

First INFOBIOMED Training Challenge,
 Viladrau, Barcelona, Spain
 12-16 September 2005



Information Society
 Technologies

REPORT on CASE STUDY 2:
**Targeting EGFR Signal
 Transduction Pathway by
 Anticancer Drugs**

Ignasi Belda ibelda@pcb.ub.es

Marilena Garefalaki marilena@aia.csd.auth.gr

Mark McAuley M.T.McAuley@newcastle.ac.uk

Eva van Soest e.vansoest@erasmusmc.nl

Karin Pike-Overzet k.pike-overzet@erasmusmc.nl

Contents

Introduction	3
Case study	4
The participants	5
Why targeting EGFR signalling pathway?	7
Initial contributions	10
Process of finding the target	11
Ligands	11
Receptor(-ligand binding domain)	12
Intracellular cascade	14
Transcription factors	16
Focus on a novel approach	17
Target 1: The ligand	17
Mathematical Modelling	17
Why use mathematical modelling	18
Overview of the model building process	18
Requirements for modelling	18
Rational behind the model	19
Tools used to build the model:	
Systems Biology Mark-up Language (SBML)	19
Drug design	21
First stage	21
Second stage	22
Third stage	23
Drug delivery approaches	23
Summary	24
Target 2: EGFR signal transduction pathway	25
Micro-array expression analysis tools	26
Micro-array expression data	27
Summary	29
Epilogue	31
References	33

Introduction

The INFOBIOMED Network of Excellence (www.infobiomed.org) aims to promote the consolidation of Biomedical Informatics (BMI). This approach of integrating medical informatics, chemoinformatics, and bioinformatics is to foster collaboration on complex case studies within a small group of researchers, who have a widely varying but complementary expertise, ensuring a way of crossing borders between disciplines. The **INFOBIOMED Training Challenge** aims to make advanced students (a) participate in a multidisciplinary research-based training environment; (b) learn the difficulties of crossing “language borders” in the context of a specific research problem; (c) become aware of contents of other disciplines and their particular approaches to the same problem; and (d) use their own expertise to advance science in the area of the case study. It is based on two groups of 5 advanced students with expertise on different biomedical informatics fields, who will collaboratively work on a single case study related to a research problem.

CASE STUDY 2

TARGETING EGFR SIGNAL TRANSDUCTION PATHWAY BY ANTICANCER DRUGS

Cancer chemotherapy has entered a new era of molecularly targeted therapeutics which is highly selective and not associated with the serious toxicities of conventional cytotoxic drugs. The first group of these novel anticancer drugs is that of targeting mutant or aberrantly expressed oncogenic growth factor receptors. HER family is a four-member family of closely related growth factor receptors, including EGFR or HER-1 (erb-B1); HER-2 (erb-B2); HER-3 (erb-B3) and HER-4 (erb-B4). EGFR receptors are essential mediators of cell proliferation and differentiation in the developing embryo and in adult tissues, and their inappropriate activation is associated with the development and severity of many cancers including breast, colon and prostate cancer. This family of receptors is involved in cell-to-cell and cell-to-stroma communication primarily through a process known as signal transduction.

The three best characterized signaling pathways induced through ErbBs are Ras-MAPK, PI3K-PKB/Akt, and PLC-PKC pathways. Ligand binding to the extracellular domain induces the formation of homo- and heterodimers of different members of the EGFR family followed by stimulation of PTK (protein tyrosine kinase) by transauto-phosphorylation. As a result, the transcription of various genes is infected. Two strategies for blocking the action of these proteins include antibodies directed against the ectodomain and drugs that inhibit protein-tyrosine kinase activity. Thus, biochemical, medical, and biological data as well as computational tools are required in order to target this complex pathway by efficiently designed drugs, based on the 3D structure of the proteins involved in the EGFR pathway, their physicochemical properties and their conformational changes during protein-protein and protein-DNA interactions.

Synergetic approaches from researchers of Medical Informatics, Bioinformatics, Chemoinformatics, Toxicology and Epidemiology are required in order to elucidate the pathology and physiology of severe diseases, such as cancer and to further achieve a joint progress in prognosis, diagnosis and treatment of these abnormalities.

The participants

The present case study (2) titled “Targeting EGFR Signal Transduction Pathway by Anticancer Drugs” was examined by 5 five PhD students with different scientific backgrounds, namely Ignasi Belda (Drug Designer), Marilena Garefalaki (Biologist), Mark McAuley (Mathematical Modeler), Eva van Soest (Epidemiologist), and Karin Pike-Overzet (Immunologist).

Ignasi is a Spanish third-year PhD student at the Scientific Park of Barcelona, Spain. He has studied Computer Engineering at the University Ramon Llull, Barcelona, focusing on artificial intelligence. The main topic of his PhD project is the development of a software platform to automated *de novo* peptide design. This software uses evolutionary computation as the main optimisation core, and some other artificial intelligence techniques. In addition, he also conducts research in other chemo-informatics areas such as molecular optimisation, molecular data-mining, or structure-based drug design. His project is fully financed by the Spanish bank BBVA.

Marilena graduated in July 2004 the Faculty of the Biology Department in Aristotle University of Thessaloniki (Greece), having obtained experience in histology processes, as well as, in molecular techniques. During her studies she practiced on the Denaturant Gradient Gel Electrophoresis (DGGE) method in the diagnostic center Locus Medicus from July to August 2003 in Athens, in order to detect mutations on the β -thalassemia gene. Since September 2004 she has been working at the Artificial Intelligence and Information Analysis (AIIA) Laboratory of Thessaloniki within the BIOPATTERN-EU Funded Network of Excellence Project. Particularly, she is working on bio-pattern data processing and content analysis, on the applicability of molecular testing techniques, especially Immunohistochemistry (IHC) and Fluorescence *in situ* Hybridization (FISH) method, in the approach of the prognostic significance and the diagnostic importance of several prognostic markers including HER2/neu gene in breast cancer.

Mark is a graduate of the University of Ulster in biochemical sciences (June 2000). Following this he completed an MSc in computing (January 2002), and then spent time working for the 5*A rated Northern Ireland Centre for Diet and Health (February 2003). For his final year project he created a

computer aided learning (CAL) software package, while his masters involved building an interface (A Java Applet), which controlled a robot via the internet. While working as a research scientist he investigated the possible role folate has in the prevention of cardiovascular disease (CVD). His doctorate studies, entitled Integrative Modelling of Human Ageing, has involved building mathematical models of lipid metabolism, in a bid to gain insights into the nutritional factors that help maintain health as we age.

Eva is a Dutch second-year PhD student at the Erasmus University Medical Center in Rotterdam, The Netherlands. She graduated in November 2003, having studied Human Nutrition and Health at the Wageningen University, Wageningen, The Netherlands, with a focus on epidemiology. The main topic of her PhD project is to investigate the prevalence, and the use of procedures and tests for diagnosis and treatment options for upper GI tract diseases in primary care. For this, a population-based general practitioner (GP) research database containing longitudinal, electronic medical records of more than 600,000 persons throughout The Netherlands, the IPCI database, is used. Her project is a collaboration between the Department of Gastroenterology and Hepatology and the Department of Medical Informatics.

Karin graduated from the University of Groningen (The Netherlands) in 1999 as a molecular biologist. After that, she spent two and a half years in Boston, MA (USA) as a research technician. Currently she is at the beginning of her fourth year as a PhD student at the department of Immunology of the Erasmus Medical Center in Rotterdam, The Netherlands. Her main project is aimed at the development of virus-mediated gene therapy for Severe Combined Immunodeficiency (SCID) patients with a defect in the RAG gene. In addition, she completed a project on the characterization of human T cell development using micro-array analysis.

Why targeting EGFR signaling pathway?

The characteristics that distinguish cancer cells from non-cancerous cells, including lack of differentiation, uncontrolled division, and propensity toward tissue invasion and metastasis, have been well described. During the past 20 years, research efforts have focused on discovering the cellular events that culminate in the transformation of normal cells to malignant cells. Clarifying the molecular basis of malignant transformation and undertaking the differences between malignant and non-cancerous cells create the potential to specifically interfere with these events. The ultimate goal is to interrupt the creation, proliferation, and/or metastasis of cancer cells, while leaving the functioning of normal cells largely undisturbed. This is in contrast to traditional cytotoxic chemotherapy, which typically does not discriminate between normal and tumour cells.

Recent research efforts have attempted to exploit biologic differences that may exist between normal and malignant cells to develop tumour-specific therapies. In theory, these therapies would avoid or minimize many of the debilitating non-specific toxicities associated with chemotherapy. By focusing on key molecules integral for cellular function, replication, or tumour genesis, such targeted therapies may exert cytostatic or cytotoxic effects on tumours.

In normal cells the family of Epidermal Growth Factor Receptors (EGFR) couples binding of extra-cellular growth factor ligands to intracellular signalling pathways regulating diverse biologic responses, including proliferation, differentiation, cell motility, and survival. The four closely related members of this receptor tyrosine kinase (RTK) family—epidermal growth factor receptor (EGFR, also known as ErbB-1 or HER1), ErbB-2 (HER2), ErbB-3 (HER3), and ErbB-4 (HER4) form a complex network of 4 receptors and more than 10 interacting ligands (figure 1). Initiation of the signalling cascade occurs upon ligand-induced receptor homo- and hetero-dimerisation. Stimulation of PTK (protein tyrosine kinase) is followed by transautophosphorylation. As a result, the transcription of various genes is affected by phosphorylation or dephosphorylation of a series of transmembrane proteins

and intracellular signaling intermediates, many of which possess enzymatic activity. Signal propagation occurs as the enzymatic activity of one protein turns on the enzymatic activity of the next protein in the pathway. The cellular outcome of activation depends upon the complement of signalling pathways induced, as well as their magnitude and duration, which in turn are determined by the composition of the receptor pair and the identity of the ligand.

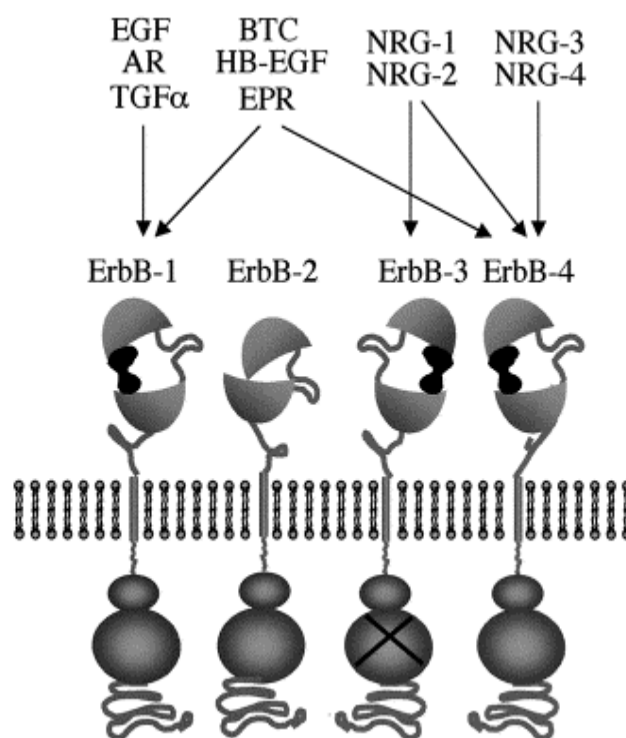


Figure 1. ErbB receptors and their ligands (Marmor *et al*, 2004)

Deregulation of the many signalling pathways (Ras-MAPK, PI3K-PKB/Akt, PLC-PKC and STAT) induced through this family can promote multiple properties of neoplastic cells, including proliferation, migration, angiogenesis, stromal invasion, and resistance to apoptosis (Figure 2). Hyperactivation of the ErbB network can occur via an autocrine secretory loop involving overproduction of ligands and receptors by the tumour cells, or paracrine

growth dependent on ErbB ligands produced by adjacent stromal cells. Alternatively, aberrant growth can ensue from constitutive receptor activation.

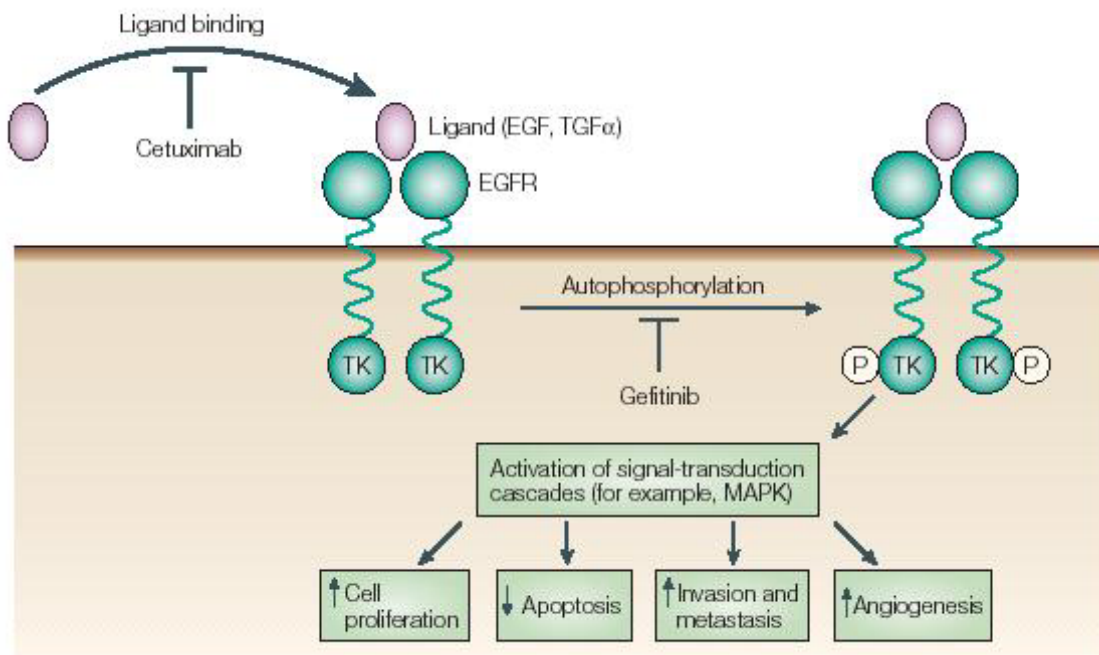


Figure 2. EGFR Signal Transduction pathway in cancer (<http://www.bufofanine.com/>)

Initial contributions

Based on their field of expertise all members of the group contributed to this case study. Marilena, as a Biologist made a general introduction to the basic principles of signal transduction, the biology of the family of Epidermal Growth Factor Receptors; and the implication of EGFR signalling in multiple human cancers. Eva, with a background in Epidemiology, reviewed the therapies targeting the EGFR signalling pathway currently under experimental and clinical investigation. Karin, as an Immunologist and having experience in micro-array data analysis introduced this method of analysis and how it could be applied in this case study. Mark overviewed basic modelling, as an expert in mathematical modelling, while Ignasi introduced the drug designing techniques because this is his field of expertise. Apart from the group-member knowledge additional sources of information were provided by Pubmed (www.pubmed.com) for literature investigation, and by the tutors that supervised the Training Challenge during the whole week. Moreover, some group members (Karin, Ignasi) used their own contacts (colleagues etc.) as well as contacts with an author of an interesting article for additional explanations. Micro-array databases and modelling software (Systems Biology Mark-up Language; SBML) were also used in this integrative research attempt.

Process of finding the target

The multi-layered network of EGFR signal transduction pathway with interacting ligands, receptors, effectors and transcription factors is a highly complex system. Targeting the EGFR pathway with anti-cancer drugs opens a broad field of potential targets, which, giving the available time and expertise, could not all be explored in detail. Enormous amounts of data exist and reductive approaches had to be followed in order to focus on a specific target. Therefore, in terms of reducing and specifying the integrative research, the multi-layered pathway of EGFR was divided into sub-layers in order to identify potential targets for therapeutic strategies. The following list was produced:

1. Extra-cellular layer:
 - Ligands
 - Receptor (-ligand binding domain)
2. Intracellular cascade (EGFR kinase domain and downstream signalling molecules)
3. Transcription factors

For each target a literature search was performed in order to identify an achievable target for this research exploration. In line with this notion, currently available therapeutic strategies were investigated and discussed in order to clarify what has been shown to be promising, what has been shown to fail and what the possible reasons were.

Ligands

Multiple ligands bind to and activate EGFR including epidermal growth factor (EGF), transforming growth factor- α (TGF- α), amphiregulin, heparin-binding EGF, betacellulin and neuregulins (NRG). They comprise the initial part of the signalling pathway. Even though biochemical data considering the crystal structure of the ligands exist, only recently the crystal structures of ligands interacting with the extra-cellular domains of EGFR have been revealed (Ogiso, Ferguson, Garrett). This might be a very challenging and novel area of research.

Therefore, the proposed strategy in this section is to block the ligands in order to impede their interaction with the receptors. Among the advantages

of choosing the ligands as a potential target, include possibly their accessibility. Since EGFR ligands are extra-cellular proteins, drugs do not need special cell crossing abilities. Hence, the *a priori* pharmacodynamics of the proposed drug could be effective. In addition, the medium-sized molecules used as drugs- possibly peptides- are not expected to raise immune response-derived side effects. Ligands share similar structural properties, which could be used in the drug design stage.

On the other hand, we have devised one important disadvantage. Since EGFRs have a very broad expression pattern on epithelial, mesenchymal and neuronal cells and signalling through these receptors plays a critical developmental role, generalized side effects could be expected.

Nevertheless, in normal cells some ligands show tissue and developmental stage specific expression pattern (Olayioye). This could contribute to targeted therapies against a certain subset of ligands, but further research is required in order to clarify if this specificity remains after the transformation of a normal cell to a malignant.

From this information a table summarizing the advantages and disadvantages of targeting this part of the signalling pathway has been built (Table 1).

Table 1:

Advantages	Disadvantages
Accessible target	Expected generalized side-effects
Not published yet	
Structural similarities	
Immune-response-derived side effects	not expected

Receptor (-ligand binding domain)

Monoclonal antibody therapy has been developed to block the binding of ligands to the receptor. With this therapy, specific antibodies bind to the

EGFR extra-cellular ligand-binding domain decreasing receptor-ligand interaction resulting in receptor down modulation and cell death.

Several antibodies are currently under investigation. Cetuximab is the most extensively investigated and was FDA-approved in February 2004 for the treatment of metastatic colorectal cancer (Erbix). Preclinical studies showed promising results, with an anti-tumour response up to 75% in mice on Cetuximab therapy. In clinical studies the drug was reasonably tolerated, with skin and allergic reactions as most common toxicities. Mild adverse reactions such as fever, chills, headache, nausea, stomatitis, vomiting, diarrhea, anemia and weight loss have been reported in 9-16% of patients. In a phase I clinical trial with Cetuximab in combination with radiation 13 of 15 patients showed a complete response, however five of them died due to a relapse after a median of 8 months. Two phase II trials examined Cetuximab in combination with chemotherapy. In the first trial 53 patients were included and the overall response rate was 17%, 66% responded with stable disease. In the other trial 22% of 18 patients had stable disease, 67% had a partial response and 11% had progressive disease. Phase III trials are currently being conducted. ABX-EGF, EMD 72000 and hR3 are three other monoclonal antibodies currently in different stages of clinical testing. All three were reasonably tolerated and response rates in small groups of patients varied between 7% and 87% (Thomas).

In summary, initial results suggest that the responses to monoclonal antibodies are modest, although trials are still ongoing.

An advantage of targeting the extra-cellular part of the receptor would be that it has good accessibility, since no cell membrane has to be crossed.

From the previously discussed information we have built a table summarizing the advantages and disadvantages of targeting this part of the signalling pathway (Table 2).

Table 2:

Advantages	Disadvantages
Clinical data available	Possible side-effects
Accessible target	Modest response rate
Promising in preclinical studies	Not novel

Intracellular cascade

Small molecule tyrosine kinase inhibitors are synthetic compounds that competitively bind, reversibly or irreversibly, to the intracellular tyrosine kinase domain of the EGFR preventing receptor autophosphorylation and initiation of the intracellular pathway.

The first to be licensed for use in this group of drugs was Gefitinib, which was FDA-approved in May 2003 for advanced lung cancer. However, the FDA withdrew the approval in December 2004 based on a phase III trial in which no survival advantage of the drug over a placebo was shown. Gefitinib was promising in preclinical studies. The drug was reasonably tolerated, with mostly grade 1 or 2 toxicities such as skin rash, nausea, diarrhoea, vomiting and asthenia. However, one study in Japan reported severe pulmonary side effects in 8.7% of 103 patients proving to be fatal in three of them. Phase II trials showed that some tumours that express EGFR do not respond to Gefitinib treatment.

Another tyrosine kinase inhibitor that was recently (November 2004) licensed for use is Erlotinib. This drug was also reasonably tolerated, with diarrhoea and cutaneous toxicities as main side effects. This drug was licensed based on a phase III trial where a small increase in overall survival and quality of life was shown. Three other tyrosine kinase inhibitors currently under investigation are PKI 166, CI-1003 and GW572016 which are all still in initial phase of clinical testing (Thomas).

In summary, although preclinical testing was promising, clinical response rates have been disappointing. Mid 2004 however two articles were published simultaneously, reporting on an association between certain mutations in the EGFR and the response to gefitinib therapy (Lynch, Paez). It

was shown that mutant receptors were more sensible to gefinitib therapy than the wild-type receptor. Later this association was also reported for erlotinib therapy (Pao). More recent reports have indicated a role for other factors in determining the response to TKIs, including increased copy number of both EGFR and HER2 (Capuzzo).

A few other substances in the intracellular cascade have been targeted in order to prevent full activation of the pathway. Results in clinical trials showed a minor or no effect. Agents interfering with or blocking the Ras pathway like Farnesyl transferase inhibitors (such as Tipifarnib and Lonafarnib) have been tested in Phase II and III clinical trials. Phase III trials using Tipifarnib showed no significant anti-tumour activity in colorectal cancer. Phase I/II studies using Lonafarnib in combination with both Gemcitabine and Paclitaxel have been carried out resulting in clinical activity mainly in pancreatic cancer and non-small cell lung cancer. However, observed anti-tumour activity was short-lived because cancer cells develop resistance to this treatment. A Protein Kinase C (PKC) inhibitor (ISIS3521) was tested in a preclinical setting but showed no effect on tumour growth. The only PI3 kinase inhibitor that was tested in a phase II trial (CCI-779) elicited initial disease stabilization in approximately 50% of patients but the durability of response was short. Some of the described compounds are currently considered for use in combination with other anti cancer drugs.

Targeting the intracellular pathway would have the disadvantage of not being as easily accessible as the two previously discussed targets (ligands and receptor) since drugs would have to be able to cross a cell membrane in order to reach the specific goal.

Furthermore, it has been shown that some of the drugs targeting the intracellular cascade only exhibit a temporary response. It is thought that this might be caused by a cell mechanism that is able to 'redirect' a specific pathway, thereby circumventing the targeted area. In other words, different pathways can interact and take over each other's functions in order to reach homeostasis.

However, an advantage could be that micro-array expression data of normal and cancerous cells is available and that mapping these data onto the

recently published extensive EGFR pathway map (Oda) could possibly give new insights in the genes incorporated in this pathway and their role in cancerous cells.

From the previously discussed information we have built a table summarizing the advantages and disadvantages of targeting this part of the signalling pathway (Table 3).

Table 3:

Advantages	Disadvantages
Clinical information available	Low response rate in clinical trials
Promising in preclinical studies	One membrane to cross
Expression data available	Alternative pathway resistance (temporary response)
Recently published pathway	Possible side-effects

Transcription factors

Transcription factors are not static proteins bound to DNA. The transcriptional activity of cells must be adjusted during tissue homeostasis and cell differentiation. The activity of most transcription factors is tightly regulated by signals from the extra-cellular milieu. Each cell is under the influence of a dense network of cross-communicating signalling pathways. Each cell type has a unique developmental history and sets of transcription factors, ensuring that different tissues respond differently to a limited number of cell-cell signals. The table below outlines some of the advantages and disadvantages for targeting EGFR signal transduction transcription factors (Table 4).

Table 4:

Advantages	Disadvantages
Novel	Two membranes to cross. Little data to proceed.

Focus on a novel approach

After discussing and weighing the advantages and disadvantages of the possible targets, we concluded that we first had to decide if we wanted to:

1. Try to find out why currently available strategies are not showing the expected results
2. Try to establish a novel approach to target the EGFR pathway.

The decision was made to focus on a novel approach. Since two possible targets seemed promising and we did not want to discard one of them in a preliminary stage, we focused on:

1. The ligands
2. The intracellular pathway

Both approaches will be discussed below.

Target 1: The ligand

Mathematical modelling

Following a novel approach of targeting the ligands of the EGFR receptor family experimental and literature based biological data are required in order to investigate the dynamics of ligand binding, internalisation and degradation. Recent research has contributed additively to this approach since based on the crystal structure of the ligands which has been revealed, the ligand-receptor interactions and the following conformational changes of the latter have been identified. In terms of interdisciplinary approaches, modelling of these interactions could provide a very intelligent predictive tool.

Within this particular context mathematical modelling involves converting a biological process such as EGFR signal transduction into mathematics. We can then use the mathematics to conduct *in silico* experiments using a computer (Figure 3).

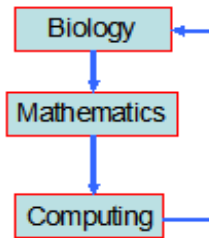


Figure 3: Interaction between disciplines.

Why use mathematical modelling?

In silico modelling offers a convenient, cheap, ethical and effective way of dealing with the complexity associated with EGFR signal transduction. Furthermore it allows hypotheses to be tested and experiments to be conducted that are difficult to do *in vivo* or *in vitro*.

Overview of the Model building Process

1. Start with a diagram.
 2. Decide on the types of reactions Involved.
 3. Parameterisation.
 4. Simulate.
 5. Compare the output with known behaviour.
-
6. *In silico* experiments/Model extension or integration.

Requirements

- Modelling building tools.
- Simulator.
- Inference Techniques.
- Experimental data.
- Data repositories.
- Knowledge – Network of experts.

Rational behind the Model

Kinetic models have been useful for representing the dynamics of ligand binding, internalization and degradation. They have predicted binding and distribution patterns of ligands within cells. Furthermore they have established links between physiological responses e.g. mitogenesis and levels of receptor activation in specific cellular compartments and have allowed us to appreciate the complexity of EGFR trafficking. They have provided significant insights into its regulation. Insights have also been gained from combining quantitative, experimental and modelling approaches to EGFR trafficking problems.

Tools Used To Build the Model: Systems Biology Mark-up Language (SBML)

SBML is software designed specifically to deal with problems in biology, such as modelling cell signalling pathways, metabolic pathways, biochemical reaction networks, genomic regulation pathways. SBML is based on an XML schema, which is ubiquitously found on the Internet as a means of representing data. Unlike computer languages such as C, C++ and Java, SBML was not designed for manual coding. Instead a number of tools are available that automatically generate the code. For this model a freely available open source software package called MathSBML was used to generate the SBML. MathSBML is a software package designed to work with Mathematica. Models are created by using a model editor that allows SBML objects such as compartments, species and reactions to be added, organized and modified. MathSBML automatically generates a system of ordinary differential equations. A function called SBMLNDSolve then solves the system of differential equations using Mathematica's NDSolve. Another function called SBMLPlot then can generate a plot of the resulting solutions.

Using the above described modelling approach, combined with biological knowledge of the EGF receptor and ligand interaction, an initial model was built (Figure 4). When adding biological constants, this model could be used to study the dynamics these interactions and the effects of interfering with one

or more of the ligands. The outcomes of the different *in silico* tests could be used to decide which ligand or ligands are best to target. *In vitro* experiments are further required in order to certify or reject the hypothesis by experiments on drug designing.

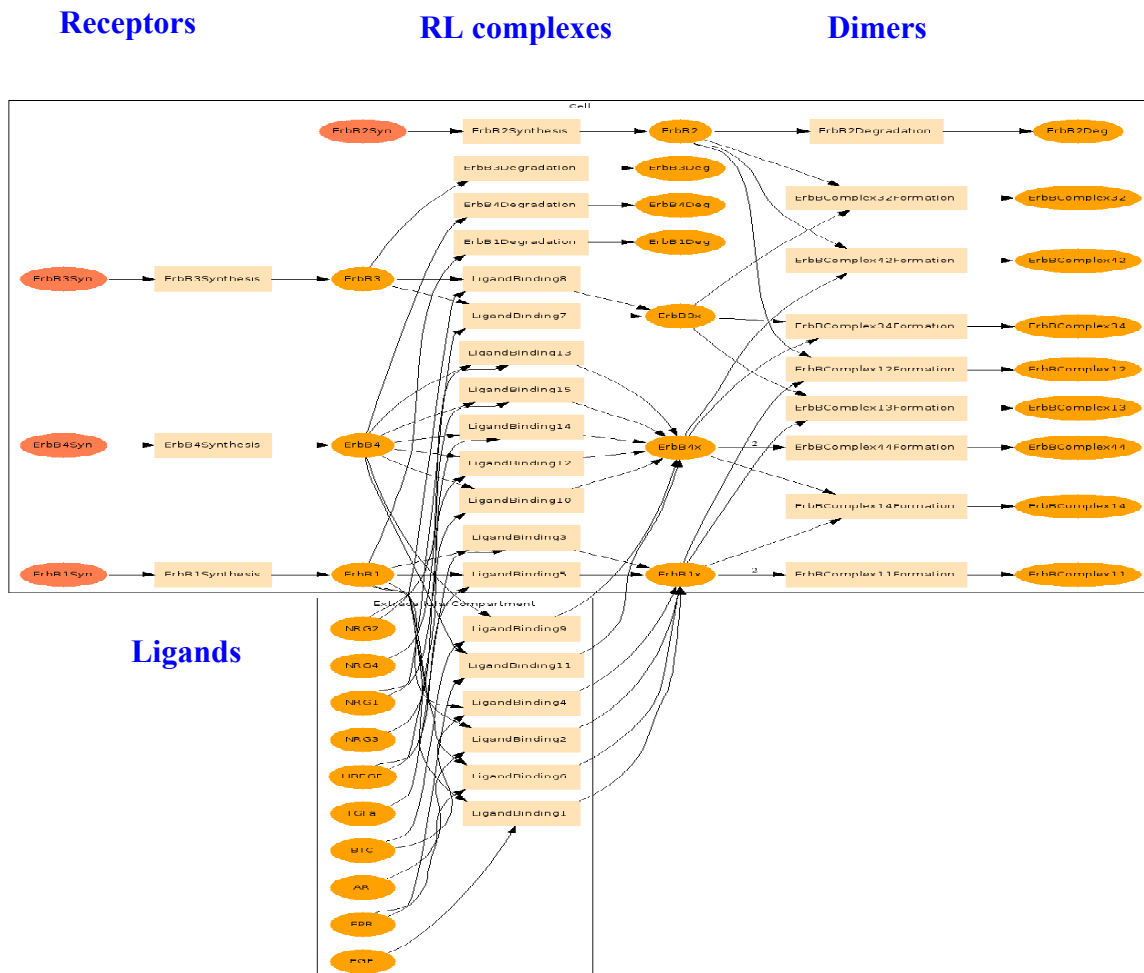


Figure 4: Diagram of the model

Resources used to build the model:

- <http://sbml.org/index.psp>- The home of the Systems Biology Markup Language
- <http://sbml.org/software/mathsbml/>- The home of MathSBML
- <http://www.wolfram.com/products/mathematica/index.html>- The home of Mathematica

Drug design

The proposed drug design and optimisation project is divided in three stages. The first stage is where the initial drug design takes place. Several strategies will be used and some *in vitro* experiments will be carried out. In the second stage, the drug will be optimised using more sophisticated *in vitro* cellular experiments. In this stage the goal will be the optimisation of the pharmacokinetic properties of the designed drug. Finally, a third stage will be performed using *in vivo* models. Here, the pharmacodynamic properties of the drug will be studied and optimised. In addition, toxicity studies will take place.

First stage

The aim of this first drug design stage is to identify one or more molecular hits. Molecular hits are those molecules that present certain activity. Since hits are difficult to obtain, several independent strategies are used. Sometimes, knowledge obtained in one research line can be used in others. Among the proposed strategies of hit identification we included: structure-based design, high-throughput screening of natural products, and *de novo* virtual design.

Structure-based design and high-throughput screening of natural products need in the latter stages some simple *in vitro* molecular model to test the proposed molecules. Basically, what is needed is a binding assay. We propose three of them, which could be used progressively. In the beginning, the cheaper techniques should be used. And when better molecules are identified, better should be the used in the binding assay. The proposed methods are: surface plasmon resonance, micro-calorimetry, and nuclear magnetic resonance. Of course, the target proteins must be synthesised or expressed in order to be able to perform those assays.

Structure-based design strategy aims to identify hits paying attention to structural facts. Since the structures and amino acid sequences of the receptors are known, we could design molecules similar to the parts of the receptor interacting with the ligands. In the very beginning, this step could be aided by some computational techniques, such as docking. Latter, in more advanced steps, *in vitro* binding assays should be used.

High-throughput screening of natural products aims to identify hits screening natural products. Natural products include a wide range of sources. The most preferred source, for its low cost, are traditional medicinal plants (for example, Chinese or Amazonic traditional products). However other natural products could be explored, for example marine animals. To carry out this drug design experiment, binding assays are needed from the very beginning. This experiments, at least in the first steps, must be quick and cheap, in order to extend the screening to a huge range of products.

Finally, the aim of the *de novo* virtual design is to identify hits by virtual screening. In here different search strategies are available. Among them, the virtual evolution of molecules is one of the preferred, since new and potent structures could be obtained easily. In this strategy is not needed an *in vitro* assay. However, high computing resources are needed.

Second stage

Once some hits have been identified in the previous stage, they should be optimised. The aim is to optimise the pharmacokinetic properties of the identified hits. For this, a more complex cellular model should be developed, where the molecules can be tested more accurately. For instance, a model could be built using neoplastic cells. The targeted EGF ligands should be placed in the medium, and then the cells start to reproduce themselves. Hence the test would be to put the designed molecules in the medium, and see if they are able to stop the cellular division.

In order to optimise the molecules, some systematic strategies must be followed. We propose combinatorial chemistry aided by the tools chemical graph identifier and QSAR (quantitative structure-activity relationship). In that manner, analogue molecules are built in a combinatorial fashion, and using statistical tools, their activity is modelled. Hence, using the statistical model, new molecules may be designed with potentially higher activities.

At the end of this stage, some simple toxicity experiments should be carried out in cells. And in addition, some simple degradation experiments in human serum.

Third stage

Finally a third stage should be carried out using *in vivo* models. The aim of this stage is not only to refine the pharmacokinetic properties of the proposed drug, but also to optimise their pharmacodynamics. For such a reason, two *in vivo* models could be built. In the first one we propose an induced cancer model, for example, mice which are exposed 16 hours each day to ultraviolet light. In this model, a quick cancer will be induced in few days. The drug can be tested and optimised in here.

In addition, another more complex model could be built. That is, a knock-in model which the selected targets are over-expressed. Hence, the effects of the drug could be analysed in a more specific way.

Both models can serve as an experiment to test the toxicity of our drugs. Side effects should be analysed and studied. In particular, potential immuno-response that could arise if the designed drug is a big molecule.

Drug delivery approaches

In order to minimize the side effects and optimise the *in vivo* activity of the designed drug, some particular issues in the delivery approaches must be taken into account. First of all, we should identify towards which tissues we want to address our drug. Hence, more efficient delivery techniques could be used for each particular case.

For instance, if we want to address the drug to lungs, we could use some modern and efficient inhaling delivery techniques. In addition, at the molecular level, we could modify our molecule to design a pro-drug, *i.e.* A drug that is only active when it loses a part of it. Human L-Capsine protease is a potent protease present in the extra-cellular medium of human cancerous lungs. Since the activity of this protease is well described, the proposed molecule will become active when this protease breaks the connection between the carrier and the drug. In this manner, the side effects are reduced in a great manner, because the drug becomes only active in lungs.

Another example is the brain. If the designed drug is addressed for cerebral cells, it must be designed taking into account their ability to cross the blood-brain barrier.

Summary

Targeting the EGFR network in its initial sub-layer, the ligands, could be a promising approach, not only due to its novelty, but existing biological, biochemical data, mathematical modelling tools and drug designing strategies offer new insights in this field of research. However, numerous tasks have to be elucidated in order to finally achieve targeted therapies combined with less side effects and resistance to therapy.

Therefore, given the tissue and developmental stage-dependent expression pattern of the ligands in normal tissue (Olayioye), one initial point is to clarify if these ligands are also over-expressed in the case of malignant transformation. This would help in more effective drug delivery.

One more important point in terms of drug delivery is to use fewer drugs with the maximum of efficacy. This could possibly be achieved due to the common structural characteristics of the EGFR ligands (EGF-like domain).

In terms of mathematical and computer assisted modelling of the signalling pathway, more quantitative data of the interacting proteins are required (kinetics), in order to refine the developed tools and increase their predictive capacity.

Target 2: EGFR signal transduction pathway

The second option we decided to explore in more detail was finding a new anti-cancer target within the signalling cascade of the EGFR. As discussed above, we decided to explore finding a new target in the cascade instead of designing a new drug for targets already described in literature and our objective was to do so using new knowledge and relatively new technology.

Recently, a comprehensive pathway map of the epidermal growth factor receptor signalling pathway was published by Kanae Oda *et al.* Among others, this pathway encompasses 202 proteins and was written in Systems Biology Mark-up Language (SBML) allowing it to be used for mathematical modelling. However, no parameters have been entered into this system at this time so no calculations can be executed as of yet. The list of proteins involved in the pathway is available and could be used for purposes other than mathematical modelling.

A technology that has developed tremendously over the past decade is RNA expression analysis using DNA micro arrays. Currently, this technique allows biologists to analyse the expression of 47,000 transcripts at the same time. This technique has been used to characterize expression in cells undergoing drug treatment (including time-courses), development, diagnosis vs. relapse and many more applications. The advantage of generating a lot of data at the same time also poses a problem. Analysing expression data of about 47,000 transcripts for each sample or each sample group requires complex calculations.

Several tools have been developed allowing researchers to extract data of interest from these huge micro-array datasets. Many of these tools are freely available on the Internet. Our idea was to map gene expression measured by DNA micro-array to the EGFR signalling pathway. The specific goal of this approach is to identify proteins within the EGFR signalling pathway that are over-expressed in cancerous cells. New drugs could be designed specifically for proteins that show a high expression level in cancerous cells.

Micro-array expression analysis tools

First, we explored some of the pathway databases and analysis programs to see if they could be used for our analysis. A short evaluation of some of the programs is reported below, mainly focusing on the comprehensiveness of data on the EGFR pathway and the ability to process data generated with Affymetrix, one of the most widely used platforms for micro-arrays.

- www.biocarta.com
Only 29 proteins are described in this database and it does not provide a tool to analyse expression data.
- stke.sciencemag.org/cm
A more comprehensive database with respect to the EGFR pathway, but only pathways and interactions are described: no option for expression analysis is given.
- www.genmapp.org
A comprehensive database, however, no information specific to the EGFR signalling pathways was found.
- www.invitrogen.com/ipath
Fairly simple pathway descriptions and no option for expression analysis is given.
- www.hprd.org/interactor_map
Very comprehensive description of the EGFR signalling pathway and many other pathways but like many other tools, options for expression analysis are not provided.

The tools discussed above were found less suitable for our objective. The following tools were considered more applicable to our specific situation.

- www.cytoscape.org
This program offers visualization of molecular interaction networks and integrating these interactions with gene expression profiles. Unfortunately, using this program requires downloading in order to see

all options. Within the scope of the Training Challenge it was not feasible to test this tool.

- <https://panther.appliedbiosystems.com>

This tool allows us to overlay expression data on pathway diagrams to visualize the relationships between genes/proteins in known pathways. The pathway description mentioned earlier (published by Kanae Oda *et al.*) has been loaded in this analysis tool. Unfortunately, using this program also requires downloading and in the timeframe given for the Training Challenge it was not feasible to test this option.

- PathwayAssist

This is another program that allows one to interpret micro-array data in the context of pathways, gene regulation networks and protein interaction maps. This program was available at the Training Challenge on the laptop of one of the tutors.

Micro-array expression data

After identifying some useful analysis programs, micro-array data were needed to see if our approach for identifying a new target would work. At the onset of the Training Challenge, gathering micro-array data did not seem like a big problem. Literally thousands of micro-array experiments have been done in the field of cancer. One source, www.oncomine.org, contained a lot of data specifically derived from cancer samples. This database even stored expression data from EGFR over-expressing cancers. However, upon taking a closer look, most data were published in such a way that anyone not directly involved in generating the data would not be able to understand what each sample represented. We also performed PubMed-searches in order to find data that would be suitable for our analysis. The main problem with most of the data we found was the lack of raw data as most publications only contain data from extensive statistical analyses.

Eventually, a publication was found in the Oncomine database that published data derived from normal and cancerous samples in such a way that an outsider could determine what the individual samples represented. The data, comparing normal tissue to various types of solid lung tumours,

were generated using relatively small arrays of about 7000 transcripts and only a selected set of transcripts was published (around 900).

Taking into account the availability of analysis tools and suitable data we decided to attempt a small analysis using the limited dataset on lung tumours and the PathwayAssist program. PathwayAssist identifies proteins using LocusLink IDs, therefore the expression data need to be converted into a format that also identifies proteins using these IDs. Existing conversion tables were used by one of the tutors to create a table with the proper format. Conversion was only successful for 92 transcripts, making our analysis even smaller.

Our original idea was to compare the expression data with the EGFR pathway only, but with the extremely small dataset only a 'general' analysis could be done. The results are depicted in figure 5 and should only be considered as an example of an analysis. The colour of a species represents the difference in expression. Red represents up-regulation and green represents down-regulation of expression in the cancerous sample versus the normal sample. The intensity of the colour corresponds to the extent of the difference in expression, with the higher intensities representing the larger differences. This figure does not represent expression within the EGFR signalling pathway and therefore does not give any information about a possible target for anticancer drugs.

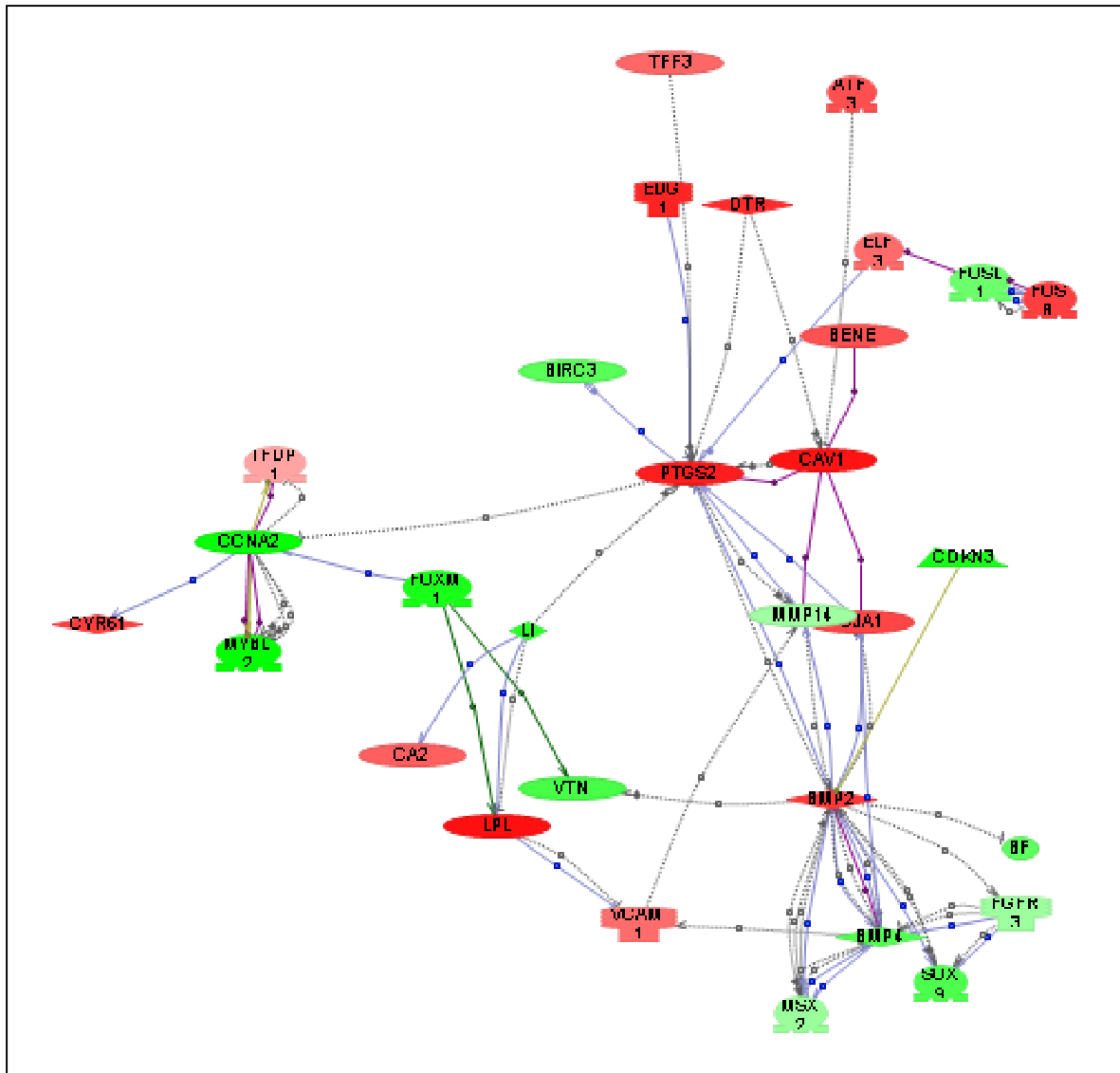


Figure 5: Example analysis

Summary

Within the time available for the Training Challenge we were not able to analyse micro-array expression data from normal vs. cancerous cells using the comprehensive pathway map of the epidermal growth factor receptor signalling pathway. However, using a small dataset to test one of the tools for analysing expression data in the context of known proteins did provide us with 'proof of principle'.

If one wanted to take this analysis further, we would advise downloading the Panther analysis tool from Applied Biosystems

(<https://panther.appliedbiosystems.com>. This program has incorporated the comprehensive pathway map of the EGFR published by Oda *et al.* The only major hurdle in this approach is the poor availability of micro-array data. In order to be able to do the really interesting analyses, a collaboration should be started with scientists that have performed the normal vs. cancerous experiments. If all these requirements are met, the proposed analysis is a realistic possibility.

Epilogue

The First INFOBIOMED Training Challenge reached all of its initial goals in integrative research. Offering a multidisciplinary research-based training environment, students had the opportunity to learn how to cross the “language borders” in the context of a specific research problem, to become aware of contents of other disciplines and their particular approaches to the same problem and finally to use their own expertise to advance science in the area of the case study.

In terms of some minor advice for future training challenges, it could be more beneficial if the participants knew the general background of the members in their study group. This way, overlapping information could be avoided. With respect to the experts that visit the ‘house’, a proper introduction of them including their scientific field of expertise could add more in this multidisciplinary approach of the study. And last but not least, the training challenge should be emphasized rather than the competition between the two groups. This way, collaboration even between the members of the two groups could expand the multidisciplinary approach, as the scientific background of the members within the two groups is not exactly identical.

After this training challenge, all students returned to their lab having gained more experience in integrative research. A good starting point of this cooperation was that all members of this study group were able to listen to each other carefully, thus crossing language borders in terms of nationality at first, and later in terms of scientific terminology. Finally it proved not only easier than expected to work with scientist from different research fields, but also creative ideas arose from interdisciplinary discussions. Furthermore, unstructured distributed leadership works, at least in a group of five members, where in turn, but in an unstructured way, everybody was taking the lead now and then, learning in this way that it is possible and pleasant to work in a group with a horizontal structure, and a defined leader is not a requirement for good collaboration.

Overall, the Infobiomed Trainings Challenge was an enjoyable experience, learning a lot about the EGFR pathway, about working together in a group, about Spanish hospitality. Working hard and long hours was combined with relaxing walks and extensive meals. Meeting new people with diverse personalities, nationalities and scientific field of expertise, lead to a fruitful and pleasant cooperation.

Therefore, all the members of this team (Mark, Eva, Karin, Marilena and Ignasi) would like to thank everybody in the organizing committee of the Trainings Challenge for the opportunity to live this experience, and the tutors and experts for their expertise, presentations, nice chats, good organization and creating a pleasant atmosphere for both work and distraction from work.

Summarizing nicely how much we laughed during the trainings week we would like to conclude with our 7C-word 'complicated cancer cell cross cascade communication complex!' and a little ligand dance, all for the benefit of multidisciplinary research and crossing language borders, of course!



References

Cappuzzo F, Varella-Garcia M, Shigematsu H, Domenichini I, Bartolini S, Ceresoli GL. Increases HER2 gene copy number is associated with response to gefitinib therapy in epidermal growth factor receptor-positive non-small-cell lung cancer patients. *J Clin Oncol* 2005; 23 (22): 5007-18.

Ferguson KM, Berger MB, Mendrola JM, Cho HS, Leahy DJ, Lemmon MA. EGF activates its receptor by removing interactions that autoinhibit ectodomain dimerization. *Mol Cell*. 2003 Feb; 11(2): 507-17.

Garrett TP, McKern NM, Lou M, Elleman TC, Adams TE, Lovrecz GO, Kofler M, Jorissen RN, Nice EC, Burgess AW, Ward CW. The crystal structure of a truncated ErbB2 ectodomain reveals an active conformation, poised to interact with other ErbB receptors. *Mol Cell*. 2003 Feb; 11(2): 495-505.

Lynch TJ, Bell DW, Sordella R, Gurubhagavatula S, Okimoto RA, Brannigan BW, et al. Activating mutations in the epidermal growth factors receptors underlying responsiveness of non-small-cell lung cancer to gefitinib. *N Eng J Med* 2004; 350 (21): 2129-39.

Marmor MD, Skaria KB, Yarden Y. Signal transduction and oncogenesis by ErbB/HER receptors. *Int J Radiat Oncol Biol Phys*. 2004 Mar 1; 58(3): 903-13.

Oda K, Matsuoka Y, Funahashi A, Kitano H. A comprehensive pathway map of epidermal growth factor receptor signalling. *Molecular Systems Biology*, published online: 25 May 2005

Ogiso H, Ishitani R, Nureki O, Fukai S, Yamanaka M, Kim JH, Saito K, Sakamoto A, Inoue M, Shirouzu M, Yokoyama S. Crystal structure of the complex of human epidermal growth factor and receptor extracellular domains. *Cell*. 2002 Sep 20; 110(6): 775-87.

Olayioye MA, Neve RM, Lane HA, Hynes NE. The ErbB signaling network: receptor heterodimerization in development and cancer. *EMBO J.* 2000 Jul 3; 19(13):3159-67.

Paez JG, Janne PA, Lee JC, et al. EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy. *Science* 2004; 304: 1497-500.

Pao W, Miller V, Zakowski M, Soherly J, Politi K, Sarkaria I, et al. EGF receptor gene mutations are common in lung cancers from “never smokers” and are associated with sensitivity of tumors to gefitinib and erlotinib. *PNAS* 2004; 101 (36): 13306-11.

Thomas SM, Grandis JR. Pharmacokinetic and pharmacodynamic properties of EGFR inhibitors under clinical investigation. *Cancer Treatment Reviews* 2004; 30: 255-268.