Symulacja w Badaniach i Rozwoju
Vol. 1, No. 2/2010

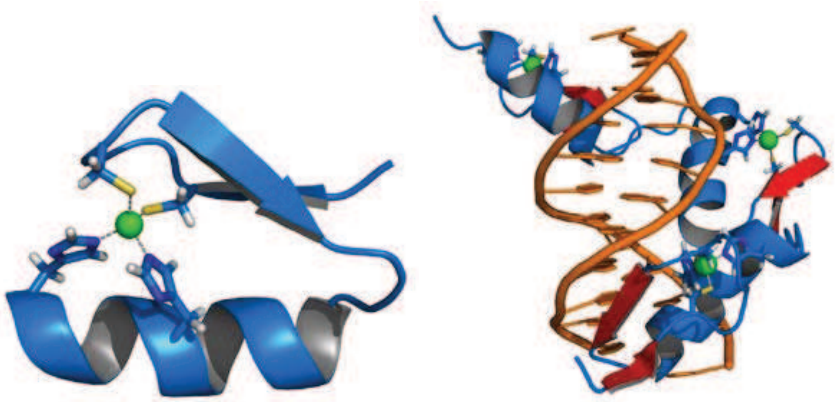
Kurt FRISCHMUTH
University of Rostock, Institute for Mathematics
E-mail: kurt.frischmuth@uni-rostock.de

Hans-Jürgen THIESEN
University of Rostock, Institute for Immunology
E-mail: hans-juergen.thiesen@med.uni-rostock.de

Modeling and Simulation in KRAB Zinc-Finger Research

1 Introduction

For a formal model of the behaviour of zinc fingers, let us assume that we are given two sets S and T. The members s_1, s_2, ... of the first set S control somehow the appearance of the elements t_1, t_2, ... in the second set T. Practically, a given controller s has the capability to suppress a given subset of targets in T. We denote this set of likely targets of s by T_s. Biologically, the elements of both sets – which may overlap - are proteins, see [3-5], Fig. 1. The aim of this article is to study the dynamics of suppression and release thereof in the presence of multiple suppressors and targets. In particular, we model limited capacities of controllers and their competition. The formulated postulates allow the implementation of a simple simulation of the temporal evolution of a hypothetical system composed of controllers and targets with random features.



Rys. 1. Palce cynkowe – schemat
Fig. 1. Zinc fingers – schematic representation

2 Model assumptions

2.1 Targeting

A first essential step towards a mathematical model capable of describing the interaction between KRAB zinc fingers and other genes is to identify which zinc finger, i.e. controller, affects which gene. Hence, we concentrate on an isolated situation – where there is only one controller and one target gene, no other players are present, [7].

Our first postulate is:

For each controller s, there is a specific set T_s of targets, whose presence is suppressed by s.

Of course, in the above statement we still need to specify what we mean by presence and how we quantify the effect on a target. It has to be mentioned that given the large number of zinc fingers and genes, experimental identification of the system components is extremely difficult and costly, [7].

In reality, we have a certain physical *carrier* of the controllers with a given limited *capacity*. This carrier can *board* only a certain quantity of copies (molecules) of each of the studied controllers. In the case of multiple controllers interacting, it is supposed that the physical arrangement of those copies is irrelevant, only the number of copies of a certain type – i.e. its share of the carrying capacity – counts.

When it comes to measuring the effect on a specific single target, we have several options. In a brutal setting, we might consider total elimination of a target *t*, cf. [6]. In a more subtle approach, we might consider a substantial reduction in its indicators, e.g. a drop of its concentration by a defined *threshold*, e.g. 60%, or similar forms of *decimation*.

Moreover, we need also to quantify the effect of a controller s on its whole target set T_s, if we assume the presence of multiple targets at the same time. By full effect we could understand again the total elimination of all of the elements in the set. A more moderate effect could be the elimination of a certain subset of the full target set T_s.

We infer that we have to identify, in a first step, a relation that expresses the relevance of the presence of a controller for that of a target in isolation. Then, in a second step, situations where controllers compete for access to the targets and / or the targets for the controllers have to be studied. The next subsection is devoted to this problem.

2.2 Competition for Limited Resources

In the case of the presence of several targets of a single controller s, it seems reasonable to imagine that there is a certain hierarchy, defining what targets will be affected first, and which targets will have smaller priority. Again, a more continuous variant based on a smooth transition from full presence to full liquidation could quantify the reduction of each of the targets. Here we would postulate different sensitivities to the presence of the controller in order to obtain a hierarchy.

If there are several controllers present at the same time, they have to compete for carrier capacity, which leads in the effect to a reduction of each single controller's priorities on the target list.

We come now to the second fundamental statement, which refines the first one:

A higher share of a given controller in the carrier capacity results in a higher effect.

As extreme example, assume that the whole carrier capacity is allocated to a single pure controller s. In that case all likely targets t from T_s are maximally reduced. By this we can mean that they are not present at all, or that their concentration is below the error margin, or that it dropped say to 5%.

If, however, s is only assigned 1/4 of the capacity, the effect might be e.g. that only the top 25% of the hierarchy of targets is affected, or that all targets are reduced by 25%.

The effect maybe over-linear or under-linear. By this we distinguish, whether a share of *p%* in the community of controllers can reduce the *controlling strength* of *s* to more or to less than *p%* of the limit value, obtained in a single controller case.

Also a combination of effects is possible: some targets are still suppressed as before, due to their high priority (or affinity) on the target list, others are no longer affected for lacking capacity.

2.3 Multiple Hits

Finally, we need to think about a third statement concerning multiple targeting. Whenever the capacity of the carrier is split between controllers, the possibility arises that several controllers are related to the same target. This might result in certain cross-effects. The obvious assumption for the roughest model would sound that *suppressed stays suppressed*. However, two competing controllers also might block each other. We adopt for this paper the straightforward assumption that the survival rates multiply.

As an example, let the carrier contain two controllers s_1 and s_2 each at 50%. The first one s_1 alone at 50% with neutral partners (controllers that do not address the same target) would reduce target t by 80%, controller s_2 would yield 90%. Together they reach 98%, because s_2 knocks out now 90% of the survivors of s_1.

Whatever rule we create, we should remember that it has to be symmetric, since we assume that the order of the controllers is not substantial to the effect. Hence we formulate our third postulate as:

Targets that are multiply addressed show stronger effects.

2.4 Feedback

So far, all model assumptions are static in the sense that certain relations between controllers and controlled targets are postulated. Once the control is on, it stays on forever. If there is partial control, i.e. suppression of a fraction of some gene, this fraction will never change.

In reality, all mentioned quantities like presence (concentration, share) of substances are functions of space and time. All substances undergo production, transportation in space, and destruction, [1,2]. It takes time to establish control, and of course, a controller itself might be controlled by some other mechanism, i.e., a suppression active at a given time may be released again later on.

In particular, we may obtain some interesting non-trivial dynamics in our system if we allow for feedback between targets and suppressors.

The suppression of certain targets implies the disappearance of selected suppressors.

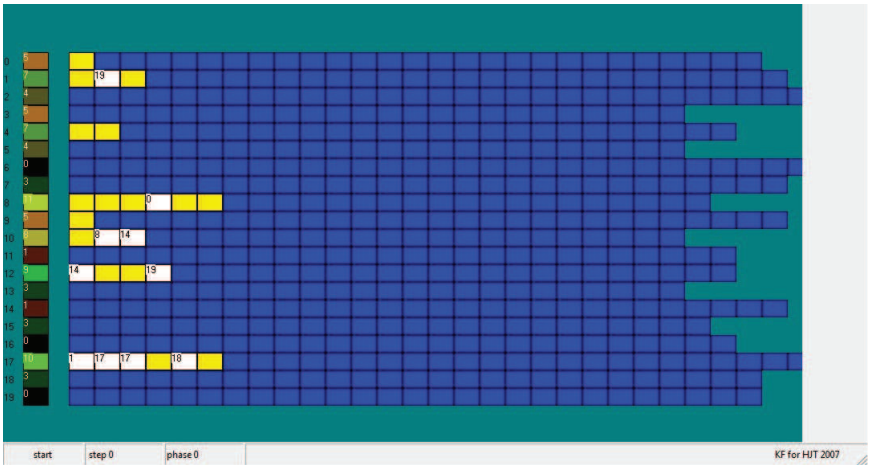
Once we introduce time into our considerations, we need to specify the time-scales at which suppressions are activated, i.e., at which the affected targets disappear. First of all, in general, this may be different for different pairs of controller and target, and it may depend on the competition (presence of other substances). For a simple model, let us assume a fairly uniform reaction time. We choose this time to define a sampling rate, at which we look at changes in *expression levels* of the genes we study.

3 Simulation

The postulates of the previous section define a rough framework for a simulation, however, they are still very general principles, which demand merely monotonicity of certain phenomenological laws.

In order to implement a nontrivial simulation, we need to choose a list of genes, their relations, how to quantify their presences and sensitivities.

We try to demonstrate the chain of events under simplifying assumptions. Assume there are n different controllers (*znfs*), each of them has a list of maximum length k_{max} target genes. The actual lengths may vary from zinc finger to zinc finger. So, we have at most $n \cdot k_{max}$ different gene types in out model's pool. This number may be smaller, since most likely not all lists exhaust their maximum length. Further, in general the lists may overlap.



Rys. 2. Narzędzie symulacyjne – zrzut ekranu
Fig. 2. Simulation tool – screenshot

Fig. 2 shows a screenshot of our simulation tool. On the left, a single colon of colored rectangles represents controllers and their levels of presence. The first postulate requires the existence of a list of dependencies – saying which genes are affected if a certain *zink finger* is present. To the right of each controller field, the corresponding targets are arranged according to the priority list.

Now, in accordance with the second postulate, with a higher expression level of a zinc finger, a longer part of the list behind the control is suppressed. We indicate this, again, by a change of color, in Fig. 2 from blue to yellow or white.

For now, targets have only binary states, suppressed or not, present or not, only the *znfs* are assigned more levels of expression.

Furthermore, we excluded overlappings, which is clear from the design of the simulation tool: the rows of squares are non-intersecting. The third postulate is of no consequence in this special case.

Finally, basing on the last postulate, we define a set of dependencies of *znfs* on target genes. In the program, this is visualized by the white fields. In those target cells the index of the dependent zinc finger is shown. Suppressing a gene means in consequence the disappearance of the corresponding zinc finger.

For the simulation, we have to decide what happens if we suppress a gene, which in turn leads to knocking out a certain zinc finger, which most likely was suppressing other genes. We lift suppressions imposed by a zinc finger in the time step following its disappearance. This way we can handle e.g. shortcuts – the case of zinc fingers which depend on a gene on their own lists of targets. In the same way, dependent zinc fingers recover – level by level – with the reappearance of the genes they depend on. This way, a cycle of life and death is established.

We performed simulations with random initial levels of zinc fingers and randomly chosen dependencies on genes. Several laws for the distribution of control power in between different zinc fingers were implemented and tested as well.

A common scenario which regularly appeared in all results was a cyclic behavior.

The simplest cycle is the mentioned *shortcut*. It is sort of a suicidal attack: zinc finger s_1 kills gene t_1 which immediately leads to the sudden death of s_1.

Obviously, such suicidal loops may have more intermediate states and substances involved, i.e., they also have a longer time period.

Indirect duels are possible as well: s_1 kills, via suppressing target gene t_1, its fellow zinc finger s_2, while at the same time s_2 kills s_1 via suppressing t_2. The question '*who is faster at the draw*?' does not arise since we imposed a uniform time step and a short delay before effects take place.

In the present implementation, the common time step assumption and the very rough state space of the genes leads to rather short cycles. We expect that longer periods will be typical if reaction times and recovery times are diversified.

4 Conclusion

In the next stage of this research, we will try to support the above postulates by data obtained from recent measurements. In particular, we will have to explain the nature of controllers, targets and carriers. We should also determine the carrier capacity (and will probably see that it is not constant in general). We assume that part of the carrying capacity can be lost to competitors, which are not from our set S and do not affect directly any targets but reduce the availability of capacity to the real controllers, [7].

The further development of the simulation tool will require the implementation of overlapping target lists, a refinement of expression levels of genes and the introduction of individual reaction and recovery times. At present, we have excluded positive effects in the simulation model – our controllers just repress, they do not enhance. This restriction also can be easily removed.

In the long run, also a spatial distribution, e.g. by introducing different cells and exchange rates between them, needs to be considered.

References

1. Murray J.D.: *Mathematical Biology I.* An Introduction. Springer Series: Interdisciplinary Applied Mathematics, Vol. 17, 2004
2. Murray J.D.: *Mathematical Biology II*. Spatial Models and Biomedical Applications. Springer Series: Interdisciplinary Applied Mathematics, Vol. 18, 2004
3. Krishna S.E.; Majumdar I.; Grishin N.V.: Survey and summary: Structural classification of zinc fingers. *Nucleic Acids Res.* **31**: 532-550, 2003
4. Jamieson A.C., Miller J.C., Pabo C.O.: Drug discovery with engineered zinc-finger proteins. *Nat. Rev. Drug Discov.* **2**: 361-8, 2003
5. Carroll D.: Progress and prospects: Zinc-finger nucleases as gene therapy agents. *Gene Therapy*, **15**: 1463-1468, 2008.
6. Ekker S.C.: Zinc finger-based knockout punches for zebrafish genes. *Zebrafish* **5**: 1121-3, 2008

7. Lorenz P., Kreutzer M., Kiseleva L., Qian Z., Frischmuth K., Li. Y., Thiesen H.-J.: *Systems biology platform established for studying KRABZNF cell biology in human tissues as well as human neoplasias*. In: M. Kellis, A. Califano, G. Stolovitzky (Eds.), Regulatory Genomics, Systems Biology DREAM3, MIT / Broad Institute / CSA, Oct 29-Nov 2, 2008

Summary

The human genome encodes about 350 KRAB zinc finger genes. The KRAB domain represents one of the strongest repression domains found in mammalian organisms. Antisera against numerous KRAB-ZNF proteins have been generated to dissect their biological functions in respect to expression patterns in normal and cancer tissues, which finally allows systems biology oriented modeling approaches. By numerical simulation we study the temporal evolution of a set of genes, assuming a certain model for interactions between their expression levels. Classification methods like SVM play an important role in the identification of the model.

Modelowanie i symulacja w badaniach KRAB palców cynkowych

Streszczenie

Geny człowieka kodują około 350 KRAB palców cynkowych. Dziedzina KRAB reprezentuje jedną z najmocniejszych dziedzin supresji pośród wszystkich znalezionych w organizmach ssaków. Wygenerowano antysera do licznych białek KRAB palców cynkowych, żeby rozróżnić ich funkcje biologiczne na podstawie poziomu ekspresji w tkankach zdrowych i rakowych. W ten sposób stworzono bazę do modelowania w ramach biologii systemowej. Za pomocą symulacji numerycznej badamy czasową ewolucję zbioru genów na podstawie pewnych założeń dotyczących interakcji między ich poziomami ekspresji. Przy identyfikacji modelu ważną rolę odgrywają metody klasyfikacji typu Maszyn Wektorów Podpierających.