Q1. $\boxed{\text{Yes.}}$ Proof:

$$\pi^*(s) = \underset{a}{\operatorname{argmax}}\ Q^*_{M}(s,a) = \underset{a}{\operatorname{argmax}}\ Q^*_{M}(s,a) + f(s) = \underset{a}{\operatorname{argmax}}\ Q^*_{M'}(s,a)$$

by definition

const. w.r.t. a

$\pi^{*\prime} = \underset{a}{\operatorname{argmax}}\ Q^*_{M'}(s,a)$ is defined as the optimal policy for $Q^*_{M'}(s,a)$, so we proved that $\pi^*(s)$ is the optimal policy for $Q^*_{M'}(s,a)$ also!

Q3. 1. Each iteration we run π_{t}, the opt. policy for iter. t, until a new state-action is found. Let's assume by contradiction that exists a pair (s',a') such that $R'_t(s',a') = 1$ we didn't discover at iteration t. In other words, the optimal policy π_t stuck on (s,a) : $R'_t(s,a) = 0$, so $V^{\pi}_t = 0$.

Given a strongly connected DDP, we know (s',a') can be visited. So a policy π' that does so will get a reward of 1 at (s',a'), thus π is not optimal, in contradiction!

Each iteration must end!

2. $\boxed{T_{iteration} = O(|S|)}$. For example, in the case that we should go over all the states that we have already visited in order to go from S_0 to (s', a')

3. every state s in the set S has a set of possible action $A(s)$, So in order to go over the entire S, a space we need $\sum_{s \in S} |A(s)| = \boxed{|S| \cdot |A| \text{ iterations}}$

4. $T_{exploitation} = T_{iteration} \cdot \text{num\_iterations} = O(|S|) \cdot |S| \cdot |A| =$

$= \boxed{O(|S|^2 \cdot |A|)}$