Join minimization examples

Problem 1 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

<table>
<thead>
<tr>
<th>$R$</th>
<th>$A$</th>
<th>$B$</th>
<th>$C$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a$</td>
<td>$c_1$</td>
<td>$-$</td>
<td></td>
</tr>
<tr>
<td>$- $</td>
<td>$b$</td>
<td>$c_1$</td>
<td></td>
</tr>
<tr>
<td>$a_1$</td>
<td>$b$</td>
<td>$-$</td>
<td></td>
</tr>
<tr>
<td>$a_1$</td>
<td>$-$</td>
<td>$c$</td>
<td></td>
</tr>
</tbody>
</table>

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 $-c_1$ 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:

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

$$\begin{array}{ccc}
answer & A & B & C \\
\hline
a & b & c \\
\end{array}$$

4. we are not yet done, because now $AC \rightarrow B$ is violated by rows (1) and (2). Chasing turns the $-c_1$ in (1) into $b$, yielding:

$$\begin{array}{ccc}
R & A & B & C \\
\hline
a & b & c_1 \\
a & b & c_1 \\
a & b & c_1 \\
a & - & c \\
\end{array}$$

$$\begin{array}{ccc}
answer & A & B & C \\
\hline
a & b & c \\
\end{array}$$

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

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

$$\begin{array}{ccc}
R & A & B & C \\
\hline
a & b & c_1 \\
a & - & c \\
\end{array}$$

$$\begin{array}{ccc}
answer & A & B & C \\
\hline
a & b & c \\
\end{array}$$

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

\[
\begin{array}{ccc}
R & A & B & C \\
\hline
a & - & 5 & \\
- & b & 5 & \\
8 & b & - & \\
8 & - & c & \\
\end{array}
\]  
\[
\begin{array}{ccc}
\text{answer} & A & B & C \\
\hline
a & b & c & \\
\end{array}
\]

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

```
select t1.A, t1.B, t2.C
from R t1, R t2
where t1.A = 8 and t2.A = 8 and t1.C = 5
```

**Problem 2** Consider a database with relations:

```
<table>
<thead>
<tr>
<th>patient</th>
<th>pID</th>
<th>name</th>
<th>illness</th>
<th>docID</th>
<th>doctor</th>
<th>docID</th>
<th>name</th>
<th>specialty</th>
</tr>
</thead>
</table>
```

Consider the SQL query:
select p.pID 
from patient p, doctor d 
where p.illness = 'Flu' and p.docID = d.docID and d.docID in
  (select d1.docID 
  from doctor d1 
  where exists
    (select * 
    from patient 
    where illness = 'Flu' and docID = d1.docID))

(i) Rewrite the query as an equivalent conjunctive SQL query

(ii) Minimize the number of joins of the query obtained in (i), using pattern minimization.

Solution
(i) An equivalent conjunctive SQL query is the following:

    select p.pID 
    from patient p, doctor d, doctor d1, patient p1 
    where p.illness = 'Flu' and p.docID = d.docID and d.docID = d1.docID 
        and p1.illness = "Flu" and p1.docID = d1.docID

(ii) The pattern corresponding to the above SQL query is:

<table>
<thead>
<tr>
<th>patient</th>
<th>pID</th>
<th>name</th>
<th>illness</th>
<th>docID</th>
</tr>
</thead>
<tbody>
<tr>
<td>x</td>
<td></td>
<td></td>
<td>Flu</td>
<td>y</td>
</tr>
<tr>
<td>z</td>
<td></td>
<td></td>
<td>Flu</td>
<td>y</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>doctor</th>
<th>docID</th>
<th>name</th>
<th>specialty</th>
</tr>
</thead>
<tbody>
<tr>
<td>y</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The minimized pattern is:

<table>
<thead>
<tr>
<th>patient</th>
<th>pID</th>
<th>name</th>
<th>illness</th>
<th>docID</th>
</tr>
</thead>
<tbody>
<tr>
<td>x</td>
<td></td>
<td></td>
<td>Flu</td>
<td>y</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>doctor</th>
<th>docID</th>
<th>name</th>
<th>specialty</th>
</tr>
</thead>
<tbody>
<tr>
<td>y</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

and the minimized SQL query is:
select pID
from patient, doctor
where illness = 'Flu' and patient.docID = doctor.docID