# Developing systems biology tools to understand nitrogen fixation and biopolymers production by gram-negative bacteria

Diego Tec-Campos[1,2], Cristal Zuñiga[1], Anurag Passi[1], John del Toro[3], JuanTibocha-Bonilla[1], Alejandro Zepeda[2], Michael J. Betenbaugh[3], Karsten Zengler[1,4,5]*

[1] Department of Pediatrics, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0760, USA

[2] Facultad de Ingeniería Química, Universidad Autónoma de Yucatán, Mérida, Yucatán, México

[3]Department of Chemical and Biomolecular Engineering, The Johns Hopkins University, 3400 North Charles Street, Baltimore, MD, 21218, US

[4]Department of Bioengineering, University of California, San Diego, La Jolla CA 92093-0412, USA.

[5]Center for Microbiome Innovation, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0403, USA

*Corresponding author. Email: kzengler@eng.ucsd.edu

## Background

Nitrogen fixation is an important metabolic process carried out by different microorganisms, which converts molecular nitrogen into inorganic nitrogenous compounds such as ammonia ($NH_3$) (Gyurján et al., 1995). These nitrogenous compounds play important roles for biogeochemical cycles on earth and for the synthesis of essential biomolecules at industrial level. *Azotobacter vinelandii* is a well-known gram-negative soil bacterium capable of converting atmospheric nitrogen gas ($N_2$) into soluble ammonia ($NH_3$) as well as into other important nitrogenous soluble compounds. *A. vinelandii* can also produce the biopolymers alginate and polyhydroxybutyrate (PHB) depending on nutritional requirements. Alginate is biodegradable exopolysaccharide (Remminghorst & Rehm, 2006) produced to reduce internal oxygen concentration necessary to fix nitrogen. PHB is synthetized by this microorganism under high carbon/nitrogen ratios as a carbon and energy reserve in the form of cysts (Zúñiga et al., 2011). Despite this metabolic versatility to use different carbon and nitrogen sources, several of the internal metabolic processes regarding carbon and nitrogen partitioning remain unknown. To comprehend the metabolic capabilities of *Azotobacter vinelandii* DJ we used a systems biology approach, which offers tools to predict the organism behavior based on mathematical representations of biological data (Campos et al., 2020). We developed the metabolic model (M-model) of *Azotobacter vinelandii* DJ to contextualize metabolic processes associated with nitrogen fixation, ammonium assimilation, and production of organic nitrogen on genome-scale. Our model was successfully validated using high-throughput phenotypic data and physiological data.

## Methods

The draft model of *A. vinelandii* DJ was generated using The COBRA (Heirendt et al., 2019) and The RAVEN (Agren et al., 2013) Toolboxes. The proteome sequence was obtained from PATRIC database (Genome ID: 322710.5) and was used as input sequence to reconstruct the draft model based on protein homology. We selected five

reference models as templates after alignment of the complete genome sequences of *A. vinelandii* DJ with all bacteria with available models in the BiGG Database (King et al., 2016). Model refinement included two major steps: manual curation/review of the GPR associations and gap-filling by adding new metabolic reactions in the model. In the first step of manual curation, we determined sequence similarity among *A. vinelandii DJ* proteins and the exogenous proteins in the GPRs to identify *A. vinelandii* (AVIN) genes closely related to the exogenous proteins. As second step, Gap-filling analysis was performed to identify the metabolites disconnected and the reactions missing in the model. gap-filling was used to connect pathways through the data retrieved. A second stage of gap-filling was accomplished to connect the metabolites from the medium conditions using literature information (Wong and Maier, 1985) and experimental data generated in the present study. The experimental data were obtained through Biolog plates, a total of 190 carbon sources and 95 nitrogen compounds were used to determine the metabolic capabilities of *A. vinelandii* to growth in different conditions. We performed *in-silico* GPR simulations to verify if the GPR associations are correctly assigned using the COBRA Toolbox algorithms. Next, we performed Mass Balance simulations on the model to check for unbalanced reactions added during the model refinement.

**Results**

*Model properties*

The initial draft model contained 2,432 metabolic reactions and 1,918 metabolites divided into three different compartments. The final *Azotobacter vinelandii* DJ metabolic model (*i*DT1278) consists of 2,003 metabolites, 2,469 reactions and 1,278. Specific metabolic capabilities of *A. vinelandii* DJ such as nitrogen fixation, PHB and alginate production represent around 3% of the metabolic reactions. *i*DT1278 includes a BOF was determined from the first reaction to predict the alginate production since *A. vinelandii* DJ (Noar et al., 2015). *i*DT1278 represents, to our knowledge, the most comprehensive M-model of the diazotroph *A. vinelandii* available to date.

*iDT1278 predicts accurately phenotypic experimental data*

The model was validated under a wide range of different growth conditions (diazotrophic and non-diazotrophic growth), using high-throughput phenotypic data as well as literature information. Initially, *i*DT1278 was tested under six different experimental conditions, specifically, carbohydrates under diazotrophic and non-diazotrophic conditions. The M-model precisely predicted the growth rates for all the carbon sources using ammonium or molecular nitrogen as nitrogen sources. Subsequently, FBA (Orth et al., 2010) was performed for a group of 38 carbon sources in diazotrophic and $H_2$-consuming conditions (Wong & Maier, 1985). Statistical results show for the subset of 38 carbon sources an accuracy of 95%, with 20 true positive predictions (100% positive predicted) and 16 true negative predicted results (89% negative predicted). Additional experimental validation was performed using Biolog plates for a set of carbon (PM1 and PM2) and nitrogen (PM3) sources to determine the

growth rate values of *A. vinelandii* DJ. Out of 190 carbon sources from the Biolog plates, 123 compounds were identified in the model; the simulations were performed under two specific conditions: diazotrophic and non-diazotrophic simulation conditions. The same procedure used in PM1 and PM2 experiments was followed to estimate the growth rates with 75 different nitrogen sources. For this case, simulations were performed using pyruvate as the carbon source. Table. 1 shows the complete analysis of the experimental and predicted data for all carbon and nitrogen sources; statistical parameters (accuracy, sensitivity, specificity, positive predicted, negative predicted and Matthews correlation coefficient) were calculated for non-diazotrophic conditions.

Ultimately, biopolymers production (alginate and PHB) was validated through experimental data retrieved from the literature. We evaluated the model accuracy to growth and alginate production using four carbon sources under diazotrophic and non-diazotrophic conditions. Simulations were confirmed to accurately predict (true positive predictions) alginate production rates with three carbon sources (glucose, mannitol, and sucrose). PHB production was also validated using metabolic modeling. Simulated flux distributions about PHB production were validated using fluxomic data retrieved from Wu et al. (2019). The metabolic fluxes of the reactions involved in the PHB synthesis and related pathways (glycolysis, pentose phosphate pathway, the Entner-Deundoroff pathway, and the TCA cycle) were calculated through FBA for diazotrophic and non-diazotrophic conditions. The simulation results were compared with the experimental measured fluxes (Wu et al., 2019) and the percent error was estimated by reaction. A general agreement in the reaction fluxes was observed under both nitrogen ($N_2$ and $NH_4$) conditions. A total of 16 out of 19 reaction flux estimations presented a global accuracy above 90% for diazotrophic and non-diazotrophic conditions.

Table 1. Statistics of the predictions under carbon and nitrogen sources: true positive (TP), true negative (TN), false positive (FP), false negative, and MCC. Statistical analysis of the estimations for 38 carbon sources under nitrogen fixation and $H_2$ consumption conditions (third column).

| Statistics | Carbon (PM1 and PM2) | Nitrogen (PM3) | Carbon+H2 (Wong and Maier, 1985) |
|---|---|---|---|
| TP | 58 | 6 | 20 |
| TN | 50 | 62 | 16 |
| FP | 3 | 6 | 0 |
| FN | 10 | 1 | 2 |
| Accuracy | 0.89 | 0.91 | 0.95 |
| Sensitivity | 0.85 | 0.86 | 0.91 |
| Specificity | 0.95 | 0.91 | 1 |
| Positive predicted | 0.95 | 0.5 | 1 |
| Negative predicted | 0.84 | 0.98 | 0.89 |
| MCC | 0.28 | 0.62 | 0.67 |

## Conclusion

Here we have created the most comprehensive genome-scale metabolic model for *A. vinelandii* DJ (*i*DT1278) deploying in great detail nitrogen assimilation, nitrogen fixation, as well as on alginate and PHB production. The model consists of 1,278 genes involved in 2,469 reactions. *i*DT1278 predicted accurately the growth ratio and production values of alginate and PHB production under diazotrophic and non-diazotrophic conditions. To our knowledge, this is the first M-model at genome-scale capable to simulate several carbon and nitrogen conditions (close to 250 conditions) with a high precision for growth values and polymer production (PHB and alginate) even when comparing internal metabolic fluxes.

## Funding

This material is based upon work supported by the National Science Foundation under Grant No. 1332344 and CBET-1804187, and the U.S. Department of Energy (DOE), Office of Science, Office of Biological & Environmental Research under Awards DE-SC0012658 and DE-SC0019388. Diego Tec-Campos was in part supported by Mexican National Research Council, CONACYT, fellowship No. 932962.

## References

Agren, R., Liu, L., Shoaie, S., Vongsangnak, W., Nookaew, I., & Nielsen, J. (2013). The RAVEN Toolbox and Its Use for Generating a Genome-scale Metabolic Model for Penicillium chrysogenum. *PLoS Computational Biology*, *9*(3), e1002980. https://doi.org/10.1371/journal.pcbi.1002980

Campos, D. T., Zuñiga, C., Passi, A., Del Toro, J., Tibocha-Bonilla, J. D., Zepeda, A., Betenbaugh, M. J., & Zengler, K. (2020). Modeling of nitrogen fixation and polymer production in the heterotrophic diazotroph Azotobacter vinelandii DJ. *Metabolic Engineering Communications*, *11*, e00132. https://doi.org/10.1016/j.mec.2020.e00132

Gyurján, I., Korányi, P., Preininger, É., Varga, S. S., & Paless, G. (1995). Artificial Plant-Azotobacter Symbiosis for Atmospheric Nitrogen Fixation. In *Azospirillum VI and Related Microorganisms* (pp. 401–413). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-79906-8_46

Heirendt, L., Arreckx, S., Pfau, T., Mendoza, S. N., Richelle, A., Heinken, A., Haraldsdóttir, H. S., Wachowiak, J., Keating, S. M., Vlasov, V., Magnusdóttir, S., Ng, C. Y., Preciat, G., Žagare, A., Chan, S. H. J., Aurich, M. K., Clancy, C. M., Modamio, J., Sauls, J. T., … Fleming, R. M. T. (2019). Creation and analysis of biochemical constraint-based models using the COBRA Toolbox v.3.0. *Nature Protocols*, *14*(3), 639–702. https://doi.org/10.1038/s41596-018-0098-2

King, Z. A., Lu, J., Dräger, A., Miller, P., Federowicz, S., Lerman, J. A., Ebrahim, A., Palsson, B. O., & Lewis, N. E. (2016). BiGG Models: A platform for integrating,

standardizing and sharing genome-scale models. *Nucleic Acids Research*, *44*(D1), D515–D522. https://doi.org/10.1093/nar/gkv1049

Noar, J., Loveless, T., Navarro-Herrero, J. L., Olson, J. W., & Bruno-Bárcena, J. M. (2015). Aerobic Hydrogen Production via Nitrogenase in Azotobacter vinelandii CA6. *Applied and Environmental Microbiology*, *81*(13), 4507–4516. https://doi.org/10.1128/AEM.00679-15

Orth, J. D., Thiele, I., & Palsson, B. Ø. (2010). What is flux balance analysis? *Nature Biotechnology*, *28*(3), 245–248. https://doi.org/10.1038/nbt.1614

Remminghorst, U., & Rehm, B. H. A. (2006). Bacterial alginates: from biosynthesis to applications. *Biotechnology Letters*, *28*(21), 1701–1712. https://doi.org/10.1007/s10529-006-9156-x

Wong, T. Y., & Maier, R. J. (1985). H2-dependent mixotrophic growth of N2-fixing Azotobacter vinelandii. *Journal of Bacteriology*, *163*(2), 528–533.

Wu, C., Herold, R. A., Knoshaug, E. P., Wang, B., Xiong, W., & Laurens, L. M. L. (2019). Fluxomic Analysis Reveals Central Carbon Metabolism Adaptation for Diazotroph Azotobacter vinelandii Ammonium Excretion. *Scientific Reports*, *9*(1), 13209. https://doi.org/10.1038/s41598-019-49717-6

Zúñiga, C., Morales, M., Le Borgne, S., & Revah, S. (2011). Production of poly-β-hydroxybutyrate (PHB) by Methylobacterium organophilum isolated from a methanotrophic consortium in a two-phase partition bioreactor. *Journal of Hazardous Materials*, *190*(1–3), 876–882. https://doi.org/10.1016/j.jhazmat.2011.04.011