

RAPORT STIINTIFIC SI TEHNIC

Etapa I a proiectului Nr. 198/2014

Sistem integrat pentru modelare biomoleculara, cu aplicabilitate la studiul bacteriilor Gram negative (SIMBAGRAN)

I. REZUMATUL ETAPEI

Etapa I, intitulata "Proiectarea gridului de calcul si a sistemului de management al fluxurilor de lucru pentru modelare in biologie moleculara", a avut drept obiectiv principal specificarea arhitecturii gridului de calcul pentru biologie computationala si elaborarea proiectului sistemului de management al *workflow*-urilor ce vor fi accesibile utilizatorilor prin intermediul portalului web RoNBio.

In cadrul etapei s-au obtinut urmatoarele rezultate:

1. Proiectul gridului de calcul pentru biologie computationala

A fost estimat necesarul de resurse de calcul pentru procesarea, stocarea si comunicarea de date in sistemul informatic integrat, pe baza analizei cerintelor utilizatorilor potentiali privind modelarea si simularea numerica a sistemelor biomoleculare. S-au stabilit metodele de autentificare a utilizatorilor in vederea accesarii si comunicarii securizate cu sistemul de resurse. S-a realizat specificatia tehnica a suportului grid pentru calcule intensive, utilizand solutii open-source si luand in considerare conditia de scalabilitate la upgrade-urile ulterioare. Sistemul urmeaza sa utilizeze middleware EMI-3 iar serviciile centrale vor fi furnizate de catre Gridul National pentru Cercetarea de Fizica si Domenii Conexe.

2. Proiectul sistemului de management al fluxurilor de lucru

Au fost evaluate functionalitatea sistemelor existente de management al *workflow*-urilor pentru bioinformatica si cerintele utilizatorilor potentiali ai portalului RoNBio privind modelarea componentelor celulelor bacteriene. Pe aceasta baza s-a elaborat design-ul sistemului, reprezentand fluxurile de lucru prin grafuri si arbori in care nodurile semnifica script-uri apelabile din interfața portalului. S-a analizat posibilitatea adaptarii unor solutii open-source in vederea realizarii unei platforme de management a *workflow*-urilor capabila de integrare rapida a programelor externe si a algoritmilor de modelare in fluxuri de lucru complexe.

3. Modelul conceptual al bazei de date RoNBio

S-au identificat si evaluat sursele externe de date necesare pentru modelarea subcomponentelor bacteriene de interes si s-au determinat cerintele pe care trebuie sa le indeplineasca datele de intrare/iesire ale sistemului RoNBio. Au fost discutate diferitele formate de date utilizate si modalitatile de conversie intre acestea. Pornind de la descrierea cerintelor utilizatorilor privind informatiile persistente necesare suportului activitatii stiintifice, s-a analizat structura datelor si asociatiile dintre acestea, elaborandu-se un model conceptual al bazei de date care este descris in cadrul formalismului entitate-relatie. Datele persistente vor fi clasificate dupa taxonomie. S-a elaborat diagrama entitate-relatie corespunzatoare tipurilor de entitati si relatiilor principale ale modelului bazei de date stiintifice asociate portalului.

4. Proiectul mediului de lucru al utilizatorilor portalului RoNBio

S-a realizat o descriere la nivel conceptual a fluxurilor de lucru, a principalelor module care vor fi implementate si a relatiilor dintre ele. Modul de interactiune a utilizatorului cu mediul de lucru oferit de portal a fost exemplificat in cazul a doua fluxuri de lucru complexe asociate construirii unui bistrat LPS - fosfolipide si, respectiv, a modelarii unui fragment de membrana format dintr-o proteina inserata in bistrat LPS - fosfolipide, care au ca finalitate simularea dinamicii moleculare a acestor structuri.

5. Design-ul unor module de modelare si parametrizare a moleculelor LPS

Au fost elaborate proceduri computationale pentru determinarea parametrilor potentialului necesari in modelarea unor componente de LPS. S-au specificat modulele corespunzatoare parametrizarii moleculelor de LPS. Design-ul general a fost detaliat in cazul particular al modulelor de modelare a bacteriei *Escherichia coli*, tulpina B4:O111.

6. Design-ul fluxului de lucru pentru modelarea interactiei proteina – ligand

Au fost concepute proceduri automate pentru modelarea interactiunii ligand - proteina tinta prin metoda de andocare moleculara. In urma analizei produselor software existente, s-a decis integrarea in RoNBio a versiunii AD Vina din suita software open-source AutoDock. S-a elaborat proiectul fluxului de modelare computationala a interactiei proteinelor bacteriene cu liganzi, realizandu-se schema generala a workflow-ului si specificandu-se modulele de modelare respective.

II. DESCRIEREA STIINTIFICA SI TEHNICA

INTRODUCERE

Raportul sintetizeaza rezultatele obtinute in prima etapa a proiectului, in cadrul careia s-au realizat, printre altele, proiectele gridului de calcul pentru biologie computationala si al sistemului de management al workflow-urilor gazduit de portalul RoNBio.

Sumarul contributiilor partenerilor la rezultatele etapei este urmatorul:

- CO - determinarea functionalitatii si a capacitatilor de calcul, de stocare si de comunicare de date ale sistemului distribuit de resurse al RoNBio; elaborarea proiectului gridului de calcul pentru biologie computationala; stabilirea cerintelor satisfacute de datele de intrare/iesire in/din sistem; contributi la elaborarea modelului conceptual al bazei de date RoNBio; elaborarea de proceduri computationale pt. determinarea parametrilor potentialului necesari in modelarea unor componente de LPS; conceperea de proceduri automate pentru modelarea interactiunii dintre liganzi si proteine tinta; contributi la proiectarea fluxului de modelare computationala a interactiei proteinelor bacteriene cu liganzi.
- P1 - estimarea necesitatilor de procesare si de comunicare de date pentru calculele de biologie computationala; evaluarea functionalitatii sistemelor de management al workflow-urilor existente si a cerintelor utilizatorilor portalului RoNBio; evaluarea bazelor externe de date necesare pentru modelarea biomoleculara; descrierea la nivel conceptual a workflow-urilor, a modulelor si a relatiilor dintre ele; parametrizarea unor componente de molecule LPS si proiectarea modulelor corespunzatoare; contributi la caracterizarea interactiunii dintre liganzi si proteine tinta.
- P2 - contributi la elaborarea modelului conceptual al bazei de date RoNBio; elaborarea proiectului sistemului de management al fluxurilor de lucru; contributi la proiectarea fluxului de modelare computationala a interactiei proteinelor bacteriene cu liganzi.

III. PARTICIPARI LA CONFERINTE

- M. Mernea, I. Vasile, D. Mihailescu, *Modeling THz spectroscopy experiments on lysozyme and camel antibody binding*, COST Action MP 1204 STSM Workshop & Management Committee Meeting, Varsovia, Polonia, 13.11.2014-15.11.2014 (prezentare orala)
- G. Necula, D. Ciobanu-Zabet, *Modeling the drug-protein interaction on DFCTI's HPC platform*, 7th RO-LCG 2014 Conference, 3-5.11.2014, Bucharest, BoA p. 36, ISBN 978-973-0-17842-5, <http://rolcg2014.ifin.ro>
- Vasile, D. Ciobanu-Zabet, M. Dulea, *Recent developments of the IFIN_BC high performance computing platform*, 7th RO-LCG 2014 Conference, 3-5.11.2014, Bucharest, BoA p. 35, ISBN 978-973-0-17842-5, <http://rolcg2014.ifin.ro>