Essay

Binding reactions: epigenetic switches, signal transduction and cancer

Mark Ptashne

Simple binding interactions lie at the heart of disparate biological functions. Multiple negative and positive 'add-ons', often with small individual effects, make elementary systems that work, work better. Cancer illustrates various of these fundamental processes gone awry.

Molecular biology continues to explode with new facts and details along with the occasional surprise. There is, I believe, an unexpected bonus: a few basic principles underlie many complex processes — signal transduction, gene expression, the maintenance or destruction of gene products, the construction of epigenetic switches, and so on. In some human diseases — cancer, for example — these processes go awry, and a conceptualization of the underlying strategies helps us understand how that can happen. Here I emphasize nature's reiterated use of the simplest of reactions: *binding*.

By binding, I mean the non-covalent interactions of macromolecules: proteins with other proteins, DNA, RNA, or membranes; of RNA with DNA, and so on. The typical interaction I refer to is reversible under physiological conditions, and its essential function is apposition, bringing one macromolecule in contact with another. In this essay I discuss a few examples of how binding reactions are deployed to different ends. Molecular details differ, but similar general strategies are found at work in these systems. The essentials are illustrated by the workings of an epigenetic switch in bacteria, my starting example.

An epigenetic switch: lessons from lambda

The bacteriophage lambda switch ensures that when one set of genes (those for lysogenic growth) are on, another set (the genes for lytic growth) are off, and *vice versa*. Once the repressor gene (*cI*) is switched on (Figure 1, left) and the lysogenic state established, that pattern of the gene expression is self-perpetuated for many bacterial divisions. The switch can be flipped by an environmental signal — such as UV light — but none of the operations of the switch entails a change in DNA sequence. Rather, the switch comprises a set of binding reactions involving two DNA-binding regulatory proteins (repressor and cro), the enzyme RNA polymerase and DNA. Here are some further salient points describing, or inferred from, the switch. These matters, as well as certain others discussed later in this article and not explicitly referenced, have been discussed previously [1,2].

- *Epigenetics.* The self-perpetuating (and hence epigenetic) character of the switch is not an inherent property of any of its components, but rather is a property of the system conferred by the pattern of binding reactions. There are two ways to make epigenetic switches, and lambda's switch includes both: a double-negative loop, in which the product of one gene (repressor) turns off expression of the other gene (*cro*) and vice versa; and a positive feedback loop, in which repressor (despite its name) activates transcription of its own gene. The original name my colleagues and I gave to this switch — we called it a 'genetic' switch — is misleading because, as just mentioned, there is no change in DNA sequence involved [3]. Epigenetic switches comprising lambda-like components are found in many developmental pathways in eukaryotes eukaryotes.
- *Cooperativity.* The switch requires that proteins bind specifically to sites on DNA. For example, in a lysogen repressor must bind to its designated sites in DNA and, more precisely, it must bind predominantly to two of three such sites as shown in Figure 1,

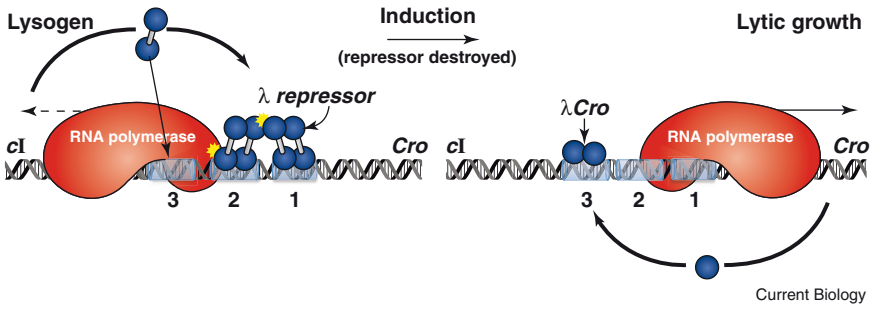

Figure 1. The lambda epigenetic switch

Two states of the switch are shown: on the left the repressor gene (*cI*) is transcribed but the *Cro* gene is not, and *vice versa* on the right. The scenario on the left is found in lambda lysogens, bacteria that carry an otherwise dormant phage lambda. Inactivation of repressor (induction) results in lytic growth of the phage, an early stage of which is shown on the right. Repressor and cro turn each other's genes off by blocking binding of RNA polymerase to the other's promoter: repressor covers the *Cro* gene promoter when bound at sites 1 and 2 as shown on the left, and cro covers the repressor-gene promoter when bound at site 3, as shown on the right. Repressor bound at sites 1 and 2 activates transcription of its own gene (*cI*), as it represses transcription of *Cro*. Repressor maintains its concentration below a specified level by binding, at higher concentrations, to site 3 (as indicated by the downwards arrow), and turning itself off. All of these effects — auto activation and repression by repressor, and the opposing effects of repressor and cro — are effected by simple binding reactions with suitably adjusted binding constants. The figure indicates that the switch can be flipped by a dose of UV light which results indirectly in cleavage of repressor. An additional set of interactions involving repressors bound here and at a site some 2000 base pairs away has been omitted.

on the left. This specificity is facilitated by cooperativity: two repressor dimers touch (bind) each other as shown, each thereby helping the other bind, and to bind specifically. All binding reactions of the sort discussed in this article face the specificity problem, and cooperativity is widely used to help solve the problem (see appendix one in [1]). I return below to a further role of cooperativity in the lambda switch.

• *Concentration control*. The individual DNA sites in Figure 1 differ only modestly in their affinities for repressor — about tenfold. And so site selectivity tends to be rather readily lost as the concentration of repressor increases. Not to worry: the switch has a 'governor' in the form of another binding reaction — as the repressor concentration increases it tends to turn off transcription of its own gene by binding to the third (lower affinity) site, as shown by the downward arrow in Figure 1, on the left. The binding reactions referred to in this article require that binding domains distinguish between related possible targets. These kinds of interactions risk losing specificity as concentrations increase.

• *Activation of transcription – the imposition of specificity by recruitment.* Lambda repressor works as an activator of transcription in another binding reaction: it simultaneously contacts DNA and RNA polymerase (as shown in Figure 1), thereby recruiting the polymerase to the adjacent promoter. Transcription of the gene is 'activated' — that is, the gene is transcribed at a higher level than it otherwise would be. The gene activated by repressor is the repressor-encoding *cI* gene itself and so, by this positive feedback loop, continuous production of repressor is ensured as these lysogenic cells divide.

We say that polymerase has been given specificity — has been instructed to transcribe a particular gene, the *cI* gene — by this recruiting reaction.The effect is modest (increasing the level of transcription some 10–50 fold) and a potentially significant level of transcription will occur in the absence of the activator. When repressor is destroyed and lysogens induced, cro, the DNA-binding protein produced early upon induction, suppresses this basal transcription as shown on the right in Figure 1.

Many eukaryotic enzymes can, like bacterial RNA polymerase, work on any of a wide array of substrates (different genes in the case of RNA polymerase), and which is chosen, under any given set of conditions, is determined by recruitment, as in the example just discussed. These enzymes include, in addition to polymerases, proteases, ubiquytilators, RNA-splicing enzymes, kinases, phosphatases, transcriptional repressing complexes, nucleosome modifying enzymes, and so on. For example, an E2 ligase can add ubiquitin to many proteins, but the choice is dictated (for one class of E2s) by recruiters called F-box proteins. Each of these recruiters simultaneously binds a specific target protein and the enzymatic machinery, and thus imposes specificity on the enzyme. Ubiquitin is added and, in a further binding reaction, the modified protein interacts with a protease and is destroyed.

Recruiting reactions typically face the problem described for activation of transcription: in the absence of the recruiter there can be an unwanted basal level of activity, and we will see a variety of strategies employed to depress that basal activity.

• *Squelching and self-squelching.* Recruiting reactions are subject to two negative effects as the concentration of the recruiter increases. *Squelching*: an over-expressed transcriptional activator, as it activates its target genes, will tend to depress transcription of other genes. The effect is attributed to competition by activators (the recruiters in this case) for binding common sites on the transcriptional machinery. The effect has been observed in transcription experiments performed with yeast and mammalian cells. *Self-squelching*: At very high expression levels, a transcriptional activator ceases to activate even its designated target genes. The result is explained as follows: successful recruitment requires that a single recruiter (a transcriptional activator in this case) simultaneously contacts the transcriptional machinery and a specific DNA binding site. At very high activator concentrations, the machinery and the DNA site will tend to be occupied by separate copies of the activator, and recruitment will be blocked. The effect has been observed in transcription experiments performed in yeast, and in proteolysis experiments in mammalian cells in which the concentration of an E3 ligase (the recruiter in this case) was varied (Pengbo Zhou, personal communication).

For any given case the extent of squelching and self-squelching will depend upon the concentrations of recruiters and targets and the affinities with which they bind each other. As a historical matter, the observation of squelching and self-squelching was one of the early indications that eukaryotic transcriptional activators work by recruitment. Another important kind of experiment in this regard is called a 'by-pass' experiment, as now outlined.

• *By-pass experiments.* A property of reactions facilitated by recruitment is that the recruiter can be dispensed with ('by-passed') if the target and enzyme are brought together in another fashion. For example, bacterial RNA polymerase constitutively transcribes a gene, at a high level, if fused to a DNA-binding domain that binds a site near the gene. (Presumably the DNA binding domain is pulled off its site as the polymerase moves along the gene, but this has not been explicitly demonstrated.) An analogous result obtains if a subunit of the eukaryotic transcription machinery is fused to a DNA-binding domain. In both of these examples the requirement for a transcriptional activator is obviated. Over-producing a target protein can suffice to drive a reaction in the absence of a recruiter. The price paid in such by-pass scenarios is that the ordinary control over the reaction — effected usually by production or modification of the recruiter — is lost.

A by-pass experiment has shown that double stranded RNA (dsRNA) works as a recruiter in the phenomenon called RNA interference (RNAi). In this experiment a 'silencing' protein (a component of the RITS complex) was fused to an RNA-binding protein. The binding site for that protein was inserted into a transcribed yeast gene, and the fusion was found to trigger silencing of that gene. [4]. Thus the sole role of dsRNA in RNAi is to direct the silencing machinery to a specific sequence.

• *Add-ons.* Sophisticated systems can be produced by the step-wise

addition of one recruiting (or simple binding) reaction to another, each addition making the system work more efficiently. The lambda switch might have evolved in this fashion, a notion we infer from the fact that certain of its features can be singly eliminated by mutation without destroying its over-all function. For example, elimination of the positive feedback loop (auto-activation of the *cI* gene), or of a negative feedback loop (auto-repression of the *cI* gene), creates a phage that can lysogenize, and these lysogens can be induced, but the processes are less efficient than those obtained with wild type phage. And there is a layer of cooperativity of repressor binding in addition to that shown in Figure 1 — the four depicted repressor monomers interact with four more repressor monomers bound some 2000 base pairs away. This high degree of cooperativity contributes to the 'switch-like' (all-or-none) character of induction. Thus three features — auto-activation and repression, and high cooperativity — are 'add-ons' that make a system that works, work better.

We find 'add-ons' working at a 'silent' region of a yeast chromosome. Here dsRNA, continuously produced from the silenced region, works together with a set of specific DNA binding protein recruiters. Together they ensure continual recruitment of the proteins required for silencing. Eliminating any one of the recruiting elements renders the silencing less stable [5].

• *A shorthand.* For convenience we say 'repressor turns off *Cro*'. This is of course a shorthanded way of describing what actually happens: repressor and polymerase compete for binding to DNA. The higher the repressor concentration the more frequently it will occupy its binding sites and the more completely it will exclude binding of RNA polymerase to the promoter of the *cro* gene. We often use shorthand — 'turn on, turn off', 'bound, free' — when describing binding reactions, but we are talking about matters of degree and graded effects. In eukaryotes specific DNA binding repressors work by recruiting repressing complexes and these repressing complexes work in graded opposition to the effects of transcriptional activators [6]. As noted in the preceding section, the imposition of ever more cooperativity can convert graded effects to something resembling 'on-off' switches.

An illustrative signal transduction pathway

Every step of the eukaryotic signal transduction pathway outlined in Figure 2, from signal to gene product, involves a binding reaction. The figure is a composite of steps found in different pathways (see figure legend) [7]. Two kinds of enzymes appear here. First, of course, there are the enzymes (enzymatic 'machines') that transcribe genes, splice RNAs, and translate mRNAs. Second, we encounter enzymes whose only role in signal transduction, so far as I know, is to make or break binding sites. These enzymes include kinases and guanine nucleotide exchange factors (GEFs). All of these enzymes have multiple possible substrates, and specificity — which substrate is chosen by each enzyme — is determined by binding/recruiting reactions.

Kinases

Three familiar kinds of kinases are involved in the pathway: two of these add phosphate to proteins, (to tyrosine residues, and to serine/threonine residues), and the third adds phosphate to a membrane lipid constituent called phosphatidyl inositol bis-phosphate (PIP_2). In each case the resulting modification presents a binding site recognized by one or another of the array of recognition modules found in eukaryotes [8,9]. For example, SH2 domains bind phospho-tyrosine; 14-3-3 domains bind phospho-serine/threonine; and certain PH domains bind the modified membrane component PIP_3. In some cases phosphorylation, rather than creating a binding site directly, does so indirectly. That is, phosphorylation can cause a conformational change in a kinase that increases its activity. But because kinases, ultimately, make or break binding sites, this consideration adds a wrinkle to, but not a reformulation of, our general rule.

As mentioned above, kinase specificity — which protein is picked to be phosphorylated — is itself often determined by binding reactions. The active sites of tyrosine kinases, for example, are notoriously similar, and specificity depends on binding reactions not involving the enzymes' active sites. In some cases recruiters ('adaptors') are used; in others, residues on the kinase, separate from the active site, direct binding to specific targets; in still others, we have 'scaffolds' that bind one or more kinase molecules plus a specific substrate; and so on. In some cases, (not represented here), addition of phosphate can block a binding reaction that would otherwise occur. Apposition of kinases with proper targets is often intimately associated with an increase in enzymatic activity, a matter I return to below.

GEFs

SOS is an enzyme (a GEF) that encourages the small GTP-binding protein Ras to exchange a GDP for a GTP and thereby assume an 'active' conformation. In that conformation it exposes a binding site for another protein. There are many different small G proteins, often present as domains attached to other protein domains and/or to a membrane. They form nearly identical overall structures when active, each displaying a different binding site. Because Ras, the small G protein in our example, is tethered to a membrane, exposure of its binding site fosters recruitment of its target (s) to the membrane. We will see that SOS itself is recruited to Ras by a binding reaction.

The steps of the pathway

The following points are detailed more fully in the Figure 2 legend. At certain points, recruited enzymes create binding sites: for example, kinases create protein–protein binding sites in (A) and (F); another kind of protein–protein interaction site is created by SOS in (B); another kinase creates a membrane-protein binding site in (D). Transcriptional activating proteins bind cooperatively to DNA in (H). In a two step recruiting process, an inhibitor is first removed from DNA, and then the transcriptional machinery is brought to the gene (I). Recruitment of the RNA splicing machinery to RNA is illustrated in (J); and, finally (K), the mRNA is translated into protein. The last step, incidentally, is the archetypical example of binding reactions determining specificity — in this case the loaded tRNAs bind specific triplets in the mRNA, and thereby present a specific array of amino acids for the translational machinery to work on.

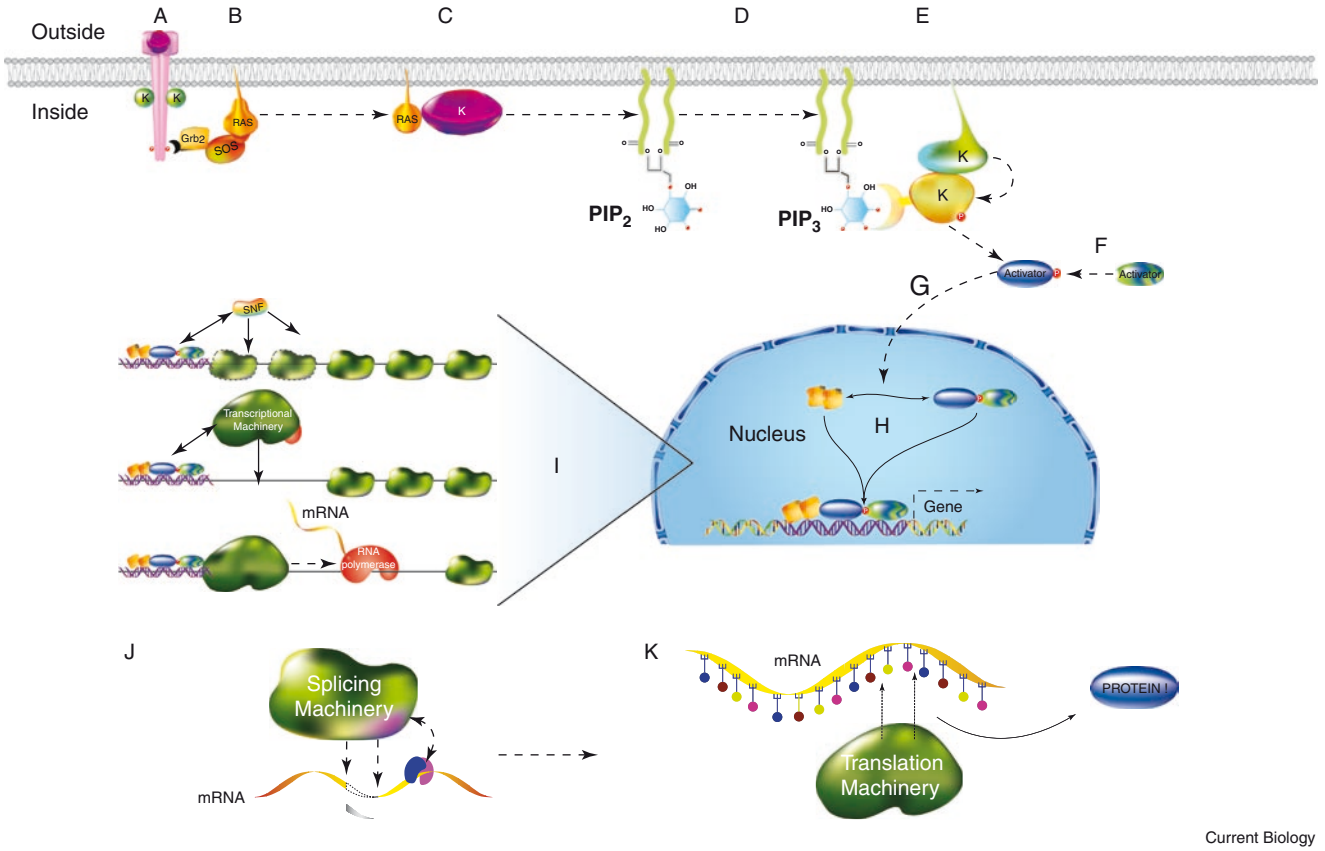

Figure 2. A representative eukaryotic signal transduction pathway.

(A). The two-chain receptor spans the plasma membrane. Its two chains have been brought together by binding of the protein ligand, a growth factor. By virtue of this apposition, each attached kinase adds a phosphate to a tyrosine residue on the partner chain (pink dots). (B)The adaptor protein Grb2, which bears an SH2 domain, simultaneously binds the modified tyrosine and SOS, thereby recruiting the latter to the membrane and thus to Ras. (C) Ras, activated by SOS, recruits to the membrane a PI 3-kinase (purple). (D) The recruited kinase coverts PIP_2 to PIP_3. (E) PIP_3 is bound by a PH domain attached to another kinase (yellow), positioning the latter so that it can be phoshorylated, and thereby activated, by the green kinase anchored in the membrane. (F) A further binding interaction, perhaps involving a scaffold, (not shown) promotes phosphorylation of a transcription factor, a modification that leads to its dimerization. (G) The dimer enters the nucleus (in a further binding reaction, not explicitly shown). (H) The protein dimer binds to DNA cooperatively with another protein dimer . (I) Activation of transcription is illustrated as a two-step process: the activator recruits a 'nucleosome remodeler' Swi/Snf (which removes nucleosomes from the promoter), and then it recruits the transcriptional machinery. The latter comprises multiple proteins, including RNA polymerase, and the third line depicts the polymerase transcribing the gene. A prior step might be involved: the activator might first recruit enzymes that facilitate recruitment of the nucleosome-remover by creating binding sites on the nucleosomes. (J) The blue/purple protein bound to a specific site on the RNA is a "splicing regulator" that recruits the splicing machinery to effect proper splicing . The mRNA is transported out of the nucleus (another binding reaction, not shown) and (K) translated into protein. Steps A–K are found in one or another growth factor pathway [7]; steps F–H are found in the Stat pathway [11]; and the steps in (I), represent a case in yeast [13].

Here are some further comments on this signal transduction pathway in light of the principles we deduced from the lambda case.

- *Inhibition of basal signaling.* At several steps there is the inevitable problem of spontaneous binding and a low level of constitutive (basal) activation of the pathway. And so, as we had anticipated, various inhibitors suppress this basal signaling: some help keep the receptor chains apart in the absence of the ligand; others bind the receptor and cover the tyrosines in the chains that are subject to phosphorylation; still others block spontaneous dimerization of the transcriptional activators, and so on [10,11]. These inhibitors work with affinities such that their effects are overcome in the presence of the signal.

The basal activities of kinases — which otherwise would tend to create binding sites spontaneously — are inhibited in various ways. As we have noted, phosphorylation of kinases is often required for their full enzymatic activity. That modification, under ordinary conditions, and absent a specific kinase-to kinase binding interaction, should occur spontaneously only rarely. Kinases often bear self-inhibitory domains whose effects are overcome by the presence of the proper substrate. For example, the tyrosine kinase Src (not shown in Figure 2) bears an SH2 domain that, by binding a phosphorylated tyrosine on another part of the kinase, holds the enzyme in an inactive state. And because the SH2 domain is buried, binding of the kinase to most phosphorylated targets is inhibited. But binding of the SH2 (probably working cooperatively with other binding domains — see the following section) to its preferred target apposes the enzyme with the proper substrate and relieves internal inhibition. Thus the kinase is active

only when it should be. The cyclin-dependent kinases (Cdks) are also controlled so as to be active only when called upon to work: in this case each cyclin, which recruits a cdk to a specific set of targets, simultaneously binds to and induces a conformational change in the kinase that greatly increases its activity. Certain other kinases, working in tandem, respond synergistically to activating signals, and this feature would also tend to depress effects of basal level signaling [12].

The effects of kinases and GEFs tend to be erased by enzymes that catalyze the opposing reactions — phosphatases and GTPase-activating proteins (GAPs). These enzymes, sometimes working as recruited functions and sometimes as background functions, thus counter basal level production of binding sites that might otherwise occur. The importance of these competing reactions is illustrated by the fact that signal transduction pathways are spontaneously activated when cells are treated with a general phosphatase inhibitor.

Basal level transcription in eukaryotes is suppressed by the wrapping of DNA into nucleosomes. As depicted in Figure 2, one of the earliest effects of a DNA-bound transcriptional activator in yeast is to recruit an enzyme that removes nucleosomes from the promoter region, thus clearing the way for recruitment, by the activator, of the transcription complex. This separate nucleosome-removal step is another 'add-on' — in its absence (in a mutant cell lacking the nucleosome-removing enzyme), recruitment of the transcriptional machinery suffices for activation, but the time required for full induction is longer. Evidently, in the absence of the nucleosome-removing step, the recruited machinery eventually out-competes nucleosomes for binding to the promoter . The delay caused by the absence of the nucleosome-removing enzyme is strikingly long, at least ten-fold (five hours versus 30 minutes for the wild type, roughly) [13], a matter I refer to in discussing artificial induction of stem cells near the end of this article.

- *Specificity and cooperativity.* In the lambda example, we noted the specificity problem faced by DNA-binding domains — they must distinguish not only between specific and non-specific sites, but also between stronger and weaker binding sites in DNA, a problem solved at least in part by cooperativity. We also noted that other kinds of binding reactions must solve similar problems. For example, like members of any given class of DNA-binding domains, SH2 domains all form similar overall structures, and yet they must distinguish one target from another — in this case one phosphorylated tyrosine from another. To some extent this selectivity is imposed by preferences for one or another of the few residues flanking the modified tyrosine residue. The preferences are weak, however, and it is likely that these domains usually find their proper targets using cooperativity. That is, one SH2 can work in conjunction with another SH2 on the same protein or in another associated component; or the SH2 domain might work together with a different protein-binding motif (an SH3 domain, for example); or with one that binds a membrane; and so on [9,14]. The typical representation of a signal transduction pathway (such as in Fig 2) ignores this likely complexity.
- *By-pass effects.* As suggested by our previous discussion of by-pass experiments, several steps in our pathway can be manipulated so as to eliminate the role of one or another recruiter that is ordinarily required. In each of these cases transcription of the target gene is elicited in the absence of the signal. Thus, for example, overproducing the receptor chain leads to spontaneous dimerization and triggering of the pathway; introducing a protein bearing SOS′ (a SOS derivative lacking an inhibitory domain) attached to a peptide sequence that anchors it in the membrane has a similar effect, as does overproducing SOS′; and, as already noted, a fusion protein bearing the appropriate DNA binding domain, attached to a component of the transcriptional machinery, activates transcription of the target gene.
- *The non-epigenetic nature of the pathway.* Our signal transduction pathway, once activated, is not self-perpetuating. Once the supply of ligand is exhausted the pathway shuts down as (among other factors) the binding sites created by the enzymes are erased. To make the signal self-perpetuating we need add positive feedback. One way to do this, if the activated gene encodes a transcriptional activator, would be to insert the DNA binding site for that activator near the gene. Another would be for the signal to cause transcription of a gene the product of which, in one way or another, leads to overproduction of the receptor chain and its concomitant spontaneous dimerization. And so on. These imagined modifications show, once again, that epigenetic (self-perpetuating) effects are properties of systems, not of individual components.

Evolving new specificities and complexities

Natural selection can readily survey, and select among, a wide array of signaling pathways with different specificities. Determinants of binding reactions are readily 'swappable' and have expanded into families of related binding domains. There are minimal stereo-specificity requirements for recruitment— a DNA-binding domain typically can be attached at either end of, or within, a transcriptional regulator; an SH2 domain can similarly be placed at any of several places in a protein, and so on, all without loss of the designated function. The evolutionary path often emphasized in 'evo-devo' discussions invokes changes in '*cis*-regulatory' sites in DNA. Thus, for example, the DNA-binding sites used in the example of Figure 2 could be put in front of any gene and, because most (perhaps all) transcriptional activators work on most genes, in that new configuration the new gene will be brought under control of the ligand used in the illustration. But the binding steps in the pathway provide many places where specificities — which genes are activated by which signals — are readily swapped and expanded. For example, any kinase bearing the recognition domains found on the yellow kinase in Figure 2 (including a PH domain, which, like the other recognition modules we have discussed, is readily attached to proteins) will be recruited to the membrane by the modification shown in Figure 2. And so on.

These systems lend themselves to constant improvement in the form of add-ons. Various inhibitory and facilitating binding reactions can be imposed on the system step-wise, each making the system work just a

bit better. And where those add-ons must themselves be controlled we need new binding reactions — inhibitors of inhibitors and facilitators, and so on. A seemingly endless regress of binding reactions.

What can go wrong

Cancer

Cancer is said to be a disease of aberrant signaling. I have emphasized how binding/recruiting reactions can lie at the heart of signaling pathways. Indeed, as a tour of a modern cancer text [7] reveals, most 'cancer-causing' mutations involve components of binding reactions. These comprise a rather restricted set: recruiters, inhibitors, inhibitors of inhibitors, receptors, and enzymes that make/break binding sites. (Obvious exceptions include mutations that affect DNA repair processes, but these presumably cause mutations in genes such as those emphasized here; and mutations of certain receptors, such as Notch, that are activated by a process more elaborate than simple apposition of receptor chains). Members of the various categories listed here, when not deleted, are found to bear point mutations, to be overexpressed, and /or to be fused to heterologous domains. Here are a few well-known examples — many more can easily be found:

- *Recruiters:* transcription factors (P53, myc, E2F, each controlling expression of sets of genes); E3 ligases (FBW7, an F-box protein that targets Myc and other transcripton factors; MDM2 which targets P53); Ras, in a mutant form that is constitutively 'active'; and cyclins.
- *Receptors:* Her2; EGF receptor; Ret
- *Inhibitors:* Rb, which binds and blocks the activating function of E2F; P27, which binds and inhibits a Cdk–cyclin complex; Ids, which bind and prevent dimerization of a class of transcription factors.
- *Inhibitors of inhibitors:* Arf, which blocks binding of MDM2 to p53;
- *Enzymes that make or break binding sites:* kinases, sometimes associated with receptors (Kit), and sometimes not (Src); phosphatases (PTEN), and GEFs (Sos).

Binding reactions, and hence signaling pathways, are easily subverted, as shown by the behavior of certain viral proteins: the Large T antigen of SV40 virus binds Rb and thereby blocks Rb binding to E2F, and similarly binds and sequesters p53; EBNA-6, encoded by Epstein-Barr virus, binds and carries to the nucleus a protein that, in turn, binds Rb and thereby prevents its interaction with E2F [15]; the E6 protein of human papilloma virus recruits an E3 ligase to P53, thereby causing its destruction. And so on.

The identification of these and other 'cancer causing' genes has, reasonably, encouraged the notion that analyses of human cancer genomes would reveal small, discrete sets of mutant genes causing, or at least strongly associated with, specific cancers. But, with a few notable exceptions, these kinds of genomic analyses, admittedly at an early stage, have run into frustrating problems. First, in general, many mutations, with small effects, contribute to tumor formation. Second, most tumors (especially sold tumors) defy classification by sequence analysis. In the typical example, breast cancer say, a small proportion of cases (~10%) is strongly associated with a common inherited defect (in this case a mutant *BRCA* gene), but most do not fit this, nor any other, obviously coherent pattern [16]. Might our depiction of a signal transduction pathway, as a series of binding reactions and reactions that create/break binding sites, give any hints as to a possible explanation for these findings?

Consider the finding of so many mutations with inferred small effects [16]. Our signaling pathway, just as the lambda switch, includes various add-ons that make the pathway work better, but are not essential — these include, for example, the various inhibitors, alluded to above, that dampen spontaneous signaling. Wouldn't one expect that these systems would work with ever-decreasing efficiencies as such add-ons were lost? And how readily would mutations causing such changes be recognized? How would one recognize, for example, the single amino acid change in a target protein that decreased its affinity for an E3, and thereby caused the protein to be present at a concentration a few-fold higher than the optimal level? At some point the accumulation of such effects could be disastrous.

Many cancer mutations are 'by-passers', changes that obviate signaling and change specificities. For example, a kinase (Abl), fused to another protein, has unfortunate effects that vary depending on just which recruiting domains have been included in the fusion; a receptor chain (Ret) fused to a dimerization domain, spontaneously dimerizes and triggers the pathway; a kinase (PI3 kinase) bearing a mutation that increases its affinity for the membrane by-passes a signal in a different way. And so on.

Many mutations result in the over-expression of one or another component of signaling pathways. One effect of such overproduction, even of a wild-type protein, would be to allow by-pass of an ordinarily-required signal. For example, an over-produced transcription factor will tend to bind its sites on DNA absent its usual partners (with which it ordinarily binds cooperatively), and even to bind sites it ordinarily never sees. Similar considerations would apply to all the binding partners that use cooperativity in finding their partners. And, to a degree that will depend on the strengths of the relevant binding sites, and on the degree of over production, any overproduced recruiter — transcription factor, E3, and so on — would be expected to cause non-specific inhibition of the targeted enzyme (squelching), and at higher concentrations, to block its own action (self-squelching), as discussed above. Wild-type cells have feedback mechanisms that discourage continuing overproduction of a protein — the lambda governor is an example, as is the action of E3 ligases. The accumulated loss of such feedbacks can render otherwise harmless changes dangerous.

The kinases present a particularly worrisome problem. The similarities in the active sites of tyrosine kinases mentioned above explain why it has been so hard to find specific tyrosine kinase inhibitors. (Gleevac, the drug that with some selectivity inhibits the bcr-abl kinase, evidently does so by trapping the enzyme in an inactive conformation, a conformation evidently not adopted by most other kinases.) Kinases depend, for specificity, upon binding reactions involving residues not associated with their active sites. Src, discussed above, is a telling example. Mutant Src was one of the earliest cancer genes discovered — and it was found

to be over-expressed in a mutant form that had lost the self-inhibitory feature described above. Such a berserk kinase would be expected to have non-specific effects, creating binding sites where it shouldn't.

We noted above that cancer is a disease of aberrant signaling disease. But it is also the result of an experiment in evolution — the cancer cell is selected to grow absent certain signals and restraints. Systems in which specificity is determined in large part by the kinds of binding reactions discussed here lend themselves not only to the unwanted activation or inhibition of this or that signaling pathway, but also to the production of 'new' pathways that allow for abnormal growth of cells. There may be many ways to interchange parts to effect any specified end.

Other diseases and processes

Perhaps it is also worth keeping these considerations in mind when analysing certain human diseases other than cancer. The genetic bases of autism and schizophrenia, for example, have so far proved difficult to pin down, with many different mutations in different combinations evidently contributing to the outcomes [17]. Perhaps matters will clarify and just a few specific pathways will turn out to be affected, and in ways we can understand. An unfortunate alternative would be that there are many different ways to elicit similar phenotypes, perhaps by the accumulation of mutations with small effects. At the opposite end of the spectrum we have certain other neurological diseases each of which is caused primarily by a defect in a specific gene Each of these cases involves a single binding protein, and in each case elimination of the protein has, evidently, an effect similar to even a two-fold over-expression of that protein: Rett syndrome (the protein is MeCP2, which binds methylated sites on DNA); fragile X syndrome (the protein binds RNA); and Angleman syndrome (the protein is a transcription factor). [18]. Might one or another of the dosage dependent effects we have discussed be relevant here?

Concerns about levels of transcription factors — specific DNA binding recruiters, in our terminology — arise in experiments in which differentiated cells are induced to form cells with stem-cell-like qualities. As originally described, this required the introduction of four such recruiters into differentiated cells, and more recent experiments indicate that too high a level of a recruiter can be as detrimental as its absence [19]. And the conversions are maddeningly slow, taking many days. But perhaps this too should not be surprising — recall our example [13] showing that simple induction of a single gene can be drastically delayed by the absence of a single factor. It is not hard to imagine that the differentiated cells respond slowly to the introduced recruiters because they are lacking one or more co-factors with which these recruiters ordinarily work more quickly.

Conclusion

Recruiting reactions have been used by natural selection to produce a wide array of complex biological processes. Just as Darwin required, these processes can be diversified and improved upon by step-wise modifications. The very simplicity of the nature of the underlying interactions accounts for much of the complexity we find in cells: rather than neat Ferrari-like engines that are switched on and off, we have binding reactions between macromolecules that must be encouraged to proceed in response to signals, and prevented from occurring spontaneously. These systems require, at least, cooperativity, control of concentrations, and inhibitory effects that come in many guises. As the lambda example shows, these requirements can be met, and a sophisticated epigenetic switch produced, with binding reactions involving just two regulatory proteins and RNA polymerase. With the appearance of enzymes that make/break binding sites — including kinases, phosphatases, GEFs, ubiqyuitilating enzymes — the possible scenarios are vastly multiplied. For a recent probing of how a series of binding reactions control, in eukaryotes, the DNA damage response, and ultimately progression through the cell cycle, see ref [20].

Natural selection was not restricted to considering, and tinkering with, the messy world of recruiting reactions. Intermediary metabolism, for example, is run by enzymatic machines that are revved up and down by allosteric responses to the binding of small molecules. And in bacteria we know of one set of genes that is not regulated by recruitment: the inactive promoters bear a special form of tightly-bound RNA polymerase, and the activator uses energy in the form of ATP to turn on transcription. In this system the basal level of transcription is vanishingly low, and so no repressor is required (or found). But as we encounter ever more complex organisms (and leave intermediary metabolism essentially unchanged) we find increasing roles played by the kinds of binding reactions discussed here. According to the following line of argument, this should not be surprising.

In *The Origin of Species* Darwin was, paradoxical as it might at first seem, looking at the simplest task evolution undertook — the elaboration of plants and animals. Unlike the evolution of bacteria that grow in disparate environments, the 'recent' evolution of these complex organisms required few new enzymatic activities — we have essentially the same set of such as do flies and other animals and plants. Its as though once evolution had produced the enzymes found in eukaryotic cells, including those that make/break binding sites, it was easy to quickly deploy these enzymes, using recruiting reactions — specificity determinants — to different ends.

Development of higher organism is made possible by elaborate programs of intercellular signaling , and the signals are usually in the form of proteins or other macromolecules. The reiterated use of binding reactions to give meaning to these signals, as we have seen, comes with unavoidable dangers. Things can go awry in many ways, and, unfortunately, it can be hard to decipher what has gone wrong in any given case, and even harder to fix it. It would be easier if we had been intelligently designed and were made of neat machines. Like Ferrari engines.

Acknowledgements

I thank for helpful discussion Robert Benezra, Helen Blau , Nicolas Buchler, Jerry Coyne, Alex Gann, Shiv Grewal, Alan Hall, Oliver Hobert, Ross Levine, Danesh Moazed, Santosh Narayan, Ken Offit, Tony Pawson, John Petrini, Bruce Stillman, Vidya Prabhu, Bert Vogelstein, and Andrew L. Wolfe.

References;

1. Ptashne, M. (2004). A Genetic Switch (Third Edition): Phage Lambda Revisited. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press.
2. Ptashne, M. and Gann, A. (2002). Genes and Signals. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press.
3. Ptashne, M. (2007). On the use of the word 'epigenetic'. Curr. Biol. *17*, R233–236.
4. Buhler, M., Verdel, A., and Moazed, D. (2006). Tethering RITS to a nascent transcript initiates RNAi- and heterochromatin-dependent gene silencing. Cell *125*, 873–886.
5. Jia, S., Noma, K., and Grewal, S.I. (2004). RNAi-independent heterochromatin nucleation by the stress-activated ATF/CREB family proteins. Science *304*, 1971–1976.
6. Gertz, J., Siggia E.D., Cohen, B.A. (2009). Analysis of combinatorial cis-regulation in synthetic and genomic promoters. Nature *457*, 215–218.
7. Weinberg, R.A. (2006). Biology of Cancer, Oxford UK, Taylor & Francis Group Ltd.
8. Pawson, T., and Nash, P. (2003). Assembly of cell regulatory systems through interaction domains. Science *300*, 445–452.
9. Yaffe, M.B. (2002). Phosphotyrosine-binding domains in signal transduction. Nat. Rev. Mol. Cell Biol. *3*, 1771–86.
10. Massagué, J., Seoane, J., and Wotton, D. (2005). Smad transcription factors. Genes Dev. *23*, 2783–2810.
11. Mertens. C., and Darnell, J.E. Jr. (2007). JAK-STAT signaling. Cell *131*, 612.
12. Ferrell, J.E. Jr. (1997). How responses get more switch-like as you move down a protein kinase cascade. Trends Biochem. Sci. *8*, 288–289.
13. Bryant, G.O., Prabhu, V., Floer, M., Wang, X., Spagna, D., Schreiber, D., and Ptashne, M. (2008). Activator control of nucleosome occupancy in activation and repression of transcription. PLoS Biol. *12*, 2928–2939.
14. Jones, R.B., Gordus, A., Krall, J.A., and MacBeath, G. (2006). A quantitative protein interaction network for the ErbB receptors using protein microarrays. Nature *439*, 168–174.
15. Kashuba, E., Yurchenko, M., Yenamandra, S.P., Snopok, B., Isaguliants, M., Szekely, L.,and Klein, G. (2008). EBV-encoded EBNA-6 binds and targets MRS18-2 to the nucleus, resulting in the disruption of pRb-E2F1 complexes. Proc. Natl. Acad. Sci. USA *105*, 5489–5494.
16. Beerenwinkel, N., Antal, T., Dingli, D., Traulsen, A., Kinzler, K.W., Velculescu, V.E., Vogelstein, B., and Nowak, M.A. (2007). Genetic progression and the waiting time to cancer. PLoS Comput. Biol *11*, e225.
17. Burmeister, M., McInnis, M.G., and Zöllner, S. (2008). Psychiatric genetics: progress amid controversy. Nat. Rev. Genet. *7*, 527–540.
18. Ramocki, M.B., and Zoghbi, H.Y. (2008). Failure of neuronal homeostasis results in common neuropsychiatric phenotypes. Nature *455*, 912–918.
19. Eminli, S., Utikal, J., Arnold, K., Jaenisch, R., and Hochedlinger, K. (2008). Reprogramming of neural progenitor cells into induced pluripotent stem cells in the absence of exogenous Sox2 expression. Stem Cells *10*, 2467–2474.
20. Usui, T., Foster, S.S., and Petrini, JH. (2009). Maintenance of the DNA-damage checkpoint requires DNA-damage-induced mediator protein oligomerization. Mol. Cell *33*, 147–159.

Sloan Kettering Institute, Memorial Sloan Kettering Cancer Center, 1275 York Avenue, Box 595, New York, NY 10021, USA.
E-mail: m-ptashne@mskcc.org