Anais da Academia Brasileira de Ciências (2019) 91(Suppl. 3): e20190252
(Annals of the Brazilian Academy of Sciences)
Printed version ISSN 0001-3765 / Online version ISSN 1678-2690
http://dx.doi.org/ 10.1590/0001-3765201920190252
www.scielo.br/aabc | www.fb.com/aabcjournal

Exploring biochemical diversity in bacteria

JEAN WEISSENBACH

Génomique Métabolique, Genoscope, Institut François Jacob, CEA, CNRS, Univ Evry,
Université Paris-Saclay, 2 rue Gaston Crémieux, 91057 Evry, France

Manuscript received on February 28, 2019; accepted for publication on April 18, 2019

How to cite: WEISSENBACH J. 2019. Exploring biochemical diversity in bacteria. An Acad Bras Cienc 91: e20190252. DOI 10.1590/0001-3765201920190252.

Abstract: The various descriptors of biochemical diversity and an evaluation of its status of knowledge are briefly outlined. Using a few examples from in house research projects, I illustrate strategies used to increase this knowledge. Because bacteria represent an extremely diverse domain of life and carry out the widest known range of biochemical transformations, this mini-review focusses on bacteria.

Key words: bacterial metabolism, biochemical function, enzyme discovery, genome annotation, bioremediation.

WHAT IS BIOCHEMICAL DIVERSITY?

Life is based on chemical transformations of the environment. To survive and to multiply, biological systems transform chemical compounds found in the environment. This enables organisms to gain energy and to obtain the elements that are part of their constituents. Chemical transformations are also used to neutralize or eliminate harmful compounds. A fraction of the chemical transformations is common to all organisms, the central metabolism in particular. But many reactions occur only in a subset of biological systems and this forms the basis of biochemical diversity (or chemical biodiversity). Environments display extremely diverse physical and chemical conditions, and only the living systems that have the appropriate characteristics will survive in a given niche. Because of biochemical diversity some specialized organisms will be able to colonize environments which can be hostile to many others, unable to perform the transformations necessary for survival. In most cases systems specialise, but in some instances they tend to keep or to acquire the ability to thrive in a broader set of conditions (e.g. aerotolerance versus strict anaerobiosis). In addition to the ability to live in particular niches, plants and microorganisms, have evolved a very wide range of secondary metabolites which are also part of chemical biodiversity.

Biochemical diversity can be seen from two points of view:

a) The ability for a given organism to thrive in a given environment. In evolutionary terms, chemical biodiversity has progressively been acquired and enabled life to colonize more and more diverse niches and environments. This notion stems back to Pasteur and the early days of microbiology, but new basic discoveries in the capacity of microorganisms to diversify sources of energy and nutrients are still being made.

E-mail: jsbach@genoscope.cns.fr
ORCid: https://orcid.org/0000-0001-6564-0840

b) In distinct organisms the pathway producing a given metabolite (or degrading a compound) may be identical but can be partly or totally different (unrelated). For instance CO2 autotrophy has been invented at least 6 times in bacteria (Fuchs 2011).

We will briefly try to outline the various descriptors of biochemical diversity, to evaluate its status of knowledge and, using a few examples from our own research projects, illustrate strategies used to increase this knowledge. Because bacteria represent an extremely diverse domain of life and carry out the widest known range of biochemical transformations, we will focus on bacteria below.

DESCRIBING THE CHEMICAL BIODIVERSITY LANDSCAPE

The extent of the repertoire of chemical transformations carried out by living systems is reflected by some indicators featured in public databases. Since many reactions occurring in cells are catalysed by enzymes, these latter represent a central category in the biochemical diversity landscape. An official catalogue of all known enzymatic activities is maintained and updated by the nomenclature committee of the International Union of Biochemistry and Molecular Biology (IUBMB) (http://www.sbcs.qmul.ac.uk/iubmb/enzyme/) which proposes recommendations regarding nomenclature. The IUBMB nomenclature committee assigns a four digit identifier, the well-known EC number, that classifies enzymes according to the reaction they catalyse. A number of other databases (https://www.ebi.ac.uk/intenz/, http://www.brenda-enzymes.info/ etc.) feature various additional information on the actual enzymes found in defined organisms catalysing the reactions, physicochemical parameters etc. With the advent of complete genome sequences, pathways from the sequenced organisms could be reconstructed from the gene composition. Several data bases such as KEGG (https://www.kegg.jp/kegg/) or MetaCyc (https://metacyc.org/) display these pathways. Last but not least inventories of metabolites which document another facet of the chemical biodiversity are also available through databases (e.g. http://www.ebi.ac.uk/chebi/). These inventories also frequently feature other compounds of non-biological origin, but exerting a biological effect. In addition to reinvention of entire pathways as mentioned above, a single reaction can be catalysed by independently evolved but isofunctional activities. A survey of isofunctional enzymes has been reported (Omelchenko et al. 2010). It is estimated that the non-homologous isofunctional activities amount between 100 and 200 entities (alternative solutions).

Altogether these resources provide very useful tools and give us a first estimate of the known chemical biodiversity. It is however also possible to enlarge the landscape using additional information. We have already seen the importance of genome sequences which establish a link with species and inform us about the coverage of the tree of life. However, the present inventories show that there are very important biases in the distribution of the species at the origin of genome sequences available (https://gold.jgi.doe.gov/). Many phylogenetic groups still remain absent in sequence databases whereas others, notably of medical interest are heavily overrepresented, thus distorting our representation of the landscape. Metagenome sequencing projects can partly remediate to this situation especially when they target specific and diverse environments. Sequence analyses from both genomes and metagenomes contribute substantially to the ever increasing gene census which accumulates more and more "function unknown" proteins, ¾ as of June 2018 (http://www.uniprot.org/). When microbial, a majority of these genes encode enzyme functions. Saturation curves from metagenomic sequence analyses indicate that the gene content of some

environments is approaching a plateau (Sunagawa et al. 2015, Li et al. 2014).

Despite such encouraging signs, this quick overview shows that the 2018 state of knowledge of metabolism remains very unsatisfactory. The ongoing sequence data deluge informs us about the tree of life coverage and the depth of our ignorance in terms of gene sequence interpretation. Several indications argue for a massive effort for improvement : (1) paradoxically there are still about 20% of experimentally validated enzymatic activities, the sequence of which remains unknown; (2) even in *Escherichia coli* K-12 new pathways are still being discovered (Denger et al. 2014); (3) similarly, the power and sensitivity of analytical methods (MS, NMR) reveals an important number of unknown metabolites as well as known orphan compounds not yet integrated in pathways; (4) substrate promiscuity (Copley 2015) of enzymes is systematically overlooked and its effect on various aspects of metabolism (e.g. flux analysis) ignored; (5) annotations are currently marred by overinterpretations and subsequent propagation of errors.

Because of insufficient efforts devoted to experimental approaches, the present situation will be long lasting. The lack of general approches to the discovery of new enzymatic activities hampers improvements. Many activities require a specific experimental design, in terms of reaction conditions, substrates etc. Therefore, once sequenced and annotated, prokaryote genomes remain systematically with a fraction of putative genes with unknown function that encompasses a third to a half of the total. The importance of this fraction reflects the type of ecological niche of the species. Despite marginal improvements resulting from inclusion of knowledge acquired from other organisms, the fraction of putative genes with no known function is essentially stable over time. There are however important incentives calling for an increased endeavour (Anton et al. 2014). The very common dream to predict a phenotype on paper, just on the basis of an extensive and accurate description of the genotype is still a naïve illusion. Similarly the construction of predictive global metabolic models rests on an exhaustive knowledge of the metabolites and the metabolic connections that is still far from completion. This modelling may also have practical consequences for biotechnological applications as do the metabolite and enzyme inventories. These latter provide a substantial repertoire for biocatalytic steps in partial or full biosynthetic pathways to industrially produce chemicals. Increasing this repertoire will add to the resources useful for bio-economy.

IMPROVING THE PRESENT KNOWLEDGE

Existing resources have not yet been used at their full power. Some resources are biological such as collections of mutants or of expressed proteins (ORFeomes). Others can be obtained from databases (sequences, 3D structures, metabolome analysis etc.). A number of bioinformatic tools can serve various purposes such as displaying genomic sequences, sequence alignments, 3D structures, phylogenetic trees etc.

A first series of actions can bear on annotations of existing genes and genome sequences. Several anciently studied pathways contain sometimes steps with enzymes with no known amino-acid sequence. As an example, the lysine fermentation pathway was established in 1972, but 3 of the enzymes of the last steps were never associated with a protein sequence. Combining gene context, similarity of enzymatic mechanisms, and molecular weight comparisons with known proteins we selected candidate genes for these three sequence-orphan proteins. The expressed recombinant protein showed the expected activities. (Kreimeyer et al. 2007).

Some systematic efforts to tackle the issue of unknown functions in genomes have been

undertaken for model organisms. These efforts consist in establishing collections of mutants by various ways and phenotyping these mutants in media containing defined C, N, S sources, and other factors susceptible to complement the mutated gene. At Genoscope we have focused on *Acinetobacter baylyi* ADP1 (de Berardinis et al. 2009). Using our genome wide mutant collection, we revisited the first step of the methionine biosynthesis pathway that consists in an acylation of L-homoserine (succinylation or acetylation) and for which we noticed a few experimental inconsistencies. According to the current opinion gene MetX encodes L-homoserine acetyl transferase (HAT) and the unrelated gene MetA L-homoserine succinyl transferase (HST). It could be experimentally shown that introducing a mutation of a single amino acid in the catalytic pocket of HST or HAT inverted both activities into HAT or HST. Both enzymes (HAT and HST) thus represent two cases of non-homologous isofunctional enzymes (NISE) (Omelchenko et al. 2010) that can be encoded by unrelated genes (MetX and MetA). This finding hence necessitated the reassignment of thousands of erroneously annotated genes encoding either HAT or HST (Bastard et al. 2017). It necessitated the development of a sequence analysis procedure of the amino acid residues of active site. The analysis was based on a method constructing hierarchical trees of the enzyme active site called "active site modelling and clustering" (ASMC) (de Melo-Minardi et al. 2010).

ASMC can also be used as a classification procedure to distinguish on a rational basis between various groups of a large protein family. The procedure was actually developed to explore a protein family corresponding to a domain of unknown function (Pfam DUF849). We had previously shown that a few members of the DUF849 family, the BKACE group, were catalysing cleavage of 3-keto-5-amino-hexanoate and transfer of the resulting amino-keto moiety onto acetyl-CoA in the lysine fermentation pathway (see above) (Kreimeyer et al. 2007). Construction of a phylogenetic tree of DUF849 also showed that the Pfam DUF849 proteins of lysine fermenters all clustered in a single branch of the tree. Moreover, most members of the Pfam DUF849 family are present in bacteria which do not ferment lysine. We thus hypothesized a generic cleavage reaction conserved within the family and applied on representatives, structural and modeling investigations (based on ASMC), analysis of genomic and metabolic context and ended with high-throughput enzymatic screening. This approach unearthed 14 potential new enzymatic activities, leading to the designation of these proteins as beta-keto acid cleavage enzymes. We propose an *in vivo* role for four enzymatic activities and suggest key residues for guiding further functional annotation. Our results illustrate also how the functional diversity within a family may be largely underestimated (Bastard et al. 2014).

Metabolites derived from particular compounds, such as xenobiotics, can be identified and reveal new transformation activities. For the last seven years or so we have tried to observe biodegradation of a polychlorinated pesticide, chlordecone or kepone. This pesticide acts against the banana black weevil. Its bishomocubane structure makes it particularly recalcitrant to biodegradation and resulted in inclusion on the list of Persistant Organic Pollutant (POP) of the Stockholm Convention. Although banned for its toxicity in the USA since 1976, it was intensely used in the French West Indies until 1993. The pollutant seems to accumulate in soil and contaminates rivers, sea, mangroves. It also accumulates in the food chain that ends with fish and seafood. The prostate cancer risk in Martinique and Guadeloupe is twice higher than the world mean value. Chlordecone also affects neurodevelopment in young children and acts as an endocrine disruptor. Until recently there was no evidence

for biodegradation. We have recently isolated a bacterial consortium and a few isolated species therefrom (notably *Citrobacter sp*) that in anaerobic conditions are able to degrade chlordecone into a number of metabolites (>30) belonging to three chemical families. The chemical structure of more than 20 components has been determined. These compounds are all partially dechlorinated, but the mechanism(s) of dechlorination remain unknown. Genome sequencing of the isolated species failed to identify known enzymes that catalyse any type of dechlorination, especially reductive (Chaussonnerie et al. 2016).

These few examples illustrate some of the long and arduous ways that are still necessary to open to increase our knowledge of the highly sophisticated chemistry performed by living systems. But without the improvements in all aspects of analytical chemistry and molecular genetics the present acquisitions would have been unthinkable a few decades ago.

REFERENCES

ANTON BP, KASIF S, ROBERTS RJ AND STEFFEN M. 2014. Objective: biochemical function. Front Genet 5: 210.

BASTARD K ET AL. 2017. Parallel evolution of non-homologous isofunctional enzymes in methionine biosynthesis. Nat Chem Biol 13: 858-866.

BASTARD K ET AL. 2014. Revealing the hidden functional diversity of an enzyme family. Nat Chem Biol 10: 42-U77.

CHAUSSONNERIE S ET AL. 2016. Microbial Degradation of a Recalcitrant Pesticide: Chlordecone. Front Microbiol 7: 2025.

COPLEY SD. 2015. An evolutionary biochemist's perspective on promiscuity. Trends Biochem Sci 40: 72-78.

DE BERARDINIS V, DUROT M, WEISSENBACH J AND SALANOUBAT M. 2009. Acinetobacter baylyi ADP1 as a model for metabolic system biology. Curr Opin Microbiol 12: 568-576.

DE MELO-MINARDI RC, BASTARD K AND ARTIGUENAVE F. 2010. Identification of subfamily-specific sites based on active sites modeling and clustering. Bioinformatics 26: 3075-3082.

DENGER K, WEISS M, FELUX AK, SCHNEIDER A, MAYER C, SPITELLER D, HUHN T, COOK AM AND SCHLEHECK D. 2014. Sulphoglycolysis in Escherichia coli K-12 closes a gap in the biogeochemical sulphur cycle. Nature 507: 114-117.

FUCHS G. 2011. Alternative pathways of carbon dioxide fixation: insights into the early evolution of life? Annu Rev Microbiol 65: 631-658.

KREIMEYER A, PERRET A, LECHAPLAIS C, VALLENET D, MEDIGUE C, SALANOUBAT M AND WEISSENBACH J. 2007. Identification of the last unknown genes in the fermentation pathway of lysine. J Biol Chem 282: 7191-7197.

LI J ET AL. 2014. An integrated catalog of reference genes in the human gut microbiome. Nat Biotechnol 32: 834-841.

OMELCHENKO MV, GALPERIN MY, WOLF YI AND KOONIN EV. 2010. Non-homologous isofunctional enzymes: a systematic analysis of alternative solutions in enzyme evolution. Biol Direct 5: 31.

SUNAGAWA S ET AL. 2015. Ocean plankton. Structure and function of the global ocean microbiome. Science 348: 1261359.