smallest mean squared error. This estimate is a weighted sum of the mean of the prior and the sensed feedback position:

$$x_{\text{estimated}} = \frac{\sigma^2_{\text{sensed}}}{\sigma^2_{\text{sensed}} + \sigma^2_{\text{prior}}}[1\,\text{cm}] + \frac{\sigma^2_{\text{prior}}}{\sigma^2_{\text{sensed}} + \sigma^2_{\text{prior}}}x_{\text{sensed}}$$

Given that we know σ^2_{prior}, we can estimate the uncertainty in the feedback σ_{sensed} by linear regression from Fig. 2a.

Resulting mean squared error

The mean squared error (MSE) is determined by integrating the squared error over all possible sensed feedbacks and actual lateral shifts

$$\text{MSE} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x_{\text{estimated}} - x_{\text{true}})^2 p(x_{\text{sensed}}|x_{\text{true}})p(x_{\text{true}})\mathrm{d}x_{\text{sensed}}\mathrm{d}x_{\text{true}}$$

For model 1, $x_{\text{estimated}} = x_{\text{sensed}}$, and this gives $\text{MSE} = \sigma^2_{\text{sensed}}$.
Using the result for $x_{\text{estimated}}$ from above for model 2 gives
$\text{MSE} = \sigma^2_{\text{sensed}}\sigma^2_{\text{prior}}/(\sigma^2_{\text{sensed}} + \sigma^2_{\text{prior}})$, which is always lower than the MSE for model 1. If the variance of the prior is equal to the variance of the feedback, the MSE for model 2 is half that of model 1.

Inferring the used prior

An obvious choice of $x_{\text{estimated}}$ is the maximum of the posterior

$$p(x_{\text{true}}|x_{\text{sensed}}) = \frac{1}{\sigma_{\text{sensed}}\sqrt{2\pi}}e^{-(x_{\text{true}} - x_{\text{sensed}})^2/2\sigma^2_{\text{sensed}}}\ p(x_{\text{true}})/p(x_{\text{sensed}})$$

The derivative of this posterior with respect to x_{true} must vanish at $x_{\text{estimated}}$. This allows us to estimate the prior used by each subject. Differentiating and setting to zero we get

$$\left.\frac{\mathrm{d}p(x_{\text{true}})}{\mathrm{d}x_{\text{true}}}\frac{1}{p(x_{\text{true}})}\right|_{x_{\text{estimated}}} = \frac{(x_{\text{estimated}} - x_{\text{sensed}})}{\sigma^2_{\text{sensed}}}$$

We assume that x_{sensed} has a narrow peak around x_{true} and thus approximate it by x_{true}. We insert the σ_{sensed} obtained above, affecting the scaling of the integral but not its form. The average of x_{sensed} across many trials is the imposed shift x_{true}. The right-hand side is therefore measured in the experiment and the left-hand side approximates the derivative of $\log(p(x_{\text{true}}))$. Since $p(x_{\text{true}})$ must approach zero for both very small and very large x_{true}, we subtract the mean of the right-hand side before integrating numerically to obtain $\log(p(x_{\text{true}}))$, which we can then transform to estimate the prior $p(x_{\text{true}})$.

Bimodal distribution

Six new subjects participated in a similar experiment in which the lateral shift was bimodally distributed as a mixture of two gaussians:

$$p(x_{\text{true}}) = \frac{1}{2\sqrt{2\pi}\sigma_{\text{prior}}}\left(e^{-(x - x_{\text{sep}}/2)^2/\sigma^2_{\text{prior}}} + e^{-(x + x_{\text{sep}}/2)^2/\sigma^2_{\text{prior}}}\right)$$

where $x_{\text{sep}} = 4$ cm and $\sigma_{\text{prior}} = 0.5$ cm. Because we expected this prior to be more difficult to learn, each subject performed 4,000 trials split between two consecutive days. In addition, to speed up learning, feedback midway through the movement was always blurred (25 spheres distributed as a two-dimensional gaussian with a standard deviation of 4 cm), and feedback at the end of the movement was provided on every trial. Fitting the bayesian model (using the correct form of the prior and true σ_{prior}) to minimize the MSE between actual and predicted lateral deviations of the last 1,000 trials was used to infer the subject's internal estimates of both x_{sep} and σ_{sensed}. Some aspects of the nonlinear relationship between lateral shift and lateral deviation (Fig. 3a) can be understood intuitively. When the sensed shift is zero, the actual shift is equally likely to be to the right or the left and, on average, there should be no deviation from the target. If the sensed shift is slightly to the right, such as at 0.25 cm, then the actual shift is more likely to come from the right-hand gaussian than the left, and subjects should point to the right of the target. However, if the sensed shift is far to the right, such as at 3 cm, then because the bulk of the prior lies to the left, subjects should point to the left of the target.

1. van Beers, R. J., Sittig, A. C. & Gon, J. J. Integration of proprioceptive and visual position-information: An experimentally supported model. *J. Neurophysiol.* **81,** 1355–1364 (1999).
2. Jacobs, R. A. Optimal integration of texture and motion cues to depth. *Vision Res.* **39,** 3621–3629 (1999).
3. Ernst, M. O. & Banks, M. S. Humans integrate visual and haptic information in a statistically optimal fashion. *Nature* **415,** 429–433 (2002).
4. Hillis, J. M., Ernst, M. O., Banks, M. S. & Landy, M. S. Combining sensory information: mandatory fusion within, but not between, senses. *Science* **298,** 1627–1630 (2002).
5. Cox, R. T. Probability, frequency and reasonable expectation. *Am. J. Phys.* **17,** 1–13 (1946).
6. Bernardo, J. M. & Smith, A. F. M. *Bayesian Theory* (Wiley, New York, 1994).
7. Berrou, C., Glavieux, A. & Thitimajshima, P. Near Shannon limit error-correcting coding and decoding: turbo-codes. *Proc. ICC'93, Geneva, Switzerland* 1064–1070 (1993).
8. Simoncelli, E. P. & Adelson, E. H. Noise removal via Bayesian wavelet coring. *Proc. 3rd International Conference on Image Processing, Lausanne, Switzerland, September,* 379–382 (1996).
9. Olshausen, B. A. & Millman, K. J. in *Advances in Neural Information Processing Systems* vol. 12 (eds Solla, S. A., Leen, T. K. & Muller, K. R.) 841–847 (MIT Press, 2000).
10. Rao, R. P. N. An optimal estimation approach to visual perception and learning. *Vision Res.* **39,** 1963–1989 (1999).
11. Hinton, G. E., Dayan, P., Frey, B. J. & Neal, R. M. The 'wake-sleep' algorithm for unsupervised neural networks. *Science* **268,** 1158–1161 (1995).
12. Sahani, M. & Dayan, P. Doubly distributional population codes: Simultaneous representation of uncertainty and multiplicity. *Neural Comput.* **15,** 2255–2279 (2003).
13. Yu, A. J. & Dayan, P. Acetylcholine in cortical inference. *Neural Netw.* **15,** 719–730 (2002).
14. Weiss, Y., Simoncelli, E. P. & Adelson, E. H. Motion illusions as optimal percepts. *Nature Neurosci.* **5,** 598–604 (2002).
15. Soechting, J. F. & Flanders, M. Errors in pointing are due to approximations in sensorimotor transformations. *J. Neurophysiol.* **62,** 595–608 (1989).
16. Krakauer, J. W., Pine, Z. M., Ghilardi, M. F. & Ghez, C. Learning of visuomotor transformations for vectorial planning of reaching trajectories. *J. Neurosci.* **20,** 8916–8924 (2000).
17. Lacquaniti, F. & Caminiti, R. Visuo-motor transformations for arm reaching. *Eur. J. Neurosci.* **10,** 195–203 (1998).
18. Van Beers, R. J., Baraduc, P. & Wolpert, D. M. Role of uncertainty in sensorimotor control. *Phil. Trans. R. Soc. Lond. B* **357,** 1137–1145 (2002).
19. van Beers, R. J., Wolpert, D. M. & Haggard, P. When feeling is more important than seeing in sensorimotor adaptation. *Curr. Biol.* **12,** 834–837 (2002).
20. Harris, C. M. & Wolpert, D. M. Signal-dependent noise determines motor planning. *Nature* **394,** 780–784 (1998).
21. Todorov, E. & Jordan, M. I. Optimal feedback control as a theory of motor coordination. *Nature Neurosci.* **5,** 1226–1235 (2002).
22. Wolpert, D. M., Ghahramani, Z. & Jordan, M. I. An internal model for sensorimotor integration. *Science* **269,** 1880–1882 (1995).
23. Vetter, P. & Wolpert, D. M. Context estimation for sensorimotor control. *J. Neurophysiol.* **84,** 1026–1034 (2000).
24. Scheidt, R. A., Dingwell, J. B. & Mussa-Ivaldi, F. A. Learning to move amid uncertainty. *J. Neurophysiol.* **86,** 971–985 (2001).
25. Fiorillo, C. D., Tobler, P. N. & Schultz, W. Discrete coding of reward probability and uncertainty by dopamine neurons. *Science* **299,** 1898–1902 (2003).
26. Basso, M. A. & Wurtz, R. H. Modulation of neuronal activity in superior colliculus by changes in target probability. *J. Neurosci.* **18,** 7519–7534 (1998).
27. Platt, M. L. & Glimcher, P. W. Neural correlates of decision variables in parietal cortex. *Nature* **400,** 233–238 (1999).
28. Carpenter, R. H. & Williams, M. L. Neural computation of log likelihood in control of saccadic eye movements. *Nature* **377,** 59–62 (1995).
29. Gold, J. I. & Shadlen, M. N. The influence of behavioral context on the representation of a perceptual decision in developing oculomotor commands. *J. Neurosci.* **23,** 632–651 (2003).
30. Goodbody, S. J. & Wolpert, D. M. Temporal and amplitude generalization in motor learning. *J. Neurophysiol.* **79,** 1825–1838 (1998).

....

Functional genomic hypothesis generation and experimentation by a robot scientist

Ross D. King[1], **Kenneth E. Whelan**[1], **Ffion M. Jones**[1], **Philip G. K. Reiser**[1], **Christopher H. Bryant**[2], **Stephen H. Muggleton**[3], **Douglas B. Kell**[4] **& Stephen G. Oliver**[5]

[1]*Department of Computer Science, University of Wales, Aberystwyth SY23 3DB, UK*
[2]*School of Computing, The Robert Gordon University, Aberdeen AB10 1FR, UK*
[3]*Department of Computing, Imperial College, London SW7 2AZ, UK*
[4]*Department of Chemistry, UMIST, P.O. Box 88, Manchester M60 1QD, UK*
[5]*School of Biological Sciences, University of Manchester, 2.205 Stopford Building, Manchester M13 9PT, UK*

....

The question of whether it is possible to automate the scientific process is of both great theoretical interest[1,2] **and increasing practical importance because, in many scientific areas, data are being generated much faster than they can be effectively analysed. We describe a physically implemented robotic system that applies techniques from artificial intelligence**[3–8] **to carry out cycles of scientific experimentation. The system automatically**

originates hypotheses to explain observations, devises experiments to test these hypotheses, physically runs the experiments using a laboratory robot, interprets the results to falsify hypotheses inconsistent with the data, and then repeats the cycle. Here we apply the system to the determination of gene function using deletion mutants of yeast (*Saccharomyces cerevisiae*) and auxotrophic growth experiments[9]. We built and tested a detailed logical model (involving genes, proteins and metabolites) of the aromatic amino acid synthesis pathway. In biological experiments that automatically reconstruct parts of this model, we show that an intelligent experiment selection strategy is competitive with human performance and significantly outperforms, with a cost decrease of 3-fold and 100-fold (respectively), both cheapest and random-experiment selection.

The branch of artificial intelligence devoted to developing algorithms for acquiring scientific knowledge is known as 'scientific discovery'. The pioneering work in the field was the development of learning algorithms for analysis of mass-spectrometric data[3]. In the subsequent 30 years much has been achieved and there are now several convincing examples in which computer programs have made explicit contributions to scientific knowledge[4–8]. However, the general impact of such programs on science has been limited. This could now change because the expansion of automation in science is making it increasingly possible to couple scientific discovery software to laboratory instrumentation[5]. Here we describe a novel system that extends the existing level of integration of scientific discovery and robotics: the 'Robot Scientist' (Fig. 1).

A widely accepted view of science is that it follows a 'hypothetico-deductive' process[1]. Scientific expertise and imagination are first used to form possible hypotheses, and then the deductive consequences of these hypotheses are tested by experiment. The Robot Scientist follows this paradigm: we employ the logical inference mechanism of abduction[10] to form new hypotheses, and that of deduction to test which hypotheses are consistent (see Methods and Supplementary Information). The system has been physically implemented and conducts biological assays with minimal human intervention after the robot is set up. The hardware platform consists of a liquid-handling robot with its control PC, a plate reader with its control PC, and a master PC to control the system and do the scientific reasoning. The software platform consists of background knowledge about the biological problem, a logical inference engine, hypothesis generation code (abduction), experiment selection code (deduction), and the Laboratory Information Management System (LIMS) code that integrates the whole system. The robot conducts assays by pipetting and mixing liquids on microtitre plates. Given a computed definition of one or more experiments, we have developed code that designs a layout of reagents on the liquid-handling platform that will allow these experiments, with controls, to be performed efficiently. In addition, the liquid-handling robot is automatically programmed to plate out the yeast and media into the correct wells. The system measures the concentration of yeast in the wells of the microtitre trays using the adjacent plate reader and returns the results to the LIMS (although microtitre trays are still moved in and out of incubators manually).
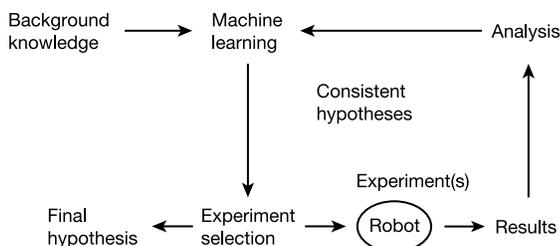
Figure 1 The Robot Scientist hypothesis-generation and experimentation loop.

The key point is that there was no human intellectual input in the design of experiments or the interpretation of data.

To test the Robot Scientist we chose the field of functional genomics. It is one of the most important in contemporary science and is an area in which laboratory automation is already mature. The current state of the art in functional genomics is to use highly automated robotics to generate data, and then to use data-mining systems to extract knowledge from that data. Within functional genomics, we selected yeast (*S. cerevisiae*) to test the methodology. Despite being one of the best understood of organisms, the functions of about 30% of the 6,000 genes in yeast are still unknown[11]. The aim was to develop a system that could automatically determine the function of genes from the performance of knockout mutants (strains in which one gene has been removed). We focused on the aromatic amino acid synthesis (AAA) pathway (Fig. 2), and used auxotrophic growth experiments to assess the behaviour (phenotype) of the mutants. The AAA pathway is relatively well understood and of sufficient complexity to make reasoning about it non-trivial, and its intermediary metabolites are commercially available. Auxotrophic growth experiments consist of growing auxotrophic mutants on chemically defined media (a defined base plus one or more intermediate or terminal metabolites in the pathway), and observing whether growth is recovered or not (see Supplementary Information for details). A knockout mutant is auxotrophic if it cannot grow on a defined medium on which the wild type can grow. Auxotrophic experiments are a classic technique for inferring metabolic pathways[9].

We needed a way of representing prior biological knowledge in the computer, so we developed a logical formalism to model cellular metabolism that captures the key relationship between protein-coding sequences (open reading frames; ORFs), enzymes and metabolites in a pathway[12]. All objects (ORFs, proteins and metabolites) and relationships (coding, reactions, transport and feedback) are described as logical formulae. The structure of the metabolic pathway is that of a directed graph, with metabolites as nodes and enzymes as arcs. An arc corresponds to a reaction. The compounds at each node are the set of all metabolites that can be synthesized by the reactions leading to it. Reactions are modelled as unidirectional transformations. Using this formalism, we have implemented a model of the AAA pathway with the logic programming language Prolog (the complete model is provided in Supplementary Information). Prolog makes it possible both to inspect the biological knowledge in the model directly and to compute the predictions of the model automatically. The model infers (deduces) that a knockout mutant will grow if, and only if, a path can be found from the input metabolites to the three aromatic amino acids. This allows the model to compute the phenotype of a particular knockout or to be used to infer missing reactions that could explain an observed phenotype (abduction). We consider that most hypothesis generation in modern biology is abductive. What is inferred are not general hypotheses, which would be inductive, but specific facts about biological entities.

The original bioinformatic information for the AAA model was taken mainly from the KEGG[13] catalogue of metabolism. The model was then tested with all possible auxotrophic experiments involving a single replacement metabolite, and was altered manually to fit the empirical results. To ensure that the model was not 'over-fitted', we carried out all possible auxotrophic experiments with pairs of metabolites. The model correctly predicted at least 98.5% of the experiments (Supplementary Information). To the best of our knowledge, no bioinformatic model has been as thoroughly tested with knockout mutants.

Machine learning is the branch of artificial intelligence that seeks to develop computer systems that improve their performance automatically with experience[14,15]. It has much in common with statistics, but differs in having a greater emphasis on algorithms, data representation and making acquired knowledge explicit. The

branch of machine learning that deals with algorithms that can choose experiments is known as 'active learning'[16,17]. If we assume that each hypothesis has a prior probability of being correct and that each experiment has an associated cost, then scientific experiment selection can be formalized as the task of selecting the optimal series of experiments (in terms of expected cost) to eliminate all except the one correct hypothesis. This problem is, in general, computationally intractable (NP-hard)[18]. However, it can be shown that the problem is structurally the same as finding the smallest decision tree—experiments are nodes and hypotheses are leaves[19]. This is significant because a bayesian analysis of decision-tree learning has shown that near-optimal solutions can be found in polynomial time[19]. This analysis leads to the following approximate recurrence formula for the expected cost[20]. Let $EC(H,T)$ denote the minimum expected cost of experimentation given the set of candidate hypotheses H and the set of candidate trials T:

$$EC(\varnothing, T) = 0$$

$$EC(\{h\}, T) = 0$$

$$EC(H,T) \approx \min_{t \in T}[C_t + p(t)(\text{mean}_{t' \in (T-t)}C_{t'})J_{H[t]} + (1 - p(t))$$

$$\times (\text{mean}_{t' \in (T-t)}C_{t'})J_{H[t]}]$$

$$J_H = -\Sigma_{h \in H}p(h)[\log_2(p(h))]$$

where C_t is the monetary price of the trial t, $p(t)$ is the probability that the outcome of the trial t is positive, and $(\ldots($ is the 'floor' function. $p(t)$ can be computed as the sum of the probabilities of the hypotheses (h) that are consistent with a positive outcome of t.

In current robot-scientist experiments, it is assumed that the system knows the biochemical equations of the AAA pathway in the wild type. What the system does not know is how this biochemistry is related to the genetics. The hypotheses are therefore bindings between the ORFs and enzymic reactions; that is, which of the 15 possible ORFs in the pathway has been deleted (although only 8 ORFs were ever actually used for technical reasons, the system was not aware of this). For example, a correct hypothesis would be that ORF YPR060c codes for chorismate mutase (which catalyses

the reaction chorismate → prephenate; Fig. 2). These hypotheses are automatically generated by abducing the different possibilities from the model[12,20–22]. Hypotheses are evaluated by comparing their predicted consequences with the actual experimental results. These logical inferences are made with our machine learning system ASE-Progol[23], where ASE stands for 'active selection of experiments' (ASE-Progol is freely available to academics, from ⟨ftp://www.comp.rgu.ac.uk/pub/staff/chb/systems/ase_progol/version_1.0⟩). We compared three experimental selection strategies: first, ASE (our approximation to choosing the experiment that minimizes the expected cost—see above); second, Naive (which chooses the cheapest experiment that has not yet been done); and last, Random (which chooses a random experiment that has not yet been done).

The performance of a strategy is defined as the average accuracy of all hypotheses not already eliminated by experiments using that strategy. The accuracy of a single hypothesis is the number of correct predictions that this hypothesis makes, for all possible single-metabolite and double-metabolite addition experiments, assuming that the model is 100% accurate. We evaluated accuracy versus the monetary price of the experiments (pounds sterling) and time (iterations; that is, the number of sequential experiments).

Using the Robot Scientist, we compared the experimental strategies ASE, Naive and Random over five experimental iterations (Fig. 3a, b). At the end of the fifth iteration, ASE had an estimated accuracy of 80.1%, in comparison with 74.0% for Naive and 72.2% for Random. At this point in the iterative cycle of experimentation, the ASE strategy was significantly more accurate than either Naive ($P < 0.05$) or Random ($P < 0.07$) using a paired t-test. Given a maximum spend of £$10^{2.26}$ (the maximum that Naive spends; note that these prices are scaled), ASE has an estimated accuracy of 79.5%, in comparison with 73.9% for Naive and 57.4% for Random. For this price, ASE is significantly more accurate than either Naive ($P < 0.05$) or Random ($P < 0.001$). Although the AAA pathway is sufficiently complex to provide an adequate test for the robot scientist system, it is not so complex that the Naive and Random strategies will not eventually achieve a high accuracy. Where ASE shows its superiority over the other two experimental strategies is in
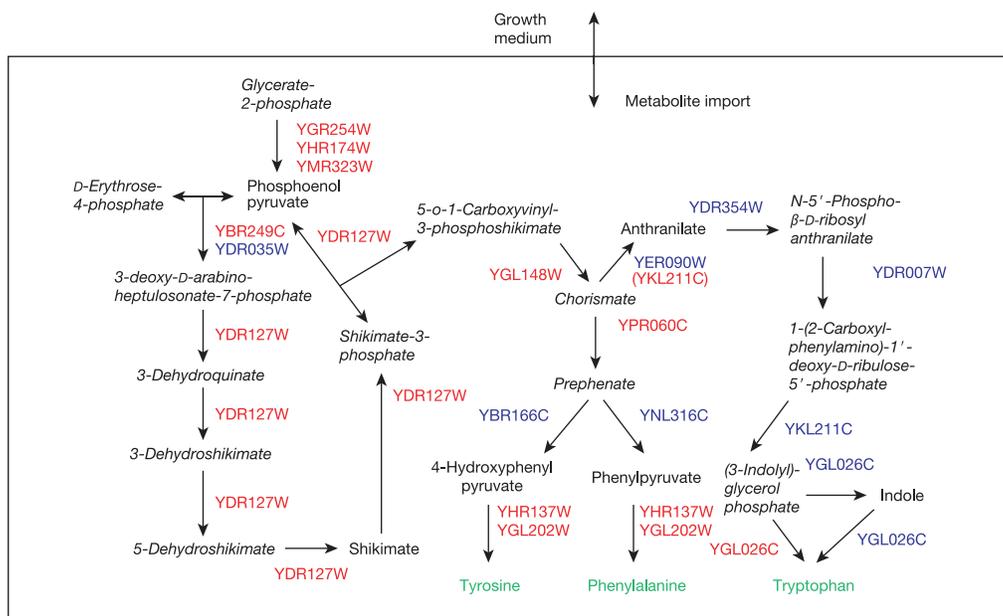
Figure 2 A schematic representation of the logical model of the aromatic amino acid pathway in yeast. The metabolites are the nodes and the enzymes the arcs in the directed graph. Metabolites that could not be used in auxotrophic experiments are in italics, and the end-product aromatic amino acids are shown in green. The ORFs encoding the enzyme proteins are blue if they are amenable to auxotrophic experimentation; otherwise they are red. The logical model also represents the import of metabolites into the cell.

249

cost-effectiveness. In science, as in industry, 'time is money' (although the conversion rate may be unclear). The experimental cost we wish to optimize will therefore be a function of price and time. Because ASE dominates for both, ASE is superior to Naive and Random for any positive function of price and time. For example, to achieve an accuracy of about 70%, ASE requires fewer trial iterations, and a hundredth of the price, of Random, and almost half the number of iterations, and a third of the price, of Naive.

We evaluated the reliability of these empirical results by running several computer simulations of the Robot Scientist. In these we used the AAA Prolog model and the ASE-Progol inference system, and we took the results of manual experiments. When necessary, we also added random noise by flipping the observation of growth to no-growth and the reverse. We compared differing types of experiment (single metabolites or double metabolites), numbers of iterations, noise, and noise reduction strategies. In Fig. 3c, d we show the results for 0% and 25% experimental noise, plotting accuracy against time and accuracy against price. In these typical experiments, the ASE strategy almost completely dominates those of Naive and Random. These simulation experiments are consistent with the results *in vivo*, because we estimate that the physical system has about 25% noise in the observations of whether growth occurred or not. The high level of noise is due to the robot's being open to the air (and therefore to microbial contamination). We were also interested to compare the performance of the Robot Scientist system with that of humans. To do this, we adapted the simulator to allow humans to choose and interpret the result of cycles of experimentation. In initial trials, using nine graduate computer scientists and biologists, we found that there was no significant difference between the robot and the best human performance in terms of the number of iterations required to

achieve a given level of accuracy.

The inference of gene function in the AAA pathway was selected to be typical of scientific inference tasks in molecular genetics. We argue that most molecular genetics differs from this model only in the much greater sophistication of the abductive process used to select a possible hypothesis and by the far greater elegance of the experiments used to discriminate among hypotheses. Although the AAA problem is relatively simple, it could still be applied directly to 'real-world' problems. For instance, if a series of AAA metabolism mutants from a related (but genetically uncharacterized) fungal species had been isolated, the *S. cerevisiae* AAA pathway could be used to guide auxotrophic growth experiments in exactly the same way as described here.

Although pioneering work by others (see ref. 22, for example) has developed automated systems to induce genetic networks from mutant data, we believe that the Robot Scientist is the first example of a closed-loop system that actually designs and executes experiments to test inferred hypotheses. Nevertheless, the Robot Scientist has currently only been demonstrated to rediscover the role of genes of known function; this initial step was essential to ensure that the system was working. We now plan to extend the system to be able to uncover the function of genes whose role is currently unknown. To enable this, the background cellular metabolism model will need to be extended. This will involve the translation of bioinformatic databases such as KEGG into our logical formalism; the resultant models are likely to be less reliable than that of our carefully checked AAA model. The abductive hypothesis generation method will also need to be extended to allow the inference of missing arcs, not just their binding. Initial work has been done on this[12].

The general robot scientist idea could be applied to many scientific problems, and we are actively investigating drug design
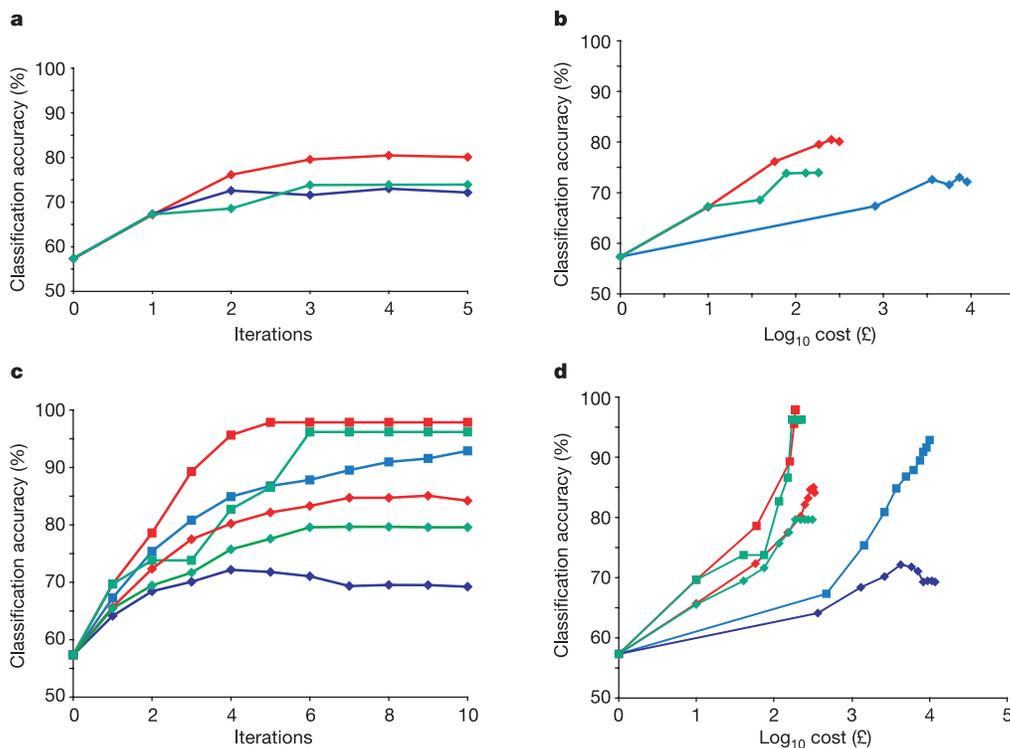
Figure 3 Actual and simulated performance of the Robot Scientist. **a**, **b**, Actual performance. **a**, Plot of average classification accuracy against experimental iteration (from four experimental repeats). **b**, Plot of average classification accuracy against average experimental scaled price (from four experimental repeats). **c**, **d**, Simulated performance. **c**, Plot of average classification accuracy against time for 0% noise (squares) and 25% noise (diamonds) (based on 100 simulation runs). **d**, Plot of average classification accuracy against average scaled price for 0% noise (squares) and 25% noise (diamonds) (based on 100 simulation runs). Strategies: red curves, ASE; blue curves, Random; green curves, Naive.

and quantum control. To apply the Robot Scientist to drug design, the robot would need to be able to perform a biological assay to perform the given optimization, and either synthesize compounds of required structure using lab-on-a-chip technology or choose them from a large library. The hypotheses would be different qualitative structure–activity relationships. The quantum control of chemical synthesis with femtosecond lasers[24] is especially challenging because of the ability and desirability of doing experiments very quickly (about 10^3 per second at present). At such speeds, human intervention is impossible, and automatic systems become essential.

The Robot Scientist extends the state of the art in integrating scientific discovery software with laboratory robotics. Moreover, the application of the Robot Scientist to functional genomics provides further evidence that some aspects of scientific reasoning can be formalized and efficiently automated. In science, experiment selection is generally done informally and without explicit regard to the cost of eliminating hypotheses; when formal experimental design[25] is employed, this is generally done to ensure that sufficient repeats and controls are in place to answer the desired question. Our results show that different experiment selection strategies can have significantly different results in terms of cost, even for the solution of a simple problem. This suggests that there remains scope to improve the general cost-effectiveness of science by developing better tools to help choose efficient experiments.

Automation was the driving force of much of nineteenth-century and twentieth-century change, and this is likely to continue. It is also becoming increasingly important in scientific research: for example, the sequencing of the human genome was made possible by factory production techniques, and modern drug discovery relies on high-throughput screening robots. We consider this trend to increased automation of science to be both inevitable and desirable. It is inevitable because it will be required to deal with the challenges of science in the twenty-first century. It is also desirable because it frees scientists to make the high-level creative leaps at which they excel. □

Methods

Growth experiments and the amino acid pathway

The mutants (of *Saccharomyces cerevisiae* strain BY4741 (ATCC201388) *MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0*; ref. 26) had the complete reading frame of each protein-encoding gene deleted by replacement with a selectable marker gene that had no phenotype in the absence of the selective agent[27]. They were therefore non-revertible null mutants. Limitations in the availability of mutant yeast strains restricted the number of genes used in the *in vivo* investigations to 15, and the auxotrophic experimental requirement for a difference in growth phenotype reduced this number to the following 8 (the other mutants always either grew or failed to grow): ybr166c, ydr007w, ydr035w, ydr354w, yer090w, ygl026c, ykl211c and ynl316c. The number of possible metabolites was limited to nine by availability and cost: anthranilate (10), indole (190), *p*-hydroxyphenolpyruvic acid (193), L-phenylalanine (53), phenylpyruvate (30), phosphoenolpyruvate (9385), shikimic acid (633), L-tyrosine (53) and L-tryptophan (53). The numbers in parentheses are the normalized true experimental costs of using each metabolite in a growth medium; note the approximate three orders of magnitude range.

Computational model of the aromatic amino acid pathway

The Prolog model of the pathway was refined in three stages. First, the original model was translated from KEGG[13] and carefully checked with the literature. We checked that all the KEGG reactions were documented in *S. cerevisiae* (consistency) and that there were no other related reactions described in the literature (completeness). Second, the predictions of the model were compared with the results of the single-metabolite experiments (see above). Whether growth or lack of growth was observed was at this point decided visually. Because certain metabolites did not seem to affect growth in the way predicted by the literature, we refined the model to make these metabolites unable to be imported into the cells efficiently. It was also necessary to add inhibition effects (see Supplementary Information for details). Last, the results of the double-metabolite experiments were tested against the model, and the automatic growth-calling software was optimized by learning a decision tree to fit the experimental results to the model (see Supplementary section 2). The model developed on the single metabolites was consistent with all except less than 1.5% of the double-metabolite experiments. The model was not further changed to include these experimental discrepancies.

Performance measures

The average performance of the hypotheses is an appropriate performance measure

because it rewards learners that discriminate between competing hypotheses. This approach is a compromise between selecting the hypothesis with the highest probability and weighting all predictions by the probability of the hypotheses that generated them. In active learning, the performance curves that have been generally used plot predictive accuracy against the number of training examples. Often, two curves are plotted on the same graph, one for active learning and one for random sampling. The accuracy of a single hypothesis is the number of correct predictions that this hypothesis makes about all possible single-metabolite and double-metabolite experiments, based on using the model as the oracle.

Such performance plots allow the difference in the number of experiments (examples per unit time) required to reach a particular level of performance to be compared. However, one drawback of such plots is that they ignore any variation in the price of obtaining individual examples. When such variation does exist, and the aim is to compare the price of attaining particular levels of performance, these plots are potentially misleading. To overcome this drawback we also plot the cumulative price of the experiments against performance[20]. For this we use the normalized price of the metabolite. At the start of the experiments, when there are eight possible hypotheses, the average accuracy is 57%.

Structure of the Robot Scientist

The Robot Scientist automates the task of liquid handling and can conduct assays by pipetting and mixing liquids on microtitre plates. The robot is controlled using Tcl (Tool Command Language), and we have written a compiler that translates Prolog commands into Tcl robot operations. Given a Prolog definition of one or more experiments, we have developed code that designs a layout of the robot that will allow these experiments, with controls, to be performed efficiently. In addition, the robot has to be automatically programmed to plate out the yeast and media into the correct wells. The microtitre plates were measured with the adjacent plate reader and the results were returned to the LIMS. Transfer of plates from the robot to the incubator, and from the incubator to the plate reader, was done manually.

1. Popper, K. *The Logic of Scientific Discovery* (Hutchinson, London, 1972).
2. Sloman, A. *The Computer Revolution in Philosophy* (Harvester, Hassocks, Sussex, 1978); available online from ⟨http://www.cs.bham.ac.uk/research/cogaff/crp/⟩.
3. Buchanan, B. G., Sutherland, G. L. & Feigenbaum, E. A. in *Machine Intelligence* Vol. 4 (eds Meltzer, B. & Michie, D.) 209–254 (Edinburgh Univ. Press, 1969).
4. Langley, P., Simon, H. A., Bradshaw, G. L. & Zytkow, J. M. *Scientific Discovery: Computational Explorations of the Creative Process* (MIT Press, Cambridge, Massachusetts, 1987).
5. Żytkow, J. M., Zhu, J. & Hussam, A. Automated discovery in a chemistry laboratory. in *Proceedings of the 8th National Conference on Artificial Intelligence (AAAI-1990)* (eds Dietterich, T. & Swartout, W.) 889–894 (MIT, Cambridge, Massachusetts, 1990).
6. King, R. D., Muggleton, S. H., Srinivasan, A. & Sternberg, M. J. E. Structure–activity relationships derived by machine learning: The use of atoms and their bond connectivities to predict mutagenicity by inductive logic programming. *Proc. Natl Acad. Sci. USA* **93,** 438–442 (1996).
7. Valdes-Perez, R. E. Discovery tools for science applications. *Commun. ACM* **42,** 37–41 (1999).
8. Langley, P. The computational support of scientific discovery. *Int. J. Hum.–Comput. Stud.* **53,** 393–410 (2000).
9. Beadle, G. W. & Tatum, E. I. Genetic control of biochemical reactions in *Neurospora. Proc. Natl Acad. Sci. USA* **27,** 499–506 (1941).
10. Pierce, C. S. *Collected Papers of Charles Sanders Pierce* (Harvard Univ. Press, 1958).
11. Mewes, H. W. *et al.* MIPS: a database for genomes and protein sequences. *Nucleic Acids Res.* **30,** 31–34 (2002); ⟨http://mips.gsf.de/proj/yeast/CYGD/db/index.html⟩.
12. Reiser, P. G. K. *et al.* Developing a logical model of yeast metabolism. *Electron. Trans. Artif. Intell.* **5,** 223–244 (2001).
13. Kanehisa, M. & Goto, S. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* **28,** 27–30 (2000); ⟨http://www.genome.ad.jp/kegg/⟩.
14. Mitchell, T. M. *Machine Learning* (McGraw-Hill, New York, 1997).
15. Langley, P. *Elements of Machine Learning* (Morgan Kaufmann, San Mateo, California, 1996).
16. Cohen, D. A., Ghahramani, Z. & Jordan, M. I. Active learning with statistical models. *J. Artif. Intell. Res.* **4,** 129–145 (1996).
17. Lin, F.-R. & Shaw, M. J. Active training of backpropagation neural networks using the learning by experimentation methodology. *Ann. Oper. Res.* **75,** 129–145 (1997).
18. Fedorov, V. V. *Theory of Optimal Experiments* (Academic, London, 1972).
19. Muggleton, S. & Page, D. in *Machine Intelligence* Vol. 15 (eds Furukawa, K., Michie, D. & Muggleton, S.) 248–267 (Oxford Univ. Press, 1999).
20. Bryant, C. H., Muggleton, S. H., Oliver, S. G., Kell, D. B., Reiser, P. G. K. & King, R. D. Combining inductive logic programming, active learning, and robotics to discover the function of genes. *Electron. Trans. Artif. Intell.* **5,** 1–36 (2001).
21. Flach, P. & Kaka, A. *Abduction and Induction* (Kluwer, London, 2000).
22. Zupan, B. *et al.* in *Proceedings of the Eighth European Conference on Artificial Intelligence in Medicine* (eds Qualini, S., Barahona, P. & Andreassen, S.) 304–313 (Springer, Berlin, 2001).
23. Muggleton, S. Inverse entailment and Progol. *New Generation Comput. J.* **13,** 245–286 (1995).
24. Rabitz, H., de Vivie-Riedle, R., Motzkus, M. & Kompa, K. Whither the future of controlling quantum phenomena? *Science* **288,** 824–828 (2000).
25. Cochran, W. G. & Cox, G. M. *Experimental Designs* (Wiley, New York, 1992).
26. Brachmann, C. B. *et al.* Designer deletion strains derived from *Saccharomyces cerevisiae* S288C: a useful set of strains and plasmids for PCR-mediated gene disruption and other applications. *Yeast* **14,** 115–132 (1998).
27. Giaever, G. *et al.* Functional profiling of the *Saccharomyces cerevisiae* genome. *Nature* **418,** 387–391 (2002).

Correspondence and requests for materials should be addressed to S.G.O. (steve.oliver@man.ac.uk).

....

Immune recognition of a human renal cancer antigen through post-translational protein splicing

**Ken-ichi Hanada[1], Jonathan W. Yewdell[2] & James C. Yang[1]**

[1]*Surgery Branch, National Cancer Institute, National Institutes of Health, 9000 Rockville Pike, Building10, Room 2B42, Bethesda, Maryland 20892, USA*
[2]*Laboratory of Viral Diseases, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland 20892, USA*

....

Cytotoxic T lymphocytes (CTLs) detect and destroy cells displaying class I molecules of the major histocompatibility complex (MHC) that present oligopeptides derived from aberrant self or foreign proteins. Most class I peptide ligands are created from proteins that are degraded by proteasomes and transported, by the transporter associated with antigen processing, from the cytosol into the endoplasmic reticulum, where peptides bind MHC class I molecules and are conveyed to the cell surface[1]. C2 CTLs, cloned from human CTLs infiltrating a renal cell carcinoma, kill cancer cells overexpressing fibroblast growth factor-5 (FGF-5)[2]. Here we show that C2 cells recognize human leukocyte antigen-A3 MHC class I molecules presenting a nine-residue FGF-5 peptide generated by protein splicing. This process, previously described strictly in plants[3] and unicellular organisms[4], entails post-translational excision of a polypeptide segment followed by ligation of the newly liberated carboxy-terminal and amino-terminal residues. The occurrence of protein splicing in vertebrates has important implications for the complexity of the vertebrate proteome and for the immune recognition of self and foreign peptides.

To identify the FGF-5-derived epitope recognized by C2, we transfected human leukocyte antigen (HLA)-A3-expressing COS-7 cells (COS-A3) with expression plasmids encoding truncated forms of the FGF-5 gene and measured C2 activation by its release of interferon-γ (IFN-γ) (Fig. 1a). Antigenicity was lost by truncation either between predicted amino acid residues 212 and 220 (G4/G5 plasmids) or residues 161 and 172 (G8/G9 plasmids) of FGF-5. The shortest stimulatory fragment identified (G8) encoded 60 amino acids (residues 161–220).

We next examined the antigenicity of synthetic peptides corresponding to all of the 8-residue, 9-residue and 10-residue peptides contained within the defined 60 amino-acid fragment, but none were recognized by C2 when incubated with APCs expressing HLA-A3 (data not shown). We similarly tested peptides encoded by the +1 and +2 reading frames of the fragment but did not find the epitope (data not shown). Post-translational structural modification of a minimal determinant within the 60 amino-acid fragment remained a possible explanation for our failure to identify a
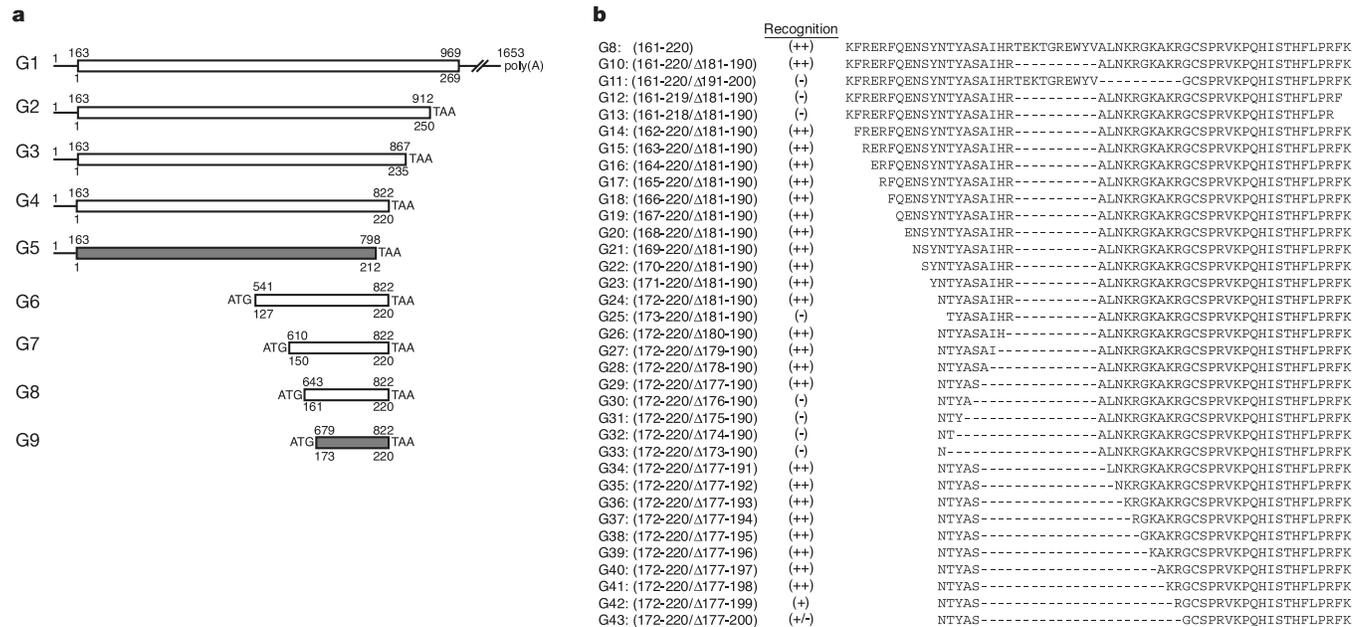
Figure 1 Genetic truncation analysis of the FGF-5-encoded determinant. **a**, A series of truncation mutants (G2–G9) were prepared by PCR from construct G1. An initiation codon and the Kozak consensus (CACCATG) were introduced into the PCR primers for constructs G6–G9. Each PCR product was cloned into the eukaryotic expression vector and transfected into COS-A3 cells; recognition by C2 was assessed by IFN-γ secretion. Nucleic acid positions are indicated above the start and finish of each construct and the corresponding amino acid positions from FGF-5 below. Constructs shown as open boxes were recognized; those shown as filled boxes were not. **b**, Analysis of internal deletion mutants. For simplicity, predicted amino acid sequences instead of DNA sequences are shown. The antigenicity of each construct was assessed as described in **a**. All constructs were designed to start with the Kozak consensus sequence and end with the termination codon TAA. Numbers in parentheses indicate the position of the first and the last amino acids. Numbers after Δ indicate internally deleted amino acid residues. The degree of CTL recognition is categorized as follows depending on IFN-γ secretion: (++), more than 1,000 pg ml^{-1}; (+), 200–1,000 pg ml^{-1}; (+/−), 100–200 pg ml^{-1}; (−), less than 100 pg ml^{-1}.