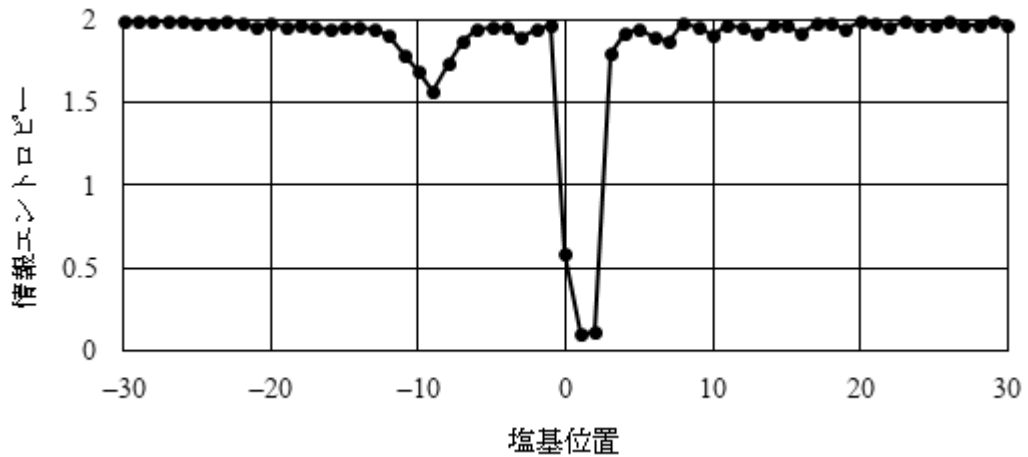


例題 3.9

以下の図は、大腸菌 K-12 株のゲノム (GenBank NC_000913) に登録されている 4,295 個の遺伝子に対して、開始コドン周辺 (-30~+30, 開始コドンの第 1 塩基の位置を 0 とする) の塩基配列の MSA を求め、情報エントロピーの値をプロットしたものである。これについて、1~3 の問いに答えよ。



1. 開始コドンは一般に ATG として知られているが、実際のパターンについて、この図からどのようなことがわかるか。
2. 開始コドンの上流に、情報エントロピーの値が小さくなっている部分が見られる。この部分はどのような領域に相当すると考えられるか。
3. 情報エントロピーの値がだいたい 2 で一定している領域の GC 含量はどれくらいか。

解答

1. すべての開始コドンが ATG であるなら、これら 3 塩基の各位置のエントロピーは 0 になるはずである。実際には、第 1 塩基にかなりのばらつきが見られ、第 2 塩基と第 3 塩基には、第 1 塩基に比べて小さなばらつきが見られることがわかる。
2. シャインダルガノ (SD) 配列に相当する。SD 配列は、-AGGAGG-のようにプリン塩基に富んだ 3~9 塩基の長さの配列であり、特徴的なパターンをもつものの、ばらつきがあり、エントロピーの値も高くなっている。また、SD 配列でそろえたアラインメントでないため、開始コドンと SD 配列の距離のばらつきによって、エントロピーの値もばらつきが見られる。
3. GC 含量はほぼ 0.5