



(12) CERERE DE BREVET DE INVENȚIE

(21) Nr. cerere: a 2010 01054

(22) Data de depozit: 05.11.2010

(41) Data publicării cererii:  
30.05.2011 BOPI nr. 5/2011

(71) Solicitant:  
• INSTITUTUL NAȚIONAL DE CERCETARE  
DEZVOLTARE PENTRU ȘTIINȚE  
BIOLOGICE, SPLAIUL INDEPENDENȚEI  
NR. 296, SECTOR 6, BUCUREȘTI, B, RO

• BUTU MARIAN, STR. DEALUL ȚUGULEA  
NR.32-36, BL.15, SC.A, AP.12, SECTOR 6,  
BUCUREȘTI, B, RO;  
• BUTU ALINA, STR. DEALUL ȚUGULEA  
NR.32-36, BL.15, SC.A, AP.12, SECTOR 6,  
BUCUREȘTI, B, RO

(72) Inventatori:

(54) SISTEM BIOINFORMATIC PENTRU OBTINEREA  
TRAIECTORIILOR DE DINAMICĂ MOLECULARĂ  
A SECVENȚELOR PROTEICE

(57) Rezumat:

Prezenta invenție se referă la un sistem bioinformatic dedicat obținerii traiectoriilor de dinamică moleculară a secvențelor proteice. Sistemul conform invenției este alcătuit din unitatea de lucru pentru pregătirea secvențelor proteice și monitorizarea simulării dinamicii moleculare, din unitatea de simulare paralelă a

dinamicii moleculare și unitatea de depozitare a traiectoriilor rezultate în urma simulării dinamicii moleculare de producție.

Revendicări: 1



OFICIUL DE STAT PENTRU INVENȚII ȘI MĂRCI	
Cerere de brevet de invenție	
Nr. ....	a 2010 01054
Data depozit .....	05-11-2010

Sistem bioinformatic descris in aceasta inventie este dedicat obtinerii traiectoriilor de dinamica moleculara a secventelor proteice.

Proteinele au structura polipeptidica, in structura peptidelor intrand 20 de aminoacizi, aceeasi pentru toate proteinele naturale de pe Pamant. La un anumit nivel de complexitate a lanturilor polipeptidice apar nivele superioare de organizare si evident, proprietati noi, neintalnite la peptidele mai mici. Proteinele sunt macromolecule care exista intr-o mare diversitate. Fiecare tip de celula, fiecare individ, fiecare specie poseda un set specific de proteine. Chiar proteinele care indeplinesc functii similare la specii diferite sunt entitati diferite, cu mase si proprietati distincte.

Proteinele este deosebit de complexe, presupune intelegerea principiilor generale de constructie a moleculelor proteice, a modului in care dintr-un numar limitat de unitati structurale se pot forma infinit de multe proteine si a factorilor care intervin in realizarea organizarii lor spatiale. In al doilea rand, studiul proteinelor necesita abordarea fiecarui tip de proteina in parte, stabilirea pentru fiecare specie moleculara a constitutiei chimice, a arhitecturii si a functiei pe care o exercita.

Structura polipeptidica este apta sa asigure proteinelor cele doua caractere, diversitate moleculara si specificitate de forma. Cu ajutorul celor 20 de radicali diferiti se pot obtine combinatii multiple, deosebite prin numarul si secventa resturilor aminoacidice. Lanturile polipeptidice, unidimensionale si flexibile se convertesc spontan in edificii cu trei dimensiuni, cu forme caracteristice pentru o anumita secventa de aminoacizi. Organizarea spatiala a lanturilor polipeptidice are loc prin interactiuni necovalente (legaturi de hidrogen, interactiuni hidrofobe, interactiuni van der Waals) la care participa atat gruparile CO si NH din catena polipeptidica, cat si diverse grupari cuprinse in actenele laterale.

In ultimii ani s-au facut progrese considerabile in tehnologia computerelor, in stiinta computerelor si in stiintele computationale. Dezvoltarea tehnologiei a condus la aparitia unor computere mai rapide si mai eficiente, a supercomputerelor si a unor microcomputere mai puternice. In acelasi timp au fost identificate noi metode de a folosi aceste resurse fizice prin dezvoltarea de algoritmi care permit o maxima eficienta a prelucrarii paralele. De asemeni, pe baza acestor noutati au aparut baze de date specifice, programe de vizualizare a moleculelor, de simulare si de modelare a structurilor moleculare. Vizualizarea structurilor si fenomenelor biochimice duce la o mai buna intelegere intuitiva a proceselor in curs de studiere si evaluare. Simularea computerizata a secventelor proteice da cercetatorului din domeniul siintelor biologice un mai mare control asupra comportamentului modelului.

Modelarea moleculară poate fi descrisă ca știința reprezentării numerice a structurilor moleculare și simularea comportamentului lor prin intermediul ecuațiilor fizicii clasice și cuantice. Aceasta permite oamenilor de știință să genereze și să prezinte datele moleculare incluzând aici datele de geometrie (lungimea legaturilor, unghiurile legaturilor, unghiurile de torsiune), energiile (caldura de formare, energia de activare, etc.), proprietatile electronice (momente, sarcini, potential de ionizare, afinitatea electronica), proprietatile spectroscopice (moduri de vibratie, deplasari chimice) si proprietati de ansamblu (volumul, aria suprafetelor, difuzia, vascozitatea, etc.).



Compararea datelor de simulare cu datele experimentale este importanta pentru a ghida atat munca de laborator, cat si abordarea computationala.

Simularea moleculara a secventelor proteice poate fi considerata o aplicatie a tehnicilor computerizate asupra metodelor fizicii teoretice si datelor experimentale. Aceasta este o punte intre teorie si experiment, permitand:

- extragera rezultatelor dintr-un model particular;
- compararea rezultatelor experimentale;
- compararea predictiei teoretice a unui sistem cu comportamentul experimental;
- o mai buna intelegere si interpretare a observatiilor experimentale;
- corelatii intre detaliile de la nivel atomic si molecular cu proprietatile macroscopice ale sistemului;
- obtinerea unor informatii care nu pot fi obtinute din experimente reale.

Pentru realizarea celor mai sus mentionate au fost metode speciale care dezvoltate indeplinesc functii specifice, cum sunt:

1. *Generarea, regāsirea sau restabilirea structurilor moleculare.* Structurile cristaline ale compușilor organici pot fi gāsitate în diferite baze de date disponibile pe internet. Coordonatele 3D ale biomoleculelor pot fi gāsitate în Cambridge Crystallographic Datafiles (<http://www.ccdc.cam.ac.uk/>), precum și în Protein Data Bank (<http://www.rcsb.org/pdb/>). Iar cele care nu pot fi gāsitate pot fi generate de către programe de transformare 3D.
2. *Vizualizarea structurilor.* Grafica computerizată este mijlocul cel mai eficient de vizualizare și manipulare interactivă a moleculelor și sistemelor moleculare. Dintre numeroasele programe de vizualizare enumerăm : VMD (Visual Molecular Dynamics), SPDBV (Swiss-PdbViewer DeepView), JMV (Java Molecular Viewer), RasMol, PyMOL, etc.
3. *Calcularea energiilor și minimizarea.* O proprietate fundamentală a moleculelor este conținutul lor energetic și nivelul energetic. Există trei metode computaționale, teoretice, de calcul a energiilor dintr-o moleculă: abordarea empirică (mecanică moleculară), semiempirică și ab initio (mecanica cuantică). Minimizarea energetică rezultă din optimizarea geometriei structurii moleculare.
4. *Simularea dinamică și căutarea conformației optime.* Rezolvarea mișcării nucleului în câmpul de forțe al electronilor se numește dinamică cuantică. Soluția ecuației lui Newton a mișcării nucleilor este cunoscută ca dinamică moleculară. Integrarea ecuației de mișcare a lui Newton pentru toți atomii dintr-un sistem generează traiectoriile moleculare. Căutarea și gāsirea conformației optime se realizează prin repetarea procesului de rotație a unor legături de referință ale moleculei de investigat pentru a găsi conformația cu energia cea mai mică.
5. *Calculul proprietăților moleculare.* Au fost dezvoltate metode de calcul sau estimare a proprietăților moleculare, de exemplu prin interpolare, extrapolare sau calcul efectiv. Unele dintre cele calculate sunt: punctul de fierbere, volumul molar, solubilitatea, capacitatea calorică, densitatea, indicele de refracție, susceptibilitatea magnetică, momentul de dipol, sarcina atomică parțială, potențialul de ionizare, potențialul electrostatic, aria suprafeței van der Waals, aria suprafeței accesibile.
6. *Superpoziția și alinierea structurilor.* Activitățile de calcul și proprietățile moleculelor implică

deseori compararea unor structuri dintr-o serie de omologie. Aceste comparații se fac cu ajutorul tehnicilor de superpoziționare și de aliniere a structurilor respective.

7. *Interacțiuni moleculare, docking.* Sunt importante de înțeles interacțiunile moleculare dintre un ligand și receptorul său și sunt necesare exerciții dificile de modelare a sistemului. De obicei receptorul este menținut rigid, iar moleculei de ligand i se permite să-și modifice conformația.

### **Pregătirea și Rularea Dinamicii Moleculare**

Intr-o simulare a dinamicii moleculare, comportamentul în funcție de timp al sistemului molecular este obținut prin integrarea ecuațiilor de mișcare ale lui Newton, folosind integratori numerici și funcția energiei potențiale. Rezultatul simulării este reprezentat de serii de conformații timp dependente; acestea poartă denumirea de traiectorii, sau caile pe care le parcurge fiecare atom în concordanță cu legile newtoniene de mișcare. Majoritatea simulărilor de dinamică moleculară sunt desfășurate în condiții constante pentru N, V, E (ansamblul microcanonic), numai că în ultimul timp, sunt folosite metode care presupun simulări în condiții constante pentru N, T și P, pentru a mima mai bine condițiile experimentale.

*Initializarea.* Pentru a porni o simulare de dinamică moleculară, trebuie aleasă o configurație inițială a sistemului, un punct de plecare,  $t=0$ . De cele mai multe ori, în simulările pe biomolecule, se folosește o structură obținută prin cristalografie cu raze X sau RMN, și care se găsește în banca de date *Brookhaven Protein Databank*; aceasta se folosește drept structură inițială. Este de asemenea posibilă utilizarea unei structuri teoretice, dezvoltată prin modelare prin omologie.

Un lucru important în simularea dinamicii moleculare îl reprezintă impunerea condițiilor la limită. Aceste condiții permit simulării să se desfășoare utilizând un număr mic de particule în așa fel încât particulele existente să resimtă forțe ca și cum s-ar afla într-un volum mare de soluție. Cutia de apă centrală este înconjurată de alte cutii vecine, iar coordonatele particulelor din cutiile vecine (particule imagine) se generează prin simplă translație a coordonatelor particulelor centrale (particule primare). Forțele care acționează asupra particulelor primare se calculează pornind de la particulele din aceeași cutie, dar și din cutiile imagine.

Înainte de a porni dinamică moleculară, este recomandat să se facă o minimizare a energiei structurii. Acest proces reduce toate interacțiunile van der Waals puternice care ar putea exista; prezenta acestora ar putea duce la distorsiuni moleculare în regiunea respectivă, și se realizează astfel o simulare instabilă.

După acest punct, molecule explicite de apă sunt adăugate pentru a solvata biomolecula. Dacă se porneste de la o structură elucidată prin cristalografie cu raze X, atunci este posibil ca o serie de molecule de apă să fie deja prezente, dar cantitatea este mai mult decât insuficientă pentru solvatare. Moleculele de apă pentru solvatare se obțin prin construirea unei cutii de apă optime, care se echilibrează înainte de a realiza solvatarea. Ulterior întreaga cutie de apă se suprapune peste proteina, iar acele molecule de apă, care se suprapun peste atomii biomoleculei, sunt îndepărtate. În această etapă, se impune realizarea unor alte minimizări. Mai întâi trebuie fixată proteina în poziția cu energie minimizată, astfel încât atomii să nu se deplaseze, iar minimizarea să se facă asupra moleculelor de apă, pentru a le permite reajustarea poziției lor în



jurul proteinei. Apoi, într-o minimizare ulterioară, în care coordonatele proteinei sunt eliberate, se supune întreg sistemul biomolecula-solvent reducerii energiei potențiale.

*Incazirea sistemului.* Se atribuie viteze inițiale la temperatura joasă pentru fiecare dintre atomii sistemului, iar ecuațiile newtoniene de mișcare sunt integrate pentru a propaga sistemul în timp. Simularea este pornită, și în mod periodic, se calculează noi viteze la o temperatură puțin mai ridicată. Acești pași se repetă până se atinge temperatura dorită.

*Echilibrarea.* Odată atinsă temperatura dorită, simularea pe sistemul biomolecula-apă continuă, iar pe parcursul acestei faze se monitorizează o serie de proprietăți, precum structura, presiunea, temperatura și energia. Rolul fazei de echilibrare este acela de a rula simularea până aceste proprietăți devin stabile în raport cu timpul. Dacă temperatura crește sau scade semnificativ, vitezele pot fi ajustate astfel încât temperatura să revină la valorile dorite.

*Faza de „Productie”.* Etapa finală a simulării este aceea de a rula simularea în faza de „productie” pentru durata de timp avută în vedere. Aceasta poate fi de ordinul sutelor de ps, până la ns sau mai mult. În timpul acestei faze se calculează parametrii termodinamici.

În contextul celor mai sus descrise invenția propune un sistem bioinformatic alcătuit din trei unități de lucru care să funcționeze astfel: pe prima unitate se realizează pregătirea secvenței peptidice pentru simularea dinamicii moleculare și monitorizarea simulării, pe a doua se realizează simularea paralelă a dinamicii moleculare și în a treia se depozitează traiectoriile obținute în urma simulărilor. Pregătirea secvenței proteice pentru simulare se realizează prin convertirea secvențelor proteice pentru utilizare în pachetele de programe de simulare a dinamicii moleculare, minimizează energia sistemului proteic, încălzirea și echilibrarea sistemului proteic. Sistemul proteic rezultat este transferat în unitatea de simulare a dinamicii moleculare unde se obțin traiectoriile care, apoi, sunt transferate în unitatea de depozitare.



### Revendicare

Sistemul bioinformatic pentru obtinerea traiectoriilor de dinamica moleculara a secventelor proteice este **caracterizat prin aceea ca** este alcatuit din unitate de lucru pentru pregatirea secventelor proteice si monitorizarea simularii dinamicii moleculare, unitate de simulare paralela a dinamicii moleculare si unitate de depozitare a traiectoriilor rezultate in urma simularii dinamicii moleculare de productie. In cadrul unitatii de lucru in vederea pregatirii pentru simulare se realizeaza urmatoarele operatii: convertirea secventele poteice pentru utilizare in pachetele de programe de simulare a dinamicii moleculare, minimizeaza energiei sistemului proteic, incalzirea si echilibrarea sistemului proteic. Sistemul proteic rezultat este transferat in unitatea de simulare a dinamicii moleculare unde se obtin traiectoriile care, apoi, sunt transferate in unitatea de depozitare.

