The Simulated Emergence of Distributed Environmental Control in Evolving Microcosms

Keith L. Downing
Department of Computer Science
Norwegian University of Science and Technology (NTNU)
7020 Trondheim
Norway
keithd@idi.ntnu.no

Abstract This work continues investigation into Gaia theory [Lovelock, (1995) *The ages of Gaia*, Oxford University Press] from an artificial life perspective [Downing, (2000) in *Proceedings of the 7th International Conference on Artificial Life*, (pp. 90–99) MIT Press], with the aim of assessing the general compatibility of emergent distributed environmental control with conventional natural selection. Our earlier system, GUILD [Downing and Zvirinsky, (1999) *Artificial Life, 5,* 291–318], displayed emergent regulation of the chemical environment by a population of metabolizing agents, but the chemical model underlying those results was trivial, essentially admitting all possible reactions at a single energy cost. The new model, METAMIC, utilizes abstract chemistries that are both (a) constrained to a small set of legal reactions, and (b) grounded in basic fundamental relationships between energy, entropy, and biomass synthesis/breakdown.

To explore the general phenomena of emergent homeostasis, we generate 100 different chemistries and use each as the basis for several METAMIC runs, as part of a *Gaia hunt.* This search discovers 20 chemistries that support microbial populations capable of regulating a physical environmental factor within their growth-optimal range, despite the extra metabolic cost. Case studies from the Gaia hunt illustrate a few simple mechanisms by which real biota might exploit the underlying chemistry to achieve some control over their physical environment. Although these results shed little light on the question of Gaia on Earth, they support the possibility of emergent environmental control at the microcosmic level.

1 Introduction

In the past few decades, the idea that an evolving microbial community can alter its physical surroundings has become less controversial, even accepted, among natural scientists. However, rousing debates begin at the mention of Gaia theory [15], which goes further by speculating that these biotic influences are homeostatic, regulating the environment within a physicochemical regime that maximally benefits the organisms themselves.

The controversy surrounding Gaia stems less from the general concept of emergent distributed control and more from Lovelock's poetic descriptions of Earth as an evolving homeostatic superorganism. The renowned evolutionary biologist Richard Dawkins [4], was quick to assault this metaphor from several directions: (a) natural selection cannot possibly apply at a planetary level, and (b) assuming natural selection at the individual level, then *cheater* organisms that do not pay the extra costs associated with Gaian

regulation could still reap its benefits and gain a selective advantage, thereby destroying cooperative regulation. Dawkins then challenged the scientific community to come up with a feasible model of Gaia that was compatible with natural selection as we know it.

One year later, Watson and Lovelock's famous Daisyworld model [24] clearly illustrated the emergence of distributed biotic control. Daisyworld has since become the poster-child for Gaia. However, from the ALife perspective, the actual emergence in Daisyworld seems too contrived, since all of the necessary genotypes are hard-wired into the initial population of daisy species. There is no emergence of new forms, only shifts in the species distribution.

Our research enlists the tools of ALife: genetic algorithms and individual-based simulations, in search of more realistic evolutionary emergent biotic control in an abstract virtual world where organisms rely on genetically determined metabolisms for energy and biomass production. In doing so, we hope to fortify the Daisyworld efforts to debunk the myth that the emergence of distributed homeostatic phenomena requires something more elaborate and fantasy filled than standard natural selection.

Many different abstract chemistries are used as the backdrop for these evolving metabolisms, in hopes of assessing the general likelihood of emergent microbial regulation of their environment about a set-point that maximizes metabolic reaction rates. Our results show that emergent regulation is quite common, although certainly not inevitable. Dawkin's cheaters turn out to be present but hardly ubiquitous, since an organism's contribution to global regulation can also have a local selective benefit. The emergence of regulation then hinges on a trade-off, brokered by natural selection, between the costs and benefits of Gaian cooperation.

It should be clear from the outset that the goal of our work is to use ALife tools to test the general *plausibility* of the evolutionary emergence of Gaian-like distributed environmental control by the biota: to investigate the reconciliation of Gaia theory with natural selection. If this plausibility can be established, then biologists and Earth scientists can begin to discuss objectively the possibility that Gaia actually occurs here on Earth. At present, the natural evidence is quite fascinating, but hardly convincing. Regardless, we feel that the theory and its implications are so important that they should not be summarily rejected on principle, due to extreme interpretations of Lovelock's writings. Rather, we believe that Gaia theory has promising analytic foundations in the theory of complex adaptive systems, as supported by Daisyworld and our own simulations.

Of course, a trillion bytes do not make a trilobite: Although ALife models typically beat standard differential equations for investigating emergent phenomena, the gap between cyberspace and nature is still immense. If the best meteorological models cannot accurately predict the weekend weather, then how can we expect to recreate a few billion years of evolution? No amount of real data in these models can significantly improve our odds vis-a-vis the Gaia-on-Earth question. However, we hope that evolutionary biologists can still see the merits of these models as evidence that emergent distributed control by evolving species (a) requires neither divine intervention nor ecosystem-level selection, and (b) is not necessarily derailed by the over-simplified concept of cheaters. From there, they can address the empirical question of Gaia on Earth from a more solid analytical base, free from the poetic misconceptions that have haunted Gaia research.

1.1 Gaia Theory

The essence of Gaia theory is trivially stated, but it evades clear proof or refutation. Volk [23] best sums up the Gaian concept as "life begetting life." The originators of Gaia theory, Lovelock and Margulis [15], took a then (1970s) controversial view and added a final eyebrow-raising twist to agitate a large sector of the natural science community. Things have only partially calmed down since.

Lovelock and Margulis' springboard was the view that biota (particularly microorganisms) can significantly affect their physicochemical environment, primarily by chemical means. Today, plenty of evidence supports that claim; geochemists and geophysicists have quantified many of the biotic influences upon the planet [20]. Now, the main controversy revolves around Gaia theory's added twist: Living organisms affect the environment *in self-beneficial ways*.

Under overzealous interpretation, this self-serving clause can conjure up distored images of bacteria performing particular metabolic activities *for the sole purpose* of altering the balance of greenhouse gases and moving global climate into a growth-optimal range, or of large climate-regulating ecosystems functioning as superorganisms and evolving as single units under a selective pressure to become better controllers. The horror stories are endless.

However, a more serious analysis [13, 14, 23, 25] uncovers the very plausible possibility that the means by which organisms collectively affect the environment can also provide a local selective advantage to the individual organisms. The global side effect is free, and relatively unrestricted in the short term. However, any large-scale side effects that prove detrimental to the species could eventually be selected against due to the species' demise.

A classic example is rock weathering, wherein tree roots break down rocks to liberate mineral nutrients for immediate gain. In addition, this process consumes CO_2 as a reactant, which is drawn down from the atmosphere into the soil. By enhancing nonbiotic rock weathering by a factor of hundreds or thousands [21], the biota exert a huge influence upon atmospheric CO_2 and, via the greenhouse effect, the global temperature.

In terms of homeostasis, one of many very plausible scenarios is the following. If global CO_2 levels rise, then tree growth is stimulated (photosynthesis consumes CO_2 to produce plant biomass) and more roots are formed. This increases rock weathering and CO_2 drawdown, thus preventing a rise in global temperature that, above certain thresholds, could eradicate most plant and animal life. Similarly, a large drop in atmospheric CO_2 would reduce plant growth, weathering and CO_2 drawdown, thus elevating CO_2 and preventing extreme temperature drops. In short, life preserves viable conditions for life. For more details of the above scenario along with a host of other examples, see [13, 14, 23].

As the ice ages and other climatic perturbations indicate, the biota can never qualify as an optimal regulator [20]. Still, life has managed to persist here on Earth through 3.6 billion years and numerous extreme geophysical disturbances. Either way, the hallmark of Gaia is not perfect environmental control, but control governed to some significant extent by the biota. In short, *life is in the loop*, and the living, dying, and evolving of organisms can drive the physical and chemical factors of the loop into different steady states—just as those states can influence life.

However, the adaptability of life injects an interesting asymmetry into these loops. While the sensitivity of atmospheric temperature to CO_2 is an immutable physical relationship (all else being equal), the sensitivity of a species to temperature may change via evolution. But the mere fact that the biota *can* adapt to physical perturbations does not imply that it *must*, that is, that life will always give in to the physical forces. Since the biota can influence the environment, there is no reason to rule out the possibility of life countering a physical push with an opposing shove. This biotic ability to fight back (albeit unconsciously, indirectly, and inadvertently) lies at the heart of Gaia theory. And the occasional absence of the previous three parenthesized adverbs in Gaian literature lies at the heart of the Gaia controversy.

Also, since the idea of nonpassive biota violates strong adaptationist principles, it is unpopular with many biologists, who also argue that cheater organisms could break the

regulatory loops by reaping the benefits of environmental stability without paying the costs [4]. However, as mentionned above, the global influences are often side effects of locally favorable behavior, so many Gaian costs that an organism pays actually purchase immediate survival benefits—the Gaian effects are accidental, and free. For more on the Neo-Darwinian critique of Gaia theory and the Gaian rebuttal, see [9, 12].

To date, Gaia theory's most convincing evidence comes from a simple elegant computer model, Watson and Lovelock's Daisyworld [24]. It clearly illustrates how a few species of organisms (daisies) can control a global physical factor (temperature) via their distributed competitive interactions and their local influences upon albedo. This competitive element is synonymous with natural selection to Daisyworld adherents, but it amounts to little more than the hardwiring of two (or in modified versions, 10 or 100) genotypes and allowing them to fight it out. In short, Daisyworld sidesteps the question of how these distributed regulatory systems might emerge via genetic permutations over evolutionary time. And without a plausible *ultimate causal explanation*, Daisyworld cannot fend off all Neo-Darwinian assaults.

1.2 Improving the GUILD Model

Our previous system, GUILD [9], borrowed some key concepts from Daisyworld, mixed in Volk's [23] ideas on biochemical guilds, and added genetic search through a large space of potential metabolisms. The emerging virtual microbial communities exhibited two key Gaian characteristics: (a) nutrient recyling, and (b) environmental regulation. Since both traits emerged over the course of tens to hundreds of evolutionary generations, these results offered support to the reconciliation of Gaia theory with natural selection.

Since Gaia theory often implies an unconscious cooperation among many populations and/or species, genera, and so forth, its relationship to standard individual-based natural selection has always been tenuous. However, by adding an abstract genetic basis to some of the key Daisyworld mechanisms, GUILD enhanced support for these ties that are so essential for bringing Gaia theory into the realm of plausible and testable biological hypotheses.

However, GUILD relied on an overly abstract chemistry that (a) permitted all intermolecular transformations, (b) conserved mass only at the highest level of abstraction, and (c) had no specific thermodynamic constraints upon reactions. For example, an organism was free to convert M units of compound A into M units of compound B, for all compounds A and B. The transformation did involve an energy cost, but one based solely on M, not on A or B. This permitted a relative metabolic free-for-all in which diversity quickly arose, along with the concomitant recycling of compounds. Once recycling was in full swing, the organisms began to control environmental nutrient concentrations and steer them toward levels that best supported life: the quintessence of Gaia. All of this was transparent to the organisms themselves and guided by the gentle hand of natural selection, which worked on the genomes that determined the metabolisms.

Although simple general models can offer vital insights into complex biological phenomena, the key details of the underlying abstraction must remain faithful to the natural situation. By allowing an unconstrained set of chemical reactions, GUILD sufficiently violated that faith to bring its results into serious question. Our new system, METAMIC (metabolically abstract microorganisms), attempts to remedy that shortcoming.

METAMIC employs the MD-CHEM (modular designer chemistry) module to generate random abstract chemistries that conform to user specifications, such as the maximum number of compounds and reactions. In addition, MD-CHEM respects a few fundamental biochemical constraints. It differentiates between atoms and molecules, and the former are conserved in all reactions. A simple thermodynamic constraint links

the synthesis (breakdown) of larger molecules from (into) smaller ones with energy consumption (release). Hence, metabolism can be partitioned into anabolic (biomass producing) and catabolic (biomass burning) stages; and an organism's fitness reflects its ability to exploit the underlying chemistry and overall ecological situation (in terms of the other biochemical guilds present in its environment) to produce energy and build biomass.

METAMIC includes a set of physical variables that (a) have a direct effect upon metabolic rates, and (b) are influenced by the relative concentrations of certain chemical compounds. For example, in the real world, atmospheric temperature is an important physical factor that both influences biotic growth and is affected by CO_2 and CH_4 levels (which the biota can alter). In METAMIC, particular global chemical ratios will move the physical variables into ranges that are optimal for metabolism and growth. The fundamental Gaian question is whether or not the evolving biota can unconsciously regulate chemical levels so as to achieve these physical optima.

To investigate the emergence of distributed environmental control as a general phenomena, we run METAMIC on a host of different MD-CHEM chemistries and look for signs of physical-variable regulation. A test of 100 chemistries that support life in METAMIC reveals 20 that satisfy a relatively conservative control metric.

This article provides a detailed description of METAMIC and MD-CHEM along with the results of a *Gaia hunt* within a restricted space of abstract chemistries. Several case studies reveal the mechanisms by which a population of microbes, whose metabolisms are subject to natural selection, unconsciously can evolve to regulate their environment for the benefit of life itself.

2 MD-CHEM: The Modular Designer Chemistry Generator

Although a wide variety of artificial chemistries already exist [5, 6, 10, 22], these generally involve complex simulations to generate a set of stable compounds. Basically, the artificial chemistry *is* the artificial life system. However, to investigate the general ability of an evolving microbial community to exploit the underlying chemistry to achieve environmental control, a different type of artificial chemistry system is needed: one that can easily produce hundreds of distinct chemistries.

To support simulations of metabolism, an artificial chemistry must incorporate some notion of energy and its relationship to the formation and breakdown of large compounds, such as the proteins, carbohydrates, and lipids that constitute biomass. In looking to real biochemistry for inspiration, one finds a host of important basic concepts such as electron valences, bond types, redox pairs, and so forth, but their analogues in an artificial chemistry may only add superfluous detail. Furthermore, the energy yields of real chemical reactions are functions of the three-dimensional physical structures of the individual compounds—again, not an easy add-on to an artificial chemistry, even in one or two dimensions. In short, real biochemistry does not appear to provide a short-list of basic concepts that are easily mapped into fully functioning realistic models at a higher level of abstraction. Most artificial chemistries capture some of the key biochemical concepts, but with only a few notable exceptions [16], they ignore energy.

MD-CHEM generates artificial chemistries by ignoring atomic-level dynamics and choosing random regroupings of random reactant sets, constrained only by user-specified parameters such as the desired numbers of compounds and reactions. Reaction energies stem from statistical entropy differences between reactants and products, as explained below. The system is only weakly constructive, since new compounds arise stochastically, but not on the basis of any first-principle physical relationships between the atoms, such as their potentials for bond formation. So chemistry construction in MD-CHEM is a purely algorithmic (albeit nondeterministic) process, guided not by an

interaction dynamic, but by simple trial-and-error search for a constraint-satisfying set of compounds and reactions.

The *designer* aspect of MD-CHEM is quite simple: A user specifies several hard constraints and biases, and the system then generates a chemistry that meets the specification. These chemistries are *modular* in the sense that they easily plug into artificial life simulators.

The key hard constraints in an MD-CHEM specification are

1. N_a—the number of legal atoms,

2. N_c—the total number of compounds to be generated,

3. N_r—the total number of reactions to be generated,

4. N_{ic}—the number of initial randomly generated compounds,

5. R_{cs}—the range of legal compound sizes,

6. R_{rs}—the range of reaction sizes, in terms of the minimum and maximum number of reactants/products.

The essential biases are

1. B_{ics}—the bias of initial compound sizes. This is a normalized list of probabilities, one for each legal size in R_{cs}, which governs stochastic size choice of the initial N_{ic} randomly generated compounds.

2. B_{rs}—the bias of reaction sizes, a probability distribution for the sizes in R_{rs}, which affects the stochastic choice of both (a) the number of reactants chosen for each random reaction, and (b) the number of product compounds that the reactant atoms are partitioned into.

The basic algorithm for chemistry generation appears in Figure 1.

For example, if the atomic set is {abc}, R_{cs} = [3, 6], R_{rs} = [2, 4], and N_{ic} = 4, then MD-CHEM begins by generating four initial compounds, such as a_2b, abc, ac_2, c_3, where subscripts denote the number of each atom and no subscript implies a single atom.

The reaction generator would then take between two and four of these compounds, with possible duplicates, such as {a_2b, ac_2, ac_2}, to form the reactant group. All atoms are then thrown into a set: {$a\ a\ b\ a\ c\ c\ a\ c\ c$} that is sent to REACT-COMP, where it is randomly permuted: {$c\ b\ a\ c\ a\ c\ a\ a\ c$}, and partitioned into two to four subsets: {$c\ b\ a\ !\ c\ a\ !\ c\ a\ a\ c$}.

Since no interatomic relationships are modeled, MD-CHEM is insensitive to the order of atoms within a compound and converts it to a canonical form. Hence, the product group for the above example becomes {abc, ac, a_2c_2}. Since the latter two compounds are new, MD-CHEM will only add the new reaction if $N_c - n_c \geq 2$.

The new reaction is then

$$a_2b + 2ac_2 \Longrightarrow abc + ac + a_2c_2 \tag{1}$$

The use of the num-attempts and max-attempts variables in Figure 1 indicates that GENCHEM makes a finite number of attempts to generate N_r reactions. Note that once $n_c = N_c$, any reaction that generates additional new product compounds will be rejected in REACT-COMP. In cases where $N_r > 2N_c$, or even $N_c \leq N_r \leq 2N_c$, the algorithm will occasionally time out before generating N_r reactions.

```
Program GENCHEM
C = { }; The set of compounds
R = { }; The set of reactions
```
$n_c = 0$; The size of C
$n_r = 0$; The size of R
```
RG, PG: reactant product groups.
```
Initialize C to N_{ic} randomly-generated compounds.
```
num-attempts = 0
Repeat
 RG = random group of compounds from C
 Collect all atoms from RG into one set, S
```
$$p = e^{\frac{-k_e(N_c - n_c)}{N_c}}$$
```
 known-compound = TRUE
 PG = { };
```
$n_p = 0$; Number of products
N_p = random-integer-in(R_{cs})
```
 While known-compound and random-number ≤ p
```
        and $n_p < N_p$
```
  known-compound = find-known-compound(S).
  if known-compound
     Add known-compound to PG.
```
     $n_p = n_p + 1$
```
  Remove atoms in known-compound from S.
 End while
```
 If REACT-COMP(S,C,R,RG,PG,$N_p - n_p$)
```
  Then Add [ RG → PG] to R, and update
```
n_r
```
  Else num-attempts = num-attempts + 1
```
Until $n_r = N_r$ or num-attempts ≥ max-attempts
```
End Program

Procedure REACT-COMP(S, C, R, RG, PG, NP)
 ; This procedure randomly completes reactions
 Randomly permute S.
```
 $PG' = PG$
```
 Randomly break S into NP groups
    and add each group to
```
PG'
```
 If
```
 All groups in PG' either:
```
    Form a pre-existing compound, or
    Form a new compound that can be
        added without
```
n_c exceeding N_c
```
```
 And $PG' \neq RG$ And $[PG' \rightarrow RG] \notin R$
```
 Then
```
    $PG = PG'$
```
    Add newly-generated compounds in $PG'$ to C
    Update
```
n_c
```
    Return TRUE.
 Else Return FALSE.
End Procedure
```

Figure 1. Algorithmic overview of MD-CHEM.

To remedy this situation, GENCHEM attempts to extract known compounds from the list of product atoms prior to calling REACT-COMP, thus decreasing the number of new compounds introduced by each reaction. It uses the probability p to determine whether to try to extract one or more known compounds from S before randomly permuting and partitioning the remaining product atoms. In the calculation of p, k_e is the extraction factor, with a typical value between 1 and 2, with lower values implying a higher extraction probability. A high value of k_e is preferable when $N_c \approx N_r$, but it should decrease as $N_r - N_c$ rises.

Regardless of k_e settings, the initial size bias, B_{ics}, is a critical success factor. If the initial compound set contains too many large compounds, then generating N_r legal reactions becomes quite difficult, particularly when R_{cs} and R_{rs} are small (i.e., tight ranges of legal reaction and compound sizes). In short, if n large molecules are chosen as reactants from which m ($m \approx n$) random products of similar size are generated, then the odds of generating preexisting compounds as products diminish rapidly as the size distribution of those compounds becomes top-heavy. Hence, most products are new, and REACT-COMP can only succeed a few times before $n_c = N_c$, after which it will predominantly fail. With a bottom-heavy size distribution, REACT-COMP has a better chance of generating preexisting compounds, thus prolonging the completion of the compound set and maintaining the flexibility to generate new legal reactions.

2.1 Reaction Energies

One critical aspect of MD-CHEM is the association of energy production and consumption with reactions in a manner that supports the basic biological fact that biomass construction demands energy, whereas biomass breakdown generally releases energy. At the reaction level, this translates into a simple qualitative MD-CHEM energy principle: building larger compounds from smaller ones (i.e., anabolism) requires energy input, whereas the reverse process (i.e., catabolism) releases energy.

Statistical entropy is the basis of size comparisons between the reactant and product sets of a reaction. Entropy essentially measures the evenness and granularity of the size distribution of molecules: many small molecules have higher entropy than a few large molecules; and m molecules of similar size have higher entropy than m molecules of diverse sizes. The fractional sizes of each compound, f_i, relative to the total number of atoms in the reaction, determine the entropy according to

$$\sum_{i=1}^{k} -f_i \log f_i \tag{2}$$

Thus, in reaction 1, which involves a total of nine atoms, the fractional sizes are $\{1/3, 1/3, 1/3\}$ for the reactants, and $\{1/3, 2/9, 4/9\}$ for the products, yielding a reactant entropy of 1.585 and a product entropy of 1.530. Since lower entropy reflects higher order, the reaction exhibits a small degree of structure formation and is considered endothermic (i.e., energy consuming). MD-CHEM maps entropy differences directly into reaction energies, so this reaction requires $1.585 - 1.530 = 0.055$ energy units in order to run. The reverse reaction is modeled as exothermic, yielding 0.055 energy units.

When deployed in an artificial-life simulator, MD-CHEM uses the law of mass action [17] to compute the rates of reactions. To wit, the product of the concentrations of the reactants along with a basal reaction constant determine the rate at which reactants are converted into products and energy is produced or consumed.

2.2   Catalysts

Catalytic relationships between compounds and reactions arise randomly during the reaction-generation process. Any compound that appears as both a reactant and product in the same reaction is considered a catalyst (i.e., enzyme). When running a reaction, MD-CHEM assumes that unless the catalyst is completely absent (concentration of 0.0), the reaction is substrate-limited and can proceed as if each catalyst molecule could instantaneously derive product from reactant molecules. Hence, the law of mass action is only applied to the concentrations of the nonenzymatic compounds in deriving the reaction rate. Of course, if the enzyme is completely absent, then the reaction cannot occur.

To model reaction catalysis, MD-CHEM uses the enzyme's molecular weight (i.e., number of atoms) as a rough estimate of catalytic enhancement, e_c. For exothermic reactions, e_c appears as an extra product in the mass-action derivation of the reaction rate, while in endothermic reactions, the energy consumption of the uphill reaction is reduced by a factor of $1/e_c$.

$$a_2b + ab_2c + c_3 \iff a_2b + ab_2c_4 \tag{3}$$

For example, if $e_c(c_i) = 1 + \text{length}(c_i)$ for any catalytic compound c_i, then the catalytic enhancement of reaction 3 is $1 + \text{length}(a_2b) = 4$. Since the left-to-right reaction is uphill/endothermic, with an entropy change from 1.571 to 0.881, or 0.69, the catalytic effect essentially reduces the energy requirement of the uphill reaction to $0.69/4 = 0.1725$. Similarly, the rate of the right-to-left downhill reaction will be enhanced by a factor of 4 in any context in which the reaction runs.

2.2.1   Using MD-CHEM

MD-CHEM is designed for investigating the abstract relationships between chemistries and the metabolisms that they support, and hence the biochemical guilds that evolve. A typical scenario is to generate hundreds of different chemistries and test each one in an artificial life simulation of metabolizing organisms. Those chemistries on which microorganism communities can survive are separated from the less supportive variants to get a general understanding of the primary life-sustaining aspects. These life-supporting chemistries can then form the backdrop for explorations into more complex phenomena such as emergent environmental control, recycling, and so forth.

Using MD-CHEM in this manner affords (a) generalizations across a wide array of phenomena-supporting chemistries to understand the essential chemical foundation of those phenomena at a level above that of the specific atoms, molecules and reactions, (b) estimates of the actual likelihood of a phenomenon under varying chemical conditions. In short, MD-CHEM permits views beyond *biochemistry as we know it* for exploration of the more general relationships between chemistry and life.

3   METAMIC: Metabolically Abstract Microorganism System

METAMIC is a simple box model with chemical inflows and outflows. The box constitutes an environment, E, for a population of individually modeled metabolizing agents (a.k.a. *metamics*) who only interact in two ways: (a) indirectly via chemical exchange with E, and (b) directly via gene swapping during conjugation. Each metamic is modeled as a cell with a genetically determined metabolism, a local chemical buffer, and a semi-permeable membrane that separates it from the environment.

METAMIC employs MD-CHEM chemistries as bases for all intra- and extracellular chemical activity. The chemical basis for a METAMIC run is defined by the pair (C, R), where C denotes the set of legal compounds, and R the legal reactions.
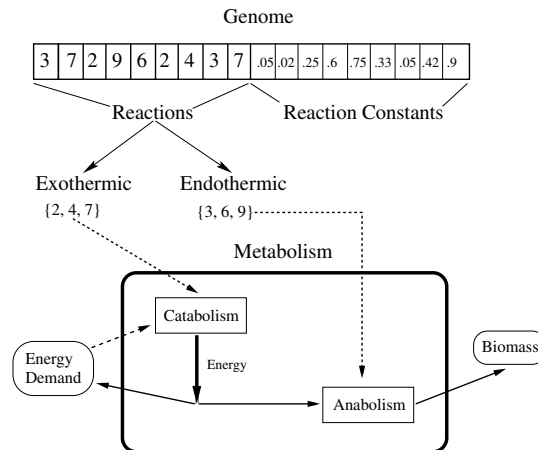
Figure 2. A METAMIC organism's genome determines the reactions and reaction constants for its metabolism. Exothermic reactions make up the catabolism, and endothermic reactions the anabolism. The energy produced by catabolism pays back the energy demand, and any remaining energy goes toward structure building (i.e., anabolism).

3.1 Physical Variables

A parameterizable set of physical variables, Φ, provides a second pathway for bidirectional interactions between the organisms and their environment. The value of each $p_i \in \Phi$ is defined as a function of the normalized concentrations of some or all of the molecules in C. Values for all p_i are computed for the global environment, based on the global concentrations; also, local values are computed for each cell, based solely on intracellular concentrations. As described below, the physical variables can influence metabolic rates, and hence growth.

3.2 Genetically Determined Metabolisms

Each metamic's genetic algorithm (GA) genome encodes r_T, a subset of R, plus base reaction constants, k_r, for each inherited reaction. All values are encoded as bit strings, where reactions are integer indices into R and rates are real numbers within a parameterized range.

From the genome, the exothermic, r_x, and endothermic, r_n, reactions are separated, where $r_T = r_x \cup r_n$; together, they compose the organism's metabolism. As shown in Figure 2, the abstract metabolic process consists of two phases: catabolism and anabolism. On each time step, metamics receive an energy request and begin catabolism, wherein the exothermic reactions, r_x, run for the maximum of two durations: (a) the official time step and (b) the estimated time needed to generate the required energy (based on previous energy-production rates of the r_x). If the former exceeds the latter, then an energy surplus results, thus triggering the anabolic processes (i.e., the endothermic reactions, r_n), which run long enough to consume the available energy and build structure by reducing internal entropy.

3.2.1 Mass Action Dynamics

Within a cell, the rate of each chemical reaction, $r \in r_T$ is the product of (a) the intracellular concentrations of the nonenzymatic reactants, I_{ne}, (b) the genetically determined reaction constant, k_r, (c) the catalytic enhancement factor (for exothermic reactions),

e_c, and (d) the physical-factor satisfaction, f_{sat}

$$k_r e_c f_{sat} \prod_{i \in I_{ne}} [i] \tag{4}$$

Reaction rates in the global environment are computed similarly, but with f_{sat} fixed at 1.0 and k_r fixed, but parameterizable (i.e., user specified).

Physical-factor satisfaction is an exponentially decaying function of the deviation of each physical factor $p_i \in \Phi$ from its optimal value (for metabolism), $p*_i$.

$$e_{sat} = \sum_{p_i \in \Phi} \left| \frac{p_i - p*_i}{p*_i} \right| \tag{5}$$

$$f_{sat} = e^{-k_{sat} e_{sat}} \tag{6}$$

The parameter k_{sat} reflects the degree to which deviations from optimal physical conditions will affect metabolism. A typical value is 1.0. The $p*_i$ and the functions for computing the p_i from chemical concentrations are user supplied and may differ considerably across experiments.

3.3 Biomass

Those compounds that constitute biomass can vary between organisms, with key restrictions. Two user-defined parameters, N_{bio} and N_{bio}^*, specify the maximum number of compounds that an organism can treat as biomass, and the number of large compounds in C that are legal biomass constituents, respectively. To choose the biomass compounds, c_{bio}, for an agent, METAMIC gathers the N_{bio} largest compounds that the organism is a net producer of (in r_T). It then intersects that set with the N_{bio}^* largest compounds in C to form c_{bio}. An organism's biomass is then the total intracellular mass of all c_{bio} compounds.

3.4 The Cell Membrane

The cell membrane (Figure 3) is semi-permeable in that diffusion rates depend upon the molecule. All compounds in c_{bio} cannot diffuse into or out of that metamic's cell: All biomass molecules, aside from those present at birth, result from internal production. Also, any compounds that the organism is a net consumer of (in r_T) can diffuse into but not out of the cell. Conversely, compounds with net production in r_T can only diffuse out, not in. All other compounds can freely diffuse into and out of the cell. Net production and consumption are estimated from r_T and the inherited reaction constants, η_k, for determining a metamic's fixed diffusion constraints. Since active transport is not modeled, all flows between cells and E are along a decreasing concentration gradient for that particular compound.

The cell membrane provides a buffer zone for the cell. This zone is not perfectly protective, since diffusive chemical exchanges with the environment will move local and global chemical concentrations, and hence p_i values, closer, but the restrictions on diffusion do allow a cell to create an interior milieu that differs from its surroundings. This buffering is critical for cell growth, since the p_i can affect metabolic rates.

3.5 External Chemical Fluxes

For each compound $c_i \in C$, the user specifies separate exogenous inflow and outflow rates for the environment, E. The inflows are simply units of chemical per time step, and the outflows are specified as fractions of the current amount of c_i in E, for each i. All reactions in R are assumed to occur in E, and since no energy is demanded of E,
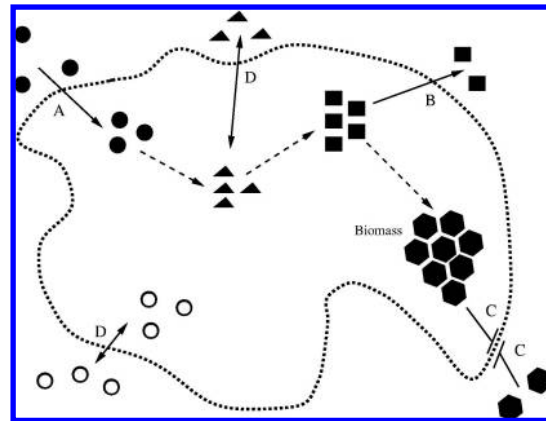
Figure 3. The semi-permeable cell membrane permits (A) only inflow of (net) consumed compounds, (B) only outflow of (net) produced compounds, (C) no diffusion of biomass, and (D) bidirectional diffusion of compounds with no significant net production nor consumption. Each icon denotes a compound, and arrows between different compounds represent chemical reactions.

the energetic fruits of r_x go directly into r_n. The user can also specify a constant energy input to the system, which is distributed among all agents.

3.6 Fitness, Reproduction, and Death

In METAMIC's GA, fitness is implicit: If an organism doubles its birth biomass, then reproduction occurs by asexual splitting, with possible mutation of both child genomes. Organisms also undergo a form of double bacterial conjugation by occasionally swapping GA chromosome segments with one another. This is essentially standard GA crossover followed by mutation, with each individual continuing its life, but with a new genome and metabolism. Both genetic operators work at the bit level, without respecting gene borders.

The same mortality rate pertains to all living organisms, regardless of age or biomass. However, agents with biomass below a critical threshold are culled from the population. Upon death, the cell's membrane dissolves and all internal chemicals are added to the global environment.

3.7 Biotic-Biospheric Interactions

METAMIC is designed to allow bidirectional interactions between the organisms (biota) and the environment (biosphere). In nature, the biota's primary route to global physical change is via a multitude of local chemical exchanges, whereas the biosphere can directly influence the biota by either chemical or physical means. In METAMIC, the only direct causal routes are chemical, since the physical variables are computed independently in the cells and in the global environment.

As summarized in Figure 4, METAMIC's primitive mechanisms support a feedback loop involving metabolism, chemical concentrations, and the physical factors. Mass-action dynamics govern the relationship between metabolic rates (and growth) and the intracellular chemical concentrations, while user-specified constraints relate concentrations to physical conditions. Deviations of physical variables from their growth optima then influence metabolic rates to close the loop.

In the environment, the physical factors have no influence upon reaction rates, since this link is intended to capture the complex relationship between physical conditions and metabolism, not chemical reactions at large.
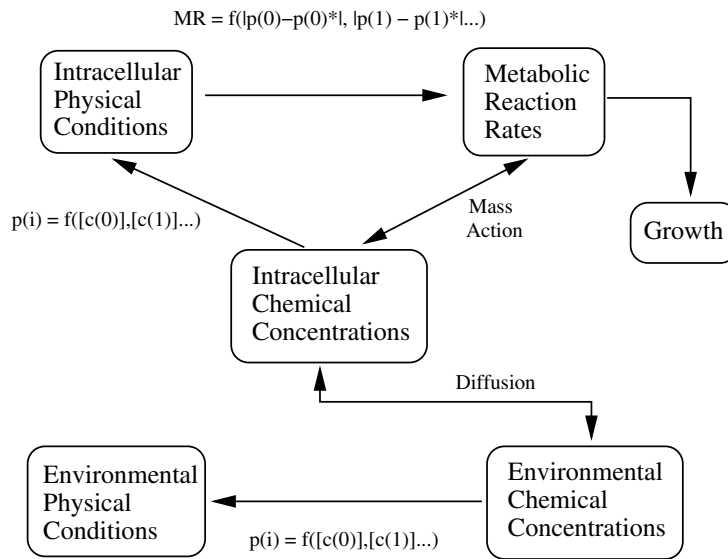
$$MR = f(|p(0){-}p(0)*|, |p(1) - p(1)*|...)$$

Figure 4. The primary interactions in METAMIC that causally interlock the biota and biosphere. Arrows indicate influence of one factor upon another, with arrow labels indicating the underlying mechanism. The $p(i)$ are physical variables, the $p(i)^*$ are optimal values for those variables, the $[c(i)]$ are chemical concentrations, and MR is the metabolic rate.
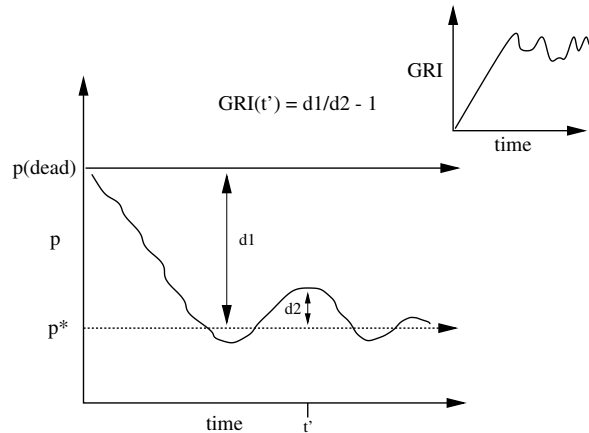

Figure 5. Illustration of the Gaian regulatory index (GRI). The large graph shows the trajectory of a physical variable p, along with the essential terms needed to compute the GRI at any time point t'. The smaller graph plots the time series of GRI values corresponding to that same trajectory of p.

3.8 GRI: The Gaian Regulatory Index

The Gaian regulatory index (GRI) is a quantitative measure of the METAMIC biota's ability to control the environment within a range that is most conducive to growth. In a nutshell, the GRI is the gain of the distributed regulator having p_{dead} as a base value and p^* as the target value, where (a) p_{dead} is the steady-state value of a focal physical variable, p, when the global environment is simulated using chemical inflows, outflows and internal chemical reactions, but without organisms, and (b) p^* is the (arbitrarily chosen) metabolic optimum value for p.

As depicted in Figure 5, the GRI at any time point t' is -1 plus the ratio of the distance from p^* to p_{dead} (i.e., the total distance that a perfect regulator would move p)

to the actual distance from $p(t')$ to p^* (i.e., the error). Higher GRI values indicate better regulation. To avoid infinity calculations in zero-error ($d2 = 0$) situations, METAMIC enforces a GRI upper bound of 20.

3.9 The Cycling Ratio

In E, the ratio of recycled atoms to the maximum of the input and output fluxes constitutes that element's cycling ratio [23]. In essence, the cycling ratio denotes the average number of times that an atom is passed from organism to organism before it is flushed out of the environment. This provides an alternate metric of Gaia, since it is a clear example of life enhancing life: An ecosystem that recycles materials can support a much larger total biomass than one that is totally dependent upon external nutrient fluxes for growth. Like GUILD, METAMIC tallies cycling ratios for all atomic elements.

3.10 Simulation Procedure

On each time step, a metamic (a) performs its metabolic activities, with all chemical reactions driven by intracellular chemical concentrations, and (b) exchanges chemicals via diffusion with the global compartment. In addition, it may asexually reproduce or die. A generation is a fixed number of time steps (e.g., 20 in all runs below), and at the end of each generation, a fixed percentage (often 30%) of organisms are chosen for conjugative gene swapping.

In E, all exogenous fluxes and chemical reactions are performed on each time step. The environmental reaction constants are typically an order of magnitude lower than their intracellular counterparts.

4 Hunting for Gaia

This reseach is motivated by a central research question: *How likely is the emergence of distributed environmental regulation in an evolving community of metabolizing organisms?* The approach is simply to generate a host of artificial chemistries, use each as the basis for a few METAMIC runs, and observe the frequency of emergent control. Since distributed environmental control is a trademark of Gaia, this procedure is referred to as *Gaia hunting*.

MD-CHEM chemistry was chosen over real biochemistry for two reasons: (a) The complexities of the latter are intimidating, particularly at the level of microbial metabolism, and (b) The former enables testing over a wide range of (albeit simple) chemistries, not just one.

The basis in MD-CHEM forces a specialization of the central research question to our basic experimental question: In chemical environments that incorporate the fundamental thermodynamic relationship between energy consumption (production) and the synthesis (breakdown) of large molecules, what is the likelihood of emergent distributed environmental regulation among a community of metabolizing microorganisms? In short, if we make this very basic thermodynamic assumption about chemistry, and little else, is emergent control expected?

As summarized in Figure 6, the Gaia-hunt loop begins with the generation of an MD-CHEM artificial chemistry, C^*. A short METAMIC simulation based on C^* then runs to determine whether C^* supports life. In these short runs, the initial population is typically 20, whereas the maximum allowable size is 100, and $k_{sat} = 0$, indicating an insensitivity of metabolisms to the physical parameter, p. After 50 generations, if the population size exceeds 20, then C^* is tested again in a 1,000-generation run, denoted $G_{base}(C^*)$. If C^* enables the population to grow to maximum size in $G_{base}(C^*)$, then C^* is deemed *life-supporting* and the homeostasis tests can begin.

```
Procedure Gaia-Hunt(Max-Chems, GRI_thresh)
   Num-Good-Chems = 0 ;; Number of life-supporting chemistries
   Num-Gaia-Chems = 0 ;; Number of life-supporting chemistries that
      also support Gaia
   While Num-Good-Chems < Max-Chems do
      Use MD-CHEM to create a chemistry, C*
      Perform a short (e.g., 50 generation) METAMIC run, based on C*
      If population growth occurs, then
         Seed = get-random-number-generator-seed
         Perform a long (e.g., 1,000 generation) METAMIC run, based
            on C*, using metabolizing organisms, and with k_sat = 0.
         If this long run supports life, then:
            Num-Good-Chems = Num-Good-Chems + 1
            Perform a short (e.g., 1,000 time step) METAMIC run,
               without organisms, to establish the baseline value,
               p_dead of the focal physical variable, p.
            p_avg = the time-average value of p from generation T_max
               to 1,000
            d = p_avg − p_dead
            p*_in = p_avg − d/2
            p*_out = p_avg + d/2
               If  Gaia-Test(C*,p_dead,p*_in,0.75,GRI_thresh,seed) or
                   Gaia-Test(C*,p_dead,p*_out,0.75,GRI_thresh,seed) then
                     Num-Gaia-Chems = Num-Gaia-Chems + 1
   End While
   Return Num-Gaia-Chems/Max-Chems
End Procedure

Procedure Gaia-Test(C*,p_dead,p*,k,GRI*,seed)
   set-random-number-generator-seed(seed)
   Perform a long (e.g., 1,000 generation) METAMIC run, based on
      C*, using metabolizing organisms, with k_sat = k, and with the
      metabolic-optimum value of the physical variable p set to p*.
   GRI_bar = the time-averaged Gaian regulatory index (based on p*
      and p_dead) for the entire simulation.
   If  GRI_bar ≥ GRI* then
      return TRUE
      else return FALSE
End Procedure
```

Figure 6. Overview of the Gaia-hunting process.

To establish base and target values of p for the regulatory tests, the value of p_{avg} is computed as the average value of p in the previous 1,000-generation run between generations T_{max} and 1,000, where T_{max} is the first generation where the population hit the maximum size. Then, p_{dead} is the steady-state p value from the aforementioned lifeless simulation. Given p_{avg} and p_{dead}, d is computed as the distance between them, and the two target values, p^*_{in} and p^*_{out}, are found by moving $d/2$ units from p_{avg} toward and away from p_{dead}, respectively (Figure 7).

These target values form the basis for two Gaian regulatory tests, $G_{in}(C^*)$ and $G_{out}(C^*)$, in which the metabolic optimum p value is p^*_{in} and p^*_{out}, respectively. These
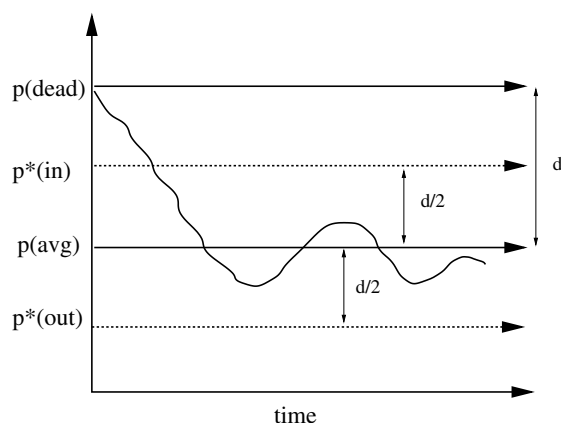
Figure 7. Establishing optimal physical conditions for Gaia tests: (1) Run a lifeless simulation to find the steady state value, p_{dead}, of the physical variable p. (2) Repeat the simulation with organisms and record the trajectory of p (curved line). (3) Take average of the trajectory, p_{avg} from point where maximum population size first occurs. (4) Let $d = p_{avg} - p_{dead}$. (5) Let $p^*_{in} = p_{avg} - d/2$ be the optimal p value for one G_{in} run. (6) Let $p^*_{out} = p_{avg} + d/2$ be the optimal p value for another G_{out} run.

three runs, $G_{base}(C^*)$, $G_{in}(C^*)$, and $G_{out}(C^*)$ are referred to as the $G(C^*)$ runs. For a given C^*, each $G(C^*)$ run begins with the same random seed. The key differences are in the values of p^* and k_{sat}, where k_{sat} is 0.0 in $G_{base}(C^*)$ and typically 0.75 in $G_{in}(C^*)$ and $G_{out}(C^*)$, reflecting a strong sensitivity of metabolic rates to the intracellular value of p. Using the two p^* values as targets, the GRI values are then computed for each generation of the regulatory tests. If $\overline{GRI}$, the average GRI value from T_{max} to 1,000, in either run exceeds GRI_{thresh}, then the simulation results are stored as an example of emergent distributed control. For the runs presented in this article, $GRI_{thresh} = 5$.

Table 1 summarizes the main parameters used in the Gaia hunt. In addition, the value of p in all $G(C^*)$ runs and the lifeless case is computed as

$$p = \tilde{c}_1 - \tilde{c}_2 + \tilde{c}_3 - \tilde{c}_4 \tag{7}$$

where $\tilde{c}_i$ $i = 1, \ldots, 4$ are the normalized (over all 10 compounds) concentrations of the four smallest compounds in C^*. This is an arbitrary choice for the functional relationship between the chemical and physical factors. Many others are clearly possible.

Figure 8 shows the results of 20 different METAMIC runs, each based on a different MD-CHEM. A total of 100 life-supporting chemistries were tested, with 20 giving $\overline{GRI}$ values above 5, indicating a reasonable degree of regulation. Of these 20 cases, 13 were G_{out} and 7 were G_{in}, with neither type showing particularly better results than the other. The average $\overline{GRI}$ of these 20 cases was 7.63, with a standard deviation of 2.71. The very best regulator had $\overline{GRI} = 14.97$, and 7 of the 20 had low values between 5 and 6.

In several scenarios, the regulatory run clearly distinguishes itself from the G_{base} case and maintains a stable value near the optimum, p^*. In other situations, the base and regulatory cases have more intertwined trajectories. This overlap does not necessarily decrease $\overline{GRI}$, if both curves are in the neighborhood of p^*. However, it weakens the homeostatic claim when the base case also oscillates near p^*. This was the whole basis for choosing p^* at a distance from p_{avg}—but not too close to p_{dead}.

In general, these results indicate that emergent distributed control is hardly a rare occurrence. Of course, it is certainly not a necessary consequence of metabolizing organisms in relatively closed environments. However, the basic phenomena arises in 20% of the tested chemistries, with 7–10% of them giving rather convincing evidence

Table 1. Key parameter values for the Gaia hunt simulations.

Parameter	Value
MD-CHEM	
Atoms	C, H, N, O
Compounds (N_c)	10
Reactions (N_r)	20
Initial compounds (N_{ic})	4
Compound size range (R_{cs})	[3, 12]
Reaction size range (R_{rs})	[1, 3]
Initial compound size bias (B_{ics})	[0.33 0.33 0.33 0 0 0 0 0 0 0 0 0]
Reaction size bias (B_{rs})	[0.14 0.72 0.14]
METAMIC genetic algorithm	
Initial population	20
Maximum population	100
Generations	1,000
Time steps per generation	20
Chromosome	10 reactions + 10 reaction constants
Mutation rate	0.04 per gene
Conjugation fraction	0.3
k_{sat}	0.0 or 0.75
METAMIC global environment	
Input fluxes	10 mass units per time step
Input compounds	The 4 smallest
Output fractional fluxes	0.001 of all 10 compounds per time step

for a deviation from normal behavior to an unconscious environment-controlling mode. Remember that in the base case, the metabolic rates are independent of the physical variable, p, so there is no selective advantage for local regulation of intracellular p values. In many G_{in} and G_{out} runs, the chemistry simply does not provide the flexibility: Organisms cannot alter their metabolisms to attain near-optimal p values while simultaneously producing enough biomass to eventually reproduce. Hence, they must stick to the standard metabolisms but operate under suboptimal p regimes.

In our GUILD simulations [9], the environmental control is robust to changes in external chemical fluxes. This stems from the lack of constraints on chemical transformations in that model. However, the reduced reaction possibilities in the above METAMIC runs strongly restrict the metabolic options. Hence, regulation often breaks down with a change in exogenous fluxes.

For example, if compound X affects the physical variable, p, and if X flows into the environment at a higher rate than desired with respect to p^*, then the chemistry may permit a metabolic pathway that both builds biomass and consumes X, thus moving the environment closer to p^*. However, if the external influx of X suddenly declines to a suboptimal value (with respect to p^*), then the chemistry may not also sanction a viable metabolism that produces X. Hence, environmental control would only occur with normal or elevated inflows of X.

4.1   Snapshots from the Hunt

A case study of a few convincing examples of emergent regulation from Gaia hunts reveals some mechanisms by which organisms manage the trade-off between producing biomass and controlling the environment.
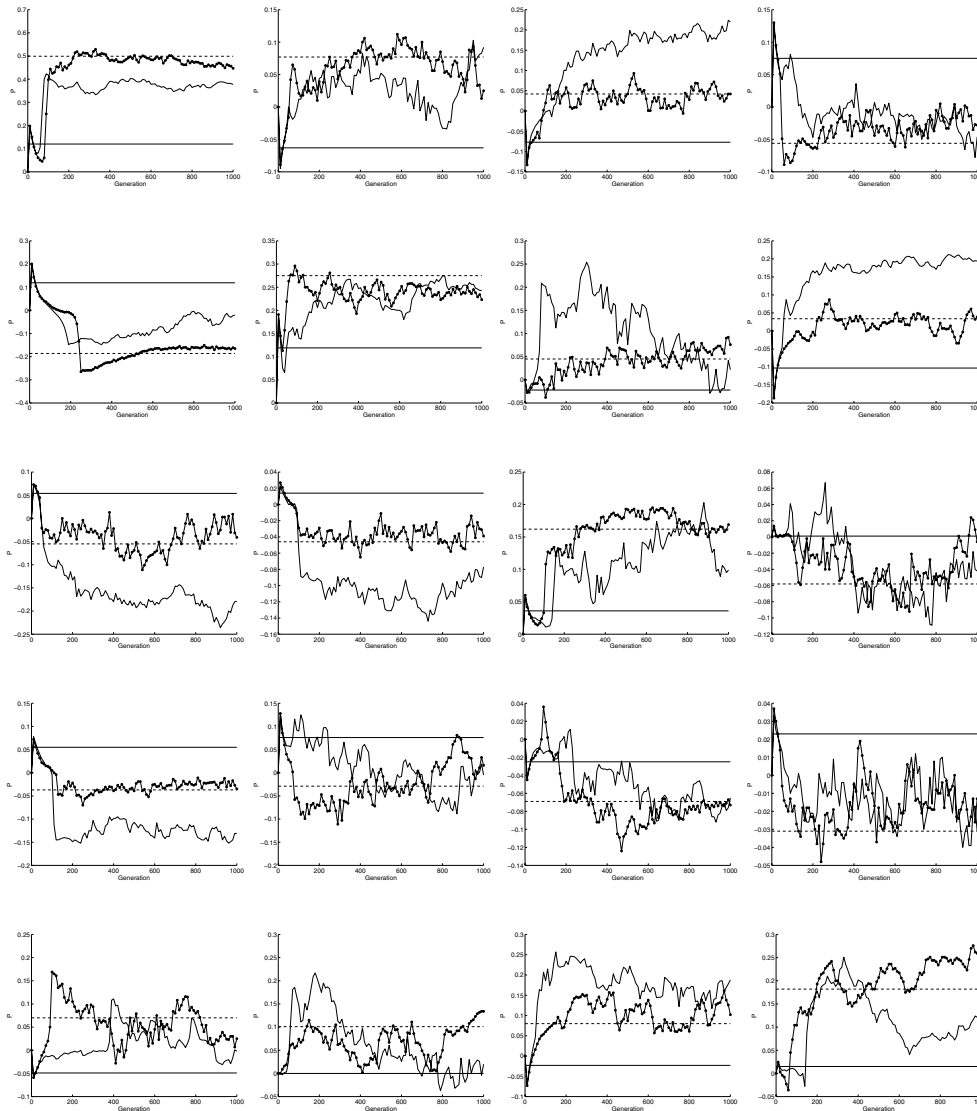
Figure 8. Emergent regulation ($\overline{\text{GRI}} > 5.0$) in 20 different METAMIC runs based on 20 different 10-compound, 20-reaction MD-CHEM chemistries. A total of 100 life-supporting chemistries were tested, and 20% gave $\overline{\text{GRI}}$ values above 5. In each graph, the straight solid line denotes p_{dead}, and the straight dashed line represents p^*. The solid curve is the trajectory of p in the G_{base} run, and the curve with asterisks shows p's trajectory in the G_{in} or G_{out} case.

4.1.1 Scenario GH1

The random chemistry of Figure 9 provides the backdrop for a METAMIC test in which the G_{out} run exhibits strong environmental regulation: $\overline{\text{GRI}} = 12.05$. Figure 10 compares p trajectories for the G_{base} and G_{out} runs; it also illustrates the relatively high $\overline{\text{GRI}}$ value after about generation 500. In these runs, the G_{base} population hits the maximum size of 100 after 184 generations, whereas the G_{out} population needed 242 generations.

A detailed comparison of the G_{base} and G_{out} runs reveals the metabolic *sacrifices* that the G_{out} organisms make to control their internal environments near p^*. First, a tally of the biomass combinations used by the 1,000th-generation organisms for both runs appears in Table 2.

Compounds:

H_2 NO C_3N C_2O_2 C_5O C_8 C_6N_2 C_4N_4 C_9N_3 C_5N_7

Reaction	Energy	Frequency
0: $C_3N + C_3N \Rightarrow C_6N_2$	1.00	(1 : 6)
1: $C_6N_2 \Rightarrow C_3N + C_3N$	−1.00	(98 : 8)
2: $C_3N + C_6N_2 \Rightarrow C_9N_3$	0.92	(1 : 2)
3: $C_9N_3 \Rightarrow C_3N + C_6N_2$	−0.92	(98 : 100)
4: $C_3N + C_9N_3 \Rightarrow C_8 + C_4N_4$	−0.19	(4 : 59)
5: $C_8 + C_4N_4 \Rightarrow C_3N + C_9N_3$	0.19	(97 : 0)
6: $C_4N_4 + C_4N_4 \Rightarrow C_3N + C_5N_7$	0.19	(1 : 6)
7: $C_3N + C_5N_7 \Rightarrow C_4N_4 + C_4N_4$	−0.19	(0 : 95)
8: $C_3N + C_2O_2 \Rightarrow NO + C_5O$	0.19	(0 : 1)
9: $NO + C_5O \Rightarrow C_3N + C_2O_2$	−0.19	(0 : 3)
10: $C_8 + C_6N_2 + C_4N_4 \Rightarrow C_9N_3 + C_9N_3$	0.58	(2 : 100)
11: $C_9N_3 + C_9N_3 \Rightarrow C_8 + C_6N_2 + C_4N_4$	−0.58	(52 : 0)
12: $C_5O + C_4N_4 \Rightarrow NO + C_9N_3$	0.39	(37 : 99)
13: $NO + C_9N_3 \Rightarrow C_5O + C_4N_4$	−0.39	(2 : 9)
14: $NO + C_3N + C_9N_3 \Rightarrow NO + C_6N_2 + C_6N_2$	−0.17	(61 : 80)
15: $NO + C_6N_2 + C_6N_2 \Rightarrow NO + C_3N + C_9N_3$	0.17	(0 : 98)
16: $C_2O_2 + C_6N_2 \Rightarrow NO + NO + C_8$	−0.33	(42 : 1)
17: $NO + NO + C_8 \Rightarrow C_2O_2 + C_6N_2$	0.33	(98 : 79)
18: $C_3N + C_6N_2 \Rightarrow C_3N + C_3N + C_3N$	−0.67	(0 : 3)
19: $C_3N + C_3N + C_3N \Rightarrow C_3N + C_6N_2$	0.67	(98 : 100)

Figure 9. Compounds and reactions for 10×20 Scenario GH1. The atoms *C*, *H*, *N*, and *O* are used only to enhance readability and recall of compound names; in no way do the atoms or compounds directly model their real-world counterparts. Energy values (prior to catalytic enhancement) appear in the second column, with positive (negative) values for endothermic (exothermic) reactions. The third column lists pairs of occurrence frequencies for the G_{base} and G_{out} cases, respectively. Each value denotes the number of organisms using that reaction in their metabolism after the final (1,000th) generation. The population size had been at or about a maximum level of 100 since very early in both simulations.

The column totals in Table 2 indicate that whereas C_6N_2 is the favorite biomass in G_{base}, the organisms of G_{out} prefer C_9N_3. The difference, one molecule of C_3N, is critical for environmental control.

In GH1, the physical variable, p, is computed as

$$p = [\tilde{H_2}] - [\tilde{NO}] + [\tilde{C_3N}] - [\tilde{C_2O_2}] \tag{8}$$

Figure 10 indicates that the base case yields a higher value of p than desired in G_{out}. Equation 8 hints at possible routes to improvement via decreasing the positive terms or increasing the negative terms. Although H_2 is inert in this chemistry (i.e., it is not involved in any reactions), the other three compounds in Equation 8 are theoretically amenable to metabolic control.

A comparison of the environmental chemical concentration profiles for G_{base} and G_{out} (Figure 11) shows that G_{out} organisms maintain a much lower value of C_3N than in G_{base}. However, this choice carries a high metabolic cost.

Although the populations in both runs exhibit a wide variety of metabolisms, a few very popular chemical reactions appear in a large majority of the organisms in each
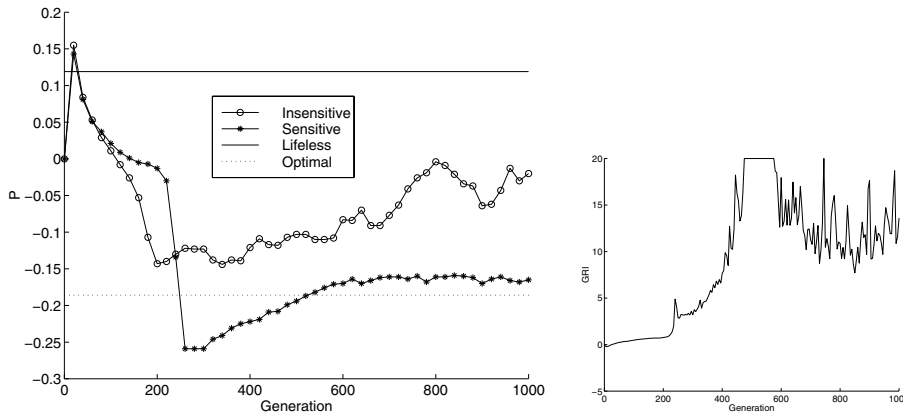
Figure 10. Homeostatic behavior in Scenario GH1. Leftmost graph displays physical variable, p, when the metabolic rates are sensitive (G_{out}) and insensitive (G_{base}) to p. Rightmost graph portrays the Gaian regulatory index (GRI) for the G_{out} run.

Table 2. Biomasses used by the 100 organisms at the end of the 1,000th generation of Scenario GH1's G_{base} and G_{out} runs. In each row, the column entries for the compounds used in the biomass combination type contain the number of organisms using that combination.

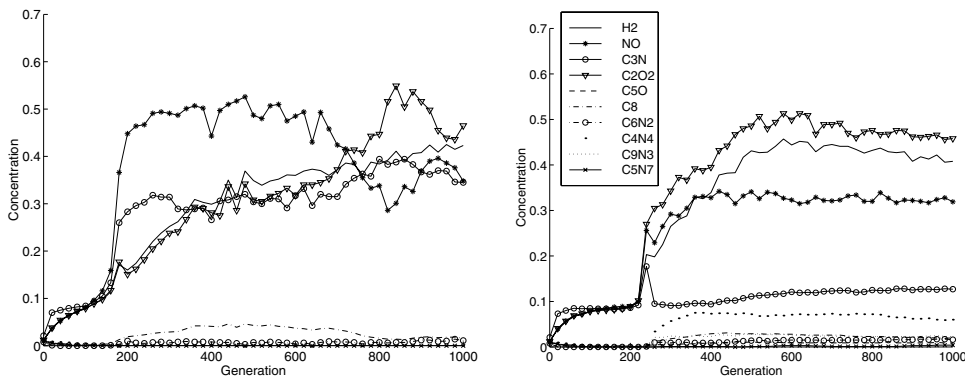| Biomass Combination Type	C_6N_2	C_4N_4	C_5N_7	C_9N_3
G_{base}				
1	67			67
2	30	30		
3	1		1	
Total	98	30	1	67
G_{out}				
1	54			54
4		41		41
5			5	5
Total	54	41	5	100

Figure 11. Environmental chemical concentrations for the G_{base} (left) and G_{out} (right) runs of Scenario GH1.

Table 3. Anabolic differences between the evolved metabolisms of the G_{base} and G_{out} runs for GH1. The exact anabolic combinations above are the most popular, appearing in 57% (67%) of the G_{base} (G_{out}) individuals at the end of generation 1,000. Energy requirements take catalytic effects into account by reducing the energy demand by the standard catalytic factor of 1 + catalyst size. Net biomass gain (rightmost column) involves five terms for the five biomass combination types from Table 2. The anabolic efficiency values for the primary biomass types of each run are in bold.

Reaction	Energy used	Net biomass gain
Common		
17: $NO + NO + C_8 \Rightarrow C_2O_2 + C_6N_2$	0.33	1,1,1,0,0
19: $C_3N + C_3N + C_3N \Rightarrow C_3N + C_6N_2$	0.133	1,1,1,0,0
G_{base} only		
5: $C_8 + C_4N_4 \Rightarrow C_3N + C_9N_3$	0.189	1, −1, 0, 0, 1
G_{base} total	0.655	3,1,2,0,1
G_{base} anabolic efficiency		**0.22**, **0.66**, 0.33,—, 0.655
G_{out} only		
10: $C_8 + C_6N_2 + C_4N_4 \Rightarrow C_9N_3 + C_9N_3$	0.585	1, −2, −1, 1, 2
12: $C_5O + C_4N_4 \Rightarrow NO + C_9N_3$	0.394	1, −1, 0, 0, 1
15: $NO + C_6N_2 + C_6N_2 \Rightarrow NO + C_3N + C_9N_3$	0.056	−1, −2, −2, 1, 1
G_{out} total	1.501	3, −3, −1, 2, 4
G_{out} anabolic efficiency		**0.50**, — , — , **0.75**, 0.38

Table 4. Catabolic differences between the evolved metabolisms of the G_{base} and G_{out} runs for GH1. The exact catabolic combinations above are the most popular, appearing in 32% (33%) of the G_{base} (G_{out}) individuals at the end of generation 1,000. Energy production ignores catalytic effects, since these alter reaction rates but not the energy yield per unit of reactant. Net biomass gain (rightmost column) involves five terms for the five biomass combination types from Table 2.

Reaction	Energy produced	Net biomass gain
Common		
3: $C_9N_3 \Rightarrow C_3N + C_6N_2$	0.92	0, 1, 1, −1, −1
14: $NO + C_3N + C_9N_3 \Rightarrow NO + C_6N_2 + C_6N_2$	0.168	1, 2, 2, −1, −1
G_{base} only		
1: $C_6N_2 \Rightarrow C_3N + C_3N$	1.0	−1, −1, −1, 0, 0
11: $C_9N_3 + C_9N_3 \Rightarrow C_8 + C_6N_2 + C_4N_4$	0.585	−1, 2, 1, −1, −2
G_{base} total	2.673	−1, 4, 3, −3, −4
G_{out} only		
4: $C_3N + C_9N_3 \Rightarrow C_8 + C_4N_4$	0.189	−1, 1, 0, 0, −1
7: $C_3N + C_5N_7 \Rightarrow C_4N_4 + C_4N_4$	0.189	0, 2, 1, 2, −1
G_{out} total	1.466	0, 6, 4, 0, −4

run. The final column of Figure 9 shows the frequencies of the different reactions at the end of both runs. Similarly, many different anabolisms and catabolisms are used, but a few variants occur repeatedly, providing a general picture of the average individual.

In Tables 3 and 4, the anabolic and catabolic reactions are listed for G_{base}(GH1) and G_{out}(GH1), with reactions common to both runs appearing first. The rightmost column, net biomass gain, considers the five basic biomass combinations, as enumerated in Table 2. From the perspective of each biomass type, the net production of both compound molecules that make up the type is calculated for each reaction and summed for the entire reaction set. For example, relative to biomass type 1, {C_9N_3, C_6N_2}, reaction 15 has a net biomass gain of −1, since two C_6N_2 molecules are consumed as reactants to produce one C_9N_3 molecule. For the endothermic reactions, the

energy input divided by the biomass gain gives a rough estimate of anabolic efficiency, with lower values indicating greater performance. These ratios are calculated relative to each biomass type, and the ratios for the most popular biomass types are indicated by bold in Table 3.

As Table 3 shows, the average G_{base} metamic has an efficiency of 0.22 for its prime biomass type, 1, which constitutes 67% of the population, and 0.66 for its secondary biomass type, 2, with a 30% frequency. Conversely, the average G_{out} metamic has efficiencies of 0.5 and 0.75 for its main biomass types, 1 and 4, which make up 54% and 41% of the population, respectively. Including the efficiencies for the other biomass types, as listed in Table 2, the average efficiencies for the two metamic populations are:

Efficiency(G_{base}(GH1)) = (0.22)(0.67) + (0.66)(0.30) + (0.33)(0.01) = 0.35

Efficiency(G_{out}(GH1)) = (0.50)(0.54) + (0.75)(0.41) + (0.38)(0.05) = 0.60

Hence, the G_{out} metamic works nearly twice as hard to produce biomass. Much of this extra effort goes to creating the larger C_9N_3 molecule, which has the side effect of tying up more C_3N in biomass, thus contributing to control of the physical variable, p. In addition, Table 4 reveals the consumption of C_3N in the two catabolic reactions that are commonly, but uniquely, found in G_{out} metamics: 4 and 7. Note that the energy yields of these two reactions are much less than those of the popular exothermic reactions in G_{base}: 1 and 11. Table 4 also shows that for its two main biomass types, 1 and 2, the G_{base} organism has minimal biomass loss (-1) and a gain (4), respectively, during catabolism, whereas the G_{out} agent has no net loss or gain for its two main biomass types, 1 and 4.

Clearly, the G_{out} metamics are selected for their ability to control their intracellular p values in exchange for otherwise less effective metabolisms. Global environmental control near p^* then emerges from this local selective advantage.

Another sign of Gaia, recycling, has a more pronounced appearance in G_{out} than in G_{base}. The former displays stable cycling ratios of between 2 and 10 for the four atomic elements, whereas the latter gives only sporadic weak recycling at ratios between 0.25 and 2. When the maximum population size is increased to 500 (with no other increases in environmental fluxes) a 5,000-generation run of G_{out} yields cycling ratios between 20 and 60 and $\overline{GRI} > 10.0$. In nature, where population sizes are in the millions and billions, and where there have been equally many years in which to evolve efficient complementary metabolisms, it is not surprising that cycling ratios in the thousands often occur [23].

Since the environment in GH1 receives (a) the input fluxes of only the four lightest compounds, and (b) no external energy injections, long-term regulation near p^* is far from a certainty due to the excessive energetic demands of the control-oriented metabolisms. However, several alternate G_{out} (GH1) runs of between 5,000 and 10,000 generations yielded persistent control with $\overline{GRI} > 10.0$.

Interestingly enough, these long runs exhibited the emergence of several different dominant biomass types. In one case, type-1 biomass, {C_9N_3, C_6N_2}, was used by 100% of the population, yielding $\overline{GRI} = 10.78$. In another, type-5 biomass, {C_9N_3, C_5N_7}, attracted 95% of the metamics, resulting in $\overline{GRI} = 10.23$. In both cases, the anabolisms ressembled those of the original G_{out}(GH1) run, with a central core of reactions 10, 12, 15, and 19. However, the catabolisms varied considerably, sometimes having more in common with the G_{base}(GH1) run.

Population convergence to a common biomass seems expected, since individuals that share a biomass type, B, with a dead organism, X, could, given the proper metabolism, more quickly assimilate the decay products of X into their own B com-

Table 5. Most popular anabolism and catabolism for an alternate $G_{out}(GH1)$ scenario, $G_{out}(GH1)^*$, which ran for 7,000 generations with $\overline{GRI}$ = 19.06. The energy values in column 2 ignore catalytic effects, while energy consumption is reduced by the standard catalytic factor: 1 + catalyst size. In this run, type-2 biomass was the most prevalent. Consequently, the anabolism uses 1.466 energy units to produce three biomass units, while the catabolic sequence yields 2.863 energy units and actually *produces* five biomass units.

Reaction	Energy produced	Net biomass gain
Anabolism		
0: $C_3N + C_3N \Rightarrow C_6N_2$	-1.00	1,1,1,0,0
17: $NO + NO + C_8 \Rightarrow C_2O_2 + C_6N_2$	-0.33	1,1,1,0,0
19: $C_3N + C_3N + C_3N \Rightarrow C_3N + C_6N_2$	-0.133	1,1,1,0,0
Anabolic total	-1.466	3,3,3,0,0
Anabolic efficiency		0.49, **0.49**, 0.49, — , —
Catabolism		
1: $C_6N_2 \Rightarrow C_3N + C_3N$	1.0	$-1, -1, -1, 0, 0$
3: $C_9N_3 \Rightarrow C_3N + C_6N_2$	0.92	$0, 1, 1, -1, -1$
4: $C_3N + C_9N_3 \Rightarrow C_8 + C_4N_4$	0.189	$-1, 1, 0, 0, -1$
11: $C_9N_3 + C_9N_3 \Rightarrow C_8 + C_6N_2 + C_4N_4$	0.585	$-1, 2, 1, -1, -2$
14: $NO + C_3N + C_9N_3 \Rightarrow NO + C_6N_2 + C_6N_2$	0.168	$1, 2, 2, -1, -1$
Catabolic total	2.862	$-2, 5, 3, -3, -5$

pounds. Organisms with a different biomass type might have to wait slightly longer for the environment to decompose B compounds into reactants that they could utilize.

However, in the most interesting of the long $G_{out}(GH1)$ runs, biomass type 2, $\{C_6N_2, C_4N_4\}$, held a steady 85–95% for over 5,000 generations, with type 1 filling in the remainder. Although type 2 had no consistent presence in the other three $G_{out}(GH1)$ runs, it made up 30% of the $G_{base}(GH1)$ individuals. This case, $G_{out}(GH1)^*$, displayed an impressive $\overline{GRI}$ = 19.06. Note that the maximum allowable GRI value at any time step was artificially set to 20, not infinity, so the $\overline{GRI}$ values are actually underestimates in the very successful regulatory runs.

In Table 5, a detailed look at the dominant metabolism in $G_{out}(GH1)^*$ explains its superior performance. Compared to $G_{base}(GH1)$'s anabolism, $G_{out}(GH1)^*$ replaces reaction 5 with reaction 0. This entails the consumption of an extra three C_3N molecules, but at an increased energy cost of $1.00 - 0.189 = 0.811$. This yields an anabolic efficiency of 0.49, midway between the efficiencies of $G_{base}(GH1)$ and $G_{out}(GH1)$. Looking back at $G_{out}(GH1)$'s anabolism, it replaced reaction 5 with reactions 10, 12, and 15. Together, these three reactions have no net regulatory influence, since they have a net production of one NO and one C_3N molecule; however, the removal of reaction 5 does reduce C_3N production by one molecule. So on the anabolic side, the regulatory enhancement of $G_{out}(GH1)$ over $G_{base}(GH1)$ is one more consumed C_3N molecule, whereas $G_{out}(GH1)^*$'s improvement is a more substantial three C_3N molecules.

On the catabolic ledger, $G_{out}(GH1)^*$ adds reaction 4 to $G_{base}(GH1)$'s set. This consumes one more molecule of C_3N but provides little extra energy, 0.189. Comparing $G_{out}(GH1)^*$ to $G_{out}(GH1)$, the former has a net catabolic *production* of one C_3N molecule, whereas the latter removes two. Hence, $G_{out}(GH1)$ appears to hold a regulatory advantage during catabolism. However, note that for its main biomass type, 2, $G_{out}(GH1)^*$ *gains* five molecules during catabolism, whereas $G_{out}(GH1)$ gains nothing for its two main biomass types. Thus, although the trade-offs may balance on the catabolic side, $G_{out}(GH1)^*$'s anabolic dominance over $G_{out}(GH1)$ appears to be the key to stellar environmental control over the entire 7,000-generation run.

Table 6. Compounds and reactions for 10×20 Scenario GH2. The atoms *C*, *H*, *N*, and *O* are used only to enhance readability and recall of compound names; in no way do the atoms or compounds directly model their real-world counterparts. Energy values (prior to catalytic enhancement) appear in the second column, with positive (negative) values for endothermic (exothermic) reactions. The third column lists pairs of occurrence frequencies for the G_{base} and G_{out} cases, respectively. Each value denotes the number of organisms using that reaction in their metabolism after the final (1,000th) generation. The population size had been at or about a maximum level of 100 since very early in both simulations. Compounds: NO CHN CN_2O N_2O_2 CN_2O_2 N_3O_2 N_3O_3 CHN_3O_2 CN_4O_3 CHN_4O_3.

Reaction	Energy	Frequency
0: $CHN + N_2O_2 \Rightarrow CHN_3O_2$	0.98	(99 : 6)
1: $CHN_3O_2 \Rightarrow CHN + N_2O_2$	−0.98	(1 : 46)
2: $CN_2O + N_2O_2 \Rightarrow CN_4O_3$	1.00	(73 : 100)
3: $CN_4O_3 \Rightarrow CN_2O + N_2O_2$	−1.00	(5 : 0)
4: $NO + N_2O_2 \Rightarrow N_3O_3$	0.92	(3 : 49)
5: $N_3O_3 \Rightarrow NO + N_2O_2$	−0.92	(33 : 23)
6: $NO + CN_4O_3 \Rightarrow CN_2O_2 + N_3O_2$	−0.28	(12 : 5)
7: $CN_2O_2 + N_3O_2 \Rightarrow NO + CN_4O_3$	0.28	(75 : 47)
8: $CHN + N_3O_3 \Rightarrow CHN_4O_3$	0.92	(4 : 98)
9: $CHN_4O_3 \Rightarrow CHN + N_3O_3$	−0.92	(9 : 1)
10: $N_3O_3 \Rightarrow NO + NO + NO$	−1.58	(40 : 68)
11: $NO + NO + NO \Rightarrow N_3O_3$	1.58	(10 : 44)
12: $NO + N_2O_2 \Rightarrow NO + NO + NO$	−0.66	(38 : 16)
13: $NO + NO + NO \Rightarrow NO + N_2O_2$	0.66	(25 : 97)
14: $CHN_3O_2 \Rightarrow NO + NO + CHN$	−1.56	(3 : 9)
15: $NO + NO + CHN \Rightarrow CHN_3O_2$	1.56	(87 : 99)
16: $CHN + N_2O_2 \Rightarrow NO + NO + CHN$	−0.57	(19 : 13)
17: $NO + NO + CHN \Rightarrow CHN + N_2O_2$	0.57	(88 : 28)
18: $N_2O_2 \Rightarrow NO + NO$	−1.00	(93 : 97)
19: $NO + NO \Rightarrow N_2O_2$	1.00	(16 : 1)

4.1.2   Scenario GH2

In a different, shorter, Gaia hunt, the random chemistry in Table 6 led to an interesting scenario, GH2. Here, environmental control emerged but could only persist for 700–800 generations. Although the graphs of Figure 12 display impressive regulation ($\overline{GRI}$ = 15.93), extended follow-up G_{out} runs of several thousand generations revealed a consistent breakdown in control.

In GH2, the physical variable, p, is computed as

$$p = [\tilde{N}O] - [C\tilde{H}N] + [C\tilde{N}_2O] - [N_2\tilde{O}_2] \tag{9}$$

As in scenario GH1, the G_{base} case yields an average p value well above p^*. Hence, control involves decreasing p via (a) increasing CHN or N_2O_2, or (b) decreasing NO or CN_2O.

As shown in Table 7, the two runs differ in their primary biomass type, with G_{base} favoring type 1 and G_{out} using type 2. The difference between these is CHN_3O_2 in type 1, and CHN_4O_3 in type 2. Hence, type 2 biomass requires one more molecule of NO, and this extra NO consumption (Figure 13) enhances regulation, but, once again, at a metabolic cost.
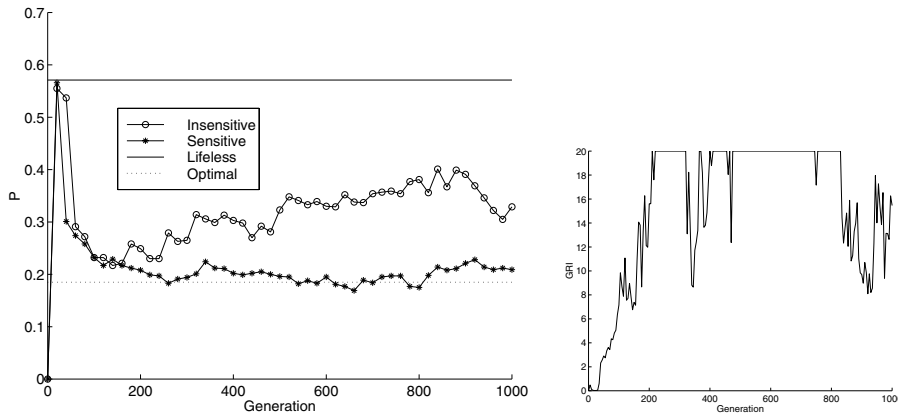
Figure 12. Homeostatic behavior in Scenario GH2. Leftmost graph displays physical variable, p, when the metabolic rates are sensitive (G_{out}) and insensitive (G_{base}) to p. Rightmost graph portrays the Gaian regulatory index (GRI) for the G_{out} run.

Table 7. Biomasses used by the 100 organisms at the end of the 1,000th generation of Scenario GH2's G_{base} and G_{out} runs. In each row, the column entries for the compounds used in the biomass combination type contain the number of organisms using that combination.

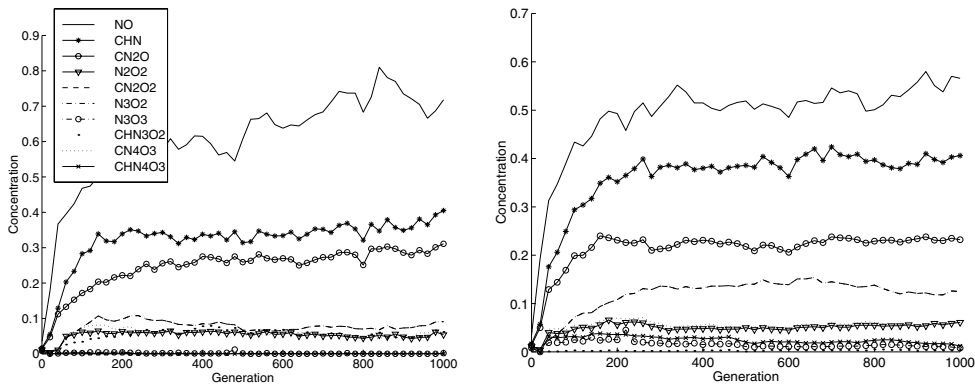| Biomass combination type	N_3O_3	CN_4O_3	CHN_3O_2	CHN_4O_3
G_{base}				
1		94	94	
2		4		4
3	2		2	
Total	2	98	96	4
G_{out}				
1		1	1	
2		99		99
Total	0	100	1	99

Figure 13. Environmental chemical concentrations for the G_{base} (left) and G_{out} (right) runs of Scenario GH2, where the main difference is in the concentrations of NO. The graphs differ slightly in scale.

Table 8. Anabolic differences between the evolved metabolisms of the G_{base} and G_{out} runs for GH2. The anabolic combinations above are the four most popular endothermic reactions for the two runs. Energy requirements take catalytic effects into account by reducing the energy demand by the standard catalytic factor of 1 + catalyst size. Net biomass gain (rightmost column) involves three terms for the three biomass combination types from Table 7. The anabolic efficiency values for the primary biomass types of each run are in bold.

Reaction	Energy used	Net biomass gain
Common		
15: $NO + NO + CHN \Rightarrow CHN_3O_2$	1.56	1,0,1
G_{base} only		
0: $CHN + N_2O_2 \Rightarrow CHN_3O_2$	0.98	1,0,1
7: $CN_2O_2 + N_3O_2 \Rightarrow NO + CN_4O_3$	0.28	1,1,0
17: $NO + NO + CHN \Rightarrow CHN + N_2O_2$	0.14	0,0,0
G_{base} total	2.96	3,1,2
G_{base} anabolic efficiency		**0.99**, 2.96, 1.48
G_{out} only		
2: $CN_2O + N_2O_2 \Rightarrow CN_4O_3$	1.00	1,1,0
8: $CHN + N_3O_3 \Rightarrow CHN_4O_3$	0.92	$0, 1, -1$
13: $NO + NO + NO \Rightarrow NO + N_2O_2$	0.22	0,0,0
G_{out} total	3.70	2, 2, 0
G_{out} anabolic efficiency		1.85, **1.85**, —

Table 9. Catabolic differences between the evolved metabolisms of the G_{base} and G_{out} runs for GH2. The exact catabolic combinations above are the three most popular exothermic reactions from the two runs. Energy production ignores catalytic effects, since these alter reaction rates but not the energy yield per unit of reactant. Net biomass gain (rightmost column) involves three terms for the three biomass combination types from Table 7.

Reaction	Energy produced	Net biomass gain
Common		
10: $N_3O_3 \Rightarrow NO + NO + NO$	1.58	$0, 0, -1$
18: $N_2O_2 \Rightarrow NO + NO$	1.00	0,0,0
G_{base} only		
12: $NO + N_2O_2 \Rightarrow NO + NO + NO$	0.66	0,0,0
G_{base} total	3.24	$0, 0, -1$
G_{out} only		
1: $CHN_3O_2 \Rightarrow CHN + N_2O_2$	0.98	$-1, 0, -1$
G_{out} total	3.57	$-1, 0, -2$

A brief comparison of Table 8 to Table 3 indicates that the underlying chemistry in GH2 offers no guarantees of metabolic success. The average anabolic efficiencies are 1.08 and 1.85 for G_{base} and G_{out}, respectively. These are both much worse (i.e., higher) than the worst efficiency, 0.6, in scenario GH1. Hence, the metamics may already be struggling for existence in the G_{base}(GH2) run. It is therefore understandable that regulation has a limited duration. In fact, in a series of G_{out}(GH2) repeats (with different random seeds), the regulatory metabolisms only arose on occasion and never persisted for more than 400–600 generations.

On the catabolic side, G_{out} organisms get slightly more energy from their three main reactions (Table 9). By using reaction 1, they reap a huge energy benefit without loss of biomass. Conversely, the G_{base} metamics use CHN_3O_2 as biomass and would therefore have trouble exploiting reaction 1 to its fullest. Instead, G_{base} catabolisms use reaction 12, which has a high energy yield but which produces NO. By using reaction 1 instead of 12, the G_{out} agents keep *NO* lower, thus enhancing control. Recycling in the two

GH2 cases is nearly identical, with cycling ratios in the 4–7 range for G_{base} and the 5–8 range for G_{out}.

4.1.3   Results Summary

The Gaia hunt and case studies therein indicate that emergent environmental control can easily occur in systems where (a) energy consumption and biomass production interrelate in a standard manner, and (b) bidirectional interactions exist between metabolisms and physical factors. Whether one uses the quantitative criteria detailed earlier (and gets 20% Gaia) or one simply views the graphs in Figure 8 (and sees 5–7% Gaia) has little bearing on the general conclusion: Although regulation is not inevitable, it is certainly within the realm of plausibility.

Even within a particular chemical environment, predictions may be ambiguous. For instance, the fact that GH2 turned up during the Gaia hunt (with the highest $\overline{GRI}$ value!) was clearly only good fortune, since repeated runs of the same scenario did not always lead to regulation. Conversely, other scenarios based on other chemistries that did not yield high $\overline{GRI}$ values may have fared better on other attempts. In short, many of the emergences of Gaian control appear contingent.

Once the population locks into a particular dominant biomass type and supporting metabolisms, the shift to another regime rarely occurs in these fixed-influx scenarios. Hence, Gaia does not appear to be a global attractor in all of the successful cases, although scenarios such as GH1 show a marked tendency toward control. In a set of isolated tests, changes to the fluxes often incurred biomass and metabolic shifts, but no extensive Gaia hunts in dynamic environments have yet been performed.

The highlighted cases, along with several others, reveal a trade-off between efficient metabolism and environmental control. Evolution in the G_{base} scenarios typically finds the chemistry's most metabolically useful reactions, whereas the G_{out} and G_{in} organisms also have a p value to contend with, thereby complicating their search in metabolism space.

5   Related Work

Daisyworld [24] and its many offspring [12, 14, 18, 19] dominate Gaian model-based thinking. All are simple and elegant, but all are based on standard differential equations and do not include models of individuals nor genetics. Since all necessary genotypes are available from the start, the regulatory behaviors that emerge are largely hardwired.

To date, few artificial life researchers have investigated Gaian issues. The EUZONE model [7] of the evolution of aquatic ecosystems is motivated by Gaian thinking and achieves the emergence of one species, vertically migrating photosynthesizers, which creates a niche for another species, aerobic bottom feeders, via its effects upon the chemical environment. However, environmental regulation does not arise in EUZONE.

GUILD and METAMIC do parallel research into the emergence of autocatalytic sets [11], metabolic systems [1], and hypercycles [2], except that we focus on (a) interaction pathways involving both organisms and chemicals, and (b) the self-organized regulation of the environment by the evolving biota. The Gaia hunt is inspired by Kauffman's [11] NK-landscape explorations, wherein many sets of randomly generated primitive interactions are simulated, and the emergent patterns categorized. But of course, the Gaia hunt reported above is only the beginning of a much needed, more comprehensive future investigation.

In general, many ALife systems involve populations of genotypes that encode for feeding, metabolizing, mating, and other strategies, but the focus is almost exclusively on the emergent structure of the evolving populations themselves, not upon the emergent environmental effects, whether regulatory or otherwise.

One notable exception is swarm intelligence [3], based on the phenomena of stigmergy, wherein simple organisms, for example, social insects, interact indirectly by environmental markings, such as pheromone trails or cell patterns in a honeycomb. In essence, the organisms communicate and coordinate via the global structure that emerges from their parallel activity. The building, and particularly the maintenance, of this structure is easily cast as a control problem. One key difference is that in Gaia, the target environmental condition is somehow optimal for the continued survival of the organisms themselves, although the beehive or ant nest is also clearly an integral part of an insect colony's well-being. Apparently, a comparitive analysis of stigmergy and Gaia could shed interesting light on both the phenomena themselves and the general principles underlying emergent environmental effects. Swarm intelligence is probably the most important contemporary ALife research area with strong links to Gaia theory.

6   Discussion

A few ardent Gaia believers envision a world in which biotic control over the biosphere is ubiquitous, optimal (they just have not quantified exactly *what* is being optimized), and the fundamental organizing principle of nature—above even natural selection itself. However, the majority of serious scientists who follow Gaian developments with interest and curiosity take a more conservative stance. To them, the biogeochemical data and computer models hint of a phenomenon that may be (a) prevalent, (b) far from optimal but generally satisfactory in a world where large-scale physical factors can easily erase much of the biota's best and worst work in a geological heartbeat, and (c) completely consistent with natural selection.

This research is grounded in the latter view. Our motivation is an understanding of some of the fundamental mechanisms by which evolving organisms could wrestle some degree of environmental control from the brute physical forces of nature. In short, how could Gaian phenomena emerge on an evolutionary scale? Hence, GUILD, and particularly METAMIC, take a step beyond Daisyworld by assuming very little about the initial phenotype pool but still providing a virtual incubator for distributed homeostatic emergence, in full harmony with individual-based natural selection.

As for ubiquity versus prevalence, METAMIC was designed to test the *general feasibility*, not the necessity, of emergent environmental control among populations of evolving, metabolizing microbes. No commitment to a particular chemistry was desired, but the chemical foundation needed (a) a restricted set of reactions, and (b) a qualitatively realistic link between energy, structural entropy, and biomass. The resulting evolution of regulation in many test runs (based on a variety of random chemistries) supports this feasibility hypothesis.

However, the trade-offs observed between efficient metabolisms and those with Gaian side effects partially support the Neo-Darwinian *cheaters-will-prosper* attack. Indeed, the p-value-controlling side effects can exact a heavy cost in METAMIC, although they do provide an immediate local selective advantage. The empirical question is the standard one: Do the benefits exceed the costs? Clearly, no right-minded Gaian nor anti-Gaian could give a definitive answer to this general question, and neither can METAMIC. In some runs in some chemical worlds, the benefits dominated and control emerged as an apparent stable attractor; in others, it arose occasionally as something akin to a saddle point; and in many cases, the chemistry simply did not offer many metabolic choices, although sustained life was possible. So if METAMIC has captured some essence of the real world, then one basic take-back-to-the-wild lesson from the computer runs is simply that Gaian homeostasis is neither tautological nor impossible, but clearly contingent on a host of physicochemical, and quite possibly historical, factors.

Of course, the real world has only one chemistry, but it also has many different chemical systems with varying constraints and dynamics based on, among others, the physical surroundings. And even within a particular chemical system, the path of microbial evolution is hardly deterministic. So the use of hundreds of abstract chemistries instead of one real chemistry should not automatically disqualify our Gaia-hunting results from legitimate biological consideration. Rather, it will hopefully inspire biochemists to explore further METAMIC-like simulators using their own chemical models.

Regarding optimality, METAMIC gives no evidence of perfect (i.e., zero-error) robust control. A population is often capable of moving the physical variable to p_{out}^* or p_{in}^*, but not both. Similarly, it may regulate under one set of environmental fluxes but completely lose control under another. In GUILD, the unrestricted metabolic pool enabled this kind of flexible regulation, but the addition of energy constraints and small reaction sets hinders that behavior in METAMIC. Many MD-CHEM chemistries appear to admit one or a few possibilities for both (a) metabolic subsistence, and (b) p-value control. But even under fixed conditions, the energetic sacrifices of the latter may support only temporary control, as in scenario GH2.

Although one prime motivation of METAMIC vis-a-vis GUILD is to test Gaian emergence under restrictive chemical situations, all chemistries in the Gaia hunt were 10x20 (10 compounds and 20 reactions), providing somewhat limited metabolic possibilities. Presumably, larger and more intricate chemistries would expand the pool of viable metabolisms, supporting a truly diverse population of organisms that could evolve to follow moving set-points and changing fluxes. A few tests with 15×30 and 20×40 chemistries have been run, and control has been observed, but so far, no indications of increased robustness have surfaced.

Although GUILD was tested in a two-dimensional world, and control emerged, METAMIC has not yet been ported to a spatial model. Without larger chemistries, the niching potential in two-dimensional situations might not be exploited, so these two improvements should be investigated simultaneously.

Finally, for more realistic Gaia tests, METAMIC should also run on a model of real biochemistry. Although easier said than done, simplified models might be sufficiently amenable to our framework. However, even then, the general approach still falls prey to attacks from a genetic angle, since the direct mapping from genes to metabolic reactions greatly simplifies reality. Thus, future genomes might encode the initial conditions for a morphogenetic process whose final result is a stable intracellular biochemical cycle: a metabolism.

Of course, one can always find deficiencies with Gaia and inaccuracies in its supporting models compared to the real world. Still, examples such as Daisyworld, GUILD, and METAMIC show that emergent distributed environmental control can occur under the guidance of standard individual-based natural selection in abstract domains that have qualitative grounding in the natural world. Essentially, all three projects had the same goal: reconciling Gaia theory with natural selection. This is the main hurdle to establishing Gaia as a viable hypothesis, amenable to scientific analysis and objective evaluation. This would also aid reconciliation at the personal level, where renowned scientists on both sides of the Gaia argument have bickered more over the new-wave ecological and spiritual interpretations that Gaia research accidentally spawned than over the biogeochemical data.

Our aim has been to exploit ALife tools in showing that the basic phenomenon of emergent distributed control of an environment by naturally selected agents is possible. Whether Gaia actually occurs on Earth is another, more complicated question, and one to which we have no strong convictions. We only hope that our results inspire further Gaian investigations from the ALife perspective. It is often said that Gaia

theory has its Aristotle (in James Lovelock) but still needs a Newton to formalize the concept thoroughly. A good ALife background could be a key prerequisite for that job.

References

1. Bagley, R., & Farmer, J. (1992). Spontaneous emergence of a metabolism. In C. Langton, C. Taylor, J. Farmer, & S. Rasmussen (Eds.), *Artificial life II* (pp. 93–140). Reading, MA: Addison-Wesley.

2. Boerlijst, M., & Hogeweg, P. (1992). Self-structure and selection: Spiral waves as a substrate for prebiotic evolution. In C. Langton, C. Taylor, J. Farmer, & S. Rasmussen (Eds.), *Artificial life II* (pp. 255–276). Reading, MA: Addison-Wesley.

3. Bonabeau, E., Dorigo, M., & Theraulaz, G. (1999). *Swarm intelligence: From natural to artificial systems*. New York: Oxford University Press.

4. Dawkins, R. (1982). *The extended phenotype*. Oxford, UK: Oxford University Press.

5. Di Fenizio, P. S. (2000). A less abstract artificial chemistry. In M. Bedau, J. McCaskill, N. Packard, & S. Rasmussen (Eds.), *Artificial life VII* (pp. 49–53). Cambridge, MA: MIT Press.

6. Dittrich, P. (1999). *Artificial chemistries*. Tutorial held at ECAL'99 (European Conference on Artificial Life).

7. Downing, K. (1998). Euzone: Simulating the emergence of aquatic ecosystems. *Artificial Life, 3*, 307–333.

8. Downing, K. (2000). Exploring Gaia theory: Artificial life on a planetary scale. In M. Bedau, J. McCaskill, N. Packard, and S. Rasmussen (Eds.), *Proceedings of the 7th International Conference on Artificial Life* (pp. 90–99). Cambridge, MA: MIT Press.

9. Downing, K., & Zvirinsky, P. (1999). The simulated evolution of biochemical guilds: Reconciling Gaia theory and natural selection. *Artificial life, 5*, 291–318.

10. Fontana, W. (1992). Algorithmic chemistry. In C. G. Langton, C. Taylor, J. D. Farmer, & S. Rasmussen (Eds.), *Artificial life II* (pp. 159–209). Reading, MA: Addison-Wesley.

11. Kauffman, S. (1993). *The origins of order*. New York: Oxford University Press.

12. Lenton, T. (1998). Gaia and natural selection. *Nature, 394*, 439–447.

13. Lenton, T., & Watson, A. (2000). Redfield revisited: Regulation of nitrate, phosphate and oxygen in the ocean. *Global Biogeochemical Cycles, 14*, 225–248.

14. Lovelock, J. (1995). *The ages of Gaia: A biography of our living Earth*. Oxford, UK: Oxford University Press.

15. Lovelock, J., & Margulis, L. (1974). Atmospheric homeostasis by and for the biosphere: The Gaia hypothesis. *Tellus, 26*, 2–10.

16. Mayer, B., & Rasmussen, S. (1998). Self-reproduction of dynamical hierarchies in chemical systems. In C. Adami, R. Belew, H. Kitano, & C. Taylor (Eds.), *Artificial life VI* (pp. 123–129). Cambridge, MA: MIT Press.

17. Riggs, D. S. (1963). *The mathematical approach to physiological problems*. Cambridge, MA: MIT Press.

18. Robertson, D., & Robinson, J. (1998). Darwinian Daisyworld. *Journal of Theoretical Biology, 195*, 129–134.

19. Saunders, P. (1994). Evolution without natural selection: Further implications of the Daisyworld parable. *Journal of Theoretical Biology, 166*, 365–373.

20. Schlesinger, W. (1997). *Biogeochemistry: An analysis of global change*. Boston: Academic Press.

21. Schwartzman, D., & Volk, T. (1991). Biotic enhancement of weathering and surface temperatures on Earth since the origin of life. *Palaeogeography, Palaeoclimatology, Palaeoecology, 90*, 357–371.

22. Suzuki, Y., & Tanaka, H. (1998). Order parameter for a symbolic chemical system. In
    C. Adami, R. Belew, H. Kitano, and C. Taylor (Eds.), *Artificial life VI* (pp. 130–139).
    Cambridge, MA: MIT Press.

23. Volk, T. (1998). *Gaia's body: Toward a physiology of Earth*. New York: Copernicus.

24. Watson, A., & Lovelock, J. (1983). Biological homeostasis of the global environment: The
    parable of Daisyworld. *Tellus, 35B*, 284–289.

25. Williams, G. (1996). *The molecular biology of Gaia*. New York: Columbia University Press.