Beilstein-Institut

Bozen 2006, May 15th – 19th, 2006, Bozen, Italy

Applying the Principle of Complementarity: Modelling Chemical and Biological Space

Stephen J. Haggarty

Broad Institute of Harvard and MIT,

7 Cambridge Center, Cambridge, MA, 02142, U.S.A.

E-Mail: haggarty@broad.harvard.edu

Received: 14th August 2006 / Published: 5th November 2007

Abstract

Chemical genomics aims to systematically explore the interactions between small molecules and biological systems at different levels of organization ranging from individual macromolecules to whole organisms. By analogy to the progression of creating genetic maps over the past century, which now provide nucleotide-level resolution of entire genomes, chemical genomics allows the annotation of 'chemomes', the full set of biologically relevant chemicals capable of interaction with a particular biological system. This article aims to discuss recent progress made toward the goal of mapping multidimensional chemical and biological descriptor spaces. The focus is on the complementary nature of these efforts and the importance of recognizing the distinction between computed versus observed descriptors. Recent examples of identifying small molecules the molecular interactions of which give rise to novel phenotypes relevant to human disease and our understanding of complex biological pathways will be described. To further advance the field, information being derived from computational studies of molecular structure and observational studies of molecular function must be integrated into global models of biological activity that are both explanatory and predictive. Unlike the complementarity principle in physics, which describes the impossibility of simultaneously observing both the wave and particle nature of light and electrons, it is possible to simultaneously observe and model

chemical and biological space. In doing so a fuller description of the interaction of small molecules with biological systems arises than if either of the two spaces is considered separately.

Introduction

Chemical genetics is modelled after classical genetics, especially with respect to the use of phenotype-based screening (the word phenotype is derived from Greek *phaino-*, from *phainein*, meaning to show or be observable) (Fig. 1) [1 – 7]. However, chemical genetics differs from classical genetics in the use of small molecules, rather than mutations, to perturb the function(s) of gene products. Thus, chemical genetics applies the principles and logic of genetics, but the analyses focus on proteins rather than genes. By extension chemical genomics differs from genomics in the focus on the systematic understanding of the interactions of small molecules at different levels of organization ranging from individual macromolecules to whole organisms. Given the temporal control offered by small molecules, and the ability to use combinations of small-molecule modulators, chemical genetics promises to complement the use of pure genetic analysis to study a wide range of biological systems and mechanisms.

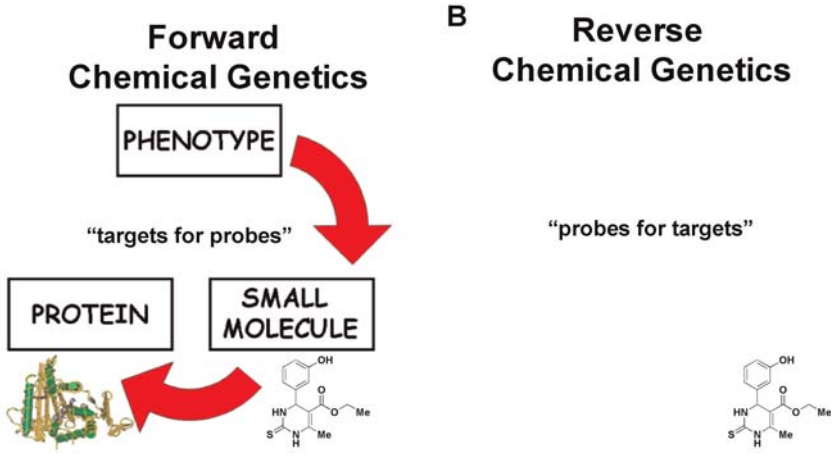


Figure 1. Chemical genetics aims to target gene products using small molecules rather than to target the genes themselves by mutating an organism's genetic material. Forward versus reverse chemical genetics. Whereas forward chemical genetics relies on a phenotype of interest to guide the selection of biologically active small molecules, reverse chemical genetics use a protein of interest to identify small molecules that can be used to probe the function of the selected protein. Both approaches require the use of small molecules and phenotypic assays but differ in the starting points of discovery.

Besides the development of high-throughput phenotypic assays for screening large collections of small molecules, chemical genetics has evolved to emulate classical genetics in a number of ways: (1) the development of high-throughput phenotypic assays compatible

with performing screens of large collections of chemicals; (2) the use of chemical-genetic modifier (suppressor and enhancer) screens to reveal connections between pathways and networks as well as epistatic relationships between gene products; (3) the use of synthetic-lethal (and synthetic-viable) screening to reveal redundant elements of pathways and networks; (4) the creation of "chemical-genetic maps" that position chemicals in a multi-dimensional space formed from phenotypic or computed descriptors (chemical space).

Probes for Chemical Genomics

In addition to examining known bioactive molecules, and therapeutically useful drugs, chemical genomics requires the efficient synthesis and screening of novel collections of small molecules having rich skeletal and stereochemical diversity in order to discover new probes of biological and disease mechanisms. Chemical Biologists are becoming increasingly adept at making small molecules that are suitable for use in forward and reverse chemical genetic studies. These methods include the use of DNA template-mediated, and target- and diversity-oriented organic synthesis, peptide and carbohydrate synthesis, and enzyme-mediated synthesis, the latter of which enables *in vitro* evolution, protein engineering, and even non-natural amino acids to be incorporated into polypeptides [8 – 13].

An important development in chemical library synthesis has been the recognition of the importance of not only creating diversity, so as to increase the likelihood of finding an active small molecule, but also to retain the potential to site- and stereoselectively attach appendages to the small molecule during a post-screening optimization stage [8 – 10]. Such chemical handles not only facilitate the addition of functionalities that increase the potency or selectivity of the small molecule, but also, equally as important, they can be used to facilitate the identification of interacting target proteins and pathways.

With access to such optimal collections of small molecules, the challenge for the field of chemical biology includes: (1) determining which of these molecules have *specific* effects upon biological systems (at various levels of resolution from proteins to whole organisms); (2) determining the structural and physiochemical properties of molecules that specify associated biological activities; and ultimately (3) directing future synthetic efforts along particular pathways in the synthetic network to produce small molecules efficiently that modulate a biological systems in any desired manner.

Molecular and Phenotypic Descriptors of Small Molecules

With growing interest in the use of chemicals as probes in basic and clinical research, the field of chemical biology in general, and chemical genomics in particular, is facing the challenge of transitioning from the *ad hoc* discovery of biologically, and therapeutically, useful small molecules to the systematic discovery and elucidation of molecular targets. The systematic discovery of biologically active small molecules and novel molecular targets, requires the development of a computational framework for analysing large, high-dimensional datasets. The recent creation of publicly accessible data repositories

and analysis environments, such as ChemBank [14], the Biomolecular Interaction Network Database (BIND) [15], and PubChem [16], now allow this work to be feasible in the academic community.

When considering chemical space there are two fundamentally different classes of descriptors that are used: computed and measured (Fig. 2). These classes differ insofar as the former are generally calculated using a computer and various algorithms designed to determine the value of a specified mathematical function, whereas the latter involve the observation of the effect of a chemical on, for example, the function of a gene product (nucleic acids, proteins) or metabolite (carbohydrate, lipid, other organic molecules). Recognizing the distinction between chemical spaces derived from computed descriptors as compared to measured descriptors is of fundamental importance. Whereas the former is definable and unambiguously definable, the latter involves the process of observation, and as such involves noise inherent to the process of measurement. Measured phenotypic descriptors are also subject to the influence of a variety of other variables, including the dose of the chemical, length of treatment, and the genotype of the biological system.

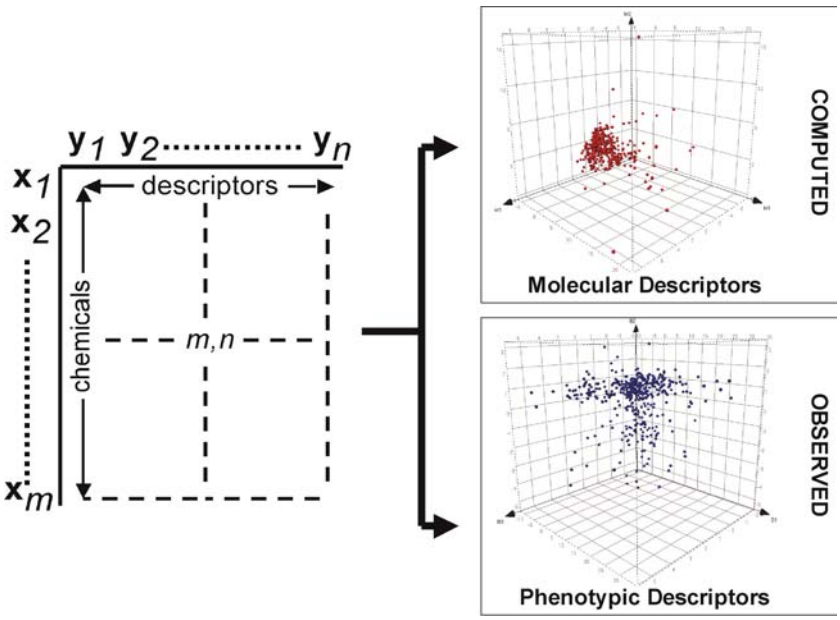


Figure 2. Computed versus observed descriptors used to create maps of chemical space. Principle component models of chemical space for 480 small molecules analysed using 24 computed molecular descriptors and 60 measured phenotypic descriptors derived from a cell-based assay of cell proliferation [data from 17].

Many excellent reviews exist on the use of computed molecular descriptors to navigate and model chemical space [18 – 22]. One challenge in the use of molecular descriptors to create maps of chemical space that can predict biological activity is that a given chemical can exist as a variety of structures corresponding to various protonation, tautomeric, and

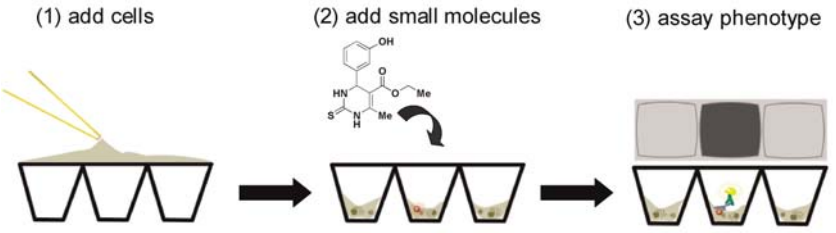
stereochemical states depending on the molecule's environment. Another major challenge is the ability of enzymes to metabolize small molecules into what might be either an active or inactive component. Together, these, and other, factors contribute to the difficulty of predicting the function of a small molecule, particularly in the context of an intact living system as complex as the human body.

In contrast to computed molecular descriptors phenotypic descriptors involve the measurement of the effects of a small molecule on a biological system. Although the term phenotype is mostly widely used in genetics to refer to any part of the observable structure, function or behaviour of a living organism, it also includes the observable physical parts: atoms, molecules, macromolecules, cells, tissues, and organs. Accordingly, phenotypic descriptors provide the opportunity to classify chemical structures by creating maps of chemical space according to biologically- or disease-based descriptors. Given the wide range of observable properties of biological systems, the challenge for mapping chemical space in this manner is to determine what the most relevant phenotypic descriptors to measure are, which in turn may depend on the biological process being studied. These descriptors may also vary if one is ultimately developing drugs rather than probes for the dissection of biological systems.

Development of Novel Phenotypic Descriptors

As discussed by Fishman and Porter [23], developing a new "grammar" for pharmaceutical discovery based upon an understanding of cellular pathways and signalling networks involved in modulating human disease is required. One component of this effort is the comprehensive mapping of chemical space through the integration of disparate medicinal chemistry structure–activity relationships and target data on drugs that have been discovered to date [24]. Another component involves assessing the properties of newly synthesized or discovered small molecules in the context of high-throughput binding assays or intact biological systems. Using the latter approach, a growing number of cell- [25 – 37] and organism-based [reviewed in 38] assays that have been developed recently and applied to screening small molecules (Fig. 3). While a number of these assays used homogeneous, plate reader-based readouts as descriptors, high-content, image-based screens obtained using automated microscopy are beginning to provide a rich source of phenotypic descriptors for classifying small molecules [35 – 37].

A



B

Assay type	Assay format	Output	Measurement class	Nature of measurement	Throughput
visual	cultured cells; whole organism	optical	optical	discrete	low-medium
enzyme-linked immunosorbent assay	cultured cells; protein extract	optical	absorbance; fluorescence; luminescence	continuous valued	medium
reporter gene assay	cultured cells; whole organism	optical	fluorescence	continuous valued	high
cytoblot	cultured cells	optical	absorbance; fluorescence; luminescence	continuous valued	medium
fluorescence imaging	cultured cells; whole organism	optical	fluorescence	continuous valued	low-medium
ion imaging	cultured cells	optical	fluorescence	continuous valued	high
viability	cultured cells; whole organism	optical	absorbance; fluorescence; luminescence	continuous valued	high
electrical recording	cultured cells	electrical	electrical	continuous valued	low-medium

Figure 3. Phenotypic assays for chemical genetics. (**A**) Types of assays that have been used for chemical genetic screening. (**B**) Example of a cell-based assay involving phospho-specific antibody-based determination of a cell-state [described in more detail in 25]. A cytoblot involves growing cells on the bottom of a well, fixing the cells and probing the cells for the presence of a particular antigen using a specific primary antibody in solution. A secondary antibody covalently linked to horseradish peroxidase is added and the presence of the entire complex is detected through the chemiluminescent reaction caused by addition of luminal and hydrogen peroxide.

Biological Space as Complementary to Chemical Space

Given the above description of chemical space, it follows that the term 'biological space' is most appropriately used to refer to a space complementary to that of chemical space (Fig. 4). This space is complementary insofar as it uses the same information encoded in the pattern of interactions of small molecules with biological systems to classify biological systems instead of chemicals. By transposing the data matrix and considering the small molecules as descriptors for the phenotypic assays, the resulting data creates an information-rich signature of the biological system being probed, which in turn can be used for pattern recognition, classification and assessing the diversity of the biological processes or systems being assayed.

As for when modelling chemical space, there is the same need to make the distinction between computed and observed descriptors when analysing biological space. When viewed in this framework, chemical and biological space as a whole are composed of a total of four quadrants (I–IV). In quadrant I, chemicals are positioned in space using computed molecular descriptors. In quadrant II, chemicals are positioned in space using measured phenotypic descriptors of biological activity. In quadrant III, biological systems

are positioned in space using measured descriptors of chemical activity. In quadrant IV, biological systems are positioned in space using computed descriptors based upon a computable model of the system.

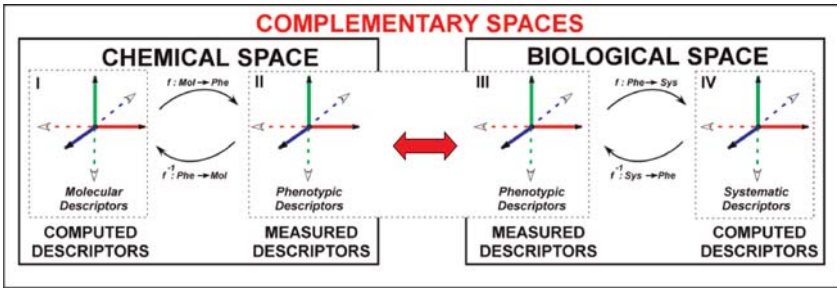


Figure 4. Chemical and biological space as complementary aspects of chemical genomics. Chemical and biological space are represented as complementary spaces composed of a total of four quadrants (I–IV) in which chemicals and biological systems, respectively, can be described quantitatively using multiple computed or measured descriptors.

To date besides the widespread computational analysis of DNA sequence, there have been only a few studies that have purposefully used the computation of biological descriptors to classify biological systems. An example, reminiscent of the use of molecular descriptors to classify chemical structures, has been described [39]. Similar perturbation-based profiling experiments may prove useful for classifying complex disease states and revealing aspects of network biology.

Computational Models of Chemical and Biological Space

The key organizing factor for analysing both chemical and biological space is derived from the multidimensional data structure generated when multiple descriptors are used to annotate a small molecule or biological system, respectively. This data structure is most often that of a two dimensional array, or matrix, denoted by **S**, consisting of an ordered array of n columns and m rows (Fig. 2). Each column ($\mathbf{y}_j$) in **S**, corresponds to a descriptor, and is denoted by a bold-face, lower case letter subscripted j (where $j = 1$ to n). Each row ($\mathbf{x}_i$) in **S** corresponds to a chemical (or biological system), and is denoted by a bold-face, lower case letter subscripted i (where $i = 1$ to m). Accordingly, an element (e_n) of **S** encodes information (m, n) about chemical (or biological system) m for descriptor n. This allows the elements of **S** to be considered as coordinates in a multidimensional space spanned by the descriptor axes, which, in turn, allows each chemical to be represented as a vector, the magnitude and direction of which is given by the corresponding values in **S**, $\mathbf{x}_i = [e_1, e_2,...,e_n]$. In this matrix-based representation of chemical space, the relative distance between chemicals (or biological systems) $\mathbf{x}_i$ becomes a measure of their similarity with respect to the particular descriptors considered.

Visualizing Interactions in Chemical and Biological Space

In order to create meaningful visual representations of high-dimensional data to allow for data exploration and to facilitate subsequent modelling efforts, a variety of dimensionality reduction and pattern finding algorithms have been developed. These algorithms allow for the creation of higher-level representations of the information inherent in the lower-level relational data inherent within matrices of data. Besides clustering, which has been widely used to group small molecules into various structural and activity classes, one method of dimensionality reduction that has been found to be particularly useful for analysing multi-dimensional chemical genetic data sets is that of principal component analysis (PCA).

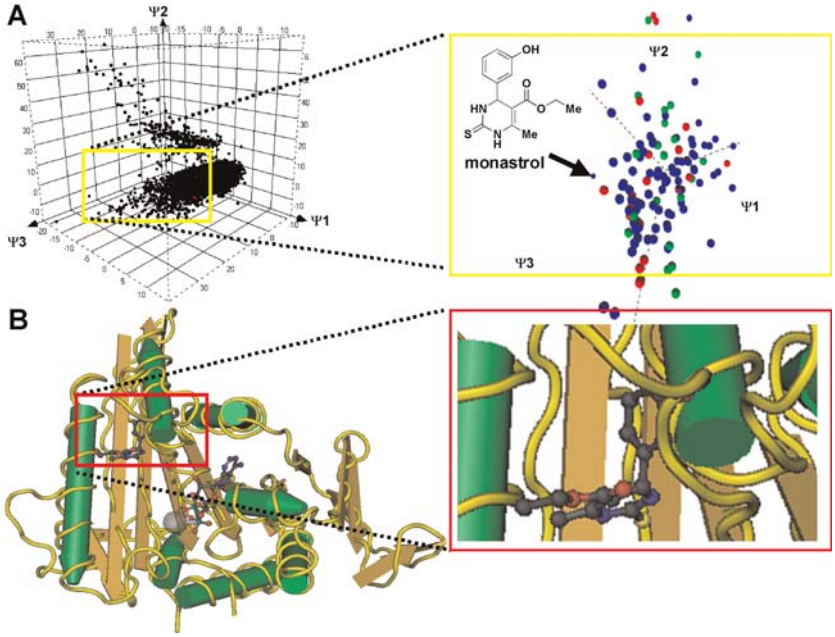


Figure 5. New activities in chemical space and the target of monastrol. (**A**) 3-dimensional representation of chemical space showing the position of 15,120 small molecules (coloured balls) in a molecular descriptor space derived from the first three principal components axes (Ψ1-Ψ3) obtained from the analysis of the corresponding structural and physiochemical descriptors (data from [27, 30, 32]). Inset shows 132 biologically active small molecules coloured based upon phenotypic data from cell-based assays for suppressors of the topoisomerase inhibitor ICRF-193 (red), suppressors of the histone deacetylase inhibitor trichostatin A (green), and anti-mitotics (blue). In total, there were 20 suppressors of ICRF-193, 21 suppressors of ITSA, 89 anti-mitotics, and 2 small molecules that scored in both the anti-mitotic and trichostatin A suppressor screen. Monastrol's location is as shown. (**B**) Co-crystal structure of monastrol with the motor domain of human KSP (Eg5) showing that monastrol confers inhibition by creating an 'induced-fit' to a pocket away from the adenosine triphosphate and magnesium binding site within the catalytic center (data from [41]).

While PCA provides a readily computable, linear dimensionality reduction, a number of algorithms with improved behaviour have been described [18, 40]. Nonetheless, PCA provides an illustrative example of the use of computational methods. At its core, PCA consists of a linear transformation of the original system of axes formed by the *n*-dimensions of the data matrix. This transformation is in the form of a Euclidean distance-preserving rotation, the directions of which are determined by computing a set of Eigenvectors and corresponding eigenvalues of a diversity matrix created by computing a standardized covariance matrix. The resulting Eigenvectors provide a new set of linearly independent, orthogonal axes, called principal components, each of which accounts for successive directions in the *n*-dimensional ellipsoid spanning the multivariate distribution of the original data. The corresponding eigenvalues account for progressively smaller fractions of the total variance in the original data. Accordingly, PCA creates a global model of chemical and biological space that minimizes the information lost upon projection into a space of reduced dimensionality, and is thus well suited for exploring complex activity patterns (Fig. 5A).

These global models of chemical space in which small-molecules, can be located are in contrast to models that depict the local interactions of a small molecule with a macromolecular target (Fig. 5B). In the case of the small-molecule monastrol, which targets a mitosis-specific, kinesin-related motor protein [26], consideration of these spaces reveal, again, complementary aspects on the interaction of small molecules with biological systems. Whereas what the crystal structure reveals suggests specific non-covalent bonding interactions and allosteric changes in the protein's structure that are involved in altering its enzymatic function [41], the PCA model reveals aspects of the specificity of monastrol and creates a landmark in the local region of the chemical space in which small molecules with similar targets can be found.

Network-Based Models of Small Molecule Interactions

As an alternative to PCA, given a multidimensional matrix of data derived from chemical genetic screens, it is also possible to use computational tools derived from the field of discrete mathematics and principles borrowed from graph theory [39, 40, 42]. For example (Fig. 6), through multiple screens, biologically active small molecules can be linked together into a network of chemical genetic interactions, which can be represented by the graph $\boldsymbol{G}=(\boldsymbol{V}, \boldsymbol{E})$, where $\boldsymbol{V}$ represents either small molecules or assays and $\boldsymbol{E}$ represents edges indicating the activity of a small molecule in a given assay. To determine that a small molecule is active, a threshold or a statistical measure based upon a control distribution of inactive or control compounds can be used. Ultimately, the topology of the chemical genetic network for a particular biological system will be determined by the selectivity of the small molecules and constrained by the properties of the underlying biological networks being studied.

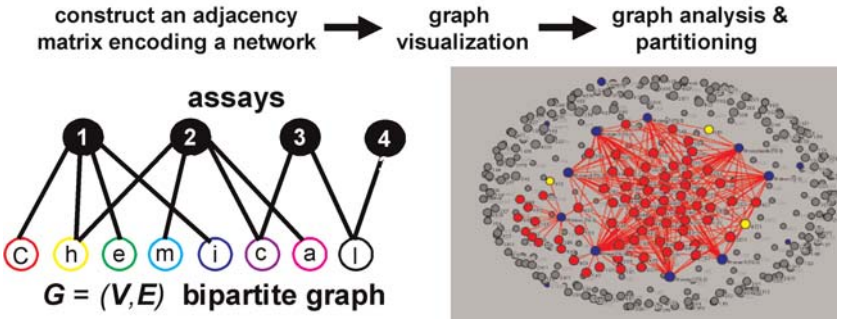


Figure 6. A chemical-genetic network representing a graph ***G***=(***V***, ***E***) (data from [39]). Each node (***V***; circles) represents a biologically active small molecule or a phenotypic assay and each edge (***E***; line) represents an observed biological activity. Shown here is an undirected, unweighted, bi-partite graph with a total of 426 nodes (***V***) and 1107 edges (***E***) between small molecule nodes (coloured red or yellow for active; grey for inactive; total of 352) and an assay node (coloured blue; total of 74 in 7 organisms). This 'energy-minimized' representation was computed using Pajek v0.72 (see http://vlado.fmf.uni-lj.si/pub/networks/pajek/).

Application of Chemical Genomics to Protein Acetylation

One of the most useful applications of chemical genetics is to reveal the gene products that function in pathways or processes in an unbiased manner. To expand further the molecular toolbox available for studying intracellular protein acetylation [reviewed in 43, 44], a number of chemical genetic screens have been performed. For example, using a panel of cell-based assays measuring the acetylation state of specific lysine residues on histones and α-tubulin using antibodies and a library of over 7,200 small molecules derived from a diversity-oriented synthesis that included 'biasing' elements to target the compounds toward the family of HDACs [45], over 600 small-molecule inhibitors of protein deacetylation were identified (Fig. 7) [31, 47].

Following the decoding of chemical tags and resynthesis, the selectivity of one inhibitory molecule (tubacin) was shown toward α-tubulin deacetylation and another (histacin) toward histone deacetylation (Fig. 8) [46]. Tubacin was found not to affect the level of histone acetylation, gene-expression patterns, or cell-cycle progression. Using immunopreciptated, recombinant enzyme, and it as determined that the class II histone deacetylase 6 (HDAC6) is the intracellular target of tubacin [46]. Through a combination of the use of catalytically inactive point mutations in each of the two catalytic domains of HDAC6 and tubacin, it was shown that only one of the two catalytic domains of HDAC6 possesses tubulin deacetylase activity, and that only that domain's deacetylase activity could be inhibited by tubacin. Collectively, the small molecules identified as suppressors of trichostatin A (ITSAs) and the selective inhibitors of protein deacetylation should facilitate dissecting the role of acetylation in a variety of cell-biological processes (Fig. 9) [30, 46, 47].

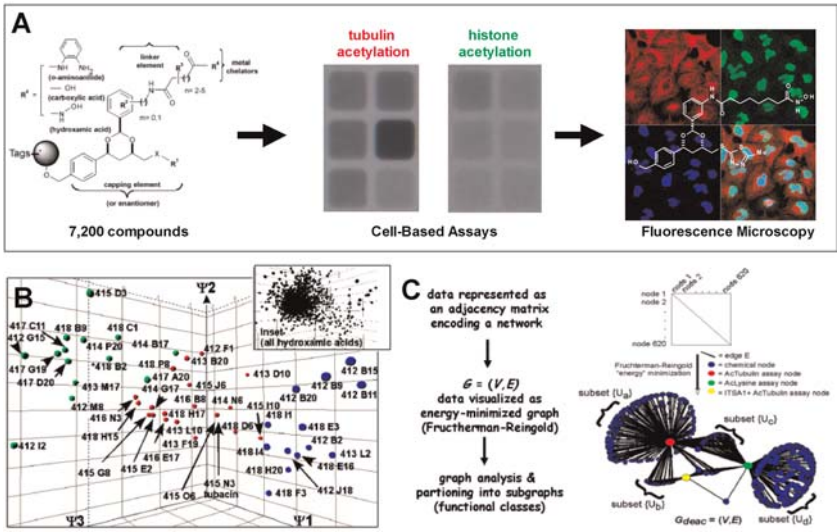


Figure 7. Forward chemical genetic screen for inhibitors of protein deacetylation (data from [31, 36]). (**A**) Overview of cell-based screens of the 1,3-dioxane-based, diversity-oriented synthesis-derived library using antibodies to measure tubulin and histone acetylation. (**B**) Relative position of selected active compounds in a 3-dimensional principal component model computed from five cell-based assay descriptors. AcTubulin-selective (red), AcLysine-selective (green), and most potent (blue). **(C)** Chemical genetic network from screening data after applying the Fruchterman-Reingold 'energy' minimization algorithm. Nodes represent either assays or small molecules according to the indicated colours. Edges (black lines) connect bioactive small molecules to the corresponding assay.

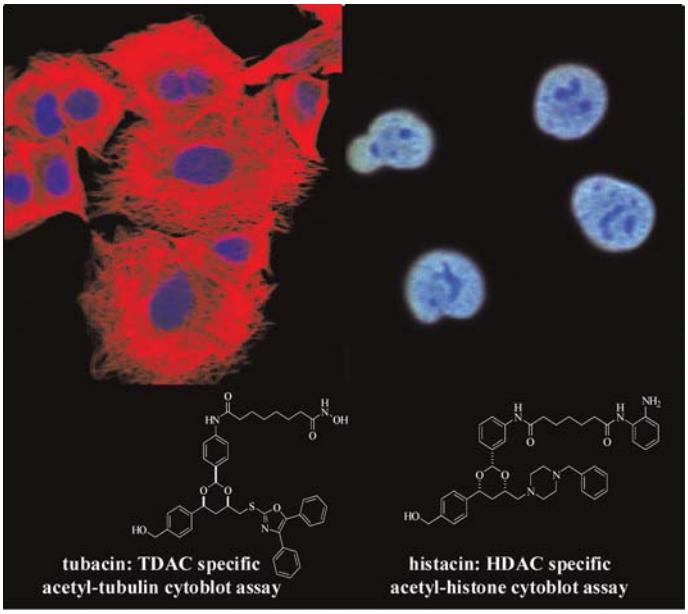


Figure 8. Selective inhibitors of α-tubulin (tubacin) and histone deacetylation (histacin) identified by chemical genetic screening [31].

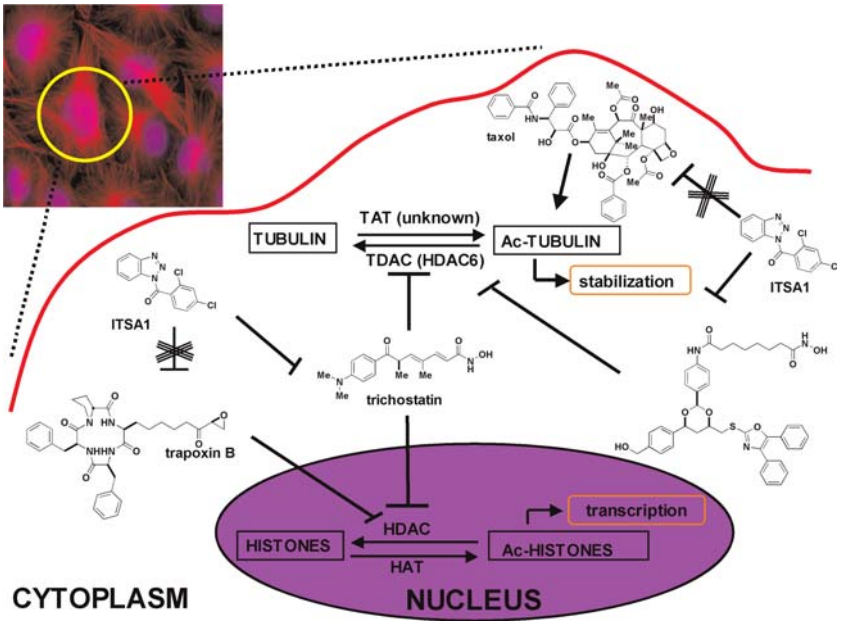


Figure 9. Molecular tools for the dissection of intracellular protein acetylation [30, 31].

Future Development

For chemical genetics to truly rival classical genetics, and for it to function as a general approach to dissecting biological mechanisms, there needs to be continued development and refinement of the techniques for screening and assessing complex patterns of phenotypic changes.

As demonstrated in the example described above for protein acetylation, it is worth noting the remarkable ability of antibodies to detect post-translational modifications of proteins and other biosynthetic events that occur intracellularly at a single-cell level. Antibodies differ from small molecules in their size, composition, and origin as they are immunoglobulins composed of both heavy and light chains, which are secreted by immune system cells. The ability to recognize epitopes as small as a single acetyl group within the context of chromatin or a single phosphate group on a protein within the cytoplasm of cells speaks to their specificity and power as markers of phenotypes. The development of an expanded collection of cell-state selective antibodies, and improved methods for multiplexing multiple probes in parallel or in series would have widespread utility for chemical genetics as part of cytoblot and image-based screens.

The further development of genetically-encoded probes that allow for imaging of signalling events and cellular processes in live cells in real-time will open up previously unexplored areas of cellular biology. In particular, the use of genetically encoded probes targeted to specific cell populations will be useful for creating more complex and physiologically relevant assays, particularly in animal models.

A prerequisite for many studies and the understanding of chemotype–phenotype proteome-wide approaches. With targets in hand, these efforts can be merged with structural biology efforts to look at atomic resolution interactions, and an examination of the degree to which specificity for targets influences the observed phenotypic effects.

It may also be possible to search for a “molecular recognition code(s)” that ultimately determines the mapping, both locally and globally, between molecules in multidimensional molecular descriptor spaces and multidimensional phenotypic descriptor spaces. These codes may be considered at a variety of levels, including more general categories that allow the prediction of properties relevant to the interaction with different subcellular structures (e. g. the mitochondria or cytoskeleton) or different biological systems (e. g. the xenobiotic transformation systems involved in drug metabolism). Knowledge of such codes would, as did knowledge of the genetic code, usher in a new era of research and medical advances that would allow the systematic modulation of gene product function based upon an understanding of molecular interactions.

Conclusion

Mendel's rules for considering the discreteness and combinatorics of inherited traits provided a foundation for classical genetics that has continued to provide insight into genotype–phenotype relations and the nature of heredity for more than a century [48, 49]. By using small molecules to perturb biological systems conditionally at the level of gene products, rather than at the level of genes themselves, chemical genomics promises to complement the use of classical genetic analysis to study a wide range of biological mechanisms and systems. Fortunately, unlike the complementarity principle discovered by the physicist Niels Bohr, which describes the impossibility of simultaneously observing both the wave and particle aspects of light and electrons, it is possible to simultaneously consider chemical and biological space. In doing so a fuller description of the interaction of small molecules with biological systems arises than if either of the two spaces is considered separately.

References

[1] Haggarty, S.J. (2005) The principle of complementarity: chemical versus biological space. *Curr. Opin. Chem. Biol*. **9:**296 – 303.

[2] Mitchison, T.J. (1994) Towards a pharmacological genetics. *Chem. Biol.* **1:**3 – 6.

[3] Schreiber, S.L. (2003) The small-molecule approach to biology: chemical genetics and diversity-oriented organic synthesis make possible the systematic exploration of biology. *Chem. Eng. News* **81:**51 – 61.

[4] Specht, K.M., Shokat, K.M. (2002) The emerging power of chemical genetics. *Curr. Opin. Cell. Biol*. **14:**155 – 159.

[5] Schreiber, S.L. (1998) Chemical genetics resulting from a passion for synthetic organic chemistry. *Bioorg. Med. Chem*. **6:**1127**–**1152.

[6] Stockwell, B.R. (2004) Exploring biology with small organic molecules. *Nature* **432:**846 – 854.

[7] Sharom, J.R., Bellows, D.S., Tyers, M. (2004) From large networks to small molecules. *Curr. Opin. Chem. Biol*. **8:**81 – 90.

[8] Clardy, J., Walsh, C. (2004) Lessons from natural molecules. *Nature* **432:**829 – 837.

[9] Schreiber, S.L. (2000) Target-oriented and diversity-oriented organic synthesis in drug discovery. *Science* **287:**1964 – 1969

[10] Arya, P., Joseph, R., Gan, Z., Rakic, B. (2005) Exploring new chemical space by stereocontrolled diversity-oriented synthesis. *Chem. Biol.* **12:**163 – 180.

[11] Shang, S., Tan, D.S. (2005) Advancing chemistry and biology through diversity-oriented synthesis of natural product-like libraries. *Curr. Opin. Chem.Biol.***9:**248 – 258.

[12] Liu, D.R., Li, X. (2004) DNA-templated organic synthesis: nature's strategy for controlling chemical reactivity applied to synthetic molecules. *Angew. Chem. Intl Ed. Engl.* **43:**4848 – 4870.

[13] Wang, L., Xie, J., Schultz, P.G. (2006) Expanding the genetic code. *Annu. Rev. Biophys. Biomol. Struct.* **35:**225 – 249.

[14] ChemBank; http://chembank.broad.harvard.edu/

[15] Biomolecular Interaction Network Database(BIND);http://www.bind.ca/Action

[16] PubChem; http://pubchem.ncbi.nlm.nih.gov/

[17] Weinstein, J.N., Myers, T.G., O'Connor, P.M. *et al*. (1997) An information-intensive approach to the molecular pharmacology of cancer. *Science* **275:**343 – 349.

[18] Agrafiotis, D.K., Lobanov, V.S., Salemme, F.R. (2002) Combinatorial informatics in the post-genomics Era. *Nature Rev. Drug Discov.* **1:**337 – 346.

[19] Oprea, T.I., Gottfries, J. (2001) Chemography: the art of navigating in chemical space. *J. Comb. Chem.* **3:**157 – 166.

[20] Oprea, T.I. (2002) Chemical space navigation in lead discovery. *Curr. Opin. Chem. Biol.* **6:**384 – 389.[

[21] Dobson, C.M. (2004) Chemical space and biology. *Nature* **432:**824 – 828.

[22] Lipinski, C., Hopkins, A. (2004) Navigating chemical space for biology and medicine. *Nature* **432:**855 – 861.

[23] Fishman, M.C., Porter, J.A. (2005) Pharmaceuticals: a new grammar for drug discovery. *Nature* **437:**491 – 493.

[24] Paolini, G.V., Shapland, R.H., van Hoorn, W.P., Mason, J.S., Hopkins, A.L. (2006) Global mapping of pharmacological space. *Nature. Biotechnol.* **24:**805 – 815.

[25] Stockwell, B.R., Haggarty, S.J., Schreiber, S.L. (1999) High-throughput screening of small molecules in miniaturized mammalian cell-based assays involving post-translational modifications. *Chem. Biol.* **6:**71 – 83.

[26] Mayer, T.U., Kapoor, T.M., Haggarty, S.J., R.W. King, R.W., S.L. Schreiber, S.L., T.J. Mitchison, T.J. (1999) Small molecule inhibitor of mitotic spindle bipolarity identified in a phenotype-based screen. *Science* **286:**971 – 974.

[27] Haggarty, S.J., Mayer, T.U., Miyamoto, D.T., Fathi, R., King, R.W., Mitchison, T.J., Schreiber, S.L. (2000) Dissecting cellular processes using small molecules: identification of colchicine-like, taxol-like, and other small molecules that perturb mitosis. *Chem. Biol.* **7:**275 – 286.

[28] Feng, Y., Yu, S., Lasell, T.K., Jadhav, A.P., Macia, E., Chardin, P., Melancon, P., Roth, M., Mitchison T., Kirchhausen, T. (2003) Exo1: a new chemical inhibitor of the exocytic pathway. *Proc. Natl Acad. Sci. U S A* **100:**6469 – 6474.

[29] Kau, T.R., Schroeder, F., Ramaswamy, S., Wojciechowski, C.L., Zhao, J.J., Roberts, T.M., Clardy, J., Sellers, W.R., Silver, P.A. (2003) A chemical genetic screen identifies inhibitors of regulated nuclear export of a Forkhead transcription factor in PTEN-deficient tumor cells. *Cancer Cell* **4:**463 – 476.

[30] Koeller, K.M., Haggarty, S.J., Perkins, B.D., Leykin, I., Wong, J.C. Kao, M.C., Schreiber, S.L. (2003) Chemical genetic modifier screens: small molecule trichostatin suppressors as probes of intracellular histone and tubulin acetylation. *Chem. Biol.* **10:**397 – 410.

[31] Haggarty, S.J., Koeller, K.M., Wong, J.C., Butcher, R.A., Schreiber, S.L. (2003) Multidimensional chemical genetic analysis of diversity-oriented synthesis-derived deacetylase inhibitors using cell-based assays. *Chem. Biol*. **10:**383 – 396.

[32] Haggarty, S.J., Koeller, K.M., Kau, T.R., Silver, P.A., Roberge, M., Schreiber, S.L. (2003) Small molecule modulation of the human chromatid decatenation checkpoint. *Chem. Biol.* **10:**1267 – 1279.

[33] Butcher, R.A., Schreiber, S.L. (2003) A small molecule suppressor of FK506 that targets the mitochondria and modulates ionic balance in Saccharomyces cerevisiae. *Chem. Biol.* **10:**521 – 531.

[34] Huang, J., Zhu, H., Haggarty, S.J., Spring, D.R., Hwang, H.,Jin, F., Snyder, M., Schreiber, S.L. (2004) Finding new components of the target of rapamycin (TOR) signaling network through chemical genetics and proteome chips. *Proc. Natl Acad. Sci. U S A* **101:**16594 – 16599.

[35] Mitchison, T.J. (2004) Small-molecule screening and profiling by using automated microscopy. *Chembiochem.* **29:**33 – 39.

[36] Perlman, Z.E., Mitchison, T.J., Mayer, T.U. (2005) High-content screening and profiling of drug activity in an automated centrosome-duplication assay, *Chembiochem.* **6:**145 – 151.

[37] Nieland, T.J., Feng, Y., Brown, J.X., Chuang, T.D., Buckett, P.D., Wang, J., Xie, X.S., McGraw, T.E., Kirchhausen, T., Wessling-Resnick, M. (2004) Chemical genetic screening identifies sulfonamides that raise organellar pH and interfere with membrane traffic. *Traffic* **5:**478 – 492.

[38] Zon, L.I., Peterson, R.T. (2005) In vivo drug discovery in the zebrafish. *Nature. Rev. Drug. Discov*. **4:**35 – 44.

[39] Haggarty, S.J., Clemons, P.A., Schreiber, S.L. (2003) Chemical genomic profiling of biological networks using graph theory and combinations of small molecule perturbations. *J. Am. Chem. Soc*. **125:**10543 – 10545.

[40] Forman, J.J., Clemons, P.A., Schreiber, S.L., Haggarty, S.J. (2005) SpectralNET – an application for spectral graph analysis and visualization. *BMC Bioinformatics* **6:**260.

[41] Yan, Y., Sardana, V., Xu, B., Homnic, C., Halczenko, W., Buser, C.A., Schaber, M., Hartman, G.D., Huber, H.E., Kuo, L.C. (2004) Inhibition of a mitotic motor protein: where, how, and conformational consequences. *J. Mol. Biol.* **335:**547 – 554.

[42] A.T. Balaban, A.T. (1976) *Chemical Applications of Graph Theory*. Academic Press, London.

[43] Kouzarides, T. (2000) Acetylation: a regulatory modification to rival phosphorylation? *E M B O J.* **19:**1176 – 1179.

[44] Grozinger, C.M., Schreiber, S.L. (2002) Deacetylase enzymes: biological functions and the use of small-molecule inhibitors. *Chem. Biol.* **9:**3 – 16.

[45] Sternson, S.M., Wong, J.C., Grozinger, C.M., Schreiber, S.L. (2001) Synthesis of 7200 small molecules based on a substructural analysis of the histone deacetylase inhibitors trichostatin and trapoxin. *Organic Lett.* **3:**4239 – 4242.

[46] Haggarty, S.J., Koeller, K.M., Wong, J.C., Grozinger, C.M., Schreiber, S.L. (2003) Domain-selective small molecule inhibitor of HDAC6-mediated tubulin deacetylation. *Proc. Natl Acad. Sci. U S A* **100:**4389 – 4394.

[47] Haggarty, S.J., Clemons, P.A., Wong, J.C., Schreiber, S.L. (2004) Mapping chemical space using molecular descriptors and chemical genetics: deacetylase inhibitors. *Comb. Chem. High Throughput. Screen*. **7:**669 – 676.

[48] Mendel, G. (1963) *Experiments in Plant Hybridization.* Harvard University Press, Cambridge.

[49] Morgan, T.H., Sturtevant, A.H., Muller, H.J., Bridges, C.B. (1915) *The Mechanism of Mendelian Heredity.* Henry Holt and Company, New York.