

- 31 Sine, S.M. (1997) Identification of equivalent residues in the γ , δ , and ϵ subunits of the nicotinic receptor that contribute to α -bungarotoxin binding. *J. Biol. Chem.* 272, 23521–23527
- 32 Czajkowski, C. and Karlin, A. (1995) Structure of the nicotinic receptor acetylcholine-binding site. Identification of acidic residues in the δ subunit within 0.9 nm of the α subunit-binding site disulfide. *J. Biol. Chem.* 270, 3160–3164
- 33 Sussman, J.L. *et al.* (1991) Atomic structure of acetylcholinesterase from *Torpedo californica*: a prototypic acetylcholine-binding protein. *Science* 253, 872–879
- 34 Zhong, W. *et al.* (1998) From *ab initio* quantum mechanics to molecular neurobiology: a cation– π binding site in the nicotinic receptor. *Proc. Natl. Acad. Sci. U. S. A.* 95, 12088–12093
- 35 Sullivan, D.A. and Cohen, J.B. (2000) Mapping the agonist binding site of the nicotinic acetylcholine receptor. Orientation requirements for activation by covalent agonist. *J. Biol. Chem.* 275, 12651–12660
- 36 Khromov-Borisov, N.V. and Michelson, M.J. (1966) The mutual disposition of cholinergic receptors of locomotor muscles, and changes in their disposition in the course of evolution. *Pharmacol. Rev.* 18, 1051–1090
- 37 Corringer, P.J. *et al.* (1998) Critical elements determining diversity in agonist binding and desensitization of neuronal nicotinic acetylcholine receptors. *J. Neurosci.* 18, 648–657
- 38 Galzi, J.L. *et al.* (1996) Identification of calcium binding sites that regulate potentiation of a neuronal nicotinic acetylcholine receptor. *EMBO J.* 15, 5824–5832
- 39 Galzi, J.L. *et al.* (1991) Allosteric transitions of the acetylcholine receptor probed at the amino acid level with a photolabile cholinergic ligand. *Proc. Natl. Acad. Sci. U. S. A.* 88, 5051–5055
- 40 Sayle, R. and Milner-White, E.J. (1995) RASMOL: biomolecular graphics for all. *Trends Biochem. Sci.* 20, 374

Thomas Grutter

Jean-Pierre Changeux*

Institut Pasteur, URA 2182 CNRS 'Récepteurs et Cognition', Dept of Biotechnology, 25 rue du Dr. Roux, 75015 Paris, France.

*e-mail: changeux@pasteur.fr

Complex networks of interactions connect genes to phenotypes

Athel Cornish-Bowden and María Luz Cárdenas

Recent work on the galactose-utilization pathway of yeast has shown how transcriptome and proteome data can be combined to deduce a network of hundreds of genes involved in protein–protein and protein–DNA interactions. Such a network leads to a new picture about how the pathway is regulated that is clearer and more complete than what was previously known.

According to William Bains, 'the genome has turned out to be a relatively poor source of explanation for the differences between cells or between people'¹. Negative statements of this sort contrast wildly with the triumphalism that characterized reports on genome projects only a few years ago. However, they are beginning to be classed as no more than realistic; hardly anyone can have failed to notice that the capacity of genome sequencing to generate mountains of new data has far outstripped our capacity to make sense of it all. For example, the WIT integrated pathway–genome database (<http://wit.integratedgenomics.com/ERGO>) now contains information about more than 80 genomes; one year ago there were a few more than 40. Clearly, there is an urgent need for more efficient and powerful methods of data analysis because without these a fully sequenced genome is hardly more useful than a list of telephone numbers without all the associated names. The genomic problem is much more complex, as assigning a

molecular function to every gene is only a beginning: afterwards the network of all the interactions between genes has to be established to see how effects on the whole system are produced; this requires integration of many different kinds of methods.

DNA microarrays already enable the mRNA expression levels of almost all the genes in an organism to be measured²; corresponding techniques to measure protein expression are currently being developed³. Applied separately, each of these is insufficient to predict new system properties. However, in a recent paper⁴ Leroy Hood and his colleagues have shown how these two measurements can be usefully combined with one another to validate and extend the existing model of how galactose utilization is regulated in yeast⁵. Their integration of mRNA and protein expression responses with the global set of protein–protein and protein–DNA interactions has allowed them to deduce a network involving hundreds of genes.

Combining deletions of each of nine galactose-utilization genes (Table 1) with two dietary states – with and without galactose – the authors compared the effects of 20 perturbations on cellular gene expression. They found that the levels of 997 mRNAs varied significantly with one or more of these perturbations – a far cry from the naive expectation, still common, that deletion (or overexpression) of one gene should affect expression of just that

gene. Although widespread pleiotropy of this kind was predicted and understood many years ago⁶, by the few who had the vision to realize that systems need to be studied as systems and not as mere collections of parts, it has only become a matter of common observation much more recently⁷. Underlining the need to look at both mRNA and protein expression patterns, and not to rely on just one or the other to provide the whole story, the correlation between these two patterns in the galactose utilization study proved to be far from perfect: some protein expression levels remained virtually unchanged despite large changes in the expression of mRNA; others changed substantially with

Table 1. GAL genes in yeast

Gene	Type of product	Function of product
GAL1	Enzyme	Galactokinase
GAL2	Transporter	Galactose transporter
GAL3	Regulator	Relieves effect of Gal80p
GAL4	Regulator	Activates expression of all other GAL genes
GAL5	Enzyme	Phosphoglucomutase
GAL6	Regulator	Acts in a drug-resistance pathway
GAL7	Enzyme	Gal-1-P Glc-1-P uridylyltransferase
GAL10	Enzyme	UDPgalactose epimerase
GAL80	Regulator	Represses effect of Gal4p

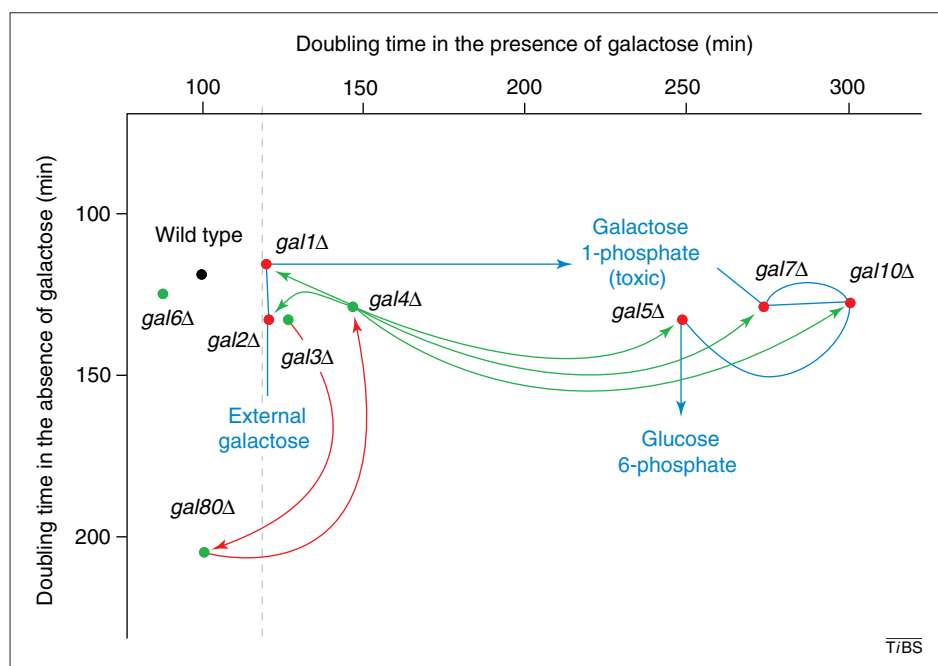


Fig. 1. Growth rates of different deletion mutants in the presence and absence of galactose. The wild type is shown by a black circle, deletions of genes for transporters or enzymes by red circles, and deletions of regulatory genes by green circles. The green arrows show the regulatory effect of Gal4p on the transporter and metabolic enzymes, and the red arrows show inhibitory effects of other genes. The pathway from galactose, via galactose 1-phosphate, to glucose 6-phosphate is superimposed on the red symbols. Note the clear separation between those that act before and after galactose 1-phosphate in the pathway. The vertical dotted line maps the growth rate of the wild type in the absence of galactose onto the scale for growth in the presence of galactose.

negligible changes in mRNA levels. By and large, however, the plot of one against the other yielded an intelligible scatter, with genes of functionally related roles appearing in clusters; one such cluster is provided by *GAL1*, *GAL7* and *GAL10*, three genes that code for three enzymes specific to galactose metabolism.

Authors of a 'proof-of-principle' paper, designed to show that a proposed new method is useful, need to steer a careful course between the Scylla of too little novelty and the Charybdis of too much. Just confirming the validity of a previous model suggests that the new approach might be valid while satisfying no obvious need; large amounts of new information with little confirmation of what was known before raises doubts about the reliability of the new approach. Hood and his colleagues⁴ have the balance about right. They used the existing biochemical model of regulation in the galactose-utilization pathway to predict the effects of their 20 perturbations on the expression levels of the *GAL* genes. Many of these predictions would have been obvious without a model, for example that deleting the *GAL1* gene would lower the expression of *GAL1*, but the majority were less obvious and were, in general, confirmed by

measurements. For example, when galactose is available the product of *GAL4* is believed to increase the expression of all the genes that code for enzymes that catalyse the transformation of galactose into glucose 6-phosphate, so deletion of *GAL4* should decrease the amounts of these enzymes synthesized during growth in the presence of galactose, but have little or no effect in the absence of galactose. This, in general is what was observed, although expression of *GAL5* did not decrease when *GAL4* was deleted (if anything it increased slightly). This makes sense when one recognizes that the product of *GAL5*, phosphoglucomutase, has important metabolic roles to fulfil regardless of galactose utilization (so much so that its classification as a *GAL* gene is surprising). The enzymes that behaved as expected, such as galactokinase (*Gal1p*), were specific to galactose metabolism.

The preliminary model contained a mechanism to suggest how the presence of galactose in the cell might stimulate its metabolism: intracellular galactose activates the Gal3p-mediated inhibition of the action of Gal80p on Gal4p, with a net result being activation of the genes involved in the galactose pathway. In

general, the *GAL* gene expression results agreed with this but, of course, the model does not predict the changes in expression observed for the hundreds of other genes. Moreover, deletion of the genes for either uridylyltransferase (*Gal7p*) or UDPgalactose epimerase (*Gal10p*), two enzymes needed to convert galactose 1-phosphate into glucose 1-phosphate, resulted in decreased expression of the other galactose enzymes as well. Although the original model did not predict this observation, it did not require radical revision to accommodate it. The immediate effect of inactivation of either enzyme would be accumulation of galactose 1-phosphate: if this had an intrinsic regulatory effect to repress synthesis of other pathway enzymes this would provide an indirect way for deletion of *GAL7* or *GAL10* to repress synthesis of these enzymes. This hypothesis was tested by deleting the gene encoding galactokinase (*Gal1p*) at the same time as deleting *GAL7* or *GAL10*, thus avoiding accumulation of galactose 1-phosphate. Sure enough, the profile over all genes of a *gal1Δgal10Δ* mutant resembled that of a *gal1Δ* mutant in the presence of galactose, not that of a *gal10Δ* mutant, confirming that the properties of the *gal10Δ* and *gal7Δ* mutants were caused by accumulation of galactose 1-phosphate, and not by a purely genetic effect such as interaction between chromosomes.

The potentially harmful effects of galactose 1-phosphate accumulation are also evident in the growth rates of the different mutants (Fig. 1). In the absence of galactose, none of the deletions apart from *gal80Δ* had very much effect on growth but, in the presence of galactose, a clear separation between two groups of genes became evident: any of the regulatory genes, or those for processes leading to formation of galactose 1-phosphate, could be deleted without producing a growth rate much less than that of the wild type in the absence of galactose; by contrast, deleting genes for enzymes needed for metabolism of galactose 1-phosphate resulted in very slow growth in the presence of galactose. We see here an example of the large effect that loss of an enzyme activity can have on metabolite concentrations, typically much larger than its effects on fluxes. This, in part, accounts for the growing interest in the metabolome (i.e. the complete set of metabolite concentrations) in studies of functional genomics, because

measurements of metabolite concentrations can often reveal the effects of genes that otherwise appear to be silent^{8,9}.

The observed pleiotropy underlines the crucial point that genes act in concert with one another and with the environment. The more complex the level at which one seeks to explain a living system, the greater the need to examine the network of interactions that lie behind the genome. This investigation of galactose metabolism⁴ represents a clear step in this direction, opening the door to studies of systems with much less prior information about the genes involved and their regulation than was available in this case.

References

- 1 Bains, W. (2001) The parts list of life. *Nat. Biotechnol.* 19, 401–402
- 2 Lander, E.S. (1999) Array of hope. *Nat. Genet.* 21, 3–4
- 3 Gygi, S.P. *et al.* (1999) Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat. Biotechnol.* 17, 994–999
- 4 Ideker, T. *et al.* (2001) Integrated genomic and proteomic analyses of a systematically perturbed metabolic network. *Science* 292, 929–934
- 5 Lohr, D. *et al.* (1995) Transcriptional regulation in the yeast *GAL* gene family: a complex genetic network. *FASEB J.* 9, 777–787
- 6 Kacser, H. (1963) The kinetic structure of organisms. In *Biological organization at the cellular and supercellular level* (Harris, R.J.C., ed.), pp. 25–41, Academic Press

- 7 Bailey, J.E. (1999) Lessons from metabolic engineering for functional genomics and drug discovery. *Nat. Biotechnol.* 17, 616–618
- 8 Raamsdonk, L.M. *et al.* (2001) A functional genomics strategy that uses metabolome data to reveal the phenotype of silent mutations. *Nat. Biotechnol.* 19, 45–50
- 9 Cornish-Bowden, A. and Cárdenas, M.L. (2001) Silent genes given voice. *Nature* 409, 571–572

Athel Cornish-Bowden*

María Luz Cárdenas

Bioénergétique et Ingénierie des Protéines, CNRS, 31 chemin Joseph-Aiguier, B.P. 71, 13402 Marseille Cedex 20, France.

*e-mail: athel@ibsm.cnrs-mrs.fr

Protein Sequence Motif

A novel superfamily of enzymes that catalyze the modification of guanidino groups

Hiroki Shirai, Tom L. Blundell and Kenji Mizuguchi

Three enzymes, peptidyl-arginine deiminase from *Porphyromonas gingivalis*, arginine deiminase and amidinotransferase are traditionally classified separately. By combining PSI-BLAST and FUGUE, data presented in this article describe how these enzymes belong to a novel superfamily, adopting a common fold and sharing similar catalytic mechanisms.

The peptidyl-arginine deiminase (PAD) enzymes catalyze the deimination of the guanidino group from carboxyl-terminal Arg residues of various peptides to produce ammonia. The bacterium *Porphyromonas gingivalis*, associated with the initiation and progression of adult-onset periodontitis, releases a PAD (Ref. 1), which is a virulence factor that prevents acidic cleansing cycles in the mouth. This PAD (PPAD) appears to be evolutionarily unrelated to mammalian PAD, which is a metal enzyme, although both catalyze the same chemical reaction.

Arginine deiminase (ADI) catalyzes the deimination of the guanidino group of

L-Arg to produce ammonia^{2,3}. ADI is found mainly in prokaryotes and has potential for use as an anti-cancer drug because of its anti-tumor activity³. Recently, however, ADI has also been discovered in a primitive eukaryote, *Giardia intestinalis*, one of the most commonly transmitted intestinal pathogens in the world. With the increasing emergence of drug-resistant strains of this organism, the involvement of ADI in a major metabolic pathway makes it highly attractive as a new drug target². Despite the similarity in substrates and reactions between ADI and PPAD, they share little sequence similarity [percentage sequence identities (PIDs) range from 8 to 12%]². Although both PPAD and ADI are attractive drug targets, attempts to design inhibitors have been hampered because of the lack of knowledge about their catalytic mechanism.

Amidinotransferase (AT) catalyzes the transamidation reaction of L-Arg to Gly in mammalian tissues for creatine synthesis⁴, or to inosamine-phosphate in *Streptomyces* sp. for antibiotic biosynthesis⁵. Both use a common mechanism that was proposed on the basis of their crystal structures^{4,5}. There are no significant sequence similarities between AT and ADI (PIDs 11–14%) nor between AT and PPAD (PIDs 10–12%).

PPAD, ADI and AT share a common fold

Of the total 556 residues of PPAD, the N-terminal 380 residues show some similarity with the full-length sequences of 11 mainly prokaryotic proteins of unknown function. This region of PPAD, as we will describe, comprises the catalytic domain and, thus, these proteins are likely to share common functions and form a new family. We call this family PPAD homologs (PPADH). A FUGUE* (Ref. 6) search against known structures predicted that PPADH would adopt a structure similar to that of AT, with a significant Z-score of 5.25 (95% confidence).

PSI-BLAST (Ref. 7) sequence search starting from a member of ADI, first detected a member of a related family, N(G),N(G)-dimethyl-L-arginine deiminase (DDAH), at the second iteration with an e-value of 4e-06. A member of AT was then detected (at the third iteration with an e-value of 4e-06), followed by most other members of AT. This was supported by FUGUE analysis, in which a sequence profile for ADI produced a significant Z-score of 9.38 (99% confidence) with the structural profile of AT.

The structure of AT is called an α/β propeller⁴. Multiple alignment shows that the structurally important residues

*FUGUE is the sequence–structure homology recognition program whose performance has been extensively benchmarked (see <http://cafasp.bioinfo.pl> for examples). It produces a Z-score, which has been shown to be a sensitive and reliable measure for distant evolutionary relationships.