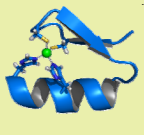
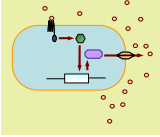
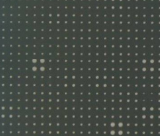
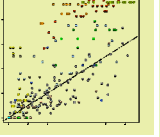


6 Conclusions

Chapter 1	Chapter 2	Chapter 3	Chapter 4	Chapter 5	Chapter 6
	Spatial	Quorum sensing	Chemo-genomics	Descriptor relationships	
Introduction					Conclusions and perspectives
	Atomic level	Pathway level	Proteome level	Cellular level	

Contents of chapter 6

Contents of chapter 6.....	1
Implications	2
Putting together the pieces.....	2
Genomic location matters	2
Implications for drug discovery and biotechnology.....	3
Drugs to disrupt quorum sensing.....	6
A potential biotechnological use of quorum sensing	6
Chemogenomic approaches to drug discovery	7
The drug-COP score	8
Towards a scoring system	9
The future of the drug-COP	10
Conclusion	11
References.....	12

Implications

In the following, I discuss the results of chapters 2 to 5 and show what directions future work may take. First, I demonstrate that the approach I have used in chapter 3 to identify novel quorum sensing systems could be extended to other cellular subsystems. What distinguishes my approach is that it takes the genomic location of genes into account. In the next section, I show what direction further work on the genomic location of genes may take. Most of the descriptors I analysed in chapter 5 show some genomic position effect. For the remainder of the chapter, I discuss how the methods and findings I have presented on chemogenomics and on quorum sensing can be used to identify leads for drug discovery.

Putting together the pieces

In chapter 3 on quorum sensing, I used the same method to identify three different quorum sensing systems in fully sequenced genomes. In the case of the γ -butyrolactone quorum sensing system in *Rhodococcus*, the *agr* system in firmicutes, and the aromatic alcohol quorum sensing system in yeasts, at least some of the components of the system needed to produce, export, sense, and respond to a quorum sensing signal were known. The basic approach to identifying new systems is to establish whether in related genomes some or all of the systems components are conserved. What inspired this approach was previous work that showed that genes that show a similar pattern of presence and absence across genomes are more likely to interact physically (Pellegrini, Marcotte et al. 1999). The second type of information to take into account is the genomic location of these components. For example, in case of the *agr* system, all its components are encoded in genomic proximity. By combining these two types of information, I was able to identify new quorum sensing systems in bacteria. I believe that this method could also be extended to cellular subsystems that are not quorum sensing systems.

Genomic location matters

When identifying cellular sub-systems such as the *agr* cell-to-cell communication system, one important piece of information to take into account is the genomic location of genes. In the case of the *agr* and γ -butyrolactone quorum sensing systems, their components tend to be encoded in close genomic neighbourhood. In the case of the *S. cerevisiae* aromatic alcohol quorum sensing system, the genes regulated by the system tend to be clustered, hence showing that genomic location also matters in eukaryotes. Although it is known that eukaryotic transcriptional

regulation is complex and requires an intricate coordination of several molecular events both in space and time (Madan Babu, Janga et al. 2008), whether the complexity of this process constrains genome organization used to be unknown. Recently, evidence for the existence of a higher-order organization of genes across and within chromosomes that is constrained by transcriptional regulation has emerged (Janga, Collado-Vides et al. 2008). This points to the possibility that specific organization of genes that allow for efficient control of transcription within the nuclear space has been selected during evolution (Janga, Collado-Vides et al. 2008).

Using the descriptors of the database assembled in chapter 5, and a similarly unbiased approach, it is possible to show that the majority of descriptors are distributed in a non-random manner across the *S. cerevisiae* chromosomes. To my knowledge, no unbiased study of the genomic location of different descriptors has been attempted until now. Correlating the values of descriptors for each gene with that of its neighbours shows that the majority of descriptors are non-randomly distributed (table 6-1). Plots similar to that in figure 6-1 showing the genomic distribution of descriptors along the *S. cerevisiae* chromosomes can be viewed on <http://www.mrc-lmb.cam.ac.uk/genomes/awuster/chrlocation/>. I have averaged values across windows of 1 gene, 5 genes, 15 genes, 45 genes, and 145 genes. I expect that uncovering such higher-order organization of genes in other eukaryotes will provide insights into nuclear architecture, have implications in genetic engineering experiments to rationally identify genomic regions for integrating a gene, and in understanding disease conditions that involve chromosomal aberrations.

Implications for drug discovery and biotechnology

Notwithstanding the fact that we now have more information on biological systems and aetiologies than ever before, the number of new drugs that are developed remains low or even declines. The pharmaceutical industry now spends more on research and development, but produces fewer new drugs, than it did 20 years ago (PWC 2007). There may be several reasons for this (Lansbury 2004).

One reason may be the deficiencies in the types of chemical structures generated using combinatorial approaches. These structures are used as candidates in large-scale drug screens, but they may not be sufficiently drug-like (Feher and Schmidt 2003). As I have shown in chapter 4, the bioactivity of a small molecule (SM) is influenced by its structure, and *in silico* screening of chemical libraries could make drug screens more efficient.

start_nt	0.996	ks_s_mikatae	0.109	groe_incoming_tfs	0.031
gnumber	0.992	ge_hypoosmotic	0.108	charged_prop	0.030
cc_target_num	0.781	mrna_copy_num	0.106	transl_noise	0.030
metabolic_centrality	0.444	protein_per_s	0.104	affinity_centrality	0.030
paralogues	0.384	ge_raffinose	0.104	affinity_degree	0.028
grd_target_num	0.335	transc_half-life	0.103	aggreg_prop	0.027
stackingen_nt	0.324	ge_irradiation	0.101	epi_degree	0.026
gc_content	0.299	ge_25c_growth	0.099	disorder	0.026
grd_incoming_tfs	0.288	h3k79me3vsh3_5p	0.098	h3k4me2vsh3_5p	0.025
cc_incoming_tfs	0.287	ge_ddt	0.098	ortho_k_lactis	0.022
metabolic_incoming	0.264	ge_steady_sorbitol	0.096	minfree_50nt_5p	0.022
metabolic_connectivity	0.254	metabolic_cc	0.095	kaks_s_mikatae	0.019
ge_pheoh	0.236	ge_diauxic_10h	0.095	tm_helices	0.019
ge_trpoh	0.218	h3k14acvsh3_h2o2_5p	0.089	aromatic_prop	0.017
ge_n_starvation	0.197	expr_div	0.082	transl_rate	0.015
ge_menadione	0.186	2hybrid_cc	0.081	pi	0.015
ge_diamide	0.182	epi_closeness	0.079	h3k4me3vsh3_5p	0.013
het_av	0.181	h3_5p	0.078	h3k14acvsh3_5p	0.012
ge_h2o2	0.177	h3k9acvsh3_5p	0.077	epi_cc	0.009
metabolic_outgoing	0.176	h4_5p	0.074	gravy	0.008
ge_aa_starvation	0.175	ribosome_num	0.068	ribo_density	0.006
het_conditions	0.174	tbp_occupancy	0.065	h4acvsh3_5p	0.004
transc_degrad	0.164	identity_sc_pombe	0.056	protein_half-life	0.004
ge_heatshock	0.159	acidic_prop	0.051	weight	-0.002
ge_ethanol	0.149	tiny_prop	0.051	basic_prop	-0.002
hom_mm	0.142	protein_level	0.048	stackingen	-0.003
hom_av	0.139	minfree_4_53nt	0.047	2hybrid_centrality	-0.003
ribo_occupancy	0.139	h3k36me3vsh3_5p	0.046	aa_length	-0.003
ge_statphaseentry	0.137	minfree_100nt_5p	0.045	kaks_average	-0.006
ge_diauxic_19h	0.136	minfree_50nt_3p	0.041	aliphatic_prop	-0.006
ge_glucose	0.135	nonpolar_prop	0.040	oe_toxicity	-0.017
ge_coldshock	0.134	polar_prop	0.040	avgpcc	-0.017
h3_h2o2_5p	0.132	identity_s_mikatae	0.040	epi_centrality	-0.020
ge_sorbitol	0.130	affinity_cc	0.038	2hybrid_closeness	-0.025
het_mm	0.128	minfree_200nt_5p	0.037	h3k4me1vsh3_5p	-0.025
h4acvsh3_h2o2_5p	0.123	cai	0.037	2hybrid_degree	-0.029
hom_conditions	0.118	affinity_closeness	0.036	complexsize	-0.041
ge_diauxic_12h	0.118	pest_motif	0.036	incoming_kinases	-0.062
ge_galactose	0.115	charge	0.032		

Table 6-1. Pearson's correlation coefficient between neighbouring genes for the continuous descriptors in the descriptor database of chapter 5. The higher the correlation, the more the value of a descriptor for a certain gene is correlated with the value of its neighbouring genes. Correlation coefficients that are statistically significant at an uncorrected p -value of less than 0.01 are in **bold**. Some of the correlations are relatively high but not significant as not enough data points are available.

Another reason could be that the current approaches to drug discovery are overly simplistic (figure 6-2) (Hopkins and Groom 2002; Hopkins, Mason et al. 2006; Hopkins 2008).

Many SMs bind multiple ligands by forming colloid-like aggregates that lead to nonspecific inhibition of protein function (McGovern, Helfand et al. 2003; Coan and Shoichet 2008). But even without this effect, which only sets in beyond a specific concentration of SMs that in turn leads to aggregation (Coan and Shoichet 2008), many, if not most, drugs bind more than one protein. This includes some of the most effective drugs. To strive for the highest possible selectivity in drug design may

therefore be misguided (Hopkins 2008). Instead, network approaches to drug discovery may prove to be more successful (figure 6-2).

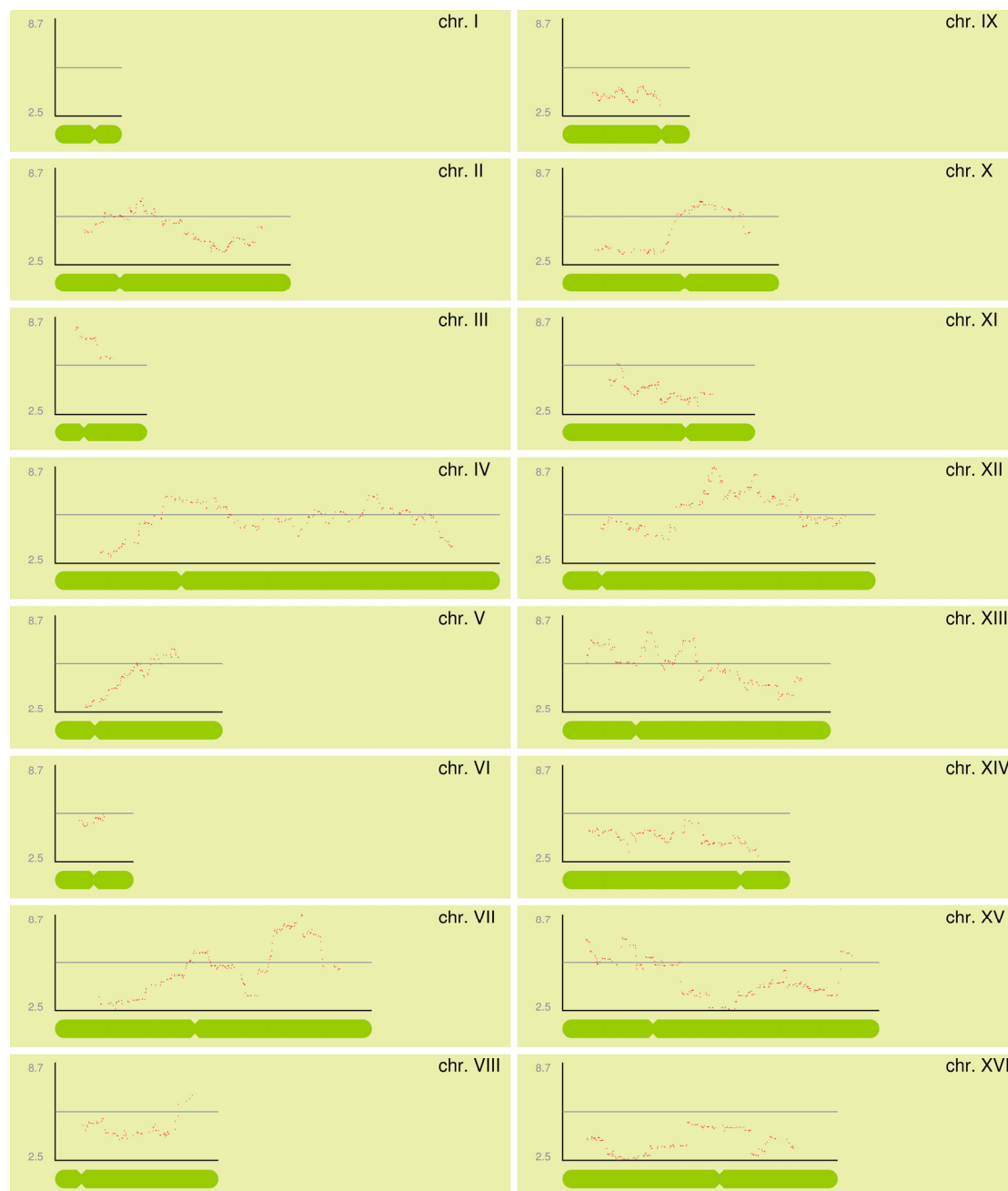


Figure 6-1. Genomic distribution of the values of the descriptor *grd_incoming_tfs*, averaged across a window of 45 genes. This descriptor informs on the number of transcription factors that regulate each gene according to transcription factor deletion experiments. It is easy to see that the values of the descriptors are not distributed across the genes of the genome randomly. For example, the genes on the left arm of chromosome X seem to be regulated by fewer transcription factors than the genes on the right arm of the same chromosome. Similar plots for other descriptors and window lengths can be viewed on <http://www.mrc-lmb.cam.ac.uk/genomes/awuster/chrllocation/>.

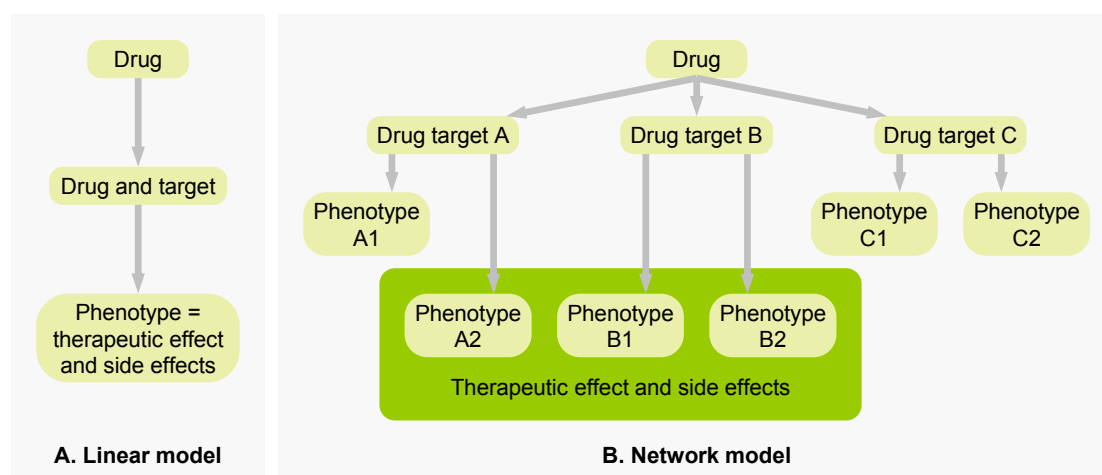


Figure 6-2. The effect of many drugs is due to their interaction with more than one target. However, a linear model of drug action (A) is the basis of most current drug discovery approaches. According to this model, a drug binds only a single target and thus has a single therapeutic effect. In the network model (B), a drug has several targets, each of which can produce more than one phenotype. Thus, the therapeutic effect and the side effects are the result of more than one molecular event. This network model is more likely to reflect reality. Figure adapted from figure 2 of (Lansbury 2004).

Drugs to disrupt quorum sensing

Multiple antibiotic resistance is increasingly becoming a problem (Dantas, Sommer et al. 2008; Sommer, Dantas et al. 2009), as can be seen by the example of Methicillin resistant *Staphylococcus aureus* (MRSA). By definition, antibiotics are of no use for killing resistant bacteria. An alternative to outright killing resistant bacteria may be to attenuate their virulence. It is known that disrupting the *agr* quorum sensing system lowers the ability of *S. aureus* to cause disease in the mouse model (Abdelnour, Arvidson et al. 1993). The *agr* system regulates genes involved in virulence not only in *S. aureus* but also in other organisms where the system is present. Therefore, the discovery of potential *agr*-like communication systems in organisms such as the pathogen *Clostridium perfringens* (see chapter 3) might have an impact on fighting disease by targeting virulence in these organisms. The same applies for the computational identification of cell-to-cell communication systems in other genomes, which I outlined above. Should this identification process be successful and lead to the discovery of other cell-to-cell communication systems in pathogenic bacteria, it might provide the basis for further research into signalling-related drug development.

A potential biotechnological use of quorum sensing

Apart from *C. perfringens*, other species where I could predict an *agr* quorum sensing system were *Moorella thermoacetica* ATCC 39070, *Desulfitobacterium hafniense* Y51, and *Thermoanaerobacter tengcongensis* (chapter 3). None of these species are pathogenic. Like all Clostridia, they are anaerobic, and *Moorella* and

Thermoanaerobacter are thermophilic (Xue, Xu et al. 2001; Drake and Daniel 2004). The genus *Desulfitobacterium* consists of bacteria which survive in a wide range of environments. Desulfitobacteria are metabolically diverse and use a variety of compounds as electron acceptors, including metals, nitrate, sulphite, and halogenated compounds. This makes the Desulfitobacteria a candidate for anaerobic bioremediation processes. In anaerobic fixed-film bioreactors *D. hafniense* has been shown to occur in a biofilm together with sulphate reducing bacteria (Villemur, Lanthier et al. 2006). If quorum sensing via the *agr* locus indeed occurs in *D. hafniense*, it can be assumed that it would modulate biofilm formation as it does in other species (Dunman, Murphy et al. 2001; Bourgogne, Hilsenbeck et al. 2006; Cassat, Dunman et al. 2006). Therefore, knowledge of the *agr* system could be used to modulate the biofilm forming behaviour of *D. hafniense* in bioreactors.

Chemogenomic approaches to drug discovery

James Black, winner of the 1988 Nobel Prize in Physiology and Medicine, stated that “The most fruitful basis for the discovery of a new drug is to start with an old drug” (Raju 2000). This way of thinking has led to a new approach in pharmaceutical research, where drug targets are no longer seen as individual proteins, but are classified into protein families. The advantage of this approach is that it allows to take full advantage of the “similar receptors bind similar ligands” paradigm (Klabunde 2007).

Chemogenomics is a relatively new field that promises to be of high relevance to the discovery of drug candidates. Although the amount of chemogenomic data continues to rise, the application of novel methods to specific biotechnological and biomedical applications (*translational research*) have not been fully exploited. In the future, new analysis tools will be required to deal with the large amount of data generated. The chemogenomic networks I introduced in chapter 4 may provide a conceptual framework for this task. Chemogenomic networks may not only be a novel way of looking at chemogenomic data, but they can also be used to gain new biological insights. Traditionally, interactome networks have considered cells under standard laboratory conditions. For the study of standard cellular processes, this has delivered valuable information. However, a frequently stated aim of obtaining ever more detailed and reliable *interactome* networks is to aid drug discovery and development. For this purpose a cell under standard growth conditions might not be an ideal model system. Rather, interactome networks under specific conditions, such as under the

influence of a drug, might deliver more relevant information. Chemogenomic networks might be one way in which this can be achieved.

A large number of chemogenomic studies to date have been carried out on the yeast *S. cerevisiae*. The primary reason for this is that genetic libraries of strains in which almost every yeast gene is deleted are available. Although yeast has been the focus of most studies, there is no reason to limit the chemogenomic approaches described here to this organism. For example, a knockout library of the bacterium *E. coli* has recently become available (Baba, Ara et al. 2006). Furthermore, a library of RNA interference (RNAi) knockouts is now available for a variety of cells (Paddison, Silva et al. 2004; Boutros, Bras et al. 2006; Horn, Arziman et al. 2007) and has been used for chemogenomic screens (Brummelkamp, Fabius et al. 2006). Recently, chemogenomic approaches have also been tried in multicellular organisms including vertebrates. For example, the fibroblast growth factor (FGF) pathway plays a role in development, proliferation, and cellular homeostasis. In order to find SMs that interfere with the activity of the FGF pathway, Molina *et al.* (2009) conducted a chemogenomic screen in *Danio rerio* (zebrafish). They tested a natural products SM library against mutants that had been engineered to glow when the FGF pathway was active. This way, they were able to identify the SM (*E*)-2-benzylidene-3-(cyclohexylamino)-2,3-dihydro-1*H*-inden-1-one (BCI) as a modulator of the FGF pathway (Molina, Vogt et al. 2009). Moreover, cell types do not always have to be as clearly defined as in mutant libraries, as for an example in a screen measuring the viability of 60 cancer cell lines in response to various drugs (Scherf, Ross et al. 2000). Taken together, this demonstrates that another major focus of chemogenomics will be to translate the methodologies and findings from yeast to other model systems.

The drug-COP score

As I mentioned above, one of the most pressing problems in chemogenomics is how to extract information from chemogenomic data. One possible method that I believe merits further refinement is the *drug-COP score*. This is a score that combines epistatic and chemogenomic data in order to predict possible SM and drug targets. The inspiration for the drug-COP score comes from the complex and pathway (COP) score. The COP score (Schuldiner, Collins et al. 2005; Collins, Schuldiner et al. 2006) refers to some type of likelihood that two gene products interact physically and is computed from an epistatic interaction dataset. For each pair of genes, the COP score integrates two types of information: The first one is the score of the epistatic interaction between the two genes, and the second one is the correlation coefficient

between the epistatic profiles of the two genes. The thinking behind this is that two genes are more likely to be members of the same protein complex if they show epistatic interaction, and if they have similar epistatic profiles (Parsons, Brost et al. 2004). Many of the gene pairs with the highest COP score were part of the same pathway or linear pathway (Collins, Schuldiner et al. 2006).

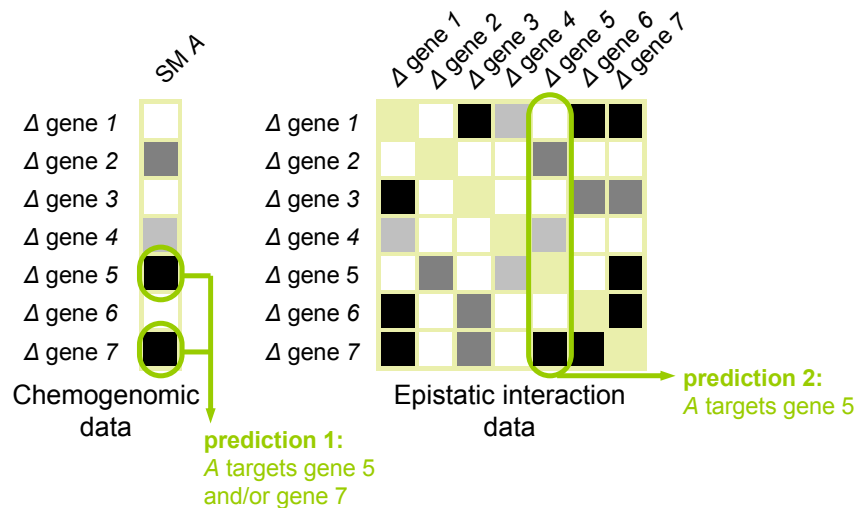


Figure 6-3. Information on possible drug targets can be extracted from chemogenomic data in two ways. First, extreme chemogenomic scores may indicate an interaction. In the case of SM A, gene 5 and 7 have extreme chemogenomic scores. Second, a large correlation between the chemogenomic profile of the SM and the epistatic interaction profile of a gene may indicate an interaction. In the case of SM A, gene 5 has a very similar epistatic interaction profile. Combining these two findings, it can be concluded that it is most likely that SM A interacts with gene 5.

Towards a scoring system

To be maximally useful, the drug-COP has to meet the following conditions: It has to provide a single score for every gene under investigation. This number refers to some type of likelihood that this gene is a target of the SM under investigation. The drug-COP has to integrate the correlation between the chemogenomic profile of the SM and the epistatic profile of the gene (r), and the chemogenomic score of the gene (s) (figure 6-3). The larger r , the larger the drug-COP. Because experience shows that both a particularly positive and a particularly negative chemogenomic score s can point to physical interactions, only intermediate values of s should be penalised. A drug-COP score computed by the following method meets all the above criteria:

$$\text{drug-COP} = r^2 / a^2 + (s + c)^2 / b^2,$$

where r is the correlation coefficient between the genetic and the chemogenomic profile if $r > 0$, s is the chemogenomic score, a the importance given to r , b the importance given to s , and c indicates how much more to count positive scores than negative scores. In the scoring model in figure 6-4, $a = 0.3$, $b = 2.5$, and $c = 5$. This

scoring method basically applies the equation for an ellipse whose centre is at $[0, -c]$ of length a and height b .

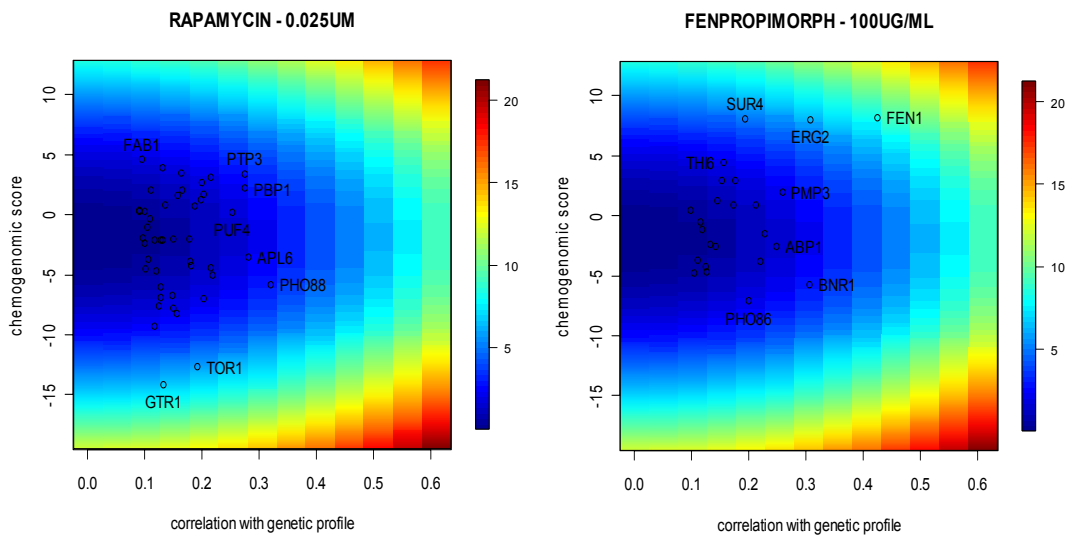


Figure 6-4. Two examples of a drug-COP scoring system applied to chemogenomic-genetic plots. Each data point corresponds to a gene. The horizontal axis shows how much the chemogenomic score profile of the SM is correlated with the epigenetic profile of the gene. The vertical axis shows how large the chemogenomic score of the gene is when treated with that particular SM. The colour in the background refers to the drug-COP score computed from the following formula: $\text{drug-COP} = r^2 / a^2 + (s + c)^2 / b^2$, where r is the correlation coefficient between the genetic and the chemogenomic profile if $r > 0$ and $p_r < 0.01$ (horizontal axis), s is the chemogenomic score (vertical axis), a is the importance given to r in the scoring, b is the importance given to s in the scoring, and c indicates how much more importance is given to positive chemogenomic scores than to negative scores. In this way, both large correlations and very large or very small chemogenomic scores are taken into account. In the example to the left, the target of rapamycin (TOR1) has a relatively high chemogenomic score. Similarly, in the example to the right, the target of fenpropimorph, an ergosterol biosynthesis gene (ERG2), also has a high chemogenomic score. FEN1 is also a known fenpropimorph resistance gene.

The future of the drug-COP

Currently, the drug-COP score is no more than a concept. Two conditions need to be met in order to ensure that it becomes a valuable tool.

Firstly, the scoring method needs to be perfected. For this, the method by which the drug-COP score is computed has to be agreed upon. But even once the method is set, it will be necessary to adjust the parameters so that the resulting scores are as informative as possible. For example, in the method introduced above, three parameters (a , b , and c) need to be set. The way in which these parameters are set strongly influences the resulting COP-score. The best way to set those parameters would be to use a gold-standard training set. The compilation of the gold standard set would be a prerequisite, as to my knowledge no such gold standard of known drug targets in *S. cerevisiae* exists at the time of writing.

Secondly, the drug-COP score has to be applied to a large chemogenomic dataset that is comparable to existing epistatic interaction datasets. Currently, such a dataset is created by Laura Schuresko-Kapitzky in the laboratory of Nevan Krogan at the University of California, San Francisco. I am currently collaborating with this laboratory for the purpose of analysing this data using various computational tools. These tools will include chemogenomic networks (chapter 4) to extract biological information from the data, and the drug-COP to identify potential drug targets.

Conclusion

For the last decade or so, there has been a demonstrable trend for more rational drug design (Yildirim, Goh et al. 2007). For example, AIDS drugs such as Agenerase and Viracept have been developed using the crystal structure of the HIV protease (Blundell, Sibanda et al. 2006). Nevertheless, the number of new drugs that are developed now is lower than it used to be 20 years ago (PWC 2007). I am convinced that the only approach that has a chance to reverse this trend is research with the aim of a better understanding of how drugs and the biological systems they modify work. It is my hope that the method and results I have presented in this thesis contribute towards this goal.

References

- Abdelnour, A., S. Arvidson, et al. (1993). "The accessory gene regulator (*agr*) controls *Staphylococcus aureus* virulence in a murine arthritis model." *Infect Immun* **61**(9): 3879-85.
- Baba, T., T. Ara, et al. (2006). "Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection." *Mol Syst Biol* **2**: 2006 0008.
- Blundell, T. L., B. L. Sibanda, et al. (2006). "Structural biology and bioinformatics in drug design: opportunities and challenges for target identification and lead discovery." *Philos Trans R Soc Lond B Biol Sci* **361**(1467): 413-23.
- Bourgogne, A., S. G. Hilsenbeck, et al. (2006). "Comparison of OG1RF and an isogenic *fsrB* deletion mutant by transcriptional analysis: the *Fsr* system of *Enterococcus faecalis* is more than the activator of gelatinase and serine protease." *J Bacteriol* **188**(8): 2875-84.
- Boutros, M., L. P. Bras, et al. (2006). "Analysis of cell-based RNAi screens." *Genome Biol* **7**(7): R66.
- Brummelkamp, T. R., A. W. Fabius, et al. (2006). "An shRNA barcode screen provides insight into cancer cell vulnerability to MDM2 inhibitors." *Nat Chem Biol* **2**(4): 202-6.
- Cassat, J., P. M. Dunman, et al. (2006). "Transcriptional profiling of a *Staphylococcus aureus* clinical isolate and its isogenic *agr* and *sarA* mutants reveals global differences in comparison to the laboratory strain RN6390." *Microbiology* **152**(Pt 10): 3075-90.
- Coan, K. E. and B. K. Shoichet (2008). "Stoichiometry and physical chemistry of promiscuous aggregate-based inhibitors." *J Am Chem Soc* **130**(29): 9606-12.
- Collins, S. R., M. Schuldiner, et al. (2006). "A strategy for extracting and analyzing large-scale quantitative epistatic interaction data." *Genome Biol* **7**(7): R63.
- Dantas, G., M. O. Sommer, et al. (2008). "Bacteria subsisting on antibiotics." *Science* **320**(5872): 100-3.
- Drake, H. L. and S. L. Daniel (2004). "Physiology of the thermophilic acetogen *Moorella thermoacetica*." *Res Microbiol* **155**(10): 869-83.
- Dunman, P. M., E. Murphy, et al. (2001). "Transcription profiling-based identification of *Staphylococcus aureus* genes regulated by the *agr* and/or *sarA* loci." *J Bacteriol* **183**(24): 7341-53.
- Feher, M. and J. M. Schmidt (2003). "Property distributions: differences between drugs, natural products, and molecules from combinatorial chemistry." *J Chem Inf Comput Sci* **43**(1): 218-27.
- Hopkins, A. L. (2008). "Network pharmacology: the next paradigm in drug discovery." *Nat Chem Biol* **4**(11): 682-90.
- Hopkins, A. L. and C. R. Groom (2002). "The druggable genome." *Nat Rev Drug Discov* **1**(9): 727-30.
- Hopkins, A. L., J. S. Mason, et al. (2006). "Can we rationally design promiscuous drugs?" *Curr Opin Struct Biol* **16**(1): 127-36.
- Horn, T., Z. Arziman, et al. (2007). "GenomeRNAi: a database for cell-based RNAi phenotypes." *Nucleic Acids Res* **35**(Database issue): D492-7.
- Janga, S. C., J. Collado-Vides, et al. (2008). "Transcriptional regulation constrains the organization of genes on eukaryotic chromosomes." *Proc Natl Acad Sci U S A* **105**(41): 15761-6.
- Klabunde, T. (2007). "Chemogenomic approaches to drug discovery: similar receptors bind similar ligands." *Br J Pharmacol* **152**(1): 5-7.
- Lansbury, P. T., Jr. (2004). "Back to the future: the 'old-fashioned' way to new medications for neurodegeneration." *Nat Med* **10** Suppl: S51-7.

- Madan Babu, M., S. C. Janga, et al. (2008). "Eukaryotic gene regulation in three dimensions and its impact on genome evolution." Curr Opin Genet Dev **18**(6): 571-82.
- McGovern, S. L., B. T. Helfand, et al. (2003). "A specific mechanism of nonspecific inhibition." J Med Chem **46**(20): 4265-72.
- Molina, G., A. Vogt, et al. (2009). "Zebrafish chemical screening reveals an inhibitor of Dusp6 that expands cardiac cell lineages." Nat Chem Biol.
- Paddison, P. J., J. M. Silva, et al. (2004). "A resource for large-scale RNA-interference-based screens in mammals." Nature **428**(6981): 427-31.
- Parsons, A. B., R. L. Brost, et al. (2004). "Integration of chemical-genetic and genetic interaction data links bioactive compounds to cellular target pathways." Nat Biotechnol **22**(1): 62-9.
- Pellegrini, M., E. M. Marcotte, et al. (1999). "Assigning protein functions by comparative genome analysis: protein phylogenetic profiles." Proc Natl Acad Sci U S A **96**(8): 4285-8.
- PWC (2007). The vision. Pharma 2020, PriceWaterhouseCoopers.
- Raju, T. N. (2000). "The Nobel chronicles. 1988: James Whyte Black, (b 1924), Gertrude Elion (1918-99), and George H Hitchings (1905-98)." Lancet **355**(9208): 1022.
- Scherf, U., D. T. Ross, et al. (2000). "A gene expression database for the molecular pharmacology of cancer." Nat Genet **24**(3): 236-44.
- Schuldiner, M., S. R. Collins, et al. (2005). "Exploration of the function and organization of the yeast early secretory pathway through an epistatic miniarray profile." Cell **123**(3): 507-19.
- Sommer, M. O., G. Dantas, et al. (2009). "Functional characterization of the antibiotic resistance reservoir in the human microflora." Science **325**(5944): 1128-31.
- Villemur, R., M. Lanthier, et al. (2006). "The Desulfitobacterium genus." FEMS Microbiol Rev **30**(5): 706-33.
- Xue, Y., Y. Xu, et al. (2001). "Thermoanaerobacter tengcongensis sp. nov., a novel anaerobic, saccharolytic, thermophilic bacterium isolated from a hot spring in Tengcong, China." Int J Syst Evol Microbiol **51**(Pt 4): 1335-41.
- Yildirim, M. A., K. I. Goh, et al. (2007). "Drug-target network." Nat Biotechnol **25**(10): 1119-1126.