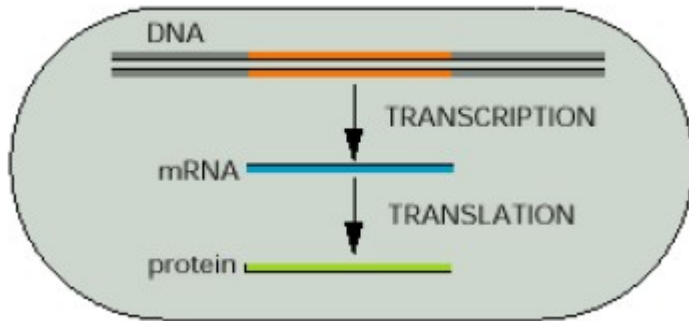


# HMM based gene finding in Prokaryots

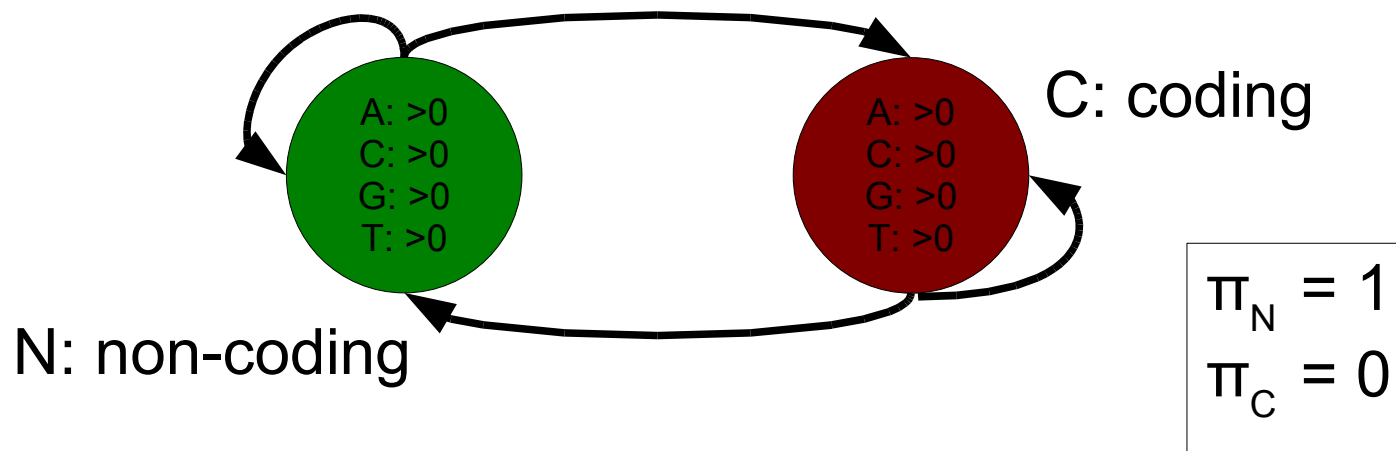


## Biological facts

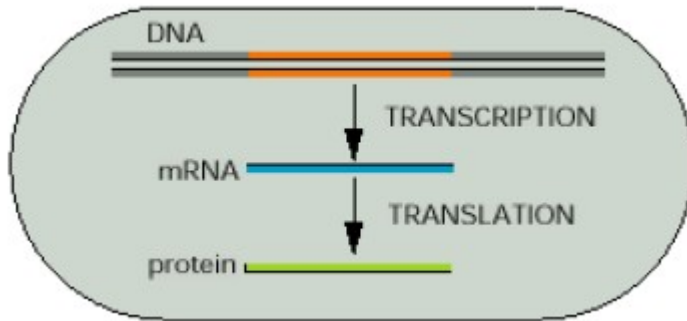
- The gene is a substring of the DNA sequence of A,C,G,T's

**Z:** NNNCCCCCCCCNNNNNNNNCCCCCCCCCCCCCCCCNNNNNNNNNNNN

**X:** acgatgcgctaataatgtccgatgacgtgagcataagcgacatgcag



# HMM based gene finding in Prokaryots

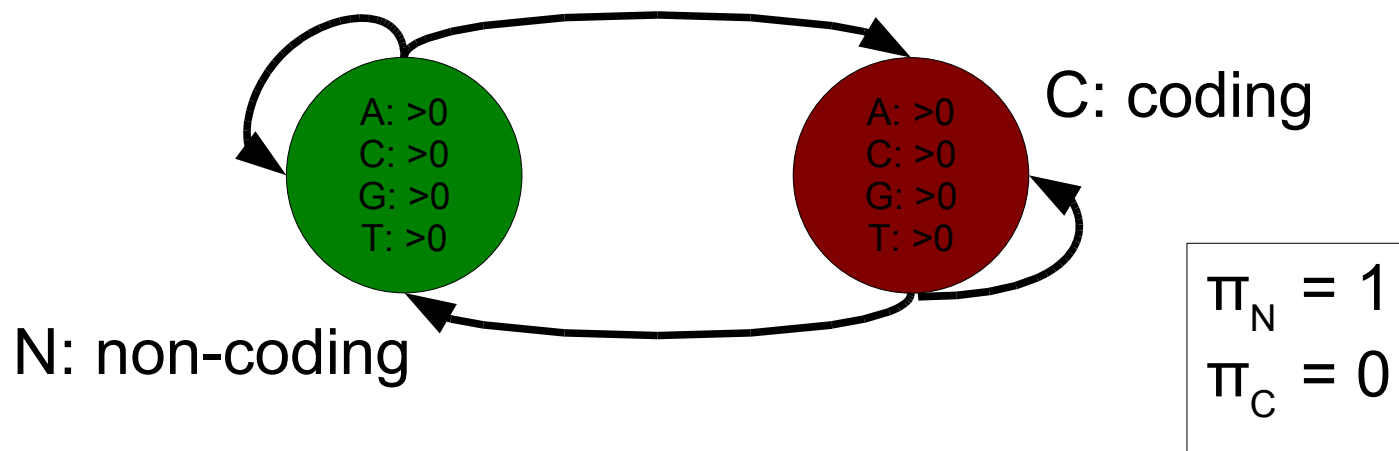


## Biological facts

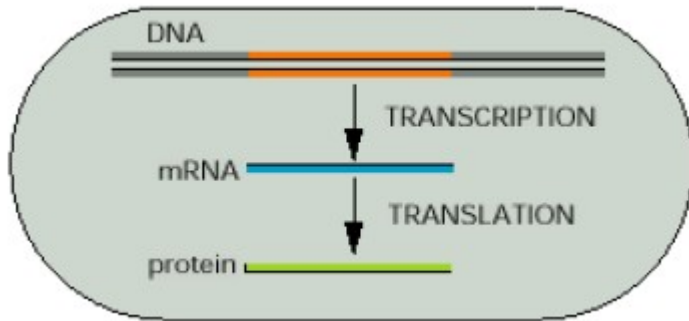
- The gene is a substring of the DNA sequence of A,C,G,T's
- The gene starts with a start-code **atg**

**Z:** NNNCCCCCCCCNNNNNNNNCCCCCCCCCCCCCCCCNNNNNNNNNNNN

**X:** acgatgcgctaataatgtccgatgacgtgagcataagcgacatgcag



# HMM based gene finding in Prokaryots



## Biological facts

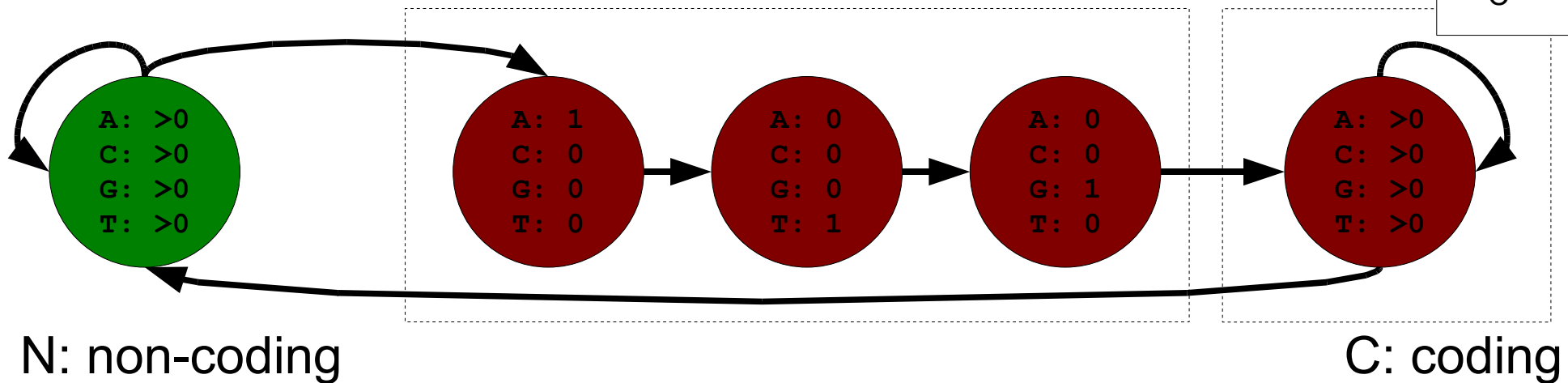
- The gene is a substring of the DNA sequence of A,C,G,T's
- The gene starts with a start-code **atg**

**Z:** NNNCCCCCCCCNNNNNNNNCCCCCCCCCCCCNNNNNNNNNN

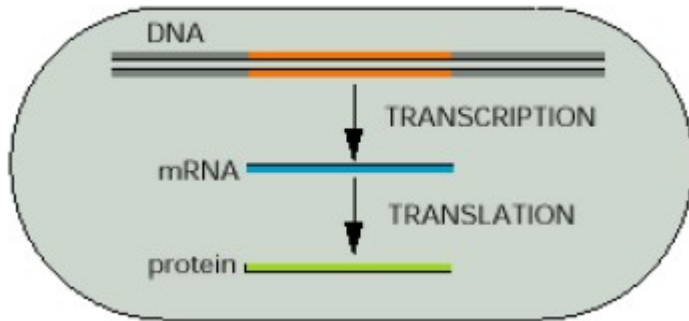
**X:** acgatgcgctaataatgtccgatgacgtgagcataagcgacat

$$\pi_N = 1$$

$$\pi_C = 0$$



# HMM based gene finding in Prokaryots



## Biological facts

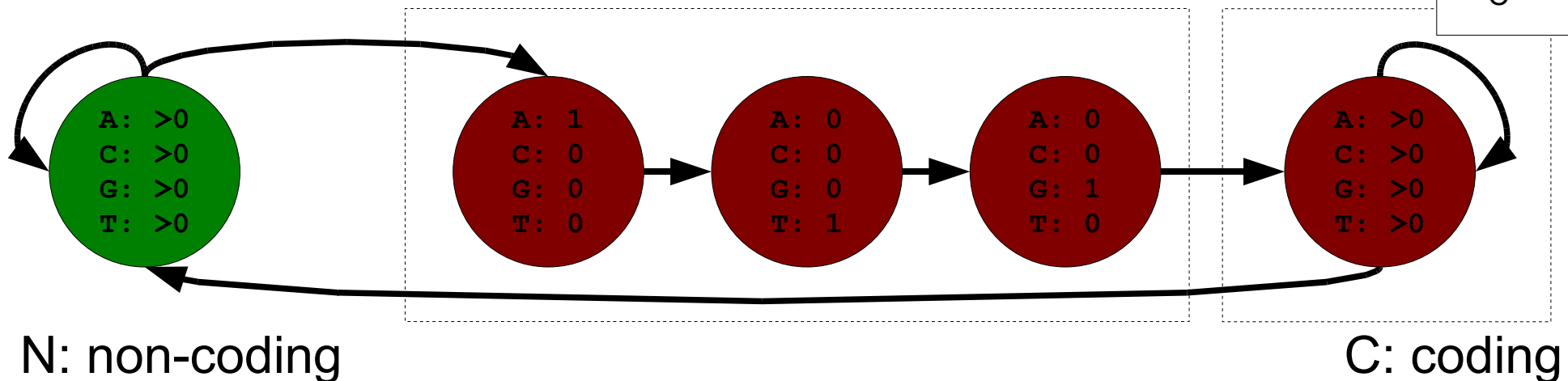
- The gene is a substring of the DNA sequence of A,C,G,T's
- The gene starts with a start-code **atg**
- The gene ends with a stop-codon **taa, tag or tga**

**Z:** NNNCCCCCCCCNNNNNNNNCCCCCCCCCCCCNNNNNNNNNN

**X:** acgatgcgctaataatgtccgatgacgtgagcataagcgacatc

$$\pi_N = 1$$

$$\pi_C = 0$$

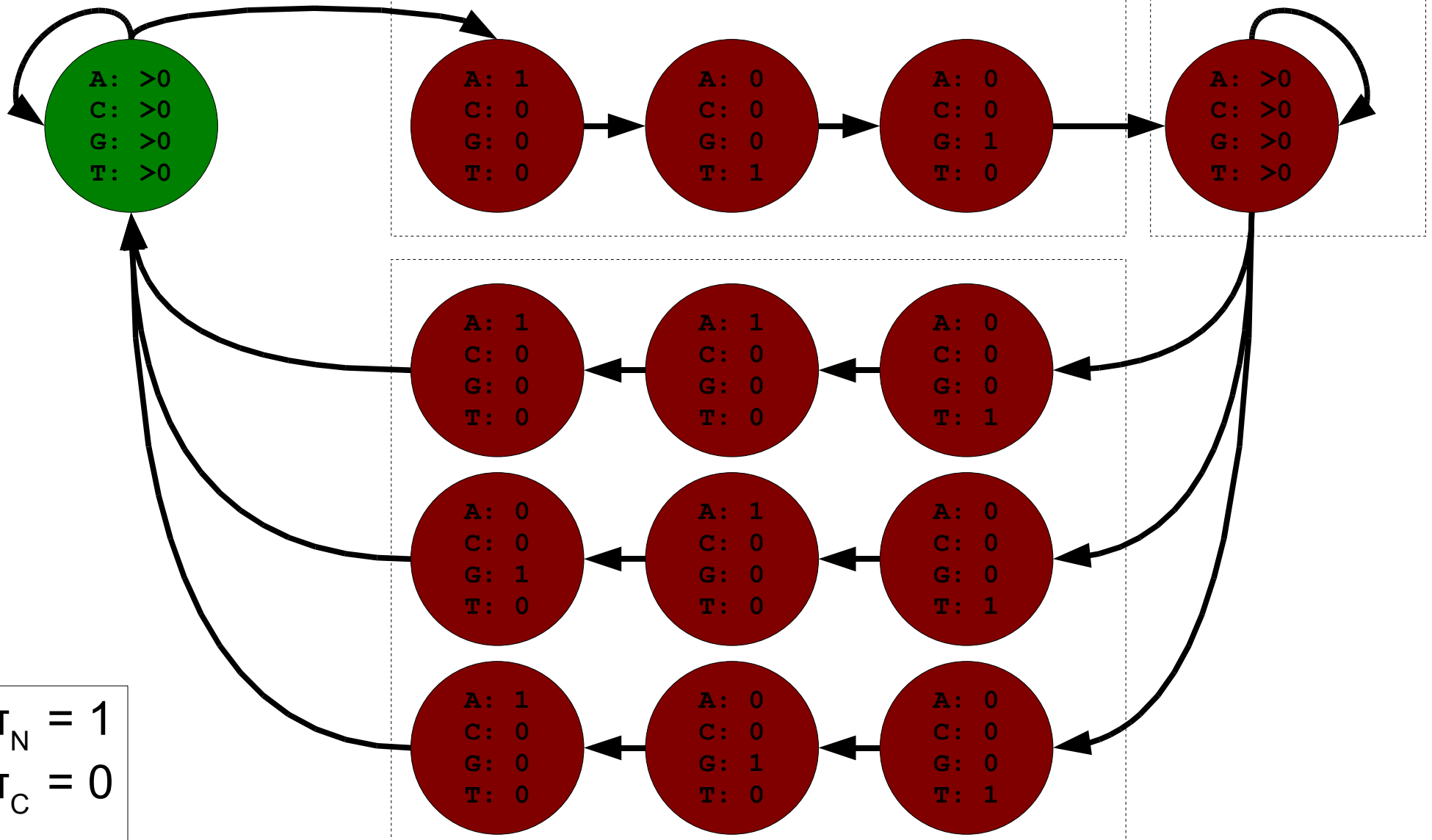


# HMM based gene

- The gene is a substring of the DNA sequence of A,C,G,T's
- The gene starts with a start-code **atg**
- The gene ends with a stop-codon **taa, tag or tga**

N: non-coding

C: coding

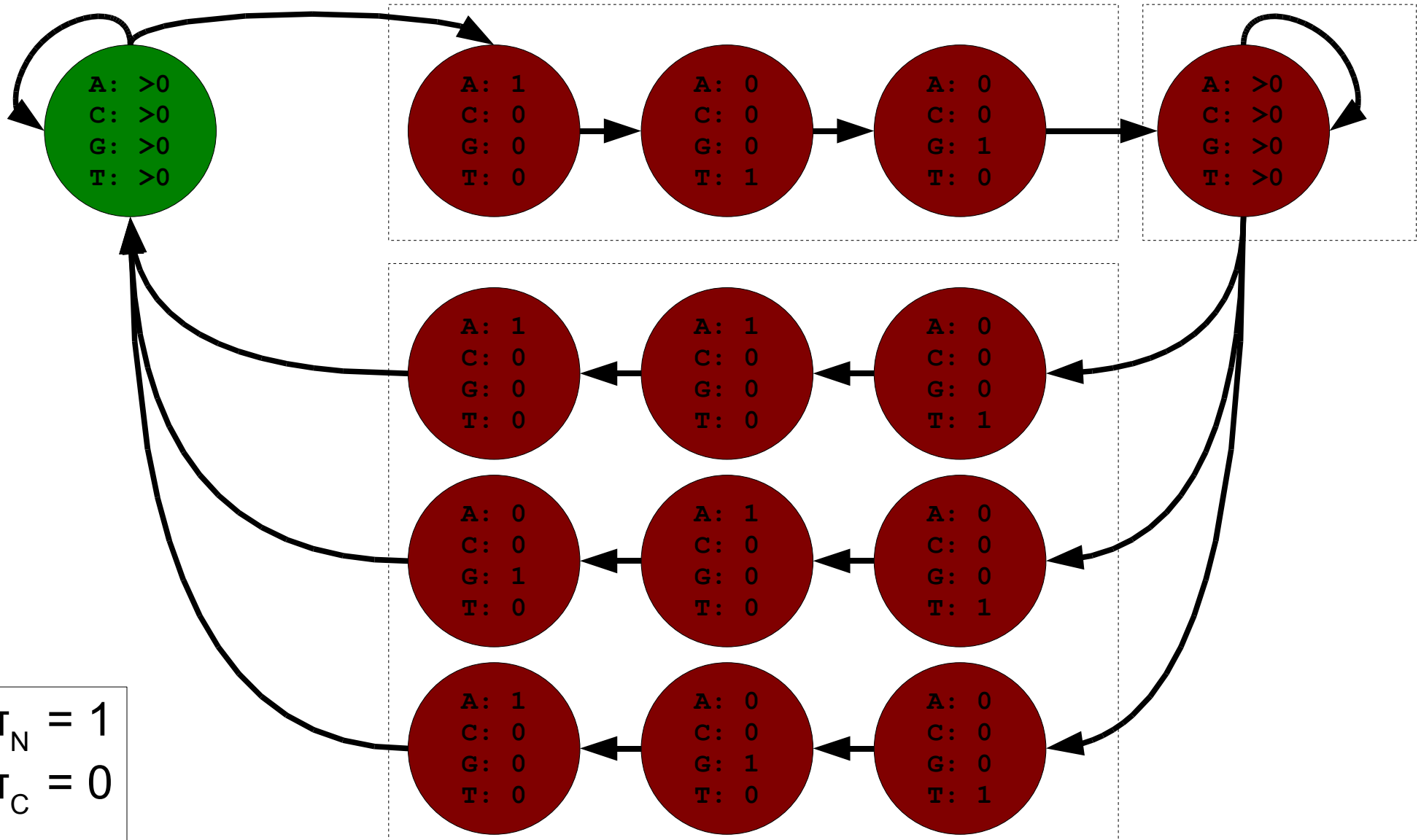


# HMM based gene

- The gene is a substring of the DNA sequence of A,C,G,T's
- The gene starts with a start-code **atg**
- The gene ends with a stop-codon **taa, tag or tga**
- The number of nucleotides in a gene is a multiplum of 3

N: non-coding

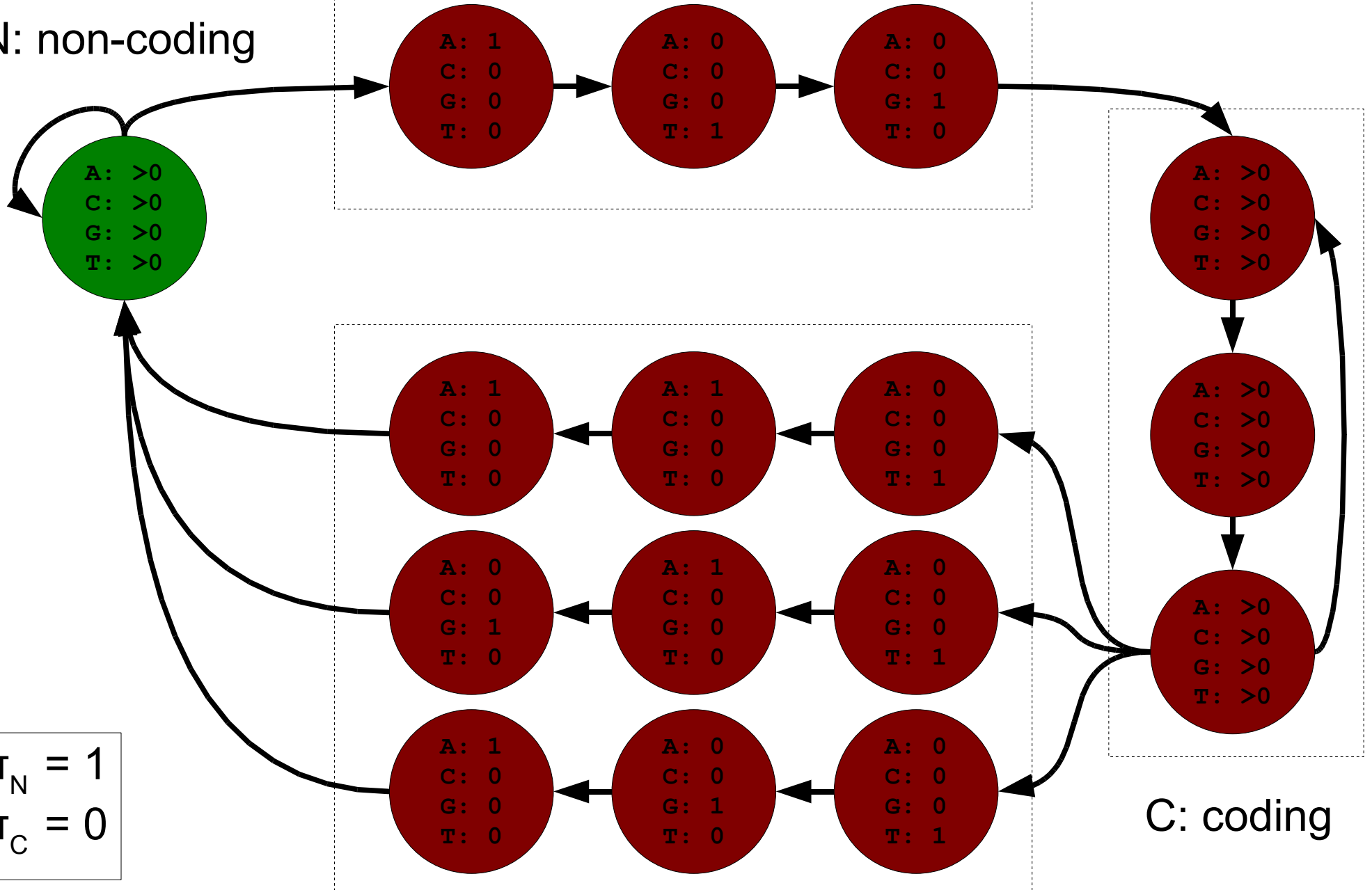
C: coding



# HMM based gene

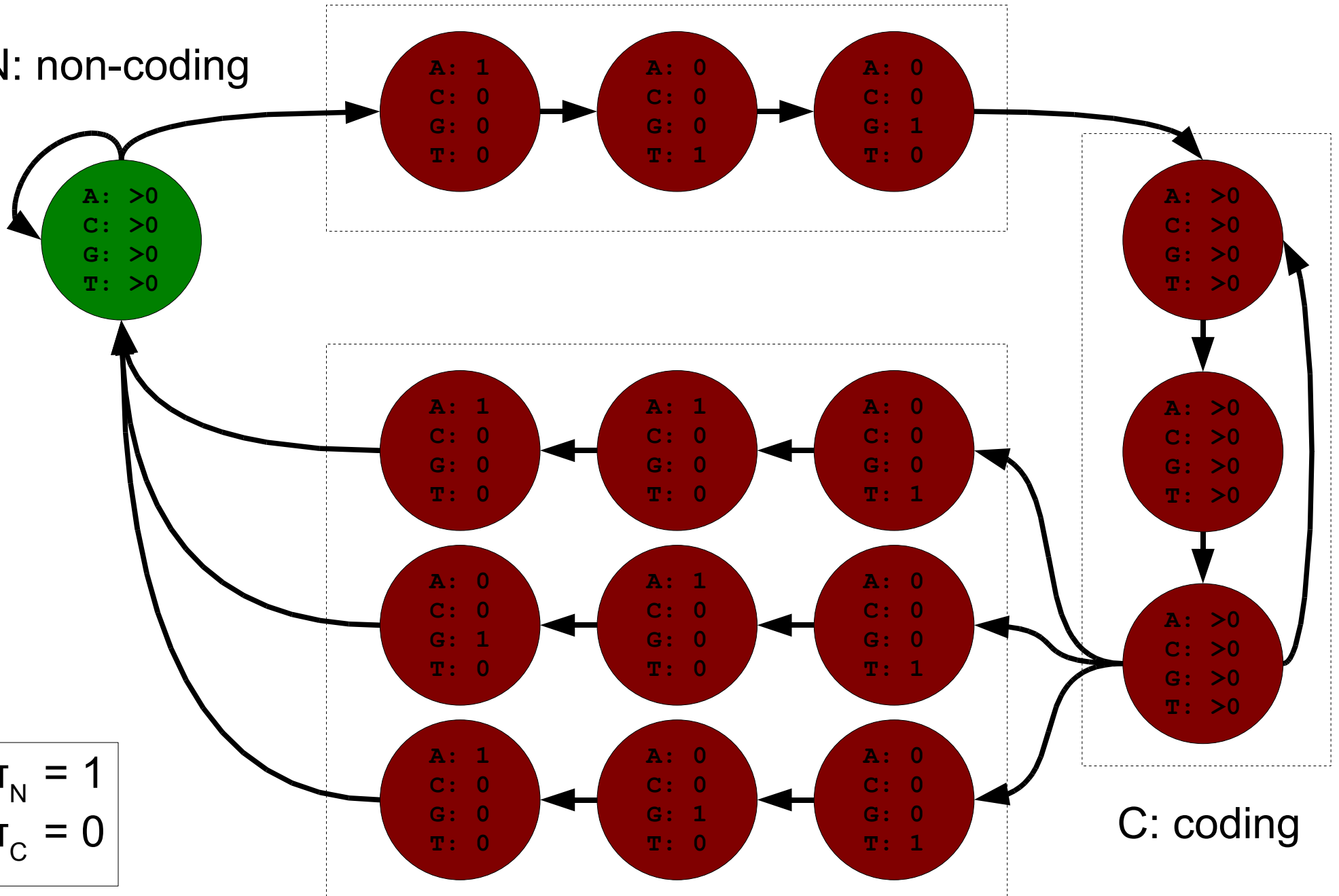
- The gene is a substring of the DNA sequence of A,C,G,T's
- The gene starts with a start-code **atg**
- The gene ends with a stop-codon **taa, tag or tga**
- The number of nucleotides in a gene is a multiplum of 3

N: non-coding



# HMM based gene finding in Prokaryots

N: non-coding



C: coding

# HMM based gene finding in Prokaryotes

N: non-coding

