.8	15.85774	96.27	0.012968043	6.239
BSK102	Δ*ldhA*	0.08347	120.3	16.47132	99.99	0.14426926	69.41
BSK103	Δ*pflA* Δ*ldhA*	0.40221	579.8	15.85774	96.7	0.012968043	6.239

Maximum uptake rates for glycerol were set to be 10 mmol g DW^{-1} h^{-1} and the corresponding oxygen uptake rate was 0.0 mmol g DW^{-1} h^{-1} for anaerobic simulation.

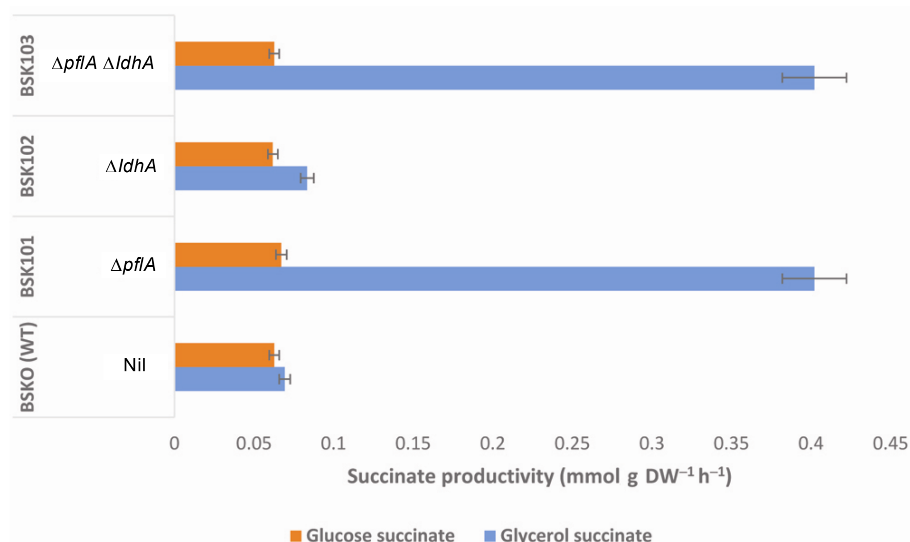

Figure 3. Maximum succinate productivity of *E. coli* BSK (WT) and mutants simulated using glucose and glycerol as substrates. Error bars indicate standard deviation of the replicates.

challenges. Despite these challenges of undergoing several mutations to increase the growth rate of *E. coli* using glycerol substrate, glycerol still represents one of the most important substrates to increase the production of succinate in *E. coli*, and it is now being generated in bulk quantities as a by-product of biofuel industries[27]. Besides its unique features of being abundant and inexpensive, it can also generate more reducing equivalents than its glucose and xylose counterparts[27].

The deletion of *ldhA* on glycerol substrate in mutant strain BSK102 led to succinate production increase of about 120.3% (20% higher) of the wild-type control strain BSKO (Table 2), while maintaining a growth rate that is 69.4% of the wild-type (Figure 3). This significant increase of about 20% by deleting one of the pyruvate dissimilating enzymes called *ldhA* could guide future experimental study to increase succinic acid production by *E. coli* from glycerol substrate (Figure 3). The acetate produced in strain BSK101 (Δ*pflA*) and BSK102 (Δ*ldhA)* is slightly lower than the wild-type control strain BSKO (Table 2). Despite the deletion of pyruvate dissimilating enzymes (*pflA* and *ldhA*), acetate is produced. This indicates that the cell might have used other alternative pathways such as pyruvate dehydrogenase complex (PDH) and/or activated *pflB* to produce acetyl-CoA, acetate and ATP (Figures 4 and 5). On the one hand, deletion of *pflA*

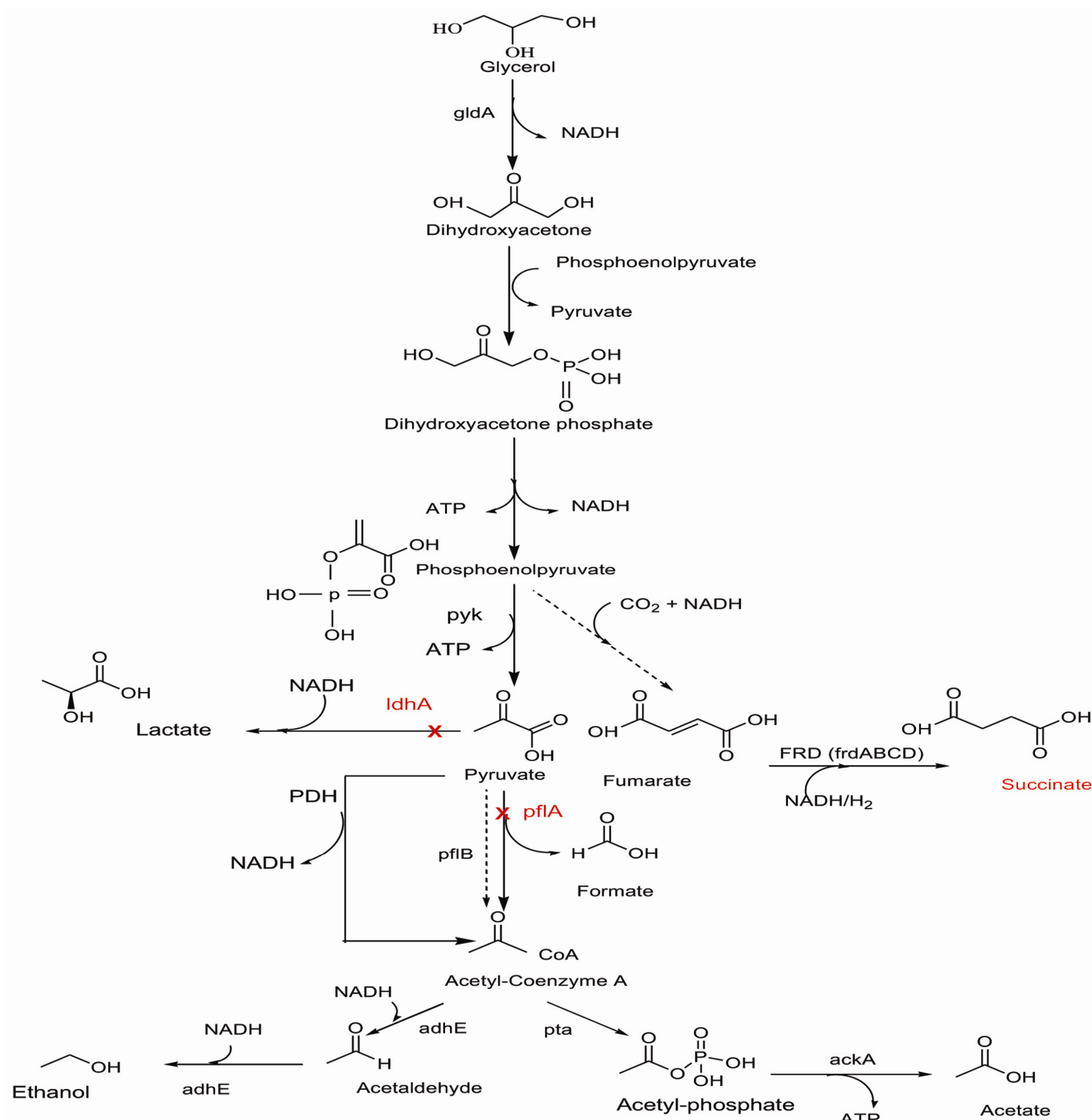

Figure 4. Main fermentative pathways involved in the anaerobic fermentation of glycerol in *E. coli* used for metabolic flux analysis, partially adapted from ref. 37. Relevant genes and corresponding enzymes are included. Ethanol, acetate, formate and succinate are the main products of the fermentative utilization of glycerol[34]. Proposed *in silico* genetic modifications are illustrated by enzymes in red (*pflA* and ldhA). ldhA, Lactate dehydrogenase A; *pflA*, Pyruvate formate lyase A; PDH, Pyruvate dehydrogenase complex; FRD, Fumarate reductase; adhE, Alcohol dehydrogenase; pta, Phosphate-acetyltransferase; pyk, Pyruvate kinase and AckA, Acetate kinase.

in mutant strain BSK101 on glucose led to no acetate production, suggesting that *pflB* is not activated under glucose substrate anaerobic conditions, while on the other, the same strain BSK101 produced acetate on glycerol substrate under anaerobic conditions (Tables 1 and 2). These findings are in agreement with the results reported elsewhere[32]. We can now hypothesize that acetate production under anaerobic condition by *E. coli* is substrate-dependent, and that the use of alternative pathways to produce acetate could be substrate-dependent as well. It can be seen conspicuously that single deletion mutant strain BSK101 (Δ*pflA*) and double deletion mutant strain BSK103 (Δ*pflA* Δ*ldhA*) have the same trend of growth rates and succinate productivities on both glucose and glycerol substrates respectively (Tables 1 and 2 and Figure 3) by deletion of *pflA*. As such, we can now hypothesize that deletion of only *pflA* in *E. coli* can led to no acetate production from glucose, lower acetate production from glycerol and increased succinate productivities from both substrates under anaerobic conditions (Tables 1 and 2 and Figure 3).

The 579% higher succinate production in strain BSK101 (Δ*pflA*) than that of the wild-type control stain BSKO, indicates that succinate production in *E. coli* using glycerol would provide a significant progress toward bio-based industrial chemical production, that is cheaper and hence environment-friendly. It was explicitly reported that crude glycerol generated by biofuel industries can be used as carbon source for microbial fermentations[27]. In addition, given the highly reduced nature of

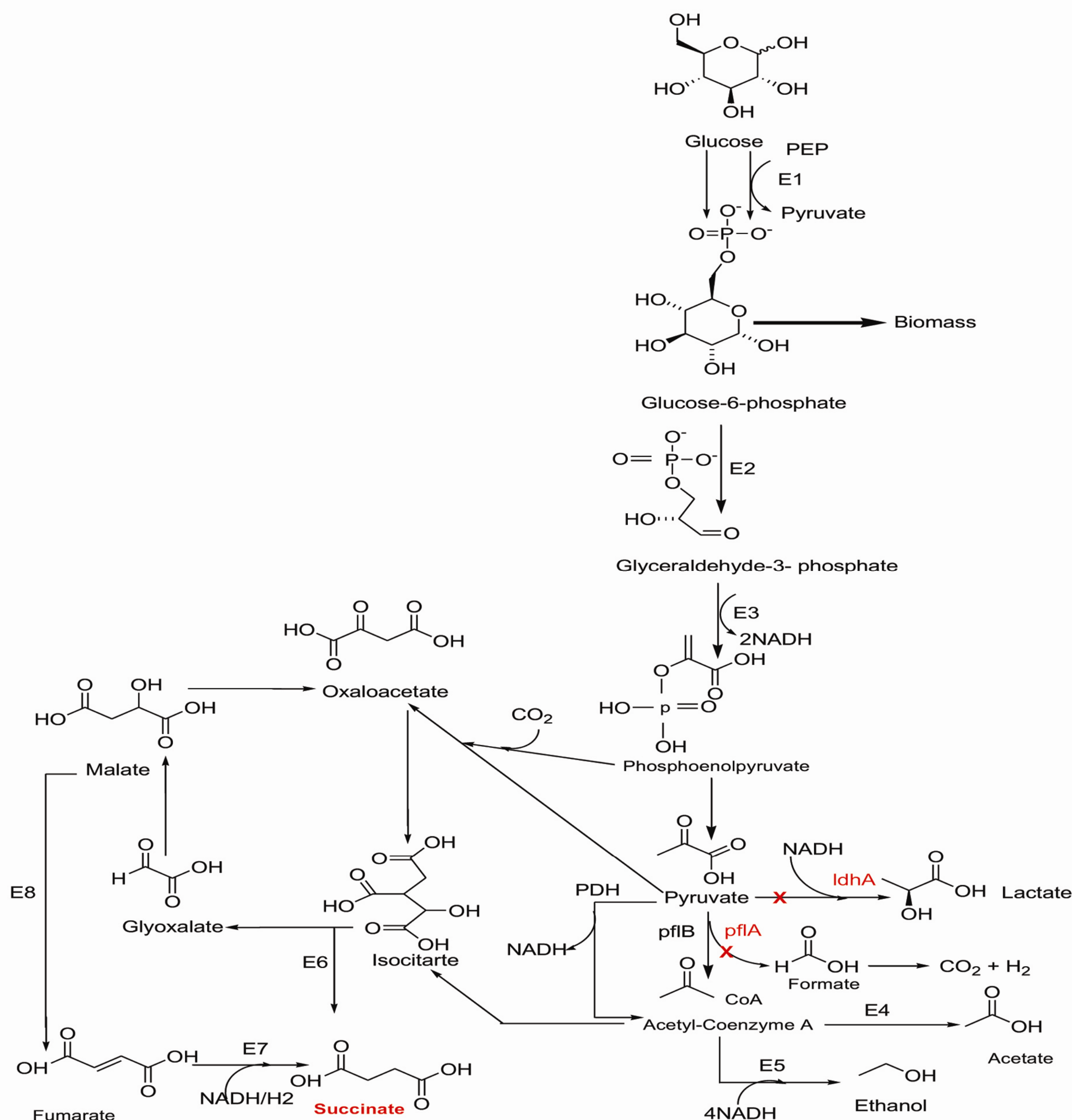

Figure 5. The central anaerobic pathway of *E. coli* used for metabolic flux analysis on glucose substrate, partially adapted from ref. 38. Relevant genes and corresponding enzymes are included. Proposed *in silico* genetic modifications are illustrated by enzymes in red (*pflA* and *ldhA*). E1, Phosphatetransferase system (PTS); E2, Fructose 6-phosphate adolase and glucose-6-phosphate isomerase; E3, Glyceraldehydes phosphodehydrogenase; E4, Phosphoacetyltransferase and acetate kinase; E5, Alcohol dehydrogenase; E6, Isocitrate lyase; E7, Fumarate reductase; E8, Fumarase; PDH, Pyruvate dehydrogenase complex; *ldhA*, Lactate dehydrogenase A; *pflA* and *pflB*, Pyruvate formate lyase A and B.

carbon in glycerol compared to sugar molecules gives it natural advantages when one considers biotransformation as a means for formation of various high-value reduced chemicals[27]. When glycerol is the substrate, twice the amount of reducing equivalent is generated for its conversion to glycolytic intermediates compared to the one generated by the metabolism of glucose or xylose[33], as such potential yields of fuels and other chemicals are higher when synthesized from glycerol as opposed to other monosaccharides[34].

In addition, the *in silico* results reported herein on the use of glycerol for succinate production in *E. coli* mutant strain BSK101 (Δ*pflA*) informed other studies that *E. coli* deficient of *pflA* could increase succinate production dramatically without over-expression of malate enzymes. This result indicates that the choice of substrates and alternative pathways (Figures 4 and 5) to be used in metabolically engineered *E. coli* to increase succinate production is critical. Other workers reported the inactivation of *pfl* and *ldh* coupled with over-expression of

malic enzymes[35,36] to increase succinate production on glucose substrate in *E. coli* mutant strain[35,36].

The two substrates (glucose and glycerol) used in this study have shown a significant predicted succinate production in *E. coli* (Figure 3). Eliminating or inactivating competing pathways of succinate such as *ldh* and *pfl* have been reported previously to increase succinate production on glucose[35,36]. These findings are in agreement with those reported in this *in silico* study. This indicates that the OptFlux software platform using MOMA can effectively predict metabolic engineering strategies based on gene knockout simulation for increased succinic acid production by *E. coli*.

E. coli carries out mixed acid fermentation reactions under anaerobic condition and produces acetate, formate, lactate, ethanol and succinate. When *ldhA* is deleted in *E. coli* under anaerobic condition on glucose and glycerol substrates, there is usually a large accumulation of pyruvate and acetyl-CoA, as previously established[29]. The pyruvate branching partitioning initiates the NADH-consuming reaction for lactate production using *ldhA*, but this reaction was blocked and so no lactate was produced (Figures 4 and 5). On the one hand, the accumulated pyruvate could initiate succinate production using glyoxalate cycle via isocitate, while on the other hand the succinate production could be channelled via phosphoenolpyruvate to oxaloacetate alternative routes (Figure 5).

This work demonstrates the feasibility of using genome-scale *in silico* metabolic model of *E. coli* to simulate gene knockout from glycerol substrates with the OptFlux software platform, which could guide future experimental works for enhanced succinic acid production. The current study further pinpoints conspicuously how significant the substrate selectivity is in relation to increasing succinate production in engineered *E. coli* model strains. Using a rational engineering approach, we report, a predicted mutant model, lacking *pflA*/*b0902* gene, capable of increased succinate production of nearly fivefold (497% higher) from the wild-type control model, when glycerol is the substrate. Its glucose counterpart showed a succinate production increase of only 20% of the wild-type control model. This would pave the way for a comprehensive model-guided experimental inquiry and/or biological discovery on the metabolic processes of the *pflA* in *E. coli* for succinate production when glycerol is the substrate.

1. Lee, S. J. *et al.*, Metabolic engineering of *Escherichia coli* for enhanced production of succinic acid, based on genome comparison and *in silico* gene knockout simulation. *Appl. Environ. Microbiol.*, 2005, **71**, 7880–7887; doi:10.1128/aem.71.12.7880-7887.2005.
2. Zeikus, J. G., Jain, M. K. and Elankovan, P., Biotechnology of succinic acid production and markets for derived industrial products. *Appl. Microbiol. Biotechnol.*, 1999, **51**, 545–552; doi:10.1007/s002530051431.
3. Atsumi, S. *et al.*, Metabolic engineering of *Escherichia coli* for 1-butanol production. *Metab. Eng.*, 2008, **10**, 305–311; doi:http://dx.doi.org/10.1016/j.ymben.2007.08.003.
4. Förster, A. H. and Gescher, J., Metabolic engineering of *Escherichia coli* for production of mixed-acid fermentation end products. *Front. Bioeng. Biotechnol.*, 2014, **2**, 1–12; doi:10.3389/fbioe.2014.00016.
5. Orth, J. D. *et al.*, A comprehensive genome-scale reconstruction of *Escherichia coli* metabolism – 2011. *Mol. Syst. Biol.*, 2011, **7**, 535; doi:10.1038/msb.2011.65.
6. Feist, A. M. *et al.*, A genome-scale metabolic reconstruction for *Escherichia coli* K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. *Mol. Syst. Biol.*, 2007, **3**, 121; doi:10.1038/msb4100155.
7. Reed, J. L., Vo, T. D., Schilling, C. H. and Palsson, B. O., An expanded genome-scale model of *Escherichia coli* K-12 (iJR904 GSM/GPR). *Genome Biol.*, 2003, **4**.
8. Feist, A. M. *et al.*, Model-driven evaluation of the production potential for growth-coupled products of *Escherichia coli*. *Metab. Eng.*, 2010, **12**, 173–186; doi:10.1016/j.ymben.2009.10.003.
9. Edwards, J. S., Ibarra, R. U. and Palsson, B. O., *In silico* predictions of *Escherichia coli* metabolic capabilities are consistent with experimental data. *Nature Biotechnol.*, 2001, **19**, 125–130; doi:10.1038/84379.
10. Monk, J. and Palsson, B. O., Genetics. Predicting microbial growth. *Science*, 2014, **344**, 1448–1449; doi:10.1126/science.1253388.
11. Mienda, B. S. and Shamsir, M. S., Thermotolerant microorganisms in consolidated bioprocessing for ethanol production: a review. *Res. Biotechnol.*, 2013, **4**, 1–6.
12. Rocha, I. *et al.*, OptFlux: an open-source software platform for *in silico* metabolic engineering. *BMC Syst. Biol.*, 2010, **4**, 45; doi:10.1186/1752-0509-4-45.
13. Mienda, B. S., Shamsir, M. S., Shehu, I., Deba, A. A. and Galadima, I. A., *In silico* metabolic engineering interventions of *Escherichia coli* for enhanced ethanol production, based on gene knockout simulation. *IIOABJ*, 2014, **5**, 16–23.
14. Mienda, B. S. and Shamsir, M. S., *In silico* gene knockout metabolic interventions in *Escherichia coli* for enhanced ethanol production on glycerol. *Res. J. Pharm. Biol. Chem. Sci.*, 2014, **5**, 964–974.
15. Mienda, B. S., Shamsir, M. S. and Salleh, F. M., *In silico* metabolic engineering prediction of *Escherichia coli* genome model for production of D-lactic acid from glycerol using the OptFlux software platform. *Int. J. Comput. Bioinform. In Silico Model.*, 2014, **3**, 460–465.
16. Segre, D., Vitkup, D. and Church, G. M., Analysis of optimality in natural and perturbed metabolic networks. *Proc. Natl. Acad. Sci. USA*, 2002, **99**, 15112–15117; doi:10.1073/pnas.232349399.
17. Burgard, A. P., Pharkya, P. and Maranas, C. D., Optknock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. *Biotechnol. Bioeng.*, 2003, **84**, 647–657; doi:10.1002/bit.10803.
18. Yim, H. *et al.*, Metabolic engineering of *Escherichia coli* for direct production of 1,4-butanediol. *Nature Chem. Biol.*, 2011, **7**, 445–452; doi:10.1038/nchembio.580.
19. Cho, A., Yun, H., Park, J. H., Lee, S. Y. and Park, S., Prediction of novel synthetic pathways for the production of desired chemicals. *BMC Syst. Biol.*, 2010, **4**, 35; doi:10.1186/1752-0509-4-35.
20. Lee, D. Y., Yun, H., Park, S. and Lee, S. Y., MetaFluxNet: the management of metabolic reaction information and quantitative metabolic flux analysis. *Bioinformatics*, 2003, **19**, 2144–2146.
21. Zhang, X., Jantama, K., Shanmugam, K. T. and Ingram, L. O., Re-engineering *Escherichia coli* for succinate production in mineral salt medium. *Appl. Environ. Microbiol.*, 2009, **75**, 7807–7813; doi:10.1128/AEM.01758-09.
22. Zhang, X., Shanmugam, K. T. and Ingram, L. O., Fermentation of glycerol to succinate by metabolically engineered strains of *Escherichia coli*. *Appl. Environ. Microbiol.*, 2010, **76**, 2397–2401; doi:10.1128/AEM.02902-09.

23. McCloskey, D., Palsson, B. O. and Feist, A. M., Basic and applied uses of genome-scale metabolic network reconstructions of *Escherichia coli*. *Mol. Syst. Biol*., 2013, **9**, 661; doi:10.1038/msb.2013.18.
24. Varma, A., Boesch, B. W. and Palsson, B. O., Stoichiometric interpretation of *Escherichia coli* glucose catabolism under various oxygenation rates. *Appl. Environ. Microbiol*., 1993, **59**, 2465–2473.
25. Fischer, E., Zamboni, N. and Sauer, U., High-throughput metabolic flux analysis based on gas chromatography–mass spectrometry derived ^{13}C constraints. *Anal. Biochem*., 2004, **325**, 308–316; doi:http://dx.doi.org/10.1016/j.ab.2003.10.036.
26. Le Novere, N. *et al*., BioModels database: a free, centralized database of curated, published, quantitative kinetic models of biochemical and cellular systems. *Nucleic Acids Res*., 2006, **34**, D689–D691; doi:10.1093/nar/gkj092.
27. Mattam, A. J., Clomburg, J. M., Gonzalez, R. and Yazdani, S. S., Fermentation of glycerol and production of valuable chemical and biofuel molecules. *Biotechnol. Lett*., 2013, **35**, 831–842; doi:10.1007/s10529-013-1240-4.
28. Jin, P., Li, S., Lu, S. G., Zhu, J. G. and Huang, H., Improved 1,3-propanediol production with hemicellulosic hydrolysates (corn straw) as cosubstrate: impact of degradation products on *Klebsiella pneumoniae* growth and 1,3-propanediol fermentation. *Bioresour. Technol*., 2011, **102**, 1815–1821; doi:10.1016/j.biortech.2010.09.048.
29. Lin, H., Bennett, G. N. and San, K. Y., Metabolic engineering of aerobic succinate production systems in *Escherichia coli* to improve process productivity and achieve the maximum theoretical succinate yield. *Metab. Eng*., 2005, **7**, 116–127; doi:10.1016/j.ymben.2004.10.003.
30. Vemuri, G. N., Eiteman, M. A. and Altman, E., Effects of growth mode and pyruvate carboxylase on succinic acid production by metabolically engineered strains of *Escherichia coli*. *Appl. Environ. Microbiol*., 2002, **68**, 1715–1727; doi:10.1128/aem.68.4.1715-1727.2002.
31. Cheng, K.-K. *et al*., Global metabolic network reorganization by adaptive mutations allows fast growth of *Escherichia coli* on glycerol. *Nature Commun*., 2014, **5**; doi:10.1038/ncomms4233.
32. Hildebrand, A., Schlacta, T., Warmack, R., Kasuga, T. and Fan, Z., Engineering Escherichia coli for improved ethanol production from gluconate. *J. Biotechnol*., 2013, **168**, 101–106; doi:10.1016/j.jbiotec.2013.07.033.
33. Yazdani, S. S. and Gonzalez, R., Anaerobic fermentation of glycerol: a path to economic viability for the biofuels industry. *Curr. Opin. Biotechnol*., 2007, **18**, 213–219; doi:10.1016/j.copbio.2007.05.002.
34. Dharmadi, Y., Murarka, A. and Gonzalez, R., Anaerobic fermentation of glycerol by Escherichia coli: a new platform for metabolic engineering. *Biotechnol. Bioeng*., 2005, **94**, 821–829; doi:10.1002/bit.21025.
35. Stols, L. and Donnelly, M. I., Production of succinic acid through overexpression of NAD+-dependent malic enzyme in an *Escherichia coli* mutant. *Appl. Environ. Microbiol*., 1997, **63**, 2695–2701.
36. Hong, S. H. and Lee, S. Y., Metabolic flux analysis for succinic acid production by recombinant *Escherichia coli* with amplified malic enzyme activity. *Biotechnol. Bioeng*., 2001, **74**, 89–95.
37. Shams Yazdani, S. and Gonzalez, R., Engineering *Escherichia coli* for the efficient conversion of glycerol to ethanol and co-products. *Metab. Eng*., 2008, **10**, 340–351; doi:10.1016/j.ymben.2008.08.005.
38. Wang, Q., Chen, X., Yang, Y. and Zhao, X., Genome-scale *in silico* aided metabolic analysis and flux comparisons of *Escherichia coli* to improve succinate production. *Appl. Microbiol. Biotechnol*., 2006, **73**, 887–894; doi:10.1007/s00253-006-0535-y.

Received 15 May 2014; revised accepted 10 December 2014

The combination of principal component analysis and geostatistics as a technique in assessment of groundwater hydrochemistry in arid environment

Yousef Nazzal[1,*], Faisal K. Zaidi[2], Izrar Ahmed[3], Habes Ghrefat[2], Muhammad Naeem[2], Nassir S. N. Al-Arifi[2], Saeed A. Al-Shaltoni[2] and Khaled M. Al-Kahtany[2]

[1]Department of Mathematics and Applied Sciences, College of Arts & Sciences, Abu Dhabi University, PO Box 59911, Abu Dhabi, UAE
[2]Department of Geology and Geophysics, King Saud University, PO Box 2455, Riyadh 11451, Saudi Arabia
[3]Colleges of Engineering, King Saud University, Riyadh 11451, Saudi Arabia

Central Saudi Arabia is one of the most arid regions of the world with very little precipitation and extreme climatic conditions. In the absence of available surface water supplies, the non-renewable groundwater resources stored in the Palaeozoic and Mesozoic sedimentary formations form the most important source for irrigation and domestic water requirements. The present study deals with 97 groundwater samples collected from Saq aquifer, which is the major aquifer in the region. The study involves the use of principal component analysis (PCA) and variogram analysis for groundwater quality mapping. PCA helped in establishing a series of factorial variables that summarize all the hydrochemical information. Efforts have been made to identify the spatial development of the principal process acting on groundwater quality by mapping it using factorial variables and ordinary kriging techniques. Two principal components (PCs) were extracted revealing that the chemical characteristics of groundwater in the region were acquired through rock–water interactions and anthropogenic influences. Finally, by applying kriging interpolation technique on the factor distribution values for the two PCs in the area under investigation, the factor distribution maps were prepared. The results concluded that both natural and anthropogenic processes contribute to the groundwater quality, but anthropogenic impacts are more important and may result in further deterioration of groundwater quality if relevant protection methodologies are not adopted.

Keywords: Arid region, geostatistics, groundwater quality, kriging, principal component analysis.

Groundwater resources worldwide are considered as precious sources for meeting the agricultural, domestic and industrial demands. This is especially true for arid

*For correspondence. (e-mail: yousef.nazal@gmail.com)