Rate matrix for an MG-style codon model

$$(\text{Rate})_{X,Y}(dt) = \begin{cases} \alpha & \pi_t dt & , & \text{one-step, synonymous substitution,} \\ \beta & \pi_t dt & , & \text{one-step, non-synonymous substitution,} \\ 0 & & , & \text{multi-step.} \end{cases}$$

X,Y = AAA...TTT (excluding stop codons),
π_t - frequency of the target nucleotide.

Example substitutions:
AAC→AAT (one step, synonymous - Asparagine)
CAC→GAC (one step, non-synonymous - Histidine to Aspartic Acid)
AAC→GTC (multi-step).

αR_{CT}
βR_{CG}

α (syn. rate) and **β** (non-syn. rate) are the key quantities for all selection analyses

Computing the transition probabilities

- In order to recover transition probabilities $\mathbf{T(t)}$ from the rate matrix $\mathbf{Q}$, one computes the matrix exponential $\mathbf{T(t) = exp\ (Qt)}$, same as with standard nucleotide models, e.g. HKY85 or GTR.

- Because the computational complexity of matrix exponentiation scales as the cube of the matrix dimension, codon based models require roughly $\mathbf{(61/4)^3 \approx 3500}$ more operations than nucleotide models.

- This explains why codon probabilistic models were not introduced until the 1990s, even though they are relatively straightforward extensions of 4x4 nucleotide models