

РЕЦЕНЗИЯ

за дисертация за присъждане на образователната и научна степен „доктор“ на тема „In silico предсказване на триизмерната структура на протеините чрез евристични алгоритми“

с автор ас. Иван Благоев Тодорин, Югозападен университет Благоевград
от проф. д-р Никола Ив. Янев, ЮЗУ, ИМИ БАН и ФМИ Софийски
университет

choby@math.bas.bg

Обща информация

Основание за написване на това становище е зап. № 1346/23.07.2019 на Ректора на ЮЗУ Благоевград, както и решението на Научното жури от неговото първо заседание.

Дисертационният труд е в област на висше образование 4. Природни науки, математика и информатика, професионално направление 4.6 Информатика и компютърни науки, докторска програма „Информатика“. Докторантът е на самостоятелна подготовка с научен ръководител доц. д-р Иван Тренчев.

Актуалност на изследването

Дисертацията е посветена на изключително предизвикателна и **актуална** задача за предсказване на 3-мерната структура (огъване, конформация и др.) на протеини въз основа на първичната им структура. Популярен подход е създаване на математически модел от клас нелинейно оптимизиране с неизвестни - координатите на централните карбонови атоми и/или центровете на страничните вериги. Целевата (авторът използва термина оценъчна) функция, чиито екстремум се търси описва енергията на конформацията и е ключов елемент на модела. В HP моделите това е

просто броя на хидрофобните контакти, а в другите случаи тя е свързана с представата на изследователя за тайнството на природата , управляващо огъването на линейната структура (опъната нишка от аминокиселинна редица) до уникално кълбо , наречено fold , което единодушно се приема за определящо функционалността на протеина. Допустимата област за променливите на моделите варира от дискретно множество от точки в 3-мерното пространство (най-често различни решетки) до извънрешетъчни модели.

Цели на дисертацията

Най общо целите на дисертацията могат да бъдат формулирани като:

използване на методите и

средствата на изследване на операциите за решаване на споменатата по-горе биоинформатична задача. Формулирани са 4 задачи на нелинейното оптимизиране, които се различават по целевите функции и ограниченията. Последната задача включва и допълнителни променливи , свързани с позициите на центровете на т.н. странични вериги (side chain). Обединяващата идея на алгоритмите за решаване на задачите е подхода на „ грубата сила “ , състоящ се в преодоляване на информатичното предизвикателство за бързо генериране на случайни допустими конформации върху кубична решетка и в 3-мерното пространство с ограничени степени на свобода. Процесът на генериране се управлява от геометрични съображения, а в случая на използване на side chains , пертурбацията на пространствените координати отчита и ефекта върху целевата функция (greedy approach).

Анализ на оформлението и на съдържанието на дисертацията

Представеният труд съдържа 137 страници, разпределени по обичайния начин в

Съдържание, Увод, 4 глави (2-5), Заключение, Приноси, Литература, Приложения.

Спазени са изискуемите от чл.27(2) на ППЗРСАРБ структурни и съдържателни атрибути на дисертацията. По отношение на езика са възниквали (предполагам) проблеми в превод на български като напр.с.16 „Много малко се знае за self-avoiding walk стриктно погледнато от математическа перспектива, въпреки че физиците са направили много предпоставки, за които се вярва, че са точни и са силно подкрепени с бройни симулации“. Изключително дразнещото е разточителното по обем представяне на решенията получени от компютърната реализация на алгоритмите при различни стойности на параметрите, което може да породи излишни съмнения за информационно стилистичните възможности на автора.

Има и проблеми в цитирането –

например няма позоваване към [8], [22], [24], [31] и др. и обратно двата основни източника на обзорната част -“Computational Methods for Protein Structure Prediction & Modeling V1 - Xu Xu and Liang” (с.25) и „Optimization Approaches to Protein Folding” - Hyun-suk Yoon (с.31) не фигурират сред източниците.

Обзорната Глава 3 подсказва за подходящо за целите на дисертацията допълване на образователната компонента на автора , въпреки че по информативно би било замяната на дългата разходка в решетки със сведения за съществуващи подходи за решаване на нагъването като задачи на нелинейната оптимизация. Следващите глави са посветени на: формулиране на задачи за свиване на протеини като задачи на нелинейната оптимизация, алгоритми за решаването им, изчислителни експерименти и анализ на резултатите.

Идентифициране и оценяване на научните и научно-приложните приноси в дисертационния труд

Получените резултати са научно-приложни , без съмнение оригинални с не добре показана значимост:

В дисертацията са предложени 4 модела с различни (авторски ?) целеви функции и допустими множества от решетъчен и извънрешетъчен тип и съответни оптимизационни алгоритми , чиято ефективност, в смисъл генериране на огъвания геометрично подобни на референтните такива, е твърде скромно демонстрирана за решетъчните модели, а за извънрешетъчния модел по-добре е показана неприложимостта на подхода за богат клас протеини, което оставя впечатление за антиреклама. Известно недоумение предизвиква и липсата на оценка на адекватността на целевите функции, чиято стойност е анонсирана за намерените от алгоритмите конформации, но не и за референтните протеини.

Публикации по дисертацията

Анонсирани са необичайно голям брой публикации за такава дисертация – 12, от които scholar google вижда само 5.

Всички са на английски, една е в авторитетно международно списание, останалите – в сборници от предимно международни научни конференции. Наличието на 5 самостоятелни публикации поставя **самостоятелността на приносите** извън всякакво съмнение.

6. Автореферат

По структура и съдържание оценявам автореферата положително, като адекватно по обем разширено резюме на дисертацията.

7. Критични бележки и препоръки (освен вече направените по-горе)

Препоръчвам – цитирам автора:

„ Ако изследването се проведе с оптимизирана оценъчна функция и с по-голям изчислителен ресурс, то най-вероятно ще се постигне и по-голямо геометрично подобие. Това оправдава продължаването на експериментите в тази посока, с цел достигането на достатъчно бързо и точно работещ модел, което да позволи реалното му приложение за целите на биохимията.

*При модела с целенасочено дооформяне на структурата, резултатите превъзхождат получените по предходните модели по степен на молекулярно геометрично подобие, а основното предимство е десетки пъти по-бързото получаване на предсказаната структура, **което ще** позволи създаването на модел с много по-детайлно и много близко до реалното представяне на протеиновата структура, който, с помощта на създадения алгоритъм за целенасочено формиране, може да работи със задоволителна скорост и може да даде възможност да се предскажат значим брой структури, което със скоростта на досега използваните в света методи на предсказване е изчислително непосилно „ продължаване на изследванията до отстраняване на условното наклонение от горното прозрение.*

Заклучение

Дисертацията има съдържание и форма, съответстващи на нормативните изисквания за дисертация за получаване на образователната и научна степен "доктор".

Критичните ми бележки не влияят на крайната оценка, която е **положителна** и да препоръчам даването на ас. Иван Благоев Тодорин образователната и научна степен „доктор“ в професионално направление 4.6

Информатика и компютърни науки, докторска програма „Информатика“.

София, 29.08.2019

проф. д-р Н. Янев