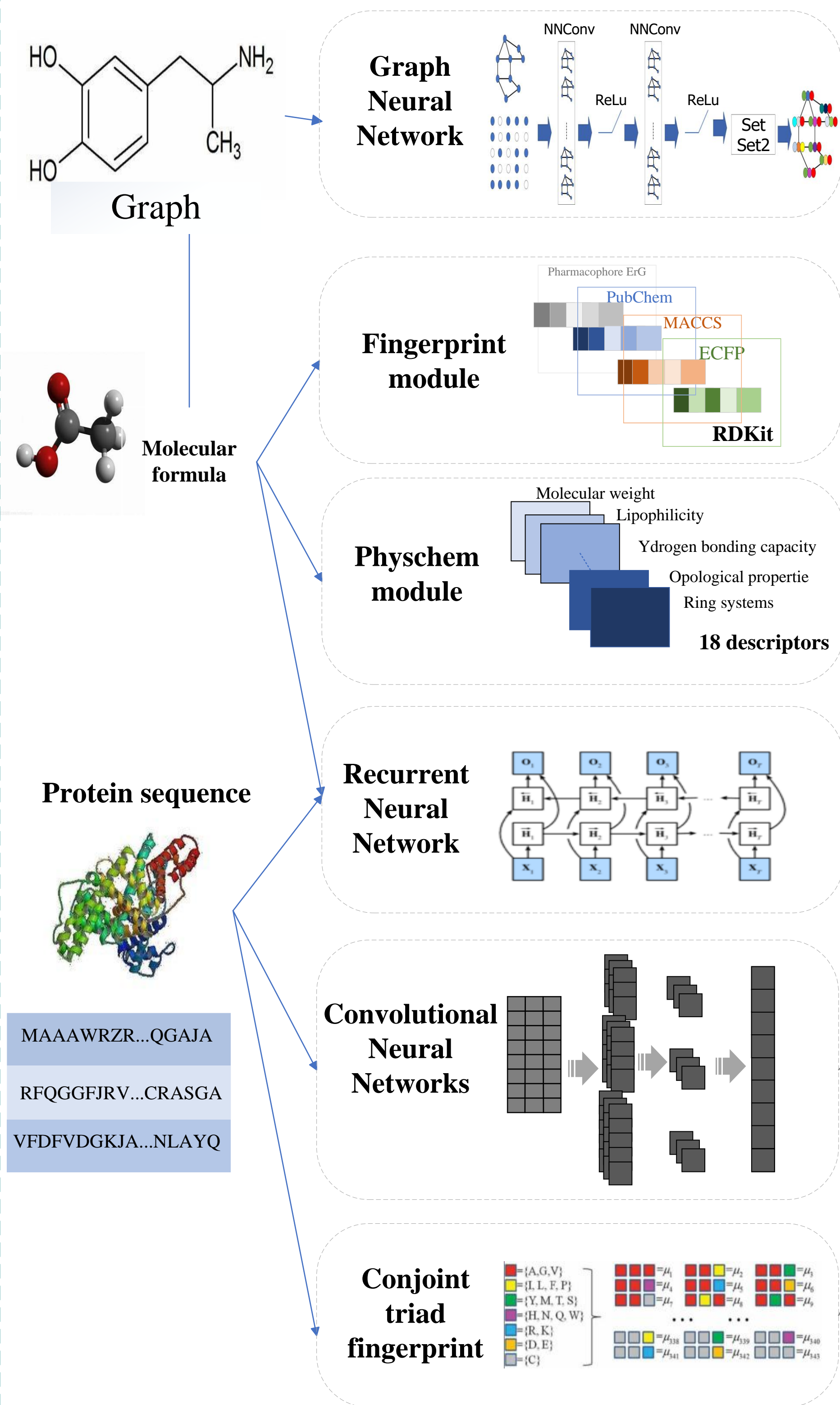


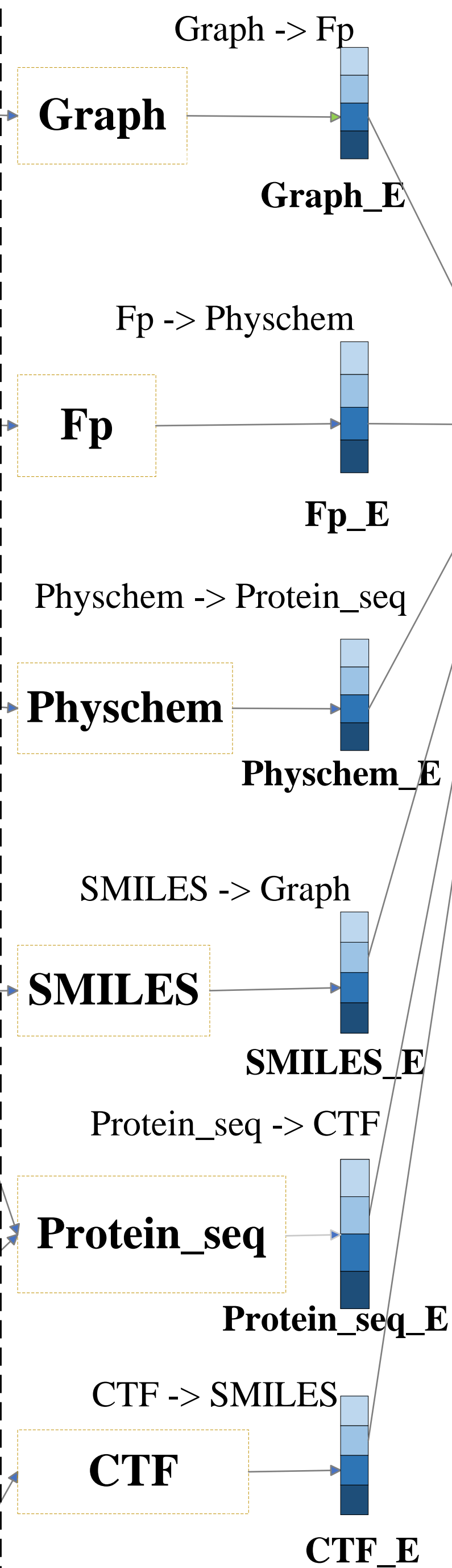
A

Feature extration



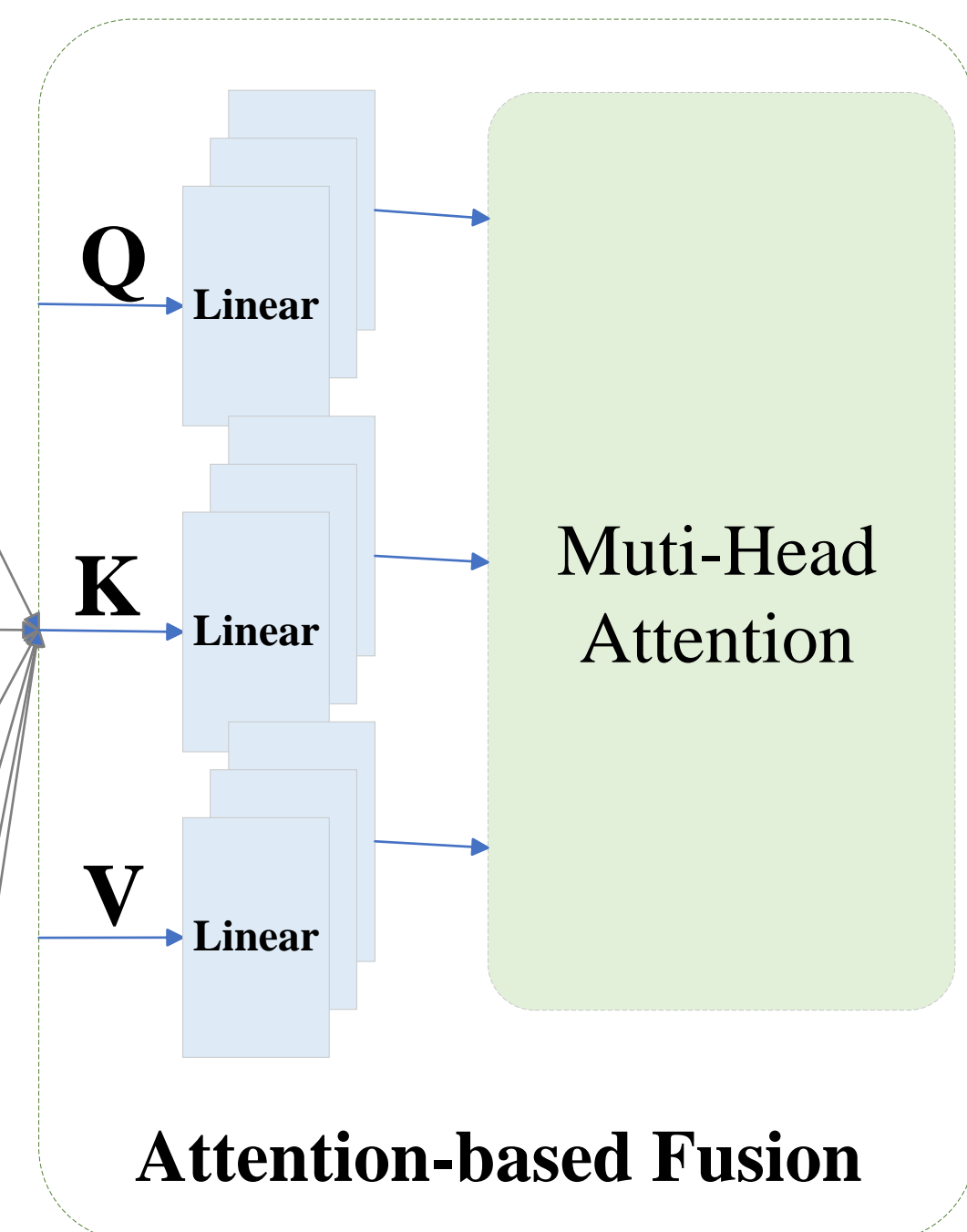
B

Feature enhanced

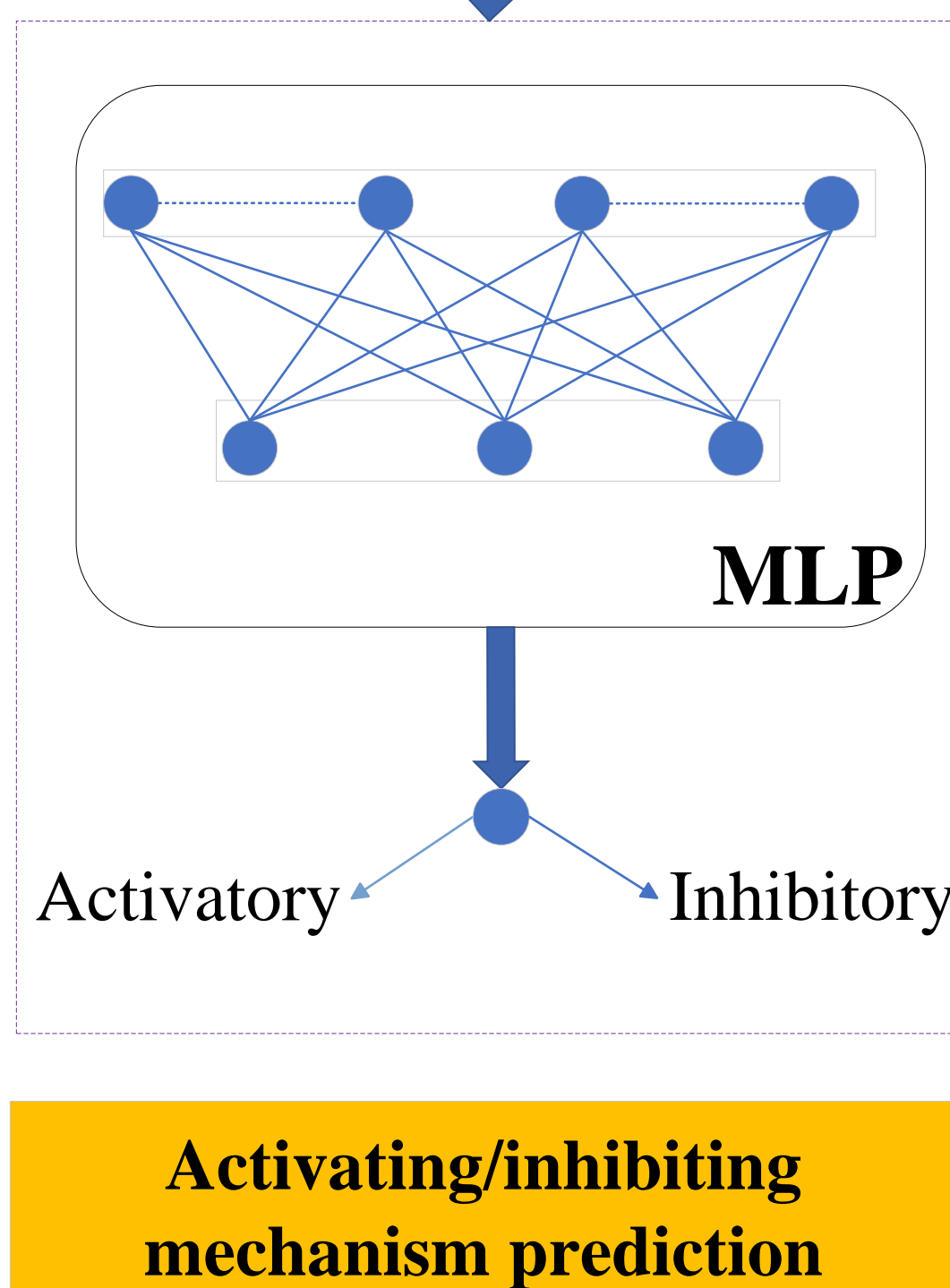


C

Fusion



D



CrossAttention: A -> B
Q = A, K = B, V = B