BIOINFORMATICS

Vol. 00 no. 00 2012
Pages 1

Supplementary Material to: GRiP: a computational tool to simulate transcription factor binding in prokaryotes

Nicolae Radu Zabet and Boris Adryan

1 SIMULATION SPEED BENCHMARK

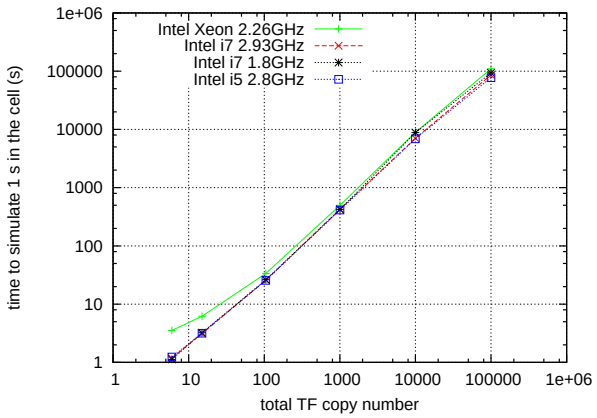

Fig. 1. *Time to simulate* 1 *second in a* E.coli *K-12 cell.* We consider *E.coli* K-12 genome and parameters of the system presented in (Zabet and Adryan, 2012). For each machine, we varied the non-cognate copy number. The simulations are run on four machines: (i) Mac Pro 2x2.26GHz quad-core Intel Xeon with 32GB memory running Mac OS X 10.6.8, (ii) PC Desktop computer i7 940 at 2.93GHz with 6GB memory running Ubuntu 10.4, (iii) Mac Book Air i7 1.8GHz with 4GB memory running Mac OS X 10.7.2 and (iv) iMAC i5 2.8GHz with 4GB memory running Mac OS X 10.6.8

Figure 1 shows how the CPU time to simulate 1 s scales with the number of molecules in the system and how this time varies on several machines. The increase in CPU time is significant. For example, an increase in TF copy number from 10^4 to 10^5 seems to generate an increase of ≈ 10 times in CPU time.

Figure 2 shows how the CPU time required to simulate 1 s scales with the size of the DNA. In particular, once can observe that the increase in CPU time is relatively small compared to the increase in genome size. For example, in our example, an increase of 10 times in DNA size results in a increase of 1.3 times in CPU time. This is due to our strategy of locating free target sites described above.

2 CONSIDERATIONS ON MEMORY

The main disadvantage of our approach is the high demand of memory. For *E.coli* K-12 and two TF species (a cognate and a non-cognate one) the application requires approximately 1.6 GB

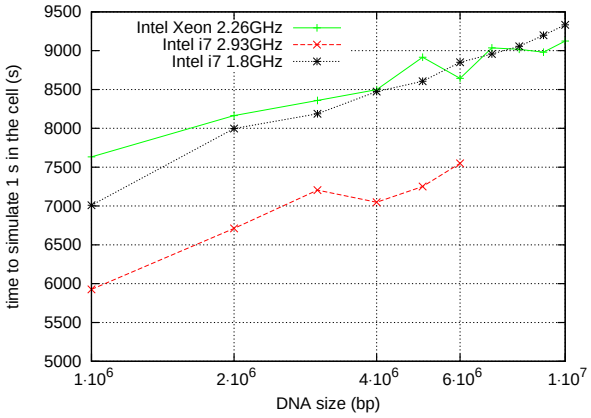

Fig. 2. *Time to simulate* 1 *second for different genomes* We consider the parameters of the system presented in (Zabet and Adryan, 2012), 10000 non-cognate TFs and 5 lacI molecules. For each machine, we varied the size of the randomly generated DNA sequences. The simulations are run on three machines: (i) Mac Pro 2x2.26GHz quad-core Intel Xeon with 32GB memory running Mac OS X 10.6.8, (ii) PC Desktop computer i7 940 at 2.93GHz with 6GB memory running Ubuntu 10.4 and (iii) Mac Book Air i7 1.8GHz with 4GB memory running Mac OS X 10.7.2

of free memory. This is a consequence of the fact that, in order to increase speed, GRiP computes a series of values *a priori* and then stores them in several arrays. In addition, our implementation strategy requires a boolean array for each TF species which indicates whether a molecule of that type can bind at that specific position. This implementation is not optimal for memory, but gives better speed results compared to other implementations. This suggests that, independent of the programming language, the application will use significant amount of memory. We also compared a 32 bit version of the application with a 64 bit one and we observed that the application used similar amount of memory on both 32 bit and 64 bit versions.

REFERENCES

Zabet, N. R. and Adryan, B. (2012). A comprehensive computational model of facilitated diffusion in prokaryotes. *Bioinformatics*.