The Type 1 growth factor receptors and their ligands considered as a complex system

W J Gullick

Department of Biosciences, University of Kent at Canterbury, Canterbury, Kent CT2 7NJ, UK; Email: W.J.Gullick@ukc.ac.uk

Abstract

The Type 1 family of growth factors and their receptors play an important role in normal development, wound healing and in diseases such as cancer. The products of the four receptor genes and the ten genes specifying ligands interact in a complex pattern. Such systems may develop emergent properties which cannot be predicted from a reductionist analysis of the interactions of individual components. New methods of determining these higher level properties are, however, being developed. These include microscopic analysis of live cells expressing fluorescently tagged ligands, receptors and second messengers which can provide positional information on components of the system. Computational simulations of the complex interactions are being developed which can help to predict the properties of the system.

Introduction

Most of the genes which can be expressed from the human genome have now been identified (Baltimore 2001, and other articles in this special ‘Human Genome’ issue of Nature vol. 409). These can, in many cases, be grouped into families based on structural similarities which may possess similar functions. In many cases, members of such groups have been shown to physically interact in cells as part of their function. One of the greatest challenges to molecular biology over the next decades is to determine the consequences of the interaction of multiple proteins within a system. Such complex systems develop emergent properties which are not necessarily predictable from an analysis of the interaction of individual components within that system (Bray 1998, Weng et al. 1999). It is essential, therefore, to develop methodologies that allow us to observe and determine the nature of these more complex interactions and their consequences. In this paper I shall briefly review the Type 1 growth factors, their tyrosine kinase receptors and their interaction with second messengers and consider the tools available or under development to explore their interactions and the emergent properties of this complex system.

Receptors

This family of molecules consists of four genes encoding receptors (epidermal growth factor (EGF) receptor, c-erbB-2/HER2, c-erbB-3/HER3 and c-erbB-4/HER4), and ten genes encoding ligands (EGF, transforming growth factor-α (TGFα), heparin binding (HB)-EGF, betacellulin, amphiregulin, epiregulin, neuregulin (NRG)-1, NRG2, NRG3 and NRG4) (Olayioye et al. 2000). Inactive receptors exist in the cell membrane in a monomeric state in the absence of ligand. As the concentration of ligand increases in the environment, an increasing proportion of receptors becomes occupied by ligand contingent on their specific affinity of interaction. This provokes the receptors to dimerise and then oligomerise in the cell membrane through a process which is, as yet, not very well understood. Dimerisation increases the catalytic activity of their intracellular tyrosine kinase domains, in this receptor family probably as a result of conformational changes (Hubbard & Till 2000). The receptors are themselves a substrate for phosphorylation and as result of this increase in enzyme activity they become phosphorylated on the hydroxyl groups of tyrosine side chains at several specific sites.

Each receptor gene can produce several different proteins due to alternative exon splicing. Some encode mRNAs which specify secreted, extracellular domains of various sizes with, as yet, essentially unknown functions. In the case of c-erbB-4, however, the receptor is expressed as four alternative full-length forms. Two of these have alternative short juxtamembrane extracellular sequences which are differentially sensitive to proteolysis (JMa and JMb) (Elenius et al. 1997).
et al. 1997). Which variant is produced appears to control the release of an extracellular sequence and the concomitant production of a fragment consisting of a short, truncated extracellular peptide, followed by a transmembrane sequence and an intracellular tyrosine kinase domain (Vecchi et al. 1998). Two further variants either possess or lack a 16 amino acid sequence in the intracellular domain (CTA and CTG), containing a sequence (YpXXM) which, when phosphorylated on the tyrosine residue, will recruit the intracellular second messenger phosphatidylinositol 3–kinase (PI3kinase) (Elenius et al. 1999). Expression of one or other of these forms will therefore probably affect the repertoire of intracellular pathways activated.

Individual receptor pairings can consist of two molecules of the same type, called homodimers, or two molecules of different types, which has been termed a heterodimer (i.e. the products of separate genes or, indeed, in the case of c-erbB-4, different splice variants) (Tzahar et al. 1997). Thus the total number of receptor proteins which can interact are therefore the three single receptor gene products (the EGF receptor, c-erbB-2 and c-erbB-3) and the four full-length splice variant encoded products of the c-erbB-4 receptor. Thus, if all these seven proteins are expressed simultaneously in the same cell, and they are capable of interacting freely in the membrane, they can form 28 possible homodimers and heterodimers (Fig. 1).

Ligands

The ten so-far identified genes which encode ligands for the Type 1 growth factor receptors each contain a structurally related element of about 50 amino acids which is sufficient to bind to and activate a receptor. These short regions contain six cysteine residues which are oxidised in the biologically active ligand to three disulphide bonds with the conserved arrangement (numbered from the N-terminus) of 1–3, 2–4, 5–6. The three-dimensional structures of EGF, TGFα and NRG1α have been determined by nuclear magnetic resonance spectroscopy and each has been shown to possess the same folded structure (Nagata et al. 1994, Jacobsen et al. 1996).

It should be noted, however, that this sequence and its fold are found very commonly in proteins which do not, as far as is known, interact with the receptors in this family. Indeed, at the Pfam web site at the Sanger Centre (http://www.sanger.ac.uk/Software/Pfam/) there are more than 3000 proteins described which contain an ‘EGF-like’ element. Thus the initial chance of a newly described protein which contains such a domain being a new ligand for this family of receptors (without any other supporting information) is at present 1 in 300. From an analysis of the evolutionary relationships of the ligand genes, however, Stein & Staros (2000) have suggested that a new EGF-like sequence may be tentatively identified as a potential ligand if there is an intron sequence in between the sequences coding for cysteine residues four and five, a characteristic shared by all known ligands and not generally found in non-ligand EGF elements.

Ligand-binding specificities have been determined using individual receptors expressed in cells or pairwise combinations of receptors or by using recombinantly expressed, purified receptor extracellular domains (Jones et al. 1999). Six of the ligands bind to the EGF receptor when expressed alone: EGF, amphiregulin, HB-EGF, TGFα, betacellulin and epiregulin. Three of these also bind to c-erbB-4 when expressed alone: HB-EGF, betacellulin and epiregulin. The products of the NRG1 and 2 genes bind to c-erbB-3 and c-erbB-4, albeit with somewhat different affinities, whereas the products so far described of the NRG3 and 4 genes appear to bind only to c-erbB-4. The situation is further complicated by the unexpected observation that EGF and betacellulin bind to c-erbB-2 and c-erbB-3 when expressed together (and thus presumably forming heterodimers) but not to either receptor when expressed alone (Fig. 1).

Ligands are also, in some cases, subject to differential splicing. Interestingly, this is confined to the more recently identified NRG genes and their products the neuregulins (also known as neu differentiation factors, heregulins, ARIA and glial growth factor). Only single mRNA forms have been described for the remaining six ligands which interact (directly) with the EGF receptor. Alternative sequences encoded by separate exons are found in the C-terminal region of the NRG EGF-like element affecting the region-contained cysteines 5 and 6 which are bonded to each other in the oxidised structure. These forms are called alpha and beta variants and the alternative sequences can affect the affinity of the ligand produced for different receptors and may therefore recruit different second receptors into heterodimers (Gullick & Srinivasan 1998). The detailed pattern of receptor recognition and the relative affinities of some of the NRG splice variants have been described experimentally, particularly in an elegant series of papers by Yarden’s group (Wen et al. 1994).

In addition to splicing of the C-terminal region of the NRG1 and 2 (and possible 3 and 4) genes each ligand possesses a more or less complex structure either side of the EGF-like sequence. All ligand genes can encode a transmembrane region and (with the exception of some NRG splice variants) are made as transmembrane proteins. These may be subject to cleavage to produce soluble forms by the same or similar proteolytic enzymes, which are responsible for cleavage of the sensitive isoform of c-erbB-4. This cleavage is regulated in cells and can be activated by phorbol esters and inhibited by metalloproteinase inhibitors, such as batimastat. Further complexity resides in whether or not the precursors contain additional protein folds (for instance, eight further EGF-like repeats are present in the EGF precursor and an immunoglobulin fold is present in the NRG
Figure 1: Diagramatic representation of the Type 1 growth factors and their receptors and their published interactions. Lilac lines indicate specific interactions. TGF, TGFα; AR, amphiregulin; BTC, betacellulin; EPI, epiregulin; JMa/JMb/CTa/CTb, splice variants of C-erbB-4. The numbers under the dimers indicate receptor pairs.

gene transcripts) and their differences in post-translational processing (the NRG1 products are known to be modified by O-linked glycosylation).

A single ligand therefore binds selectively to a subset of these receptors provoking a particular pattern of receptor dimerisation. Other ligands bind to different subsets of receptors and induce other dimer combinations. In the presence of multiple ligands there is therefore a complex interplay between ligand binding and the pattern of dimers present. As discussed below, however, although all the possible homodimeric and heterodimeric combinations of receptors and their interacting ligands have been described, there are several limiting factors which impinge upon the system and which may control the selection of these possible combinations.

Second messengers

Within the cell the phosphorylated receptors now can interact with proteins called second messengers. Again, a complex array of these proteins has been discovered, containing various structures which allow them to recognise receptors phosphorylated at specific sites. These fall into two main classes, those that possess structures called SH2 domains, and a smaller group of proteins which interact with phosphorylated receptors via a structure called a PTB domain (Pawson 1995). A number of studies have examined and catalogued the ability of various isolated receptors to bind selectively to particular second messengers in vitro and to a lesser extent in vivo. There are, however, several important issues which have yet to be adequately addressed, such as whether one receptor can bind to more than one second messenger at once and whether one second messenger can bind more than one receptor at one time (by multiple SH2 domains or by combinations of SH2 with PTB domains). It is not known whether the recruitment of second messengers occurs to receptor dimers or if receptor clustering is required or indeed the sequence and detailed timing of the second messengers which are recruited.
Position is important

I have described above some of the molecules involved in this system and their possible interactions. Whether or not these take place depends on two additional factors, the amount of each protein present and its position in the cell. The control of receptor and ligand expression occurs at the level of transcription and is influenced by a variety of factors including the activation of signalling by EGF family ligands themselves via Type 1 receptors as well as extrinsic factors such as steroid hormones which suppress transcription of the c-erbB-2 and c-erbB-3 genes. The level of receptor protein in a cell is also affected by a process called downregulation, which is activated by ligand binding and receptor phosphorylation and is effectuated by receptor internalisation and by ubiquitination which targets modified receptors for degradation by lysosomes (Waterman & Yarden 2001).

An important emerging concept is that ligands and receptors may be directed to, or sequestered in, different parts of the cell membrane which will affect their ability to interact. One example in human tissues is in polarised epithelia such as the enterocytes of the colon where the EGF receptor is expressed only on the basolateral membranes of the cells. How this localisation is controlled is not yet entirely clear but the equivalent protein in the worm C. elegans, called Let23, has been shown to interact with a protein called Lin7 which restricts its localisation (Simske et al. 1996). Some of the ligands and receptors contain sequences at their C-termini which allow them to interact with structures present in other proteins called PDZ domains. The TGFα protein has recently been shown to bind to a protein called p59/GRASP55 via a PDZ domain (Kuo et al. 2000) and the c-erbB-2 protein to interact with a PDZ domain protein called ERBIN (Borg et al. 2000). c-erbB-4 also contains a C-terminal sequence which interacts with a PDZ domain containing protein called PSD 95 which is present in the post-synaptic membrane of neuromuscular junctions (García et al. 2000).

The Type 1 receptors (and possibly their ligands) may not be randomly distributed in cell membranes due to the lipid environment. Several reports have suggested that these sorts of signalling molecules may be concentrated in structures described as caveolae or lipid rafts (Carpenter 2000). Such local concentrations of components of the system may radically affect responses to ligands both in terms of concentration/response profiles and in speed and duration of responses (Bray et al. 1998).

New methods of exploring complex interacting protein systems

It is useful to those trying to understand this system in more detail to seek to represent it in models. In a rather simple model, the system can be viewed as existing in three layers. The first layer, which is extracellular, represents the pattern of relative concentrations of ligands and their bioavailability. The second layer represents the receptors embedded in the cell membrane which can be provoked to dimerise in different patterns depending upon the information contained in the first layer. The third layer represents the intracellular proteins or second messengers proteins, which can bind to particular activated receptors at specific sites and convey intracellular signals affecting properties such as growth and cell shape. Clearly, however, such a description represents an over-simplification of the whole system since, as described above, there are many other points at which additional complexities exist. For instance, not all ligands are available to a cell in a soluble form; most are indeed first expressed bound to the cell surface, for instance by a transmembrane sequence. The release of the ligands is via a proteolytic enzyme whose activity is also under control. Some ligands are then active in solution while others bind with high affinity to proteoglycans present on the surface of cells, where they may be available for receptor binding or may exist in a form of ligand reservoir. Other levels of complexity include the phosphorylation of receptors at other positions, notably, serine and threonine residues by other kinases. Such modifications can affect the rate of enzyme activity of the receptor and its affinity for its ligand. Nonetheless, several models have been published which consider the ligands, receptors and second messengers as a form of closed system. With the recent success in sequencing the human genome it appears that there are only 90 tyrosine kinases of which 58 are of the receptor type and that there are no candidate sequences which could represent additional Type 1 receptors (Robinson et al. 2000) although we may not have seen all the ligands as yet.

This system can be considered as resembling a neural network (Hopfield 1982, Schamel & Dick 1996) which contains an input layer, a hidden layer and an output layer. In this model, the input represents the ligands, the hidden layer the receptors and the output the activity or location of second messenger proteins. The strengths of interaction are determined at the input level by ligand affinity and at the output level by variable phosphorylation levels and the use of different phosphorylation sites. Emergent properties of such a network arise as a consequence of varying the interaction strengths or information traffic between nodes. In this system, an example would be a cell expressing equal amounts of each receptor which are not segregated in compartments and can therefore interact. Addition of a ligand such as EGF will induce clustering of the EGF receptor and the formation of a group of homo- and heterodimeric receptor combinations. As a consequence, some receptors will be internalised and lost to the system. The intracellular signals generated will, however, over time induce expression of the...
A GFP-tagged EGF receptor cDNA was microinjected into NR6 mouse fibroblasts (which lack EGF receptor expression). (A) Cell prior to addition of EGF and (B) same cell 5 min after addition of 500 nM EGF to the medium.

EGF receptor itself, possibly other receptor gene products and several of the ligands. Thus, if a second stimulus is given to the cell it will be received and interpreted by a different composition of receptors in a different ligand environment. Moreover, the ability of receptors to interact and to recruit second messengers may be altered. This represents an emergent property, in effect, a learned response.

Such complex interactions are difficult to address by standard experimental approaches which tend to over-simplify by considering only subsets of the molecules involved. Experimental design has also tended to vary only single parameters. Such parallel information processing is, however, amenable to computational simulation. The problem with such models is that they are difficult to simulate unless there are known strengths of interactions between components which can be quantified numerically. At present, there are few methods which can produce such numbers. An example would be to determine the affinity of interaction of two receptor types in a cell membrane in response to addition of a ligand. Indeed, conventional interaction constants are of limited value in a system constrained in two dimensions. As an example of such an approach, with two colleagues, Colin Johnson and Jacki Goldman, we have developed a prototype computational simulation of growth factor receptor clustering (http://www.cs.ukc.ac.uk/people/staff/cgj/research/receptors.html). Our intention is to use this model and more complex simulations to search for solutions of interaction constants which are consistent with observed behaviour. To this end, and as described below, we have been employing green fluorescent protein (GFP)-labelled growth factors, receptors and second messengers to obtain experimental data on receptor clustering (Figs 2 and 3). Others have developed computational simulations of intracellular signalling which are more dependent on known interaction strengths (http://bms-mudshark.brookes.ac.uk/frances/fabweb5.htm). No doubt combinations of such approaches based on more elaborate models will be employed in the future. Such simulations may
have some use in searching for sensitive aspects of the system which may be vulnerable to inhibitors, with applications in cancer and in chronic diseases. There is an evident need, however, to provide more data on the system at the level of multiple protein interactions.

Digital microscopy as a means to analyse complex interacting protein systems

In the last few years there has been an explosion in the availability of new microscopy techniques and physical reporters which may prove very useful for monitoring complex protein interactions. The discovery of GFP and the development of optimised derivatives with different spectral properties and the more recent use of red fluorescent protein from the IndoPacific sea anemone relative Discosoma (see http://www.clontech.com/archive/OCT99UPD/RFP.html) has allowed the tagging of multiple protein types in live cells. Several second messenger proteins such as PI3 kinase (Gillham et al. 1999 and see http://www.bio.ukc.ac.uk/gullick.html) and SHC (Sato et al. 2000) have been labelled and have been shown using low light, digital fluorescence microscopy to relocate to the cell membrane in response to growth factor receptor activation. Growth factor receptors themselves have also been tagged with GFP, including the EGF receptor (Fig. 2). Carter & Sorkin (1998) have shown that such a modification does not appear to affect signalling and experiments by Kim’s group have shown that reintroduction of GFP-tagged Let23 into C. elegans in which the normal alleles were inactivated fully complemented the mutant phenotype (Simske et al. 1996), suggesting that the normal functions of the receptor are not compromised. A note of caution, however, is derived from work by Yancopoulos and colleagues who showed that mutations to the C-terminus of members of the Eph receptor family receptor prevented their interaction with PDZ domain-containing protein(s) and affected their membrane localisation and ability to cluster (Torres et al. 1998). As noted above, the c-erbB-2 receptor is thought to interact with ERBIN via such a C-terminal motif and c-erbB4 with PSD-95. Thus care must be taken in interpreting such experiments so that the tag does not affect the subcellular

Figure 3 A GFP-tagged cDNA clone of the p85 subunit of PI3 kinase-α was microinjected into the breast cancer cell line MCF-7. (A) Cell prior to the addition of insulin and (B) same cell 5 min after the addition of 500nM insulin to the medium.

Gullick: Type 1 growth factor receptors
location of these molecules, even if it does not evidently affect the physical properties of the receptors in assays on solubilised material.

Additional techniques are available or under development which employ fluorescently tagged proteins which can determine the proximity of differently tagged protein pairs. These include fluorescent resonance energy transfer (FRET) (Pollok & Heim 1999) and fluorescent lifetime imaging microscopy (Bastiaens & Squire 1999). Recently, using the FRET technique, GFP-tagged EGF receptor has been shown to interact with the Grb2 protein fused to yellow fluorescent protein (Sorkin et al. 2000). Finally, there have been considerable developments in atomic force microscopy which does not require fluorescent protein tagging and which may allow the description of protein complexes at a higher resolution than light microscopy (Engle et al. 1999).

In summary, it is becoming evident that cells are made up of complex sets of interacting proteins which perform specialised functions, such as regulation of gene expression, intracellular signalling and metabolic processing. In the same way that proteins have been found to be made up from a selection of different members of a limited number of folds, it may be that these interacting systems have a limited number of logical designs. One such system which is important in development and disease is the Type 1 growth factors and their receptors. New technologies are under development; these will allow us to analyse such systems in which many components are changing their position, level of expression and activity simultaneously. A combination of these approaches to gain experimental data and computational simulations to facilitate analysis of these data and to test predictions and search for otherwise unobtainable interaction strengths may help us to understand more fully how these systems operate and how they adapt to events and acquire new properties.

Acknowledgements
I am grateful to Helen Sheldon (née Gillham) for the use of Figs 2 and 3. I am also grateful to Jacki Goldman and Colin Johnson for permission to refer to the computer simulation website.

References
Guillick WJ & Srivivasan R 1998 The Type 1 growth factor receptor family: new ligands and receptors and their role in breast cancer. Breast Cancer Research and Treatment 52 43–53.
undergoes translocation and mediates up-regulation of the tyrosine kinase c-src in EGF-stimulated A431 cells. *Genes to Cells* **5** 749–764.


