Join minimization examples

Let $R$ be a relation over attributes $ABC$.
(i) Simplify the following conjunctive SQL query, knowing that it is applied only to relations $R$ satisfying the set of FDs $F = \{ AC \rightarrow B, B \rightarrow C, C \rightarrow A \}$ (use pattern minimization and the chase):

\[
\text{select } t_1.A, t_2.B, t_4.C \\
\text{from } R t_1, R t_2, R t_3, R t_4 \\
\text{where } t_3.A = t_4.A \text{ and } t_2.B = t_3.B \text{ and } t_1.C = t_2.C
\]

(ii) Redo (i) for the query:

\[
\text{select } t_1.A, t_2.B, t_4.C \\
\text{from } R t_1, R t_2, R t_3, R t_4 \\
\text{where } t_2.C = 5 \text{ and } t_3.A = t_4.A \text{ and } t_2.B = t_3.B \text{ and } t_1.C = t_2.C \text{ and } t_4.A = 8
\]

Solution
(i) The pattern $P$ corresponding to $Q$ is

\[
\begin{array}{c|ccc}
R & A & B & C \\
\hline
\begin{array}{c}
a \\
b \\
a_1 \\
a_1
\end{array} & \begin{array}{c}
c_1 \\
c_1 \\
- \\
-
\end{array} & \begin{array}{c}
- \\
- \\
- \\
-
\end{array} \\
\end{array}
\]

We now chase the pattern with $F = \{ AC \rightarrow B, B \rightarrow C, C \rightarrow A \}$. The steps are as follows:

1. we first consider $AC \rightarrow B$. However, there is no violation of this FD at this point.

2. we consider $B \rightarrow C$, which is violated by rows (2) and (3). Chasing leads us to identify the $-$ in (3) with $c_1$, yielding:
3. we consider $C \rightarrow A$, which is violated by rows (1),(2),(3). Chasing leads us to identify $a$, the $-$ in (2), and $a_1$, which all become equal to $a$. Note that this includes the $a_1$ in row (4), which is the same $a_1$ as in row (3). This yields:

\[
\begin{array}{|c|c|c|c|} 
\hline
R & A & B & C \\
\hline
a & b & c_1 \\
- & b & c_1 \\
a_1 & b & c_1 \\
a & c_1 \\
\hline
\end{array}
\]

4. we are not yet done, because now $AC \rightarrow B$ is violated by rows (1) and (2). Chasing turns the $-$ in (1) into $b$, yielding:

\[
\begin{array}{|c|c|c|c|} 
\hline
R & A & B & C \\
\hline
a & b & c_1 \\
a & b & c_1 \\
a & b & c_1 \\
a & c_1 \\
\hline
\end{array}
\]

The above pattern satisfies $F$, so the chase is done.

Eliminating duplicate rows from $CHASE_F(P)$ yields the following pattern, which is minimal:

\[
\begin{array}{|c|c|c|c|} 
\hline
R & A & B & C \\
\hline
a & b & c_1 \\
a & c_1 \\
\hline
\end{array}
\]

and the minimal SQL query is:
(ii) The pattern $P$ corresponding to $Q$ is

$$
\begin{array}{c|ccc}
R & A & B & C \\
\hline
a & - & 5 & \\
b & 5 & \\
8 & b & - \\
8 & - & c \\
\end{array}
$$

Chasing with respect to $F = \{AC \rightarrow B, B \rightarrow C, C \rightarrow A\}$ yields (after eliminating duplicate rows):

$$
\begin{array}{c|ccc}
CHASE_F(P) = & R & A & B & C \\
\hline
 & 8 & b & 5 \\
 & 8 & - & c \\
\end{array}
$$

This pattern is minimal, and a corresponding SQL query with minimum number of joins is

$$
\begin{align*}
\text{select } & t_1.A, t_1.B, t_2.C \\
\text{from } & R t_1, R t_2 \\
\text{where } & t_1.A = t_2.A \\
\text{and } & t_1.C = 5 \\
\text{and } & t_2.A = 8 \\
\end{align*}
$$