

BNF101 - Exercice interactions protéine-protéine

Enoncé

Un laboratoire souhaite construire une base de données de protéines et de complexes protéiques.

Dans le cadre de cette base, nous décrirons une protéine comme possédant :

- un id (protID),
- un nom (protNom),
- une fonction (protFonc),
- une séquence d'acides aminés (protSeq),
- un code PDB (protCodePDB),
- un numéro d'accès SWISSPROT (protNumSwissProt).

Règles de gestion associées aux protéines :

- Chaque protéine interagit avec une ou plusieurs protéines.
- Chaque interaction concerne 2 protéines et est caractérisée par un attribut typeInteraction.
-

Ce complexe est décrit par :

- un id (compID),
- un nom (compNom),
- un nombre sous-unités (compNbSousU),
- un code PDB (compPDB),
- un numéro d'accès SWISSPROT (compNumSwissProt).

Règles de gestion associées aux complexes et aux protéines :

- Une protéine peut faire partie d'1 ou plusieurs complexes.
- Un complexe peut contenir un nombre variable de protéines.

Un site actif est représentée par :

- un ID (siteID),
- un nom (siteNom).
- Une séquence (siteSeq)

Règles de gestion associées aux complexes et sites actifs :

- Un complexe peut avoir un plusieurs sites actifs,
- Un site actif appartient à 1 seul complexe.

Questions :

1/ Représentez le MCD correspondant à cet énoncé.

2/ Représentez le MPD correspondant à cet énoncé.

3/Soulignez les clés primaires et les clés étrangères avec un code couleur différent.

**4/ Nous considérons que le laboratoire utilise un SGBD Oracle.
Ecrivez les scripts SQL de création des tables «siteActif» et «complexe».**

Les colonnes de la table «siteActif» sont définies comme suit :

- siteID NUMBER(5)
- siteNom VARCHAR2(30)
- siteSeq VARCHAR2(200)

Les colonnes de la table «complexe» sont définies comme suit :

- compID NUMBER(5)
- compNom VARCHAR2(30)
- compNbSousU NUMBER(2)
- compPDB VARCHAR2(8)
- compNumSwissProt NUMBER(5)

Dans quel ordre faut-il créer ces 2 tables?
Justifiez votre réponse.

Remarque :

Dans les tables ci-dessus, les éventuelles clés étrangères définies dans la modélisation du MPD ne sont pas indiquées.
Il faudra les ajouter si-besoin.

5/ Ecrivez les requêtes suivantes :

- obtenir (extraire) la liste complète des protéines,
 - obtenir (extraire) la liste complète des complexes,
 - obtenir (extraire) le nom des protéines, le nom et la séquence des sites actifs de ces protéines,
 - obtenir (extraire) le nom des protéines et le nom des complexes pour les complexes qui ont 3 sous-unités,
 - obtenir (extraire) le nom des protéines dont l'interaction est de type «coiled-coil ».
- question subsidiaire : peut-on optimiser l'enregistrement des types d'interaction ?