BioSystems: PDF for review

Journal	BioSystems
Article ID	BIO_306
Title	Boolean Networks with biologically relevant rules show ordered behavior
Version	3
Article type	Full-length article
Submitted	18 Nov 05

Files submitted

Name	Fig No	Format	Use	Description
biosystems2.tex		Manuscript (Tex/LateX document)		
fig1.eps	1 a)	Figures (Postscript)	Yes	
fig2.eps	1 b)	Figures (Postscript)	Yes	

Boolean Networks with biologically relevant rules show ordered behavior ⋆

S. Nikolajewa [a], M. Friedel [a] and T. Wilhelm [a,*]

[a] *Theoretical Systems Biology, Leibniz Institute for Age Research - Fritz Lipmann Institute (former Institute of Molecular Biotechnology), Beutenbergstr. 11, Jena, D-07745, Germany*

Abstract

It was found recently that natural gene regulatory systems are governed by hierarchically canalyzing functions (HCFs), a special subclass of Boolean functions. Here we study the HCF class in detail. We present a new minimal logical expression for all HCFs. Based on this formula, we calculate the cardinality of the HCF class. Moreover, we define HCF subclasses and calculate their cardinality as well. Using the well-known critical connectivity condition $2K_c p(1-p) = 1$, we discuss order-chaos transitions of Boolean networks (BNs) regulated by functions of given HCF subclasses. Finally, analysing real gene regulatory rules we show that nearly all of the biologically relevant functions belong to the simplest HCF subclasses. This restriction is important for reverse engineering of transcription regulatory networks and for ensemble approach studies in systems biology. It is shown that Boolean networks with functions belonging to the biologically realized HCF subclasses show ordered behavior.

Key words: Boolean Network, canalyzing function, hierarchically canalyzing function, nested canalyzing function

1 Introduction

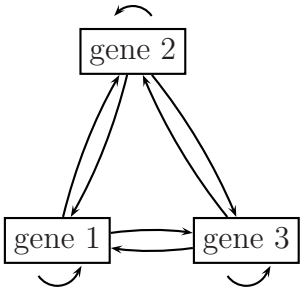
One of the outstanding problems in contemporary systems biology is the understanding of the multifariously interwoven networks underlying cellular reg-

⋆ This work has been supported by the Bundesministerium für Bildung und Forschung Grant 0312704E.

* Corresponding author. Adress: FLI, Beutenbergstr. 11, Jena, D-07745, Germany; tel:+49-3641 65 6208, fax:+49-3641 65 6191

Email address: wilhelm@fli-leibniz.de (T. Wilhelm).

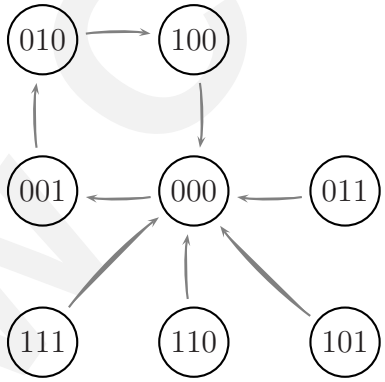
ulation. Boolean networks (BNs) (Kauffman, 1969, 1993) play a prominent role to elucidate and simulate cellular regulatory systems. In these simple models the nodes (for instance genes) are either on or off. Consider, for example, a BN representing a gene regulatory system comprising three genes, with the wiring diagram and the corresponding Boolean updating rules:



$$x_1^{t+1} = \bar{x}_1^t \wedge x_2^t \wedge \bar{x}_3^t$$
$$x_2^{t+1} = \bar{x}_1^t \wedge \bar{x}_2^t \wedge x_3^t$$
$$x_3^{t+1} = \bar{x}_1^t \wedge \bar{x}_2^t \wedge \bar{x}_3^t$$

The BN can also be defined by a truth table or a state space flow diagram, mapping each possible input state to the corresponding output state:

x_1^t	x_2^t	x_3^t	x_1^{t+1}	x_2^{t+1}	x_3^{t+1}
0	0	0	0	0	1
0	0	1	0	1	0
0	1	0	1	0	0
0	1	1	0	0	0
1	0	0	0	0	0
1	0	1	0	0	0
1	1	0	0	0	0
1	1	1	0	0	0



This system has one attractor of length 4.

It was stated that up to now BNs represent the only class of dynamical models which led to nontrivial results about cellular organization on a large scale (Szallasi and Liang, 1998). Meanwhile different interesting large-scale results were published using other formalisms, such as differential equations (Heinrich et al., 2002). Nevertheless, many authors study BNs for a better understanding of cellular systems (Albert and Barabasi, 2000; Bornholdt and Sneppen, 2000; Glass and Hill, 1998; Huang, 2001; Stauffer, 1987; Lähdesmäki et al., 2003; Huang et al., 2005; Shmulevich et al., 2005), others try to evaluate in detail the inexactness due to such an abstract approach (Buchler et al., 2003; Setty et al., 2003).

Real genetic networks do not use all Boolean rules with the same probability (Kauffman, 1993; Shmulevich et al., 2003; Gat-Viks and Shamir, 2003).

Harris et al. (2002) collected the updating rules of 139 different real genes. A corresponding analysis confirmed earlier speculations: it was shown that nearly all of these rules belong to the class of canalyzing functions (CFs) (Kauffman, 1993), also denoted as canalizing functions (Shmulevich et al., 2004) or forcing functions (Stauffer, 1987). A Boolean function is canalyzing if already one input alone can determine the output. The other inputs play a role only if this canalyzing input takes its non-canalyzing value (Kauffman, 1993). Moreover, the cardinality of the CF class was calculated (Just et al., 2004). A later analysis of Harris' data revealed that 133 of the 139 rules belong to a special subclass of CFs: to hierarchically canalyzing functions (HCFs), also known as nested canalyzing functions (Kauffman et al., 2003), a class first introduced some years ago by Szallasi and Liang (1998).

Section 2 deals with HCFs in general. We present the minimal logical expression for HCFs. Based on this result we calculate the number of HCFs. Moreover, we define subclasses of HCFs and calculate their cardinality as well. Based on the well-known critical connectivity formula $2K_c p(1-p) = 1$ (Derrida and Pomeau, 1986), the order-chaos transitions are discussed for BNs regulated by functions of a given HCF subclass. Section 3 discusses biological applications. Analyzing Harris' data (Harris et al., 2002) we show that 128 of the 133 hierarchically canalyzing gene regulatory rules belong to the two simplest HCF subclasses. It is shown that BNs with functions of these biologically relevant subclasses show ordered behavior.

2 Hierarchically Canalyzing Functions (HCF)

Some years ago the idea of CFs was extended to hierarchically canalyzing functions (Szallasi and Liang, 1998). In HCFs all inputs are canalyzing in a hierarchical manner: if the first input takes on its non-canalyzing value, a second input is canalyzing for the remaining states. If the second input takes on the non-canalyzing value, a third input is canalyzing, etc. HCFs represent an important subclass of CFs. It was shown that HCFs enhance order even more than simple CFs (Szallasi and Liang, 1998; Kauffman et al., 2004). Studying the transcriptional regulation of real genes, it was found that just 6 of 139 are not HCFs (Kauffman et al., 2003). Thus, in genetic regulatory networks there seems to be a very strong tendency towards HCFs. In the following we give a formal definition of HCFs and present the corresponding minimal logical formula representation. Moreover, we calculate the exact number of hierarchically canalyzing functions (depending on the number of inputs), define HCF subclasses and calculate their cardinality as well.
Let k denote the number of inputs of a Boolean function $f(x) = f(x_1, \ldots, x_k)$ and $\mathcal{B}^k$ the set of all Boolean functions on k variables. The symbol σ stands for a possible negation of Boolean variables, so x^σ can be x or $\bar{x}$. The input x_i is

called essential if $f(x_1,\ldots,x_{i-1},0,x_{i+1},\ldots,x_k) \neq f(x_1,\ldots,x_{i-1},1,x_{i+1},\ldots,x_k)$.

Definition 1 (Hierarchically Canalyzing Function) *Let f be a canalyzing function with canalyzing input x_i and canalyzing input value a_i. If the function $g(x_1,\ldots,x_{i-1},x_{i+1},\ldots,x_k)$=$f(x_1,\ldots,x_{i-1},\bar{a}_i,x_{i+1},\ldots,x_k)$ on $k-1$ inputs is canalyzing, then the function $f(x)$ is a 2-times canalyzing function. A Boolean function f on k essential inputs is called hierarchically canalyzing if $f(x)$ is k-times canalyzing with the canalyzing inputs $x_1,\ldots,x_k$ and the canalyzing input values $a_1,\ldots,a_k$.*

2.1 Minimal formula representation of a Hierarchically Canalyzing Function

A Boolean function $f(x)$ on k essential inputs is a hierarchically canalyzing function if and only if $f(x)$ can be written with k different inputs and $k-1$ binary operations $\wedge$ or $\vee$, where the operation priority is ordered from left to right ($\odot \in \{\wedge,\vee\}$):

$$f_i(x_1,x_2,\ldots,x_k) = {x_{i_1}}^{\sigma_{i_1}} \odot ({x_{i_2}}^{\sigma_{i_2}} \odot (\ldots \odot ({x_{i_{k-1}}}^{\sigma_{i_{k-1}}} \odot {x_{i_k}}^{\sigma_{i_k}}))...). \quad (1)$$

This can constructively be proven as follows: let $f(x_1,x_2,\ldots,x_k)$ be a hierarchically canalyzing function with the first canalyzing input x_1 (index i is omitted w.l.o.g.), then there exists a function $f_2 = g(x_2,\ldots,x_k)$ such that f can be written as $f = {x_1}^{\sigma_1} \odot f_2(x_2,x_3,\ldots,x_k)$, $\odot \in \{\wedge,\vee\}$. From the HCF definition it follows that function $f_2 \in \mathcal{B}^{k-1}$ is canalyzing again. Repeating this procedure k times leads to the minimal logical formula (1).

Example 1 *TGF-β is a gene regulated by a HCF. The transforming growth factor-β controls growth differentiation and apoptosis of cells. It is regulated by five transcription factors (Harris et al., 2002): the negative regulator SnoN, the receptor regulated proteins Smad2/4 and Smad3/4, histone deacetylases HDAC, and the nuclear transcriptional corepressor N-CoR. The gene updating rule can be defined by a binary string (11111111000000001111111111101111), where the order of digits corresponds to the order of transcription factors indicated above, or the minimal logical formula $TGF^{(t+1)} = Smad2/4^{(t)} \wedge (not\ SnoN^{(t)} \vee Smad3/4^{(t)} \vee not\ N-CoR^{(t)} \vee not\ HDAC/Sin3^{(t)})$.*

2.2 Number of Hierarchically Canalyzing Functions

Szallasi and Liang (1998) numerically calculated the number of HCFs up to $k=5$ inputs. Here we present the exact formula for arbitrary k. The number

of hierarchically canalyzing functions is

$$N = 2 + 2k + \sum_{i=2}^{k} \binom{k}{i} a_i 2^{i+1}, \tag{2}$$

with $a_2 = 1$
$a_3 = 1 + \binom{3}{2} a_2$
$a_4 = 1 + \binom{4}{2} a_2 + \binom{4}{3} a_3$
...
$a_k = 1 + \sum_{j=2}^{k-1} \binom{k}{j} a_j$.

This number is deduced by counting all different HCF representations (1) for a given number of inputs k. For each k two constant functions (0,0,...,0 and 1,1,...,1) and $2k$ functions that depend on just one input exist (first two summands in (2)). Each summand in (2) counts the number of functions with i essential inputs, a_i denotes the number of different parenthesis patterns (Table 1), separating $\wedge$ and $\vee-$ operators. For each parenthesis pattern, there are two possibilities of operator assignments: starting with $\wedge$ or with $\vee$.

i	parenthesis pattern	operator assignments		a_i
2	$(x_1 \odot x_2)$	$x_1 \vee x_2$; $x_1 \wedge x_2$	1	$a_2 = 1$
3	$x_1 \odot x_2 \odot x_3$	$x_1 \vee x_2 \vee x_3$; $x_1 \wedge x_2 \wedge x_3$	1	$a_3 = 1 + \binom{3}{2} a_2$
	$(x_1 \odot x_2) \odot x_3$	$(x_1 \vee x_2) \wedge x_3$; $(x_1 \wedge x_2) \vee x_3$		
	$x_1 \odot (x_2 \odot x_3)$	$x_1 \vee (x_2 \wedge x_3)$; $x_1 \wedge (x_2 \vee x_3)$	$\binom{3}{2}$	
	$x_2 \odot (x_1 \odot x_3)$	$x_2 \vee (x_1 \wedge x_3)$; $x_2 \wedge (x_1 \vee x_3)$		
4	$x_1 \odot x_2 \odot x_3 \odot x_4$	...	1	
	$(x_1 \odot x_2) \odot x_3 \odot x_4$	...	$\binom{4}{2}$	$a_4 = 1 + \binom{4}{2} a_2 + \binom{4}{3} a_3$
	...	...		
	$x_1 \odot (x_2 \odot (x_3 \odot x_4))$	...	$\binom{4}{3}$	
i	...	...	...	$a_i = 1 + \sum_{j=2}^{i-1} \binom{i}{j} a_j$

Table 1
Number (a_i) of different possibilities to set parentheses in minimal logical expressions of HCFs with i essential inputs.

The number of different σ patterns (negations of inputs) is 2^i. Therefore, the number of different HCFs on i essential inputs is $a_i * 2 * 2^i$.

2.3 Subclasses of Hierarchically Canalyzing Functions

The definition of HCF subclasses S_l^k is based on the minimal formula representation of a HCF (1), containing k essential inputs and $k-1$ logical operations in a fixed order. These operations can be encoded by a binary number of length $k-1$, where 1 and 0 correspond to OR and AND, respectively. For example,

the operations of function $\bar{x}_1 \wedge (x_2 \vee (\bar{x}_3 \wedge x_4))$ are encoded by 010 (or decimal $l = 3$), thus this function belongs to the class S_3^4.

Definition 2 (S_l^k: subclasses of HCFs) *Let l_b be the binary representation of the decimal number l: $l_b = l_0 l_1 \dots l_{k-2}$ codes for the operations order in the minimal logical HCF formula (1). All hierarchically canalyzing functions $f_i(x_1, \dots, x_k) = {x_{i_1}}^{\sigma_{i_1}} \odot_1 ({x_{i_2}}^{\sigma_{i_2}} \odot_2 (\dots ({x_{i_{k-1}}}^{\sigma_{i_{k-1}}} \odot_{k-1} {x_{i_k}}^{\sigma_{i_k}}))...)$, where $\odot_j$ is $\wedge$, if $l_{j-1} = 0$, or $\odot_j$ is $\vee$, if $l_{j-1} = 1$ with the same operation order $l_b = l_0 l_1 \dots l_{k-2}$ belong to the S_l^k class.*

Based on the different operator patterns the class of HCFs on k inputs can be divided into 2^{k-1} subclasses: $S_0^k,\ S_1^k,\ \dots,\ S_{2^{k-1}-1}^k$. The class S_0^k contains all rules $x_1^{\sigma_1} \wedge x_2^{\sigma_2} \wedge \dots \wedge x_k^{\sigma_k}$. In class S_1^k the last operation is $\vee$ and all other operations are $\wedge$, for instance $x_k^{\sigma_k} \wedge x_{k-1}^{\sigma_{k-1}} \wedge \dots \wedge (x_2^{\sigma_2} \vee x_1^{\sigma_1})$. Functions with only $\vee$ belong to the class $S_{2^{k-1}-1}^k$.

Appendix A shows an example of all subclasses of HCFs on $k = 4$ inputs. In appendix B a procedure to calculate the cardinality of the HCF subclasses S_l^k is given.

2.4 Order and chaos of Boolean Networks with S_l^k functions (S_l^k networks)

The phase transition between the ordered and chaotic regimes can be defined with help of the average sensitivity

$$\bar{S} = 2Kp(1-p), \tag{3}$$

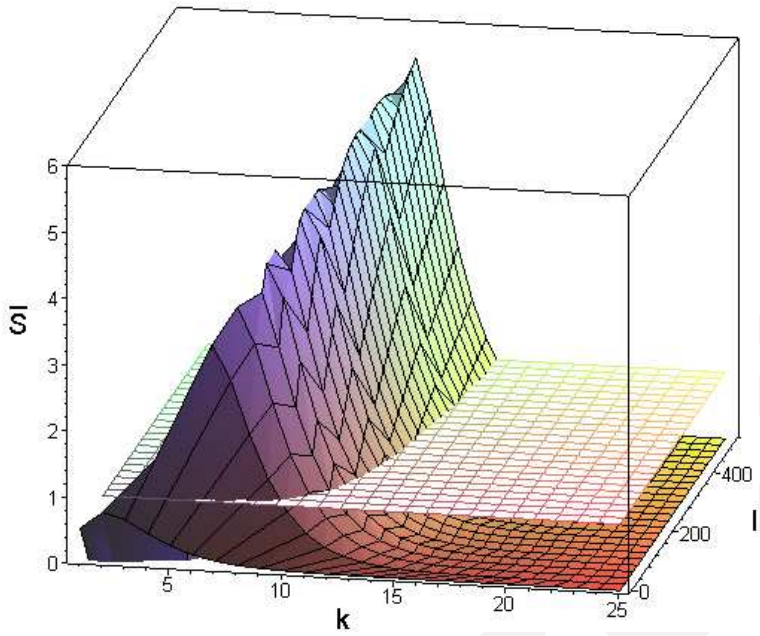
where K is the average connectivity of the BN and p is the probability of choosing 1 rather than 0 for the transition function output values (Derrida and Pomeau, 1986; Shmulevich et al., 2005). For $\bar{S} < 1$ the network is in the ordered regime, for $\bar{S} > 1$ the network shows chaotic behavior. The number of ones of a given S_l^k function is $2l+1$ (proof is given in appendix C). Therefore,

$$p = \frac{2l+1}{2^k}. \tag{4}$$

A BN where each node is regulated by a S_l^k function (S_l^k network) has the average sensitivity: $\bar{S}(k,l) = 2k\frac{2l+1}{2^k}(1 - \frac{2l+1}{2^k})$.

Figure 1a shows $\bar{S}(k,l)$ and figure 1b $\bar{S}(k)$ for some given l.

a)



b)

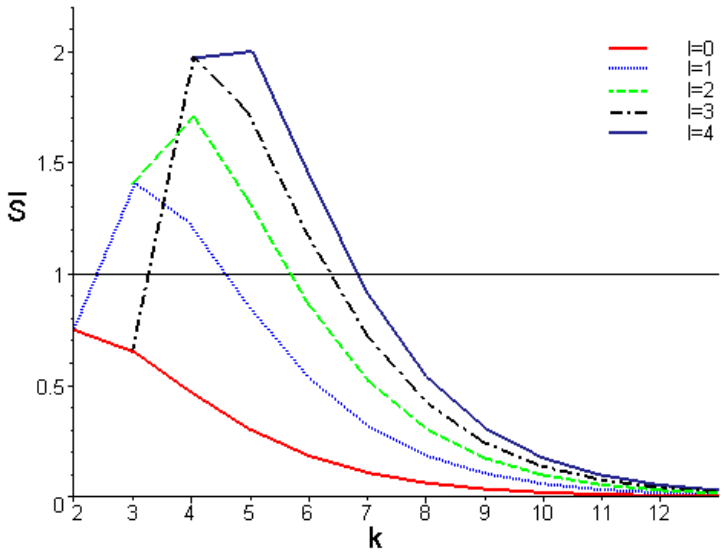


Fig. 1. Sensitivity of S_l^k networks. a) $\bar{S}(k,l)$. The cut between $\bar{S}(k,l)$ and $\bar{S}=1$ defines the order - chaos transition. b) $\bar{S}(k)$ for $l=0,1,2,3,4$

S_0^k networks are always stable, because $\bar{S}<1$ for arbitrary k. Generally, for S_l^k networks intervals of chaotic behavior exist (Table 2). Interestingly, for many l (for instance $l=1,3$) ordered behavior is shown for small and large k and chaotic behavior for medium k.

l	chaos	order
0 and $2^{k-1}-1$	–	$k \geq 2$
1 and $2^{k-1}-2$	$3 \leq k \leq 4$	$k = 2,\ k \geq 5$
2 and $2^{k-1}-3$	$3 \leq k \leq 5$	$k \geq 6$
3 and $2^{k-1}-4$	$4 \leq k \leq 6$	$k = 3,\ k \geq 7$
...		
$2^{k-2}-2$ and $2^{k-2}+1$	$k \geq 4$	$k = 3$
$2^{k-2}-1$ and 2^{k-2}	$k \geq 3$	–

Table 2
Ordered and chaotic regimes for S_l^k networks.

3 Biological Importance of Hierarchically Canalyzing Functions

Analyzing natural gene regulatory rules it was first found that all of them belong to the class of canalyzing functions (Harris et al., 2002). Later it was shown that 133 of the 139 analyzed functions are also hierarchically canalyzing (Kauffman et al., 2003). We translated all these 133 HCFs into the corresponding minimal logical expressions and found that nearly all belong to two special subclasses: S_0^k $66,39\%$ and S_1^k $29,41\%$.
Based on the average sensitivity (3) and the probability (4) it can be shown that BNs, which are made up of $2/3$ S_0^k and $1/3$ S_1^k rules, show always ordered behavior. Such a network has a mean probability $p = \frac{2}{3} * \frac{1}{2^k} + \frac{1}{3} * \frac{3}{2^k} = \frac{5}{3} * \frac{1}{2^k}$, and therefore a mean sensitivity $\bar{S}(k) = \frac{10}{3} * \frac{k}{2^k} * (1 - \frac{5}{3*2^k})$ with $\bar{S}(k) < 1\ \forall k$ (maximum $\bar{S}_{max} = \bar{S}(3) = \frac{95}{96}$).

HCFs can most easily be realized by gene regulatory systems. Here only one single operator ($\wedge$ or $\vee$) per transcription factor exists, no additional intermediate TFs are needed. In contrast, non-canalyzing functions seem to be unfavorable for the design of gene regulatory systems. Consider, for instance, the simplest non-canalyzing function x_1 XOR x_2 (Table 3): the result is false, if both inputs have the same value. The implementation of XOR needs intermediate transcription factors, which implies synchronization problems (Buchler et al., 2003). The XOR-gate cannot be implemented by the minimal formula (1).

The specific $\wedge$, $\vee$ pattern in the minimal expression of a given HCF provides interesting information about the output control. In the simplest cases all operations are the same. If the minimal logical expression contains only AND (S_0^k), then all (transcription) factors have the same importance and cannot be replaced by other factors. In the 'only OR case' ($S^k_{2^{k-1}-1}$) the factors are independent from each other, each factor alone can determine the output. Generally, the number of ANDs and ORs and their position in the minimal

inputs		1st step		2nd step
x_1	x_2	$x_1 \wedge \bar{x}_2$	$\bar{x}_1 \wedge x_2$	$(x_1 \wedge \bar{x}_2) \vee (\bar{x}_1 \wedge x_2) = x_1\ XOR\ x_2$
0	0	0	0	0
0	1	0	1	1
1	0	1	0	1
1	1	0	0	0

Table 3
Implementation of XOR based on AND and OR gates.

expression (1) provide information about (gene) regulation: the number of $\vee$ and $\wedge$ corresponds to the number and size of factor groups, respectively, that determine the output independently of each other (four such groups in example 2.1).

4 Discussion

We presented a minimal logical expression for Boolean functions of the biologically important HCF class. Based on this result the exact number of HCFs was calculated. Moreover we defined meaningful subclasses of HCFs, calculated their cardinality as well, and discussed the stability of S_l^k networks. Analyzing biological data on gene regulation we found that nearly all genes are regulated by functions of the two simplest subclasses of HCFs: S_0^k and S_1^k. This is important for corresponding ensemble approach studies (Kauffman, 2004a,b; Kauffman et al., 2004) and for reverse engineering of gene regulatory networks (Liang et al., 1998; Akutsu et al., 1999, 2000; D'haeseleer et al., 2000; Yeung et al., 2003). The smaller the number of possible functions, the fewer data is needed for reverse engineering. Table 4 demonstrates that S_0^k and S_1^k comprise just a small subset of all HCFs.

k	S_0^k	S_1^k	HCF	CF	$Total$
1	0	0	4	4	4
2	4	4	14	14	16
3	8	24	96	120	256
4	16	96	1050	3514	65536
5	32	320	15036	1292276	4294967296

Table 4
The number of Boolean functions on k inputs in different biologically meaningful subclasses in comparison to the total number of Boolean functions (2^{2^k}).

We have shown that Boolean networks with functions belonging to the biologically observed subclasses (Harris et al., 2002) are always stable. This corresponds to recently found numerical results (Kauffman et al., 2003; Rämö et al. , 2005). Recently, a special stability measure for HCF networks was proposed, which is based on particular canalyzing rule distributions (Kauffman et al., 2004). In contrast, we used the well-known Derrida-formula (3) and our calculation of true points for the different S_l^k classes 4. S_0^k-networks are always stable, but already S_1^k-networks can be chaotic (if $k = 3$ and/or 4). We have shown that networks with the biologically realistic distribution of 2/3 S_0^k and 1/3 S_1^k functions are always stable. Interestingly, 72% of all observed HCFs (Harris et al., 2002) have 2 or 3 inputs. For these k the average sensitivity $\bar{S}$ is close to 1, which implies that real gene regulatory networks may operate close to the order/chaos transition. This corresponds to the often discussed 'edge-of-chaos' theory stating that evolution drives natural systems towards the critical regime (Kauffman, 1993). Analysing time series of mRNA expression it was recently found that eukaryotic cells are in the ordered or critical regime (Shmulevich et al., 2005).

One may question the biological relevance of our analysis of Harris' data (Harris et al., 2002) with respect to (i) the quality of the data and (ii) the validity of the Boolean approach itself. We have argued that simple biological reasons are responsible for the overwhelming dominance of simple HCF functions because these functions are easily realisable at the level of cis-regulatory transcription control. However, one cannot exclude that the analysed data are biased in the sense that biologists simply do not consider all complicated possibilities of logical regulation but tend to define simple activation or inhibition rules. Concerning the Boolean approach it was often argued that although being a crude simplification it nevertheless captures the essential aspects of the gene-regulation dynamics (Shmulevich et al., 2005). Corresponding detailed analyses show that transcription factor concentrations are important and Boolean descriptions are simplifications only valid for special values of the kinetic parameters (Setty et al., 2003; Buchler et al., 2003; Mayo et al., 2006). However, it was shown that important aspects of the dynamical behavior of networks, such as multistationarity and oscillations, are mainly influenced by the network structure (topology), which holds for both, logical and differential formalisms (Thieffry and Thomas, 1998).

The minimal logical expression of HCFs has different additional advantages for the analysis of gene regulation and beyond. Membership in the HCF class can be proven in polynomial time for any Boolean function, thus for these functions the famous minimization problem (important for chip design, for instance) is solved. For genes regulated by HCFs the minimal expression helps to quantify the importance of the different transcription factors. Moreover, groups of TFs that operate together can easily be identified. These combined operations could be proven by clustering of corresponding mRNA expression

data.

Here we exclusively analysed HCFs and HCF networks because nearly all known gene regulatory functions belong to this class (Kauffman et al., 2003, 2004). However, it may be possible that in future also other functions will be discovered where, for instance, only a subset of transcription factors act in a canalyzing way. It would be interesting to analyse the biological roles of such different classes of transcription factors.

Acknowledgment. We thank S. Harris for providing us the data on gene regulation.

APPENDIX A. The 8 subclasses of HCFs with $k = 4$.

For the sake of simplicity possible negations are neglected. Negations only shift the true points and do therefore not alter the number of the true/false points. The second line of the table shows the binary representation l_b of the corresponding operator pattern, given in the third line.

	Class				S_0^k	S_1^k	S_2^k	S_3^k	S_4^k	S_5^k	S_6^k	S_7^k
					000	001	010	011	100	101	110	111
n	x_1	x_2	x_3	x_4	$x_1(x_2(x_3x_4))$	$x_1(x_2(x_3 \vee x_4))$	$x_1(x_2 \vee (x_3x_4))$	$x_1(x_2 \vee (x_3 \vee x_4))$	$x_1 \vee (x_2(x_3x_4))$	$x_1 \vee (x_2(x_3 \vee x_4))$	$x_1 \vee (x_2 \vee (x_3x_4))$	$x_1 \vee (x_2 \vee (x_3 \vee x_4))$
0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	1	0	0	0	0	0	0	0	1
2	0	0	1	0	0	0	0	0	0	0	0	1
3	0	0	1	1	0	0	0	0	0	0	1	1
4	0	1	0	0	0	0	0	0	0	0	1	1
5	0	1	0	1	0	0	0	0	0	1	1	1
6	0	1	1	0	0	0	0	0	0	1	1	1
7	0	1	1	1	0	0	0	0	1	1	1	1
8	1	0	0	0	0	0	0	0	1	1	1	1
9	1	0	0	1	0	0	0	1	1	1	1	1
10	1	0	1	0	0	0	0	1	1	1	1	1
11	1	0	1	1	0	0	1	1	1	1	1	1
12	1	1	0	0	0	0	1	1	1	1	1	1
13	1	1	0	1	0	1	1	1	1	1	1	1
14	1	1	1	0	0	1	1	1	1	1	1	1
15	1	1	1	1	1	1	1	1	1	1	1	1
	Number of ones				1	3	5	7	9	11	13	15

APPENDIX B. The number of functions in HCF subclasses S_l^k.

The number of different functions in the S_0^k (and $S_{2^{k-1}-1}^k$) class (k essential inputs) is easily counted: each negation just shifts the only 1 (or 0). Because there are 2^k negations (each input can be negated or not), there are 2^k different S_0^k (and $S_{2^{k-1}-1}^k$) functions. There are $\binom{k}{2}$ possibilities to choose the two inputs which are combined by $\vee$. Therefore, the cardinality of the S_1^k class is $\binom{k}{2}2^k$. The determination of the number of exchange possibilities for an arbitrary HCF subclass is based on the corresponding operator pattern given by the minimal logical expression. The pattern is divided into blocks of equal operations (the order of operations is fixed). Then the number of exchanges leading to different functions is calculated for each block. Finally, these numbers are multiplied. The procedure to calculate the number of different functions of a given S_l^k class is:

Input: k is the number of essential inputs; $l_b = l_0 l_1 \dots l_{k-2}$, ($l_i \in \{0,1\}$) denotes the subclass.
Output: N is the number of elements in the S_l^k class.

PROCEDURE N(k,l_b);
  $N := 2^k$;
  $\widehat{k} := k$;
  $n_{op} := 2$;
  $i := k - 3$;
  repeat
    if $l_{i+1} = l_i$ **then** $n_{op} := n_{op} + 1$
    else
      $N := N * \binom{\widehat{k}}{n_{op}}$;
      $\widehat{k} := \widehat{k} - n_{op}$;
      $n_{op} := 1$;
    $i := i - 1$;
  until ($i < 0$)
  return N;
END;

APPENDIX C. The number of true points of any function of the S_l^k class is $2l + 1$.

The binary representation $l_b = l_0 l_1 \dots l_{k-2}$ of the decimal number $l = l_0 2^{k-2} + l_1 2^{k-3} + \dots + l_{k-2} 2^0$ codes for the order of the $k-1$ operations. If the first bit $l_0 = 1$ (first operator is $\vee$), the first input is canalyzing to 1, which implies $\frac{2^k}{2}$ ones in the truth table. Generally, each $l_i = 1$ yields $\frac{2^k}{2^{i+1}}$ ones in the truth table. Because the last input is always canalyzing to 1 we have one additional 1 in the truth table output. Summing the total number of ones leads to $l_0 \frac{2^k}{2} + l_1 \frac{2^k}{2^2} + \dots + l_{k-2} \frac{2^k}{2^{k-1}} + 1 = 2l + 1$.

References

Albert,R., Barabasi,A.-L., 2000. Dynamics of complex systems: Scaling laws for the period of Boolean networks. Phys. Rev. Lett. 84, pp. 5660-5663.

Akutsu,T., Miyano,S., Kuhara,S., 2000. Algorithms for inferring qualitative models of biological networks Pacific Symposium on Biocomputing 5, 293-304.

Akutsu,T., Miyano,S., Kuhara,S., 1999. Identification of genetic networks from a small number of gene expression patterns under the Boolean network model. Pacific Symposium on Biocomputing 4, pp. 17-28.

Bornholdt,S., Sneppen,K., 2000. Robustness as an Evolutionary Principle. Proc. R. Soc. London B. 267, pp. 2281-2286.

Buchler,N.E., Gerland,U., Hwa,T., 2003. On schemes of combinatorial transcription logic. Proc. Natl. Acad. Sci. 100, pp. 5136-5141.

Derrida,B., Pomeau,Y., 1986. Random networks of automata - a simple annealed approximation. Europhysics letters 1(2), pp. 45-49.

D'haeseleer P., Liang S., Somogyi R., 2000. Genetic network inference: from co-expression clustering to reverse engineering. Bioinformatics. 16(8),pp. 707-726.

Gat-Viks,I., Shamir,R., 2003. Chain functions and scoring functions in genetic networks. Bioinformatics 19(1), pp. 108-117.

Glass,L., Hill,C., 1998. Ordered and Disordered Dynamics in Random Networks. Europhys. Letts. 41, pp. 599-604.

Harris,S., Sawhill,B., Wuensche,A., Kauffman,S., 2002. A Model of Transcriptional Regulatory Networks Based on Biases in the Observed Regulation Rules. Complexity 7(4), pp. 23-40.

Heinrich,R., Neel,B.G., Rapoport,T.A., 2002. Mathematical models of protein kinase signal transduction. Mol Cell. 9(5), pp. 957-70.

Huang,S., 2001. Genomics, complexity and drug discovery: insights from Boolean network models of cellular regulation. Pharmacogenomics 2(2), pp. 203-221.

Huang,S., Eichler,G., Bar-Yam,Y., Ingber,D.E., 2005. Cell fates as high-dimensional attractor states of a complex gene regulatory network. Phys. Rev. Lett. 94, pp. 128701.

Just,W., Shmulevich,I., Konvalina,J., 2004. The number and probability of canalizing functions. Physica D. 197, pp. 211-221.

Kauffman,S., 1969. Metabolic Stability and Epigenesis in Randomly Constructed Genetic Nets, J. Theor. Biol. 22, pp. 437-467.

Kauffman,S., 1993. The Origins of Order: Self-Organization and Selection in Evolution. Oxford University Press, New York.

Kauffman,S., 2004. A proposal for using the ensemble approach to understand genetic regulatory networks. J. Theor. Biol. 230(4), pp. 581-590.

Kauffman,S., 2004. The ensemble approach to understand genetic regulatory networks. Physica A, 340, pp. 733-740.

Kauffman,S., Peterson,C., Samuelsson,B., and Troein,C., 2004. Genetic net-

works with canalyzing Boolean rules are always stable. Proc. Natl. Acad. Sci. 101, pp. 17102-17107

Kauffman,S., Peterson,C., Samuelsson,B., Troein,C., 2003. Random Boolean Network Models and the Yeast Transcriptional Network. Proc. Natl. Acad. Sci. 100, pp. 14796-14799.

Lähdesmäki,H., Shmulevich,I., Yli-Harja,O., 2003. On learning gene regulatory networks under the Boolean network model. Machine Learning 52, pp. 147-167.

Liang,S., Fuhrman,S., and Somogyi,R., 1998. Reveal, a general reverse engineering algorithm for inference of genetic network architectures. in: Pacific Symp. on Biocomputing 3, pp. 18-29.

Mayo,A.E., Setty,Y., Shavit,S., Zaslaver,A., Alon,U., 2006. Plasticity of the cis-regulatory input function of a gene. PLoS Biol. 4(4), e45.

Rämö, P., Kesseli,J., and Yli-Harja,O. 2005. Stability of functions in Boolean models of gene regulatory networks. Chaos 15, 034101.

Setty,Y., Mayo,A.E., Surette,M.G., Alon,U., 2003. Detailed map of a cis-regulatory input function. Proc. Natl. Acad. Sci. 100, pp. 7702-7707.

Shmulevich,I., Lähdesmäki,H., Dougherty,E.R., Astola,J., Zhang,W., 2003. The role of certain Post classes in Boolean network models of genetic networks. Proc. Natl. Acad. Sci. USA 100(19), pp. 10734-10739.

Shmulevich,I., Lähdesmäki,H., and Egiazarian,K., 2004. Spectral Methods for Testing Membership in Certain Post Classes and the Class of Forcing Functions. IEEE Signal Processing Let. 11(2), pp. 289-292.

Shmulevich, I., Kauffman,S.A., Aldana,M., 2005. Eukaryotic cells are dynamically ordered or critical but not chaotic. Proc. Natl. Acad. Sci. 102, pp. 13439-13444.

Stauffer,D., 1987. On forcing functions in Kauffman random boolean networks. J. Stat. Phys. 46(3-4), pp. 789-794.

Szallasi,Z., Liang,S., 1998. Modeling the Normal and Neoplastic Cell Cycle With Realistic Boolean Genetic Networks: Their Application for Understanding Carcinogenesis and Assessing Therapeutic Strategies. in: Pacific Symp. on Biocomputing 3, pp. 66-76.

Thieffry,D., Thomas,R., 1998. Qualitative analysis of gene networks. in: Pacific Symp. on Biocomputing 3, pp. 77-88.

Yeung,M.K.S., Tegner,J., Collins,J.J., 2003. Reverse engineering gene networks using singular value decomposition and robust regression. PNAS 99, pp. 6163-6168.