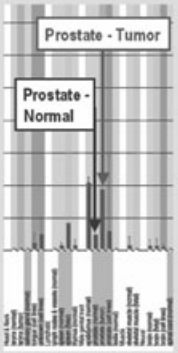
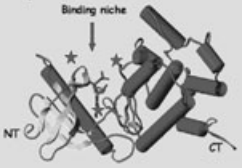
The Target Discovery Process: From Gene to Target

DNA microarray



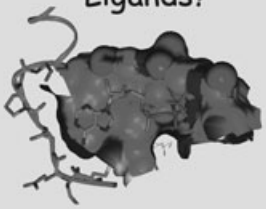
Bioinformatics
Gene Expression Profiling (GEP)
GEP data mining

Homologous crystal structure?



Structure-based sequence alignment

Binding niche? Ligands?



Selectivity & drugability analysis

Identifi-cation

Candidate Gene

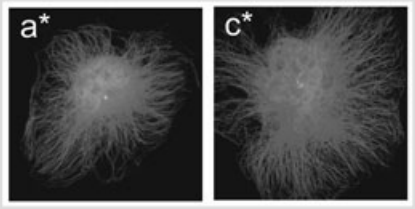
Assess-ment

Drugable Target Candidate

Two independent approaches to analyse phenotypic changes

1) RNAi-based knockdown studies

2) Overexpression studies



Vali-dation

Validated target involved in disease-relevant process

DOI: 10.1002/cbic.200400158 *ChemBioChem* **2005**, *6*, 468 – 479

The Target Discovery Process

Ursula Egner,* Jörn Krätzschmar, Bertolt Kreft, Hans-Dieter Pohlenz, and Martin Schneider[a]

In order to minimise attrition rates in drug development projects, a target discovery process is implemented to select and characterise the most suitable candidate kinase targets, before lead identification and lead optimisation are embarked upon. The process consists of 1) target selection, 2) target assessment, and 3) target validation. This rational approach to target discovery, as a prerequisite for lead discovery, ensures that new therapeutic targets fulfil a set of general criteria, as well as indication-specific, descriptive and functional ones. The approach should ultimately maximise the likelihood of achieving target-selective inhibition by small-molecule inhibitors with minimal in vivo side effects and a therapeutic effect based on a sound biological hypothesis.

Introduction

During the past decade we have seen fundamental changes in the way pharmaceutical companies approach the identification and validation of new drug targets. Historically, lead compounds were identified by screening small molecules for the induction of a desired phenotype, for example, anticancer agents blocking cell proliferation. However, many such drug candidates ultimately failed in clinical development, either due to poor pharmacokinetic compound characteristics or because of intolerable side effects, which may reflect insufficient specificity of the compound or unsuitability of the target. In fact, a significant number of drug development projects have failed because the underlying biological hypothesis about the target has been incorrect.[1] To reduce the attrition rate during development, most, if not all, pharmaceutical companies have adopted a target-directed molecular approach, which aims at understanding the cellular mechanisms underlying a given disease phenotype. This change in the drug discovery process has been supported substantially by the completion of the sequencing of the human genome and by the introduction of novel genomics and proteomics technologies. Overall, identifying novel drug targets, that is, proteins that are critically involved in the development and/or progression of a disease, is a multistep endeavour involving various disciplines, including large-scale expression profiling and bioinformatics, structural biology, traditional cell biology, and ultimately functional in vivo studies.

The selection of candidate targets is greatly supported by the use of systematic gene expression profiling to enable the discovery of genes/proteins with a desirable tissue distribution that are regulated in a model system of pathophysiology and/or differentially expressed in clinical patient samples in comparison to normal samples. Gene expression profile information, put in the context of functional information on a target protein or one of its close homologues, is today guiding selection of candidate targets as the first step in the target discovery process. The cellular function of a potential target is analysed in detail in the process of target validation. In a series of in vitro assays, including knock-down and over-expression studies in appropriate model systems, target validation aims at demonstrating that a candidate target protein—and ultimately its enzymatic activity—does play a critical role in a disease-relevant cellular process.

In addition to its critical contribution to a disease condition, the target protein of interest should be drugable, that is, it should have the potential to bind a small molecule with an appropriate binding affinity and with appropriate chemical properties.[2] Drews identified about 5000–10 000 potential target proteins, whereas Hopkins and Groom estimated the number to be 600–1500 drugable targets within the human genome.[2,3] In order to identify drugable targets, Schering examines each target protein for its potential to selectively bind a small molecule in a well-defined binding niche. A prerequisite for such an analysis is the availability of 3D-structure information about the target or a homologous protein. Target assessment has become an important decision point before entering target validation, as an early decision for or against a project can influence the costs in Research & Development dramatically: up to 40% of all costs in Research & Development arise during the target discovery stage.

In summary, pharmaceutical companies have developed different target discovery strategies according to their individual needs and the therapeutic areas they serve. In this article, rather than summarising all of the current methods used in drug discovery, we provide an overview of the Schering-specific target discovery process (Figure 1). This overview focuses on the individual steps of target identification, target assessment and target validation, including lessons that we have learnt while establishing and optimising these processes that are useful across all therapeutic areas.

[a] *Dr. U. Egner,[+] Dr. J. Krätzschmar,[+] Dr. B. Kreft,[+] Dr. H.-D. Pohlenz, Prof. M. Schneider*
Research Center Europe, Enabling Technologies
Schering AG, 13342 Berlin (Germany)
Fax: (+49) 30-468-91522
E-mail: ursula.egner@schering.de
[+] *These authors contributed equally to this work.*

 DOI: 10.1002/cbic.200400158

Martin Schneider obtained his PhD in medicinal chemistry in 1981. He was a postdoctoral researcher in molecular biology at the National Cancer Institute, Bethesda, MD, USA, and he thereafter performed his Habilitation in oncology research at the University of Regensburg. He joined Schering AG in Berlin in 1987 and later became Head of Experimental Oncology until 1998. Since then he has been Head of Enabling Technologies, including Genomics & Bioinformatics, Protein Chemistry, Assay Development & HTS and Structural Biology, in the Research Center Europe.

Hans-Dieter Pohlenz, born in 1953, studied biochemistry at the University Tübingen and subsequently moved to the Ludwig-Maximilians University in Munich where he received his PhD in 1985 for his work on the genome organisation of human immunoglobulin light-chain genes of the kappa type in the laboratory of Hans-Georg Zachau at the Institute of Physiological Chemistry. After a short postdoctoral period at the same laboratory, he joined Schering AG in Berlin in 1986 where he established a laboratory for plant biotechnology within Schering's agrochemical division. In 1992 he became head of the department of plant biochemistry and molecular biology within agrochemical research. His work focused on the identification and validation of novel targets for insecticides and herbicides and on the identification of genes conferring herbicide resistance to plants. In 1996 he moved to preclinical drug research within Schering AG and in 1999 he became responsible for the department of Genomics & Bioinformatics.

Ursula Egner, born in 1958, studied physics at the Ruprecht-Karls University in Heidelberg with an emphasis in biophysics and protein crystallography. During her PhD she determined the crystal structure of a yeast adenylate kinase at the Albert-Ludwigs University in Freiburg. After receiving her PhD in 1987, she joined the protein crystallography group of Wolfram Saenger at the Free University of Berlin, specialising in sequence analysis and homology modelling of protein–ligand complexes. In the early 1990s she moved to Schering AG, where she established a homology modeling platform and appropriate analysis tools. In 2001 she became Head of Structural Biology. The current research focus of Structural Biology is in the X-ray structure analysis of low-molecular-weight compounds and protein–ligand complexes, homology modeling and target assessment.

Bertolt Kreft, born in 1965, studied biochemistry at the Free University of Berlin. He received his PhD in 1997 for studies on cadherin-mediated cell–cell adhesion in epithelial cells performed at the University Hospital Rudolf Virchow, Charité, in Berlin. He was a postdoctoral fellow in the laboratory of Keith Burridge at the University of North Carolina, Chapel Hill, USA, supported by the Deutsche Forschungsgemeinschaft (DFG), where he studied the regulation of the actin cytoskeleton through small GTPases of the Rho family. In late 1999 he joined the department of Genomics & Bioinformatics at Schering AG, where he established a platform of cell-based assays for target validation studies in vitro. His main research interest today is to understand signalling pathways in tumour cells involved in the regulation of apoptosis and cell proliferation.

Jörn Krätzschmar, born in 1965, received his diploma in biochemistry from the Free University of Berlin in 1989 and his PhD in 1994 for the cloning, expression and functional characterisation of the vampire bat plasminogen activators in the research laboratories of Schering AG. He worked on the ADAM metalloprotease family in the laboratory of Carl Blobel at the Sloan–Kettering Institute, New York, USA, in 1994. Since 1998, he has pioneered and established the use of array gene expression profiling for various preclinical research applications at Schering AG and has directed a systematic effort to build and exploit a database of gene expression profiles to support target discovery. His main research interest in recent years has been the molecular characterisation of neoplastic and inflammatory diseases, both in animal models and in clinical situations.

Target Identification and Selection

To complement the continuous search for novel targets emanating from descriptions in the literature, we have conducted a systematic effort aimed at the identification of novel candidate targets based on the following set of gene expression criteria:

1) tissue selectivity (for example, predominant or exclusive expression in endothelial cells in reproductive organs),
2) gene regulation in ex vivo and in vitro model systems (for example, induction during activation or differentiation of lymphocytes), and
3) differential expression in samples representing human disease (for example, tumour versus normal organ, lesional skin versus normal skin, peripheral blood monocytic cells (PBMCs) of inflammatory disease versus PBMCs from healthy donors.

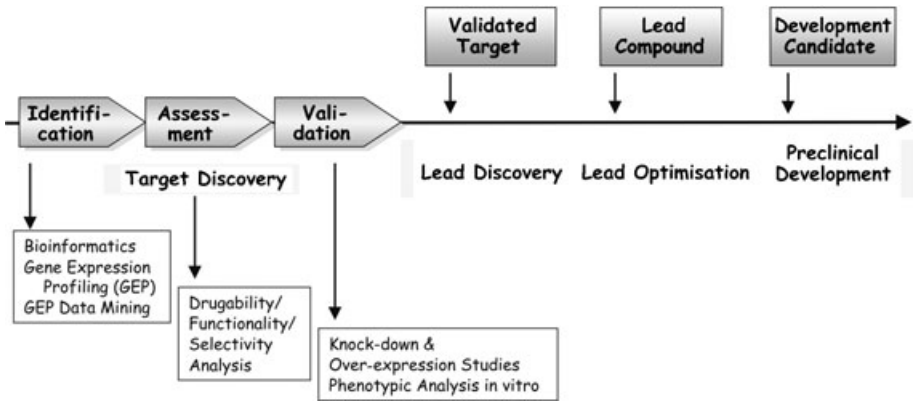


Figure 1. *The drug discovery process at Schering. The target discovery process comprises three consecutive steps: target identification, target assessment and target validation.*

The relative importance of these descriptive criteria is variable between the individual indications. As an example, the need for target-tissue specificity is linked to the severity of the disease to be treated, and it has paramount importance in the search for novel approaches to fertility control.

Obviously, additional levels of information qualifiying proteins as candidate targets exist, such as the posttranslational activation status in disease models (for example, tyrosine or serine/threonine phosphorylation and proteolytic processing resulting in activation), that are not directly accessible by large-scale gene expression profiling. However, array- or bead-based antibody technologies for multiplex analysis of protein activation status are only emerging, so array-based mRNA analysis is the only mature tool for systematic ("holistic") target selection efforts.

A Gene expression array focused on target protein families

To focus our target selection efforts on drugable target candidates, we have designed a proprietary AFFYMETRIX gene expression array representing those protein families for which small-molecule inhibitors are most likely to be found. To this end, prior to the availability of the human genome sequence, we had conducted a major bioinformatics effort to collect from various sequence databases, including nonpublic expressed sequence tag collections (INCYTE), all sequences encoding human protein kinases, protein phosphatases, proteases, nuclear receptors, and G protein-coupled receptors. Sequence redundancy was reduced by clustering of sequences; however, the criteria for cluster building were set to minimise the number of false chimeric clusters at the expense of minimising redundancy. An essential part of the sequence assembly for the final array design was the identification of major 3′-end variants for the 7500 genes to be represented, to account for transcript polymorphisms critically impacting on the detection of mRNA species by using the AFFYMETRIX technology. When our custom array was compared to the AFFYMETRIX standard arrays available at the onset of this systematic effort, the major advantages were the ability to address all relevant genes on a single array and also to cover genes that were not well represented in public sequence databases.

Database of target-gene expression profiles and data views

We have established a suite of programs, including the EXPRESSIONIST software (GeneData AG, Switzerland), as well as proprietary software modules for quality control, normalisation and condensation over large sets of AFFYMETRIX array raw data.

Using the custom array produced by AFFYMETRIX, we have to date profiled more than 600 human samples, thereby creating a database of gene expression profiles for the major target protein families. This database has become a valuable resource for systematic data-mining campaigns to identify candidate targets in a number of indications.

A broad panel of human cell lines was profiled, to enable identication of suitable model systems for in vitro target validation (see below). A large set of normal human tissue samples was included in the analysis, with multiple replicates representing major organs. This part of the data collection can best be viewed as a so-called "Array Northern", where the mean values over all samples representing one specific organ are displayed in a bar-graph format (Figure 2). This tool is also used for a summary display of differential gene expression, when mean values over multiple samples for an individual pathology, for example, breast carcinoma, are represented alongside the corresponding normal organ sample mean values.

In vitro and ex vivo systems modelling specific aspects of human pathophysiology, such as stimulation of endothelial cells, activation of T cells, and differentiation and maturation of monocytes to dendritic cells, have been profiled by using the custom array. In addition to supporting the identification of candidate targets, the value of such in vitro model data becomes most obvious when the data are overlaid with the information on differential gene expression derived from complex clinical samples, thereby allowing the possibility of assigning differential expression events seen in the clinical samples to a particular cell type or a specific cellular process. Thus, upregulation of specific genes in skin inflammation, for instance, can be attributed to the recruitment or activation of T cells in a skin lesion.

Our approach is to use large-scale gene expression profiling as a first filter and to conduct additional descriptive analyses on preselected candidate targets, including in situ hybridisation or, if a suitable antibody is commercially available, immunohistochemical analyses, to elucidate the localisation of the mRNA to a specific cell type of an organ. Therefore, in the area

of clinical samples, which has seen many applications of gene expression profiling, we have sought to obtain and profile significant numbers of intact biopsy or tissue samples representing a specific pathology, rather than conducting labour-intensive microdissection studies on a limited number of samples. Also, it has been argued that focusing gene expression profiling on the transformed epithelium of a carcinoma sample, by using microdissection, will inevitably miss disease-associated events in the neighbouring stroma that are potentially valuable, for example, for the identification of novel targets for molecular-diagnostics applications.

For the exemplary candidate target shown (Figures 2 and 3), the protease hepsin, the tissue-distribution profile, with predominant expression in the liver, and the differential mRNA expression in several carcinoma types compared to the corresponding undiseased tissues, especially in prostate carcinomas, is in line with published information, thus underlining the validity of the data derived from large-scale expression-profiling efforts[30,31] The current challenge lies therefore more in the downstream process, the analysis and expert assessment of large numbers of gene expression profiles, which is supported by dedicated data-display tools tailored to the specific purpose (Figures 2 and 3).

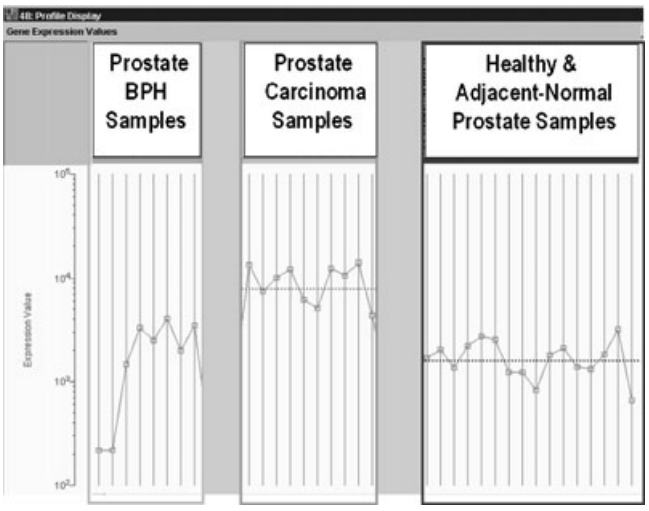


Figure 3. *EXPRESSIONIST profile display view of hepsin gene expression values across a set of prostate samples. The gene expression values for the serine protease hepsin are displayed on a logarithmic scale, by using the profile display tool of the EXPRESSIONIST software package (GeneData AG, Switzerland), for three sets of samples, namely, prostate samples from healthy individuals along with disease-free tissue adjacent to the carcinoma samples ("adjacent-normal"; right), benign prostate hyperplasia (BPH) samples (left) and prostate carcinoma samples (middle). The line connecting the individual datapoints helps in the inspection of complex profiles over hundreds of samples.*

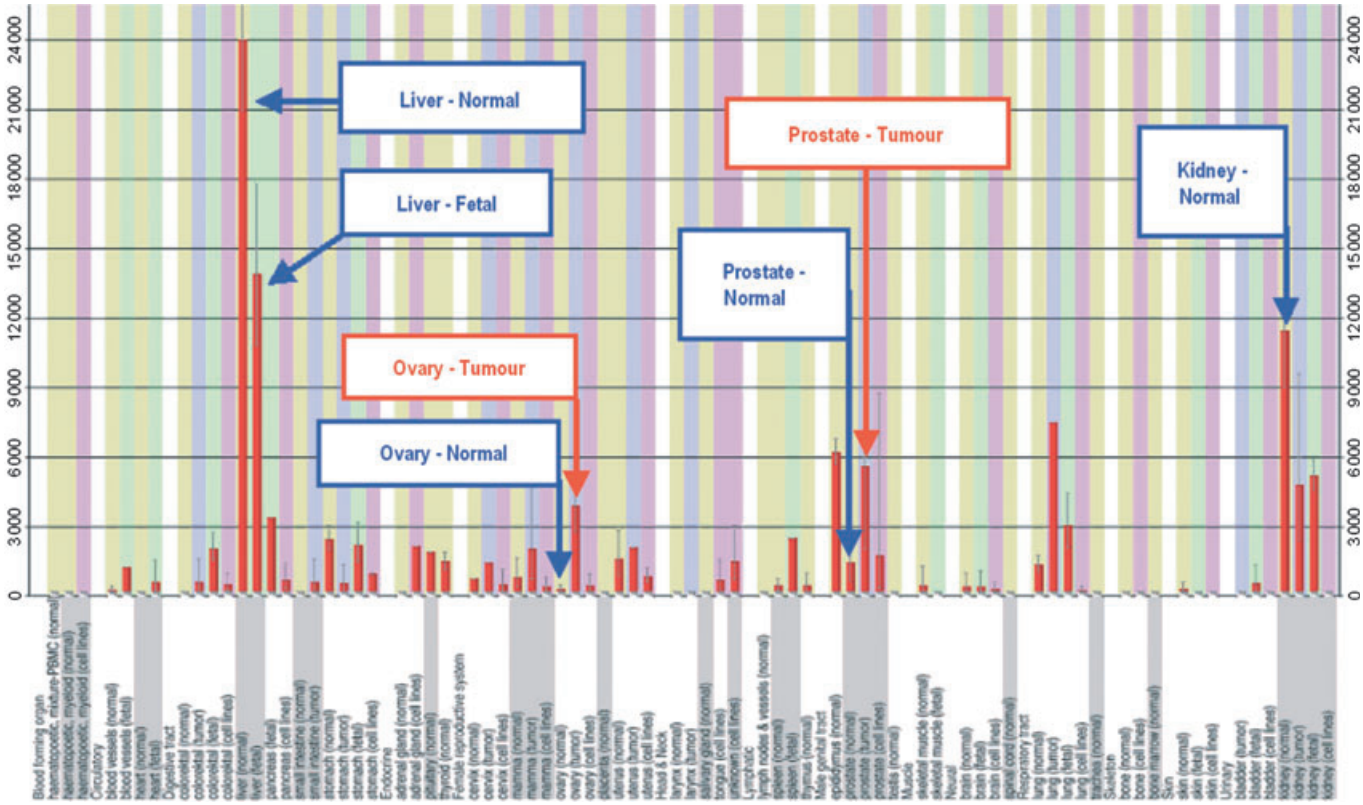


Figure 2. *An "Array Northern" view of the gene expression profile of hepsin, a serine protease. The gene expression values for a specific gene in more than 600 samples are displayed as a bar graph of the geometric mean values of the expression value on an arbitrary scale over all samples belonging to a specific class (for example. "normal adult prostate"). For a number of organs, in addition to normal adult organ samples (beige background), fetal organ samples (light green background) and cell lines (light magenta background) have been analysed and are displayed in separate bars. Where available, tumour sample data (light blue background) are displayed next to the corresponding adult normal sample data. The serine protease hepsin has been described as over-expressed in human prostate and ovarian carcinoma samples.[33,34] The profile based on our target-gene expression database clearly recapitulates the predominant expression in the liver as well as the differential expression in the carcinomas.*

Target selection data-mining campaigns

In a typical data-mining campaign for the identification of candidate targets, based on the specific target criteria for an indication, we would conduct a bioinformatics analysis by using the EXPRESSIONIST software package to identify all genes showing a specific expression pattern in sets of relevant samples, for example, searches for genes over-expressed in prostate carcinoma samples, or showing predominant or even specific expression in the normal organ (that is, the unaffected prostate). From several hundred genes fulfilling such initial criteria, an individual review of the primary list of candidates is conducted, which involves the inspection of profile display views of the expression values of such genes across all individual normal and disease samples (Figure 3) as well as the analysis of tissue distribution by using the "Array Northern" tool (Figure 2). As a result, a few dozen preselected candidate targets remain, which then become the subject of review and discussion in expert teams comprising biochemists, cell biologists and pharmacologists. Known biological activities of the candidate targets, or of closely related proteins, and their position in signal-transduction pathways are considered at this stage of the selection process; this results in the final selection of a handful of candidate targets that will be studied in more detail. The further descriptive characterisation generally includes verification of gene expression array data by quantitative PCR or Northern blotting, along with in situ or immunohistochemical analysis of cell-type distribution of the candidate mRNA or protein in normal and diseased tissue samples, in parallel with a target assessment (see below), before the candidate target ultimately enters functional in vitro validation studies (see below).

Notably, in our systematic gene expression data-mining work, we have also on several occasions found information on a potential role of existing drug targets in a previously unknown indication context; this results in the extension of target validation studies to new disease models or prompts the use of existing tool compounds in an animal model for a potential secondary indication.

In summary, the expression-profile-driven target selection process described here leads to the identification of proteins with a potential functional role in disease initiation and progression or with suitability as molecular diagnostic targets. Further descriptive and functional analyses are usually required.

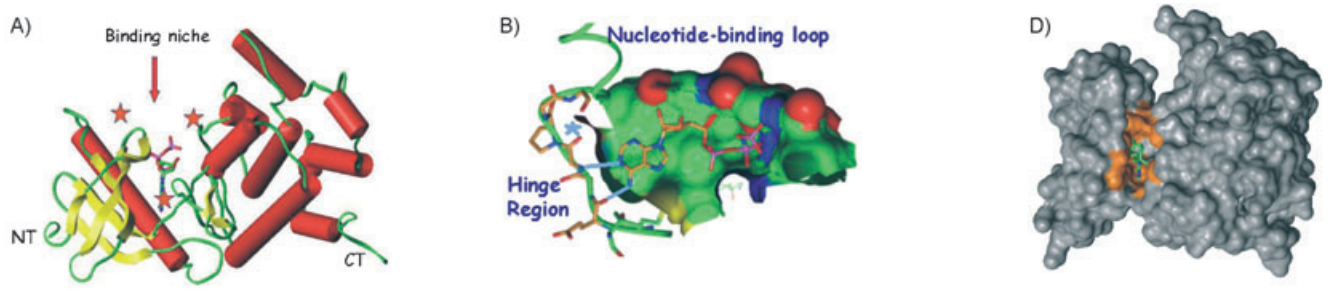
Target Assessment

After potential new targets are identified, the next step in the target discovery process at Schering is the drugability assessment of the target protein. A prerequisite for such an analysis is that 3D-structure information of the target protein itself or of a homologous protein is available. With presently more than 23 000 crystal and NMR spectroscopy structures deposited in the Protein Data Bank, the probability of finding, if not the target structure itself, than at least a homologous structure with sequence identity of 30% or higher is relatively good.[4] The assumption holds if the target belongs to one of the major enzyme and protein classes, such as nuclear receptors, protein kinases, proteases and protein phosphatases. It has to be noted, however, that there is still a lack of experimental structures for other important drugable proteins like G protein-coupled receptors.

Target assessment comprises, in addition to the prediction of drugability, the analysis of catalytic and/or functional aspects and an analysis of selectivity issues. In order to assess the drugability, the probability of identifying small-molecule inhibitors for the target protein is estimated. As a first step, we examine whether a binding niche can be identified in the target protein that—judging from its shape and size—would allow the accommodation of small-molecule inhibitors. From a structural biologist's point of view, "nondrugable" refers to a binding site that is too small, too flat, and/or too hydrophobic to allow tight and specific binding of a small molecule. With respect to protein functionality, the presence of sequence motifs contributing to enzyme activity is reviewed. In order to evaluate the potential to achieve selectivity, residues contributing to a potential binding niche are examined for their possible interactions with a ligand and for the sequence conservation of these residues in the protein family of interest. Target assessment in this sense links information from sequence space and structure space (Figure 4). Related approaches evaluating drug targets with respect to genomic and structural data have been described in the literature.[2,5,6]

Provided that a binding niche in a protein can be deduced by using information from crystal/NMR spectroscopy structures or from homology models, the amino acid residues lining this niche can be identified and mapped back to the primary sequence level (Figure 4C). When sequences are aligned for a group of related proteins, the mapping of residues from structure to sequence level allows the deduction of which residues are likely to contribute to the binding niche in the entire family. The protein kinase sequence alignment in Figure 4 depicts only those residues that have been identified as interacting with a bound small-molecule ligand in at least one 3D kinase structure. Highlighted in grey are residues that interact with the inhibitor in the respective kinase–inhibitor complex. Thus, residues listed in Figure 4C are those involved in ligand binding in various kinases. Although these residues might not have the same orientation in all structurally known kinases, they all are taken into account when analysing a target kinase with unknown structure. Sequence variablility in the (assumed) binding niche of the target protein to other related proteins of interest can thus be detected and located in structure space. This enables the identification of key areas in the 3D structure, where—upon inhibitor binding—selectivity is likely to be achieved. Inhibitors forming strong interactions with residues that are nonconserved in the protein family are more likely to be selective than inhibitors that interact predominantly with conserved side chains or protein backbone atoms.

At Schering, target assessments are performed for all potential target proteins for which a structural homologue with an overall sequence identity $>30\%$ is available. In the following section the target assessment approach is exemplified for the target family of protein kinases. Target assessments of the cata-



C)

Protein	34	35	36	37	38	39	40	42	62	63	64	89	91	92	93	105	125	127	128	129	130	131	132	134	135	138	195	196	197	198	200	202	203	205	220	221	222
p_1pkg_c-kit_hu	L	G	A	G	A	F	G	V	A	V	K	E	K	V	L	V	V	T	E	Y	C	C	Y	G	D	N	I	H	R	D	A	R	N	L	C	D	F
p_1m52_mm_abl1	L	G	G	G	Q	Y	G	V	A	V	K	E	A	V	M	V	I	T	E	F	M	T	Y	G	N	D	I	H	R	D	A	R	N	L	A	D	F
p_1fpu_mm_abl1	L	G	G	G	Q	Y	G	V	A	V	K	E	A	V	M	V	I	T	E	F	M	T	Y	G	N	D	I	H	R	D	A	R	N	L	A	D	F
p_1iep_mm_abl1	L	G	G	G	Q	Y	G	V	A	V	K	E	A	V	M	V	I	T	E	F	M	T	Y	G	N	D	I	H	R	D	A	R	N	L	A	D	F
p_1jpa_mm_epb2	I	G	A	G	E	F	G	V	A	I	K	E	S	I	M	I	I	T	E	F	M	E	N	G	S	S	V	H	R	D	A	R	N	L	S	D	F
p_1agw_hu_fgf	L	G	E	G	e	F	G	V	A	V	K	E	E	M	M	I	V	V	E	Y	A	S	K	G	N	E	I	H	R	D	A	R	N	L	A	D	F
p_1fgi_hu_fgf	L	G	E	G	e	F	G	V	A	V	K	E	E	M	M	I	V	V	E	Y	A	S	K	G	N	E	I	H	R	D	A	R	N	L	A	D	F
p_2fgi_hu_fgf	L	G	E	G	e	F	G	V	A	V	K	E	E	M	M	I	V	V	E	Y	A	S	K	G	N	E	I	H	R	D	A	R	N	L	A	D	F
p_1qcf_hu_hck	L	G	A	G	Q	F	G	V	A	V	K	E	N	V	M	V	I	T	E	F	M	A	K	G	S	D	I	H	R	D	R	A	N	L	A	D	F
p_2hck_hu_hck	L	G	A	G	Q	F	G	V	A	V	K	E	N	V	M	V	I	T	E	F	M	A	K	G	S	D	I	H	R	D	R	A	N	L	A	D	F
p_1ad5_hu_hck	L	G	A	G	Q	F	G	V	A	V	K	E	N	V	M	V	I	T	E	F	M	A	K	G	S	D	I	H	R	D	R	A	N	L	A	D	F
p_1k3a_hu_igf1	L	G	Q	G	S	F	G	V	A	I	K	E	S	V	M	V	V	M	E	L	M	T	R	G	D	S	V	H	R	D	A	R	N	M	G	D	F
p_1i44_hu_irk	L	G	Q	G	S	F	G	V	A	V	K	E	S	V	M	V	V	M	E	L	M	A	H	G	D	S	V	H	R	D	A	R	N	M	G	D	F
p_1ir3_hu_irk	L	G	Q	G	S	F	G	V	A	V	K	E	S	V	M	V	V	M	E	L	M	A	H	G	D	S	V	H	R	D	A	R	N	M	G	D	F
p_1qpd_hu_lck	L	G	A	G	Q	F	G	V	A	V	K	E	N	L	M	V	I	T	E	Y	M	E	N	G	S	D	I	H	R	D	R	A	N	L	A	D	F
p_1qpc_hu_lck	L	G	A	G	Q	F	G	V	A	V	K	E	N	L	M	V	I	T	E	Y	M	E	N	G	S	D	I	H	R	D	R	A	N	L	A	D	F
p_2src_hu_src	L	G	Q	G	C	F	G	V	A	I	K	E	Q	V	M	V	I	T	E	Y	M	S	K	G	S	D	V	H	R	D	R	A	N	L	A	D	F
p_1m17_hu_egfr1	L	G	S	G	A	F	G	V	A	I	K	E	Y	V	M	C	L	T	Q	L	M	P	F	G	C	D	V	H	R	D	A	R	N	L	T	D	F
p_1qpe_hu_lck	L	G	A	G	Q	F	G	V	A	V	K	E	N	L	M	V	I	T	E	Y	M	E	N	G	S	D	I	H	R	D	R	A	N	L	A	D	F
ZAP-70	L	G	C	G	N	F	G	V	A	I	K	E	Q	I	M	V	L	M	E	.	M	A	G	G	P	K	V	H	R	D	A	R	N	L	S	D	F

Figure 4. *Target assessment approach at Schering. When a crystal structure or homologous structure of the target protein (for example, ZAP-70) is available (for example, kinase lck; PDB entry code: 1qpe;[13] A), the binding niche of this protein and other members of the same family are analysed (B) with respect to binding niches, inhibitor binding and contact residues of the inhibitor to the protein. Thus, a list of the binding niches is compiled that highlights the kind of interaction in the proteins, for example, interaction through hydrogen bonds or van der Waals contacts (C). Amino acids on a grey background interact through side chains with the inhbitors in the respective protein–inhibitor complex, while those in bold and underlined contact the ligands with their main-chain atoms. As a final step, a model of the target protein is calculated (D; example shows the model for Zap-70 on the basis of the lck–PP2 complex (PDB entry code: 1qpe)) and the interacting residues are highlighted in a surface presentation (orange on an overall surface of grey).*

lytic domain of protein kinases have the advantage that a broad basis of structural information is available. In December 2003, 224 structures of the catalytic domain of protein kinases were deposited in the Protein Data Bank.[4] Of these 224 structures, 159 represent kinases in complex with a low-molecular-weight compound (54 with adenosine triphosphate (ATP) and derivatives, 12 with staurosporine or derivatives thereof, and 93 with other small-molecule inhibitors). These 159 kinase structures comprise complex structures of 17 different serine/threonine (Ser/Thr) protein kinases and of 12 different tyrosine (Tyr) protein kinases. We have carefully analysed the binding niches of these structures to determine a) which amino acid residues contribute to the binding niche, b) the flexibility of residues and loops contributing to the binding niche, and c) whether the residues interact with the ligand through main-chain or side-chain atoms.

Protein kinases have a highly conserved catalytic core of about 300 residues (Figure 4).[7–9] This core structure consists of two lobes, with the binding site for ATP or ATP-competitive inhibitors located between these lobes. The orientation of the two lobes relative to each other and the flexibility of the nucleotide-binding loop in the active site (the name reflects the interactions with the phosphate moieties of ATP) often varies between different ligand scaffolds.

Another region of interest in inhibitor binding is the so-called hinge region between the N- and C-terminal lobes of the catalytic domain. Both the natural ligand ATP and the ATP-competitive inhibitors interact with this region through up to three hydrogen bonds. The residue preceding the hinge region, the so-called gate-keeper residue, is often of small size (Thr or Val; if not stated otherwise, amino acids are described by their three-letter code) in Tyr kinases whereas it is larger (often Phe, Met or Leu) in Ser/Thr kinases.

In addition to contributing to the orientation of ATP, the catalytic loop and the activation loop in kinases are responsible for the positioning of the substrate. In almost all kinases, sequence motifs such as His–Arg–Asp at the start of the catalytic loop and Asp–Phe–Gly at the beginning of the activation loop are conserved, as they are involved in the correct orientation of the natural ligand ATP in the binding site.

The target assessment approach is demonstrated in more detail by using the protein kinase ZAP-70 as an example. ZAP-70 is involved in T-cell activation and belongs to the family of nonreceptor tyrosine kinases. The 70 kDa protein (619 residues,

Swissprot entry code: za70_human, Isakov1996) contains two SH2 domains, arranged in tandem, and a C-terminal catalytic protein kinase domain. The NMR spectroscopy and crystal structures of the apo and peptide-bound SH2 tandem domains have been determined and several examples of ZAP-70 inhibitors targeting the SH2 domains have been published.[10–12] From sequence alignments, the kinase domain of ZAP-70, for which no 3D-structure information is available so far, is thought to be located between residues 338 and the C-terminal end of the protein.

In Table 1 the results of the target assessment for the ZAP-70 kinase domain are summarised. In ZAP-70, all sequence motifs contributing to the kinase catalytic machinery (for example, the Asp–Phe–Gly and His–Arg–Asp motifs, etc) are conserved and classify ZAP-70 as a Tyr kinase. The only exception is the gate-keeper residue, which is large in ZAP-70 (Met414) but usually small in other Tyr kinases. The highest sequence homology to structurally known kinases is found for the catalytic domain of FAK (focal adhesion kinase; 43% sequence identity; PDB entry code: 1mp8), followed by Eph-A2 receptor tyrosine kinase (41%, 1mqb), EGF receptor kinase (40%, 1m14) and lck (40%, 1qpe).[13–15] Given these high levels of sequence identity and the highly conserved common fold of protein kinases, a binding niche similar in size and shape to other protein kinases can be expected with high probability for ZAP-70.

Calculating sequence identities based only on residues in the binding niche, the closest homologues to ZAP-70 among a list of 160 randomly selected kinases are members of the Eph-B receptor family, which exhibit sequence identities of 50% in this region, while the sequence identity of lck is only 44%. Figure 4 shows a sequence alignment of residues in the binding niche of the tyrosine receptor subfamily, which includes FAK, c-abl, Eph-A2 receptor, Eph-B2 receptor and EGFR. The similarity in the binding niche of ZAP-70 compared to a randomly selected list of 165 kinases is relatively low (<50%), a fact suggesting that it should be possible to find selective inhibitors for ZAP-70. A detailed analysis of the active-site residues differing between ZAP-70 and these homologous kinases (not presented here) may give further insights into whether it will be possible to develop inhibitors selective for ZAP-70 alone.

Many of the residues involved in coordinating ATP are conserved among kinases, thereby leading to a higher level of sequence identity in the binding niche than the overall sequence identity of protein kinases of approxiamtely 30%. For example, the sequence identity between the binding niche of the Ser/Thr kinase cdk2 and the Tyr kinase c-src is only 42% while that between c-src and the Tyr kinase abl amounts to 65%. Accordingly, the target assessment will predict that the probability of encountering selectivity problems between either cdk2 and c-src or abl and c-src is rather low. Examples of inhibitors binding to c-src but not to cdk2 are PP1 and SU6656, which exhibit IC_{50} values for c-src of 170 nM and 280 nM, respectively, while no inhibition of cdk2 can be detected ($IC_{50} > 10\,000$ nM).[16,17] With a sequence identity of 65% in the binding niche between c-src and abl, inhibitors can be found with similar IC_{50} values, like PP1 (170 nM and 250 nM for c-src and abl, respectively), or already differing IC_{50} values, like the compound SU6656 (280 nM and 1740 nM for c-src and abl, respectively). A careful analysis is necessary in order to obtain selective compounds. An exact prediction of which compound might lead to selectivity problems is beyond the scope and capabilities of a target assessment. Still, a target assessment in the early phase of a project can alert scientists to critical issues relating to the functionality, selectivity and drugability aspects of a potential target protein. It can thus contribute to the decision

Table 1. *Summary of the structural biology target assessment for the protein kinase ZAP-70.*[a]

Characteristics	Comments relating to target kinase ZAP-70	Conserved in ZAP-70/Ser/Thr or /Tyr?
Functionality		
sequence motif: nucleotide-binding loop, GxGxxG	**G**C**G**N**F**G (aa 345–350)	yes (both)
sequence motif: conserved interaction between Lys and Glu in helix C	VAI**K** (aa 366–369) and **E** (aa 386)	yes (both)
hinge region: gate-keeper residue	Met414	Ser/Thr
hinge region: size of hinge region	large, loop size as in lck	Ser/Thr
catalytic loop: Prosite motif[b]	**FVH**R**DLAA**N**VLL** (aa 457–469)	Tyr
activation loop: DFG	**DFG** (aa 479–481)	yes (both)
activation loop: xxxxxAPE	P**LKWYAPE** (aa 502–509)	Tyr
Drugability		
structural homology: overall sequence identity	43% FAK (PDB: 1mp8), 41% Eph-A2 RTK (PDB: 1mqb), 40% EGFR kinase (PDB: 1m14), 40% lck (PDB: 1qpe)	Tyr
existence and size of binding niche	niche available, similar size to that in other protein kinases	
Selectivity		
sequence identity in binding site	50% to Eph-B1, -B2, -B3, -B4 and -B8 RTK[c] 44% to lck (calculated from a set of 165 kinases)	
homologues in kinase kinome[36]	Tyr kinase family: syk subfamily (followed by abl and FAK subfamilies)	Tyr

[a] Single-letter amino acid (aa) codes are used in the sequences, x = any amino acid, bold type indicates a residue conserved in the consensus sequence. [b] Consensus pattern for Tyr kinases from the Prosite database:[35] [LIVMFYC]-x-[HY]-x-D-[LIVMFY]-[RSTAC]-xx-N-[LIVMFYC]$_3$ where D is an active-site residue. [c] RTK = receptor tyrosine kinase.

about whether a target candidate should enter the target validation process.

Since the establishment of target assessments at Schering three years ago, a number of potential target proteins did not proceed to target validation due to a (predicted) lack of functionality (which in addition was confirmed experimentally). For targets with selectivity problems, that is, with binding sites that are more than 95% identical to that of homologous proteins, additional data relating to the target's tissue distribution are analysed before a "no go" decision is taken.

Target Validation

Impressive technical advances such as large-scale sequencing efforts and systematic functional genomics studies in model organisms have led to a massive increase in potential drug targets during recent years.[2] As a result, many target discovery groups are facing the same dilemma, that is, they are confronted with a large number of targets which, in most cases, have been identified solely by virtue of their differentiated regulation under pathophysiological conditions (see above). Even though many of these potential targets may contribute in some way to disease phenotypes, further functional characterisation is required to identify key switches in biochemical pathways as appropriate intervention points for drug treatment. The process of target validation aims at identifying these switches exactly by demonstrating that a target plays an essential role in a disease-relevant cellular process.

It is needless to say that there is no standard process for target validation and pharmaceutical companies are applying various techniques according to their indication-specific needs (reviewed in ref. [1]). Five years ago, we had already agreed upon sets of essential criteria that have to be fulfilled by a candidate gene in order for it to constitute a valid target for a particular indication. In principle, there are two ways to establish a link between a target gene and a disease-relevant phenotype.[1] In contrast to the identification of potential targets through random phenotypic screens, that is, "forward genetics", Schering has settled for the alternative strategy of "reverse genetics", which seeks to unravel the specific molecular function of a candidate gene within a given physiological process of interest. In this approach, the function of a target protein is first blocked, either by the introduction of a dominant-interfering mutation or by the specific suppression of gene expression. The resulting loss-of-function phenotypes are subsequently monitored, thereby making it possible to link the inhibition of precisely one target to the observed phenotypic changes.

One of the main challenges in target validation is the establishment of appropriate model systems, which mimic the in vivo situation and, thus, are indeed predictive of disease. Furthermore, these model systems have to be amenable to medium- or even high-throughput applications in order to analyse a sufficient number of potential targets in parallel. For this reason, we have decided to conduct target validation studies in vitro in cellular systems, which have been very helpful in the past in elucidating signalling pathways that govern essential physiological processes such as cell proliferation or the regulation of survival and apoptosis. Suitable in vitro models for target validation have to meet certain criteria, that is, 1) the cells have to express the gene of interest to detectable levels, 2) the cellular model has to mimic the in vivo situation, thereby allowing the analysis of a well-defined and disease-relevant phenotype, and 3) the cells need to be amenable to experimental manipulation. In this respect, much has been learned from studies in immortalised cell lines, which are either derived from tumours or which have been established by the stable introduction of viral oncogenes. Still, these stable cell lines frequently exhibit aberrant properties, since the process of immortalisation correlates with gross alterations of the karyotype and the acquirement of a dedifferentiated phenotype.[18] Therefore, we prefer to perform target validation studies in primary cells derived from relevant human tissue whenever possible, as the human situation is best reflected by these nontransformed cells.

Independently of the cell system used, one major hurdle in functional genomics studies is the delivery of target validation tools like antisense oligonucleotides, short interfering RNAs (siRNAs), expression vectors or proteins (for example, blocking antibodies) into cells. The introduction of oligonucleotides and plasmids is typically achieved by using cationic lipids or polymers.[19] Gene delivery through lipofection is simple and fast, and there are numerous specialised transfection reagents available from different vendors. However, this delivery system is restricted to dividing cells and exhibits a number of limitations, including sometimes severe toxicity and, most importantly, only mediocre transfection efficiencies, especially in primary cells. In our hands, the Nucleofector technology developed by Amaxa Biosystems (Cologne, Germany) turned out to be the most powerful delivery system. This gentle electroporation method is able to deliver DNA or RNA molecules directly into the nucleus without inducing severe cell damage and apoptosis. As a result, it is possible to transfect even resting cells[20] and the time between electroporation and phenotypic analysis is reduced significantly. Importantly, nucleofection has been optimised for the efficient delivery of DNA and RNA molecules into hard-to-transfect cells, and we have consistently experienced high transfection efficiencies (up to 80%) in primary human cells of various origin (endothelial cells, prostate stromal cells, T cells). Another means of delivering genes into cells is viral transduction by using adeno-, retro- or lentiviruses.[21] While retroviruses can only be used to transduce proliferating cells, adeno- and lentiviruses are also able to drive gene expression in fully differentiated and nondividing cells. Viral transduction is restricted, however, to certain cells or tissues, depending on the tropism of the virus. Furthermore, viruses are known to trigger cellular defence mechanisms, such as the interferon (IFN) response, which may obscure the results of further phenotypic analyses.[22]

Loss-of-function phenotypes are frequently induced by the inhibition of mRNA expression and the resulting knock-down of the endogenous protein. This approach requires no further knowledge of the potential target besides limited nucleic acid sequence information, and there are several methods available to suppress gene expression.[1,23] In our experience, the most

effective knock-down technology is RNA interference (RNAi), which is part of a universal defence mechanism triggering generalised translational repression and apoptosis in response to viral infections.[24] When siRNAs of about 21 nucleotides in length are used, it is possible to induce the specific degradation of the corresponding mRNAs without activation of the IFN response.[23,25] Although recent reports suggest that siRNAs and short-hairpin (sh) RNAs might be able to activate mediators of the IFN response,[26,27] we have not encountered similar problems in our internal studies with synthetic siRNAs (B.K., unpublished data). Compared to antisense oligonucleotides, the identification of functional siRNAs is relatively easy—in our experience about 30–50% of siRNAs are effective—and several commercial vendors are offering a wide variety of RNAi tools, including functionally validated siRNAs. As this method can be applied to the simultaneous characterisation of multiple candidate genes, Schering is using RNAi as a first filter to sift through a larger number of target candidates in parallel. Still, one has to keep in mind that, similarly to other knock-down technologies, RNAi is associated with a number of uncertainties. There is no general rule for which level of mRNA knock-down has to be achieved in order to induce a sufficient reduction of the target protein. Depending on the rate of transcription and the stability of a particular target protein, the knock-down kinetics, and thus the optimal time for the observation of phenotypic changes, can vary substantially from gene to gene.

Due to the aforementioned limitations, target validation at Schering does not rely solely on the results of RNAi experiments (Figure 5). Instead, a complementary approach is used

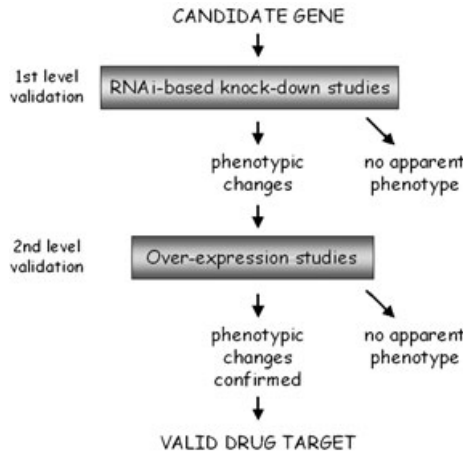


Figure 5. *The Schering target validation process.*

to further characterise those candidate genes for which a desired phenotype has been observed consistently in knock-down studies. In this approach, the function of the endogenous protein is blocked following over-expression of a functionally impaired, that is, enzymatically inactive, "dominant-negative" (dn) mutant, which competes for the interaction with upstream activators and/or downstream effectors/substrates of the target protein. Since the expression of transgenes to very high levels can produce artefacts in some cases, it is important to compare the effects induced by either the wild-type or the mutant target protein. For enzymes such as kinases the generation of dn mutants is straightforward, because critical amino acid residues required for the enzymatic activity are well conserved and mutation of a only single lysine residue within the catalytic domain is generally sufficient to abolish the kinase function.[28] Ideally, the phenotype previously observed in RNAi studies can be confirmed in cells over-expressing a dn form of the target protein. In contrast to knock-down experiments, where the entire protein is missing, this approach provides information on whether the enzymatic function of the potential target is indeed required for its role in a particular cellular process. Through this analysis it is now possible to discriminate between scaffold proteins acting merely as structural components of larger signalling complexes and valid targets whose activity can be influenced in a desired way by a low-molecular-weight inhibitor.

For instance, the kinase suppressor of Ras (KSR), a protein originally identified in genetic screens for molecular components downstream of Ras, contains a predicted C-terminal kinase domain, even though KSR lacks several key properties of a protein kinase, including a conserved lysine residue in the ATP-binding niche. It has been proposed that—rather than acting as a kinase—KSR may function as a scaffolding protein that coordinates Ras/Raf/MAP kinase signalling through the assembly of an activated signalling complex.[29] This is only one example for a protein fulfilling its cellular function independently from the predicted enzymatic activity, and we have encountered several similar cases in our own search for novel drug targets.

The analysis of cells expressing either wild-type or mutant forms of a target protein is performed in transiently transfected cells. As opposed to the generation of stable cell lines, transient transfection is fast, yields higher levels of transgene expression and prevents secondary effects, such as negative selection for cells expressing no or low amounts of the target protein. In addition, the transient over-expression of dn mutants eliminates the possibility that cells escape this immediate impact through the activation of compensatory mechanisms. As transient transfections result in mixed populations of cells expressing the respective transgene to different levels, it is necessary to analyse potential phenotypic changes in single cells. This is achieved through the application of fluorescence microscopy-based assays, which allow the analysis of a wide variety of disease-relevant phenotypes (for example, cell proliferation, mitosis/cell-cycle regulation, senescence, cell death or apoptosis, morphological changes, cell migration, activation of signalling pathways). As an example, the phenotypic changes induced after inhibition of a target protein regulating critical steps in mitosis are depicted in Figure 6.

It is obvious that there are indications for which in vitro target validation is not an option. Physiological processes, particularly those involving complex interactions of different cell types over time, can only be analysed within the context of intact organisms, and mouse genetics, that is, the generation of transgenic or knockout (KO) mice, has become the method

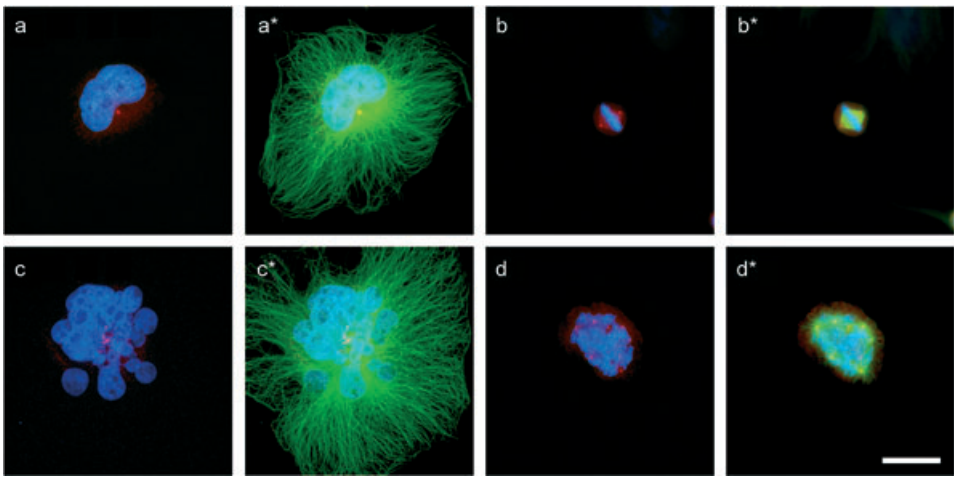


Figure 6. *Inhibition of a target protein regulating mitosis that induces mitotic slippage, centrosome amplification and the formation of multinuclear cells. Depicted are control cells (a, a*, b, b*) and cells in which the target has been inhibited (c, c*, d, d*). Cells are shown either in the interphase (a, a*, c, c*) or in mitosis (b, b*, d, d*). Nuclei have been stained with Hoe33258 (blue) and centrosomes were visualised by staining the centrosomal marker γ-tubulin (red). Microtubules (α-tubulin) were stained in green and an overlay of all stainings is shown in pictures marked with an asterisk. The scale bar represents 20 μm.*

of choice.[30] However, this approach is facing a number of problems, such as embryonic lethality, developmental phenotypes, the induction of compensatory mechanisms and the more general question of whether a KO throughout development indeed reflects the consequences of blocking a target protein's function in the adult organism with specific inhibitors. Some of these problems can be overcome by using conditional KO models and inducible or tissue-specific promoters,[31] but target validation through mouse genetics remains an expensive and rather time-consuming approach. Recently, mouse knock-in technology has been developed that enables the effect of drug inhibition of a protein kinase in vivo to be mimicked, by replacing the endogenous wild-type enzyme with a specific ATP-binding pocket variant that can specifically be inhibited by a tool compound inert against all natural kinases.[28]

One attractive alternative approach to in vivo target validation is the generation of transgenic mice expressing shRNAs that are subsequently processed to yield functional siRNAs, thereby inducing the mRNA knock-down of specific genes. Lentiviral delivery systems have successfully been used to analyse the phenotype of mouse embryos completely derived from embryonic stem cells stably expressing shRNAs.[32] Since the expression of such shRNA constructs can be controlled by inducible and tissue-specific inhibitors, it is conceivable that these transgenic RNAi systems will soon become the technology of choice for the validation of potential targets in vivo.

Summary and Outlook

To minimise the attrition rates of drug development projects in later phases, pharmaceutical companies have developed strategies and processes in the field of target discovery to select and characterise the most suitable candidate targets, before embarking on lead identification and lead optimisation for only the validated targets.

We have outlined the target discovery process implemented at Schering AG for work on kinases over the last few years. This process consists of the following three areas:

1) target selection, based on a combination of gene expression criteria and relying on a dedicated data resource of gene expression profiles for clinical samples and indication-relevant in vitro model systems, to identify candidate targets with a specific tissue distribution and presence in human pathology,
2) target assessment, exploiting the three-dimensional structure of proteins for detailed binding-site analysis to estimate the drugability of the protein for small-molecule inhibitor binding as well as selectivity profiles, and
3) target validation, providing evidence for a functional role in an in vitro model system of human disease, thus corroborating the biological hypothesis underlying the therapeutic concept around the candidate target.

This rational approach to target discovery, as a prerequisite for lead discovery, ensures that new therapeutic targets fulfil a set of general criteria, as well as indication-specific, descriptive and functional ones, and should ultimately maximise the likelihood for achieving target-selective inhibition by small-molecule inhibitors with minimal in vivo side effects and a therapeutic effect based on a sound biological hypothesis.

Note added in proof

After submision of the manuscript, the crystal structure of the catalytic domain of ZAP-70 was solved (L. Jin, S. Pluskey, E. C. Petrella, S. M. Cantin, J. C. Gorga, M. J. Rynkiewicz, P. Pandey, J. E. Strickler, R. E. Babine, D. T. Weaver, K. J. Seidl, *J. Biol. Chem.* **2004**, *279*, 42818). The available structure does not change the predictions.

Acknowlegements

The authors would like to thank Bertram Weiss and Henrik Seidel for their pioneering work in sequence collection and assembly for the custom-array design, Roman Hillig and Martina Schäfer for their valuable input into the target assessment process, and Anette Sommer and Luisella Toschi for their valuable contributions to the target selection and target validation projects.

Keywords: functional genomics · gene technology · protein structures · target discovery · target validation

[1] M. A. Lindsay, *Nat. Rev. Drug Discovery* **2003**, *2*, 831.
[2] A. L. Hopkins, C. R. Groom, *Ernst Schering Res. Found. Workshop* **2003**, 11.
[3] J. Drews, *Science* **2000**, *287*, 1960.
[4] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, P. E. Bourne, *Nucleic Acids Res.* **2000**, *28*, 235.
[5] P. R. Caron, M. D. Mullican, R. D. Mashal, K. P. Wilson, M. S. Su, M. A. Murcko, *Curr. Opin. Chem. Biol.* **2001**, *5*, 464.
[6] M. Murcko, P. Caron, *Drug Discovery Today* **2002**, *7*, 583.
[7] M. Huse, J. Kuriyan, *Cell* **2002**, *109*, 275.
[8] G. Scapin, *Drug Discovery Today* **2002**, *7*, 601.
[9] M. Cherry, D. H. Williams, *Curr. Med. Chem.* **2004**, *11*, 663.
[10] R. H. Folmer, S. Geschwindner, Y. Xue, *Biochemistry* **2002**, *41*, 14176.
[11] M. H. Hatada, X. Lu, E. R. Laird, J. Green, J. P. Morgenstern, M. Lou, C. S. Marr, T. B. Phillips, M. K. Ram, K. Theriault, M. J. Zoller, J. L. Karas, *Nature* **1995**, *377*, 32.
[12] C. Garcia-Echeverria, *Curr. Med. Chem.* **2001**, *8*, 1589.
[13] X. Zhu, J. L. Kim, J. R. Newcomb, P. E. Rose, D. R. Stover, L. M. Toledo, H. Zhao, K. A. Morgenstern, *Structure Folding Des.* **1999**, *7*, 651.
[14] J. Nowakowski, C. N. Cronin, D. E. McRee, M. W. Knuth, C. G. Nelson, N. P. Pavletich, J. Rogers, B. C. Sang, D. N. Scheibe, R. V. Swanson, D. A. Thompson, *Structure (Cambridge, MA, US)* **2002**, *10*, 1659.
[15] J. Stamos, M. X. Sliwkowski, C. Eigenbrot, *J. Biol. Chem.* **2002**, *277*, 46265.
[16] J. Bain, H. McLauchlan, M. Elliott, P. Cohen, *Biochem. J.* **2003**, *371*, 199.
[17] M. Warmuth, R. Damoiseaux, Y. Liu, D. Fabbro, N. Gray, *Curr. Pharm. Des.* **2003**, *9*, 2043.
[18] C. Horrocks, R. Halse, R. Suzuki, P. R. Shepherd, *Curr. Opin. Drug Discovery Dev.* **2003**, *6*, 570.
[19] M. E. Davis, *Curr. Opin. Biotechnol.* **2002**, *13*, 128.
[20] H. I. Trompeter, S. Weinhold, C. Thiel, P. Wernet, M. Uhrberg, *J. Immunol. Methods* **2003**, *274*, 245.
[21] C. Mah, B. J. Byrne, T. R. Flotte, *Clin. Pharmacokinet.* **2002**, *41*, 901.
[22] N. Grandvaux, B. R. tenOever, M. J. Servant, J. Hiscott, *Curr. Opin. Infect. Dis.* **2002**, *15*, 259.
[23] K. S. Lavery, T. H. King, *Curr. Opin. Drug Discovery Dev.* **2003**, *6*, 561.
[24] G. N. Barber, *Cell Death Differ.* **2001**, *8*, 113.
[25] D. M. Dykxhoorn, C. D. Novina, P. A. Sharp, *Nat. Rev. Mol. Cell Biol.* **2003**, *4*, 457.
[26] C. A. Sledz, M. Holko, M. J. de Veer, R. H. Silverman, B. R. Williams, *Nat. Cell Biol.* **2003**, *5*, 834.
[27] A. J. Bridge, S. Pebernard, A. Ducraux, A. L. Nicoulaz, R. Iggo, *Nat. Genet.* **2003**, *34*, 263.
[28] F. R. Papa, C. Zhang, K. Shokat, P. Walter, *Science* **2003**, *302*, 1533.
[29] T. Raabe, U. R. Rapp, *Sci. STKE* **2002**, *136*, PE28. http://stke.sciencemag.org/
[30] B. P. Zambrowicz, A. T. Sands, *Nat. Rev. Drug Discovery* **2003**, *2*, 38.
[31] E. Bockamp, M. Maringer, C. Spangenberg, S. Fees, S. Fraser, L. Eshkind, F. Oesch, B. Zabel, *Physiol. Genomics* **2002**, *11*, 115.
[32] D. A. Rubinson, C. P. Dillon, A. V. Kwiatkowski, C. Sievers, L. Yang, J. Kopinja, D. L. Rooney, M. M. Ihrig, M. T. McManus, F. B. Gertler, M. L. Scott, L. Van Parijs, *Nat. Genet.* **2003**, *33*, 401.
[33] J. Luo, D. J. Duggan, Y. Chen, J. Sauvageot, C. M. Ewing, M. L. Bittner, J. M. Trent, W. B. Isaacs, *Cancer Res.* **2001**, *61*, 4683.
[34] H. Tanimoto, Y. Yan, J. Clarke, S. Korourian, K. Shigemasa, T. H. Parmley, G. P. Parham, T. J. O'Brien, *Cancer Res.* **1997**, *57*, 2884.
[35] N. Hulo, C. J. Sigrist, V. Le Saux, P. S. Langendijk-Genevaux, L. Bordoli, A. Gattiker, E. De Castro, P. Bucher, A. Bairoch, *Nucleic Acids Res.* **2004**, *32*, D134.
[36] G. Manning, D. B. Whyte, R. Martinez, T. Hunter, S. Sudarsanam, *Science* **2002**, *298*, 1912.

Received: May 17, 2004

