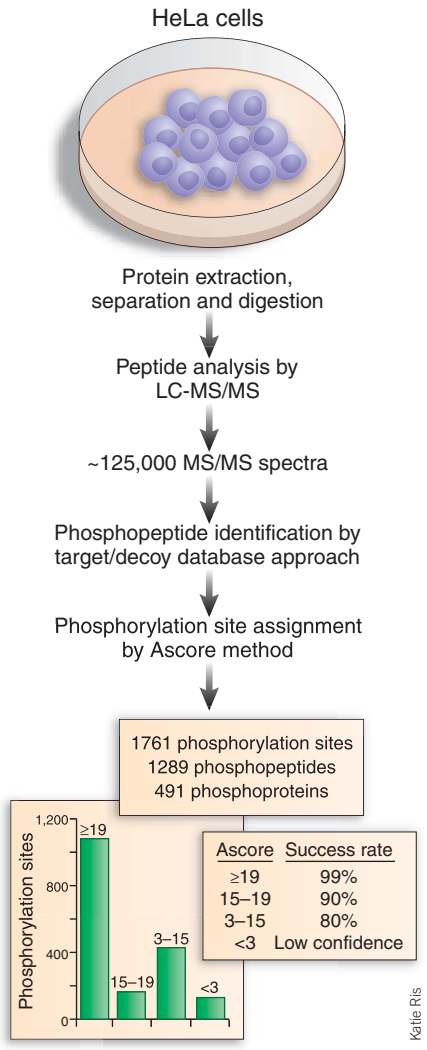

Figure 1 Phosphoproteomics strategy. Automated, large-scale identification of phosphopeptides and localization of phosphorylation sites using MS and computational methods. MS/MS data are used with a target/decoy database search strategy. Phosphorylation sites are automatically assigned by a probabilistic approach (Ascore) that considers the most prominent phosphopeptide sequence-specific fragment ions. The certainty of assignment is readily estimated by the Ascore value.

of post-translational modifications in proteins and peptides by database searching. In fact, using Fourier-transform MS, the authors showed that the majority of intact phosphopeptide masses were determined with a mass deviation of less than 8 p.p.m. This level of peptide mass accuracy is achievable by most modern mass analyzers, including quadrupole time-of-flight (QTOF), Orbitrap and Fourier transform ion-cyclotron resonance (FT-ICR) instruments, and leads to more confident phosphopeptide identification.

The key component of the work by Beausoleil *et al.* is the probabilistic Ascore algorithm. It provides an objective measure of the confidence of phosphorylation site determination. The Ascore algorithm was developed using phosphopeptide MS/MS data obtained with rather low mass accuracy (100–200 p.p.m.), which nevertheless led to high-confidence assignments of approximately 60% of phosphopeptides. The remaining 40% of phosphopeptide assignments are less certain, requiring additional interpretation and validation.

It is likely that a higher fraction of phosphopeptides and phosphorylation sites can be assigned with certainty when the Ascore algorithm is used in combination with high–mass-accuracy MS/MS technology and/or greater phosphopeptide fragmentation efficiency to generate more sequence-specific fragment ions to aid in phosphorylation site assignments. Furthermore, the Ascore algorithm does not consider phosphorylation-specific fragment ions in MS/MS, such as 'neutral loss' of H_3PO_4 and the presence of phosphotyrosine immonium ions, for assignment of phosphorylation sites, which might prove useful in some cases.

So, is the phosphoproteomic pipeline nearly complete? The answer is certainly no. The work of Beausoleil *et al.* and other recent phosphoproteomic studies reveal some of the shortcomings of current analytical technology. The ~2,800 phosphopeptides identified in this study represent less than 10% of the tens of thousands of peptide MS/MS spectra that were acquired in the experiments. Thus, improved sample preparation methods and more efficient MS techniques are required to access deeper layers of the phosphoproteome for identification of phosphoproteins and for sequencing and annotation of additional phosphorylation sites. Furthermore, efficient methods for phosphopeptide quantification are a requirement for functional studies of cellular signaling events. The present contribution by Beausoleil *et al.* catalyzes this research by introducing important computational tools for rapid evaluation of large-scale phosphoproteomic datasets.

1. Beausoleil, S.A., Villén, J., Gerber, S.A., Rush, J. & Gygi, S.P. *Nat. Biotechnol.* **24**, 1285–1292 (2006).
2. Johnson, S.A. & Hunter, T. *Nat. Methods* **2**, 17–25 (2005).
3. Seet, B.T. *et al. Nat. Rev. Mol. Cell Biol.* **7**, 473–483 (2006).
4. Steen, H. & Mann, M. *Nat. Rev. Mol. Cell Biol.* **5**, 699–711 (2004).
5. Jensen, O.N. *Nat. Rev. Mol. Cell Biol.* **7**, 391–403 (2006).
6. Ptacek, J. & Snyder, M. Charging it up: global analysis of protein phosphorylation. *Trends Genet.* published online 12 August 2006.
7. Reinders, J. & Sickmann, A. *Proteomics* **5**, 4052–4061 (2005).
8. Schmelzle, K. & White, F.M. *Curr. Opin. Biotechnol.* **17**, 406–414 (2006).

Deciphering bioplastic production

Sang Yup Lee

The genome sequence of a bioplastic producer opens new avenues for improving polyester biosynthesis.

The shift toward sustainable production systems has focused attention on bioplastics, which can be produced by many microorganisms from renewable raw materials. Polyhydroxyalkanoates (PHAs) are natural polyesters synthesized in the form of discrete intracellular granules in abundances that can reach 90% of dry cell weight[1,2]. In this issue, Pohlmann *et al.*[3] report the long-awaited genome sequence of the most widely used PHA producer, *Ralstonia eutropha* H16. Thus far, commercialization of PHAs has been hampered by their relatively poor material properties and high production costs[2]. The genome sequence provides a foundation for optimizing PHA biosynthesis and for developing metabolic-engineering strategies to generate bioplastics with improved properties.

Currently, several polymers are chemically synthesized at an industrial scale from monomers produced by microbial fermentation, such as lactic acid, 1,3-propanediol and succinic acid. PHAs are unique in that they are synthesized as polymers and are thus

Sang Yup Lee is in the Department of Chemical and Biomolecular Engineering (BK21 program), Department of BioSystems, BioProcess Engineering Research Center, Bioinformatics Research Center, Korea Advanced Institute of Science and Technology (KAIST), 373-1 Guseong-dong, Yuseong-gu, Daejeon 305-701, Republic of Korea.
e-mail: leesy@kaist.ac.kr

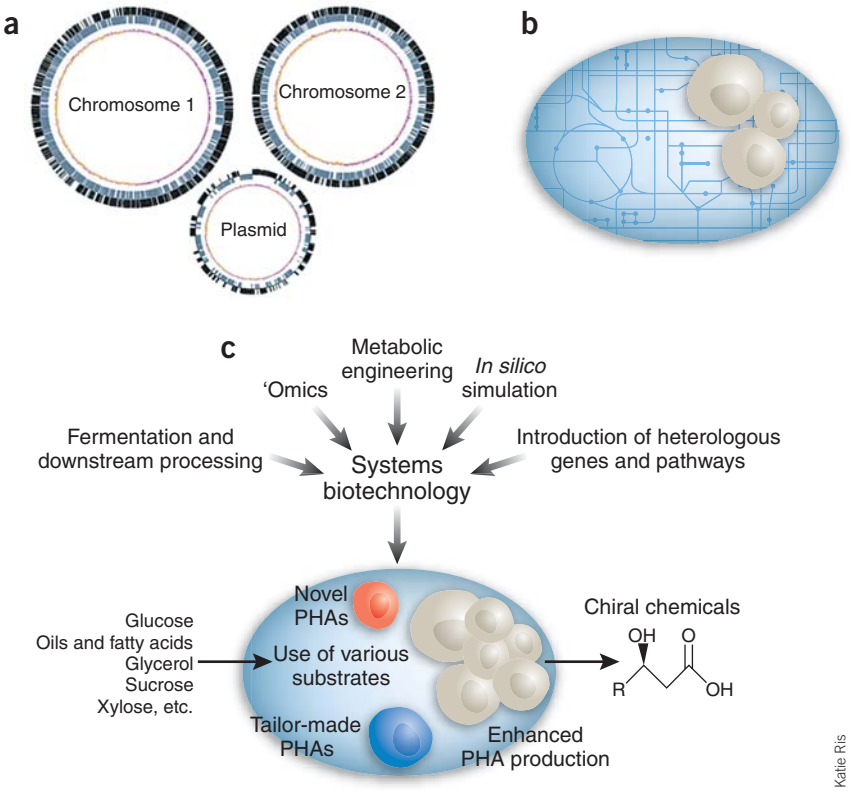

Figure 1 Systems biotechnology of PHA production in *Ralstonia eutropha* H16. (**a**–**c**) Availability of the complete genome sequence (**a**) and a metabolic blueprint (**b**) of this microorganism provides the necessary tools for engineering improved strains (**c**). Improvements include increased production of PHAs, biosynthesis of new PHAs and the production of chiral chemicals for other applications.

amenable to metabolic engineering to generate new polymers. Since the discovery of poly(3-hydroxybutyrate), a polymer of (*R*)-3-hydroxybutyrate, more than 140 monomers have been identified in various PHAs synthesized in microorganisms growing on different carbon substrates[1,2,4]. This wide variety of monomers allows PHAs to have diverse physicochemical properties that make them suitable as thermoplastics (short-chain-length PHAs consisting of 3- to 5-carbon monomers), elastomers (medium-chain-length PHAs consisting of 6- to 16-carbon monomers) and their copolymers showing intermediate properties[4].

In *R. eutropha*, β-ketothiolase (encoded by the gene *phaA*) condenses two acetyl-coenzyme A (acetyl-CoA) molecules to form acetoacetyl-CoA, which is reduced to (–)-3-hydroxybutyryl-CoA by an NADPH-dependent reductase (encoded by *phaB*). The latter is incorporated into the growing chain of poly(3-hydroxybutyrate) by PHA synthase (encoded by *phaC*)[4]. Other types of PHA synthase genes, as well as the genes and metabolic pathways that generate short- and medium-chain-length monomers, have been identified in many other microorganisms[4].

Using these genes, various metabolic engineering and enzyme evolution studies have been carried out to develop improved PHA producers[4,5]. However, most of these studies have been limited to manipulating local pathways directly involved in PHA biosynthesis and precursor supply. The availability of the complete genome sequence of a model PHA producer will now enable system-wide approaches.

R. eutropha H16 possesses a previously sequenced megaplasmid, pHG1, and two chromosomes (4.0 and 2.9 Mbp) encoding more than 6,100 putative genes[3]. Pohlmann *et al.* identified many more apparent PHA biosynthetic genes than they expected. In addition to the PHA synthase gene present in the well-characterized *phaCAB* operon, there are another PHA synthase gene, 37 *phaA* isologs and 15 *phaB* isologs. These genes deserve further study to determine their possible roles in providing precursors for various PHA homo- and copolymers. Furthermore, four genes for phasins, the proteins found on the surface of PHA granules that regulate the number of subunits and size of the granules, and several genes for PHA depolymerase and oligomer hydrolases were identified. These results suggest that PHA metabolism in *R. eutropha* is flexible and more complex than previously appreciated, a finding that may also be true for other PHA producers.

System-level analysis of metabolic, regulatory and signaling networks has revealed new targets and strategies for metabolic engineering[6]. This approach, termed systems biotechnology, considers many components in an integrated manner during the development of strains and processes for optimal production of bioproducts (**Fig. 1**). In the case of PHA, the objective is to design a strain and process that produce PHA of a desired composition and molecular mass at sufficiently high concentration, productivity and yield from the most inexpensive available carbon substrate (which will vary with the production site). The latter is especially important considering that the cost of the carbon substrate in large-scale PHA production processes can be as high as 50% of total operating costs[4].

To achieve these goals, one could carry out metabolic optimization studies using a genome-scale metabolic model based on the *R. eutropha* sequence. Such studies, which include constraints-based flux analysis[7] and comparative 'omics analyses under various genotypic and environmental conditions, would reveal candidate genes and pathways to be engineered[6] that are no longer restricted to those directly involved in PHA biosynthesis.

Improved PHAs with desired monomer compositions should address existing deficiencies in material properties and polymer processing. A good example is the recent development of short- and medium-chain-length PHA copolymers with improved ductility and toughness similar to those of low-density polyethylene[8]. Such copolymers are usually not produced as efficiently as poly(3-hydroxybutyrate), a problem that might be solved through system-level metabolic engineering.

The lessons learned from analysis of *R. eutropha* can most likely be extended to other PHA producers. Thus, the genome sequence is a starting point for developing truly optimized bioplastic production processes through systems biotechnology. The sequence will also facilitate other exciting applications, such as the production of enantiomerically pure hydroxycarboxylic acids[9], biological fuel cells and hydrogen production systems using *R. eutropha*'s robust hydrogenase system[3] and biocatalytic and bioremediation platforms based on its ability to efficiently dissimilate aromatic compounds[10]. Last but not least, new biological information deciphered from the genome will advance our understanding of many aspects of important hydrogen-oxidizing lithoautotrophic bacteria.

1. Steinbuchel, A. & Fuchtenbusch, B. *Trends Biotechnol.* **16**, 419–427 (1998).
2. Lee, S.Y. *Biotechnol. Bioeng.* **49**, 1–14 (1996).
3. Pohlmann, A. *et al. Nat. Biotechnol.* **24**, 1257–1262 (2006).
4. Steinbuchel, A & Doi, Y. (eds.) *Biotechnology of Biopolymers—From Synthesis to Patents* (Wiley-VCH, Weinheim, Germany, 2005).
5. Aldor, I.S. & Keasling, J.D. *Curr. Opin. Biotechnol.* **14**, 475–483 (2003).
6. Lee, S.Y., Lee, D.-Y., & Kim, T.Y. *Trends Biotechnol.* **23**, 349–358 (2005).
7. Price, N.D., Reed, J.L. & Palsson, B.O. *Nat. Rev. Microbiol.* **2**, 886–897 (2004).
8. Noda, I. *et al. Biomacromolecules* **6**, 580–586 (2005).
9. Lee, S.Y., Lee, Y. & Wang, F. *Biotechnol. Bioeng.* **65**, 363–368 (1999).
10. Johnson, B.F. & Stanier, R.Y. *J. Bacteriol.* **107**, 468–475 (1971)

Wastewater genomics

Takashi Mino & Hiroyasu Satoh

A metagenomic sequencing effort sheds light on the biology of wastewater treatment.

Wastewater treatment around the world relies largely on microbial ecosystems, yet little is known about the organisms involved. This 'black box' system mediates aerobic and anaerobic processes that diminish high phosphate loads in wastewaters, a process known as enhanced biological phosphate removal (EBPR)[1,2]. In this issue, Martin *et al.*[3] report the complete genome sequence of "*Candidatus* Accumulibacter phosphatis"[4,5], recently identified as one of the most prevalent organisms in EBPR sludges. The authors use a metagenomic approach to analyze the genome *in situ*—a necessity given that no pure cultures of this organism are available. Knowledge of the genome sequence will improve understanding of how EBPR systems work and provide engineers with the tools needed to optimize EBPR-based wastewater treatment.

For over 30 years, EBPR-based systems have been used in various settings, particularly municipal wastewater treatment plants. Because their design and operation has relied largely on trial-and-error manipulation of poorly characterized sludges, engineers have been unable to understand and correct unpredictable failures or underperformance in phosphate removal.

Microbiologists have also been interested in such systems because of the unique metabolic characteristics and ecological function of the polyphosphate-accumulating organisms (PAOs) responsible for EBPR. Questions in the field have included identifying the key players in phosphate removal, understanding their roles in the process and determining how such systems emerge in the first place. The metabolic and ecological aspects of EBPR have been studied using mixed cultures enriched with PAOs. More recently, a dominant PAO, "*Candidatus* Accumulibacter phosphatis"[4,5], was identified in acetate-fed lab-scale EBPR reactors.

Martin *et al.* used a metagenomic approach to determine the complete genome of *A. phosphatis* and thus obtain a full picture of its metabolic capabilities. Several findings should be of interest to engineers. These include the presence of a high-affinity orthophosphate (P_i) transporter gene ensuring low P_i concentration in treated wastewater, the discovery of a cytochrome supporting an anaerobic functionality of the tricarboxylic acid (TCA) cycle, and the presence of a fumarate reductase that could account for the biosynthesis of the polyhydroxyalkanoate (PHA) polyhydroxyvalerate, a biodegradable plastic substance.

Additional findings relevant to the optimization of EBPR are the operation of the Embden Meyerhof pathway rather than the Entner Doudoroff pathway for glycogen degradation, which affects theoretical predictions of released phosphate and contradicts common assumptions; the formation of extracellular polymeric substances leading to dense flocculation; the surprising lack of a nitrate reductase to catalyze reduction of nitrate to nitrite; the presence of full sets of genes for nitrogen and CO_2 fixation; and strong cobalt dependency.

Among all the findings of Martin *et al.*, however, it is probably the presence of both succinate dehydrogenase and fumarate reductase that best explains one of the key phenomena observed during EBPR, namely the uptake of carbon under anaerobic conditions and its channeling into PHA production. Succinate dehydrogenase is active when the TCA cycle is producing reducing power in the form of NAD(P)H. This provides an alternative pathway to glycolysis that supplies reducing power necessary for the anaerobic

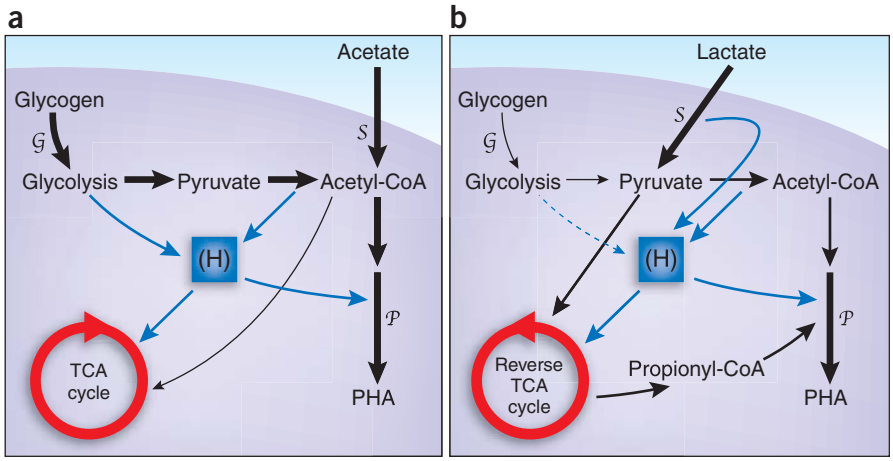

Figure 1 Control of the redox balance by *A. phosphatis* in an EBPR system under anaerobic conditions. (**a**) Reducing power required for the conversion of acetate into 3-hydroxybutyrate for PHA synthesis is provided by glycolysis via the Embden Meyerhof pathway and by acetate oxidation through the TCA cycle. If all reducing power is supplied by glycolysis, then the ratio $\mathcal{S}/\mathcal{G}/\mathcal{P}$ will be 1:(1/6):(2/3). On the other hand, if all reducing power is supplied by the TCA cycle, then $\mathcal{S}/\mathcal{G}/\mathcal{P}$ will be 1:0:(4/9). (**b**) When the TCA cycle operates in reverse in the presence of fumarate reductase, the cell can take up lactate from the environment. When half of the lactate is converted to PHA via acetyl-CoA and the rest via propionyl-CoA, the redox balance is maintained, and $\mathcal{S} = 2\mathcal{P}$. The reverse operation of the TCA cycle is needed to consume excess reducing power produced during the conversion of lactate into acetyl-CoA. (H), reducing power or hydrogen in such forms as NAD(P)H and FADH2; $\mathcal{S}$, molar amount of acetate or lactate taken up; $\mathcal{G}$, molar amount of glucose unit in glycogen consumed; $\mathcal{P}$, molar amount of monomeric units of PHA produced.

Takashi Mino and Hiroyasu Satoh are in the Department of Socio-Cultural Environmental Studies, Graduate School of Frontier Sciences, The University of Tokyo, 5-1-5, Kashiwanoha, Kashiwa City 277-8563, Japan.
e-mail: mino@mw.k.u-tokyo.ac.jp