When life is like a slice of almond bread

by Yves Konigshofer

The outcomes of some experiments are extremely dependent on experimental conditions that may not be thought of as important for, or related to, the processes being studied. In addition, the outcomes of some experiments do not always agree with the predicted outcomes. In order to determine whether or not chaotic events could be part of some cellular processes, the well-known fractal known as the Mandelbrot set was analyzed. Interested? Read on…

Part 1: The Mandelbrot Set

The equation for the Mandelbrot set is simply

$$z_{n+1} = z_n^{\,2} + m$$

where z_0 is equal to zero and m is a constant complex number. A given value of m is part of the Mandelbrot set if the value of z does not start increasing towards infinity.

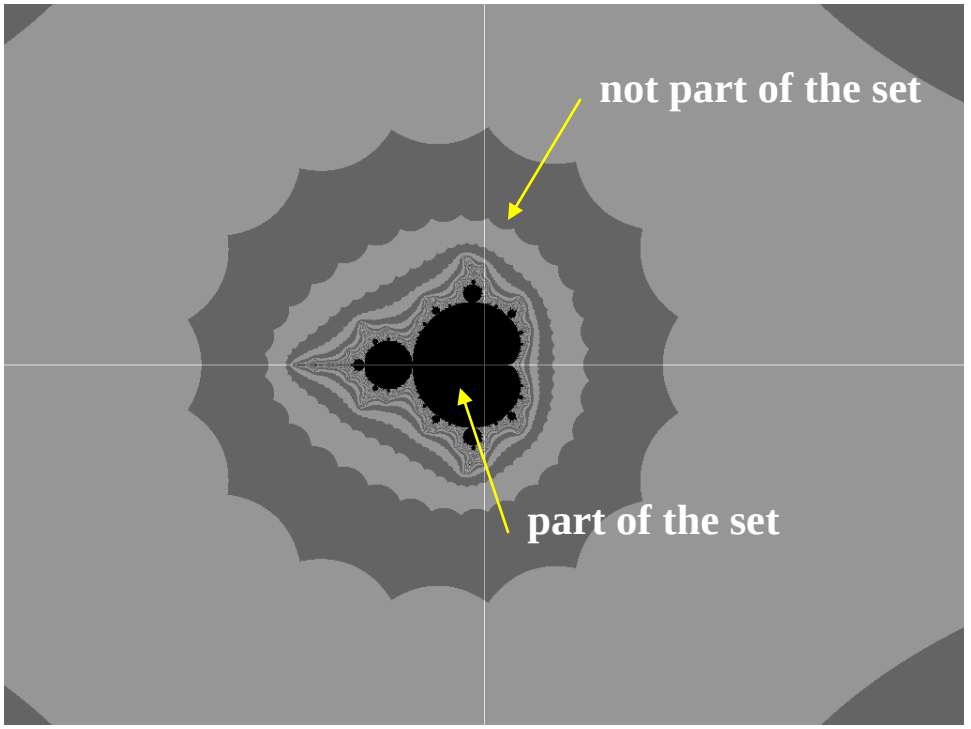
Thus, if m has a value of 0, it is quickly apparent that 0 is part of the set. 1, on the other hand, is not part of the set ($z_0 = 0$, $z_1 = 1$, $z_2 = 2$, $z_3 = 5$, $z_4 = 26$ …).

In order to draw the Mandelbrot set, the real and imaginary components of m represent the coordinates of a point on a graph. If the value of z increases above 30, then one can conclude that the point is not part of the set. If, after much iteration, it does not, then one can conclude that the point is probably part of the set.

If x and y are the real and imaginary numbers of z, and g and h are the real and imaginary numbers of m, then the above equation can be rewritten as:

$$x_{n+1} = x_n^{\,2} - y_n^{\,2} + g$$
$$y_{n+1} = 2x_n y_n + h$$

This leads to the following plot where points that are part of the set are colored black and points that are not part of the set are colored either light or dark gray depending on whether an odd or even number of iterations was needed before the absolute value of either x or y became greater than 30.



If g and h represent the constant rates of synthesis of two proteins, x and y, the Mandelbrot set illustrates the following concepts:

1) The complex appearance of this set suggests that for some constant rates of synthesis, *g* and *h*, it is nearly impossible to predict whether they will lead to large amounts of *x* and *y*.

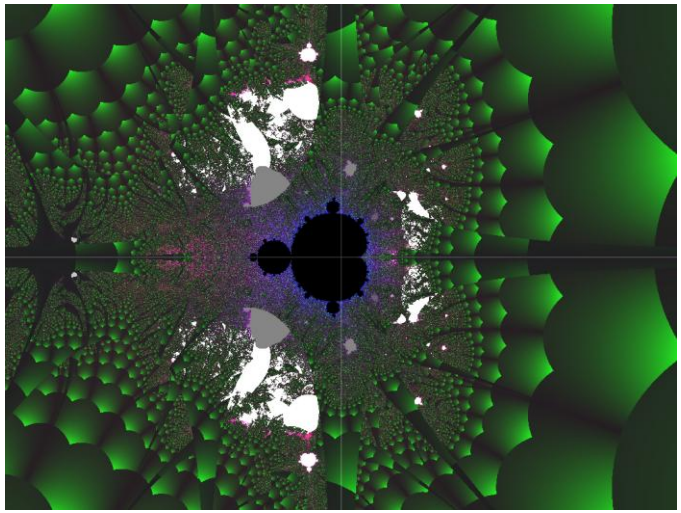
2) The knowledge of the results of all constant rates of synthesis of *x* and *y* provides little help in determining the underlying set of equations.

Part 2: Adding Cell Division

Some rates of synthesis lead to a surge in concentrations, where values of x and y increase towards infinity, while others do not.

We now introduce cell division into this model to combat the problem of ballooning concentrations by stating that if the absolute value of the product of x and y is larger than a fixed amount then both x and y are divided by two.

In the following figure, if the absolute value of the product of x_n and y_n is larger than 0.75, then both x_n and y_n are divided by two. Points that are part of the set and never lead to division are colored black. Points that are part of the set but had their x_n and y_n values divided by two one or more times are colored from black to gray to white, where the intensity reflects the amount of times x_n and y_n were divided by two. Points that are not part of the set are colored so that red levels reflect the amount of times their x_n and y_n values were divided by two, their green levels reflect the current absolute value of y and their blue values reflect the amount of iterations that took place before either x_{n+1} or y_{n+1} became greater than 30.



The points that are not part of the set can be seen as representing rates of synthesis that, in the long run, are not compatible with life. We can now introduce two more concepts:

3) The appearance of gray and white patches indicates that additional rates of synthesis are compatible with life as long as cell division takes place.

4) The difference in green intensities shows both that y can have many different values at the time when it is determined that a point is not part of the set and that it is essentially impossible to infer what this value will be when knowing just the constant rate of synthesis.

The two equations, while perfectly good for plotting the Mandelbrot Set, have a minor problem in that values of x and y can end up becoming negative. This is a problem because we are going to assume that x and y represent the concentrations of two proteins.

If nature abhors a vacuum, then nature surely abhors negative proteins…

Part 3: Dealing With Negative Protein Concentrations

By subtracting of one protein concentration from another, one can obtain a negative result this is still be perfectly compatible with life.

So, we can replace *x* with the difference in concentrations between proteins A and B and *y* with the difference in concentrations between proteins C and D. This yields the following set of equations where [A], [B], [C] and [D] represent the concentrations of A, B, C and D, respectively:

$$[A]_{n+1} - [B]_{n+1} = ([A]_n - [B]_n)^2 - ([C]_n - [D]_n)^2 + g$$
$$[C]_{n+1} - [D]_{n+1} = 2([A]_n - [B]_n)([C]_n - [D]_n) + h$$

This can be expanded to:

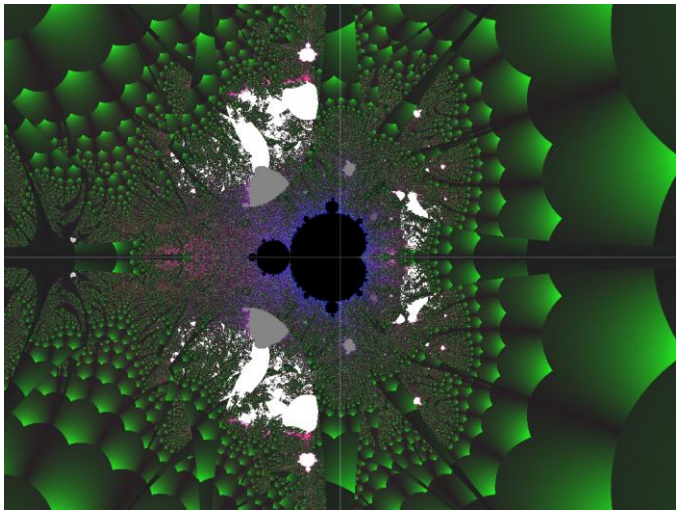
$$[A]_{n+1} - [B]_{n+1} = [A]_n^{\,2} - 2[A]_n[B]_n + [B]_n^{\,2} - [C]_n^{\,2} + 2[C]_n[D]_n - [D]_n^{\,2} + g$$
$$[C]_{n+1} - [D]_{n+1} = 2[A]_n[C]_n - 2[A]_n[D]_n - 2[B]_n[C]_n + 2[B]_n[D]_n + h$$

Next, we assign all of the positive terms to A and C and all of the negative terms to B and D and split *g* and *h* into g_a, g_b, h_c and h_d where $g = g_a - g_b$ and $h = h_c - h_d$. This allows us to obtain the following four equations that can be used to determine the values of $[A]_{n+1}$, $[B]_{n+1}$, $[C]_{n+1}$ and $[D]_{n+1}$.

$$[A]_{n+1} = [A]_n^{\,2} + [B]_n^{\,2} + 2[C]_n[D]_n + g_a$$
$$[B]_{n+1} = 2[A]_n[B]_n + [C]_n^{\,2} + [D]_n^{\,2} + g_b$$
$$[C]_{n+1} = 2[A]_n[C]_n + 2[B]_n[D]_n + h_a$$
$$[D]_{n+1} = 2[A]_n[D]_n + 2[B]_n[C]_n + h_b$$

We now have four separate equations that describe how the concentrations of four proteins change that are based on the single equation that gives rise to the Mandelbrot Set. In order to obtain the same image as before, we divide all concentrations by two if the absolute value of the product of [A] minus [B] and [C] minus [D] is larger than 0.75. In addition, we set g_a to g/2, g_b to –g/2, h_a to h/2 and h_b to –h/2.

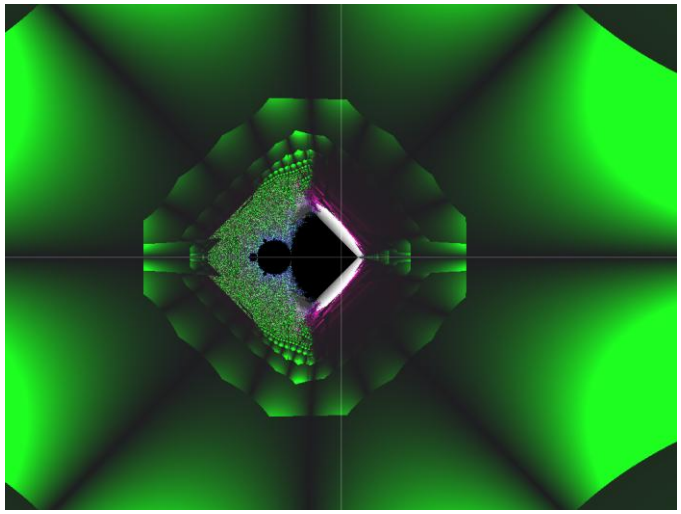
In the following figure, cell division takes place if the product of [A] minus [B] and [C] minus [D] is greater than 0.75. Points that are part of the set but had their values divided by two one or more times are colored from black to gray to white where the intensity reflects the amount of times *x* and *y* were divided by two. Points that are not part of the set are colored so that red levels reflect the amount of times their values were divided by 2, green levels reflect the current absolute value of [C] minus [D] and blue values reflect the amount of iterations that took place before either [A] minus [B] or [C] minus [D] became greater than 30.



Part 4: The Removal Of Proteins

In the case of –g/2 and –h/2 we are still subtracting and, in fact, values of [A], [B], [C] and [D] may become negative because of that. For this reason, we will now take a look at what happens if we do not subtract. In this case, we set g_a to g and h_a to h and keep g_b and h_b at zero.

The four-equation Mandelbrot set without subtraction. Cell division takes place if the product of [B] and [D] is greater than 0.1. Points are considered to be part of the set if the absolute values of [A], [B], [C] and [D] never become greater than 30.



When g_a and h_a are greater than zero, which is in the lower right quadrant, it appears as if there is a triangular region where cells rest that is surrounded by a region where cells divide. If even larger values for the constant rates of synthesis are used, then the concentrations of one or more species rise to the point that the cells die.

This is not as interesting as before…

Extending these equations to biologically-relevant processes still has the problem that if [A], [B], [C] and [D] represent the concentrations of proteins, then it is difficult to imagine how their concentrations will change from one iteration to the next to values that are partially the products of existing concentrations. Thus, we modify the equations to add to the existing concentrations instead of replacing them entirely. This leads to:

$$[A]_{n+1} = [A]_n + [A]_n^{\,2} + [B]_n^{\,2} + 2[C]_n[D]_n + g_a$$
$$[B]_{n+1} = [B]_n + 2[A]_n[B]_n + [C]_n^{\,2} + [D]_n^{\,2} + g_b$$
$$[C]_{n+1} = [C]_n + 2[A]_n[C]_n + 2[B]_n[D]_n + h_a$$
$$[D]_{n+1} = [D]_n + 2[A]_n[D]_n + 2[B]_n[C]_n + h_b$$

Given that proteins degrade, we can also include the concept of a protease that is capable of degrading A, B, C and D. The following reaction denotes what happens when a protease, P, acts on a substrate, S, and produces a degraded product, D.

$$P + S \rightarrow P + D$$

Adding the effects of a protease to this system leads to:

$$[A]_{n+1} = [A]_n e^{-k_A \times t \times [P]} + [A]_n^{\,2} + [B]_n^{\,2} + 2[C]_n[D]_n + g_a$$
$$[B]_{n+1} = [B]_n e^{-k_B \times t \times [P]} + 2[A]_n[B]_n + [C]_n^{\,2} + [D]_n^{\,2} + g_b$$
$$[C]_{n+1} = [C]_n e^{-k_C \times t \times [P]} + 2[A]_n[C]_n + 2[B]_n[D]_n + h_a$$
$$[D]_{n+1} = [D]_n e^{-k_D \times t \times [P]} + 2[A]_n[D]_n + 2[B]_n[C]_n + h_b$$

Part 5: Life Is Not Exactly Like The Mandelbrot Set

As we have introduced a parameter for time into the equation for degradation, we will also make synthesis time dependent. This leads to:

$$[A]_{n+1} = [A]_n e^{-k_A \times t \times [P]} + t \times \left([A]_n^{\,2} + [B]_n^{\,2} + 2[C]_n[D]_n + g_a\right)$$
$$[B]_{n+1} = [B]_n e^{-k_B \times t \times [P]} + t \times \left(2[A]_n[B]_n + [C]_n^{\,2} + [D]_n^{\,2} + g_b\right)$$
$$[C]_{n+1} = [C]_n e^{-k_C \times t \times [P]} + t \times \left(2[A]_n[C]_n + 2[B]_n[D]_n + h_a\right)$$
$$[D]_{n+1} = [D]_n e^{-k_D \times t \times [P]} + t \times \left(2[A]_n[D]_n + 2[B]_n[C]_n + h_b\right)$$

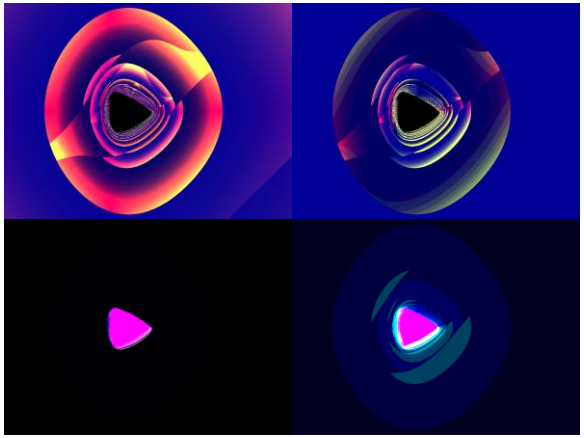
There are still factors of 1 and 2 in front of some of the products. As there is no compelling reason why these factors need to be 1 or 2, they are replaced by additional factors denoted by *f*:

$$[A]_{n+1} = [A]_n e^{-k_A \times t \times [P]} + t \times \left(f_{A[A][A]}[A]_n^{\,2} + f_{A[B][B]}[B]_n^{\,2} + f_{A[C][D]}[C]_n[D]_n + g_a\right)$$
$$[B]_{n+1} = [B]_n e^{-k_B \times t \times [P]} + t \times \left(f_{B[C][C]}[C]_n^{\,2} + f_{B[D][D]}[D]_n^{\,2} + f_{B[A][B]}[A]_n[B]_n + g_b\right)$$
$$[C]_{n+1} = [C]_n e^{-k_C \times t \times [P]} + t \times \left(f_{C[A][C]}[A]_n[C]_n + f_{C[B][D]}[B]_n[D]_n + h_a\right)$$
$$[D]_{n+1} = [D]_n e^{-k_D \times t \times [P]} + t \times \left(f_{D[A][D]}[A]_n[D]_n + f_{D[B][C]}[B]_n[C]_n + h_b\right)$$

In order for these equations to lead to the original Mandelbrot set, *t* should be equal to 1 and [A], [B], [C] and [D] need to be fully degraded in each of the iterations. Thus, the product of each *k* and [P] needs to be large. Importantly, it is still necessary to subtract in order to get a result that looks like the original Mandelbrot set and this is not compatible with protein concentrations. **Thus, the Mandelbrot set, while illustrative of the concepts, may not be a possible example of a cellular process.**

However, this equation can indeed use plausible positive values and generate the following result:

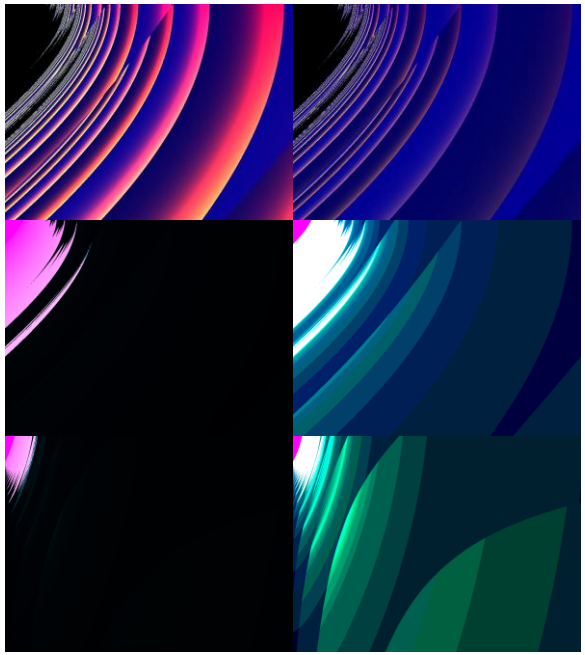
In this figure, the following constants are used: t = 1.0, h_a = 0.01, h_b = 0.01, [P] = 1.0, k_A = 0.7, k_B = 0.7, k_C = 0.7, k_D = 0.7, $f_{A[A][A]}$ = 1.5, $f_{A[B][B]}$ = 1.1, $f_{A[C][D]}$ = 0.2, $f_{B[C][C]}$ = 4.1, $f_{B[D][D]}$ = 4.0, $f_{B[A][B]}$ = 1.2, $f_{C[A][C]}$ = 2.0, $f_{C[B][D]}$ = 1.0, $f_{D[A][D]}$ = 1.3 and $f_{D[B][C]}$ = 1.3. g_a and g_b are varied along the x and y axes for g_a values between -5 and 5 and g_b values between -3.75 and 3.75. Up to 400 iterations are performed as long as [A], [B], [C] and [D] remain below 300. Cell division takes place if [D] rises above 0.1. For points that are not part of the set, the final [A] and [B] values (top left panel) and [C] and [D] values (top right panel) are depicted in red and green, respectively. Blue values alternate between bright and medium intensities to show the number of iterations that were needed before it was determined that the point is not part of the set. Points that are part of the set are left black in the top panels. In the bottom panels, points that are part of the set are bright red and the blue intensity is indicative of the number of iterations that were performed (400 being bright blue that coupled with bright red lead to magenta). The green intensity is indicative of the amount of cell divisions that have taken place and leads to white when combined with magenta. The intensities shown in the lower left panel show green and blue intensities where 0 iterations or divisions result in the darkest color while 400 iterations or divisions result in the brightest color. The green and blue intensities shown in the lower right panel have been increased so that even low numbers of iterations or divisions are visible.



Part 6: Minor Changes Lead To Major Differences

When g_a and g_b are only positive, a complex pattern emerges of when cell division takes place and when it does not. In addition, this pattern becomes susceptible to minor changes in the factors used in the equation.

A magnification of the region where g_a and g_b are positive in the previous figure showing what happens if some of the factors in the equations are slightly changed. g_a ranges from 0 (left) to 0.6 (right) and g_b ranges from 0 (top) to 0.45 (bottom). The top and middle panels are colored like the four panels in the previous figure. The bottom two panels show the result when some of the values are subtly changed. $f_{A[A][A]}$ = 1.5 (vs. 2.0), $f_{A[B][B]}$ = 1.5 (vs. 1.1), $f_{A[C][D]}$ = 0.4 (vs. 0.2), $f_{B[C][C]}$ = 3.1 (vs. 4.1), $f_{B[D][D]}$ = 3.0 (vs. 4.0), $f_{C[A][C]}$ = 4.0 (vs. 2.0), $f_{D[A][D]}$ = 1.0 (vs. 1.3), $f_{D[B][C]}$ = 1.0 (vs. 1.3), $h_a = g_a$ (vs. 0.01), h_b = 0.008 (vs. 0.01).



The irregularity of this shape and its susceptibility to changes in the factors used in the equation supports the following concept:

5) In this model it is difficult to predict which constant rates of production (i.e. g_a, g_b, h_a and h_b) lead to conditions that do or do not lead to infinite protein production and which of these rates also led to cell division.

Naturally, this now raises the question of how this could be applicable to *in vivo* cellular processes.

g_a, g_b, h_a and h_b could be seen as constant rates of production that are each due to the concentration of one transcription factor. Thus, [A] would increase by g_a through the constant activity of one transcription factor.

There are many ways through which transcription factors can regulate transcription. For instance, their concentrations can be important as increasing concentrations drive their associations with targets. In addition, mutations affecting their ability to bind DNA or recruit additional transcriptional machinery can alter their ability to regulate transcription.

Part 7: The Relevance To Diseases, Transcription Factors And Microarrays

While potential changes in transcription factor concentrations can be detected by microarrays, changes in the activity thereof cannot. As g_a, g_b, h_a and h_b relate to the ability of transcription factors to produce A, B, C and D and not necessarily to the concentration of these transcription factors, this leads to the following concept:

6) The observed levels of some proteins or their transcripts may differ from normal conditions due to changes that are not related to the concentrations of transcription factors or the transcripts thereof.

Now that g_a, g_b, h_a and h_b can be related to cellular processes, this raises the question of how the products and squares of [A], [B], [C] and [D] could affect transcription.

Collision theory provides the answer as the relative number of collisions between molecules is proportional to the product of their concentrations.

Therefore, A, B, C and D can affect transcription in this model if their ability to do so is dependent on their interaction with one another.

Ideally, these interactions would have high dissociation rates so that the binding of two monomers to form a complex only temporarily reduces the concentrations of the monomers. These complexes can then either temporarily act as transcription factors or catalyze the activation of transcription factors, directly or indirectly. The transcription factors then lead to the production of additional monomers. This leads to the following concepts:

7) The previously mentioned equations and/or modifications thereof can apply to physiological processes.

8) The outcomes of physiological processes that are influenced by the interactions of two or more molecules are nearly impossible to predict unless one knows exactly how the molecules interact.

This explains why the underlying cause of a disease may be difficult to detect as it may be a subtle alteration in the concentration or activity of a single previously unknown protein. While potential changes in concentration can be detected by microarray, the difference between a healthy and a diseased state may be too small to detect; and a change in activity is impossible to detect directly. Thus when studying processes using microarrays, many of the genes responsible for the phenotype of a differentiated cell will be detectable while genes that are responsible for controlling the differentiation may not be. This leads to the following concept:

9) Microarrays are perfectly good for identifying differences while they may give little insight into the underlying cause or causes of the differences.

Part 8: Determining The Duration Of Receptor (e.g. TCR) Ligation

In order for activities to be dependent on the products and squares of concentrations, the interactions need to be short lived so that the concentrations of monomers are not greatly lowered through the formation of complexes.

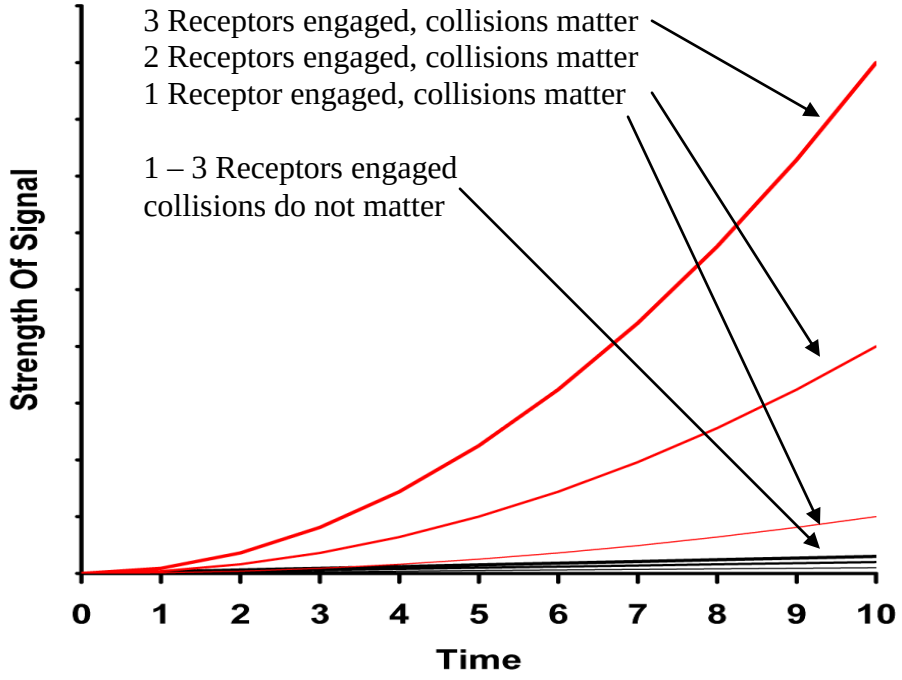
If an interaction has a high association rate and a low dissociation rate then the concentration of the complexes will be linearly dependent on the lower monomer concentration. That is, if B is in excess, then each additional A will bind to form AB and thus the amount of AB is dependent on the amount of A that is produced.

If, on the other hand, an interaction has a high dissociation rate then concentrations of short-lived AB complexes are dependent on the products of the concentrations of A and B.

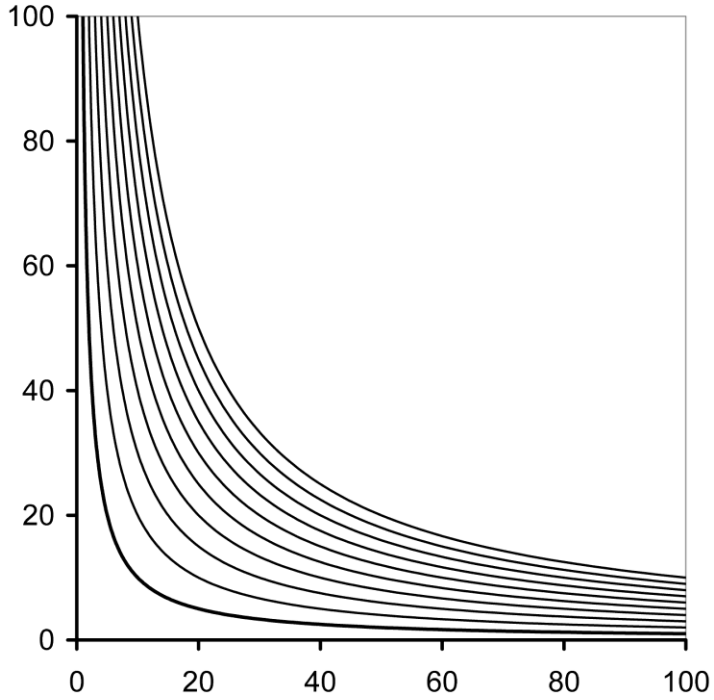
This has implications as such complexes, while potentially capable of influencing cellular processes, are impossible to detect through techniques such as immunoprecipitation that rely on the formation of stable complexes.

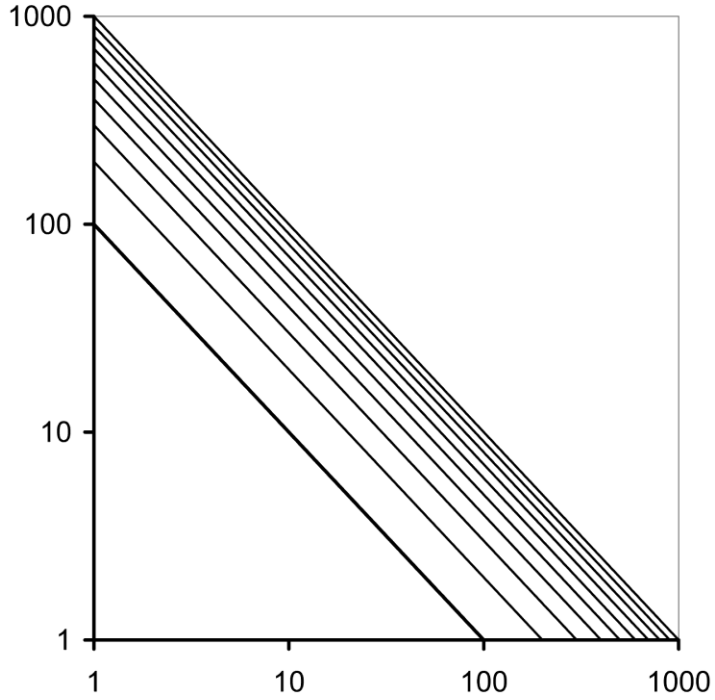
However, these complexes may be very important…

Time-dependent processes may be dependent on such fast dissociation rate interactions. For a cell, determining how long a receptor has been engaged is important. If engagement leads to a catalytic activity, then the amount of product formed from substrate is proportional to the duration of the engagement **(black lines)**. At the same time, it is also proportional to the amount of engaged receptors. However, as products are formed linearly and diffuse away from their receptors, product collisions should increase based on the square of product concentrations. Thus, even if many catalytic receptors are briefly engaged, there will not be many product collisions. However, if even a few are engaged for a longer time, then product collisions will begin to increase rapidly **(red lines)**. If such product collisions can be measured, perhaps through a catalytic activity that is only present when two colliding products briefly form a complex, then these collisions can be used to determine whether or not a long-duration receptor engagement has taken place.



Part 9. What To Look For By Intracellular FACS





The rate of collision of two molecules is proportional to the product of the concentrations of these two molecules.

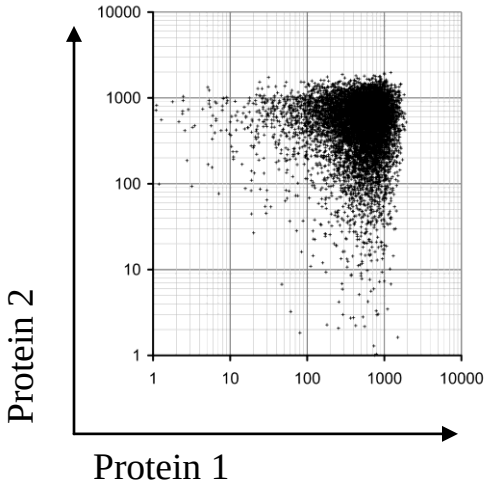
If a constant rate of collision between two molecules, or variants thereof (e.g. phosphorylated), is important for cellular survival, then their concentrations in population of cells may distribute as above. The concentrations of the two molecules are represented by the two axes and the thick curved line represents values where the product is 100. The thin lines represent values where the product is 200, 300, 400, 500, 600, 700, 800, 900 and 1000.

Thus, if the distribution of concentrations of two molecules in a population of cells takes the shape of a downward diagonal on a logarithmic plot, then this suggests that the rate of collision between these two molecules, or variants thereof, is kept constant and may be important.

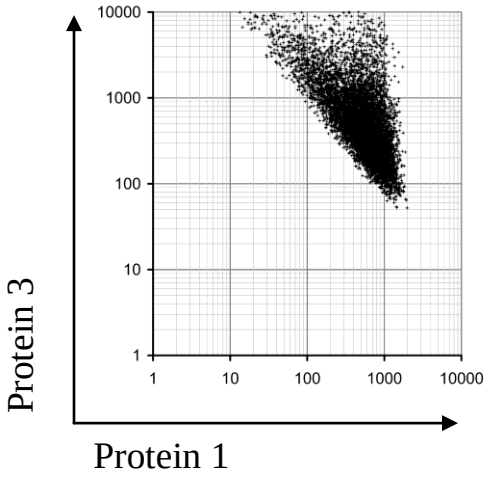
10. But It Might Not Be That Easy

One can now easily imagine situations where collisions between three (instead of two) proteins matter…

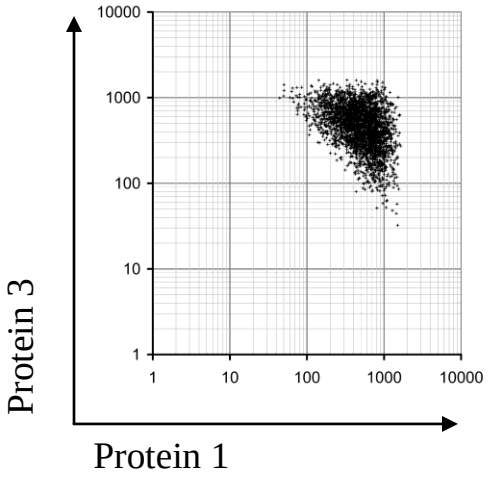
For this system, assume that the expression of two proteins in a population of cells has a normal distribution (i.e. cells express a certain amount on average but might also express a bit more or less of each protein). In addition, the expression of each protein has no effect on the expression of the other. By intracellular FACS, this would lead to a plot that looks similar to the following:



If collisions between these two proteins and a third protein are important and if the expression of the third protein is adjusted dependent on the expression of the other two proteins, then the resulting intracellular FACS plot of the third protein in relation to the first or second protein would look as follows:



On the other hand, if the expression of all three proteins has a normal distribution and only certain numbers of collisions between these proteins are compatible with life, then the intracellular FACS plot would look as follows:



Conclusions

- The Mandelbrot set is illustrative of many important concepts but may not be directly descriptive of any cellular process
- Interactions with fast dissociation rates may be very important in cellular processes
- Interactions with fast dissociation rates may be the reason why some experiments are highly dependent on seemingly irrelevant experimental conditions
- Interactions with fast dissociation rates could allow cells to determine how long their receptors are bound to a ligand
- If a signaling pathway branches, collisions between molecules in different downstream braches can be used to determine how long upstream events are taking place
- Molecules that interact with fast dissociation rates are difficult to identify
- The phenotypes caused by molecule interactions cannot be used to identify the equations of the underlying interactions
- It will be necessary to identify the association and dissociation rates for most proteins with most other proteins in order to obtain the equations
- With only 30,000 proteins, that leads to only about a billion experiments that need to be performed … assuming that we are only concerned with collisions between two (and not three) molecules