Thinking combinatorially

Stuart Kauffman* and Andrew D Ellington†

Biopolymers and chemical compounds with novel functions can be selected or screened from randomized libraries. Recently, it has become possible to augment the functions of biopolymers via the conjugation or incorporation of unnatural chemical moieties. In the future, it should prove possible to engineer systems that can self-evolve and thereby reveal unexpected emergent properties.

Addresses
*Santa Fe Institute, 1660 Old Pecos Trail, Suite A, Santa Fe, NM 87501, USA; e-mail: stu@santafe.edu
†CIStem Molecular and University of Texas, Department of Chemistry and Biochemistry, 26th Street and Speedway, Austin, TX 78712, USA; e-mail: andy.ellington@mail.utexas.edu

Current Opinion in Chemical Biology 1999, 3:256–259

http://biomednet.com/elecref/1367593100300256

 ISSN 1367-5931

Introduction

Two seminal events occurred in the war-heavy year of 1944. Erwin Schrodinger published his famous and famously small book, 'What is Life?' [1], and at the University of Pennsylvania, the first digital computer crackled into action. Slightly over fifty years later, molecular biology has fully come of age, and the computer-mediated information revolution has transformed much of the economy.

In 1990, with the near simultaneous publication of several articles on phage display and nucleic acid selection [2–5], molecular diversity and, soon after, combinatorial chemistry sprang into practical existence. Fifty years from 1990 is 2040. Can we begin to guess what shall have come into being, unleashed by the stunning novelty of our capacity to create and work with libraries of over 10^{15} molecular species, a number that rivals the diversity of proteins and nucleic acids already present in nature? Our own initial guess is that we have crossed a threshold as fundamental as the two traversed in 1944. Molecular diversity, writ large, already promises new medicines, enzymes, diagnostics, structures and new approaches to structure/function relationships, but even more importantly, points us towards a post-genomic era in medicine and biology, towards the development of increasingly complex networks of chemical reactions, towards new forms of self-reproducing molecular systems, and thus towards a 'general biology' freed from the constraints of terrestrial biology to consider the characteristics of new living systems both man-made and throughout the Cosmos.

This much we can begin to glimpse, and if history is any guide, we cannot yet glimpse the half of it. Nonetheless, in this review we can perhaps begin to map the major eras that are now upon us and those that are the next to arise. First and foremost, of course, is the cornucopia of molecular diversity that is now being sieved for biopolymers and chemicals with novel functions; next and nascently is the augmentation of biopolymer libraries with unnatural chemistries, and the augmentation of chemical libraries with biological properties such as replicability; and, finally, is the design and evolution of systems whose emergent properties are far greater than the sum of their parts, which is on the horizon.

Diversity

The recent past has been a necessary prologue to the prosperous present: Speigleman and co-workers' [6,7] early experiments with Qbeta replicase and the parallel insights of leaders such as Eigen and Schuster [8] and Orgel [9] clearly and cleanly foreshadowed the development of *in vitro* selection methods; Geysen *et al.*'s [10] fine work with random peptides on pin arrays (and the concomitant patent [P1]) is the intellectual forebear of most combinatorial chemistry that exists today; Smith's [5] first phage peptide libraries have spawned a rash of surface display methods; and finally the Ballivet–Kauffman patent (filed first in 1985 [P2]) bespeaks the power of combinatorial methods to transform chemistry and biology. Similarly powerful innovations in combinatorial chemistry and biochemistry continue to this day, such as Stemmer's inspired development of artificial recombination methods for protein engineering (DNA shuffling [11]).

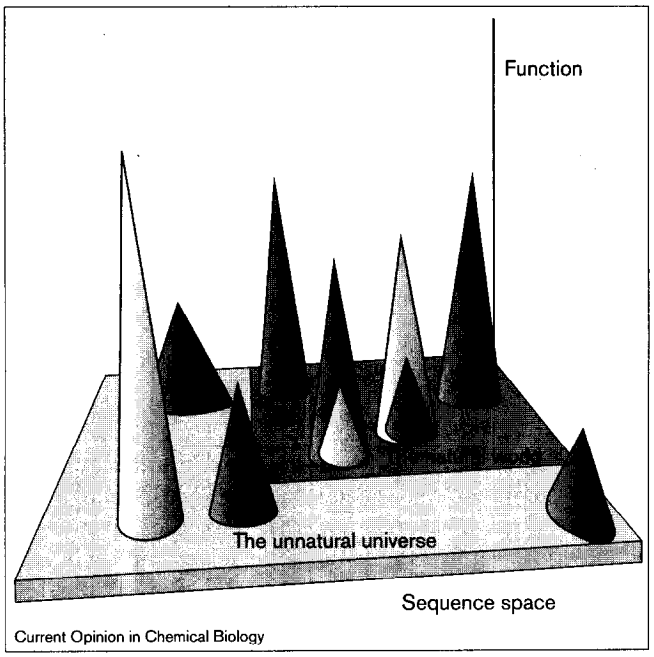
In the present, these advances have allowed nucleic acid binding species and catalysts to be selected *in vitro* from almost unimaginably large random sequence populations, peptides and proteins to be quickly optimized by selection, and a wealth of chemical compounds to be modularly constructed. When coupled with high-throughput screening methods, these technologies enable the genesis of a new 'chemical biology' that pairs compounds with phenotypes [12]. As the human genome project comes to completion and fine-grained genetic variations in populations are elucidated, the enhanced capacity in drug lead discovery enabled by molecular diversity will help create increasingly sculpted drugs to treat subsets of patients with precision for what might otherwise be orphan diseases. The ease of lead generation also implies the ease of diagnostic libraries, arrays and kits.

Nonetheless, no matter how impressive the selection of a ribozyme amide synthetase [13,14] or the identification of antibodies that can bind magnetite [15] may be, the fact remains that combinatorial biochemistry and chemistry are essentially short cuts for rational design. While the combinatorial methods may have for the moment outpaced rational design strategies, they cannot in the end overcome them. Just as computers quickly caught up with

and surpassed several thousand years of paper accounting, computer-assisted rational design will probably soon catch up with and surpass brute-force combinatorial efforts. For example, it has already proven possible to *de novo* design rather than screen for novel protein catalysts [16••]. To hasten the day when research costs are further pared by compound identification *in silico* rather than selection or screening *in vitro*, the link between rational design and combinatorial experimentation needs to be made manifest. There are currently relatively few efforts to explain why a particular combinatorial biopolymer or chemical is efficacious, nor to determine if the results of screens might be *a posteriori* recapitulated and eventually *a priori* mimicked by computers.

Encouragingly, some of the tools necessary to begin melding screening and design are already available. The concept of sequence space, popularized by Smith in 1970 [17], now underlies many forays into biochemical diversity (Figure 1). The notion of shape space (a multidimensional space of physical properties, including the three spatial dimensions and physical features such as charge, dipole moment, hydrophobicity and so on, where a point represents shape), introduced by Perelson and Oster in 1979 [18], is also paying off. For example, their observation that a sufficient diversity of molecular species, such as the human immune repertoire, could 'saturate' shape space in turn suggests that it may be possible to use automated, high-throughput selections of antibodies [19,20] to develop therapeutic reagents for known and unknown chemical and biological weapons. Understanding the nature of molecular fitness landscapes [21–23] (Figure 1) is becoming increasingly pertinent to increasing the speed and efficiency of lead optimization.

Figure 1



Molecular fitness landscapes. The x–y axis represents the potential diversity of molecular species, of whatever composition. For example, for a nucleic acid or protein landscape, the plane might represent all possible sequences (obviously such a space would be n-dimensional, depending on the length of the nucleic acid or protein, but can be most conveniently represented as plane). The z axis represents functionality. Each sequence in the plane will have an associated function along the z axis. Hills or menhirs on this surface thus represent molecular species of varying functionality. The 'natural world' comprises all molecules derived from organisms, whereas the 'unnatural universe' comprises all possible molecules. Ongoing experimental explorations in the era of diversity have begun to explore what molecules may lie outside the natural world. New lines of inquiry that have begun to define the era of augmentation change the landscape, leading to the modulation of the functionality of old molecules or to the appearance of completely new ones. The era of emergent properties cannot be properly represented on such a static landscape, since complex systems that begin to traverse a fitness landscape will of necessity change it and themselves in the process.

Augmentation

For the most part, combinatorial biochemistry and chemistry are simple extensions of the core Darwinian dogma of vary, select and amplify. The methods that have been developed for searching sequence and fitness landscapes may allow man to ponder a larger number of variants than nature can, but the throughput increase is a quantitative change, not a qualitative one. However, technological innovations are also beginning to give rise to new types of landscapes with new types of molecules, molecules with properties that deviate significantly from what has previously been observed in either nature or industry. These innovations can generally be described as 'augmenting' the existing chemistries of biopolymers or organic compounds.

For example, just as natural products have long been chemically modified to improve their pharmacokinetic properties, biopolymers are now being augmented with new chemistries. Nucleic acids have proven surprisingly adept at incorporating and utilizing unnatural nucleotides and appended chemical moieties [13,14,24]. Proteins have been somewhat more resistant, but methods for the introduction of unnatural amino acids now exist [25,26]. Conversely, whereas organic chemists have long copied nature's structures, they are now also beginning to incorporate some of the properties of biopolymers into organic compounds. For example, to assist with high-throughput screening, individual compounds in chemical libraries have been tagged with encoded or replicable biopolymers [27,28].

Most interestingly, the properties of entire systems are now being 'bred' with one another. While a cell has previously been the only intermediary between nucleic acid information and protein phenotype, by coupling the translation-terminating antibiotic puromycin to mRNAs, Roberts and Szostak [29••] have developed a system by which nucleic acids can be directly connected to their translated products. This system should soon herald the direct selection of protein function, and it is likely that the selected nucleoproteins will have unique properties of their own

that are an admixture of each partner. Furthermore, entire biological systems may soon serve as combinatorial libraries for the presentation of novel chemistries. For example, phage display libraries can potentially provide diverse, replicable structural scaffolds for the pharmacophores in chemical combinatorial libraries. Bertozzi and co-workers' [30] recent advances in remodeling cell surfaces could similarly lead to the development of combinatorial chimeras of chemistry and biology.

Emergent properties

So far, we have considered only systems whose wholes are more or less the sum of their parts, even when those parts are quite chemically distinct from one another. While the augmentation of chemistry with biology and *vice versa* may lead to new landscapes, those landscapes are nonetheless still a fixed function of the molecules that underlie them. As the combinatorial revolution progresses, though, we can next reach out to systems that have the capability to actually innovate and reinvent themselves and their associated landscapes. In a simple sense, this will merely mean systems of increasing functional complexity. In a more profound sense, however, it will mean systems that incorporate self-reproduction and feedback loops during their evolution.

To help envision the emergent properties of combinatorial chemistries, we suggest a parallel with the preceding digital revolution. Early, electronically implemented search algorithms identified patterns in data based on preset criteria. Today, however, neural networks employ feedback loops that allow weighting between the criteria to be altered, and thus can learn over time to identify patterns. In analogy, screens or selections can search for particular compounds that meet preset criteria, but a more complex system such as a cell employs multiple feedback loops that allows it to develop new compounds while adapting to new environments.

Some chemical and biological systems are just beginning to approach this level of complexity. Wright and Joyce [31••] have described continuous evolution systems that yielded catalysts similar to those produced in more canonical selection experiments. By incorporating the feedback loop of replication directly onto and into a selection for catalytic function, however, these systems have also shown an emergent, unexpected property — the evolution of replication parasites [32]. While such emergent properties may not always be desirable, they do illustrate the ability of truly complex systems to change the very nature of the landscapes on which screening or selection is taking place, and thus to change 'the rules of the game.' Given that discrete peptides and organic compounds have been shown to be capable of self-replication [33], the emergence of novel properties from combinatorial libraries of such replicators can be readily imagined. As an example, Ghadiri and coworkers [34] have shown the emergence of coupled hypercycles of replicating peptides.

The most obvious venue for exploring emergent chemical and biochemical properties, though, is within and between cells. Combinatorial experimentation on a grand scale has frequently occurred during the evolution of cells. Even ignoring the obvious example of the diversity of secondary metabolites, animals and plants are the result of sustained exchange and experimentation involving the genes, proteins and metabolic pathways of eukaryotes and their mitochondrial and chloroplast endosymbionts [35•].

Just as analyses of fitness landscapes and hypercycles anticipated the results of selection experiments, analyses of genomes and gene regulatory networks may now assist in the development of novel 'combinatorial' metabolisms and organisms. For example, analyses of artificial genetic regulatory networks already suggest that eukaryotic genomes will typically partition into genes with fixed activities in all cell types (i.e. housekeeping genes) and into isolated islands of genes whose activities vary between cell types [36] . As these theoretical methods are applied to real world problems, it may prove possible to guide the next level of combinatorial syntheses, such as which enzymes should be optimized in parallel to generate emergent metabolic pathways, which regulatory and structural genes should be combined to yield emergent cell types (such as artificial endosymbionts), and even which regulatory elements should be co-stimulated to produce an emergent organismal phenotypes (such as a new limb or organ).

Conclusions

At root, though, all of the eras we have described — diversity, augmentation and emergent properties — harken back to an operational understanding of selection pressure and a theoretical understanding of landscapes. Given enough diversity and a big enough selective hammer, researchers can quickly pound out the details of new molecules and systems; however, even the smallest protein landscapes are already well out of reach of complete experimental searches and can quickly beggar even the largest supercomputers. To overcome these limitations, organisms, with a fine appreciation of the natural landscapes they live on, invented recombination to jump between functional peaks, and technologists have followed with artificial recombination, in the guise of DNA shuffling. Now, we stand on the verge of attempting to engineer systems that, even if they are not defined as life (a term more suited to poetry than science), will have the characteristics of organisms, such as adaptation, learning and communication. In this respect, our comparison to the digital world comes full circle, as these are precisely the same characteristics that many electronic devices are now imbued with. Indeed, as molecular 'alligator clips' between the biochemical and digital worlds are constructed (e.g. peptides that interface with semiconductors, ribozyme logic gates and neurons wired to operational amplifiers) the most efficient algorithm for screening and selection may involve establishing feedback loops between the molecules that

carry out combinatorial experiments and the devices (other than man) that evaluate them.

Acknowledgement

Andrew D Ellington would like to thank David Conrad of the Naval Research Laboratory for helpful discussions.

References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Schrodinger E (Ed): *What is Life?* Cambridge: Cambridge University Press; 1944.
2. Ellington AD, Szostak JW: **In vitro selection of RNA molecules that bind specific ligands.** *Nature* 1990, **346**:818-822.
3. Tuerk C, Gold L: **Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase.** Science 1990, **249**:505-510.
4. Cwirla SE, Peters EA, Barrett RW, Dower WJ: **Peptides on phage: a vast library of peptides for identifying ligands.** *Proc Natl Acad Sci USA* 1990, **87**:6378-6382.
5. Scott JK, Smith GP: **Searching for peptide ligands with an epitope library.** *Science* 1990, **249**:386-390.
6. Mills DR, Peterson RL, Spiegelman S: **An extracellular Darwinian experiment with a self-duplicating nucleic acid molecule.** *Proc Natl Acad Sci USA* 1967, **58**:217-224.
7. Kramer FR, Mills DR, Cole PE, Nishihara T, Spiegelman S: **Evolution in vitro: sequence and phenotype of a mutant RNA resistant to ethidium bromide.** *J Mol Biol* 1974, **89**:719-736.
8. Eigen M, Schuster P: **The hypercycle. A principle of natural self-organization. Part A: Emergence of the hypercycle.** Naturwissenschaften 1977, **64**:541-565.
9. Orgel LE: **Selection *in vitro*.** *Proc R Lond B Biol Sci* 1979, **205**:435-442.
10. Geysen HM, Meloen RH, Barteling SJ: **Use of peptide synthesis to probe viral antigens for epitopes to a resolution of a single amino acid.** *Proc Natl Acad Sci USA* 1984, **81**:3998-4002.
11. Stemmer WP: **Rapid evolution of a protein *in vitro* by DNA shuffling.** *Nature* 1994, **370**:389-391.
12. Zuhlsdorf MT: **Relevance of pheno- and genotyping in clinical drug development.** *Int J Clin Pharmacol Ther* 1998, **36**:607-612.
13. Wiegand TW, Janssen RC, Eaton BE: **Selection of RNA amide synthases.** *Chem Biol* 1997, **4**:675-683.
14. Zhang B, Cech TR: **Peptide bond formation by *in vitro* selected ribozymes.** *Nature* 1997, **390**:96-100.
15. Barbas CFD, Rosenblum JS, Lerner RA: **Direct selection of antibodies that coordinate metals from semisynthetic combinatorial libraries.** *Proc Natl Acad Sci USA* 1993, **90**:6385-6389.
16. •• Hellinga HW: **Computational protein engineering.** *Nat Struct Biol* 1998, **5**:525-527.

This is an excellent review of recent advances in rational computational design techniques. While rational design has often proven less successful than screening or selection techniques, this review reveals why digital methods may soon overtake analogue.

17. Smith JM: **Natural selection and the concept of a protein space.** *Nature* 1970, **225**:563.
18. Perelson A, Oster G: **Theoretical studies of clonal selection: minimal antibody repertoire size and reliability of self-nonself discrimination.** *J Theo Biol* 1979, **81**:645.
19. Barry MA, Lai WC, Johnston SA: **Protection against mycoplasma infection using expression-library immunization.** *Nature* 1995, **377**:632-635.
20. Johnston SA, Barry MA: **Genetic to genomic vaccination.** *Vaccine* 1997, **15**:808-809.
21. Wright S: **The shifting balance theory and macroevolution.** *Annu Rev Genet* 1982,**16**:1-19.
22. Kauffman SA, Levin S: **Towards a general theory of adaptive walks on rugged landscapes.** *J Theo Biol* 1987, **128**:11-45.
23. Eigen M, McCaskill J, Schuster P: **The molecular quasi-species.** *J Phys Chem* 1988, **92**:6881.
24. Verma S, Eckstein F: **Modified oligonucleotides: synthesis and strategy for users.** *Annu Rev Biochem* 1998, **67**:99-134.
25. Noren CJ, Anthony-Cahill SJ, Griffith MC, Schultz PG: **A general method for site-specific incorporation of unnatural amino acids into proteins.** *Science* 1989, **244**:182-188.
26. Lemeignan B, Sonigo P, Marliere P: **Phenotypic suppression by incorporation of an alien amino acid.** *J Mol Biol* 1993, **231**:161-166.
27. Needels MC, Jones DG, Tate EH, Heinkel GL, Kochersperger LM, Dower WJ, Barrett RW, Gallop MA: **Generation and screening of an oligonucleotide-encoded synthetic peptide library.** *Proc Natl Acad Sci USA* 1993, **90**:10700-10704.
28. Ohlmeyer MH, Swanson RN, Dillard LW, Reader JC, Asouline G, Kobayashi R, Wigler M, Still WC: **Complex synthetic chemical libraries indexed with molecular tags.** *Proc Natl Acad Sci USA* 1993, **90**:10922-10926.
29. •• Roberts RW, Szostak JW: **RNA–peptide fusions for the in vitro selection of peptides and proteins.** *Proc Natl Acad Sci USA* 1997, **94**:12297-12302.

In a biotechnology tour-de-force, RNA molecules were engineered to attach themselves to nascently translated proteins. The resultant 'coupled translation' system brilliantly circumvented the necessity of cellular (or phage) compartmentalization for peptide and protein selection.

30. Mahal LK, Yarema KJ, Bertozzi CR: **Engineering chemical reactivity on cell surfaces through oligosaccharide biosynthesis.** *Science* 1997, **276**:1125-1128.
31. •• Wright MC, Joyce GF: **Continuous *in vitro* evolution of catalytic function.** *Science* 1997, **276**:614-617.

Ribozyme catalysis was coupled with a continuous amplification reaction to generate a system in which only the fastest ribozymes preferentially survived and gained access to reproductive resources. The behavior of this *in vitro* system was similar to that of many living systems, and begs the question of what constitutes an organism.

32. Breaker RR Joyce GF: **Emergence of a replicating species from an in vitro RNA evolution reaction.** *Proc Natl Acad Sci USA* 1994, **91**:6093-6097.
33. Lee DH, Granja JR, Martinez JA, Severin K, Ghadiri MR: **A self-replicating peptide.** *Nature* 1996, **382**:525-528.
34. Lee DH, Severin K, Yokobayashi Y, Ghadiri MR: **Emergence of symbiosis in peptide self-replication through a hypercyclic network.** *Nature* 1997, **390**:591-594.
35. • Martin W, Schnarrenberger C: **The evolution of the Calvin cycle from prokaryotic to eukaryotic chromosomes: a case study of functional redundancy in ancient pathways through endosymbiosis.** *Curr Genet* 1997, **32**:1-18.

An outstanding review of the molecular exchanges, deletions and symbioses that entrenched chloroplasts in plants and in consequence bestowed the ability to fix carbon dioxide.

36. Kauffman SA: *Origins of Order: Self-Organization and Selection in Evolution.* Oxford: Oxford University Press; 1993.

Patents

P1. Geysen H: **Antigenically active amino acid sequences.** November 8 1984; US4708871.

P2. Ballivet X, Kauffman X: **Process for obtaining DNA, RNA, peptides, polypeptides or proteins by recombinant DNA techniques.** March 30 1985; SW137985.