# Complexity in Biological Systems

**Alessandro P. S. de Moura (Aberdeen, UK)**

# Why is biology so hard to model?

- Coexistence of widely different relevant spatial scales (a few nanometres to metres)
- Coexistence of widely different time scales (microseconds to years)
- Out-of-equilibrium systems with many degrees of freedom, highly coupled to the environment
- The role of evolution in shaping existing biological systems means that they have a history, which may be relevant for modelling

# Consequence of complexity of biological systems: no clear-cut theoretical framework

Many areas of mathematics and theoretical physics have been used in biology:

- Deterministic models with ODEs and PDEs
- Stochastic processes (Markov chains, etc)
- Statistical physics (equilibrium and non-equilibrium)
- Concepts from systems and control engineering
- Bayesian statistics and other inference and statistical methods
- Machine learning and other computer science ideas
- …...

# Mathematical Modelling of DNA Replication

**Alessandro P. S. de Moura (Aberdeen, UK)**
Jens Karschau (Aberdeen, UK)
Conrad Nieduszynski (Nottingham, UK)
Renata Retkute (Nottingham, UK)
Michelle Hawkins (Nottingham, UK)

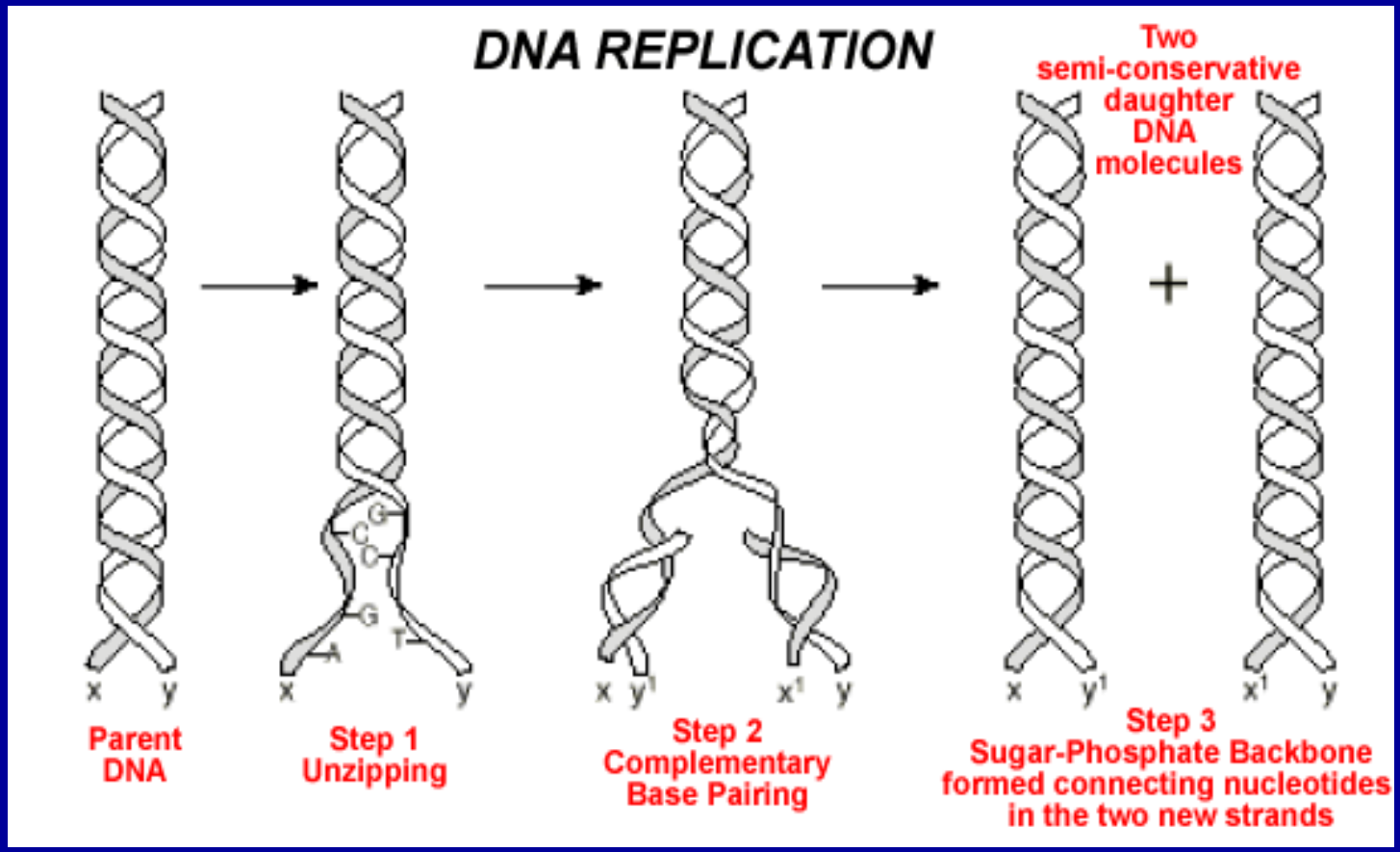
# References


- “Mathematical modelling of whole chromosome replication”, A. de Moura, R. Retkute, M. Hawkins, C. Nieduszynski, *Nucleic Acids Research* **38,** 5623 (2010).
- “Dynamics of DNA replication in yeast”, R. Retkute, C. Nieduszynski, A. de Moura, *Physical Review Letters* **107,** 068103 (2011).
- “Optimal placement of origins for DNA replication”, J. Karschau, J. Blow, A. de Moura, *Physical Review Letters* **108,** 058101 (2012).

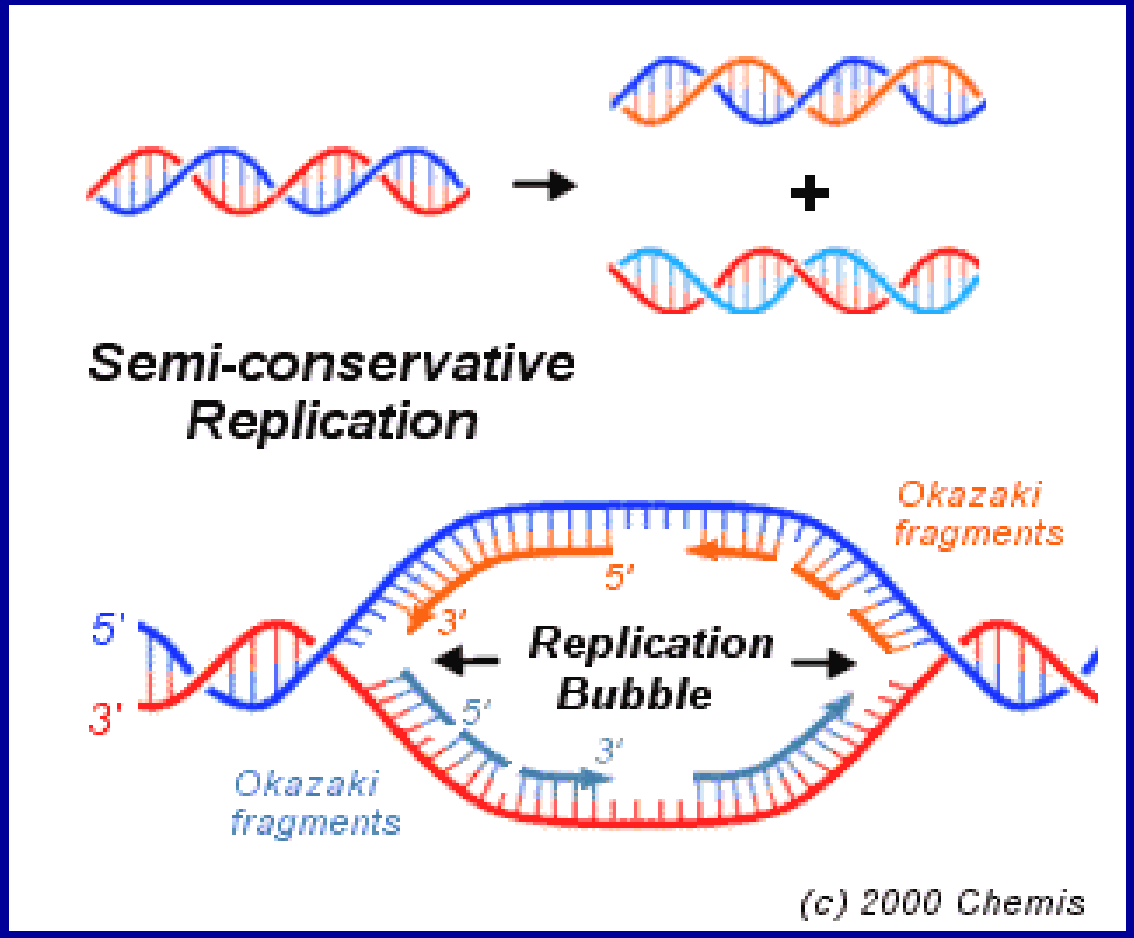
# Summary

- A physicist's version of the biology of DNA replication.
- Building a model of DNA replication.
- Simple consequences and predictions of the model.
- Applying the model for Chromosome VI: (preliminary) comparison with the data.
- Next step: optimization algorithm to determine intrinsic activation times and other parameters.
- Conclusion and perspectives.

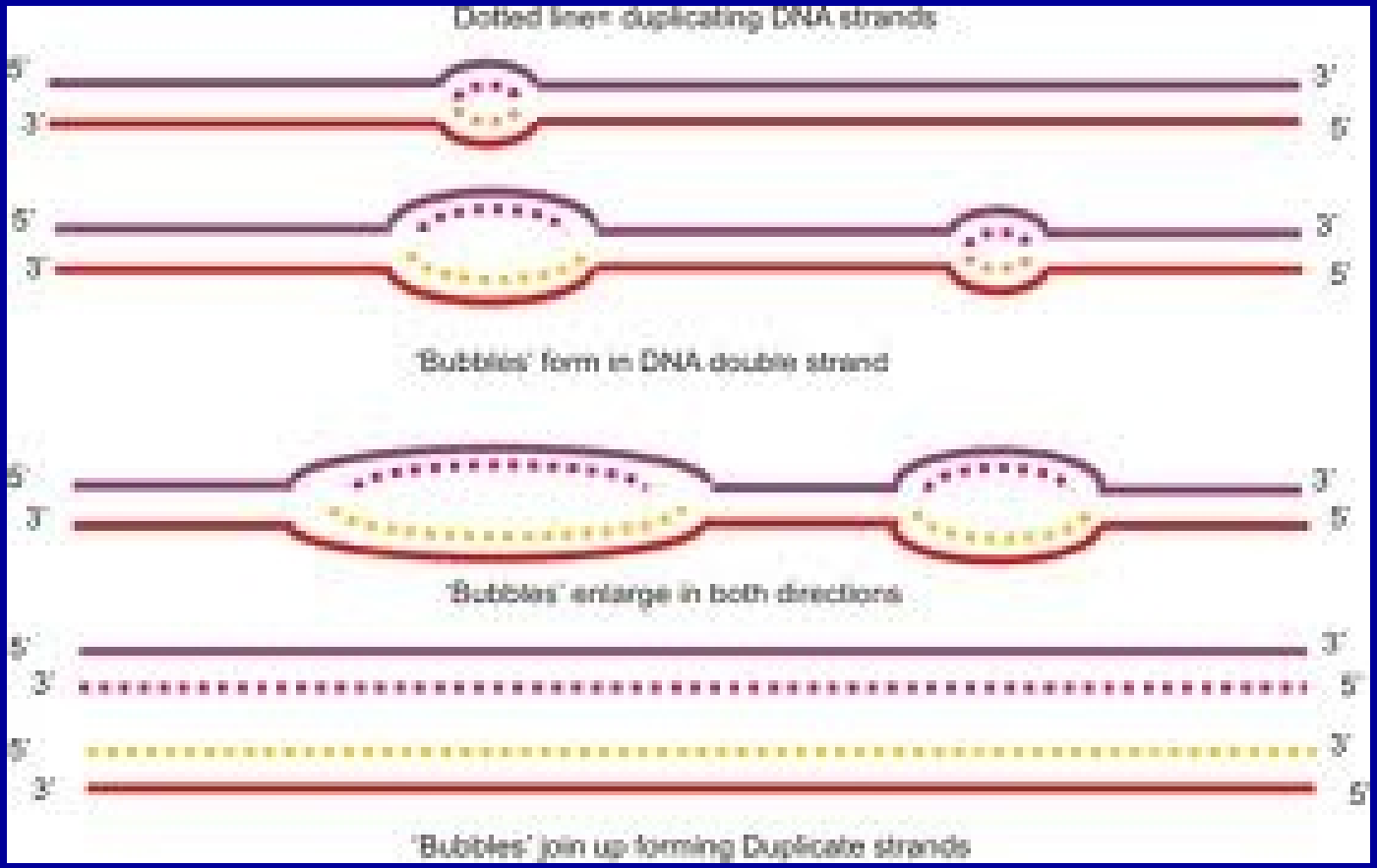
# Basic biology of DNA replication



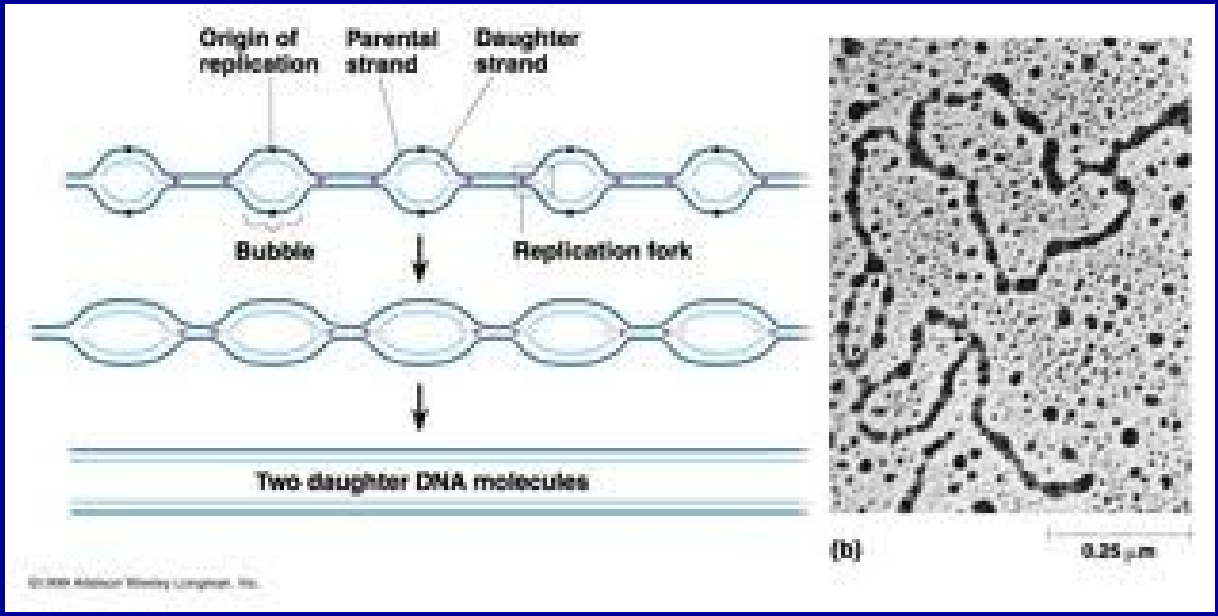
# Basic biology of DNA replication



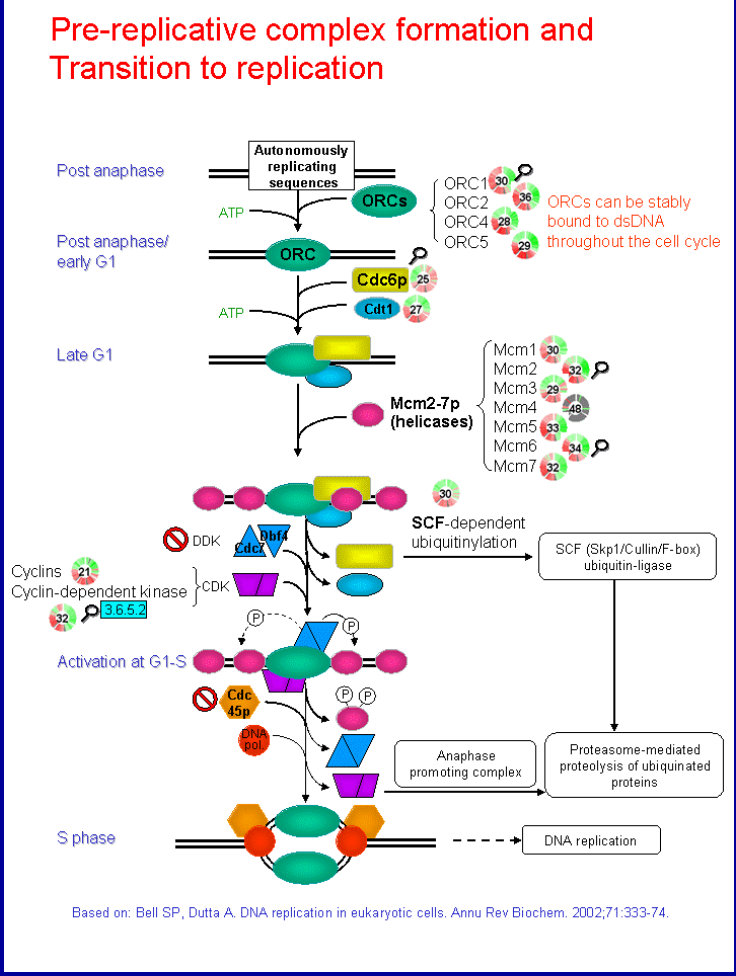
# Replication forks



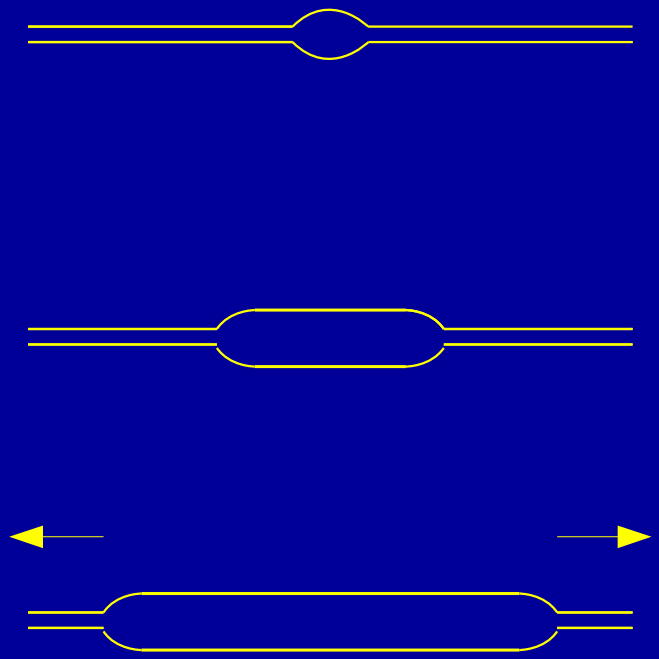
# Replication forks caught in action



# Complexity of fork activation

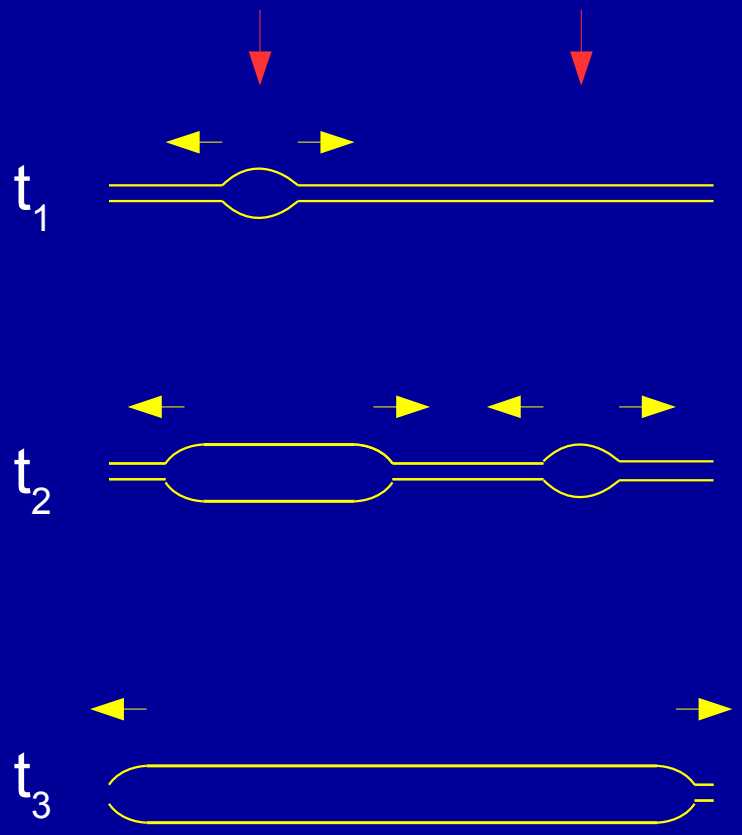


# Basic biology of DNA replication



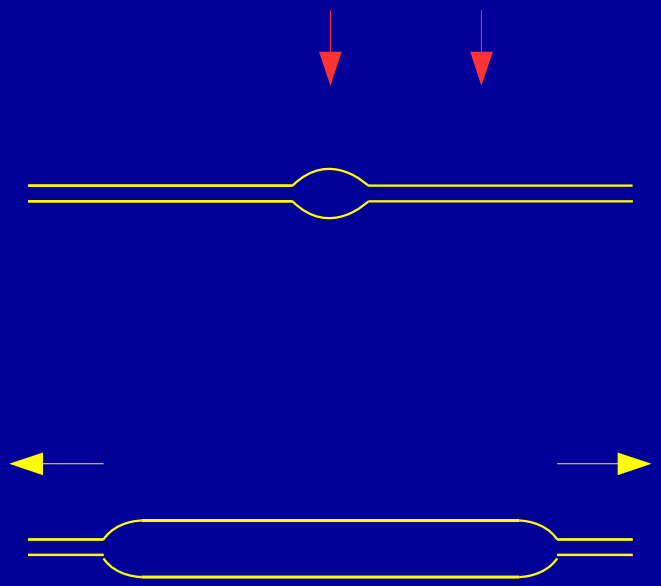
- Replication starts in well defined points in the chromosome – the **origins**.
- Once started, the replication fork travels at **constant average speed** through the chromosome.
- Speed: 3.7 kilobases per minute.

# Basic biology of DNA replication



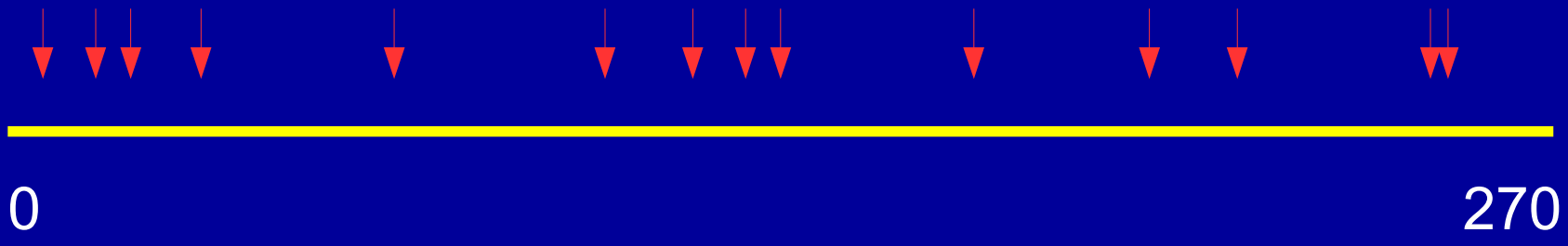

- There are many origins in each chromosome.
- Replication starts in each origin independently. Usually, multiple replication forks move simultaneously in each chromosome.
- The starting time for replication is random at each origin.
- Each origin has its average starting time – some start early, some late.
- When two replication waves meet, they merge.

# Basic biology of DNA replication



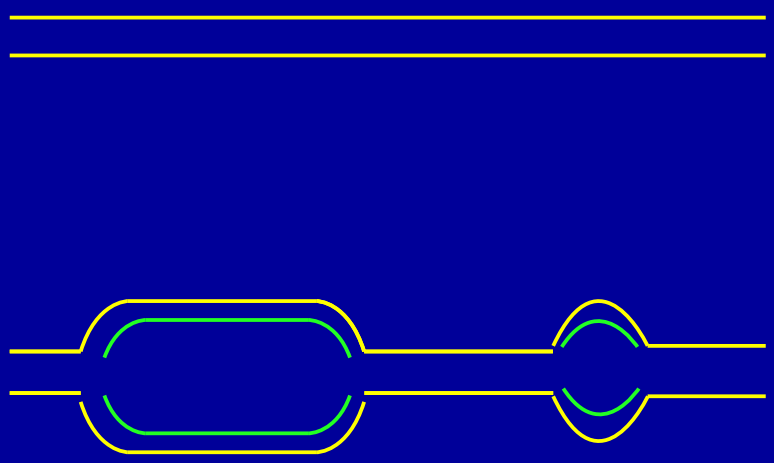
- If an origin is overrun by a replication wave, it is inactivated, and does not have the chance of starting another wave.
- Origins also have an intrinsic probability of not activating at all, independently of the action of other origins.

# Map of origins in Chromosome VI of *S. cerevisiae*




**Important question:** what is the replication time at each point in the chromosome?

# Measuring the replication time

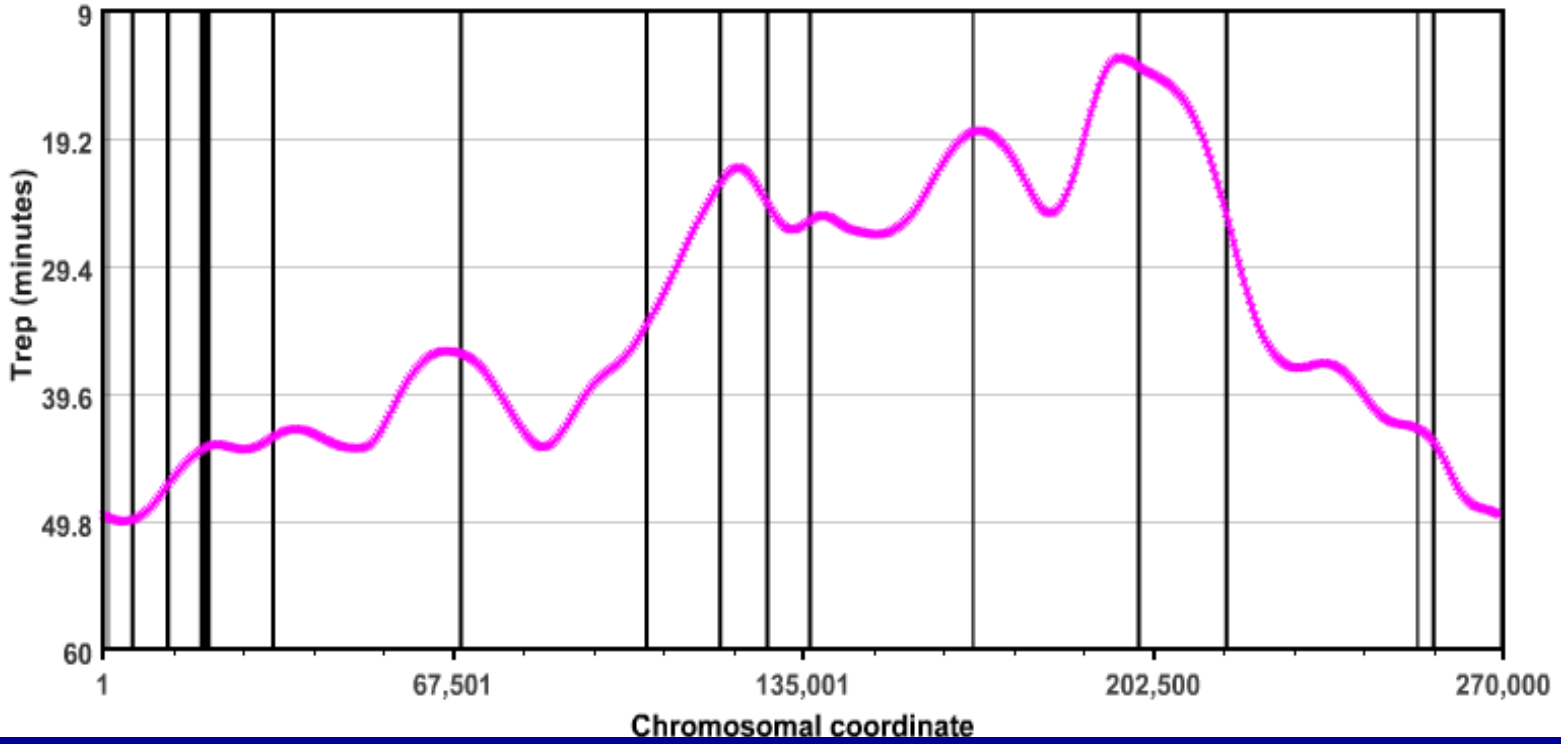


Heavy isotopes

Light isotopes

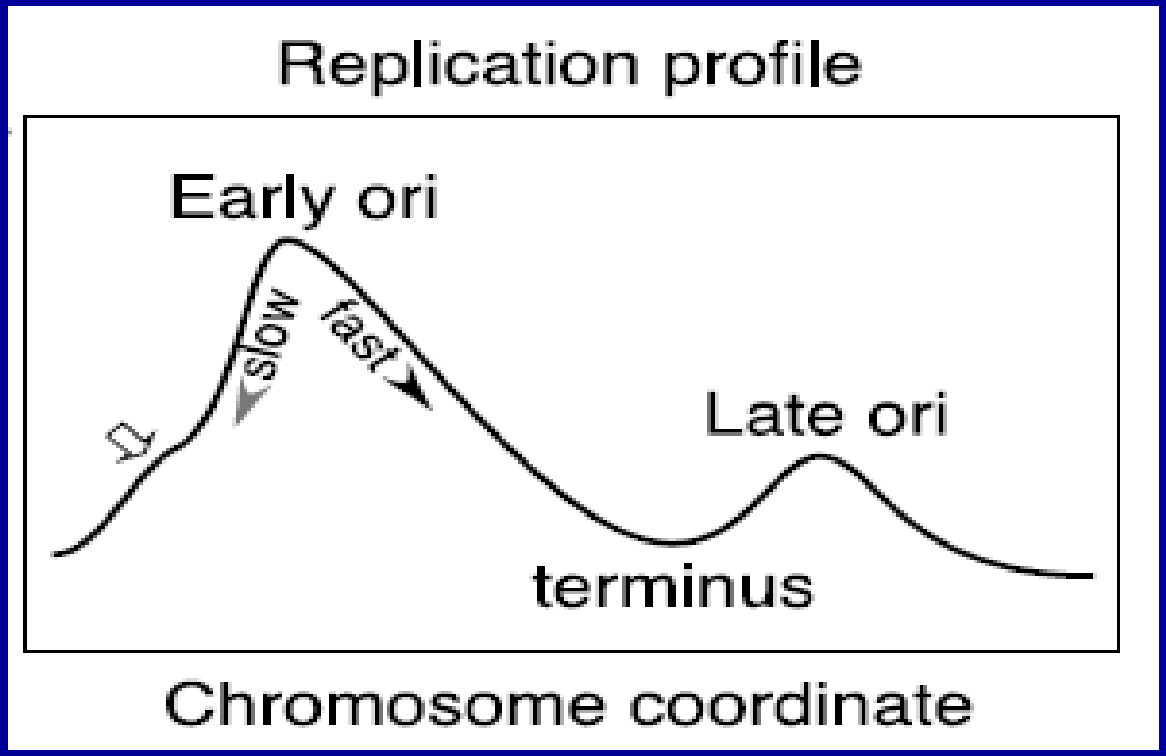
Cells are grown for many generations in a medium with the heavy isotopes $C^{13}$ and $N^{15}$. They are then transferred to a medium with normal isotopes. The old DNA strands are then isotopically heavy, and the new strands added by replication are light. The DNA is then cut at specific points by a restriction nuclease, and from the proportions of light-to-heavy isotopes for each fragment, the fraction of the population where replication occurred at that point in the chromosome can be found.

# Result for chromosome VI of *S. cerevisiae*




Raghuraman *et al.*, 2001

# The interpretation of time data can be tricky...




Raghuraman *et al.*, 2001

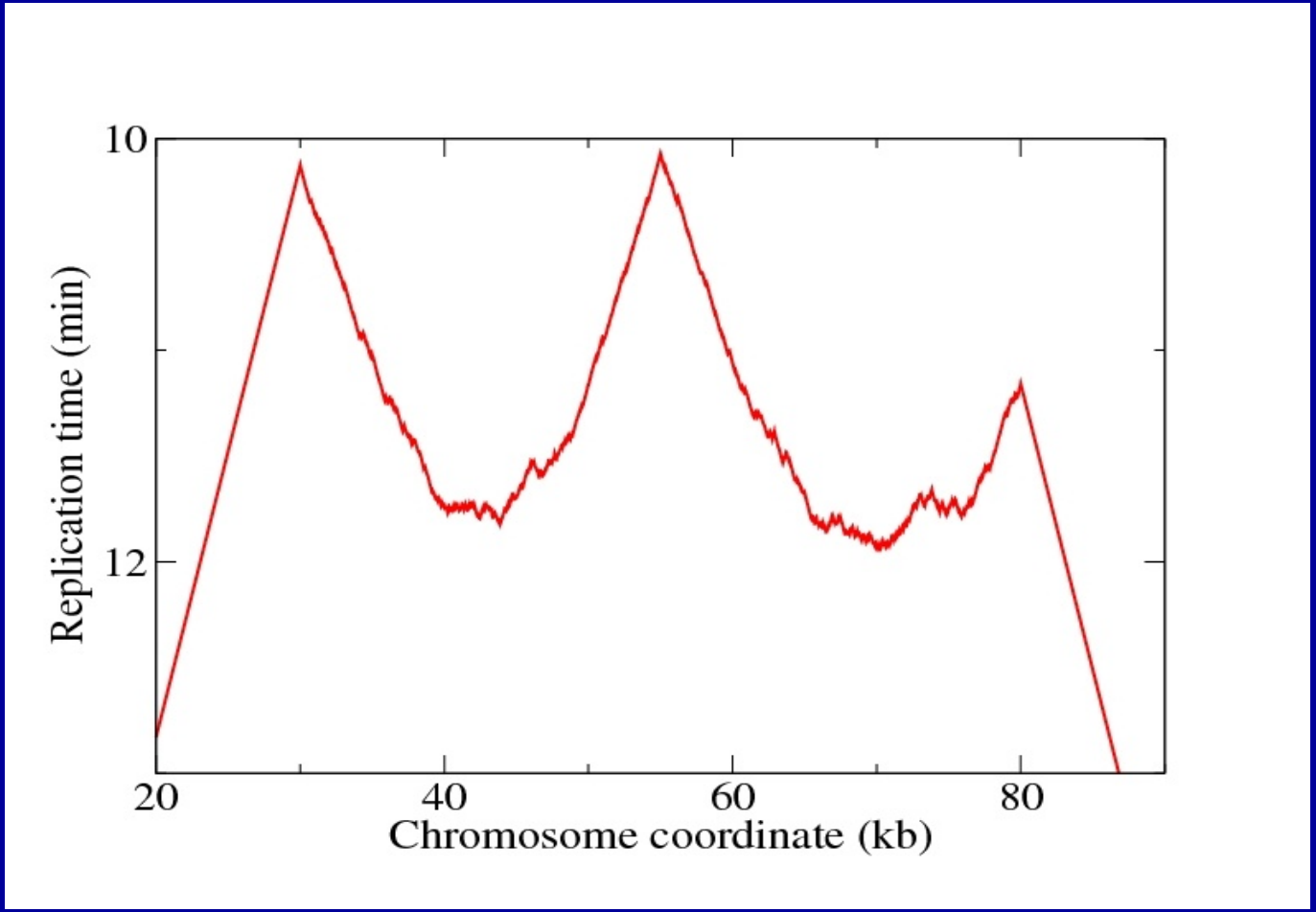
# Why a model is useful

- The relation between the observed replication time and the intrinsic quantities characterizing the origins is not straightforward.
- In particular:
  - The slope of the curve is **not** the speed of the replication forks.
  - The peaks do **not** give directly the replication time. This implies that a higher peak does **not** imply an earlier activation.
  - The averaging process used in smoothing the curve has important and sometimes unforeseen consequences.
- A mathematical model would clarify these issues, and once built, it would make new predictions.

# Stochastic model of DNA replication

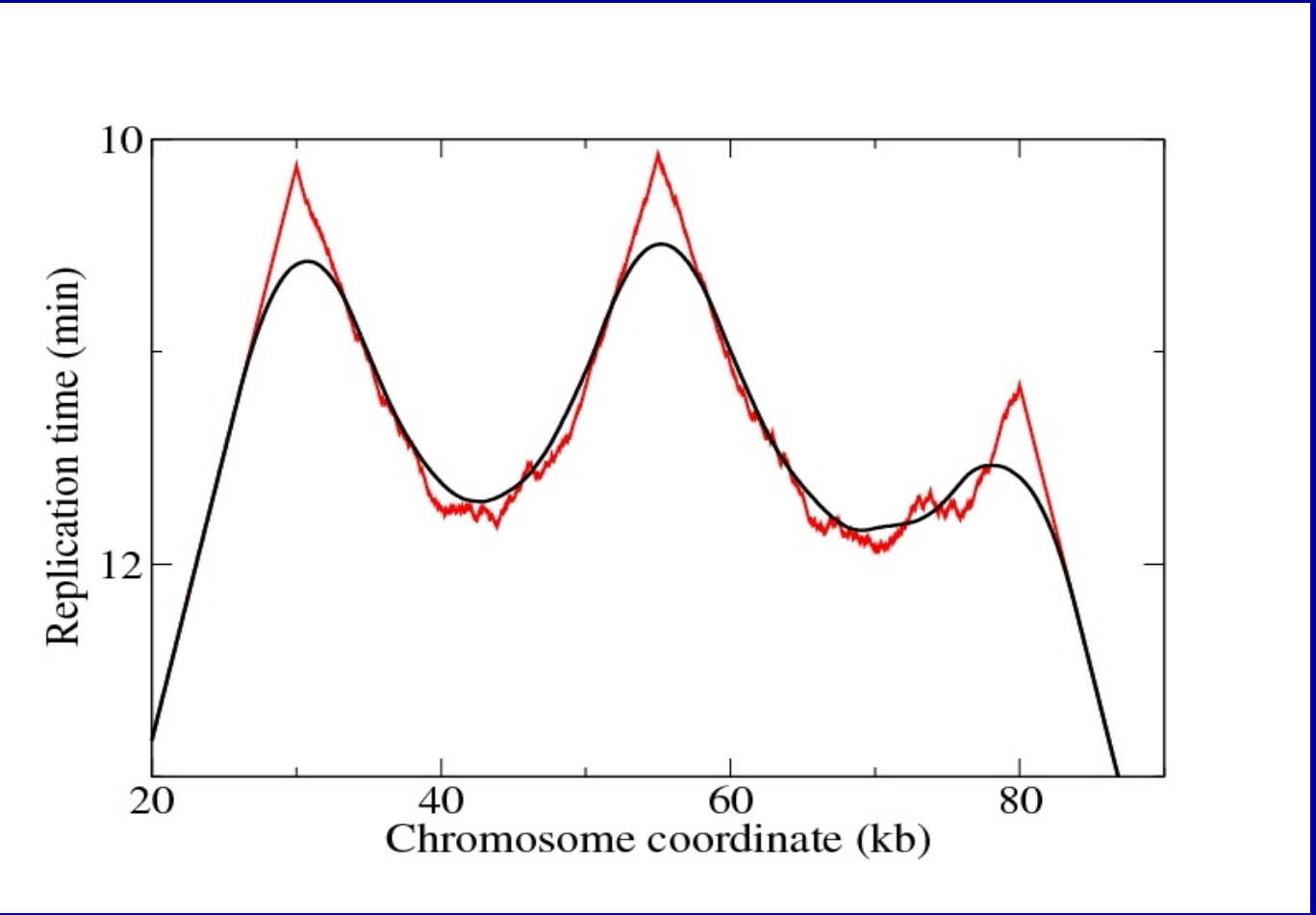
- Each individual replication fork travels with an average speed which is constant throughout the chromosome.
- Each chromosome replicates independently.
- Each origin is activated in a probabilistic way, independently of the other origins.
- Each origin *i* is characterized by the parameters: position $x_i$, average (intrinsic) activation time $T_i$ , standard deviation of activation time $s_i$ , and (intrinsic) activation probability $p_i$ .
- In simulations, a virtual population of chromosomes is created in the computer memory, and in each member the origins are activated randomly according to their statistical parameters.

# A test case... an “artificial” chromosome with 3 origins



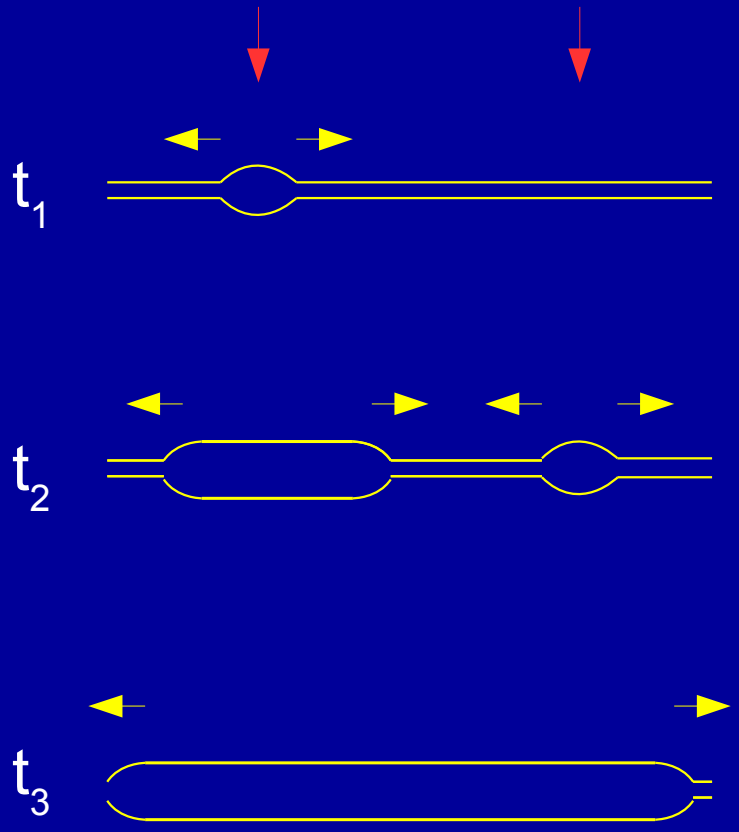

Origins at 35kb, 50kb, and 80kb. All have the same intrinsic activation time of 10 min. The origin at 80kb has an intrinsic efficiency of 0.8, that of the other two is 0.9. The population is 1000.

# Now performing an average on a window of 8kb...



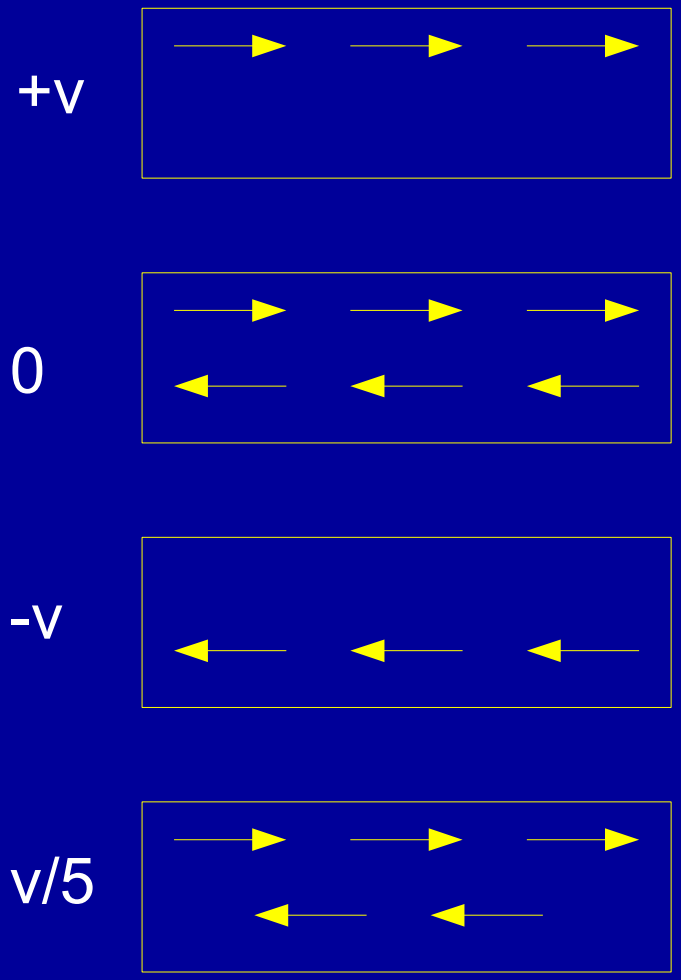

Origins at 35kb, 50kb, and 80kb. All have the same intrinsic activation time of 10 min. The origin at 80kb has an intrinsic efficiency of 0.8, that of the other two is 0.9. The population is 1000.

# The slope is **not** the velocity



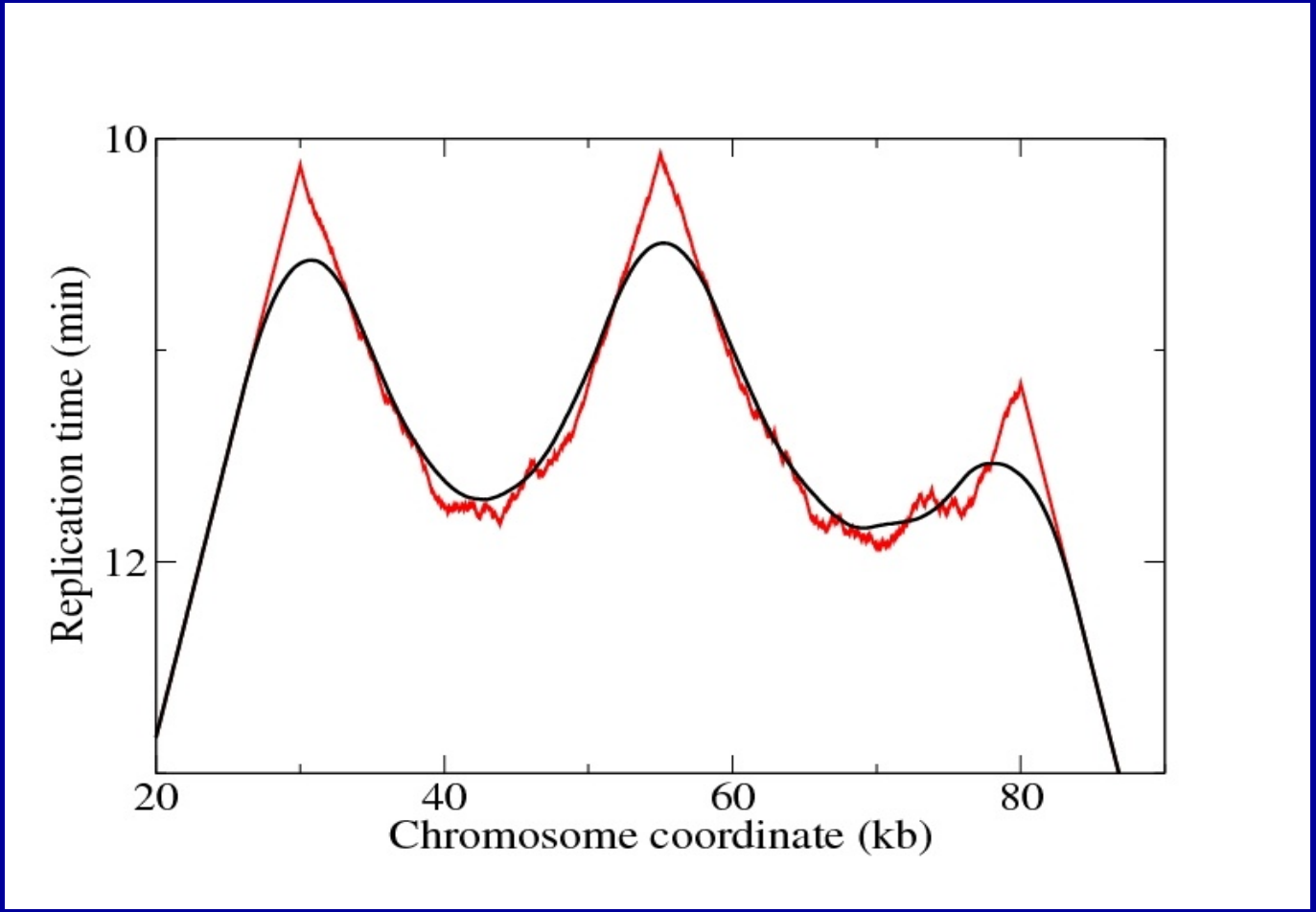

- The replication time is actually a population average.
- At any given position, the replication wave in some cells move left, and in some it moves right, depending on the random activation times of the origins.
- The slope is therefore the average of a positive and a negative velocity; the average is weighted by the proportion of the population whose waves are moving left or right.

# The slope is **not** the velocity



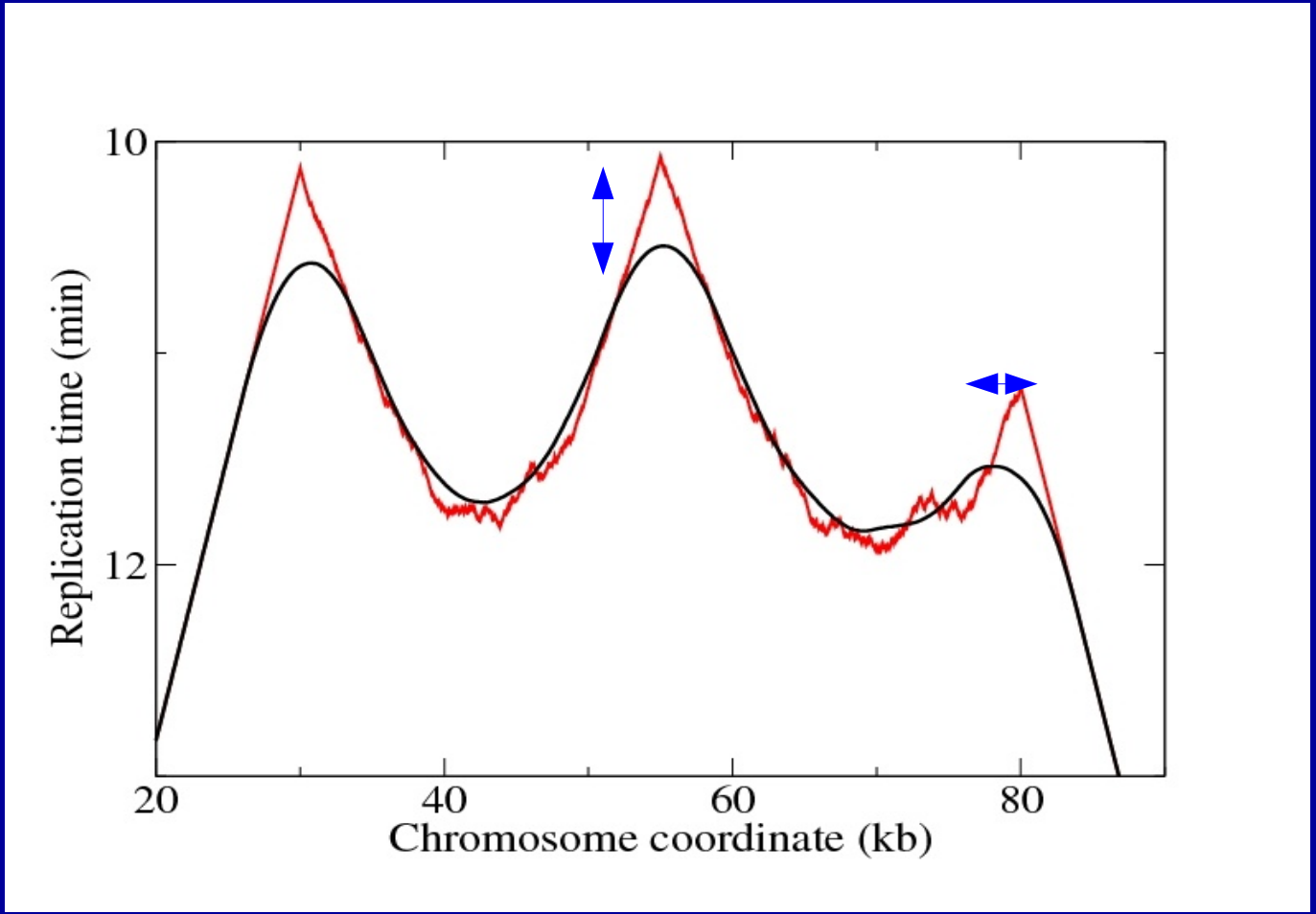

- The replication time is actually a population average.
- At any given position, the replication wave in some cells move left, and in some it moves right, depending on the random activation times of the origins.
- The slope is therefore the average of a positive and a negative velocity; the average is weighted by the proportion of the population whose waves are moving left or right.

# Artificial chromosome with 3 origins



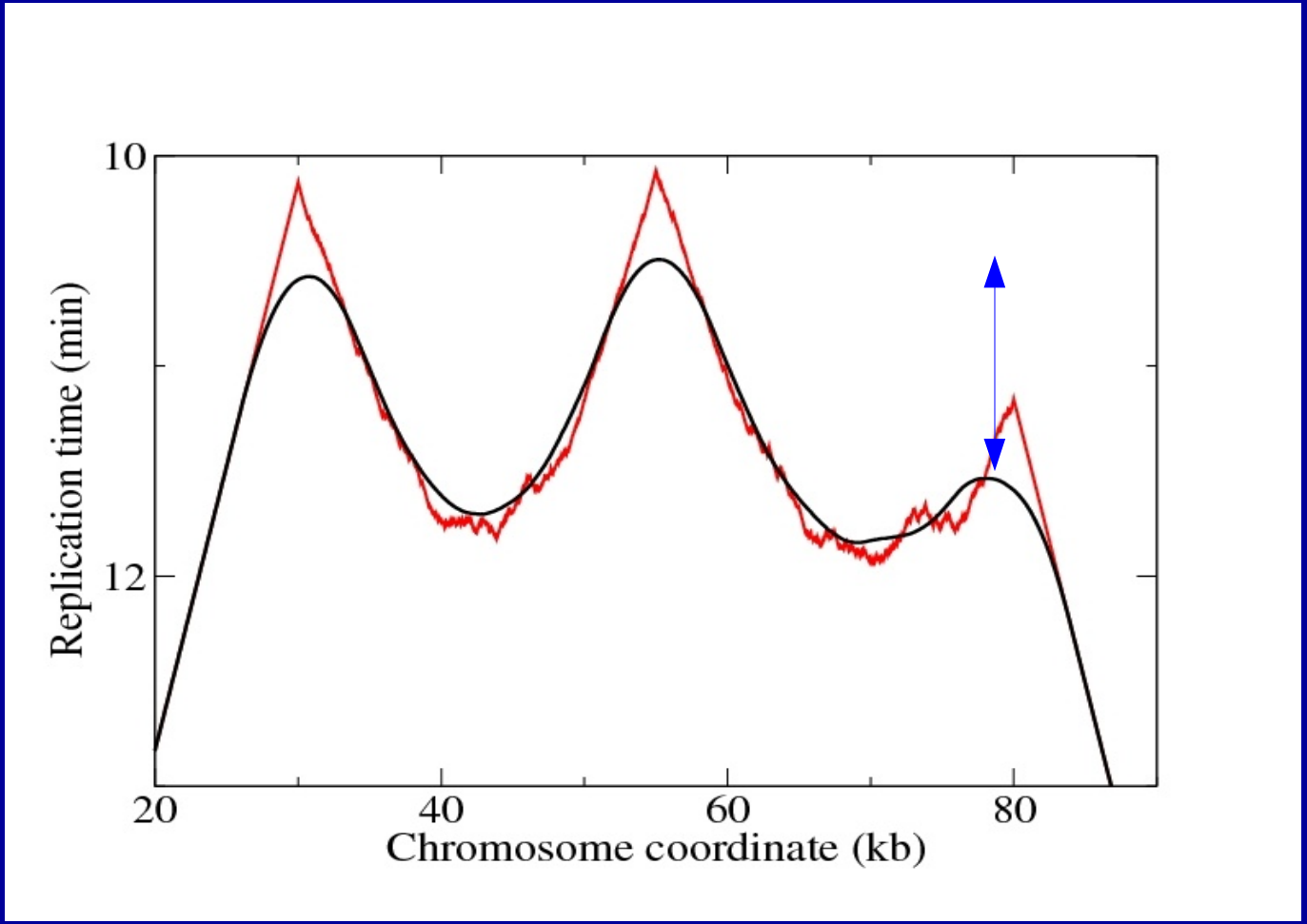

Origins at 35kb, 50kb, and 80kb. All have the same intrinsic activation time of 10 min. The origin at 80kb has an intrinsic efficiency of 0.8, that of the other two is 0.9. The population is 1000.

# The averaging changes the position and the heights of the peaks



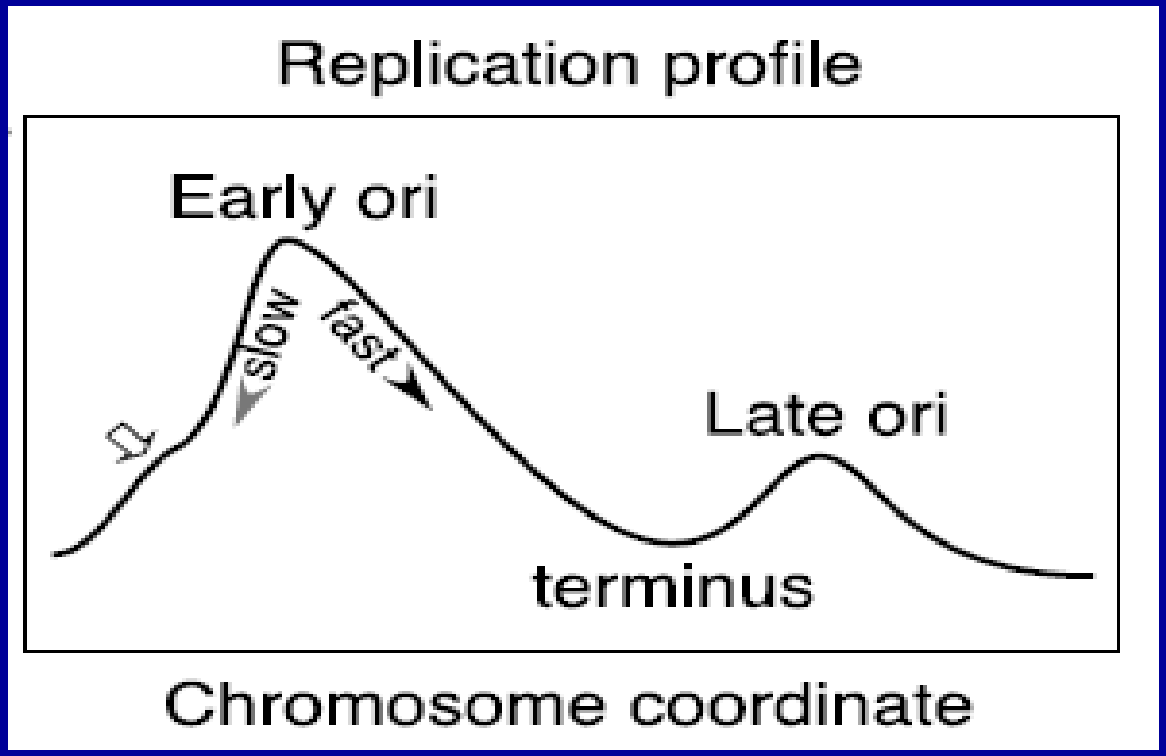

Origins at 35kb, 50kb, and 80kb. All have the same intrinsic activation time of 10 min. The origin at 80kb has an intrinsic efficiency of 0.8, that of the other two is 0.9. The population is 1000.

# The peaks do **not** give the activation times, and a higher peak does not always mean early activation




Origins at 35kb, 50kb, and 80kb. All have the same intrinsic activation time of 10 min. The origin at 80kb has an intrinsic efficiency of 0.8, that of the other two is 0.9. The population is 1000.

# The interpretation of time data can be tricky...

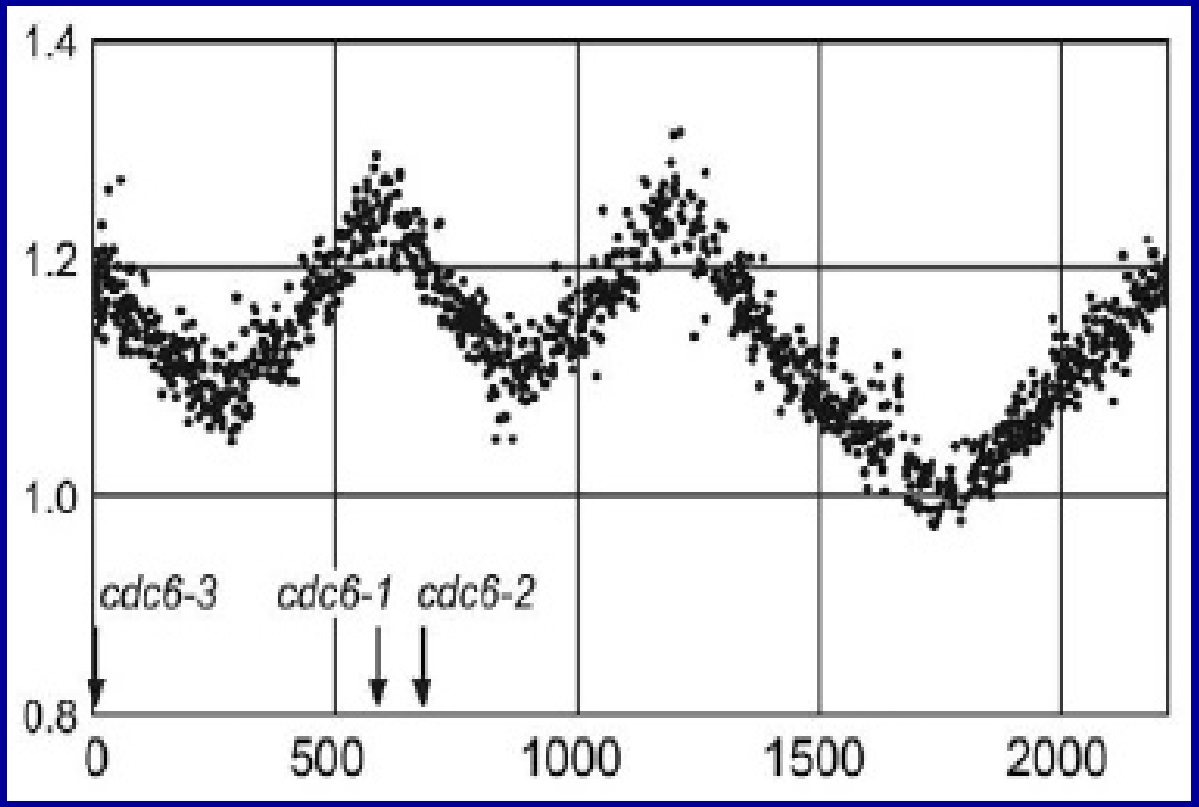


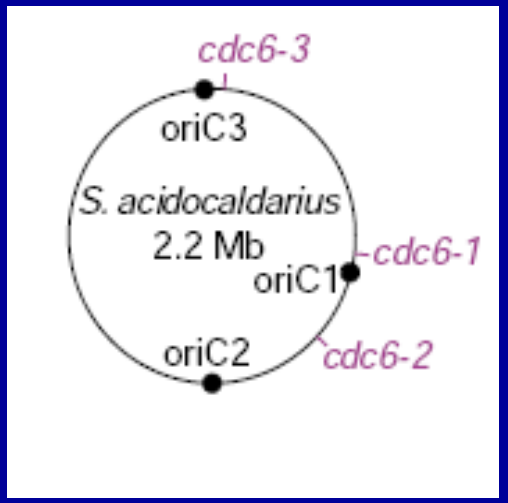

Raghuraman *et al.*, 2001

With “perfect” measurements, and without the averaging, the origins would be sharp peaks, and the valleys between them would be smooth due to the stochasticity in the activation times – the waves would meet at different points in different cells.

If the activation times were not random, the replication profile would be a series of straight lines connected at sharp corners (without averaging).

# Replication times in a simple case: *S. acidocaldarius*




3 origins, starting synchronously

Lundgren M, Andersson A, Chen L, Nilsson P, Bernander R, 2004.

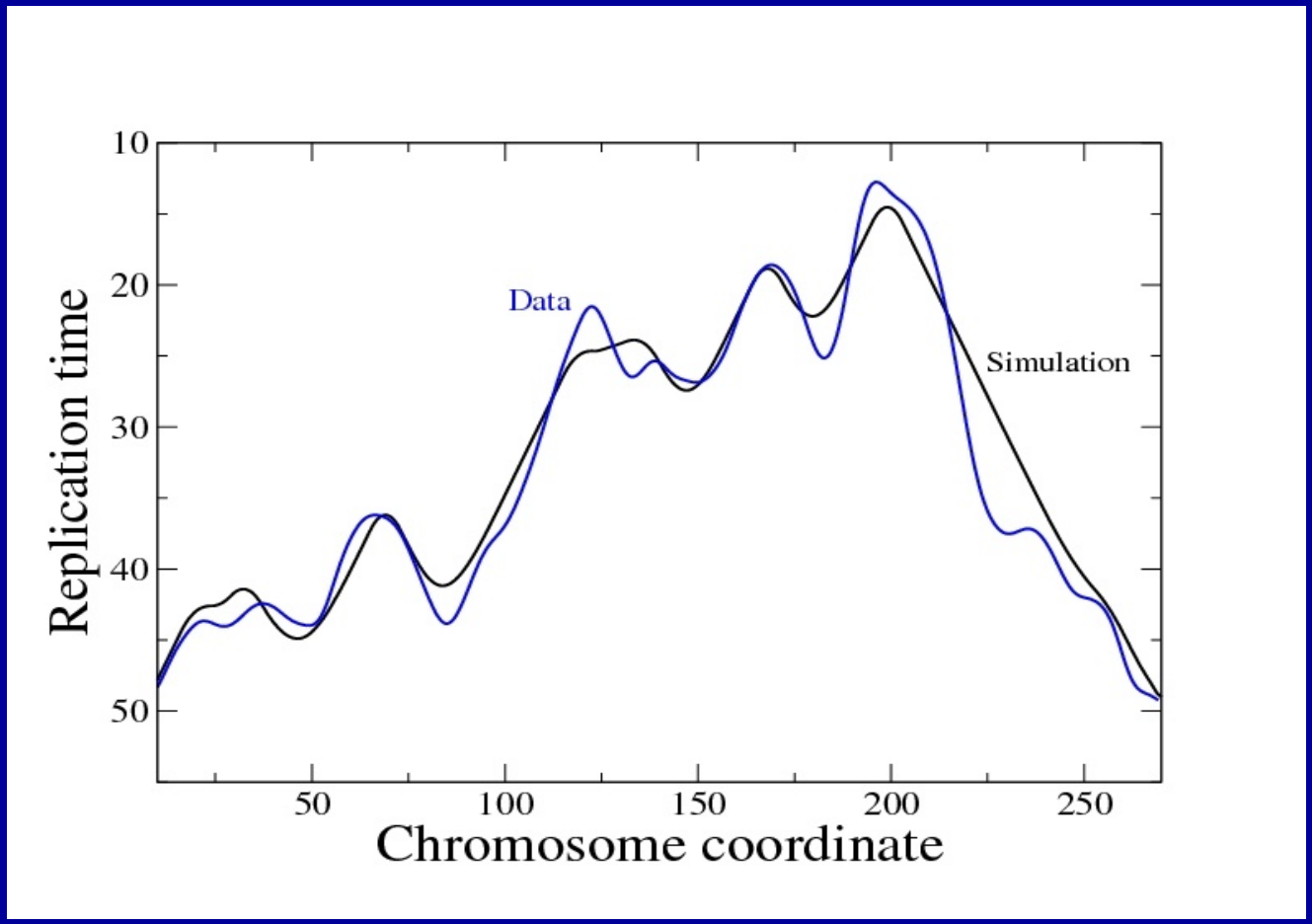
# Parameter estimation methods:

Determining parameters (origin activation times, etc.) from data by using global optimization techniques.

# Estimating activation times

- $t_1 = 42.6$
- $t_2 = 40.4$
- $t_3 = 34.5$
- $t_4 = 25.0$
- $t_5 = 21.2$
- $t_6 = 23.7$
- $t_7 = 17.5$
- $t_8 = 13.1$
- $t_9 = 41.8$
- $t_{10} = 44.7$
- $t_{11} = 55.0$

# Comparison with data



# Estimating activation times: the issue of identifiability

- $t_1 = 42.6$
- $t_2 = 40.4$
- $t_3 = 34.5$
- $t_4 = 25.0$
- $t_5 = 21.2$
- $t_6 = 23.7$
- $t_7 = 17.5$
- $t_8 = 13.1$
- $t_9 = 41.8$
- $t_{10} = 44.7$
- $t_{11} = 55.0$

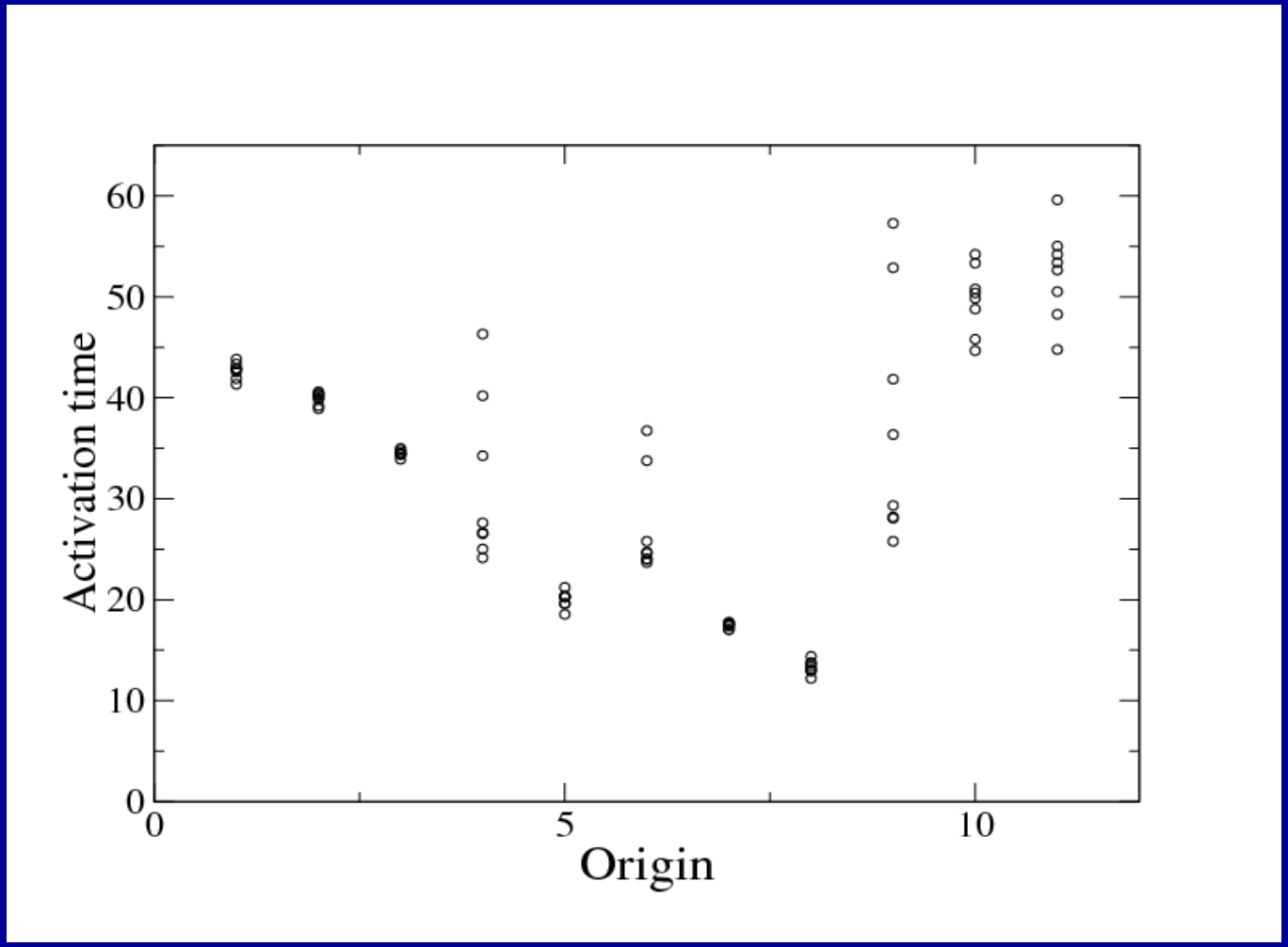
- $t_1 = 42.8$
- $t_2 = 39.9$
- $t_3 = 34.5$
- $t_4 = 46.3$
- $t_5 = 19.6$
- $t_6 = 24.7$
- $t_7 = 17.6$
- $t_8 = 13.1$
- $t_9 = 29.3$
- $t_{10} = 49.8$
- $t_{11} = 52.6$

# Estimating activation times: the issue of identifiability

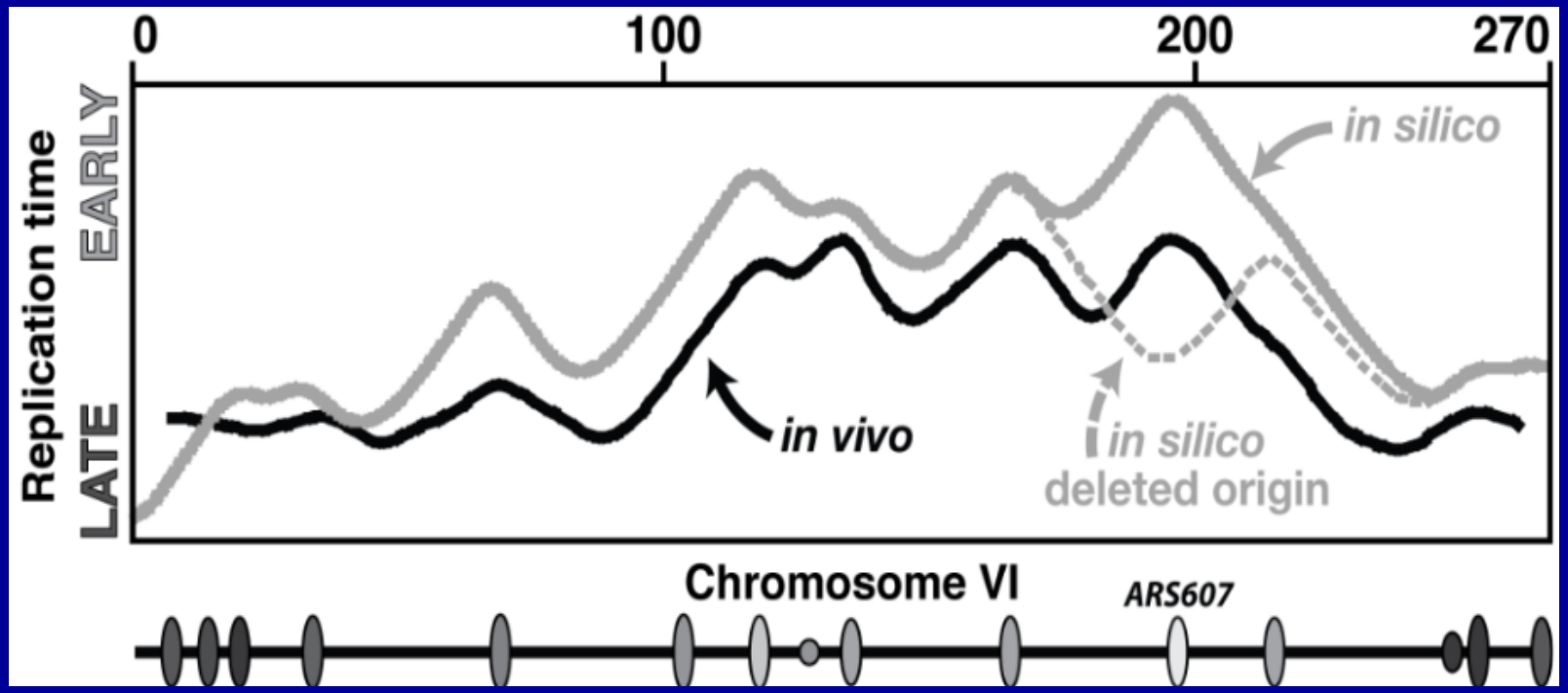
- $t_1 = 42.6$
- $t_2 = 40.4$
- $t_3 = 34.5$
- $t_4 = 25.0$
- $t_5 = 21.2$
- $t_6 = 23.7$
- $t_7 = 17.5$
- $t_8 = 13.1$
- $t_9 = 41.8$
- $t_{10} = 44.7$
- $t_{11} = 55.0$

- $t_1 = 42.8$
- $t_2 = 39.9$
- $t_3 = 34.5$
- $t_4 = 46.3$
- $t_5 = 19.6$
- $t_6 = 24.7$
- $t_7 = 17.6$
- $t_8 = 13.1$
- $t_9 = 29.3$
- $t_{10} = 49.8$
- $t_{11} = 52.6$

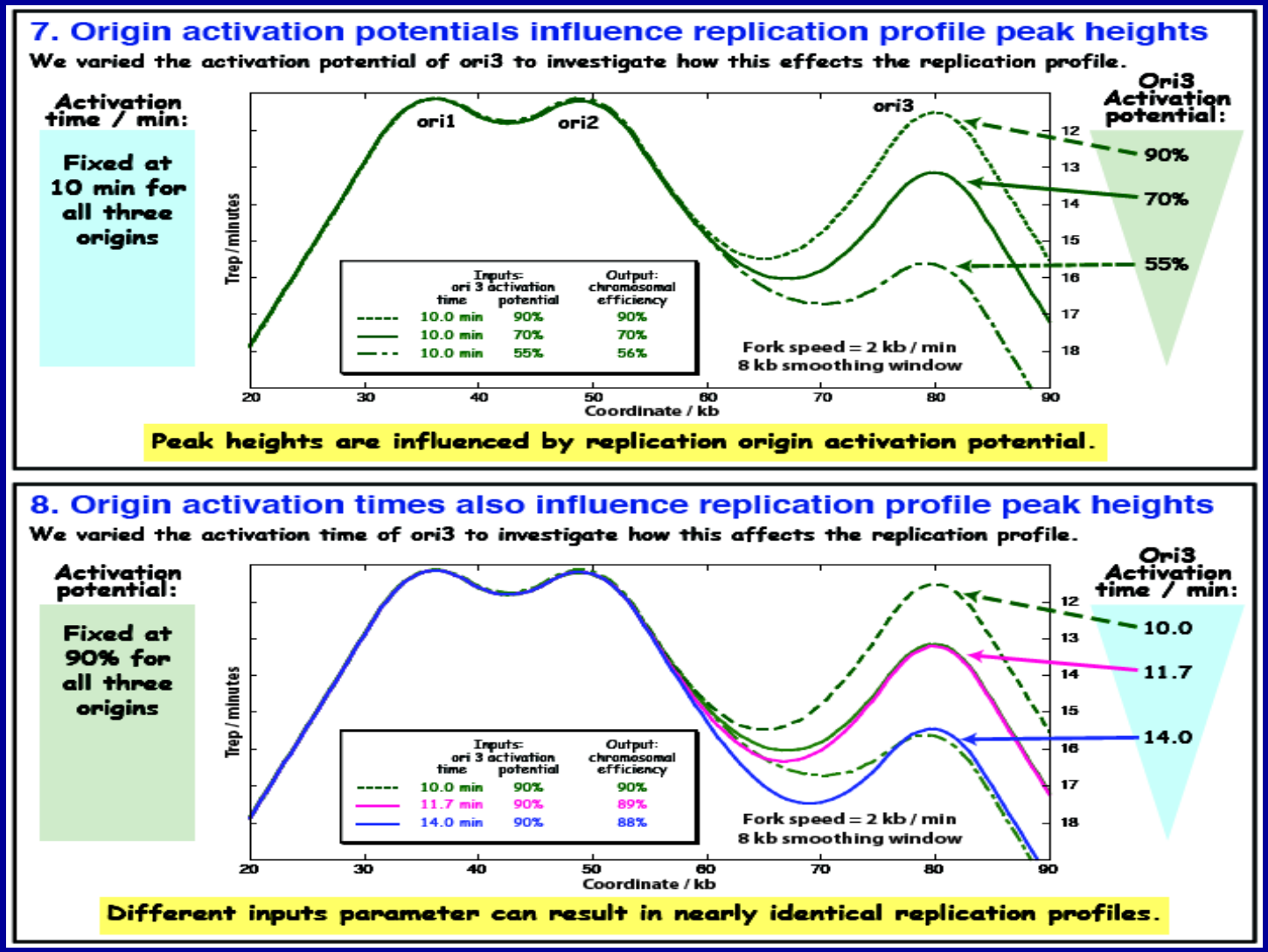
# Non-uniqueness in the determination of activation times



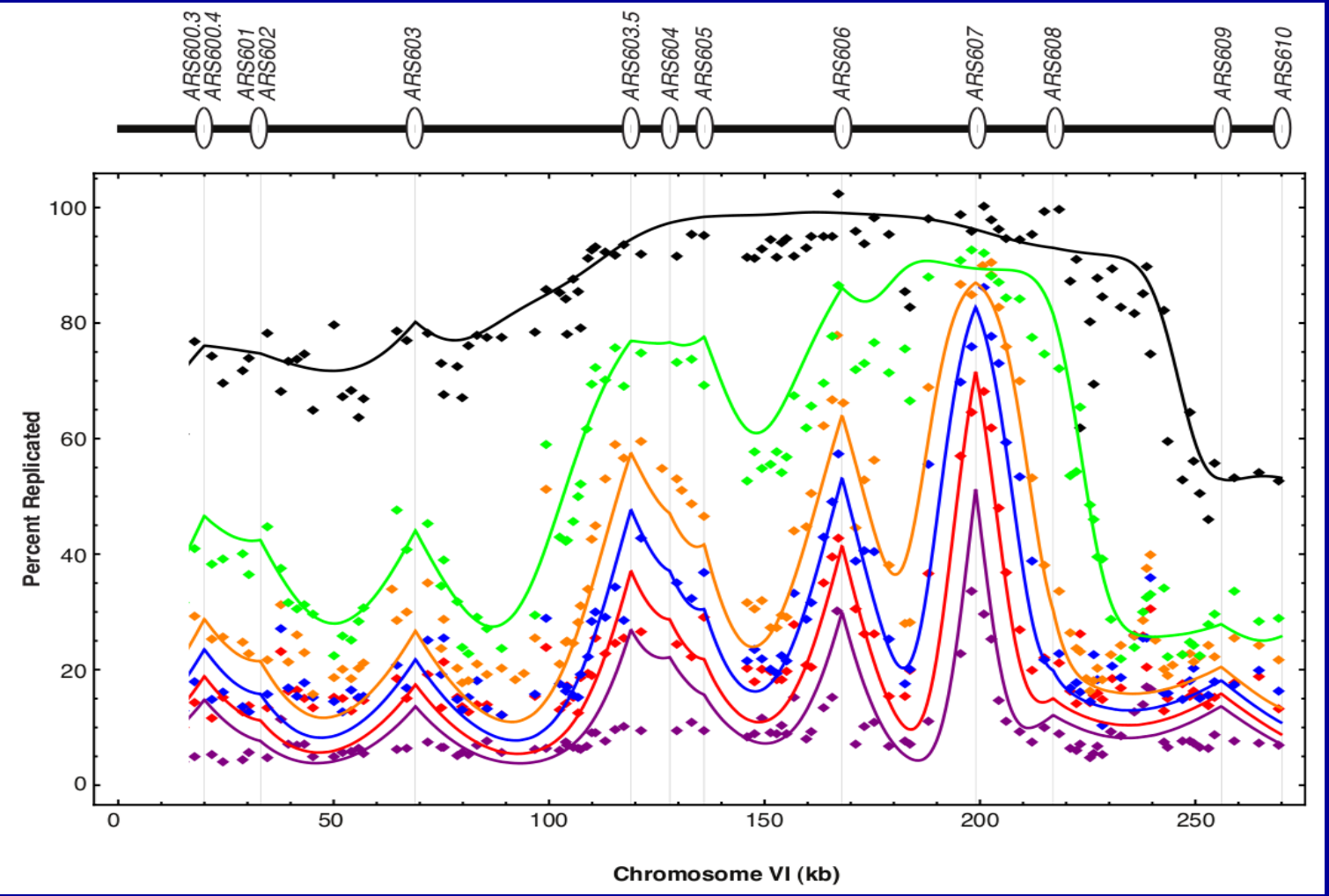
# Proposed experimental test of the model: why “dormant origins” matter



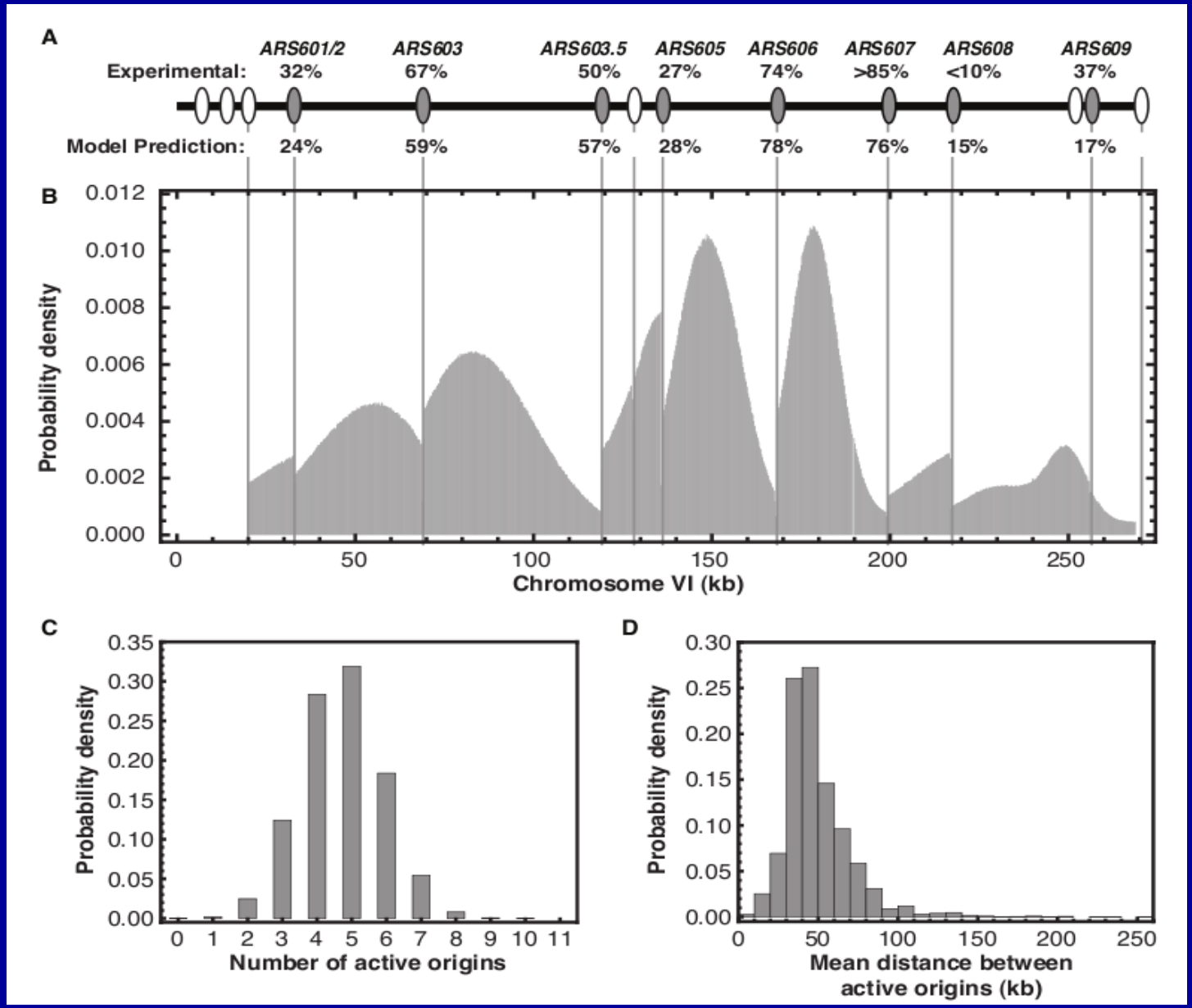
# Non-uniqueness in activation potential / activ. time determination



# Replicated fractions



# Some predictions of the model



# References


- “Mathematical modelling of whole chromosome replication”, A. de Moura, R. Retkute, M. Hawkins, C. Nieduszynski, *Nucleic Acids Research* **38,** 5623 (2010).
- “Dynamics of DNA replication in yeast”, R. Retkute, C. Nieduszynski, A. de Moura, *Physical Review Letters* **107,** 068103 (2011).
- “Optimal placement of origins for DNA replication”, J. Karschau, J. Blow, A. de Moura, *Physical Review Letters* **108,** 058101 (2012).