Metabolomics: Open Access

Editorial Open Access

Behind Every Good Metabolite there is a Great Enzyme (and perhaps a structure)

Garry W. Buchko[1]*, Isabelle Phan[2], Lisabeth Cron[2], Robin Stacy[2], Lance J. Stewart[3], Bart L. Staker[3], Thomas E. Edwards[3], Gabriele Varani[4], Wesley C. Van Voorhis[5] and Peter J. Myler[2,6]

[1]Biological Sciences Division, Pacific Northwest National Laboratory, Richland, WA, USA
[2]Seattle BioMed, Seattle, WA, USA
[3]Emerald BioStructures, Bainbridge Island, WA, USA
[4]Departments of Chemistry and Biochemistry, University of Washington, Seattle, WA, USA
[5]Department of Medicine, University of Washington, Seattle, WA, USA
[6]Department of Biomedical Informatics & Molecular Education and Department of Global Health, University of Washington, Seattle, WA, USA

Editorial

In the early days of modern molecular biology the focus of scientific research was to simplify problems as much as possible and study one gene/protein at a time [1]. Eventually, in part due to technology advances and in other part due to necessity, groups of genes and proteins were studied and biochemical pathways were identified. Today, due to yet greater number of technological advancements including the ability to store and analyze large amounts of data, it is possible to study everything at the same time. These "totality" studies gave birth to the fields of genomics, transcriptomics, proteomics, and metabolomics. In turn, the combined study of all these global analyses gave birth to the field of systems biology [2].

Another "totality" field brought to life with new emerging technologies is structural genomics, an effort to determine the three-dimensional structure of every protein encoded in a genome. The Seattle Structural Genomics Center for Infectious Disease (SSGCID) is a specialized structural genomics effort composed of academic (University of Washington), government (Pacific Northwest National Laboratory), not-for-profit (Seattle BioMed), and commercial (Emerald BioStructures) institutions that is funded by the National Seattle Structural Genomics Center for Infectious Diseases (Federal Contract: HHSN272200700057C and HHSN27220120025C) to apply genome-scale approaches in solving protein structures from bio-defense organisms, as well as those causing emerging and re-emerging disease [3,4]. The SSGCID's target selection strategy focuses on drug targets, essential enzymes, virulence factors, and vaccine candidates from a number of bacterial (*Bartonella, Brucella, Ehrlichia, Anaplasma, Rickettsia, Burkholderia, Borrelia* and *Mycobacterium*) and eukaryotic (*Babesia, Cryptosporidium, Toxoplasma, Giardia, Entamoeba, Coccidioides* and *Encephalitozoon*) pathogens, as well as ssDNA and negative-strand ssRNA viruses. Community input (in the form of target requests for entry into the structure determination pipeline) is actively solicited at http://www.ssgcid.org/home/Community.asp. Target genes are PCR amplified, cloned, and screened for expression in *Escherichia coli*. Soluble proteins are purified in milligram quantity, screened for crystallization, and crystals analyzed by X-ray diffraction using in-house sources or off-site synchrotron beam-lines. Small proteins (<25 kDa) that fail to crystallize are queued for structure determination by NMR-based methods. All expression clones, purified proteins, and protein structures produced by SSGCID are freely available to the scientific community and lay the groundwork for research in more than 40 collaborative projects. In five years over 60 manuscripts describing SSGCID structures and methods have been published, including an entire edition of *Acta Crystallographica F* (Sept 2011) devoted to SSGCID [5]. Over 540 structures have been deposited into the Protein Data Bank (PDB) by SSGCID. About one third of all SSGCID structures contain bound ligands, many of which are metabolites or metabolite analogues present in the cell. These proteins structures are the blueprints for the structure-based design of the next generation of drugs against bacterial pathogens and other infectious diseases [6].

Many of the selected SSGCID targets are annotated enzymes from known metabolomic pathways essential to cellular vitality, suggesting selectively "knocking-out" one of the enzymes in an important pathway with a drug may be fatal to the organism. Because the active site can differ between close homologues [7], SSGCID selects identically annotated enzymes from several genera and species in order to increase the structural coverage of metabolomic pathways. One reason why metabolomic pathways are important is because of the small molecules, or metabolites, produced at various steps in these pathways and identified by metabolomic studies. Unlike genomics, transcriptomics, and proteomics that may be influenced by epigenetic, post-transcriptional, and post-translational modifications, respectively, the metabolites present in the cell at any one time represent downstream biochemical end products, and therefore, metabolite profiles may be most closely associated with a phenotype [8] and provide valuable information for infectious disease research. Metabolomic data would be even more useful if it could be linked to the vast amount of structural genomics data. Towards this goal SSGCID has created an automated website tool (http://apps.sbri.org/SSGCIDTargetStatus/Pathway) that assigns selected SSGCID target proteins to MetaCyc pathways (http://metacyc.org/).

To date, SSGCID has selected at least one target from 936 of the 1790+ pathways in the Metacyc database. To illustrate some of the information present on the SSGCID-Pathway website, the tabulated information for PWY-4981 (Proline Biosynthesis II (from Arginine) as of 10/21/12 is shown in Figure 1.

This pathway contains five enzymes and SSGCID has targeted at least one enzyme in this pathway from five different genera. For each genus the number of enzymes selected, enzymes in PDB, enzymes

*Corresponding author: Garry W. Buchko, Biological Sciences Division, Pacific Northwest National Laboratory, Richland, WA 99352, USA, Tel: 509-371-6543; E-mail: garry.buchko@pnnl.gov

solved by SSGCID in PBD, SSGCID clones available, and SSGCID proteins available are listed. Information on SSGCID's genus-specific coverage may be pulled-down by selecting a genus as shown for *Burkholderia* in Figure 2.

Here, the current status (updated weekly) of each target in the SSGCID structure determination pipeline is provided. Yet more specific information for each enzyme may be pulled-down by selecting the EC number to the left of the enzyme's name as shown in Figure 3 for ornithine carbamoyltransferase (2.1.3.3).

Here, every clone generated by SSGCID is identified with an SSGCID ID number and the status of each clone provided in more detail. For this enzyme, orthologues of the gene from seven different *Burkholderia* species were selected to increase the odds of success (*B. pseudomallei, B. ambifaria, B. cenocepacia, B. multivorans, B. phymatum, B. thailandensis, B. vietnamiensis, B. xenovorans*). Additional information may be pulled-down by selecting the target SSGCID ID number. For example, if ButhA.00088.a is selected then the amino acid sequence, cloning information, and in this case, the PDB ID number (4F2G) is obtained. Note that the available clones may be ordered free of charge from the BEIR repository (http://www.beiresources.org) and available protein (left over primarily following crystal screening) are available free of charge from SSGCID (www.SSGCIDproteins.org). To guide the elicitation of new targets currently not in the SSGCID pipeline for structure determination, there is a user-friendly "box" for the scientific community to request a protein's entry into the SSGCID pipeline as illustrated in Figure 4.
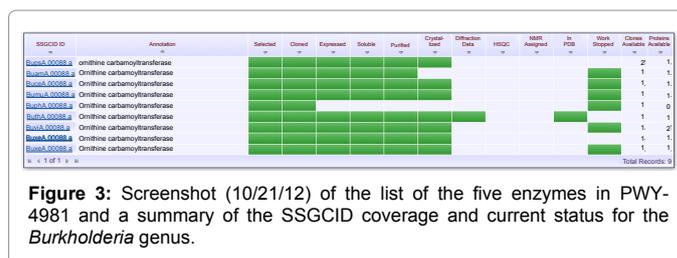

Figure 1: Screenshot (10/21/12) of the list of genera for PWY-4981 (Proline Biosynthesis II (from Arginine) on the SSGCID-Pathway website (http://apps.sbri.org/SSGCIDTargetStatus/Pathway/PWY-4981). The metabolomic pathway contains five enzymes and SSGCID has targeted at least one of these enzymes for structure determination from this genus list of infectious disease organisms.

Figure 2: Screenshot (10/21/12) of the list of the five enzymes in PWY-4981 and a summary of the SSGCID coverage and current status for the *Burkholderia* genus.
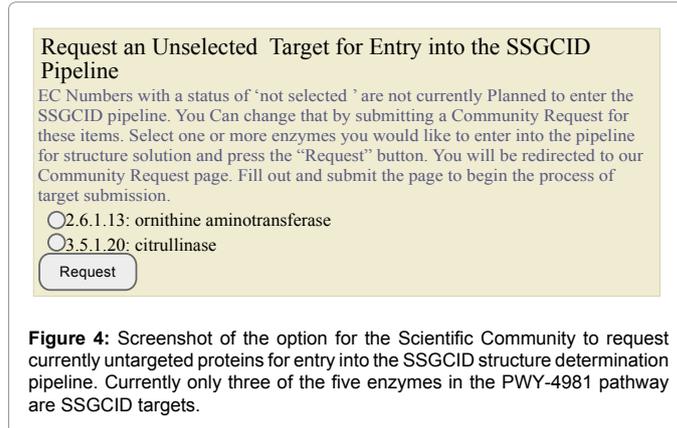

Figure 3: Screenshot (10/21/12) of the list of the five enzymes in PWY-4981 and a summary of the SSGCID coverage and current status for the *Burkholderia* genus.

Figure 4: Screenshot of the option for the Scientific Community to request currently untargeted proteins for entry into the SSGCID structure determination pipeline. Currently only three of the five enzymes in the PWY-4981 pathway are SSGCID targets.
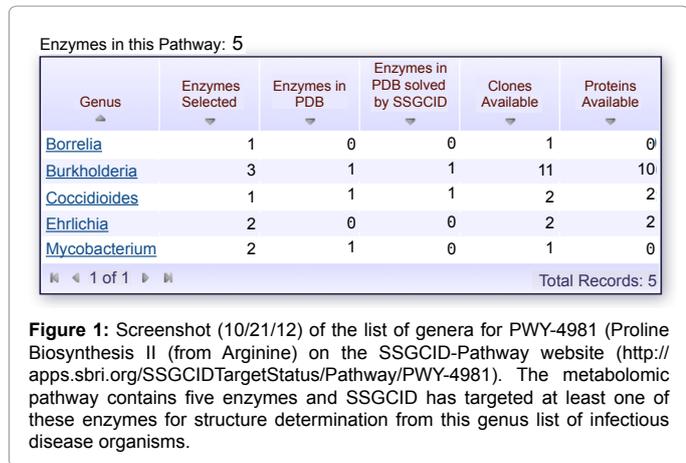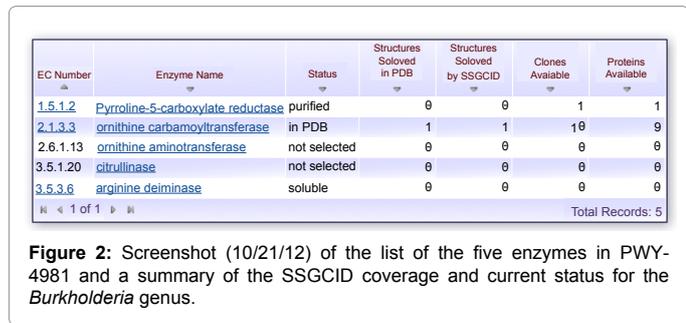
In summary, the rapidly progressing field of metabolomics is poised to make significant strides in identifying important metabolites generated by infectious disease organisms. The SSGCID-Pathway website tool represents a first big step towards linking metabolites and metabolic pathways to structural genomic data with the goal of accelerating the discovery of new agents to battle infectious diseases.

Acknowledgments

References

1. Beadle GW, Tatum EL (1941) Genetic control of biochemical reactions in neurospora. Proc Natl Acad Sci USA 27: 499-506.

2. Kirschner MW (2005) The meaning of systems biology. Cell 121: 503-504.

3. Myler PJ, Stacy R, Stewart L, Staker BL, Van Voorhis WC, et al. (2009) The Seattle Structural Genomics Center for Infectious Disease (SSGCID). Infect Disord Drug Targets 9: 493-506.

4. Stacy R, Begley DW, Phan I, Staker BL, Van Voorhis WC, et al. (2011) Structural genomics of infectious disease drug targets: the SSGCID. Acta Crystallogr Sect F Struct Biol Cryst Commun 67: 979-984.

5. Buchko GW (2011) Structural genomics - A goldmine of blueprints for structure-based drug design. Metabolomics 1:104e.

6. Van Voorhis WC, Hol WG, Myler PJ, Stewart LJ (2009) The role of medical structural genomics in discovering new drugs for infectious diseases. Plos Comput Biol 5: e1000530.

7. Erlandsen H, Abola EE, Stevens RC (2000) Combining structural genomics and enzymology: completing the picture in metabolic pathways and enzyme active sites. Curr Opin Struct Biol 10: 719-730.

8. Patti GJ, Yanes O, Siuzdak G (2012) Innovation: Metabolomics: the apogee of the omics trilogy. Nat Rev Mol Cell Bio 13: 263-269.