D Enzymatic computation

This lecture is based primarily on:
(1) Yaakov Benenson, Tamar Paz-Elizur, Rivka Adar, Ehud Keinan, Zvi Livneh, and Ehud Shapiro. Programmable and autonomous computing machine made of biomolecules. *Nature*, 414:430–434, 2001.
(2) Yaakov Benenson, Rivka Adam, Tamar Paz-Livneh, and Ehud Shapiro. DNA molecule provides a computing machine with both data and fuel. *Proceedings of the National Academies of Science*, 100(5):2191–2196, 2003.

¶1. The molecular computation processes that we have seen so far are externally controlled by a person or conventional automatic controller sequencing the chemical operations.

¶2. **Autonomous molecular computation:** In *autonomous molecular computation* the chemical processes seqence themselve so that they do not require external control.
This is also called "one-pot" molecular computation.
Autonomous molecular computation is essential for, example, controlling drug delivery in the body.

¶3. Shapiro and his colleagues have demonstrated how to implement FSMs by autonomous molecular computation.

¶4. In addition to DNA, it uses a restriction enzyme, ligase, and ATP (for fuel).

¶5. **fokI:** The implementation is based on the *fok*I restriction enzyme.

¶6. "Once the protein is bound to duplex DNA via its DNA-binding domain at the 5′-GGATG-3′ : 5′-CATCC-3′ recognition site, the DNA cleavage domain is activated and cleaves, without further sequence specificity, the first strand 9 nucleotides downstream and the second strand 13 nucleotides upstream of the nearest nucleotide of the recognition site."[8]

[8] wikipedia, s.v. fokI.

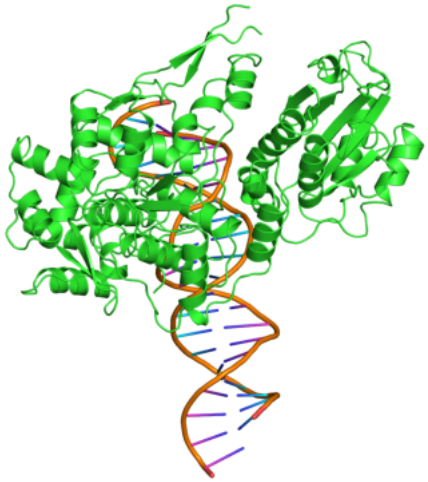

Figure IV.10: The *fok*I restriction enzyme bound to DNA. [source: wikipedia]

¶7. It leaves 4-nucleotide sticky ends.

```
GGATGNNNNNNNNN NNNNNNNNNN
CCTACNNNNNNNNNNNNN NNNNNN
```

The 'N's can be any nucleotides (respecting Watson-Crick complementarity, of course).

¶8. Both the current state of the FSM and the input string are represented by a double DNA strand.

¶9. **Recognition:** *fok*I operates at the beginning of this string and leaves a sticky end that encodes both the current state and the next input symbol.
See ¶20, p. 244 below.

¶10. **Transition molecules:** The state transitions of the FSM are encoded in *transition molecules*, which have sticky ends complementary to the state-symbol code at the beginning of the string.
The rest of a transition molecule ensures that the string properly encodes the new state, including adding a new recognition site.

¶11. **Transition:** A matching transition molecule binds to the string's sticky end, providing a new opportunity for *fok*I to operate, and so the process continues.

¶12. A state transition $(q, s_1) \to q'$ can be visualized:

$$[q, s_1]s_2s_3 \cdots s_n t \Longrightarrow [q', s_2]s_3 \cdots s_n t$$

where $[q, s]$ represents a DNA sequence encoding both state q and symbol s, and t is a *terminator* for the string.

¶13. The *fok*I enzyme cleaves off $[q, s_1]$ in such a way that a transition molecule can bind to the sticky end in a way that encodes $[q', s_2]$.

¶14. **Terminator:** A special *terminator* symbol marks the end of the input string.

¶15. **Example:** As an example we will consider a two-state FSM on {a, b} that accepts strings with an even number of 'b's.

¶16. **State-symbol encoding:** Ignoring the terminator, DNA codes are assigned to the two symbols 'a' and 'b' as follows:

$$\begin{aligned} \text{a} &\mapsto AA\alpha\alpha aa \\ \text{b} &\mapsto BB\beta\beta bb \end{aligned}$$

where $A, \alpha, a, B, \beta, b$ are unspecified (by me) bases.[9]
The bases are selected in such a way that either the first four bases ($AA\alpha\alpha$, $BB\beta\beta$) or the last four bases ($\alpha\alpha aa$, $\beta\beta bb$) encode the symbol.

¶17. The transition molecules are constructed so that the distance between the recognition site (for *fok*I) and the next symbol depends on new state.
As a consequence, when *fok*I operates it cleaves the next symbol code at a place that depends on the state.

[9]Note that repeated letters might refer to different bases.

Therefore the sticky end encodes the state in the way that it represents the next symbol:

$$\begin{aligned} [q_0, \mathrm{a}] &\mapsto \alpha\alpha aa \\ [q_1, \mathrm{a}] &\mapsto AA\alpha\alpha \\ [q_0, \mathrm{b}] &\mapsto \beta\beta bb \\ [q_1, \mathrm{b}] &\mapsto BB\beta\beta \end{aligned}$$

¶18. **Transition molecules:** The transition molecules are:

$$\begin{array}{ll} (q_0, \mathrm{a}) \to q_0 & \texttt{GGATG}NNN \\ & \texttt{CCTAC}\overline{NNN\alpha\alpha aa} \\ (q_1, \mathrm{a}) \to q_1 & \texttt{GGATG}NNN \\ & \texttt{CCTAC}\overline{NNNAA\alpha\alpha} \\ (q_0, \mathrm{b}) \to q_1 & \texttt{GGATG}NNNNN \\ & \texttt{CCTAC}\overline{NNNNN\beta\beta bb} \\ (q_1, \mathrm{b}) \to q_0 & \texttt{GGATG}N \\ & \texttt{CCTAC}\overline{NBB\beta\beta} \end{array}$$

N represents any base.

¶19. **Spacers:** The Ns are used as *spacers* to adjust the restriction site to represent the new state.

¶20. After transition to the new state the sense strand will look like this (for convenience assuming the next symbol is 'a'):

$$\begin{array}{ll} q_0 & \mathrm{GGATG}NXXYYyyAA.\alpha\alpha aa \\ q_1 & \mathrm{GGATG}NNNXXYYyy.AA\alpha\alpha aa \end{array}$$

Here XX represents either spacers or the first two bases of the previous first symbol, and $YYyy$ represents the last four bases of this symbol. The cleavage site is indicated by a period.

¶21. **Problem size:** The longest strings processed in the PNAS experiments were 12.

¶22. **Speed:**[10] About 20s per step.

[10] Benenson et al., PNAS **100** (5), March 4, 2003.

¶23. **Parallel speed:** About 6.6×10^{10} ops/s/μl.

¶24. **Energy:** About $34kT$ per transition.

¶25. **Operating temperature:** "Reaction rates were surprisingly insensitive to temperature and remained similar over the range of 2–20°C."

¶26. **Nondeterministic FSM:** This implementation also handles ND FSMs (just put in all the transition molecules), but the yield decreases exponentially (due to following out all the ND paths, breadth-first). Therefore it doesn't seem to be practical.