

СОФИЙСКИ УНИВЕРСИТЕТ
“СВ. КЛИМЕНТ ОХРИДСКИ”

ФИЗИЧЕСКИ ФАКУЛТЕТ
КАТЕДРА “АТОМНА ФИЗИКА”



Елена Боянова Лилкова

АВТОРЕФЕРАТ

на дисертация

за присъждане на образователната и научна степен “Доктор”,
в професионално направление “4.1 Физически науки”,
научна специалност “01.03.19 Физика на атомите и молекулите”

**ИЗСЛЕДВАНЕ НА
ЧОВЕШКИ ИНТЕРФЕРОН ГАМА
ЧРЕЗ МОЛЕКУЛНО-ДИНАМИЧНИ
СИМУЛАЦИИ**

Научен ръководител:
доц. д-р Леандър Литов

София, 2015 г.

Авторката е редовен докторант в катедра “Атомна физика” на Физически факултет на Софийски университет “Св. Климент Охридски” за периода 01.07.2011–01.07.2015 г.

Дисертацията съдържа списък със съкращения, увод, пет глави, заключение, списък с научните приноси, списък с научните статии, на които се базира трудът, благодарности, две приложения и списък с цитираната литература. Работата е изложена на 118 страници и включва 9 таблици и 53 фигури. Цитирани са 109 литературни източника.

Дисертационният труд е обсъден на заседание на разширен катедрен съвет на катедра “Атомна физика” на Физически факултет на Софийски университет “Св. Климент Охридски” на 29.06.2015г.

Защитата на дисертационния труд ще се проведе на в аудитория във Физическия факултет на Софийския университет ”Свети Климент Охридски”, пред научно жури в състав:

Съдържание

Цели и структура на дисертацията	1
1 Въведение	3
2 Молекулно моделиране на биологични макромолекули	7
3 Анализ на скалируемостта на алгоритмите за едромасщабни симулации на биологични макромолекули	14
4 Изследване на стабилността на локалната структура на мутагенни форми на hIFN- γ	19
5 Молекулно моделиране на комплекса hIFN1- γ -hIFN γ R α -производни на хепарина олигозахариди	25
Научни приноси	36
Научни статии, на които се базира дисертация	37
Благодарности	39
Библиография	40

Цели и структура на дисертацията

Дисертацията е посветена на моделирането на човешки интерферон гама с методите на молекулната динамика. Интерферон- γ (IFN- γ) е сигнална молекула, която играе основна роля в регулацията на защитата на организма и имунопатологичните процеси, като проявява антивирусни, антипролиферативни и имуномодулаторни свойства. Аномалната експресия на IFN- γ се свързва с множество автовъзпалителни и автоимунни заболявания. Ето защо се търси начин за контролиране на неговата активност. Целта на дисертацията е да се изгради модел на взаимодействията на цитокина, който да обясни как той взаимодейства с рецепторите си и с други лиганди. Такъв модел би подпомогнал усилията да се намери начин за инхибиране активността на IFN- γ .

Компютърните симулации играят много важна роля в съвременната наука. Те са междинното звено между теория и експеримент. От една страна, те се основават на теоретични модели и се използват за предсказване на свойства и процеси въз основа на теоретични разглеждания. От друга страна, компютърните симулации могат да се използват за тестване на теоретични модели, което до голяма степен ги нарежда сред експерименталните методи. В много случаи компютърните симулации могат да служат за изследване на проблеми, които не могат или е прекалено скъпо да бъдат изследвани експериментално.

Молекулната динамика е метод за компютърно симулиране на многочастични класически системи, в който се проследява взаимодействието на атомите и молекулите за определен период от време под действие на законите на физиката. Молекулно-динамичните симулации предоставят много подробна информация, която често е недостъпна за лабораторните експерименти. Ето защо те могат да се разглеждат като виртуални *in silico* експерименти. Днес молекулната динамика се използва в много области от физиката – физика на кондензираната материя и твърдото тяло, материалознание и биофизика.

Дисертацията съдържа увод, пет глави и заключение. В първата глава е направено кратко описание на функциите и структурата на човешки интерферон гама и двата му известни лиганда – извънклетъчния домейн на α -рецептора sIFN γ R α и гликозаминогликаните хепарин и хепаран сулфат. Втора глава въвежда основни понятия и методи в молекулно-динамичните симулации на биологични макромолекули.

Въпреки прогреса в съвременните компютърни технологии и изчислителните методи, изследването на свойства и процеси в биологичните системи изисква огромни изчислителни ресурси и време. Това налага да се изследва влиянието на различни параметри на симулацията върху производителността ѝ. Серия от такива изследвания са представени в глава трета. Целта е да се оптимизира методологията на симулация за максимално ефективно оползотворяване на достъпните ресурси.

В четвърта глава са представени резултатите от молекулно-динамични симулации и метадинамични симулации на сто мутагенни форми на човешки интерферон гама. Целта на изследванията е да се открие подходяща мутагенна форма на този белтък, която да е в състояние да се свързва с клетъчния му рецептор, но да не предизвиква биологична реакция в клетката. Описани са и методите за анализ на измененията във вторичната структура

на спиралата, съдържащи мутираните аминокиселинни остатъци. Въведени са критерии за оценка на стабилността на локалната структура, в която мутираните остатъци участват, въз основа на анализ на повърхнината на свободната енергия на системата в пространството на подходящо избрани колективни променливи.

Последната пета глава е посветена на изследвания на взаимодействията на IFN- γ с протеогликаните хепарин и хепаран сулфат, както и тяхното влияние върху свързването на цитокина с рецепторите му. Множество предварителни симулации показаха, че чистото взаимодействие между IFN- γ и IFN γ R α не е достатъчно за бързо и правилно формиране на комплекс. Това ни доведе до хипотезата за необходимостта от участие на гликозамингликани като помощни молекули в този процес. За проверката на тази хипотеза се наложи разработването на силово поле за един от монозахаридите, изграждащи хепарин и NS-областите на хепаран сулфат, с които IFN- γ взаимодейства, а именно 2-N-сулфатиран глюкозамин 2-сулфат. Получените параметри позволиха да се изследва влиянието на производни на хепарина олигозахариди с различна степен на полимеризация върху свързването на цитокина и извънклетъчните му рецептори.

Дисертацията завършва с изброяване на научните приноси на дисертанта, списък на залегналите в основата ѝ публикации и списък на използваната литература. В две приложения са представени получените параметри и топология за хепарин.

Глава 1

Въведение

1.1 IFN- γ и неговите рецептори

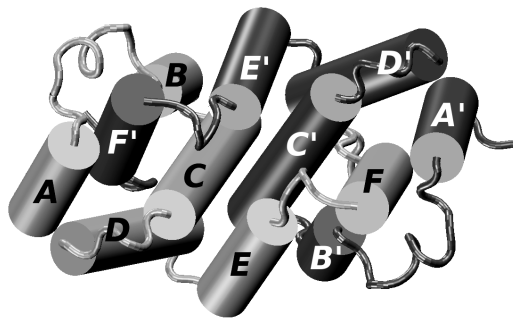
Интерферон- γ (IFN- γ) е плейотропен цитокин, който играе основна роля в регулацията на защитата на организма и имунопатологичните процеси [1]. Установено е, че стотици гени могат да се регулират чрез стимулация с IFN- γ . Много от тях съответстват на транскрипционни фактори, регулиращи от своя страна други гени, което обяснява извънредната плейотропност на цитокина [2, 4]. Въпреки, че е необходим за овладяването на инфекции, високи нива на цитокина са свързани с патогенната му роля в автоимунни и автовъзпалителни заболявания, включително възпалителни заболявания на червата, множествена склероза, захарен диабет [3, 5], и оказват негативно влияние върху някои видове тумори [6]. Това налага търсенето на начин да се контролира биологичната активност на IFN- γ при повишена експресия на белтъка.

IFN- γ е нековалентен хомодимер, всяка поединица е с молекулна маса 17kDa и се състои от 143 аминокиселини, организирани в 6 α -спирали, които обхващат около 66% от молекулата (фиг. 1). Спиралите са свързани помежду си с къси примки от по три до пет аминокиселинни остатъка, с изключение на примка АВ, която се състои от 13 остатъка, които обграждат спирала F. Останалата част от веригата след спирала F (а.к. остатъци 121-143) се разпростира извън глобулата на молекулата в разтворителя [7–9].

Човешкият IFN- γ (hIFN- γ) е базичен протеин, богат на титруеми групи. Всеки мономер съдържа 54 йоногенни групи, което прави около една трета от общия брой аминокиселинни остатъци. Сред тях 28 са лизини и аргинини, 10 от които са локализиращи в С-края. Това дава на цитокина нетен заряд $+16e^1$ при физиологични условия [10].

Един от най-неясните и спорни въпроси около IFN- γ е свързан с ролята, която последните 22 аминокиселини, разположени в С-края, играят в биологичната активност на цитокина. Това до голяма степен се дължи на липсата на данни от рентгенова дифракция за тази част от молекулата [7–9]. С-краят на IFN- γ е силно заредена, изложена на разтворителя опашка без твърда конформация [11], която е изключително чувствителна към протеолитично действие [12, 13]. С-опашката на биологично активния естествен IFN- γ е хетерогенна [12], като от нея могат да липсват между 5 и 16 а.к. остатъка [13]. Отстраняването на къси аминокиселинни последователности не намалява биологичната активност на цитокина, дори напротив [14, 15]. С-краят съдържа два силно заредени домейна — D1 със заряд $+5e$, обхващащ а.к. остатъци 126-132, и D2 със заряд $+3e$, а.к. 138-141. Участъкът D2 по-скоро има модулираща функция и не е от критично значение за свързването с рецептора. Изглежда, че участието на С-края в свързването с рецептора не се осъществява чрез

¹ e е зарядът на електрона.



Фигура 1: Структура на човешки интерферон гама. С цилиндри са представени α -спиралите на белтъка.

структурно дефиниран интерфейс. Приносът на С-опашката в свързването се обяснява с неспецифични взаимодействия между базичните аминокиселини на IFN- γ и киселинните клъстери в рецептора му, включващи множество конформационни състояния [9].

Експериментално е показано, че при свързване на IFN- γ с рецептора, комплексът се интернализира в клетката [16] чрез така наречените сигнали за ядрена локализация (NLS), които служат за белязване на протеини за транспортиране в клетъчното ядро чрез ядрени порови комплекси. В hIFN- γ са открити две предполагаеми NLS – горна в спирала Е (остатъци 83-89) [17] и долна, локализирана в неструктурирания С-регион (остатъци 126-132) [18]. Установено е, че NLS последователността играе важна роля в предизвикването на биологичен отговор в клетките [18, 19], но няма единно мнение коя от двете вериги е по-съществена [20, 21].

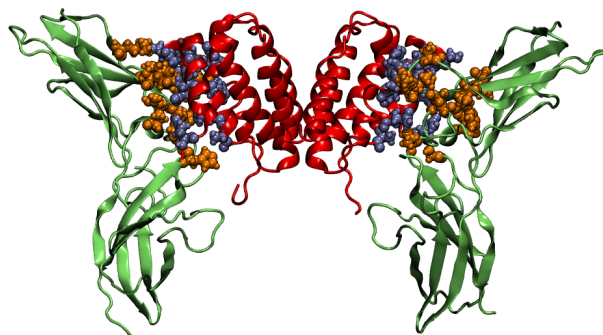
Функционалният рецепторен комплекс на IFN- γ се състои от два мономера на лиганд-свързващата верига IFN γ R α и два на помощен фактор IFN γ R β . [22]. IFN γ R α е отговорен не само за взаимодействието с цитокина, но и участва във вътреклетъчната сигнализация. IFN γ R β не взаимодейства директно с IFN- γ , но е необходим за предаването на сигнала [22, 23].

IFN- γ взаимодейства специфично и с висок афинитет ($K_d = 10^{-10}$ M) с извънклетъчния домейн на α -рецептора на човешки интерферон (shIFN γ R α), [22]. Въпреки високия афинитет, формирането на комплекс само между тези белтъци не е достатъчно за трансдукция на сигнала и предизвикване на биологичен отговор [24]. Този процес изисква задължителното участие на 5 белтъка: IFN γ R α , IFN γ R β , Jak-1, Jak-2 и STAT1 [23].

Кристалографската структура на комплекса hIFN- γ -shIFN γ R α е получена с рентгенова дифракция и е достъпна в базата данни PDB [25] с ID 1FG9 [9]. Две shIFN γ R α молекули свързват двустранно двете идентични повърхнини на димера на hIFN- γ и така формират 2:1 hIFN- γ -shIFN γ R α комплекс със симетрия от втори порядък (фиг. 2).

IFN- γ се свързва с извънклетъчния домейн от веригата на IFN γ R α чрез N-края си [9, 26]. Аминокиселинните остатъци в IFN- γ , които образуват повърхнината, контактуваща с всеки от рецепторите, са групирани в два сегмента – спирала А, примка АВ и спирала В на единия мономер (а.к. остатъци 1-42) и спирала F' на другия мономер (а.к. остатъци 108-120). Този интерфейс е изключително стабилен, като от всяка страна се формират по 18 водородни връзки. Не се наблюдават контакти с рецептора след Ala¹¹⁸ на цитокина. Структура 1FG9 съдържа координати на атомите до 126-тата аминокиселина на hIFN- γ . С-краища не се виждат в кристалографската структура, защото са много подвижни и неструктурирани.

Предложен е алтернативен на класическия модел на сигналната каскада [27], който започва с взаимодействието на цитокина с α -рецепторите, асоциирането на β -рецепторите към комплекса, активиране на асоциираните към рецепторите Jak кинази, фосфорилиране



Фигура 2: Кристалографска структура на комплекса hIFN- γ -shIFN γ R α . Цитокинът е изобразен в червено, а рецепторите – в зелено. Със сини и оранжеви сфери съответно са представени аминокиселинните остатъци формиращи контактния интерфейс в IFN- γ и IFN γ R α .

на Tyr⁴⁴⁰ в молекулите на IFN γ R α , скачване на латентни STAT1 мономери, последвано от тяхното фосфорилиране и димеризация. Следващият етап в този модел е ендоцитозата на така формирания комплекс. Молекулите на IFN γ R β се дисоциират от комплекса и излизат отново на клетъчната повърхност. Останалата част от комплекса, IFN- γ -IFN γ R α -Jak1-Ja2-STAT1 димер, влиза в клетъчното ядро посредством активен ядрен транспорт, като транспортният протеин импортин α се свързва с една от NLS веригите на IFN- γ и така транслокира целия сигнал комплекс в ядрото [27]. Този модел предлага директно участие на IFN- γ и IFN γ R α в активацията на различни гени.

1.2 Глюкозаминогликаните хепарин и хепаран сулфат

Глюкозаминогликаните (ГАГ) са неразклонени полизахаридни вериги, съставени от повтарящи се дизахаридни единици. Единият монозахарид винаги е аминокиселина, която често е сулфатирана, а вторият обикновено е уронова киселина. Тъй като захаридите съдържат сулфатни или карбоксилни групи, ГАГ са силно отрицателно заредени и са най-анионните молекули, които животинските клетки могат да синтезират. ГАГ се свързват ковалентно с протеини, наречени сърцевина (core protein), и формират т. нар. протеоглигани (ПГ). Те са закачени по клетъчната мембрана, където формират извънклетъчна обвивка и участват в извънклетъчния матрикс.

Хепаран сулфат (ХС) ПГ се намират по повърхността на почти всички клетки на бозайниците и имат разнообразни мотиви във веригите си. Това се отразява на обхвата на взаимодействията и предполагаемите им функции. Основният повтарящ се дизахарид в ХС е изграден от 1 \rightarrow 4 свързана последователност от глюкозамин и уронова киселина [28]. Глюкозаминът е или N-ацетилян (GlcNAc) или N-сулфатилян (Glc2NS), като последната група се намира само в ХС и хепарин. Аминокиселината може да бъде естер(О)-сулфатирана при С6. Уроновите остатъци са представени или като глюкуронова киселина, или като идурунова киселина, която може да бъде О-сулфатирана при С2. ХС ГАГ имат специфична структура с разделени клъстери, които или са несулфатирани (NA-области), или сулфатирани (NS-области).

Хепаринът е съставен от същите дизахариди като NS-областите на ХС, но въпреки това се води отделно ГАГ семейство. За разлика от ХС ГАГ, които се произвеждат от почти всички клетки, хепаринът се произвежда само от мастоцитите и базофилните гранулоцити и се отделя като ГАГ в кръвоносната система в близост до наранени тъкани. Той е силно сулфатилян (над 80% от глюкозамините са N-сулфатирани) и винаги има моларен излишък

на O- над N-сулфати, като средно един дизахарид има по 2.7 сулфатни групи [28–30]. Подобно на NS-областите в ХС ГАГ, хепаринът притежава едновременно необичайна подвижност благодарение на идуроновите остатъци и забележителна конформационна стабилност на полимера заради гликозидните връзки, които поддържат цялостната форма на молекулата [28].

1.3 IFN- γ и ГАГ

IFN- γ има три базични последователности, наподобяващи консенсусните вериги за свързване на ГАГ – K⁸⁷KKR⁹⁰, K¹²⁶TGKRKR¹³² и R¹³⁸GRR¹⁴¹. Първите две обхващат а.к., участващи в NLS веригите на протеина, а последната е D2 домейна в С-края. И трите последователности се намират по повърхността на протеини и са достъпни за разтворителя и за ГАГ.

Установено е, че IFN- γ се свързва с ХС ПГ по повърхността на клетките с висок афинитет ($K_d \approx 1$ nM) [31]. Взаимодействието с хепарин е с афинитет $K_d = 1.4 \pm 0.2$ nM, а с ХС ГАГ $K_d = 2.4 \pm 0.24$ nM [32]. Предполага се, че скоро след секретитрането, IFN- γ се захваща от ХС ГАГ в близост до секретиращите клетки, осигурявайки локално действие на цитокина. Интерфероновите молекули, които не са свързани с ХС, бързо се инактивират от протеази, което не допуска нежелана дифузия на цитокина [33].

Експерименталните данни показват, че във взаимодействието с ХС участват двете С-крайни базични вериги D1 и D2 [33]. Те допринасят различно за свързването. D1 е критичен за свързването, като синтетичен пептид, в който тази последователност липсва, не се свързва с хепарин. Този факт показва още, че D2 сам по себе си не осигурява свързване на цитокина с хепарин, дори и при високи концентрации. Въпреки това, в присъствието на D1, D2 играе решаваща роля, като отстраняването му води до 60-кратно намаление на афинитета на свързване. Следователно D1 е необходим за формирането на комплекс IFN- γ -HS и представлява действителното място на свързване. D2 не взаимодейства директно с хепарина, но повишава скоростта на асоциация, като чрез неспецифични електростатични взаимодействия спомага за оптимално благоприятна ориентация на двете молекули за формиране на комплекс [32].

IFN- γ -свързващите ХС ГАГ са съставени от две крайни сулфатирани вериги с дължина шест или осем дизахарида, разделени от предимно N-ацетилирана последователност, богата на GlcA. Оптималната дължина на тази част е около 5 nm, което дава обща дължина на ХС олигозахарида от 11.5 nm [34].

Глава 2

Молекулно моделиране на биологични макромулекули

2.1 Молекулна динамика

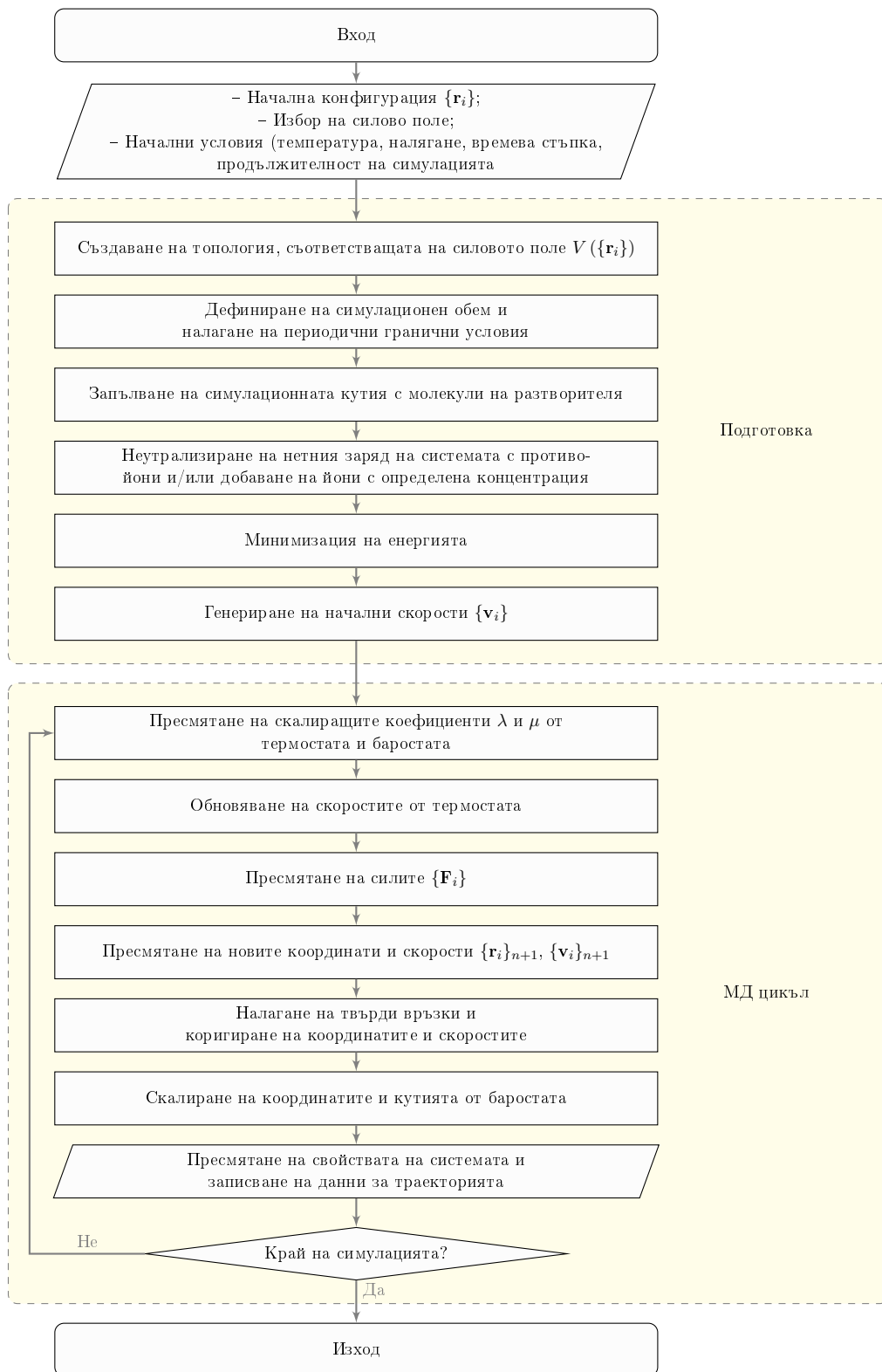
Един от основните инструменти в теоретичното изследване на биологични молекули е молекулната динамика (МД). Това е техника за пресмятане на равновесни и транспортни свойства на класически многочастични системи [35,36]. Физическата система се състои от дискретни частици и взаимодействия между тях и може да бъде описана чрез техните координатите, скоростите и действащите им сили. В МД се пренебрегва движението на електроните и се проследява само еволюцията във времето на ядрата. За потенциал на взаимодействие между ядрата се използва т. нар. силово поле (СП), което е приближение на повърхнината на потенциалната енергия в основно състояние, която се дължи на електронната структура и на междуйдрените взаимодействия. Микросъстоянието на системата се нарича конфигурация $(\{\mathbf{r}_i\}, \{\mathbf{v}_i\})$, $i = 1, \dots, N$, където N е броят на атомите в системата, която се описва с набор от $3N$ координати и $3N$ скорости.

Еволюцията във времето на системата от N частици се определя чрез интегриране на уравненията на движение на Нютон. Уравненията се дискретизират и се решават едновременно за малки времеви стъпки. Ако е необходимо, температурата и/или налягането се поддържат постоянни. Координатите като функция на времето представляват траекторията на системата. След известен период от време системата достига до равновесие и чрез усредняване по равновесната траектория могат да се пресметнат макроскопичните ѝ свойства.

Принципната схема на една МД симулация е представена на фиг. 3. Отделните стъпки от нея са описани по-подробно в следващите секции.

2.1.1 Молекулно-механично силово поле

За да се реши уравнението на движение на Нютон, трябва да са известни силите, действащи на всеки атом. Силите са отрицателният градиент на потенциалната енергия. Затова изборът на потенциална функция е от изключително значение за МД симулациите. За симулации на биологични молекули се използват полу-емпирични СП, които обикновено се параметризират с помощта на експериментални данни и квантово-механични симулации на малки моделни съединения. Днешното поколение силови полета представляват разумен компромис между точност и изчислителна ефективност и имат доказана способност да възпроизвеждат физични свойства, които могат да бъдат измерени чрез експеримент,



Фигура 3: Схема на МД симулация

например чрез рентгенова кристалография, ЯМР спектроскопия и нееластично разсейване на неутрони [37].

В биомолекулните СП атомите са масивни частици с Ван дер Ваалсов радиус и частичен заряд (заряд, по-малък от заряда на електрона), а потенциалната енергия се представя като сума от прости на вид дву-, три- и четиричастични взаимодействия. Има два типа приноси към потенциалната енергия – ковалентни взаимодействия, дължащи се на химичните връзки между атомите, и нековалентни взаимодействия. Типичният вид на потенциалната функция е:

$$U(\{\mathbf{r}_i\}) = \sum_{bonds} K_l (l - l_0)^2 + \sum_{angle} K_\theta (\theta - \theta_0)^2 + \sum_{dihedrals} K_\phi (1 + \cos(n\phi - \delta)) + \sum_{i=1}^N \sum_{j=i+1}^N \left(4\varepsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] + \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}} \right) \quad (1)$$

където K_l , K_θ , K_ϕ са силовите константи съответно на химичната връзка, на валентния ъгъл и на ъгъла на усукване; l , θ и ϕ са съответно дължината на химичната връзка, валентния ъгъл и торсионния ъгъл, като с нулев индекс са означени съответните равновесни стойности на отделните променливи; n и δ са мултиплетността и фазовото отместване на ъгъла на усукване; ε_{ij} е дълбочината на ямата на потенциала на Ленард-Джоунс, σ_{ij} е разстоянието на което този потенциал достига нулева стойност, q_i е частичният заряд на i -тия атом, а r_{ij} е разстоянието между атоми i и j .

Първият член в (1) е хармоничен потенциал, моделиращ вътрешно-молекулното взаимодействие на двойки атоми, свързани с ковалентна химична връзка като функция на отклонението на дължината на връзката от идеалната дължина l_0 . Силовата константа K_l характеризира еластичната сила и е специфична за всяка двойка атоми.

Вторият член на потенциала също е хармонична функция, описваща вибрациите около равновесната стойност θ_0 на ъгъла между три атома, свързани с две ковалентни връзки. K_θ и θ_0 зависят от химичната природа на атомите, участващи във взаимодействието и се параметризират по същия начин.

Третият член описва енергията от ротация с порядък на симетрия $n=1,2,3$ около средната връзка на четири последователно свързани атома. Този потенциал трябва да е периодичен, поради което се изразява като косинусова функция.

Нековалентните взаимодействия представляват последните два члена на силовото поле и описват енергията на системата, която не се дължи на химични връзки – Ван дер Ваалсово и електростатично взаимодействия. Първото обикновено се моделира с потенциал на Ленард-Джоунс, а за пресмятането на последното в симулациите на биологични макромолекули най-често се използват методи, производни на сумите на Евалд (секция 2.1.2).

Важно е да се отбележи, че силовото поле представлява не само аналитичната функция на потенциалната енергия (1). То включва всички параметри, които участват в нея за всички т. нар. “типове атоми”, поддържани от полето. Преди започването на каквато и да е симулация, трябва да се конструира топология, която да описва системата в термини на типове атоми, заряди, дължини на химичните връзки и др. Тя е специфична за всяко конкретно силово поле.

2.1.2 Електростатично взаимодействие

Електростатичното взаимодействие е далекодействащо и бавно сходящо, което не позволява просто обрязване на потенциала както при Ван дер Ваалсовото взаимодействие. По тази причина в МД симулациите за пресмятане на електростатиката основно се ползват методи,

базиран на сумите на Евалд – техника, разработена за пресмятане на електростатичното взаимодействие в периодични и псевдо-периодични системи. Идеята е бавно сходящата сума в последния член на уравнение (1) да се преобразува в две бързо сходящи суми, които да се обрежат, като се разделят бързо вариращата част на потенциала за малки r и гладката част за големи r .

Първо всеки заряд q_i ефективно се неутрализира чрез добавяне на сферичен гаусов облак с противоположен заряд с център в q_i . Сумирането на точкови и гаусови заряди се извършва в реалното пространство и представлява директната сума на Евалд, която сега е с по-къс радиус на действие. След това се налага втори набор от гаусови заряди със същия знак като на точковите, които отново са центрирани в q_i , за да компенсират ефекта на първия набор гаусови заряди. Потенциалът, който се създава от тези заряди, може да се пресметне от уравнението на Поасон и се решава чрез Фурие преобразование в реципрочното пространство – реципрочна сума на Евалд. Пълната електростатична енергия съдържа и корекция, която да елиминира енергията, породена от взаимодействието на гаусиана сам със себе си. Така методът на Евалд замества една (при периодични гранични условия безкрайна) бавно сходяща сума с две бързо сходящи суми, в реалното и реципрочното пространство, и корекция за самовзаимодействието:

$$\begin{aligned}
U_{el.stat.} &= U_{dir} + U_{rec} + U_0 \\
U_{dir} &= \frac{1}{4\pi\epsilon_0} \frac{1}{2} \sum_{\mathbf{n}} \sum_{i,j}^N \frac{q_i q_j}{|\mathbf{r}_{ij} + \mathbf{n}L|} \operatorname{erfc}(\beta |\mathbf{r}_{ij} + \mathbf{n}L|) \\
U_{rec} &= \frac{1}{2V\epsilon_0} \sum_{\mathbf{k} \neq 0} \sum_{i,j}^N \frac{q_i q_j}{k^2} \exp\left(\frac{-k^2}{4\beta^2}\right) \exp(-i\mathbf{k} \cdot \mathbf{r}_{ij}) \\
U_0 &= -\frac{1}{4\pi\epsilon_0} \frac{\beta}{\sqrt{\pi}} \sum_i^N q_i^2
\end{aligned} \tag{2}$$

където \mathbf{n} е векторът на кутията в реалното пространство, а \mathbf{k} – вълновият вектор в реципрочното пространство и $\operatorname{erfc}(x) = \frac{2}{\sqrt{\pi}} \int_x^\infty \exp(-t^2) dt$.

Методът PME (Particle Mesh Ewald) [38] подобрява ефективността на изчисление на реципрочната сума. Вместо директно да се сумират вълнови вектори, гаусовите заряди се разпределят по решетка, като се използва B-сплайн интерполация. След това се прави Фурие образ на решетката с помощта на алгоритъм за бързо 3D Фурие преобразование (FFT) и реципрочната сума се получава като единична сума по решетката в k-пространството. Потенциалът във възлите на решетката се смята като обратно преобразование и от коефициентите на интерполацията се получават силите, действащи върху всеки атом. Алгоритъмът PME изисква $\mathcal{O}(N \log N)$ пресмятания, което го прави предпочитан за симулации на средни и големи молекулни системи.

2.1.3 Интегратори

Уравненията на движение на частиците се дискретизират на малки времеви стъпки и се решават числено. Един от най-популярните методи за интегриране на уравненията на движение е интеграторът прескочи-кобила, при който пресмятането на координатите и скоростите

се разминава с половин стъпка:

$$\begin{aligned}
\mathbf{v}_i^{n+1/2} &= \mathbf{v}_i^{n-1/2} + \frac{\mathbf{F}_i^n}{m_i} \Delta t + O(\Delta t^3) \\
\mathbf{r}_i^{n+1} &= \mathbf{r}_i^n + \mathbf{v}_i^{n+1/2} \Delta t + O(\Delta t^3) \\
\mathbf{v}_i^n &= \frac{\mathbf{v}_i^{n-1/2} + \mathbf{v}_i^{n+1/2}}{2}
\end{aligned} \tag{3}$$

Алтернативата е интеграторът на Верле със скорости, при който координати и скорости се обновяват в един и същи момент от време:

$$\begin{aligned}
\mathbf{v}_i^{n+1/2} &= \mathbf{v}_i^n + \frac{\mathbf{F}_i^n}{2m_i} \Delta t \\
\mathbf{r}_i^{n+1} &= \mathbf{r}_i^n + \mathbf{v}_i^{n+1/2} \Delta t + O(\Delta t^3) \\
\mathbf{v}_i^{n+1} &= \mathbf{v}_i^{n+1/2} + \frac{\mathbf{F}_i^{n+1}}{2m_i} \Delta t + O(\Delta t^3)
\end{aligned} \tag{4}$$

2.1.4 Молекулна динамика с фиксирана дължина на връзките

В молекулно-динамичните симулации времевата стъпка е ограничена от необходимостта да се интегрират всички движения в молекулата, дори и тези с най-високата честота – вибрациите на ковалентни връзки, особено включващи водородни атоми, които ограничават $\Delta t < 0.5\text{-}1.0$ fs. При големите биологични молекули обикновено вибрационните степени на свобода не са свързани с транслационните и ротационните. Освен това те рядко представляват интерес. Ето защо често в МД симулациите вибрационните степени на свобода се “замразяват” и химичните връзки в молекулите се разглеждат като “твърди”, т.е. са с фиксирана дължина чрез въвеждане на холономни връзки в Лагранжиана на системата.

Най-разпространеният метод за налагане на връзки в МД симулациите е алгоритъмът SHAKE [39], при който всяка връзка се разглежда отделно, уравненията на връзките се третират последователно и силите на реакция на връзките се изчисляват в итеративна процедура, докато не се удовлетворят всички връзки в рамките на зададен толеранс. Когато за интегриране на уравнението на движение се използва метода на Верле със скорости, скоростите на всеки атом също трябва да се коригират с аналогичен на SHAKE алгоритъм, наречен RATTLE [40]. SHAKE и RATTLE схождат много бързо, типично в две до три итерации. Използването на химични връзки с фиксирана дължина и алгоритмите SHAKE и RATTLE позволяват времевата стъпка, с която интегрираме уравнението на движение, да се увеличи 2-4 пъти.

2.1.5 Интегратори с променлива времева стъпка

Друг начин да се ускори пресмятането на еволюцията на системата е да се разделят взаимодействията на бързо и бавно изменящи се. След това се използва т.нар. интегратор с променлива времева стъпка, който интегрира бавните степени на свобода с по-голяма времева стъпка. Най-разпространеният такъв интегратор r-RESPA [41]. Той може да се използва в комбинация със замразяване на вибрационните степени на свобода, което позволява $\delta t = 2\text{fs}$. Основният му недостатък е, че води до нелинейни нестабилности [42]. R-RESPA може да се запише в следния псевдокод:

Algorithm 1 r-RESPA

```
1:  $\mathbf{v}_i = \mathbf{v}_i + \frac{\Delta t}{2m_i} \mathbf{F}_{1i}$ 
2: for  $k = 1, \dots, n$  do
3:    $\mathbf{v}_i = \mathbf{v}_i + \frac{\delta t}{2m_i} \mathbf{F}_{si}$ 
4:    $\mathbf{r}_i = \mathbf{r}_i + \delta t \mathbf{v}_i$ 
5:    $\mathbf{F}_{si} = \text{update\_F}_s(\mathbf{r}_i)$ 
6:    $\mathbf{v}_i = \mathbf{v}_i + \frac{\delta t}{2m_i} \mathbf{F}_{si}$ 
7:  $\mathbf{F}_{1i} = \text{update\_F}_1(\mathbf{r}_i)$ 
8:  $\mathbf{v}_i = \mathbf{v}_i + \frac{\Delta t}{2m_i} \mathbf{F}_{1i}$ 
```

Вместо него може да се използва надстройка на този алгоритъм, MOLLY (Mollified impulse method) [43]. При него за обновяването на далечните (бавните) сили се използват усреднени по времето координатите, генерирани от референтния пропагатор. Ако усредняващата функция е A , в псевдокод MOLLY изглежда така:

Algorithm 2 MOLLY

```
1:  $\mathbf{v}_i = \mathbf{v}_i + \frac{\Delta t}{2m_i} \mathbf{F}_{1i}$ 
2: for  $k = 1, \dots, n$  do
3:    $\mathbf{v}_i = \mathbf{v}_i + \frac{\delta t}{2m_i} \mathbf{F}_{si}$ 
4:    $\mathbf{r}_i = \mathbf{r}_i + \delta t \mathbf{v}_i$ 
5:    $\mathbf{F}_{si} = \text{update\_F}_s(\mathbf{r}_i)$ 
6:    $\mathbf{v}_i = \mathbf{v}_i + \frac{\delta t}{2m_i} \mathbf{F}_{si}$ 
7:  $\bar{\mathbf{r}}_i = A(\mathbf{r}_i^k)$ 
8:  $\mathbf{F}_{1i} = \text{update\_F}_1(\bar{\mathbf{r}}_i)$ 
9:  $\mathbf{v}_i = \mathbf{v}_i + \frac{\Delta t}{2m_i} \mathbf{F}_{1i}$ 
```

Интеграторът MOLLY не се използва в комбинация с твърди връзки, тъй като алгоритъмът на практика усреднява вибрациите.

2.2 Метадинамика

Метадинамиката е мощен метод за изследване на свойствата на многомерни повърхнини на свободната енергия на сложни многочастични системи чрез не-Марковска динамика в пространство, дефинирано от няколко колективни променливи [45]. При този метод еволюцията на системата се повлиява от зависещ от времето външен потенциал, конструиран като сума от гаусиани, центрирани по траекторията на подходящо избран набор от колективни променливи. Това насърчава системата да изследва области от ограниченото фазово пространство, които преди това не са били посещавани и ѝ пречи да се връща към вече посетени области. Така на системата се позволява да излезе от минимума по път с ниска свободна енергия и да изследва съседни минимума в повърхнината на свободната енергия. Сумата от гаусианите се използва за итеративно реконструиране на профила на свободната енергия в пространството на избраните КП.

Нека имаме система, описвана от набор от координати на частиците $\{\mathbf{r}_i\} = \mathbf{r}$, еволюираща под действие на потенциала $V(\mathbf{r})$. Интересуваме се от изследване на свойствата на системата като функция на краен брой колективни променливи $s_\alpha(\mathbf{r})$, $\alpha = 1, \dots, d$, където d е малко число. Предполага се, че КП дават добро обобщено описание на системата. Те

трябва да са явни функции на координатите, например разстояния, ъгли, координационни числа или потенциалната енергия. Свободната енергия на системата $F(s)$ като функция на КП е:

$$F(s) = -k_B T \ln \left(\int d\mathbf{r} e^{-\frac{1}{k_B T} V(\mathbf{r})} \delta(s - s(\mathbf{r})) \right) \quad (5)$$

Метадинамиката добавя външен потенциал, действащ на системата в момента t :

$$V_G(s(\mathbf{r}), t) = \omega \sum_{\substack{t'=\tau_G, 2\tau_G, \dots \\ t' < t}} \exp \left\{ - \sum_{\alpha=1}^d \frac{(s_\alpha(\mathbf{r}) - s_\alpha(t'))^2}{2\delta s_\alpha^2} \right\} \quad (6)$$

където d на броят на КП, а $s_\alpha(t) = S_\alpha(\mathbf{r}(t))$ е стойността на α -та КП в момента t . Целта на този допълнителен потенциал е да извади системата от локалния минимум и да ѝ позволи да изследва области от повърхнината на свободната енергия, които иначе системата би посетила с много малка вероятност.

След достатъчно дълъг период от време $V_G(s, t)$ може да се използва за статистическа оценка на повърхнината на свободната енергия в пространството на КП:

$$\lim_{t \rightarrow \infty} V_G(s, t) = -F(s) + C \quad (7)$$

В дефиницията на V_G влизат следните три параметъра:

- височината на гаусианите ω
- ширината на гаусианите δs
- периода τ_G , на който се добавят гаусианите.

Те оказват влияние върху точността и производителността на реконструкцията на профила на свободната енергия (ПСЕ). Ако гаусианите са широки и високи, повърхнината на свободната енергия ще се обходи бързо, но няма да могат да се реконструират черти на ПСЕ, които са по-малки от δs . От друга страна, ако гаусианите са малки и се добавят по-рядко, реконструкцията ще бъде по-точна, но ще изисква много дълго време. Затова параметрите ω и δs се избират така, че да се намери баланс между точност и ефективност на обхождането. Добавянето на Гаусиани с голям обем за кратко време може да доведе до напълно грешна реконструирана ПСЕ [46].

Идеалният набор от КП трябва да отговаря на следните условия:

- да разграничава ясно началното и крайното състояние на системата (евентуално и междинните състояния);
- да описва всички най-бавни събития, които имат отношение към изследвания процес;
- да се състои от възможно най-малък брой КП, типично до 3.

Глава 3

Анализ на скалируемостта на алгоритмите за едромащабни симулации на биологични макромолекули

МД симулациите на големи системи обикновено се пресмятат паралелно и изискват високопроизводителни машини, както и софтуер, който е в състояние да използва ефективно предоставените ресурси. Системите, които се изследват с МД, непрекъснато нарастват на размер и става очевидно, че програмните пакети не успяват да се справят с нарастващия обем на изчисленията. Затова е необходимо да се изследва скалируемостта на алгоритмите, които се използват, както и да се идентифицират проблематичните моменти в симулациите.

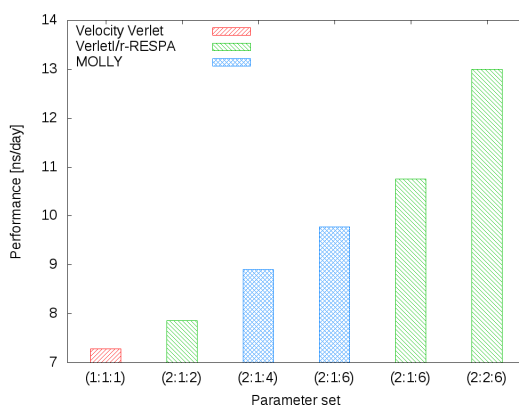
Всички представени симулации са направени на суперкомпютъра IBM BlueGene/P към Националния център за суперкомпютърни приложения [47] на различен брой изчислителни ядра. Молекулно-динамичните пакети, чиято производителност беше изследвана и профилирана, са GROMACS 4.5.3 [48] и NAMD [49] версия CVS 2011-02-19.

3.1 Интегратори с променлива стъпка на интегриране в пакета NAMD

Пакетът GROMACS не предлага интегратор с променлива стъпка на интегриране. За разлика от него, в пакета NAMD е имплементиран такъв алгоритъм – Verlet-I/r-RESPA [41]. Тъй като при него се проявяват линейни и нелинейни нестабилности [42], в NAMD е имплементирана и надстройка на този алгоритъм, MOLLY (Mollified impulse method) [43]. Потребителят може да задава стойности на три параметъра, определящи структурата на интеграционния алгоритъм – `timestep`, `nonbondedFreq` и `fullElectFrequency`. Това са съответно времевата стъпка, честотата на пресмятане (в брой стъпки) на близките нековалентни взаимодействия с малък радиус и честотата на пресмятане на далечните взаимодействия. Например при набор от стойности $\{1,1,1\}$, всички взаимодействия се изчисляват на всяка стъпка, която е една фемтосекунда. Изборът `nonbondedFreq=1` и `fullElectFrequency=1` съответства на стандартния интегратор на Верле със скорости. При стойности $\{2,1,2\}$ късододействащите нековалентни взаимодействия се пресмятат на всяка стъпка, а далекодействащите нековалентни на всеки две стъпки, което при базова стъпка 2 fs означава на всеки 2 fs и 4 fs съответно.

Беше изследвана зависимостта на производителността на пакета NAMD (в ns/day) при

система с около 35 хиляди атома (човешки гама-интерферон във вода) от броя на използваните изчислителни ядра. Получените резултати са представени на фиг. 4. За получаването на първата точка е използван стандартния алгоритъм на Верле със скорости с интеграционна стъпка от 1 fs, а за втората – алгоритъм с променлива стъпка със стандартни параметри, с които се смята в NAMD. Повечето симулации се правят именно с тези параметри. За точка 6 е използван Verlet-I/r-RESPA интегратор, като симулация с параметри {2,2,6} с интегратора MOLLY прекъсва с грешка на 4500-та стъпка. Всички симулации са проведени за 20 000 стъпки. С параметрите {2,1,6} и {2,2,6}, допълнително са проведени и дълги симулации (2 000 000 стъпки, т.е. 40 ns), първата с интегратора MOLLY, втората – с алгоритъм Verlet-I/r-RESPA. Като правило, симулациите с Verlet-I/r-RESPA прекъсват със съобщение за грешка в алгоритъма RATTLE (секция 2.1.4), който коригира скоростите на частиците при фиксиране на дължината на химичните връзки. Това означава, че алгоритъмът Verlet-I/r-RESPA води до нестабилност на траекторията. Интересно е, че симулациите с параметри {2,1,6} и {2,2,6} с интегратор MOLLY също прекъсват с грешка, при това по-рано от тези с Verlet-I/r-RESPA.



Фигура 4: Производителност на NAMD при използването на различни алгоритми за интегриране и различни параметри.

Използването на алгоритъма с променлива стъпка Verlet-I/r-RESPA, с пресмятане на далекодействащи електростатични взаимодействия на всяка 6-та стъпка, води до 1.5-кратно ускоряване на симулацията в сравнение със стандартния алгоритъм. При това пълната енергия на системата на практика не се изменя. Трябва да се отбележи, обаче, че стабилността на алгоритъма е ограничена във времето, най-вероятно в зависимост от размера и структурата на симулираната система.

3.2 Пространствено разделяне на системата и пресмятане на електростатичните взаимодействия в пакета GROMACS

При провеждане на паралелна симулация на молекулна система е необходим алгоритъм, който да раздели системата на части, които да се разпределят по отделните процесорите. В GROMACS има два алгоритъма за разделяне на системата – разделяне по частици (particle decomposition) и пространствено разделяне (разделяне по домени – domain decomposition) [48].

Разделянето по домени се основава на факта, че повечето взаимодействия в системата са локални. Ако се използва този алгоритъм, трябва да се реши как ще се раздели симулационния обем на домени и как ще се разпредели изчислението на силите по проце-

сорите. В общия случай на клиногнална симулационна кутия, пространството се разделя на 1-, 2- или 3-мерна решетка от паралелепипеди, които се наричат домени. При този алгоритъм на всеки процесор се назначава пространствен домен от системата. Процесорът решава уравненията на движение за частиците, които в момента се намират в съответния домен. Системата се разделя на домени отново и се преразпределя по процесорите на всяка молекулно-динамична стъпка, при която се обновяват списъците с най-близките съседи на частиците (обикновено на 10 времеви стъпки).

Електростатичното взаимодействие е далекодействащо, а не локално. Затова за пресмятането му се използват специални алгоритми. В GROMACS обикновено се използва алгоритъмът PME [38], при който всяка частица взаимодейства с всички останали частици, което изисква глобална комуникация. За да се ограничи ефекта на този проблем, част от изчислителните ядра се използват само за изчисления на електростатичното взаимодействие с алгоритъма PME (наречени *pme* ядра), а останалите (*pp* ядра) пресмятат всички останали взаимодействия [48].

По подразбиране GROMACS разделя *pp:pme* изчислителните ядра в съотношение 3:1 и ги подрежда в режим *interleave*. В този режим ако съотношението *pp:pme* ядра е $n:1$, след всеки n *pp* ядра, следва 1 *pme* ядро. GROMACS предлага и други два режима за разпределяне на *pp:pme* изчислителните ядра. В режима *pp_pme* първо се нареждат *pp* ядрата и накрая следват *pme* ядрата. Режимът *cartesian* е комбинация от предните два и е специално създаден за изчислителни машини с реална тримерна тороидална комуникационна система, като BlueGene.

За изследване зависимостта на производителността на GROMACS 4.5.3 от режима на подреждане на *pp/pme* изчислителните ядра беше използван пакета за профилиране Scalasca 1.3.1 [50]. Тестовата система се състои от 103079 атома, симулацията е с продължителност 10000 стъпки (20ps, времева стъпка е 2 fs). Системата е с периодични гранични условия и се симулира при постоянна температура, като се използва термостатът на Берендсен. Няма наложени холономни връзки за замразяване на вибрационни степени на свобода. За електростатичното взаимодействие се използва алгоритъма PME с радиус на обрязване на директната сума 1.4 nm. Списъците със съседи се обновяват на всеки 10 стъпки.

Профилирането със SCALASCA показва, че разделянето на изчислителните ядра на *pp* и *pme* води до много неравномерно разпределение на комуникацията. Около 40% от комуникацията се пада на $\frac{1}{4}$ от процесорите. Това се дължи на факта, че електростатичното взаимодействие е далекодействащо и *pme* изчислителните ядра трябва първо да получат координатите и зарядите на частиците, след това да извършат PME изчисленията и накрая да изпратят на *pp* ядрата силите, действащи на частиците.

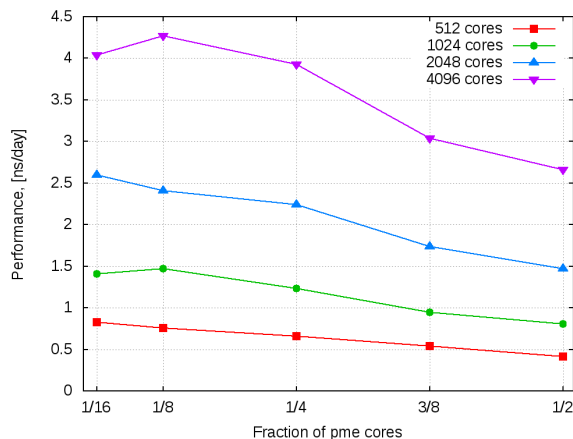
Резултатите за производителността (табл. 1) показват, че за система от около 100000 атома, режимът по подразбиране на подреждане на *pme* ядрата *interleave* има известно предимство пред останалите режими до 2048 ядра. При 4096 изчислителни ядра, обаче, той е най-нископроизводителен. Най-добри резултати се наблюдават при режим на подреждане на *pp/pme* ядрата *cartesian*.

Увеличаването на броя на изчислителните възли, които да пресмятат само далекодействащите електростатични сили, би ускорило тяхното пресмятане, но на цената на по-малко на брой изчислителни възли за процедури, които поне 65% от общото време за пресмятане. Освен това с нарастване на *pme* ядрата, нараства и общата комуникация, което може да доведе до значително забавяне на изчислението, особено ако това внася необходимост от допълнителни синхронизиращи операции.

Това наложи да се изследва зависимостта на производителността от отношението на *pme* към *pp* изчислителни възли. За да се изключи дори и минимално влияние от страна на инструментацията пакет, изследването беше проведено на базата на симулации с неин-

Таблица 1: Зависимост на производителността на GROMACS 4.5.3 от режима на подреждане на pp/pme ядрата при повишаване на броя на изчислителните ядра.

# изчислителни ядра	interleave [ns/day]	pp_pme [ns/day]	cartesian [ns/day]
512	6.67	6.59	6.60
1024	12.12	11.90	11.97
2048	20.86	20.63	20.43
4096	27.99	31.31	31.54



Фигура 5: Производителност на GROMACS в зависимост от дела на pme ядра при различен брой изчислителни ядра.

струментиран GROMACS 4.5.3, като данните за производителността за взети направо от изходните log-файлове.

На фиг. 5 е показана зависимостта на производителността на GROMACS (в ns/day) като функция на дела на pme ядрата при различен брой изчислителни ядра. Вижда се, че с намаляване на частта на pme ядрата производителността нараства почти монотонно с изключение на кривите при 1024 и 4096 ядра, които имат слабо изразен максимум при 1/8 pme ядра, независимо от нарастването на броя комуникации или други фактори. Понататъшно намаляване на броя на pme ядрата – например до 1/32 – не е възможно.

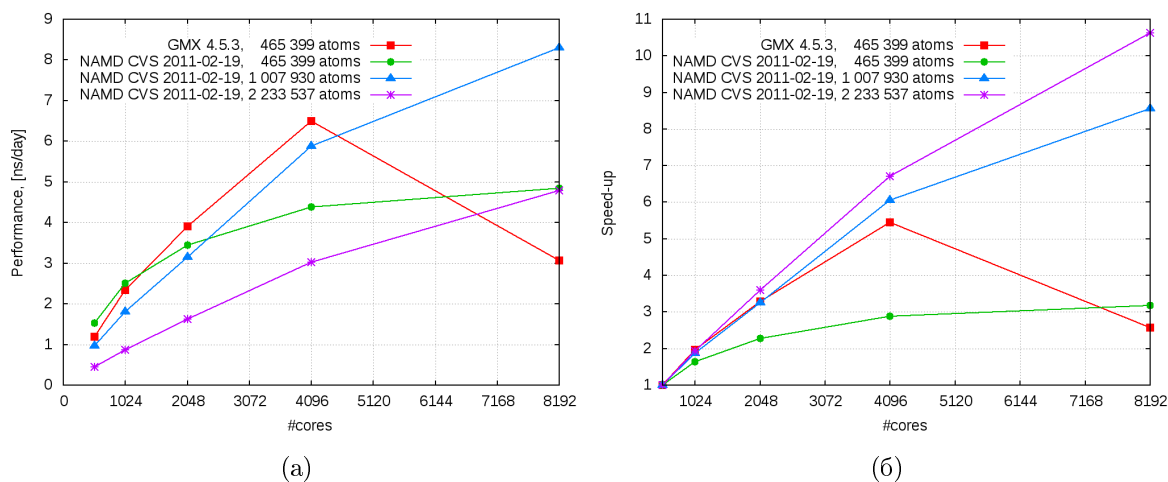
3.3 Зависимост на производителността от размера на симулираната системата

За изследване на производителността в зависимост от големината на симулираната система бяха използвани три тестови системи, съдържащи съответно 465 399 атома, 1 007 930 атома и 2 233 537 атома.

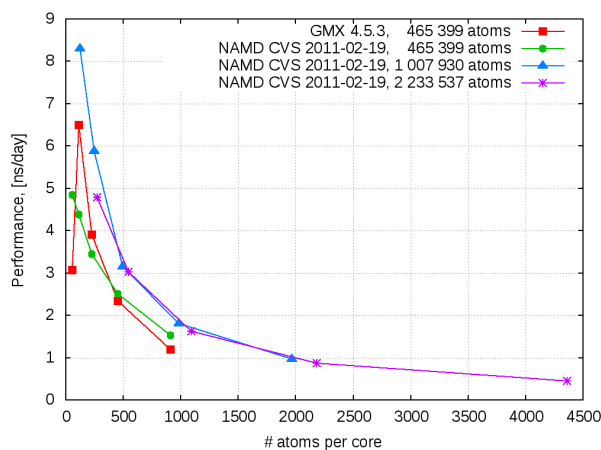
Изследването беше проведено с пакетите GROMACS 4.5.3 и NAMD CVS 2011-02-19. И двата пакета са компилирани с XL компилаторите на IBM за архитектурата на изчислителните възли на BlueGene/P. NAMD CVS компилацията позволява да се включи възможността да се използват компресирани входни данни. Изчислителните възли на IBM BlueGene/P разполагат с 2GB памет за 4 ядра, т.е. по 512MB на ядро. Оказва се, че това не е достатъчно за първоначално зареждане на координатите на двете по-големи тестови системи. Затова скалируемостта на пакета GROMACS 4.5.3 беше изследвана само за тестова система

с половин милион частици.

Резултатите са представени на фиг. 6а. Данните за производителността са взети директно от изходните файлове на симулациите, за да се елиминират артефакти от профилирането. Интеграционната стъпка и в двата случая е 2 fs. Скалируемостта на NAMD се подобрява с увеличаването на размера на системата, въпреки че се забавя след 4096 ядра (фиг. 6б). Производителността на GROMACS рязко спада след 4096 ядра за малката система. Това вероятно се дължи на рязкото увеличаване на комуникациите. Изобщо, производителността на GROMACS се влошава драстично, когато средният брой на атомите, които се падат на едно изчислително ядро падне до около и под 100, защото времето за осъществяване на комуникациите вече доминира времето за изчисление. Такова поведение не се забелязва при пакета NAMD (фиг. 7).



Фигура 6: Производителност (а) и ускорение (б) на пакетите GROMACS 4.5.3 и NAMD CVS 2011-02-19 на архитектурата IBM BlueGene/P.



Фигура 7: Зависимост на производителността на пакетите GROMACS 4.5.3 и NAMD CVS 2011-02-19 от средния брой атоми на изчислително ядро.

Глава 4

Изследване на стабилността на локалната структура на мутагенни форми на hIFN- γ

4.1 Цел

Анормалната експресия на IFN- γ се свързва с множество автовъзпалителни и автоимунни заболявания. Основната идея за контролиране на такива състояния е да се блокира или молекулата на цитокина, или на неговия рецептор. Това може да се постигне чрез намирането на подходяща мутагенна форма на hIFN- γ , която да е способна да се свързва с рецептора, но да не притежава биологична активност. IFN- γ се свързва с hIFN γ R α предимно чрез N-края си. Ето защо мутации на остатъци, които не участват в активния център на молекулата, отговорен за разпознаването на рецептора, не би следвало да повлияват свързването, освен ако не предизвикват значителни изменения в структурата на белтъка. Тъй като IFN- γ предизвиква биологичните си въздействия чрез активиране на Jak/STAT сигнализацията, за блокиране на биологичната активност на протеина е разумно да се мутират остатъци, които участват в сигналния път – аминокиселини от NLS веригите.

Колегите от Института по молекулярна биология на БАН идентифицираха като най-подходящи за мутиране аминокиселини от горната NLS последователност – остатъци Lys⁸⁶ – Lys – Lys⁸⁸, тъй като NLS веригата в C-края на IFN- γ има съществен принос във взаимодействието с рецептора. Те ни предоставиха списък със 100 мутагенни форми на hIFN- γ , в които тези три остатъка са заменени с други. Целта беше да бъдат определени мутантите, които запазват максимално локалната вторична структура на спирала E.

4.2 Анализ на вторичната структура на белтъци

4.2.1 Диаграма на Рамачандран

Вторичната структура на полипептид или протеин може да се характеризира чрез ъглите на усукване на гръбнака, които описват ротацията на равнините на две съседни пептидни връзки една около друга. Използват се два ъгъла, които дефинират наклона на две съседни амидни равнини (равнината на пептидната връзка), като C'_α е центърът на ротация. Това са ъгълът ϕ , включващ карбоксилния въглерод C на предишния а.к. остатък, азотът от амидната група N', алфа въглеродния атом C'_α и карбоксилния въглерод C'; и ψ , включващ атоми N' – C'_α – C' – N''.

Позволените конформации на гръбнака на пептид могат да се изчислят от ротации около връзките, дефинирани от ъглите на усукване ϕ и ψ , и стеричните ограничения на страничните вериги, определени от Ван дер Ваалсовите им радиуси. Получената конформационна карта се нарича диаграма на Рамачандран [51, 52]. Поради възникването на голям брой водородни връзки между близко разположени пептидни групи и поради пространствено препречване на страничните вериги, ъглите ϕ и ψ не могат да имат произволно значение между -180° и $+180^\circ$. Двойките ъгли ϕ и ψ се струпват в т.нар. разрешени области, като повтарящи се последователно ϕ и ψ ъгли винаги водят до някакъв тип подредена структура, стига тя да е стерично позволена. Областите в диаграмата на Рамачандран съответстват на α спирали, леви α -спирали и разгънати (β) конформации.

4.2.2 DSSP

Кабш и Зандер разработват по-прецизен метод за разпознаване на мотиви и извличане на структурни характеристики от координатите на атомите на протеини, наречен DSSP (dictionary of protein secondary structure) [53]. Алгоритъмът се базира основно на откриването на мотиви от водородни връзки. Дефиницията на структури, базирани на Н-връзки, е йерархична. Първо се дефинират Н-връзки, а въз основа на тях се разпознават “п-врътки”, които имат Н-връзка между C=O групата на аминокиселинен остатък i и N-H групата на остатък $i+n$, където $n=3,4,5$, и “мостове”, при които два непоследователни във веригата остатъка са свързани с Н-връзка. Последователни 3-врътки образуват 3_{10} -спирали, 4-врътки дефинират α -спирали, 5-врътки – π -спирали, а повтарящи се мостове образуват β -структури. Всяка структурна характеристика се дефинира независимо от останалите и припокриването им се отстранява като накрая се прави обобщение на вторичната структура и се назначава едно единствено състояние на всеки аминокиселинен остатък.

В основата на метода е дефиницията на Н-връзка. “Добрата” водородна връзка има енергия от около -3 kcal/mol. За компенсирание на разклонени Н-връзки и експерименталните грешки в координатите, прага за обрязване по енергии е по-висок. Приема се, че Н-връзка между карбонилната (C=O) група на остатък i и amino групата (N-H) на остатък j съществува, ако енергията на взаимодействие е по-малка от -0.5 kcal/mol. Електростатичната енергия на взаимодействие между двете групи се изчислява като на C,O атомите се приписват частични заряди $(+q_1, -q_1)$, а на N,H атомите – $(-q_2, +q_2)$:

$$E = q_1 q_1 \left[\frac{1}{r(ON)} + \frac{1}{r(CH)} - \frac{1}{r(OH)} - \frac{1}{r(CN)} \right] f, \quad (8)$$

където $q_1 = 0.42e$ и $q_2 = 0.20e$, e е елементарния заряд на електрона, $r(AB)$ е разстоянието между атоми A и B в Å, $f = 332$ е безразмерен коефициент и E е в kcal/mol.

4.3 Подготовка

Структурите на мутантите бяха създадени с помощта на пакета за визуализация и манипулация на молекули РуМОЛ [54]. Страничните вериги на трите лизина бяха заменени със страничните вериги на трите заместващи аминокиселини, като гръбнакът беше запазен непроменен. По този начин в началото всички мутагенни форми на цитокина имаха α -спирална локална структура на мястото на мутациите.

След това бяха направени кратки МД симулации, за да се релаксират получените структури, както и на референтната структура на IFN- γ от комплекса с PDB ID 1FG9 [9]. Симулациите бяха направени с програмния пакет NAMD 2.6 [49], като беше използвано силовото поле CHARMM 22 с модифициран TIP3P модел на водата [55]. Молекулите на протеините

бяха разтворени в кубични кутии с размери $80 \times 80 \times 80 \text{ \AA}$ (около 48 000 атома) с наложени периодични гранични условия. Симулациите са направени в каноничен ансамбъл, като за поддържане на постоянна температура от 310K бе използван термостат на Ланжевен [44] с коефициент на триене от 5/ps. За пресмятане на електростатиката беше използван методът PME с радиус на обрязване на правата сума от 12 \AA и минимална константа на решетката за реципрочната сума 0.625 \AA^{-1} . За Ван дер Ваалсовите взаимодействия беше използвана изглаждаща функция, която се включва на радиус от 10 \AA и плавно ги нулира на радиус от 12 \AA . Списъците с най-близки съседни се конструираха на всеки 10 стъпки с радиус на обрязване от 13.5 \AA . Използван беше интегратора с променлива времева стъпка Verlet-I/rRESPA с времева стъпка от 2fs и пресмятане на далечните електростатични взаимодействия на всеки 2 стъпки. Първо беше направена кратка минимизация на енергията от 1000 стъпки по метода на спрегнатите градиенти, последвана от същинската МД симулация с продължителност от 10 ps. Изчисленията бяха направени на 512 изчислителни ядра на системата IBM BlueGene/P към Националния център за суперкомпютърни приложения [47].

4.4 Метадинамичен модел

Крайните конформации от МД симулациите бяха използвани за вход на метадинамични симулации за оценка на стабилността на крайната част на α -спирала E. За колективни променливи бяха избрани ъглите на усукване на гръбнака на аминокиселината на позиция 86. Целта беше да се оцени големината на бариера, който разделя в диаграмата на Рамачандран областта на α -спиралите от областта на разгънатите конформации за най-външната за спирала E аминокиселина.

Симулациите бяха направени с програмния пакет NAMD 2.7b1 [49]. Бяха използвани същото силово поле, симулационна кутия, параметри на термостата и методи и радиуси на обрязване за нековалентните взаимодействия, както в МД симулациите. За метадинамичните симулации е използван интеграторът на Верле със скорости с времева стъпка от 1fs и пресмятане на далечните електростатични взаимодействия на всяка стъпка. Симулациите бяха с продължителност 0.1 ns, като допълнителните гаусиани бяха добавяни на всеки 10 fs. Височината на гаусианите е 0.1 kcal/mol, а ширината им – 6 °. Изчисленията бяха направени на 512 ядра на системата IBM BlueGene/P към Националния център за суперкомпютърни приложения [47].

4.4.1 Анализ на данните от метадинамични симулации

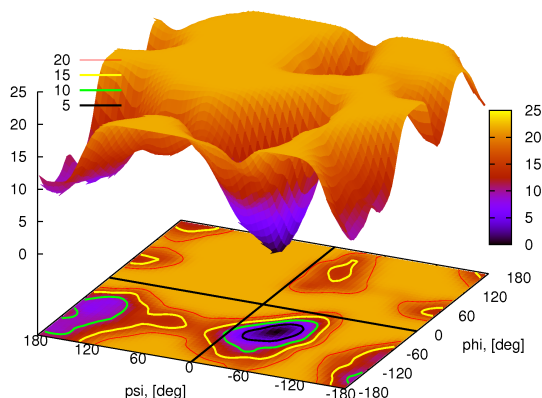
Получените повърхнини на свободната енергия в пространството, дефинирано от торсионните ъгли на гръбнака на аминокиселинен остатък 86, бяха анализирани с помощта на няколко параметъра. Беше пресметната средно-квадратичната разлика в профилите на свободната енергия между референтния профил на нативната форма на hIFN- γ (фиг. 8) и съответния мутант в областта на α -спиралите ($\phi \in [0; 30], \psi \in [-180; 0]$):

$$S = \sqrt{\frac{\sum_i^n (E(\phi_i, \psi_i) - E(\phi_{ref}, \psi_{ref}))^2}{n}} \quad (9)$$

По-малка стойност на S означава, че двата профила съвпадат по-добре в областта на α -спиралите.

Освен това беше оценена големината на бариера в свободната енергия ΔG , разделящ областите на α -спиралите и разгънатите конформации в диаграмата на Рамачандран. Това ни дава оценка за стабилността на вторичната структура, в която участва аминокиселинен остатък на позиция 86 (спирала E) спрямо тази за природния интерферон. За него този

барьер е 16.3 ± 0.2 kcal/mol. Височината на бариерите беше пресметната като разликата в енергиите на най-ниския минимум в областта на α -спиралите и енергията в най-близката, най-ниска седловинна точка, която разделя двете ями - на α -спиралите и на разгънатите конформации.



Фигура 8: Повърхнина на свободната енергия в пространството на ъглите на усукване на гръбнака на аминокиселина Lys⁸⁶ на нативен hIFN- γ .

4.5 Резултати

4.5.1 Молекулно-динамични симулации

За всяка от траекториите бяха изследвани поведението и стабилността на α -спиралата, съдържаща остатъци 86-88, чрез диаграмата на Рамачандран. Когато се наблюдаваше тенденция към прекъсване на спиралата, траекторията на съответния мутант беше изследвана с по-прецизния, но и по-времеемък метод DSSP.

Наблюдаваха се два случая на прекъсване на спиралата. В първия случай аминокиселинен остатък Lys⁸⁸ (най-вътрешния за спиралата) а заменен с пролин или глицин. Тези аминокиселини са известни с като “терминатори на спирали”, тъй като обикновено се намират в краищата на спиралите. Измененията във вторичната структура при такива мутанти бяха наблюдавани както на диаграмата на Рамачандран, така и с DSSP.

Във втория случай на позиция 88 няма “терминатор”, но въпреки това спиралата се прекъсва. В тези случаи конформационните изменения не настъпват веднага, а постепенно след средата на МД симулацията.

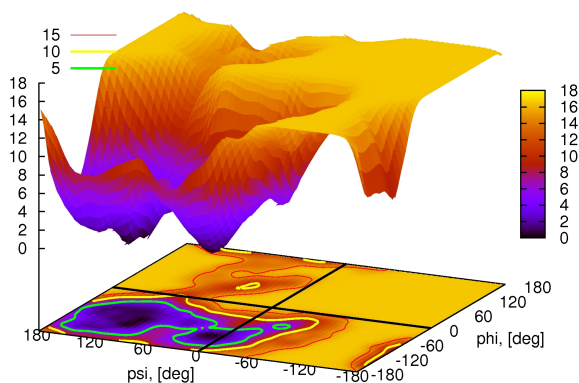
При мутациите, които не предизвикват изменения във вторичната структура, аминокиселинни остатъци 86-88 остават в по-строга дефинираната зона на α -спиралите в диаграмата на Рамачандран през цялата траектория.

4.5.2 Метадинамични симулации

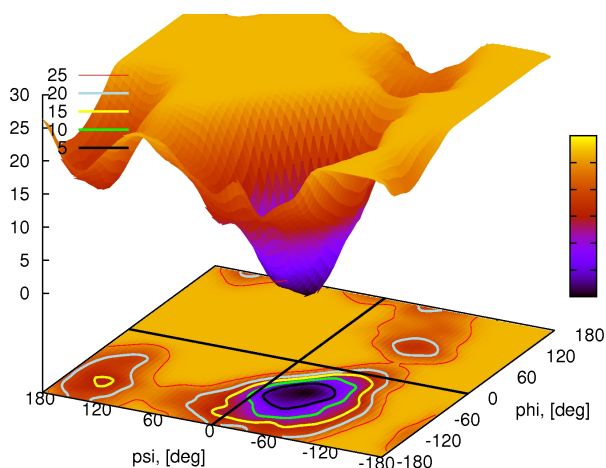
Целта на метадинамичните симулации беше от 100-те изследвани мутанта да се отсеят само няколко на брой, които не само да запазват локалната вторична структура, но и тя да остава с подобна или по-добра стабилност от тази на нативната форма на hIFN- γ . Метадинамиката не само дава допълнителна информация, но и предлага обяснение за стабилността на локалната вторична структура, в която участват мутираните аминокиселинни остатъци и евентуалното настъпване на изменения в нея.

При мутанти, които са се разплели още по време на МД симулациите обикновено реконструиранят ПСЕ показва, че Gln⁸⁶ изобщо не участва в α -спирала, защото е излязъл от тази област на диаграмата на Рамачандран още по време на МД симулацията. Средно-квадратичното отклонение S на профила на свободната енергия в пространството ϕ, ψ на такива мутагенни форми от този на референтната нативна структура е много висок (> 0.23 kcal/mol).

За разлика от тях, ПСЕ на мутагенни форми като мутант №61 е много близък до този на нативната форма на hIFN- γ , като бариерът, който разделя α -областта от областта на разгънатите конформации, е дори по-висок (24.26 ± 0.2 kcal/mol) от тази на нативната форма (16.3 ± 0.2 kcal/mol). Такива мутагенни форми имат ниски стойности на S (< 0.17 kcal/mol).



Фигура 9: Повърхнина на свободната енергия в пространството на ъглите на усукване на гръбнака на аминокиселина Gln⁸⁶ на мутант №120.



Фигура 10: Повърхнина на свободната енергия в пространството на ъглите на усукване на гръбнака на аминокиселина His⁸⁶ на мутант №61.

Има и такива случаи, при които резултатът от стандартната МД симулация е, че мутацията запазва α -спиралата, но реконструираният с метадинамика профил на повърхнината

на свободната енергия показва, че мутацията води до нестабилна спирала. Такива мутанти имат голям S .

Част от мутагенните форми изглеждат стабилни по време на МД симулацията и имат ниско S при метадинамичната симулация, но бариерът, разделящ областите на α -спиралите и на разгънатите конформации в диаграмата на Рамачандран, е много нисък. Причината спиралата да остава стабилна по време на МД симулацията е краткото симулационно време. За 10 ps системата не е успяла да преодолее тези близо 8 kcal/mol и да приеме нова конформация.

Има няколко мутагенни форми, при които по време на МД симулациите спирала E в единия мономер на интерферона остава стабилна, а в другия се разгъва. Реконструираният профил на свободната енергия дава обяснение на това поведение. Обикновено ямата е доста разлята и обхваща части и от двете области в диаграмата на Рамачандран - на α -спиралите и на разгънатите конформации. Това е причината остатък 86 в едната поединица да навлезе в областта на разгънатите конформации, а в другата поединица да остане в спиралата.

Като запазващи локалната структура на спирала E и като достатъчно стабилни са избрани тези мутагенни форми, които отговарят на следните условия:

- Да имат S , отчитащо сходността на профилите на свободната енергия в пространството на ъглите на усукване ϕ и ψ на аминокиселина на позиция 86 на мутагенната форма и нативната форма на hIFN- γ < 0.155 kcal/mol;
- Големината на бариера ΔG , разделящ областите на α -спиралите и на разгънатите конформации в диаграмата на Рамачандран, да е > 13 kcal/mol.

За всички 100 мутагенни форми параметърът S варира в границите от 0.10 до 0.30 kcal/mol. Радиусът на обрязване по S е избран така, че да се изберат мутагенни форми с максимално подобен на нативния профил на свободната енергия. Радиусът на обрязване по ΔG е 13 kcal/mol, за да се изберат само мутанти, чиято спирала E е с близка или по-голяма стабилност от тази в нативната молекула ($\Delta G^{hIFN\gamma} = 16.3 \pm 0.2$ kcal/mol), като все пак се позволи скъсването на една водородна връзка ($E^{H-bond} \approx 3$ kcal/mol). При този избор на радиуси на обрязване 12 от 100-те мутагенни форми на hIFN- γ удовлетворяват условията, посочени по-горе.

Глава 5

Молекулно моделиране на комплекса hIFN γ -hIFN γ R α -производни на хепарина олигозахариди

Установено е, че IFN- γ взаимодейства с гликозаминогликани и по-конкретно с хепарин и хепаран сулфат, което обяснява интереса към симулирането на това взаимодействие. За тази цел, обаче, са необходими параметри на силовото поле за всички съединения, участващи в изследваната система. Нито едно от известните силови полета не съдържа параметри за монозахарида N-сулфатиран глюкозамин сулфат, който участва в структурата на дизахаридите, изграждащи гликозаминогликана хепарин и NS-областите на хепаран сулфат. Това наложи да се разработи нов набор от параметри, който да описва монозахарида и да е съвместим с конкретно силово поле.

5.1 Разработване на силово поле хепарин

За параметризация на N-сулфатиран глюкозамин сулфат беше избрано полето CHARMM 36 [56]. Файловете с различните набори топологии и параметри могат да се свалят свободно от сайта на групата на Маккеръл [57]. Полето за въглехидрати [58] съдържа набор от параметри за множество различни захариди, гликозидни връзки и функционални групи.

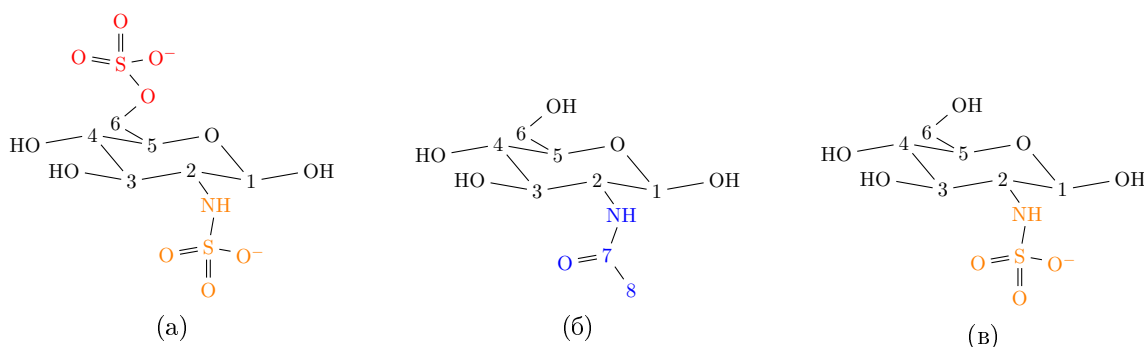
5.1.1 Начални параметри за D-Glc2NS

Въглехидратното поле CHARMM36 предлага топология и параметри за 2-ацетил-2-дезоксид- α -D-глюкозамин (RESI AGLCNA във файла `top_all136_carb.rtf`, фиг. 116) [58]. Освен това има и параметри за O-сулфатната група при C_6 (PRES SUF6) [59].

В литературата няма разработени параметри за сулфатната анионна група NSO_3^- при C_2 (представена в оранжево на фиг. 11a). За параметризацията на тази група беше използвано моделно съединение – 2-N-сулфатиран глюкозамин (фиг. 11b).

Особеност на силовото поле CHARMM е, че то има модулен характер. Това означава, че отделно се параметризират малки моделни съединения и различни функционални групи, които могат да се добавят към тях. Не е необходимо да се параметризират цели сложни съединения, като например N-сулфатиран глюкозамин. Достатъчно е да се параметризира само новата функционална група – атоми C_2 , H_2 , N , HN , S , OS_1 , OS_2 , OS_3 . За всички останали атоми бяха използвани вече готовите топология и параметри от RESI AGLCNA.

За дефиниране на начални параметри за новата функционална група беше използван сървърът Ramachem [60–62]. На него върви програмата CGenFF от пакета за МД



Фигура 11: Различни производни на D-глюкозамин. Ацетилната група е оцветена в синьо, сулфатната – в червено и N-сулфатните – в оранжево. (а) 2-N-сулфатиран-D-глюкозамин-6-сулфат. Това е монозахарида, който участва в основната дизахаридна последователност, която изгражда хепарин и NS-областите на XC; (б) 2-ацетил-2-дезоксид-D-глюкозамин. Този монозахарид участва в NA-областите на XC. Освен това силовото поле CHARMM36 за въглехидрати съдържа топология и параметри за него под запис RESI AGLCNA. Част от тях бяха използвани като начални за моделното съединение; (в) Моделното съединение 2-N-сулфатиран-D-глюкозамин, което беше използвано, за да се разработи топология и параметри за сулфатната анионна група NSO_3^- .

CHARMM [63] за автоматично определяне на типовете атоми и назначаване на параметри и частични заряди по аналогия с вече параметризирани в CGenFF съединения със сходна химична природа. Към всеки назначен параметър се дава наказателна оценка, която трябва да ориентира потребителя за качеството на съответния параметър.

5.1.2 Параметризация на D-Glc2NS

За оптимизиране на параметрите с висока наказателна оценка (> 12) от Paramchem беше използвана добавката Force Field Toolkit [64] (FFTK) на пакета за визуализация и манипулация на молекули VMD [65].

Първата стъпка в параметризационната процедура е определянето на типовете атоми, участващи в новата функционална група и приписването на Ван дер Ваалсови параметри на всеки тип атом. Типовете за атоми C_2 и H_2 бяха запазени същите като в N-ацетилян глюкозамин RESI AGLCNA. За да не се получи объркване с вече готови параметри във файла `par_all36_cgenff.prm`, бяха дефинирани нови типове атоми на всички атоми, изграждащи NSO_3^- групата. Ван дер Ваалсовите параметри на новите типове атоми бяха назначени по аналогия с две съединения, които вече са параметризирани в полето. Новите типове атоми, аналогичните им атоми от CGenFF и съответните Ван дер Ваалсови параметри са представени в табл. 2.

Следващата стъпка е да се оптимизира началната геометрия на молекулата. Началните координати бяха получени от структурата на производен на хепарина додеказахарид dp12 (dp – степен на полимеризация), остатък SGN. Структурата е получена с ЯМР спектроскопия и може да се сваля от базата данни PDB с идентификационен код 1HPN [66]. Ръчно беше изтрита SO_3^- групата при атом O_6 и беше заменена с водороден атом. Освен това ръчно бяха добавени хидроксилна група към атом C_3 и един водороден атом към атом O_1 .

Тази начална структура беше оптимизирана с помощта на квантово-механични симулации с пакета GAUSSIAN 09 [67]. За симулацията беше използван методът на Молер-Плесет от 2-ри ред [68] (MP2) с набор от базисни функции 6-31G* (който описва вътрешните електрони с една слейтърова орбитала, съставена от 6 примитивни гаусови орбитали, а ва-

Нов тип атом	CGenFF тип атом	ϵ , [kcal/mol]	$r_{min}/2$, [\AA]
NNSO3	NG2S3	-0.200	1.8500
HNSO3	HGP1	-0.046	0.2245
SNSO3	SG3O1	-0.470	2.1000
ONSO3	OG2P1	-0.120	1.7000

Таблица 2: Типове атоми в сулфаматната анионна група. В първата колона са дадени новите типове атоми, във втората – аналогичните им типове в полето CGenFF, в трета колона е дадена дълбочината на ямата на Ленард-Джоунс потенциала, а в последната колона – половината от разстоянието до минимума на Ленард-Джоунс потенциала.

лентните електрони с 2 слейтъррови орбитали – една от три примитивни гаусиана и една самостоятелна примитивна GTO. Добавени са и набор от d-тип поляризационни функции за всички тежки атоми в системата).

В СП CHARMM за определяне на частичните заряди на дадена група се набляга на възпроизвеждане на квантово-механични взаимодействия с TIP3P водни молекули [69]. Всеки достъпен за разтворителя атом се добавя в списък с потенциални донори, акцептори или и двете на водородни връзки. За всеки такъв атом (място на взаимодействие) FFTK конструира автоматично комплекс на съединението и водна молекула, като водата е ориентирана в идеална за формиране на водородна връзка геометрия. Началните координати на водата се определят от геометрията на взаимодействието така, че да се минимизира стеричното отблъскване между водната молекула и съседните атоми, свързани ковалентно с мястото на взаимодействие. За всеки комплекс с водна молекула се генерира отделен входен файл за GAUSSIAN. По конвенция в CHARMM алифатните водороди винаги са със заряд $+0.09e$, затова зарядът на атом H_2 беше изключен от оптимизацията.

Тъй като водородните връзки са линейни, трябва да се определят два свободни параметъра – разстоянието между водната молекула и прицелния атом и ъгъла на завъртане на водната молекула спрямо линията, която я свързва с атома. Тези параметри се оптимизират с квантово-механични симулации, в които всички други степени на свобода са ограничени. За съвместимост с полето CHARMM симулациите се правят с метода на Хартри-Фок с базисен набор 6-31G*.

Крайните оптимизирани заряди на новата функционална група, получени след фитиране към КвМ пресметнатите оптимални разстояния и енергии на взаимодействия и диполен момент на съединението, са дадени в табл. 3. Пълният заряд на групата е $-1e$.

Атом	Частичен заряд
C_2	-0.02
H_2	0.09
N	-0.33
HN	0.26
S	0.47
$O1S$	-0.49
$O2S$	-0.49
$O3S$	-0.49

Таблица 3: Оптимизирани заряди на новата функционална група (части от заряда на електрона).

Следващата стъпка е да се определят ковалентните параметри, т.е. равновесни стой-

ности и силови константи за всички химични връзки и валентни ъгли, които подлежат на параметризация.

FFTK оптимизира силовите константи за връзките и ъглите като изчислява и съпоставя квантово-механичните и молекулно-механичните повърхнини на потенциалната енергия (ППЕ) директно във вътрешни координати, вместо в нормални моди и декартови координати. За всяка вътрешна координата, т.е. за всяка връзка или ъгъл, се генерират малки отклонения в две противоположни посоки и се изчислява енергията на съответната изкривена конформация спрямо тази на минимизираната конформация. Така получените три точки служат за локално описание на формата на ППЕ. ММ ППЕ се пресмята с помощта на плъгина `NAMD Energy` [70] на `VMD` за всяка генерирана конформация. КвМ ППЕ се пресмята с помощта на хесиана, който представлява локално хармонично приближение на ППЕ за вибрационни движения на молекулата около минимизираната ѝ геометрия. Хесианът се смята чрез вътрешните координати. FFTK определя ефекта на всяко усукване на конформацията върху енергията по следната формула:

$$\Delta E = \sum_j \frac{1}{2} \frac{\partial^2 E}{\partial q_i \partial q_j} \delta q_i \delta q_j \quad (10)$$

където q_i е прицелната координата, сумата е по всички q_j координати, а производната е съответният елемент в хесиана. В съответствие с процедурата за параметризация на CHARMM хесианът се пресмята с метода `MP2/6-31G*`.

При оптимизацията на връзките и ъглите теглото на компонентата в целевата функция на възпроизвеждането на геометрията на съединението е увеличено от 1 на 2 по препоръка на членове от екипа, който разработва полето CHARMM [71]. Пак по същата препоръка отклонението в ъглите от минимизираната структура беше намалено от 10° на 2° . Началните и оптимизираните параметри на връзките и ъглите са представени в табл. 4.

Връзка	Начални параметри		Оптимизирани параметри	
	K_b [kcal/(mol Å ²)]	b_0 [Å]	K_b [kcal/(mol Å ²)]	b_0 [Å]
NNSO3 SNSO3	180.00	1.7920	204.03	1.6570
Ъгъл	K_θ [kcal/(mol rad ²)]	θ_0 [°]	K_θ [kcal/(mol rad ²)]	θ_0 [°]
CC3161 NNSO3 SNSO3	110.00	118.30	126.68	106.24
NNSO3 SNSO3 ONSO3	130.00	109.47	116.13	112.40
SNSO3 NNSO3 HNSO3	30.00	123.60	112.78	108.86

Таблица 4: Начални и оптимизирани параметри за химичните връзки и валентните ъгли на новата функционална група.

Параметризационната процедура завършва с оптимизация на параметрите на ъглите на усукване, които описват ротациите около химични връзки. Параметрите се получават от КвМ ППЕ, генерирани чрез явно сканиране на двустенните ъгли за параметризация с метода `MP2/6-31G*`. При сканирането се пресмята енергията за фиксирани стойности на ъгъла на усукване, като останалата част от съединението се оставя да релаксира, за да се изолира приносът на съответния двустенен ъгъл към енергията.

Не е необходимо да се сканират всички двустенни ъгли, които трябва да се параметризират. Достатъчно е да се сканира само по един от всички, които описват ротация около една и съща връзка. FFTK генерира файлове за две посоки на сканиране – една положителна и една отрицателна. Това се прави, за да се елиминират проблеми с рязко променяне на конформацията, което може да доведе до нестабилна симулация. При обработката на КвМ

симулации данните от отрицателната посока на сканиране се обръщат, за да се получи непрекъснат профил на усукването.

За всяка оптимизирана конформация се трябва да се пресметне ММ енергия. Преди това обаче всяка КвМ оптимизирана конформация се минимизира с ММ силово поле като се фиксират само параметрите на съответния двустенен ъгъл. След това параметрите на двустенните ъгли за параметризиране се оптимизират по метода на симулираното каляване. При тази процедура се варират само силовите константи. Мултиплетността на всеки ъгъл на усукване се задава предварително от потребителя. Фазовото отместване в CHARMM може да има стойности 0° и 180° , като това също се оптимизира. След симулираното каляване потребителят може да направи допълнителна оптимизация на получените параметри по метода на най-бързото спускане като се обреже ППЕ до зададен от потребителя праг. Целта на тази допълнителна стъпка е да се оптимизират параметрите така, че да възпроизвеждат по-добре ППЕ при по-ниски енергии.

В случая бяха сканирани по един двустенен ъгъл, описващ ротациите около връзките $C_1 - C_2$, $C_3 - C_2$, $C_2 - N$ и $N - S$. Сканирани бяха конформации през 5° . Средно-квадратичната разлика между квантово-механичния профил на потенциалната енергия и молекулно-механичния, получен с оптимизираните параметри на ъглите на усукване е 0.59 kcal/mol. Крайните стойности на подлежащите на параметризиране ъгли на усукване са дадени в табл. 5.

Ъгъл на усукване				K_χ [kcal/mol]	n	δ [$^\circ$]
CC3161	CC3161	NNSO3	HNSO3	0.040	1	180.0
CC3162	CC3161	NNSO3	HNSO3	0.569	1	180.0
CC3161	CC3161	NNSO3	SNSO3	2.030	1	180.0
CC3161	CC3161	NNSO3	SNSO3	1.025	2	180.0
CC3161	CC3161	NNSO3	SNSO3	0.326	3	0.0
CC3162	CC3161	NNSO3	SNSO3	2.261	1	0.0
CC3162	CC3161	NNSO3	SNSO3	1.180	2	180.0
CC3162	CC3161	NNSO3	SNSO3	0.263	3	0.0
CC3161	NNSO3	SNSO3	ONSO3	0.631	3	0.0
CC3161	NNSO3	SNSO3	ONSO3	0.746	4	0.0
NNSO3	CC3161	CC3162	OC3C61	2.995	3	180.0
HCA1	CC3161	NNSO3	SNSO3	0.242	3	0.0
HNSO3	NNSO3	SNSO3	ONSO3	0.091	3	180.0

Таблица 5: Оптимизираните параметри на подлежащите на параметризиране ъгли на усукване .

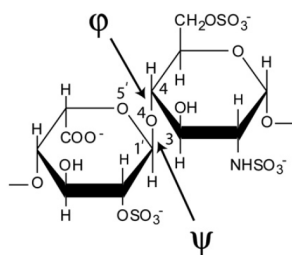
5.1.3 Валидиране на новите параметрите

Оценката на качеството на новите параметри беше изключително проблематична поради липсата на достатъчно експериментални данни за хепарин. Затова критерият дали новите параметри са достатъчно добри беше дали структурата на dp12 се запазва достатъчно стабилна по време на МД симулация.

Началните координати на произведен на хепарина додеказахарид бяха получени от магнитно-резонансната структура с код 1HPN [66] в базата данни PDB. За симулациите беше използван вторият кадър в структурата, при който идуроновата киселина е в 1C_4 конформация. Симулациите бяха направени с пакета за МД GROMACS 4.6.7. Беше използван интеграторът прескочи-кобила, като дължините на химичните връзки бяха замразени,

за да се позволи времева стъпка от 2fs. Списъците с най-близки съседни се обновяваха на всеки 10 стъпки с радиус на обрязване 12\AA . Радиусът на обрязване на нековалентните взаимодействия също беше 12\AA , като Ван дер Ваалсовите взаимодействия бяха постепенно изключени от 10\AA , а за електростатичните беше използван методът PME. Температурата беше поддържана на 300K от термостата v-rescale с константа на свързване 0.25ps. Постоянно налягане от 1atm беше поддържано от баростат на Паринело-Раман с константа на свързване 1.0ps. Симулациите са с продължителност от 50ns.

Първо беше проверено дали новите параметри деформират гликозидния мост, който свързва два последователно монозахарида. За целта бяха пресметнати ъглите на усукване ϕ и ψ , дефинирани на фиг. 12.



Фигура 12: Дефиниране на ъглите на усукване на гликозидната връзка между два монозахарида. Ъгълът ϕ се дефинира от атоми $O_5'-C_1'$ от първия монозахарид и O_4-C_4 от втория захарид. Ъгъл ψ се дефинира от атоми $C_1'-O_4-C_4-C_3$. Фигурата е от [72].

Получените от симулацията ϕ, ψ -диаграми са представени на фиг. 13. Те бяха сравнени с разпределението на тези ъгли, получено от Кан и съавтори [72] (фиг. 14). Както се вижда има много добро съвпадение между експерименталното разпределение на ϕ и ψ и това получено от симулациите. Средните стойности и стандартните отклонения на ъглите са дадени в табл. 6.

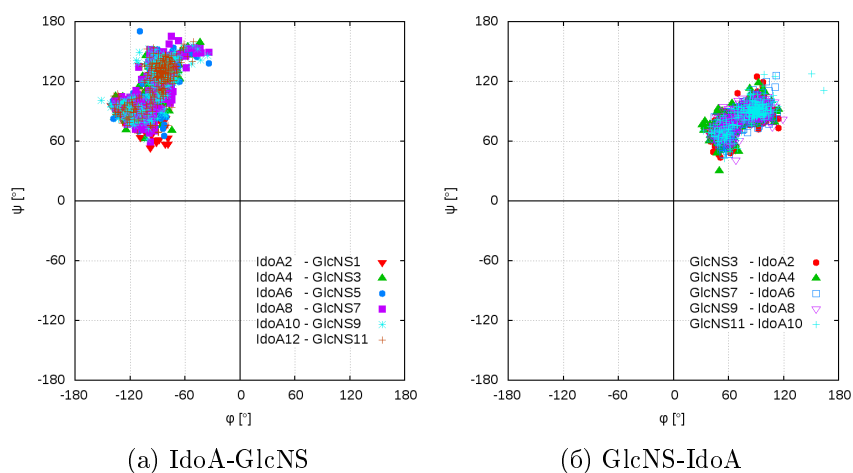
	$\phi_{IdoA-GlcNS} [^\circ]$	$\psi_{IdoA-GlcNS} [^\circ]$	$\phi_{GlcNS-IdoA} [^\circ]$	$\psi_{GlcNS-IdoA} [^\circ]$
Симулация	-87.74 ± 16.82	122.36 ± 22.77	76.31 ± 19.00	83.38 ± 12.68
1HPN	-77.06 ± 0.09	132.95 ± 0.27	78.67 ± 0.12	88.40 ± 0.43

Таблица 6: Средни стойности на ъглите ϕ и ψ за всички връзки IdoA-GlcNS и GlcNS-IdoA: експерименталните стойности и получените от симулация с новите параметри за NSO_3^- групата се съгласуват в рамките на стандартното отклонение.

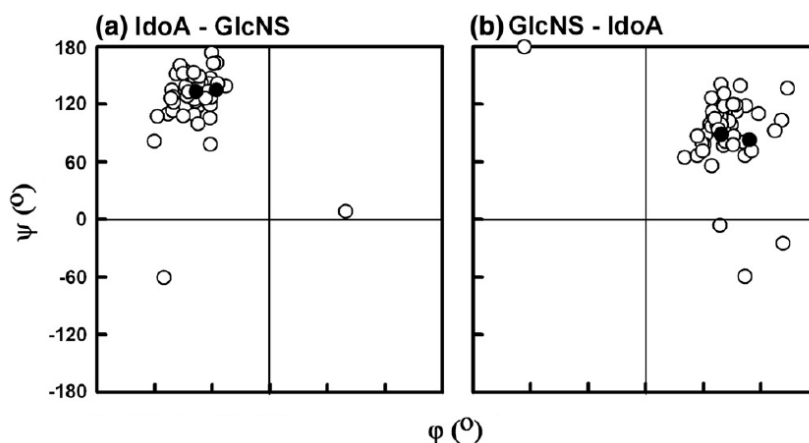
Тези данни ни дават основание да считаме, че новоразработените топология и параметри за сулфаматната анионна група NSO_3^- при атом C_2 са достатъчно качествени. Те бяха използвани за следващи симулации на взаимодействието на IFN- γ , IFN γ R α и производни на хепарина олигозахариди с различна степен на полимеризация.

5.2 Изследване на взаимодействието на hIFN- γ и hIFN- γ R α с МД симулации

Основната цел на това изследване беше да се създаде модел на взаимодействието на IFN- γ и двата му извънклетъчни рецептора IFN γ R α , който да обясни някои специфики в свързването. С-краищата на цитокина модулират афинитета му към рецепторите. В кристалографската структура на комплекса hIFN- γ -hIFN γ R α (PDB ID 1FG9 [10]) присъстват само координатите на атомите на първите 126 аминокиселини на цитокина. Координатите на



Фигура 13: Разпределение на ъглите ϕ , ψ по време на симулациите за различните гликозидни връзки.



Фигура 14: Разпределение на ъглите ϕ , ψ в експериментални структури на фрагменти хепарин. С незапълнени кръгчета са представени 67 двойки ъгли за връзката IdoA-GlcNS и 62 двойки за GlcNS-IdoA връзката, получени от 19 комплекса на протеин-хепаринни комплекси (PDB кодове 1AZX, 1AXM, 1BFB, 1BFC, 1E00, 1E03, 1FQ9, 1GMN, 1GMO, 1G5N, 1RID, 1QQP, 1XMN, 2AXM, 2GD4, 2HYU, 2HYV, 3DY0, и 3EVJ). С пълни кръгчета са изобразени ψ и ψ на додеказахаридата от структурата с PDB код 1HPN. Фигурата е от [72].

последните 18 аминокиселинни остатъка от С-края не са известни. Така, първата задача се състоеше в изграждане на цялостен модел на цитокина, в който са налични всичките 144 аминокиселинни остатъка на всеки от мономерите.

5.2.1 Нагъване на С-краищата на hIFN- γ

Експерименталните данни сочат, че С-краищата на hIFN- γ са много подвижни и не притежават конкретна твърда конформация [11]. Затова първоначално липсващите във всеки от мономерите 18 аминокиселини бяха изградени в напълно разгъната конформация с програмата за манипулация и визуализация на молекули PyMOL [54]. След това големините на ъглите на усукване на гръбнака на добавените аминокиселини бяха променени със случайни стойности. Така получената структура беше минимизирана и уравновесена и беше

използвана за начална структура за МД симулациите по нагъване на C-опашките на hIFN- γ .

МД симулациите бяха направени с пакета GROMACS 4.6.7 [48]. За подобряване на обхождането на фазовото пространство бяха проведени по три симулации с различни начални скорости с полето CHARMM 36 с модифициран TIP3P модел на водата. Всички симулации бяха с продължителност от 150 ps (т.е. 450 ps общо симулационно време) и бяха проведени при периодични гранични условия в симулационна кутия с размери $100 \times 120 \times 130 \text{ \AA}^3$. Използвани бяха твърди връзки, за да се позволи времева стъпка от 2 fs на интегратора прескочи-кобила. Системата беше симулирана в NPT ансамбъл при температура 310K, поддържана от термостат v-rescale с константа на свързване 0.25 ps и налягане от 1atm, поддържано от баростат на Паринело-Раман с константа на свързване 1ps. Списъците с най-близки съседи бяха обновявани на всеки 10 стъпки, електростатичното взаимодействие беше пресмятано с метода PME, а радиусите на обрязване на Ван дер Ваалсовото взаимодействие, директната PME сума и списъците със съседи бяха 14 \AA . Симулациите бяха направени на клъстера BIOSIM¹, както и на клъстера PHYSON [73] към Физически факултет на Софийски университет "Св. Климент Охридски".

За да се анализират симулациите и да се намерят най-вероятните и представителни конформации на C-краищата, получените траектории бяха подложени на клъстерен анализ. За отстраняване на всички ротационни и транслационни движения на центъра на масите на глобуларната част на hIFN- γ (а.к. 1-122), траекториите бяха фитирани и наложени върху началната структура по метода на най-малките квадрати. След това траекториите бяха обработени с програмата `g_rms` на GROMACS и беше пресметнат безразмерният параметър на сходство ρ [74] за конформацията на гръбнака на C-опашките (а.к. 123-144). Клъстерите бяха формирани по метода за клъстеризиране GROMOS [75] при радиус на обрязване по параметъра ρ 1.

Анализът показва, че C-опашките на hIFN- γ не остават в разгънато състояние. Във всички симулации те веднага започват да се нагъват и след около 15-20 ps вече са доста близо до глобуларната част на белтъка, като така се намалява общата повърхнина на молекулата, достъпна за разтворителя, както и радиусът на въртене. През останалото време, C-краищата продължават да се приближават до глобулата, но циклично променят ориентацията си спрямо нея.

5.2.2 Влияние на производни на хепарина олигозахариди върху взаимодействието на hIFN- γ и hIFN γ R α

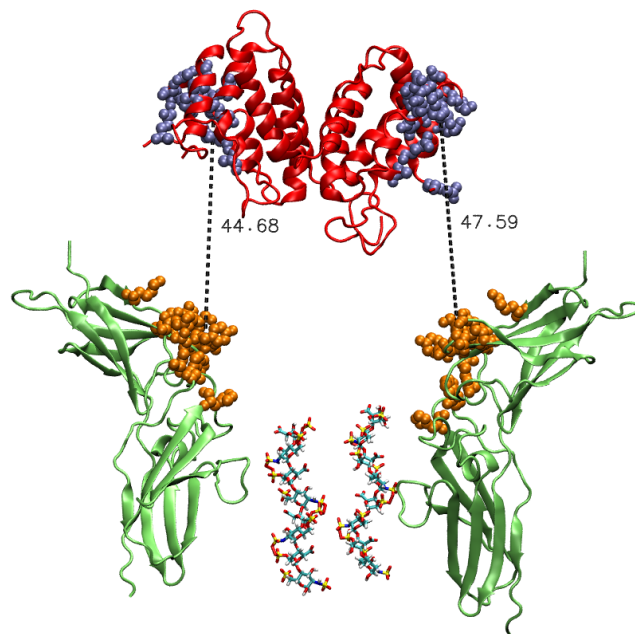
Получените представителни структури с различни конформации на човешки интерферон γ с напълно изградени C-краища бяха използвани за серия молекулно-динамични изследвания. Конформациите на пълната нагъната молекула на IFN- γ бяха наложени върху кристалографската структура на hIFN- γ в комплекс с рецепторите с PDB ID 1FG9, след което бяха отдалечени от рецепторите на няколко нанометра, за да се изследва процесът на формиране на комплекс.

Паралелно беше направена референтна симулация на кристалографския комплекс hIFN- γ -hIFN γ R α [9], за да се получи информация за средното разстояние между центровете на свързване в цитокина и двата рецептора.

Конформацията на напълно изградената и нагъната молекула на IFN- γ получена от центроида на най-посещавания клъстер по време на симулациите по нагъване, отдалече-

¹ Локален Linux клъстер, който разполага с 24 процесора Intel Xeon E55402.53GHz и 8 процесора Intel Xeon E5-2620v2.10GHz с общо 288 нишки и 544 GB RAM, както и с 4 графични ускорителя Tesla K20m, 8 ТВ дисково пространство.

на на известно разстояние от рецепторите, беше използвана и за други две симулации, в които бяха добавени по два олигозахарида между двата рецептора. Олигозахаридите бяха с различна степен на полимеризация – хекса- и октазахариди. Целта беше да се провери дали тяхното присъствие в близост до рецепторите ще спомогне за по-бързото и правилно формиране на комплекса hIFN- γ -hIFN γ R α . Началната конформация на системата цитокин–рецептори–олигозахариди е представена на фиг. 15).



Фигура 15: Начална конфигурация за симулациите по свързване на hIFN- γ и hIFN γ R α . Цитокинът е представен с червени ленти, рецепторите – със зелени ленти, а олигозахаридите – с цветни пръчици. Със сини и оранжеви сфери съответно са представени аминокиселините остатъци формиращи контактния интерфейс в IFN- γ и IFN γ R α . С черни пунктирани линии са представени разстоянията $d_1 = 47.6 \text{ \AA}$ и $d_2 = 44.7 \text{ \AA}$ между центровете на свързване от двете страни на молекулата на цитокина и тези в двата рецептора.

Всички симулации бяха направени с пакета GROMACS 4.6.7, със силово поле CHARMM 36 и модифициран TIP3P модел на водата. Симулациите бяха проведени с периодични гранични условия в симулационни кутии с размери $178 \times 164 \times 104 \text{ \AA}^3$, които съдържат над 300000 атома. Беше използван интеграторът прескочи-кобила с времева стъпка от 2 fs и твърди връзки. Изотермно-изобарният ансамбъл беше поддържан при температура от 300 K и налягане от 1 atm от термостат v-rescale с константа на свързване 0.25 ps и баростат на Паринело-Раман с константа на свързване 1 ps. За електростатиката беше използван методът PME с радиус на обрязване на правата сума 12 \AA . За Ван дер Ваалсовите взаимодействия беше използвана изглаждаща функция, която започваше да действа на разстояние 10 \AA с радиус на обрязване 12 \AA . Симулациите бяха направени на локалния клъстер BIOSIM, както и на клъстера PHYSON [73] към Физически факултет на Софийски университет “Св. Климент Охридски”. Симулациите са с продължителност от 50 ps поради сравнително големите размери на симулираните системи.

Разстоянията d_1 и d_2 между центровете на масите на центровете на свързване от двете страни на hIFN- γ и тези в двата рецептора за всички симулации са представени на фиг. 16. Средното разстояние между центровете на свързване в молекулата на интерферона и рецепторите в комплекса от референтната симулация е $6.5 \pm 0.4 \text{ \AA}$.

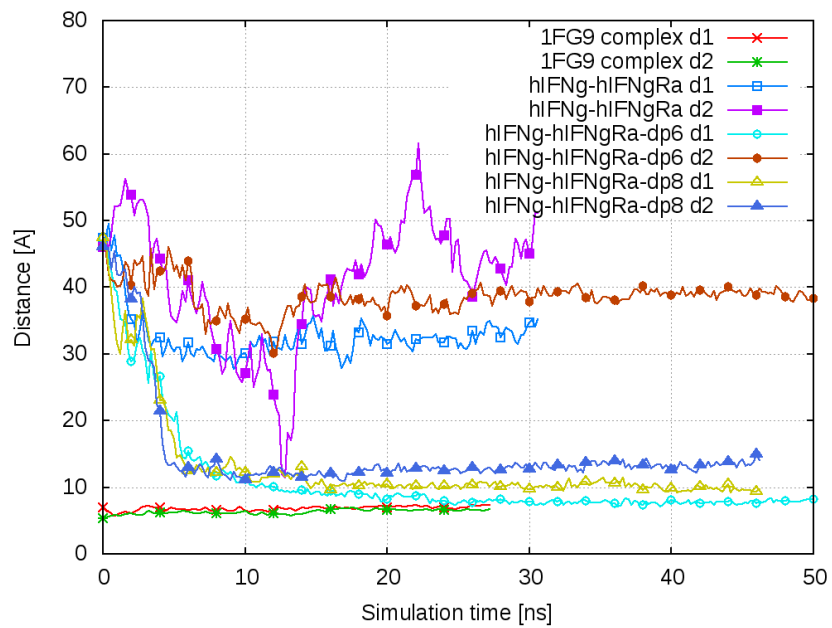
В симулацията, в която участват само с $IFN-\gamma$ и $IFN\gamma R\alpha$, силно положително заредените опашки на цитокина активно си взаимодействат с отрицателно заредени домейни в молекулите на всеки от рецепторите. Това взаимодействие пречи на глобуларната част на $IFN-\gamma$, където се намират рецептор-свързващите участъци, да се доближи достатъчно до рецепторите. Така опашките на цитокина всъщност пречат на правилното формиране на комплекс. Това поведение беше наблюдавано и в поредица други симулации с различни начални разстояния и ориентации между трите белтъка, както и с различни конформации на С-краищата на $IFN-\gamma$. Въпреки това С-краищата са важни за взаимодействието с извънклетъчните домейни на рецепторите. Резултати от предварителни симулации, в които молекулата на цитокина с липсващи по 18 ак. от опашките ($hIFN-\gamma$ от кристалографската структура) е отдалечена по аналогичен начин от рецепторите, показват, че цитокинът изобщо не взаимодейства с рецепторите и остава приблизително на мястото, на което е бил поставен в началото. Това се дължи на липсата на големия положителен заряд, който е концентриран в С-краищата на $IFN-\gamma$ ($+18e$). Всеки от рецепторите има заряд $-9e$. Налага се заключението, че електростатичното привличане между изцяло изградената молекула на цитокина и двата рецептора не е достатъчен фактор за бързо и правилно формиране на комплекс.

Това ни накарва да предположим, че бързото и правилно формиране на цитокин – рецепторен комплекс зависи от други помощни молекули. Хипотезата ни е, че такава роля играят протеогликаните, съдържащи хепарин и хепаран сулфат, които се намират в изобилие по повърхността на всички животински клетки. Проверката на тази хипотеза наложи разработването на параметри за силовото поле за хепарин. Полученото в резултат силово поле за Glc2NS ни позволи да проверим дали производни на хепарина олигозахариди биха подпомогнали свързването на $IFN-\gamma$ с рецепторите му.

Експерименталните данни сочат, че цитокинът взаимодейства най-силно с хекса- и октазахариди. По тази причина бяха направени две симулации с тези два олигозахарида. Както се вижда от фиг. 16, и в двата случая $d1$ намалява драстично и значително се доближава до референтните стойности за комплекса. И в двата случая това се дължи на електростатичното взаимодействие между олигозахаридите и С-краищата на $IFN-\gamma$, което е по-силно от това между опашките и отрицателно заредените домейни в рецепторите. На практика това привличане успява да придърпа цялата молекула на цитокина между рецепторите така, че свързващите центрове да се приближат значително и да имат възможност да си взаимодействат.

В симулацията с $dr6$ се оказва, че хексазахаридите не успяват да привлекат достатъчно силно С-краищата и на двата мономера на $IFN-\gamma$. Единият остава да взаимодейства с отрицателно заредения домейн на един от рецепторите и, както в случая на липса на олигозахариди, това пречи на рецептор-свързващия център от тази страна на цитокина да се доближи достатъчно до рецептора (виолетовата крива на фиг. 16).

Тези резултати дават качествено потвърждение на хипотезата, че формирането на комплекс между $IFN-\gamma$ и $IFN\gamma R\alpha$ е бавен и несигурен процес, който изисква участието на допълнителни молекули. Както се вижда от фиг. 16 производните на хепарина олигозахариди подпомагат приближаването на цитокиновата молекула към рецепторите поради силното електростатично привличане между отрицателно заредените олигозахариди и положително заредените С-краища на $IFN-\gamma$. Освен това беше установено, че октазахаридите оказват по-силно благоприятно действие върху формирането на цитокин-рецепторния комплекс от хексазахаридите. За по-подробен анализ на тези процеси и количествена оценка на влиянието на производни на хепарина олигозахариди с различна степен на полимеризация върху свързването на $IFN-\gamma$ и $IFN\gamma R\alpha$ са необходими допълнителни симулации, с по-голяма продължителност, набор от няколко различни начални конфигурации на участ-



Фигура 16: Разстояния $d1$ и $d2$ между центровете на масите на центрoвете на свързване в молекулата на hIFN- γ и в рецепторите IFN γ R α .

ниците и вероятно използване на методи за ускорено обхождане на фазовото пространство като, например, метадинамика.

Научни приноси

- Създаден е физичен модел на човешки интерферон гама и неговите извънклетъчни алфа-рецептори, с помощта на който са изследвани процесите и механизмите на свързване на белтъците в комплекс;
- Разработени са модел и методология за виртуално мутиране и *in silico* предсказване на стабилността на мутагенни форми на hIFN- γ чрез молекулно-динамични и метадинамични симулации;
- Въз основа на тези модел и методология са селектирани 12 от 100 мутагенни форми на човешки интерферон гама, които запазват локалната структура на спирала E и показват близка или по-голяма стабилност на вторичната структура на мутираната област от нативната форма на протеина;
- Издигната е обоснована хипотеза за критичната роля на трети участник в процеса на свързване на IFN- γ и неговите рецептори – гликозаминогликаните хепарин и NS-хепаран сулфат. Проведените изследвания доказват качествено валидността на тази хипотеза;
- Създадено е силово поле за гликозаминогликаните хепарин и NS-хепаран сулфат. За целта са разработени топология и параметри, съвместими със силовото поле CHARMM 36, за сулфаматната анионна група NSO_3^- , присъстваща в монозахариди, изграждащи тези гликозаминогликани. Получените с новите параметри свойства на захаридите се съгласуват добре с оскъдните експериментални данни. Те позволяват да се изследва взаимодействието на тези захариди с множество протеини;
- Проведено е методологично изследване на пакетите за МД симулации GROMACS и NAMD с цел оптимизиране на симулационните протоколи за изследване на биологични макромолекули и комплекси на голям брой изчислителни ядра;
- Изследвани са различни възможности за оптимизиране на далекодействащите електростатични взаимодействия. Показано е, че използването на интегратори с променлива времева стъпка значително повишава производителността на пакета NAMD, но влошава стабилността на симулацията. Установено е оптималното от гледна точка на производителност и стабилност съотношение на честотите на пресмятане на близките и далечните взаимодействия;
- Проведени са оптимизационни изследвания за съотношението на ядрата, отделени за пресмятане на близки и далечни взаимодействия. Показано е, че двукратното намаляването на съотношението по подразбиране (1/8 вместо 1/4) води до увеличаване на производителността, макар и за сметка на влошена скалируемост.

Научни статии, на които се базира дисертация

Публикации в научни журнали

1. **G. Nacheva, E. Lilkova, P. Petkov, P.St. Petkov, N. Ilieva, I. Ivanov, and L. Litov**, In silico studies on the stability of human interferon-gamma mutants, *Biotechnology & Biotechnological Equipment*, 2012, **26**(1), 200-204, DOI: 10.5504/50YRTIMB.2011.0036.
2. **E. Lilkova, G. Nacheva, P. Petkov, P.St. Petkov, S. Markov, N. Ilieva, and L. Litov**, Metadynamics study of mutant human interferon-gamma forms, *Computers and Mathematics with Applications*, 2012, **64**(3), 272-277, DOI: 10.1016/j.camwa.2012.01.061.
3. **E. Lilkova, L. Litov, P. Petkov, P.St. Petkov, S. Markov, and N. Ilieva**, Computer simulations of human interferon gamma mutated forms, *AIP Conference Proceedings*, 2010, **1203**, 914-919, DOI: <http://dx.doi.org/10.1063/1.3322582>.

Публикации в сборници на конференции

1. **D. Grancharov, E.Lilkova, N. Ilieva, P. Petkov, L. Litov**, Open Problems in High-Performance Molecular-Dynamics Simulations, *Information technologies and control*, 01/2012, Year X(2), 23-29.
2. **P. Petkov, E. Lilkova, P.St. Petkov, N. Ilieva, and L. Litov**, Application of metadynamics for investigation of human interferon gamma mutated forms, *Proc. 6th International Scientific Conference Computer Science 2011*, 2011, ISBN: 978-954-438-914-7, 347-352.

Изпратени за печат

1. **E. Lilkova, P. Petkov, N. Ilieva, and L. Litov**, Molecular Modelling of the interactions of Heparin-Derived Oligosaccharides, hIFN1- γ and its extracellular receptors, submitted to *AIP Conference Proceedings*.

Доклади на конференции

1. **E. Lilkova, P. Petkov, P. S. Petkov, S. Markov, N. Ilieva, N., and L. Litov**, Metadynamics study of mutated human interferon gamma forms, *International Conference on Mathematical Methods and Models in Biosciences, BIOMATH 2011*, Sofia, Bulgaria, 15-18 June 2011.

2. **P. Petkov, E. Lilkova, P. S. Petkov, N. Ilieva, and L. Litov**, Application of metadynamics for investigation of human interferon gamma mutated forms, *International Scientific Conference Computer Science'2011*, Ohrid, Macedonia, 01-03 September 2011.
3. **P. Petkov, E. Lilkova, P. S. Petkov, N. Ilieva, and L. Litov**, Application of metadynamics for investigation of human interferon gamma mutated forms, *Second Regional Conference "Supercomputing Applications in Science and Industry"*, Sunny Beach, Bulgaria, 20-21 September 2011.
4. **G. Nacheva and E. Lilkova**, Stability of mutant human interferon-gamma derivatives – a complex approach study, *Anniversary Molecular Biology Conference "50 years Institute of Molecular Biology Roumen Tsanev"*, Sofia, Bulgaria, 6-7 October 2011.
5. **P. Petkov, E. Lilkova, N. Ilieva, and L. Litov**, Modeling the Structure of Human Interferon Gamma C-Termini., *International Conference on Bioinformatics and Computational Biology, BIOCOMP 2012*, Varna, Bulgaria, 20-21 September 2012.
6. **P. Petkov, E. Lilkova, D. Grancharov, P. S. Petkov, N. Ilieva, and L. Litov**, Calculating Binding Free Energies of Variants of hIFN-g and their Extracellular Receptor., *International Conference on Mathematical Methods and Models in Biosciences, BIOMATH 2013*, Sofia, Bulgaria, 16-21 June 2013.
7. **E. Lilkova, P. Petkov, N. Ilieva, D. Grancharov, L. Litov, P. S. Petkov**, Molecular Dynamics Simulations of Human Interferon Gamma., *Втори Национален Конгрес по Физически Науки*, София, България, 25-29 септември 2013.
8. **E. Lilkova, P. Petkov, N. Ilieva, and L. Litov**, Molecular Modelling of the Complex hIFN γ -hIFN γ R-Heparin-Derived Oligosaccharides, *International Conference on Mathematical Methods and Models in Biosciences, BIOMATH 2015*, Blagoevgrad, Bulgaria, 14-19 June 2015.
9. **E. Lilkova, P. Petkov, N. Ilieva, and L. Litov**, Molecular Modelling of the interactions of Heparin-Derived Oligosaccharides, hIFN γ and its extracellular receptors, *Seventh Conference of the Euro-American Consortium for Promoting the Application of Mathematics in Technical and Natural Sciences*, Albena, Bulgaria, June 28-July 3 2015.

Благодарности

Бих искала да благодаря на научния ми ръководител доц. д-р Леандър Литов за предоставената възможност да работя в тази интересна област и да напиша тази дисертация, както и за полезните съвети и наставления, които ми дава през цялото време. Изказвам специални благодарности на д-р Пейчо Петков за вложените в мен време и усилия, както и за оказаната помощ, съдействие и проявеното търпение. Бих желала сърдечно да благодаря за помощта на д-р Невена Илиева, която допринесе съществено за оформянето на дисертацията с ценни съвети и редакции. Благодаря и на д-р Венелин Кожухаров, Георги Георгиев, Дамян Грънчаров и на всички останали колеги от групата на доц. Литов, че винаги се отзоваваха, когато имах нужда от помощ.

Благодаря на проф. д-р Иван Иванов и д-р Геновева Начева от Институт по Молекулярна Биология на БАН за предоставянето на списъка с мутагенните форми на hIFN- γ , за съвместната ни работа и ценните консултации относно биологични въпроси, които са извън моята сфера на компетентност.

Изказвам дълбоки благодарности и признателност към цялото ми семейство и близките ми, които неизменно ме търпяха и подкрепяха през всички тези години.

Изследванията, представени в дисертацията, бяха финансирани от Фонд научни изследвания по проекти “Развитие на център за върхови постижения “Суперкомпютърни приложения” (Super CA++)” (ДЦВП02/1/29.12.2009г.) и “Виртуално и експериментално изследване на взаимодействието на цитокини с техните рецептори” (ДМУ 23/03-12.12.2011г.), както и по проект “Изграждане на съвременна образователна и научноизследователска среда за развитието на докторанти, постдокторанти и млади учени във Физическия факултет на СУ “Св. Климент Охридски”” (BG051 PO001-3.3.06-0057), финансиран по Оперативна програма “Развитие на човешките ресурси”, съфинансирана от Европейския социален фонд на Европейския съюз.

Библиография

- [1] Farrar M. A., and Schreiber R. D., *Annu Rev Immunol*, 1993, **11**, 571-611.
- [2] Schroder K., Hertzog P. J., Ravasi T., and Hume D. A., *J Leukoc Biol*, 2004, **75**, 163-189.
- [3] Schoenborn J. R., and Wilson C. B., In: Alt F. W., Editor(s), *Adv Immunol*, 2007, Academic Press, **96**, 41-101.
- [4] Hertzog P., Forster S., and Samarajiwa S., *J Interferon & Cytokine Res*, 2011, **31**(1), 5-11.
- [5] Pollard K. M., Cauvi D. M., Toomey C. B., et al., *Discov Med*, 2013, **16**(87), 123-131.
- [6] Zaidi M. R. and Merlino G., *Clin Cancer Res*, 2011, **17**(19), 6118-6124.
- [7] Ealick S.E., Cook W. J., Vijay-Kumar S., et al., *Science*, 1991, **252**, 698-702.
- [8] Walter M. V., Windsor W. T., Nagabhushan T. L., et al., *Nature*, 1995, **376**, 230-235.
- [9] Thiel D. J., le Du M.-H., Walter R. L., et al., *Structure*, 2000, **8**(9), 927-936.
- [10] Altobelli G., Nacheva G., Todorova K., et al., *PROTEINS*, 2001, **43**, 125-133.
- [11] Grzesiek S., Döbeli H., Gentz R., et al., *Biochemistry*, 1992, **31**, 8180-8190.
- [12] Rinderknecht E., O'Connor B. H., and Rodriguez H., *J Biol Chem*, 1984, **259**, 6790-6797.
- [13] Pan Y.-C. E., Stern A. S., Familletti P. C., et al., *Euro J Biochem*, 1987, **166**, 145-149.
- [14] Haelewyn J., Michaels L., Verhaert P., et al., *Biochem J*, 1997, **324**, 591-595.
- [15] Slodowski O., Böhm J., Schöne B., and Otto B., *FEBS Lett*, 1991, **202**, 1133-1140.
- [16] Celada A., and Schreiber, R. D., *J Biol Chem*, 1987, **139**(1), 147-153.
- [17] Zu X., Jay F. T., *J Biol Chem*, 1991, **266**, 6023-6026.
- [18] Subramaniam P. S., Mujtaba M. G., Paddy M. R., and Johnson H. M., *J Biol Chem*, 1999, **274**(1), 403-407.
- [19] Johnson H. M., Torres B. A., Green M. M., et al., *Biochem Biophys Res Comm*, 1998, **244**, 607-614.
- [20] Larkin III J., Subramaniam P. S., Torres B. A., and Johnson H. M., *J Interferon & Cytokine Res*, 2001, **21**(6), 341-348.
- [21] Petrov S., Boyanova M., Berzal-Herranz A., et al., *Biotech Biotechnol Equip*, 2009, **23**(1), 1058-1062.
- [22] Marsters S. A., Pennica D., Bach E., et al., *PNAS*, 1995, **92**, 5401-5405.
- [23] Bach E. A., Tanner J. W., Marsters S., et al., *Mol Cell Biol*, 1996, **16**(6), 3214-3221.
- [24] Pestka S., Kotenko S. V., Muthukumaran G., et al., *Cytokine & Growth Factor Revs*, 1997, **8**(3), 189-206.
- [25] <http://www.pdb.org>
- [26] Jarpe M. A., and Johnson H. M., *J Interferon Res*, 1990, **10**, 243-252.
- [27] Johnson H. M., Noon-Song E. N., Dabelic R., and Ahmed C. M., *Front Immunol*, 2013, **4**, 202.
- [28] Gallagher J. T., Lyon M., and Steward W. P., *Biochem J*, 1986, **236**, 313-325.

- [29] Hileman R. E., Fromm J. R., Weiler J. M., and Linhardt R. J., *BioEssays*, 1998, **20**, 156-167.
- [30] Mulloy B., and Forster M. J., *Glycobiology*, 2000, **10**(11), 1147-1156.
- [31] Lortat-Jacob H., Turnbull J. K., and Grimaud J.-A., *Biochem J*, 1995, **310**, 497-505.
- [32] Saesen E., Sarrazin S., Laguri C., et al., *J Am Chem Soc*, 2013, **135**, 9384-9390.
- [33] Lortat-Jacob H., *Biochem Soc Trans*, 2006, **34**(Pt 3), 461-464.
- [34] Sanchez H. P., Tatarenko K., Nigen M., et al., *Biochemistry*, 2006, **45**, 13227-13238.
- [35] Frenkel D., and Smit B., *Understanding Molecular Simulation*, 2nd Ed., 2002, Academic Press, San Diego.
- [36] Leach, A., *Molecular Modelling: Principles and Applications*, 2nd Ed., 2001, Prentice-Hall.
- [37] Beauchamp K. A., Lin Y.-S., Das R., and Pande V. S., *J Chem Theory Comp*, 2011, **8**, 1409-1414.
- [38] Darden T., York D., and Pederson L., *J Chem Phys*, 1993, **98**, 10089-10092.
- [39] Ryckaert J.-P., Ciccotti G., and Berendsen H. J. C., *J Comp Phys*, 1977, **23**, 327-341.
- [40] Andersen H. C., *J Comp Phys*, 1983, **52**, 24-34.
- [41] Tuckerman M., Berne B. J., and Martyna G. J., *J Chem Phys*, 1992, **97**(3), 1990-2001.
- [42] Ma Q., Izaguirre J. A., and Skeel R. D., *SIAM J Sc Comp*, 2003, **24**, 1951-1973.
- [43] Garcia-Archilla B., Sanz-Serna J. M., and Skeel R. D., *SIAM J Sc Comp*, 1998, **20**(3), 930-963.
- [44] Grest G. S., and Kremer K., *Phys Rev A*, 1986, **33**(5), 3628-3631.
- [45] Laio A., and Gervasio F. L., *Rep Prog Phys*, 2008, **71**, 126601.
- [46] Ensing B., Laio A., Parrinello M., and Klein M. L., *J Phys Chem B*, 2005, **109**, 6676-6687.
- [47] <http://scc.acad.bg/ncsa/index.php/bg/>
- [48] Hess B., Kutzner C., van der Spoel D., and Lindahl E., *J Chem Theory Comp*, 2008, **4**(3), 435-447.
- [49] Phillips J. C., Braun R., Wang W., et al., *J Comp Chem*, 2005, **26**, 1781-1802.
- [50] Geimer M., Wolf F., Wylie B. J. N., et al., *Concurr Comp Pract E*, 2010, **22**(6), 702-719.
- [51] Ramachandran G. N., Ramakrishnan C., and Sasisekharan V., *J Mol Biol*, 1963, **7**, 95-99.
- [52] Lovell S. C., Davis I. W., Arendall W. B. 3rd, et al., *Proteins*, 2003, **50**(3), 437-450.
- [53] Kabsch W., and Sander C., *Biopolymers*, 1983, **22**, 2577-2637.
- [54] The PyMOL Mol Graphics System, Version 1.7.4 Schrödinger, LLC., <https://www.pymol.org>
- [55] MacKerell Jr., A. D., Bashford D., Bellott M., et al., *J Phys Chem B*, 1998, **102**, 3586-3616.
- [56] Best R. B., Zhu X., Shim J., et al., *J Chem Theory Comp*, 2012, **8**(9), 3257-3273.
- [57] http://mackerell.umaryland.edu/charmm_ff.shtml
- [58] Guvench O., Mallajosyula S. S., Raman E. P., et al., *J Chem Theory Comp*, 2011, **7**(10), 3162-3180.
- [59] Mallajosyula S. S., Guvench O., Hatcher E. R., and MacKerell Jr. A. D., *J Chem Theory Comp*, 2012, **8**(2), 759-776.
- [60] Vanommeslaeghe K., and MacKerell Jr. A. D., *J Chem Inf Model*, 2012, **52**, 3144-3154.
- [61] Vanommeslaeghe K., Raman E. P., and MacKerell Jr. A. D., *J Chem Inf Model*, 2012, **52**, 3155-3168.
- [62] <https://cgenff.paramchem.org>

- [63] Brooks B. R., Brooks III C. L., MacKerell Jr. A. D., et al., *J Comp Chem*, 2009, **30**(10), 1545-1614.
- [64] Mayne C. G., Saam J., Schulten K., et al., *J Comp Chem*, 2013, **34**(32), 2757–2770.
- [65] Humphrey W., Dalke A., and Schulten K., *J Mol Graph*, 1996, **14**, 33-38.
- [66] Mulloy B., Forster M. J., Jones C., and Davies D. B., *Biochem J*, 1993, **293**, 849-858.
- [67] Frisch M. J., Trucks G. W., Schlegel H. B., et al., Gaussian 09, Revision A.02, Gaussian, Inc., Wallingford CT, 2009.
- [68] Møller C., and Plesset M. S., *Phys Rev*, 1934, **46**, 0618-22.
- [69] Vanommeslaeghe K., Hatcher E., Acharya C., et al., *J Comp Chem*, 2010, **31**, 671-690.
- [70] <http://www.ks.uiuc.edu/Res/vmd/plugins/namdenergy/>
- [71] <http://www.charmm.org/ubbthreads/ubbthreads.php?ubb=showflat&Number=33420>
- [72] Khan S., Gor J., Mulloy B., and Perkins S. J., *J Mol Biol*, 2010, **395**, 504-521.
- [73] <http://physon.phys.uni-sofia.bg/>
- [74] Maiorov V. N., and Crippen G. M., *PROTEINS*, 1995, **22**, 273-283.
- [75] Daura X., Gademann K., Jaun B., et al., *Angew Chem Int Ed*, 1999, **38**, 236-240.