Софийски университет
"Св. Климент Охридски"

Физически факултет
Катедра "Атомна физика"

Елена Боянова Лилкова

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.

Съдържание

Съкращения

а.к.	аминокиселини
ГАГ	глюкозаминогликани
КП	колективни променливи
МД	молекулна динамика
ПГ	протеогликани
ПСЕ	повърхнината на свободната енергия
ПСП	потенциал на средното поле
СП	силово поле
ХС	хепаран сулфат

BE-Meta	метадинамика с обмяна на метадинамичния потенциал
CMAP	контактна карта
FGF	фибробластен растежен фактор
GAS	гама-активирана последователност
Gal	галактоза
GalNAc	N-сулфатиран галактозамин
GalNAc(4S)	4-сулфатиран N-сулфатиран галактозамин
GalNAc(6S)	6-сулфатиран N-сулфатиран галактозамин
GlcA	глюкуронова киселина
GlcA(2S)	2-сулфатирана глюкуронова киселина
Glc2NS	N-сулфатиран глюкозамин
GlcNAc	N-ацетилиран глюкозамин
GlcNAc(6S)	6-сулфатиран N-ацетилиран глюкозамин
CS	хондроитин сулфат
H3	хистон 3
hIFN-γ	човешки IFN-γ
HS	хепаран сулфат
IduA	идуронова киселина
IFN-γ	интерферон-γ
IFNγR	интерферон-γ рецепторен комплекс
IFNγRα	α-рецептор на IFN-γ
IFNγRβ	β-рецептор на IFN-γ
IL	интерлевкин
IRF1	интерферон-регулиращ фактор 1
Jak-1	янус активационна киназа 1
Jak-2	янус активационна киназа 2
K_d	константа на дисоциация
NA	N-ацетилиран

NLS	сигнал за ядрена локализация
NPC	ядрен поров комплекс
NS	N-сулфатиран
shIFNγRα	извънклетъчния домейн на α-рецептора на hIFN-γ
sIFNγRα	извънклетъчния домейн на α-рецептора на IFN-γ
STAT1	сигнален трансдукционен фактор и активатор на транскрипцията
UA	уронова киселина

Увод

Дисертацията е посветена на моделирането на човешки интерферон гама с помощта на 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. Анормалната 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 sIFNγRα 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 критериите 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, дефинирано от торсионните ъгли на гръбнака на аминокиселинния остатък на позиция 86.

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 IFNγRα не е достатъчно за бързо и правилно формиране на комплекс. Това ни доведе до хипотезата, че гликозаминогликаните вероятно участват като помощни молекули в този процес. За да бъде тествана 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 дисертацията, списък със залегналите в основата ѝ публикации. В две приложения са представени получените параметри и топология за хепарин. Направен е и списък с използваната литература.

Глава 1

Въведение

1.1 IFN-γ

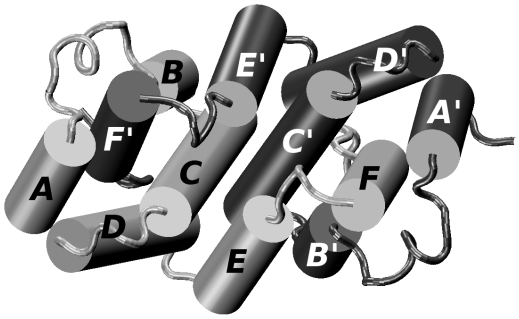
Интерферон-γ (IFN-γ) е плейотропен цитокин, който играе основна роля в регулацията на защитата на организма и имунопатологичните процеси [1]. Въпреки че първоначално е бил изолиран въз основа на антивирусната си активност, IFN-γ проявява основно имуномодулаторни свойства, които включват силно антипролиферативно действие, предизвикване на антивирусно състояние, подготвяне на макрофагите за унищожаване на туморни клетки, модулиране на производството на антитела, както и стимулиране на растежа, диференциацията и узряването на различни клетки [2,3]. 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].

IFN-γ се секретира от клетките на имунната система, основно NK-клетки (отговорни за вродения имунен отговор), CD8+ лимфоцити и CD4+ Th1 Т-клетки (придобит имунен отговор) [2, 3]. Той осъществява ефектите си чрез взаимодействие с видово-специфичен рецепторен комплекс (IFNγR), експресиран върху клетъчната повърхност на почти всички видове клетки [5]. Рецепторът е изграден от по два мономера на два различни протеина – IFNγRα и IFNγRβ.

Генетични заболявания, свързани с намаляване или липса на производството на IFN-γ или намаляване на чувствителността към него, водят до висока податливост на бактериални и вирусни инфекции, както и повишени честота, разнообразие и агресивност на тумори [3]. 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 [3,6], и оказват негативно влияние върху някои видове тумори [7]. Това налага търсенето на начин да се контролира биологичната активност на IFN-γ при повишена експресия на белтъка.

1.1.1 Структура

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. Всички 12 спирали са почти успоредни на локална некристалографска ос от 2-ри порядък [8–10]. Спирали E и F от карбоксилния край на всяка верига асоциират спирали A, B, C и D от другата верига (фиг. 1). Спиралите са свързани помежду си с къси примки от по три до пет аминокиселинни остатъка, с изключение на примка AB, която се състои от 13 остатъка, които обграждат спирала F. Останалата част от веригата след спирала F (а.к.



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

остатъци 121-143) се разпростира извън глобулата на молекулата в разтворителя [8–10].

Установено е, че примката AB и C-края не притежават вторична структура [8,10]. Свързаните за рецептор IFN-γ мономери се различават от свободните форми основно в примката AB (остатъци 16-27). В свободната форма тя е неструктурирана [8]. Когато IFN-γ се свърже с извънклетъчния домейн на α-рецептора (sIFNγRα), примката се структурира и остатъци 20-23 в центъра ѝ оформят добре очертана врътка от 3_{10} спирала, перпендикулярна на спирала F [10].

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 локализирани в C-края. Това дава на цитокина нетен заряд $+16e$[1] при физиологични условия [11]. В зрялата молекула на IFN-γ няма цистеинови остатъци, поради което в протеина липсват вътрешно-молекулни дисулфидни връзки.

1.1.2 N-край

IFN-γ се свързва с извънклетъчния домейн от веригата на IFNγRα чрез N-края си [10,12]. Синтетичен пептид, обхващащ остатъците 1-39 на цитокина, е в състояние да се свърже с α единиците на рецептора [12,13].

Установено е, че N-краят участва в поддържането на антипаралелна α-спирална глобуларна структура, както и на целостта на молекулата. Аминокиселинните остатъци на позиции от 1 до 7 са абсолютно необходими за повторното нагъване на IFN-γ от денатурираното му състояние [14]. Отстраняването на 8–10 остатъка от N-края предизвиква изменения във вторичната структура на IFN-γ, които водят до драстично намаляване на α-спиралното съдържание (6%) и съответно до загуба на функционална активност. Мутанти с премахнат N-край се характеризират с високо съдържание на β-листи (36%) и случайно нагънати вериги [14].

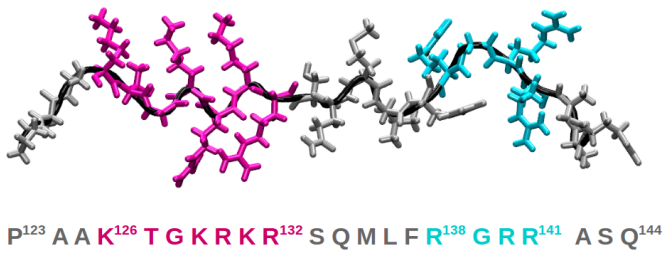
1.1.3 C-край

Един от най-неясните и спорни въпроси, засягащи IFN-γ, е свързан с ролята, която последните 22 аминокиселини, разположени в C-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 [8–10].

[1] e е зарядът на електрона.

C-краят на IFN-γ е силно заредена, изложена на разтворителя опашка без твърда конформация [15], която е изключително чувствителна към протеолитично действие [16, 17]. C-опашката на биологично активния естествен IFN-γ е хетерогенна [16], като от нея могат да липсват между 5 и 16 а.к. остатъка [17]. Отстраняването на къси аминокиселинни последователности не намалява биологичната активност на цитокина, дори напротив [18, 19].

C-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 (фиг. 2). Участъкът D2 по-скоро има модулираща функция и не е от критично значение за свързването с рецептора. Известно е, че тетрапептидът Arg^{129}-Lys-Arg-Ser^{132} допринася съществено за високия афинитет на белтъка към извънклетъчния му рецептор и съответно за капацитета на IFN-γ да предизвиква множество реакции в прицелните клетки [18, 20].



Фигура 2: C-краят на човешки интерферон гама. Участъкът D1 е оцветен в розово, а D2 – в синьо.

Изглежда, че участието на C-края в свързването с рецептора не се осъществява чрез структурно дефиниран интерфейс. Приноса на C-опашката в свързването се обяснява с неспецифични взаимодействия между базичните аминокиселини на IFN-γ и киселинните клъстери в рецептора му, включващи множество конформационни състояния [10].

Експериментални данни сочат, че C-опашките на цитокина взаимодействат с вътреклетъчната част на IFNγRα (а.к. остатъци 252-291). Синтетичен пептид, обхващащ аминокиселинната последователност 95-132 на hIFN-γ, експресиран вътреклетъчно, успява да предизвиква антивирусно състояние на клетката [21, 22].

1.1.4 NLS

Експериментално е показано, че при свързване на миши IFN-γ с рецептора, комплексът се интернализира в клетката [23]. Впоследствие се установява, че както молекулата на миши, така и на човешки интерферон съдържат така наречените сигнали за ядрена локализация (NLS), които служат за белязване на протеини за транспортиране в клетъчното ядро чрез ядрени порови комплекси (NPC).

В hIFN-γ са открити две предполагаеми NLS – горна в спирала E (остатъци 83-89) [24] и долна, локализирана в неструктурирания C-регион (остатъци 126-132) [21]. Установено е, че NLS последователността играе важна роля в предизвикването на биологичен отговор в клетките [21, 22], но няма единно мнение коя от двете вериги е по-съществена [25, 26].

1.2 IFNγR

Функционалният рецепторен комплекс на IFN-γ се състои от два мономера на лиганд-свързващата верига IFNγRα и два на помощен фактор IFNγRβ. [27]. IFNγRα е отговорен не само за взаимодействието с цитокина, но и участва във вътреклетъчната сигнализация. IFNγRβ не взаимодейства директно с IFN-γ, но е необходим за предаването на сигнала [27,28]. И двата протеина спадат към семейството на клас II цитокинови рецептори.

В нестимулирани клетки молекулите на α и β рецепторите не са асоциирани едни с други, а по-скоро са свързани с неактивни форми на ензимите – съответно, Янус активационна киназа 1 и 2 (Jak-1 и Jak-2). Стимулация с IFN-γ води до димеризация на IFNγRα, последвана от асоцииране на две единици IFNγRβ. Тази лиганд-индуцирана олигомеризация доближава цитоплазмените домейни на α и β рецепторите и асоцииираните към тях Jak-1 и Jak-2, при което те се активират. Активираните кинази фосфорилират аминокиселинен остатък Tyr^{440} във вътреклетъчните части на IFNγRα, като така се формира двойка места за скачване на латентни форми на транскрипционния фактор STAT1. При асоцииране с α-рецепторите в молекулите на STAT1 се фосфорилират тирозинни остатъци, което им позволява да димеризират. Накрая STAT1 димера се транслокира в клетъчното ядро, където индуцира транскрипция на индуцируеми от IFN-γ гени [28].

Взаимодействията между IFN-γ и IFNγRα, както и между извънклетъчните домейни на IFNγRα и IFNγRβ са видово специфични. Цитоплазмените части на IFNγRβ и на IFNγRα не са видово специфични. IFN-γ–IFNγRα взаимодействието не зависи от това между IFNγRα и IFNγRβ [29].

1.2.1 IFNγRα

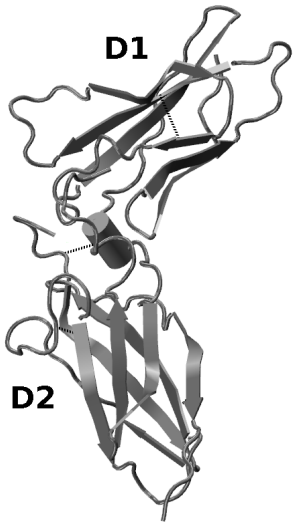
IFN-γ активира прицелните клетки чрез специфично взаимодействие с висок афинитет ($K_d = 10^{-10}$ M) с извънклетъчния домейн на α-рецептора на човешки интерферон (shIFNγRα), [27]. Въпреки високия афинитет, формирането на комплекс само между тези белтъци не е достатъчно за трансдукция на сигнала и предизвикване на биологичен отговор [29]. Този процес изисква задължителното участие на 5 белтъка: IFNγRα, IFNγRβ, Jak-1, Jak-2 и STAT1 [28].

IFNγRα е гликопротеин, който има извънклетъчна, трансмембранна и вътреклетъчна (цитоплазмена) част. Гликозилацията на белтъка не оказва влияние на свързването на лиганда и не променя афинитета му [29].

Молекулата на shIFNγRα се състои от два имуноглобулино-подобни β-листни домейна (фиг. 3). D1 обхваща N-края на белтъка (а.к. 14-102) и е оформен като β-сандвич с три β-ленти подредени върху слой от четири β-ленти. D2 е в C-края (остатъци 114-221) и също има структурата на β-сандвич, но при него и двата листа се стоят от по четири ленти. Двата домейна са свързани чрез подвижен линкер (а.к. 103-113), който съдържа къса α-спирала от 6 остатъка [8,10].

shIFNγRα съдържа 8 цистеинови остатъка, които формират 4 дисулфидни моста. Установено е, че и четирите са изключително важни за поддържането на структурата и съответно за биологичната активност на протеина [30].

Цитоплазмената част на рецептора има две функционални области, важни за предаването на сигнала. Първата е близо до клетъчната мембрана и обхваща сегмента Leu^{266}-Pro-Lys-Ser^{269}. На това място се свързва Jak-1 [29]. Другата ключова за функциите на рецептора област е в C-края му и се състои от а.к. остатъци Tyr^{440}-Asp-Lys-Pro-His^{444}. Тази последователност е напълно консервирана в човешкия и мишия рецептор и служи за свързване на STAT1 след фосфорилиране на Tyr^{440} при свързване на лиганда. Останалите четири аминокиселини от този пентапептид са отговорни за специфичността на взаимодействието



Фигура 3: Структура на извънклетъчния домейн на α-рецептора на човешки интерферон гама. С цилиндри са представени α-спиралите на белтъка, а с ленти – β-листите. С черна пунктирана линия са означени дисулфидните мостове.

на SH2-домейна на STAT1 и фосфорилирания Tyr^{440}. Веригата след His^{444}, както и тази между 290-тия и 439-тия остатък нямат отношение към биологична функция на IFNγRα, но изглежда служат за пространствено разделяне на двете функционални области. Синтетичен мутант, в който тези вериги са премахнати, няма биологична активност, тъй като няма достатъчно място да се асоциират Jak-1 и STAT1 [29].

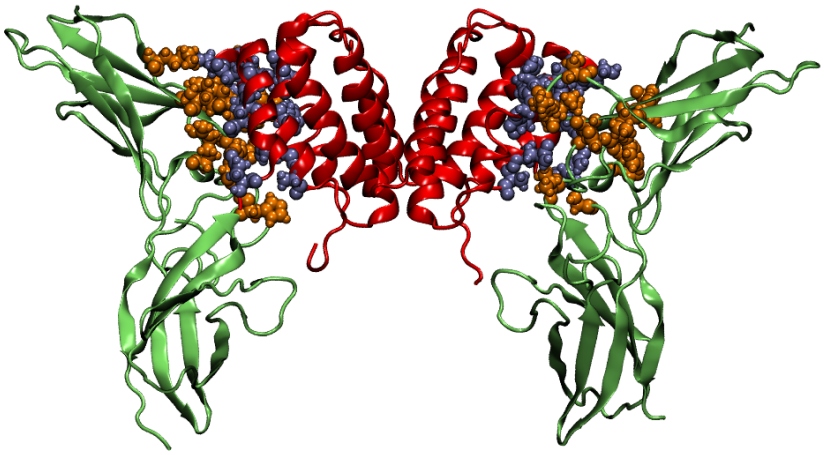
1.2.2 Структура на комплекса hIFN-γ–shIFNγRα

Кристалографската структура на комплекса hIFN-γ–shIFNγRα е получена с рентгенова дифракция и е достъпна в базата данни PDB [31] с ID 1FG9 [10]. Две shIFNγRα молекули свързват двустранно двете идентични повърхнини на димера на hIFN-γ и така формират 2:1 hIFN-γ–shIFNγRα комплекс със симетрия от втори порядък (фиг. 4).

Аминокиселинните остатъци в IFN-γ , които образуват повърхнината, контактуваща с всеки от рецепторите, са групирани в два сегмента – спирала A, примка AB и спирала B 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 (табл. 1). Част от взаимодействащите с лиганда а.к. остатъци (Lys^{47}, Trp^{82}, Glu^{101} и Trp^{207}) са консервирани в човешкия и мишия IFNγRα [29]. Не се наблюдават контакти с рецептора след Ala^{118} 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.

1.2.3 IFNγRβ

Вторият протеин в рецепторния комплекс на интерферон гама, IFNγRβ, не взаимодейства директно с цитокина, а играе ролята на помощен фактор за трансдукция на сигнала [27,28]. Открито е, че вариациите в експресията на този белтък имат модулиращ ефект върху формирането и функцията на рецепторния комплекс и насочват подадения от hIFN-γ сигнал към различни клетъчни отговори. При стимулация с цитокина, високата експресия на IFNγRβ предизвиква бързо активиране на STAT1 и високо ниво на IRF1 (интерферон-



Фигура 4: Кристалографска структура на комплекса hIFN-γ–shIFNγRα. 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 IFNγRα.

Таблица 1: Водородни връзки между мономерите на hIFN-γ и рецептора.

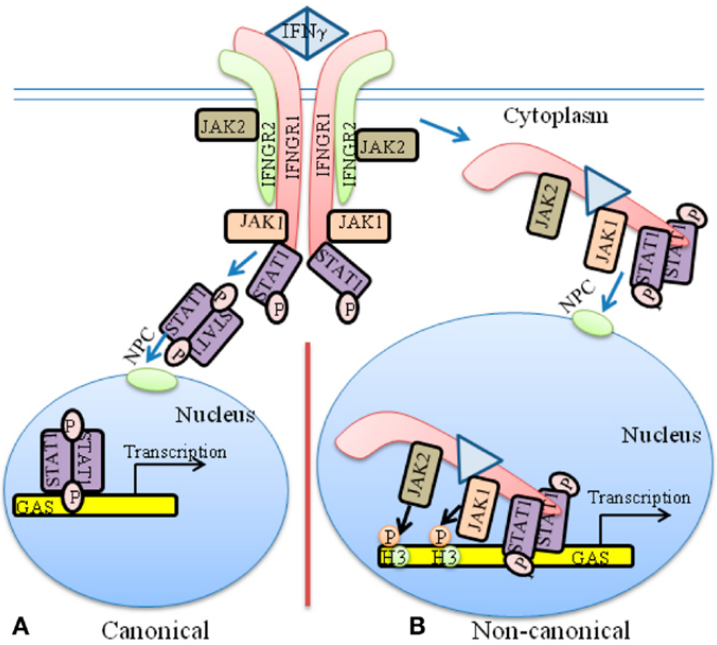
Мономер	А.к. от hIFN-γ	А.к. от shIFNγRα
I1	Gln^{1}	Trp^{207}
I1	Glu^{9}	Arg^{106}
I1	Glu^{9}	Trp^{207}
I1	Lys^{12}	Glu^{101}
I1	Gly^{18}	Trp^{82}
I1	Ser^{20}	Lys^{98}
I1	Ala^{23}	Gly^{50}
I1	Ala^{23}	Val^{51}
I1	Asp^{24}	Lys^{47}
I1	Asp^{24}	Ser^{54}
I1	Asn^{25}	Val^{51}
I1	Asn^{25}	Asn^{53}
I1	Gly^{26}	Val^{51}
I2	Lys^{108}	Tyr^{49}
I2	His^{111}	Glu^{101}
I2	Glu^{112}	Tyr^{49}
I2	Gln^{115}	Asn^{79}
I2	Ala^{118}	His^{205}

регулиращ фактор 1), което води до апоптоза. Обратно, при ниска клетъчно-мембранна експресия на IFNγRβ се наблюдава бавно активиране на STAT1, ниско ниво на IRF1 и усилена пролиферация [2].

Установено е, че единствената роля на β-веригата на рецептора е да осигури близостта на асоциираната към нея Jak-2 до комплекса, за да може да се фосфорилира Tyr^{440} от α рецептора [28]. Само близката до клетъчната мембрана половина от вътреклетъчната верига на IFNγRβ е достатъчна за провеждане на сигнала. В молекулата на този рецептор не се фосфорилират тирозинови остатъци. Jak-2 се асоциира към две близки последователности – Pro^{263}-Pro-Ser-Ile-Pro^{267} и Ile^{270}-Glu-Glu-Tyr-Leu^{274} [28].

1.3 Сигнална каскада

Класическият модел на сигналната каскада, която запуска интерферон гама, е представена на фиг. 5A. Първата стъпка е взаимодействието между цитокина и две hIFNγR рецепторни единици. В резултат се образува симетричен сигнален комплекс, което активира асоциираните към рецепторите Jak кинази. Веднъж активирани, те фосфорилират критична тирозин-съдържаща аминокиселинна последователност (а.к. 440 - 444 на IFNγRα), намираща се в близост до C-края на рецепторите. С местата на фосфорилиране се свързват два латентни до този момент STAT1 белтъка, като SH2 домейна на всеки от тях разпознава Tyr-фосфорилираните последователности. След това двата STAT белтъка образуват хомодимер, който се транслокира в ядрото и се свързва със специфични GAS (gamma-activated sequence) елементи на интерферон-индуцируемите гени, стимулирайки тяхната транскрипция.



Фигура 5: Сигнална каскада на hIFN-γ. A) Класически модел на Jak/STAT сигнална пътека. B) Алтернативен модел, при който целият комплекса (IFN-γ–IFNγRα–Jak1–Ja2–STAT1 димер) се транслокира в клетъчното ядро посредством NLS веригата в C-опашката на цитокина. Графиката е от [32].

За да се транслокира в ядрото, STAT-хомодимерът трябва да притежава специфичен сигнал за ядрена локализация (NLS). Белтъците STAT обаче нямат такива последова-

телности, а моделът не може да обясни как STAT димерът влиза в ядрото. Освен това Jak/STAT пътеката се използва и от други цитокини и растежни фактори. Например IL-10 също използва STAT1. Този модел на сигнализацията не може да даде обяснение как активирани от различни цитокини STAT1 димери знаят кои гени да стимулират.

Експериментално е показано, че IFN-γ , hIFNγRα и Jak киназите се транслокират в ядрото. Освен това е установено, че hIFNγRα действа като транскрипционен/ко-транскрипционен фактор, а киназите имат епигенетична функция и фосфорилират ключова тирозинова аминокиселина (Tyr^{41}) в молекулата на хистон 3 (H3) [32].

Предложен е алтернативен модел на сигналната каскада [32], който отново започва с взаимодействието на цитокина с α-рецепторите, асоциирането на β-рецепторите към комплекса, активиране на Jak киназите, фосфорилиране на Tyr^{440} в молекулите на IFNγRα, скачване на латентни STAT1 мономери, последвано от тяхното фосфорилиране и димеризация. Следващият етап в този модел е ендоцитозата на така формирания комплекс. Молекулите на IFNγRβ се дисоциират от комплекса и излизат отново на клетъчната повърхност. Останалата част от комплекса, IFN-γ–IFNγRα–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 [32]. Този модел предлага директно участие на IFN-γ и IFNγRα в активацията на различни гени.

1.4 Глюкозаминогликани

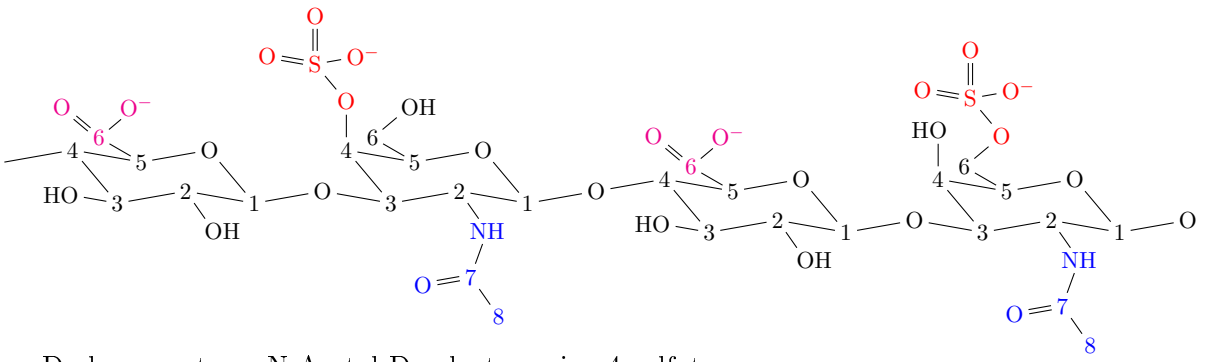
Глюкозаминогликаните (ГАГ) са неразклонени полизахаридни вериги, съставени от повтарящи се дизахаридни единици. Единият монозахарид винаги е аминозахар, която често е сулфатирана, а вторият обикновено е уронова киселина (UA). Тъй като захаридите съдържат сулфатни или карбоксилни групи, ГАГ са силно отрицателно заредени и са най-анионните молекули, които животинските клетки могат да синтезират. ГАГ се делят на шест основни групи в зависимост от захаридите, които ги изграждат, гликозидна връзка между тях, както и броя и местата на сулфатните групи. Галактозамин-съдържащите ГАГ са хондроитин 4-сулфат, хондроитин 6-сулфат и дерматан сулфат. Глюкозамин-съдържащите ГАГ включват хиалуронан, хепарин, хепаран сулфат и кератан сулфат, който вместо уронова киселина съдържа галактоза (фиг. 6) [33].

ГАГ се свързват ковалентно с протеини, наречени **сърцевина (core protein)**, и формират т. нар. протеогликани (ПГ). Те са закачени по клетъчната мембрана, където формират извънклетъчна обвивка и участват в извънклетъчния матрикс. ПГ играят важна роля в биохимичната сигнализация между клетките [33]. Те свързват и регулират активността на протеолитични ензими и техните инхибитори. Освен това ПГ взаимодействат с множество сигнални молекули, включително цитокини и растежни фактори, и могат да модулират активността им. Някои от плазмените ПГ могат да служат като ко-рецептори, които подпомагат конвенционалните клетъчни рецепторни протеини [34–36].

ПГ осъществяват регулация по един от следните начини: (1) имобилизират протеина в близост до мястото му на производство и така ограничава обхвата на действието му; (2) блокират стерично активността му; (3) действат като резервоар и забавят действието на протеина; (4) предпазват протеина от протеолитично разграждане; (5) повишават концентрацията на протеина и спомагат за неговата олигомеризация, оптимизирайки взаимодействието му с рецепторите [33].

Например хепаран сулфатните вериги на ПГ се свързват с фибробластен растежен фактор, който стимулира пролиферацията на много типове клетки. Това взаимодействие води до олигомеризацията на молекулите на растежния фактор и позволява формирането на

кръстосани връзки между тях и последващото активиране на клетъчните им рецептори. Известно е, че ГАГ взаимодействат още с гранулоцит/макрофаг колония-стимулиращ фактор, моноцит колония-стимулиращ фактор, интерлевкин 1β, интерлевкин 3, интерлевкин 8, макрофаген инфламаторен протеин, тромбоцитен растежен фактор, туморен некрозен фактор α и интерферон гама. Със специфични ГАГ се свързват още с липопротеини с ниска плътност, липопротеин липаза, тромбин, антитромбин и тромбомодулин [37,38].

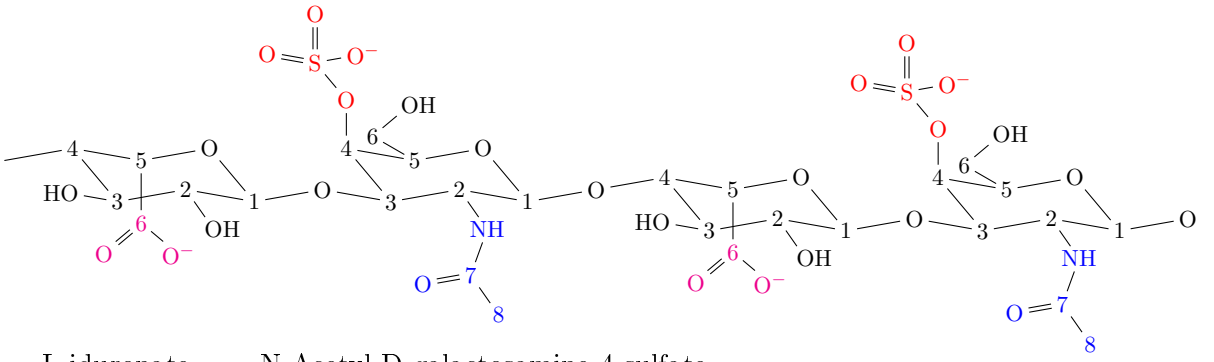


D-glucoronate N-Acetyl-D-galactosamine-4-sulfate D-glucoronate N-Acetyl-D-galactosamine-4-sulfate

$\left(\text{–4) D-GlcA (1→3) D-GalNAc(4S) (1–}\right)_n$

$\left(\text{–4) D-GlcA (1→3) D-GalNAc(6S) (1–}\right)_n$

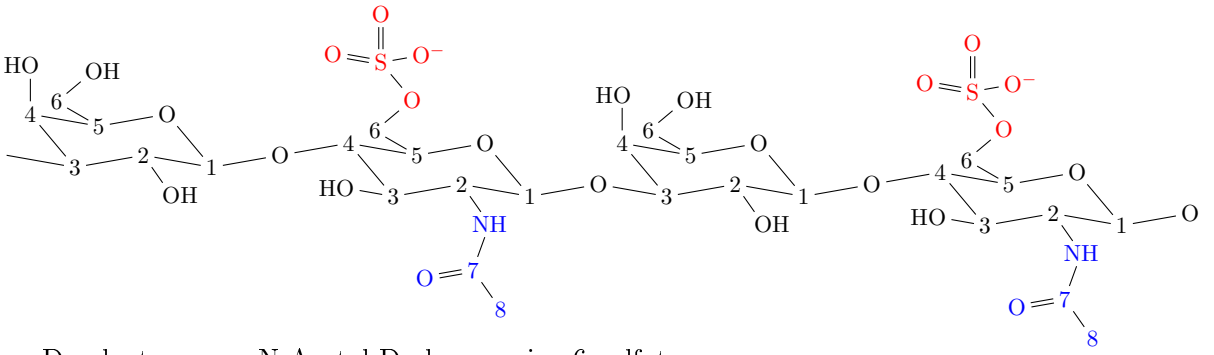
(а) Хондроитин сулфат



L-iduronate N-Acetyl-D-galactosamine-4-sulfate

$\left(\text{–4) L-IdoA (1→3) D-GalNAc(4S) (1–}\right)_n$

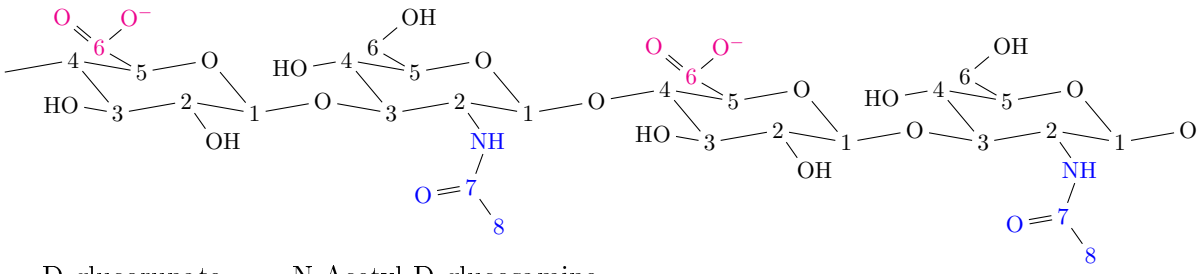
(б) Дерматан сулфат



D-galactose N-Acetyl-D-glucosamine-6-sulfate

$\left(\text{–3) D-Gal (1→4) D-GlcNAc(6S) (1–}\right)_n$

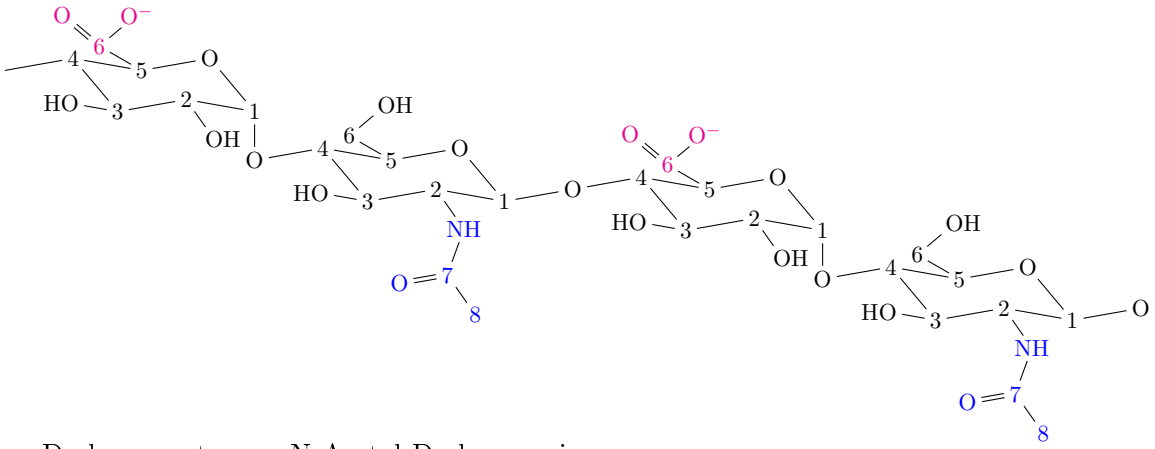
(в) Кератан сулфат



D-glucorunate N-Acetyl-D-glucosamine

$\left(\text{–4) D-GlcA (1→3) D-GlcNAc (1–}\right)_n$

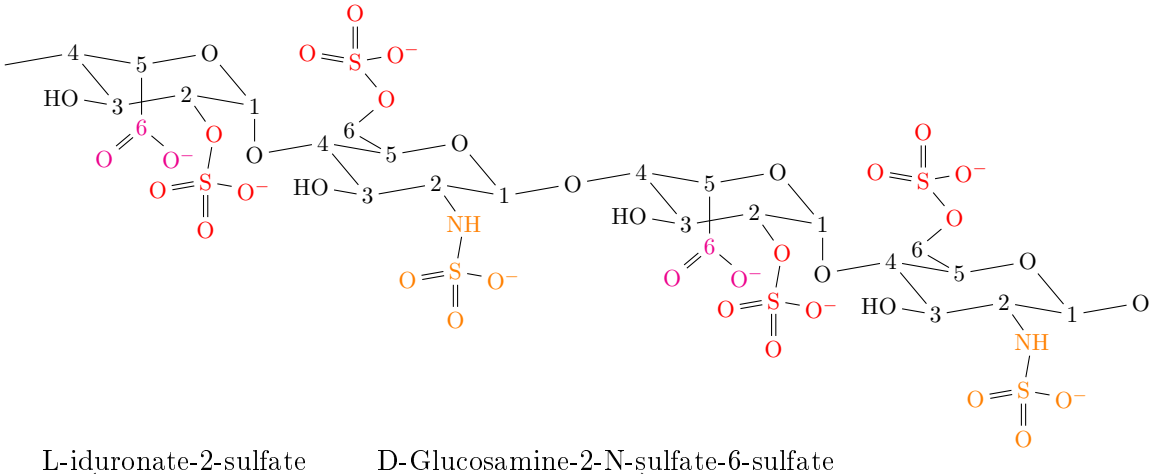
(г) Хиалуронан



D-glucoronate N-Acetyl-D-glucosamine

$\left(\text{–4) D-GlcA (1→4) D-GlcNAc (1–}\right)_n$

(д) Хепаран сулфат, NA-области



L-iduronate-2-sulfate D-Glucosamine-2-N-sulfate-6-sulfate

$\left(\text{–4) L-IdoA(2S) (1→4) D-Glc2NS(6S) (1–}\right)_n$

(е) Хепарин / Хепаран сулфат, NS-области

Фигура 6: Глюкозаминогликани. Карбоксилните групи са оцветени в лилаво, ацетилните – в синьо, сулфатните – в червено и N-сулфатните – в оранжево.

1.4.1 Хепаран сулфат

Хепаран сулфат (ХС) протеогликаните 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→4 свързана последователност от глюкозамин и уронова киселина [39]. Глюкозаминът е или N-ацетилиран (GlcNAc) или N-сулфатиран (Glc2NS), като последната група се намира само в ХС и хепарин. Аминозахарта може да бъде естер(O)-сулфатирана при C6. Уроновите остатъци са представени или като глюкуронова киселина, или като идуронова киселина, която може да бъде O-сулфатирана при C2 (фиг. 6д,е) [39].

Понеже всички съставящи захари на ХС могат да бъдат O-сулфатирани, а Glc2NS може да бъде ди-O-сулфатирана (при C3 и C6, фиг.6е), теоретично това прави общо десет различни монозахарида. В естествените полимери обаче не се наблюдава такова разнообразие и са установени следните ограничения на комбинациите от захари:

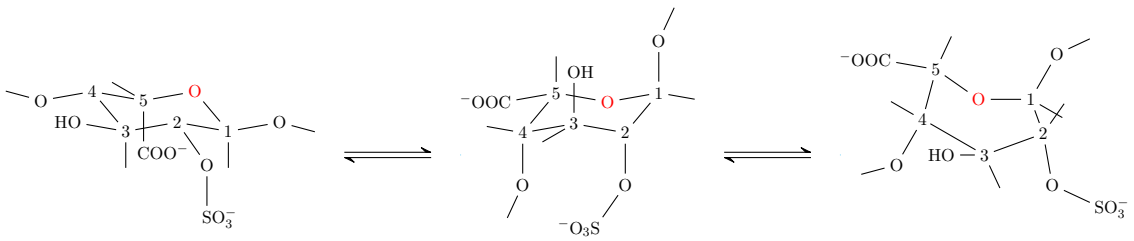
- IduA захари се закачат само за C1 на Glc2NS захари (т.е. Glc2NS (1→4) IduA), но няма такова ограничение за GlcA;

- O-сулфатни групи имат изключително дизахариди, закачени директно за Glc2NS. O-сулфатирани производни на GlcA и IduA се намират предимно в непосредствена близост до N-сулфатирани групи в следните последователности:

  - Glc2NS (1→4) GlcA(2S);
  - Glc2NS (1→4) UA (1→4) GlcNAc(6S)

Тези ограничения означават, че идуроновите остатъци и O-сулфатните групи до голяма степен са съсредоточени в райони, богати на N-сулфатирани глюкозамини (като O-сулфатите не са раздалечени от N-сулфатите на повече от един монозахарид). В същото време разпределението на ацетилирани глюкозамини корелира с ниска степен на сулфатация и висока честота на GlcA [39]. По тази причина ХС ГАГ имат специфична структура с разделени клъстери, които или са несулфатирани (NA-области, фиг. 6д), или сулфатирани (NS-области, фиг. 6е). Това разделяне обаче не е абсолютно и около 30% от веригите на ХС ГАГ са смесени, с редуващи се N-ацетилирани и N-сулфатирани дизахариди [39]. Най-често се срещат несулфатираните последователности на дизахарида GlcA-GlcNAc, разделени от къси NS-сулфатирани сегменти от два до дванадесет дизахарида, последвани от 16 до 18 смесени NA/NS дизахарида.

Идуроновата киселина има необичайно гъвкав пиранозен пръстен. Това оказва съществено влияние върху конформационните и динамични свойства на полимера. Останалите галакто- и гликопиранозни монозахариди имат стабилна конформация "$^4C_1$ стол", но IdoA може да заема няколко конформации - $^4C_1$ стол, $^1C_4$ стол и усукана $^2S_0$ вана (фиг. 7). Равновесието между конформациите стол и усукана $^2S_0$ вана зависи от 2-O-сулфатацията на идуроновата киселина, както и от заместванията в съседните глюкозамини. Въпреки населването на различни конформации на пръстена, експериментални данни показват една конформация на гликозидната връзка със съседните захари. Следователно преходът стол–вана се осъществява без да се предизвиква огъване на полизахаридната верига [39]. Това означава, че гликозидна връзка има сравнително твърда конформация и поддържа пръчковидната форма на полимера. Интересно е, че всички специфични протеин-свързващи ГАГ последователности съдържат поне един идуронов остатък. Вероятно конформационната подвижност на киселината играе важна роля в свързването [34].



Фигура 7: Конформации на пиранозния пръстен на идуронова киселина

1.4.2 Хепарин

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 сулфатиран (над 80% от глюкозамините са N-сулфатирани) и винаги има моларен излишък на O- над N-сулфати (фиг. 6е), като средно един дизахарид има по 2.7 сулфатни групи [34, 39, 40]. 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 [39].

Хепаринът е най-подробно изследваният ГАГ и често се използва като моделно съединение за изследване на свойствата на останалите ГАГ и техните взаимодействия с други биологични молекули. Това се дължи на няколко причини, включително защото се получава лесно, съставът и структурата му се характеризират относително по-лесно в лабораторни условия от другите ГАГ и се свързва с висок афинитет към почти всички протеини, които взаимодействат с извънклетъчен ХС при физиологични условия [39].

Хепаринът се свързва стабилно с някои протеини, като основен принос за това имат електростатични взаимодействия между силно отрицателно заредените захариди на хепарина и базични аминокиселинни остатъци по повърхността на белтъците. Типичен пример за доминирано от електростатиката взаимодействие на хепарина е това с антитромбин. То обаче изисква допълнително в олигозахаридната верига да присъства необичаен пентазахарид, съдържащ специфичен монозахарид (N-сулфатиран,3,6-O-сулфатиран глюкозамин), който има висок афинитет към антитромбин. Свързването води до конформационни изменения в молекулата на протеина, които го правят по-ефективен инхибитор на активирани фактори в коагулационната система на плазмата [39, 40].

Едно от изключенията е взаимодействието с фибробластен растежен фактор (FGF), като хепаринът повишава афинитета на белтъка към рецептора и участва във формирането на рецепторния комплекс на клетъчната повърхност. Това конкретно взаимодействие не е изцяло електростатично, значителна част от свободната енергия на свързване се дължи на некулонови взаимодействия, включително водородни връзки и ван дер Ваалсови взаимодействия [34, 40]. Кристалографската структура на FGF в комплекс с хепаринов хексазахарид показва, че плътно свързана 2-сулфат-идуронова киселина, която заема конформацията усукана $^2S_0$ вана, а слабо взаимодействащ идуронов остатък е в стандартната конформация $^4C_1$ стол [34].

1.4.3 Консенсусна последователност на свързване на протеини с ГАГ

Въз основата на анализ на множество кристалографски структури на ГАГ-протеинни комплекси и експерименти със синтетични пептиди е установено, че съществува консенсусна аминокиселинна последователност, свързваща дългите, хидрофилни и силно отрицателно заредени ГАГ вериги на ПГ. Тези физични особености на ГАГ обуславят ролята им в свързването на протеини с положително заредени домейни, богати на лизини и аргинини.

Пептиди, богати на аргинини и лизини, но не и хистидини, се свързват с най-висок афинитет към хепарин. Пептиди с висок афинитет към ХС съдържат също и много полярни а.к., включително серини и глутамини. Освен това ГАГ свързващите домейни са богати на остатъци, които могат да образуват водородни връзки с хидроксилните групи на захаридите, важни за специфичността на свързването. Аргинин се свързва по-силно с хепарин в сравнение с лизин, което се дължи на по-здравата Н-връзка между гуанидиновата група в страничната верига на аргинин и сулфатните групи на хепарин [34].

Консенсусните последователности за свързване на хепарин имат вида XBBXBX или XBBBXXBX, където B е лизин или аргинин, а X-овете в консенсусната последователност са определени въз основа на анализ на честотата появяване на даден остатък на конкретната позиция в анализираните хепарин-белтъчни комплекси. Аминокиселините Asn, Ser, Ala, Gly, Ile, Leu, и Tyr се срещат по-често на позиции X, докато Cys, Glu, Asp, Met, Phe, и Trp изключително рядко. Установена е и трета, по-дълга консенсусна последователност, която се наблюдава в някои растежни фактори. Тя не е линейна и има вида TXXBXXTBXXXTBB, където T е извивка на гръбнака на протеина [34].

Оптималните за свързване на ХС последователности са от вида BBXBB. Възможно е хидрофобни аминокиселини да взаимодействат хидрофобно е N-ацетилните групи на несулфатираните домейни на ХС. Установено е, много протеини, включително IFN-γ , тромбоцитен растежен фактор 4, IL-8, RANTES и ендостатин, взаимодействат с ХС ГАГ, които се състоят от две NS-области, разделени от дълга NA-верига [34].

1.4.4 IFN-γ и ГАГ

IFN-γ има три базични последователности, наподобяващи консенсусните вериги за свързване на ГАГ – $K^{87}KKR^{90}$, $K^{126}TGKRKR^{132}$ и $R^{138}GRR^{141}$. 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) [38]. 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 [43] *In vivo* свързването с ХС контролира кръвния клирънс, последващото насочване към тъкани и локалната акумулация на интерферон гама. Освен това, когато се свърже с цитокина, ХС ограничава степента на протеолитично разграждане на С-краищата до по-малко от 10 а.к. остатъка и съответно намалява или повишава активността на протеина [38]. Предполага се, че скоро след секретитрането, IFN-γ се захваща от ХС ГАГ в близост до секретиращите клетки, осигурявайки локално действие на цитокина. Интерферонови молекули, които не са свързани с ХС, бързо се инактивират от протеази, което не допуска нежелана дифузия на цитокина [41].

Експерименталните данни показват, че във взаимодействието с ХС участват двете С-крайни базични вериги D1 и D2 [41]. 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 [43].

IFN-γ-свързващите XC 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 XC олигозахарида от 11.5 nm [42].

IFN-γ взаимодейства и с друг ГАГ - хондроитин сулфат (CS), който е вторият най-често разпространен ГАГ след XC и е основен компонент на извънклетъчния матрикс в артериите [37]. Афинитетът на това свързване е среден ($K_d \approx 90$ nM). Експерименти със синтетични пептиди показват, че взаимодействието на IFN-γ и CS отново се дължи до голяма степен на D1 домейна. Освен това обаче и заредената последователност $K^{87}KKR^{90}$ в горната NLS верига на цитокина има известен принос, докато D2 домейна не допринася за свързването [37].

Глава 2

Молекулно моделиране на биологични макромулекули

2.1 Молекулна динамика

Тъй като изследването на биологични макромолекули включва огромен брой атоми, не е възможно, а в повечето случаи не е и необходимо, тези системи да се разглеждат квантово. 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 [44, 45]. 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 междуядрените взаимодействия. Това класическо описание може да се използа при следните условия:

- приближението на Борн-Опенхаймер е валидно;
- не се изследва електронната структура;
- не се формират и разкъстват ковалентни връзки;
- електроните са локализирани (т.е. неподходящо за метали)

Микросъстоянието на системата се нарича конфигурация $(\{\mathbf{r}_i\}, \{\mathbf{v}_i\})$, $i = 1, \ldots, N$, където N е броят на атомите в системата, която се описва от набор 3N координати и 3N скорости. Потенциалната енергия зависи само от координатите, а кинетичната – само от скоростите. Пълната енергия на системата се запазва и се задава от Хамилтониана:

$$H(\{\mathbf{r}_i\}, \{\mathbf{v}_i\}) = K(\{\mathbf{v}_i\}) + U(\{\mathbf{r}_i\}) = \sum_i^N \frac{m_i |\mathbf{v}_i|^2}{2} + U(\{\mathbf{r}_i\}) \tag{1}$$

Еволюцията във времето на системата от N частици се определя чрез интегриране на уравненията на движение на Нютон:

$$m_i \frac{d^2\mathbf{r}_i}{dt^2} = \mathbf{F}_i \tag{2}$$

Силите са отрицателните производни по координатите на потенциалната енергия:

$$\mathbf{F}_i = -\nabla_i U\left(\{\mathbf{r}_i\}\right) \tag{3}$$

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. 8. Отделните стъпки от нея са описани по-подробно в следващите секции.

.

2.1.1 Молекулно-механично силово поле

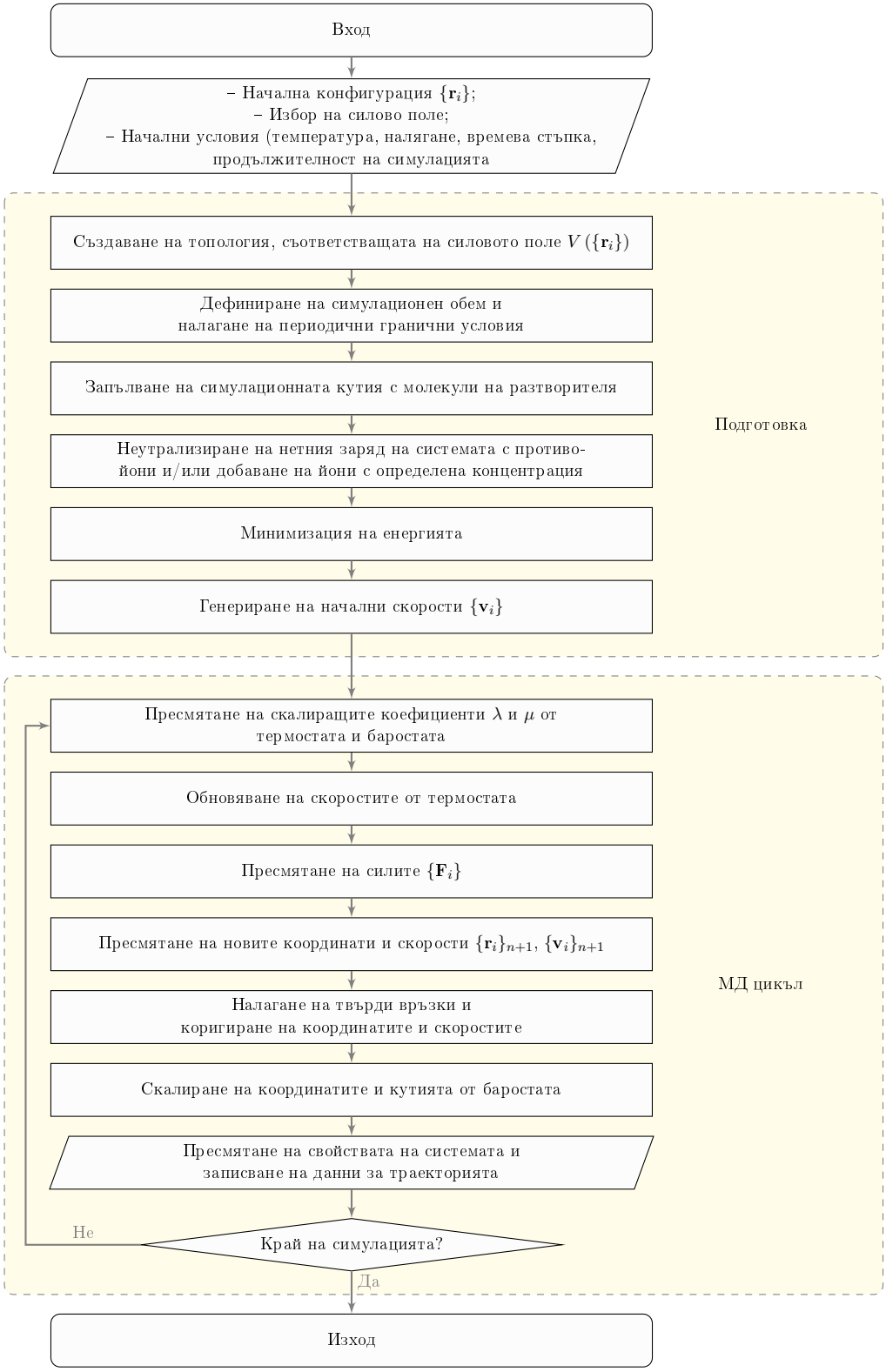
За да се реши уравнението на движение на Нютон, трябва да са известни силите, действащи на всеки атом. Силите са отрицателния градиент на потенциалната енергия (3). Затова изборът на потенциална функция е от изключително значение за МД симулациите. За симулации на биологични молекули се използват полу-емпирични 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, например чрез рентгенова кристалография, ЯМР спектроскопия и нееластично разсейване на неутрони [46].

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 двучастични 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:

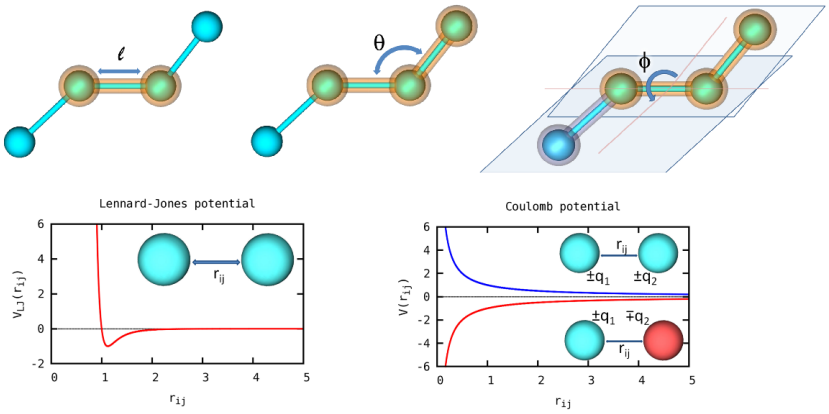
$$\begin{aligned} U\left(\{\mathbf{r}_i\}\right) = &\sum_{bonds} K_l\left(l-l_0\right)^2 + \sum_{angle} K_\theta\left(\theta-\theta_0\right)^2 \sum_{dihedrals} K_\phi\left(1+cos\left(n\phi-\delta\right)\right) \\ &+ \sum_{i=1}^{N}\sum_{j=i+1}^{N}\left(4\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12}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

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. Различните членове на това уравнение са илюстрирани на фиг. 9.

Първият член в (4) 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, т.е. зависи от химичния характер на атомите, участващи във взаимодействието. Обикновено K_l и l_0 се параметризират въз основа на *ab initio* изчисления.



Фигура 8: Схема на МД симулация



Фигура 9: Ковалентни и нековалентни членове на силовото поле.

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 е потенциалната функция на торзионния ъгъл, която описва наличието на стерични бариери между 1,4-двойки атоми, разделени от три ковалентни връзки. Движението, свързано с този член, е ротация с порядък на симетрия n=1,2,3 около средната връзка. Този потенциал трябва да е периодичен, поради което се изразява като косинусова функция. За параметризация на K_{phi}, ϕ_0, n и δ обикновено се използват параметри от квантови симулации, които после се донагласят чрез фитиране към експериментални данни.

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 2.1.3). За определянето на Ван дер Ваалсовите параметри и частичните заряди често се използват квантови пресмятания и фитиране към експериментални данни. Основен проблем при пресмятането на нековалентните взаимодействия е, че те включват всички двойки атоми в системата и броят им расте като $\mathcal{O}(N^2)$. Затова за тяхното пресмятане са разработени специални методи, разгледани в следващите секции.

Важно е да се отбележи, че силовото поле представлява не само аналитичната функция на потенциалната енергия (4). То включва всички параметри, които участват в нея за всички т. нар. “типове атоми”, поддържани от полето. По този параметър СП се делят на следните видове:

- СП с описание на всички атоми (all atom force fields), при които има отделни параметри за всеки атом в системата. Типовете атоми не се изчерпват с вида на атома, а зависят от химичното обкръжение на съответния атом. Например някои полета поддържат различни типове атоми за различните хибридизации на въглеродните атоми;

- СП с обединени атоми (united atom force fields), при които неполярните водородни атоми се обединяват с атома, за който са свързани, в един нов тип обединен атом с ефективни Ван дер Ваалсови параметри и частичен заряд;

- едрозърнести СП (coarse grained force fields), при които молекулата се разглежда по-обобщено и в по-малко детайли. Цели функционални групи или групи от атоми се

обединяват в обобщени частици, наречени зърна, които се характеризират с някакви ефективни параметри.

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. Важно е топологията да съответства на структурата, което означава, че самата структура също трябва да бъде преработена и да съдържа видовете атоми (например обединени атоми), съответстващи на използваното силово поле.

2.1.2 Ван дер Ваалсово взаимодействие

Потенциалът на Ленард-Джоунс (предпоследният член в (4)) се използва за описание на Ван дер Ваалсово взаимодействие. Ленард-Джоунс параметрите ε_{ij} и σ_{ij} са специфични за всяка двойка типове атоми и се получават от правилата за комбиниране на Лоренц-Бертло: $\varepsilon_{ij} = \sqrt{\varepsilon_{ii}\varepsilon_{jj}}$ и $\sigma_{ij} = 1/2(\sigma_{ii} + \sigma_{jj})$.

И двата члена на потенциала зависят силно от разстоянието и са бързо сходящи. Ето защо за пресмятане на Ван дер Ваалсовото взаимодействието може да се използва обрязване на потенциала по сферична повърхност с център в i-тия атом и радиус r_{cutoff}. За да се елиминира прекъсването на потенциала при $r = r_{cutoff}$ и за да имаме непрекъснати производни в тази точка, към силите се добавя изглаждаща функция $S(r)$, която има за цел да направи $U_{LJ}(r_{cutoff}) = 0,\ F_{LJ}(r_{cutoff}) = 0$ [47]. Изгладените сили имат вида:

$$\begin{aligned} F_s(r) &= F_\alpha(r) && r < r_1 \\ F_s(r) &= F_\alpha(r) + S(r) && r_1 \le r < r_{cutoff} \\ F_s(r) &= 0 && r_{cutoff} \le r \end{aligned} \tag{5}$$

където $F_\alpha(r) = r^{-(\alpha+1)}$, α е степенният показател на взаимодействието ($\alpha = 6$ за дисперсионната и $\alpha = 12$ за отблъскващата компонента на потенциала на Ленард-Джоунс), а r_1 е разстоянието, от което започва изглаждането.

Изглаждащата функция има вида:

$$\begin{aligned} S(r) &= A(r - r_1)^2 + B(r - r_1)^3 \\ A &= \frac{(\alpha+4)\, r_{cutoff} - (\alpha+1)\, r_1}{r_{cutoff}^{\alpha+2}\,(r_{cutoff} - r_1)^2} \\ B &= \frac{(\alpha+3)\, r_{cutoff} - (\alpha+1)\, r_1}{r_{cutoff}^{\alpha+2}\,(r_{cutoff} - r_1)^3} \end{aligned} \tag{6}$$

Пълните изгладена сила и потенциал са:

$$\begin{aligned} F_s(r, r_1, r_{cutoff}) &= \frac{1}{r^{\alpha+1}} + A(r - r_1)^2 + B(r - r_1)^3 \\ U_s(r, r_1, r_{cutoff}) &= \frac{1}{r^\alpha} - \frac{A}{3}(r - r_1)^3 - \frac{B}{4}(r - r_1)^4 - C \\ C &= \frac{1}{r_{cutoff}^\alpha} - \frac{A}{3}(r_{cutoff} - r_1)^3 - \frac{B}{4}(r_{cutoff} - r_1)^4 \end{aligned} \tag{7}$$

2.1.3 Електростатично взаимодействие

Електростатичното взаимодействие е далеко-действащо и бавно сходящо, което не позволява просто обрязване на потенциала както при Ван дер Ваалсовото взаимодействието.

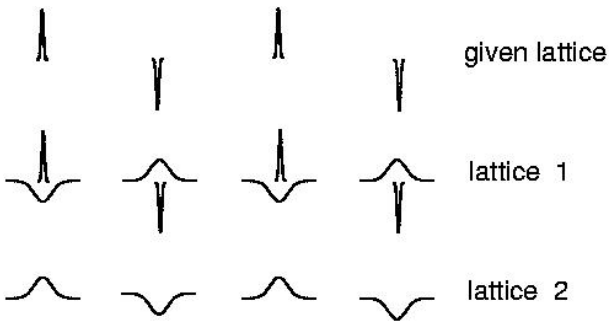
Затова в МД симулациите за пресмятане на електростатиката основно се ползват методи, базирани на сумите на Евалд – техника, разработена за пресмятане на електростатичното взаимодействие в периодични и псевдо-периодични системи. Идеята е бавно сходящата сума в последния член на уравнение (4) да се преобразува 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:

$$\frac{1}{r} = \frac{f(r)}{r} + \frac{1 - f(r)}{r} \tag{8}$$

като се избира $f(r) = \mathrm{erfc}(\beta r)$. Тук β е параметър, който определя относителното тегло на правата и реципрочната сума. 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 сумата 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 (фиг. 10). 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:

$$\begin{aligned}
U_{el.stat.} &= U_{dir} + U_{rec} + U_0 \\
U_{dir} &= \frac{1}{4\pi\varepsilon_0}\frac{1}{2}\sum_{\mathbf{n}}\sum_{i,j}^{N}\frac{q_i q_j}{|\mathbf{r}_{ij} + \mathbf{n}L|}\mathrm{erfc}\left(\beta\,|\mathbf{r}_{ij} + \mathbf{n}L|\right) \\
U_{rec} &= \frac{1}{2V\varepsilon_0}\sum_{\mathbf{k}\neq\mathbf{0}}\sum_{i,j}^{N}\frac{q_i q_j}{k^2}\exp\left(\frac{-k^2}{4\beta^2}\right)\exp\left(-i\mathbf{k}\cdot\mathbf{r}_{ij}\right) \\
U_0 &= -\frac{1}{4\pi\varepsilon_0}\frac{\beta}{\sqrt{\pi}}\sum_{i}^{N} q_i^2
\end{aligned} \tag{9}$$

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



Фигура 10: Метод на Евалд – директна и реципрочна суми

Бързата сходимост на сумите позволява да се използват малки радиуси на обрязване в реалното пространство (от порядъка на нанометри) и в реципрочното пространство (около 10 вълнови вектора във всяка посока), но дори при оптимизирани алгоритми пресмятането на реципрочната сума расте като $\mathcal{O}(\mathrm{N}^{3/2})$, което не е практично за големи системи.

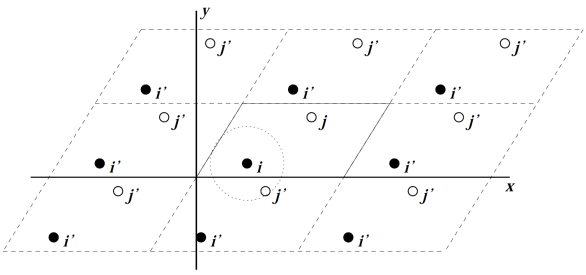
Методът PME (Particle Mesh Ewald) [48] подобрява ефективността на изчисление на реципрочната сума. Вместо директно да се сумират вълнови вектори, гаусовите заряди се разпределят по решетка, като се използва B-сплайн интерполация. След това се прави Фурие образ на решетката с помощта на алгоритъм за бързо 3D Фурие преобразувание (FFT) и реципрочната сума се получава като единична сума по решетката в k-пространството. Потенциалът във възлите на решетката се смята като обратно преобразувание и от коефициентите на интерполацията се получават силите, действащи върху всеки атом. Алгоритъмът PME изисква $\mathcal{O}(N \log N)$ пресмятания, което го прави предпочиран за симулации на средни и големи молекулни системи.

2.1.4 Периодични гранични условия

За да се запази макроскопичния характер на системата, която се изследва, е необходимо граничните повърхности на системата да се третират по правилен начин. Колкото и голяма система да симулираме, броят на частиците в нея винаги ще е много по-малък от термодинамичната граница и отношението на частиците от повърхността към общия брой ще е много голямо, което води до завишена оценка на повърхностните ефекти.

Решението на този проблем е да се наложат периодични гранични условия – дефинира се централна симулационна кутия, която след това се възпроизвежда чрез транслация по трите пространствени оси до безкрайност, изпълвайки цялото пространство. Ако по време на симулацията някоя частица излезе от симулационната кутия, периодичният ѝ образ във всяка от съседните кутии ще се движи по абсолютно същия начин и следователно един от съседните образи ще влезе в централната симулационна кутия. Така броят на атомите в централната кутия се запазва постоянен, елиминират се граничните повърхности и съответно повърхностните ефекти.

Налагането на периодични гранични условия води до рязко увеличаване на броя на взаимодействащи си частици. Този проблем може да се реши чрез прилагането на т.нар. конвенция за най-близкия образ (фиг. 11) - отделните частици взаимодействат само с най-близките образи на останалите частици в системата.



Фигура 11: Периодични гранични условия и конвенцията за най-близкия образ. Атом i си взаимодейства не с атом j, а с най-близкия му образ j'. Графиката е от [49].

2.1.5 Минимизация на енергията

Често енергията на началната конфигурация е доминирана от няколко неблагоприятни Ван дер Ваалсови контакта. Тогава силите между част от атомите може да се окажат прекалено големи, което да доведе до огромни измененния в скоростите и координатите на тези атоми. На следващата интеграционна стъпка тези атоми може да се ударят в други и да доведат до каскада от огромни сили и неконтролируемо движение на частици, при което в крайна сметка МД симулацията да се провали. Ето защо преди действителната МД е необходимо да

се направи минимизация на енергията на конформацията и геометрията на молекулата да се оптимизира, докато попадне в най-близкия до началната конформация локален минимум.

Най-често за минимизация се използват градиентни алгоритми. При метода на най-бързото спускане координтите се преместват в посока, обратна на градиента на потенциалната енергия. Единственият параметър, който се задава, е максималната стъпка, с която се преместват координатите. Стъпката обикновено се наглася динамично за по-бърза сходимост. Всяка следваща посока на спускане е ортогонална на предишната. Алгоритъмът е много бърз и стабилен, когато началната конформация е далече от минимум. Основните недостатъци на този метод са, че не е много ефективен и може да проявява осцилиращо поведение, когато системата е близо до минимума.

Когато сме близо до минимум по-подходящо е да се използва методът на спрегнатите градиенти. Алгоритъмът започва минимизацията по посоката на най-бързото спукане и продължава в тази посока, докато се достигне минимум по нея, след което продължава по посока, перпендикулярна на нея. Ако минимумът има формата на дълга, тясна долина, по този метод минимумът може да бъде достигнат с много по-малко стъпки, отколкото с метода на най-бързото спускане [45].

2.1.6 Инициализация

За да започне интегрирането на уравненията на движение са необходими начални условия. Началните координати обикновено се взимат от база данни на структури, получени чрез дифракция на рентгенови лъчи или ЯМР спектроскопия (например Brookhaven Protein Data Bank [31]), или чрез хомоложно моделиране.

Началните скорости се генерират чрез даване на случайни скорости на всеки атом така, че пълният импулс на системата да е нула и средната кинетична енергия правилно да възпроизвежда Максуел-Болцманово разпределение за дадената температура T:

$$p\left(v_i\right) = \sqrt{\frac{m_i}{2\pi k_B T}} exp\left\{-\frac{m_i v_i^2}{2k_B T}\right\} \tag{10}$$

където k_B е константата на Болцман.

2.1.7 Интегратори

Уравненията на движение на частиците се дискретизират на малки времеви стъпки и се решават числено. Развиваме координатите на i-тия атом в ред на Тейлър по степени на Δt:

$$\mathbf{r}\left(t+\Delta t\right) = \mathbf{r}\left(t\right) + \dot{\mathbf{r}}\left(t\right)\Delta t + \frac{1}{2}\ddot{\mathbf{r}}\left(t\right)\Delta t^2 + \frac{1}{6}\dddot{\mathbf{r}}\left(t\right)\Delta t^3 + O\left(\Delta t^4\right) \tag{11}$$

$$\mathbf{r}\left(t-\Delta t\right) = \mathbf{r}\left(t\right) - \dot{\mathbf{r}}\left(t\right)\Delta t + \frac{1}{2}\ddot{\mathbf{r}}\left(t\right)\Delta t^2 - \frac{1}{6}\dddot{\mathbf{r}}\left(t\right)\Delta t^3 + O\left(\Delta t^4\right) \tag{12}$$

Събираме у-ния (11) и (12):

$$\mathbf{r}\left(t+\Delta t\right) + \mathbf{r}\left(t-\Delta t\right) = 2\mathbf{r}\left(t\right) + \ddot{\mathbf{r}}\left(t\right)\Delta t^2 + O\left(\Delta t^4\right) \tag{13}$$

Изваждаме (12) от (11):

$$\mathbf{r}\left(t+\Delta t\right) - \mathbf{r}\left(t-\Delta t\right) = 2\dot{\mathbf{r}}\left(t\right)\Delta t + O\left(\Delta t^3\right) \tag{14}$$

Или:

$$\begin{aligned} \mathbf{r}_i^{n+1} &= 2\mathbf{r}_i^n - \mathbf{r}_i^{n-1} + \frac{\mathbf{F}_i^n}{m_i}\Delta t + O\left(\Delta t^4\right) \\ \mathbf{v}_i^n &= \frac{\mathbf{r}_i^{n+1} - \mathbf{r}_i^{n-1}}{2\Delta t} + O\left(\Delta t^2\right) \end{aligned} \tag{15}$$

Уравнения (15) представляват метода на Верле за интегриране на уравненията на движение. Той е много бърз и има предимството, че не са нужни скорости за пресмятане на координатите. Освен това е обратим по построение. Недостатък е голямата грешка при изчисляване на скоростите, което е проблем при симулации с постоянна температура. На практика в МД симулациите се използват други два интегратора, които са еквивалентни на Верле – методът прескочи-кобила (leap-frog integrator) и Верле със скорости.

Нека дефинираме скорости, отместени с половин времева стъпка:

$$\mathbf{v}_i^{n-1/2} = \frac{\mathbf{r}_i^n - \mathbf{r}_i^{n-1}}{\Delta t}; \quad \mathbf{v}_i^{n+1/2} = \frac{\mathbf{r}_i^{n+1} - \mathbf{r}_i^n}{\Delta t} \tag{16}$$

Заместваме у-ния (16) в у-ния (15) и получаваме интегратора прескочи-кобила, при който пресмятането на координатите и скоростите се разминава с половин стъпка:

$$\begin{aligned} \mathbf{v}_i^{n+1/2} &= \mathbf{v}_i^{n-1/2} + \frac{\mathbf{F}_i^n}{m_i}\Delta t + O\left(\Delta t^3\right) \\ \mathbf{r}_i^{n+1} &= \mathbf{r}_i^n + \mathbf{v}_i^{n+1/2}\Delta t + O\left(\Delta t^3\right) \\ \mathbf{v}_i^n &= \frac{\mathbf{v}_i^{n-1/2} + \mathbf{v}_i^{n+1/2}}{2} \end{aligned} \tag{17}$$

Предимството на този интегратор е по-високата точност при изчисляване на скоростите, а директното им пресмятане позволява да се контролира по-точно температурата. Недостатъкът е, че моментните скорости в цяла стъпка се апроксимират.

Алтернативата е интергграторът на Верле със скорости. От второто уравнение в (15) следва:

$$\mathbf{r}_i^{n-1} = \mathbf{r}_i^{n+1} - 2\mathbf{v}_i^n\Delta t \tag{18}$$

Заместваме го в израза за координатите на Верле и получаваме:

$$\mathbf{r}_i^{n+1} = \mathbf{r}_i^n + \mathbf{v}^n\Delta t + \frac{\mathbf{F}_i^n}{2m_i}\Delta t^2 \tag{19}$$

Отново от второто уравнение в (15) получаваме:

$$\mathbf{v}_i^{n+1} = \frac{\mathbf{r}_i^{n+2} - \mathbf{r}_i^n}{2\Delta t} \tag{20}$$

От интегратора на Верле следва:

$$\mathbf{r}_i^{n+2} = 2\mathbf{r}_i^{n+1} - \mathbf{r}_i^n + \frac{\mathbf{F}_i^{n+1}}{m_i}\Delta t^2 \tag{21}$$

Заместваме това уравнение в (20):

$$\mathbf{v}_i^{n+1} = \frac{\mathbf{r}_i^{n+1} - \mathbf{r}_i^n}{\Delta t} + \frac{\mathbf{F}_i^{n+1}}{2m_i}\Delta t \tag{22}$$

Заместваме (19) в горното уравнение и получаваме:

$$\mathbf{v}_i^{n+1} = \mathbf{v}_i^n + \frac{\mathbf{F}_i^{n+1} + \mathbf{F}_i^n}{2m_i}\Delta t \tag{23}$$

Уравнения (19) и (23) представляват интергратора на Верле със скорости. Той е сравнително бърз, стабилен и обратим метод и до голяма степен запазва енергията на системата,

освен това е самостартиращ се и позволява да се променя времевата стъпка. Недостатъкът му е, че изисква повече памет за изчисленията от метода прескочи-кобила [44]. На практика методът на Верле със скорости за решаване на уравнението на движение за i-тия атом се имплементира най-често така:

$$\begin{aligned} \mathbf{v}_i^{n+1/2} &= \mathbf{v}_i^n + \frac{\mathbf{F}_i^n}{2m_i}\Delta t \\ \mathbf{r}_i^{n+1} &= \mathbf{r}_i^n + \mathbf{v}_i^{n+1/2}\Delta t + O\left(\Delta t^3\right) \\ \mathbf{v}_i^{n+1} &= \mathbf{v}_i^{n+1/2} + \frac{\mathbf{F}_i^{n+1}}{2m_i}\Delta t + O\left(\Delta t^3\right) \end{aligned} \tag{24}$$

2.1.8 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 Δt $<$ 0.5-1.0 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:

$$L^* = K\left(\{\mathbf{v}_i\}\right) - U\left(\{\mathbf{r}_i\}\right) - \sum_\alpha \lambda_\alpha \sigma_\alpha = L\left(\{\mathbf{r}_i\}, \{\dot{\mathbf{r}}_i\}\right) - \sum_\alpha \lambda_\alpha \sigma_\alpha\left(\{\mathbf{r}_i\}\right) \tag{25}$$

където λ_α е множителят на Лагранж за връзката σ_α:

$$\sigma\left(\mathbf{r}_i, \mathbf{r}_j\right) = \frac{1}{2}\left(r_{ij}^2 - d^2\right) = 0; \quad d = const. \tag{26}$$

От у-нията на Лагранж $\frac{d}{dt}\frac{\partial L}{\partial \dot{\mathbf{r}}_i} - \frac{\partial L}{\partial \mathbf{r}_i} = 0$, получаваме модифицираните уравнения на движение:

$$m_i \ddot{\mathbf{r}}_i = -\frac{\partial U}{\partial \mathbf{r}_i} - \sum_\alpha \lambda_\alpha \frac{\partial \sigma_\alpha}{\partial \mathbf{r}_i} = \mathbf{F}_i + \mathbf{G}_i \tag{27}$$

От метода на Верле за изменението на координатите на i-тия атом имаме:

$$\mathbf{r}_i'\left(t + \Delta t\right) = 2\mathbf{r}_i\left(t\right) - \mathbf{r}_i\left(t - \Delta t\right) + \Delta t^2 \frac{\mathbf{F}_i}{m_i}\left(t\right) \tag{28}$$

Силата на реакция на връзката с j-тия атом коригира координатите:

$$\mathbf{r}_i\left(t + \Delta t\right) = \mathbf{r}_i'\left(t + \Delta t\right) + \Delta t^2 \frac{\mathbf{g}_{ij}\left(t\right)}{m_i} \tag{29}$$

където $\mathbf{r}_i'$ е обновената координата на i-тия атом без наложени връзки. Аналогично за j-тия атом имаме:

$$\mathbf{r}_j\left(t + \Delta t\right) = \mathbf{r}_j'\left(t + \Delta t\right) + \Delta t^2 \frac{\mathbf{g}_{ji}\left(t\right)}{m_j} \tag{30}$$

$$\mathbf{r}_{ij} = \mathbf{r}_i - \mathbf{r}_j; \quad \mathbf{g}_{ij} = -\mathbf{g}_{ji} \tag{31}$$

Векторът по направление на химичната връзка в следващата стъпка е:

$$\begin{aligned} \mathbf{r}_{ij}\left(t + \Delta t\right) &= \mathbf{r}_{ij}'\left(t + \Delta t\right) - \Delta t^2 \mathbf{g}_{ij}\left(t\right)\left(\frac{1}{m_j} + \frac{1}{m_i}\right) = \\ &= \mathbf{r}_{ij}'\left(t + \Delta t\right) - \lambda_{ij}\left(t\right)\Delta t^2 \mathbf{r}_{ij}\left(t\right)\left(\frac{1}{m_j} + \frac{1}{m_i}\right) \end{aligned} \tag{32}$$

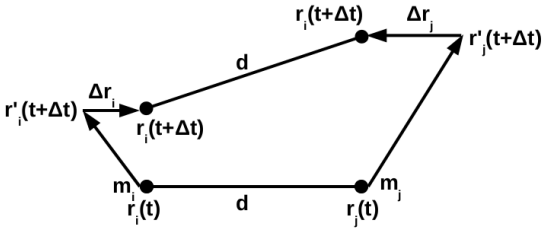
Уравнението на връзките трябва да бъде удовлетворено и в следващата стъпка:

$$\begin{aligned}
&\frac{1}{2}\left(r_{ij}^{2}\left(t+\Delta t\right)-d^{2}\right)=0\\
&\frac{1}{2}\left(\left[\mathbf{r}_{i}^{\prime}\left(t+\Delta t\right)-\mathbf{r}_{j}^{\prime}\left(t+\Delta t\right)+\Delta t^{2}\left(\frac{1}{m_{j}}+\frac{1}{m_{i}}\right)\lambda_{ij}\left(t\right)\mathbf{r}_{ij}\left(t\right)\right]^{2}-d^{2}\right)=0
\end{aligned} \tag{33}$$

Това е квадратно уравнение спрямо множителите на Лагранж. Най-разпространеният метод за налагане на връзки в МД симулациите е алгоритъмът SHAKE [50]. SHAKE пренебрегва квадратичния член по λ, 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.

От горното уравнение за λ получаваме:

$$\lambda_{ij}\left(t\right)=\frac{d^{2}-r_{ij}^{\prime 2}\left(t+\Delta t\right)}{4\Delta t^{2}\left(\frac{1}{m_{i}}+\frac{1}{m_{j}}\right)\left[\mathbf{r}_{ij}\left(t\right)\cdot\mathbf{r}_{ij}^{\prime}\left(t+\Delta t\right)\right]} \tag{34}$$



Фигура 12: Принцип на действие на SHAKE: интеграторът обновява координатите на атоми i и j ($\mathbf{r}_i(t+\Delta t)$ и $\mathbf{r}'_j(t+\Delta t)$), след което реакцията на силите на връзките $\mathbf{g}_{ij}$ ги коригират с $\Delta\mathbf{r}_i$ и $\Delta\mathbf{r}_j$, за да запазят предварително зададената дължина на химичната връзка d.

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 [51]:

$$\begin{aligned}
\mathbf{v}_{i}\left(t+\Delta t\right)=&\mathbf{v}_{i}\left(t\right)+\frac{1}{2}\Delta t\left(\frac{\mathbf{F}_{i}}{m_{i}}\left(t\right)-\frac{2}{m_{i}}\lambda_{ij}\left(t\right)\mathbf{r}_{ij}\left(t\right)\right)+\\
&\frac{1}{2}\Delta t\left(\frac{\mathbf{F}_{i}}{m_{i}}\left(t+\Delta t\right)+\frac{2}{m_{i}}\lambda_{ij}\left(t+\Delta t\right)\mathbf{r}_{ij}\left(t+\Delta t\right)\right)
\end{aligned} \tag{35}$$

SHAKE и RATTLE схождат много бързо, типично в две до три итерации. Използването на химични връзки с фиксирана дължина и агоритмите SHAKE и RATTLE позволяват времевата стъпка, с която интегрираме уравнението на движение, да се увеличи 2-4 пъти.

2.1.9 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 е методът r-RESPA [52].

В МД се изследват системи, които еволюират съгласно уравненията на Хамилтон:

$$\dot{\mathbf{p}} = -\frac{\partial H}{\partial \mathbf{q}}(\mathbf{q},\mathbf{p}) \qquad \dot{\mathbf{q}} = \frac{\partial H}{\partial \mathbf{p}}(\mathbf{q},\mathbf{p}) \tag{36}$$

където $\mathbf{q}$ и $\mathbf{p}$ са обобщените координати и импулси на частиците, а $H(\mathbf{q},\mathbf{p})$ е хамилтонианът на системата. Такива системи са симплектични, т.е. динамиката запазва обема на фазовото пространство и е обратима във времето. Необходимо е интеграторът, който ще се използва, да описва адекватно тези свойства. Интегратори с такива свойства могат да се генерират с помощта на оператора на Лиувил и теоремата на Тротер.

Операторът на Лиувил е линеен ермитов оператор и за система с N_{df} степени на свобода се дефинира като:

$$iL = \{\cdots, H\} = \sum_{j=1}^{N_{df}} \left[\frac{\partial H}{\partial \mathbf{p}_j}\cdot\frac{\partial}{\partial \mathbf{q}_j} - \frac{\partial H}{\partial \mathbf{q}_j}\cdot\frac{\partial}{\partial \mathbf{p}_j}\right] \tag{37}$$

където $\{\cdots,\cdots\}$ са скобките на Поасон. В декартови координати операторът на Лиувил е:

$$iL = \sum_{j=1}^{N_{df}} \left[\frac{\mathbf{p}_j}{m_j}\cdot\frac{\partial}{\partial \mathbf{r}_j} + \mathbf{F}_j\cdot\frac{\partial}{\partial \mathbf{p}_j}\right] \tag{38}$$

Състоянието на системата в момента t е:

$$\Gamma(t) = U(t)\,\Gamma(0) \tag{39}$$

където $\Gamma(t) = \{\mathbf{q}_i(t), \mathbf{p}_i(t)\}$ е наборът от координати и импулси на частиците, а $U(t)$ е класическият пропагатор:

$$U(t) = \mathrm{e}^{itL} \tag{40}$$

Нека разложим оператора на Лиувил на две части:

$$iL = iL_1 + iL_2 \tag{41}$$

Теоремата на Тротер дава:

$$\mathrm{e}^{it(L_1+L_2)} = \left[\mathrm{e}^{i(t/P)(L_1+L_2)}\right]^P = \left[\mathrm{e}^{i(\Delta t/2)L_2}\mathrm{e}^{i(\Delta t)L_1}\mathrm{e}^{i(\Delta t/2)L_2}\right]^P + \mathcal{O}\left(\Delta t^3\right) \tag{42}$$

където $\Delta t \equiv t/P$. Това позволява да се дефинират дискретни пропагатори:

$$G(\Delta t) = U_2(\Delta t/2)\,U_1(\Delta t)\,U_2(\Delta t/2) = \mathrm{e}^{i(\Delta t/2)L_2}\mathrm{e}^{i(\Delta t)L_1}\mathrm{e}^{i(\Delta t/2)L_2} \tag{43}$$

Разлагаме средния пропагатор по следния начин:

$$\mathrm{e}^{i\Delta t L_1} = \left(\mathrm{e}^{i\delta t L_1}\right)^n \tag{44}$$

където $n\delta t = \Delta t$. По този начин може да се пресметне еволюцията на система, която еволюира по две различни времеви скали. Вътрешният пропагатор, наречен референтен, се асоциира с бързо изменящите се взаимодействия и тези степени на свобода се пропагират с по-малка времева стъпка. Външният пропагатор, наречен коригиращ, се използва за еволюиране на по-бавно изменящите се степени на свобода. Състоянието на системата в момента t се определя формално така:

$$\begin{aligned}\Gamma(\Delta t) &= U_2(\Delta t/2)\,U_1(\Delta t)\,U_2(\Delta t/2)\,\Gamma(0) + \mathcal{O}\left(\Delta t^3\right)\\ &= \mathrm{e}^{i(n\delta t/2)L_2}\left[\mathrm{e}^{i(\delta t)L_1}\right]^n \mathrm{e}^{i(n\delta t/2)L_2}\Gamma(0) + \mathcal{O}\left(\Delta t^3\right)\end{aligned} \tag{45}$$

За класическа система в декартови координати с бързо и бавно изменящи се (близки и далечни) взаимодействия разложението на оператора на Лиувил е:

$$iL = \dot{\mathbf{r}} \cdot \frac{\partial}{\partial \mathbf{r}} + \mathbf{F_s}(\mathbf{r}) \cdot \frac{\partial}{\partial \mathbf{p}} + \mathbf{F_l}(\mathbf{r}) \cdot \frac{\partial}{\partial \mathbf{p}} = iL_s + \mathbf{F_l}(\mathbf{r}) \cdot \frac{\partial}{\partial \mathbf{p}} \tag{46}$$

Това дава следния пропагатор:

$$G_{lsl}(\Delta t) = \mathrm{e}^{(\Delta t/2)\mathbf{F_l}(\mathbf{r})\cdot\partial/\partial\mathbf{p}} \mathrm{e}^{iL_s\Delta t} \mathrm{e}^{(\Delta t/2)\mathbf{F_l}(\mathbf{r})\cdot\partial/\partial\mathbf{p}} \tag{47}$$

Използвайки теоремата на Тротер за референтния пропагатор получаваме:

$$\mathrm{e}^{iL_s\Delta t} = \left[\mathrm{e}^{(\delta t/2)\mathbf{F_s}(\mathbf{r})\cdot\partial/\partial\mathbf{p}} \mathrm{e}^{\delta t\dot{\mathbf{r}}\cdot\partial/\partial\mathbf{r}} \mathrm{e}^{(\delta t/2)\mathbf{F_s}(\mathbf{r})\cdot\partial/\partial\mathbf{p}}\right]^n \tag{48}$$

където $\delta t = \Delta t/n$, а n се избира така, че да гарантира стабилна симулация.

Интеграторът r-RESPA може да се запише в следния псевдокод:

Algorithm 1 r-RESPA

1: $\mathbf{v}_i = \mathbf{v}_i + \frac{\Delta t}{2m_i}\mathbf{F_l}_i$
2: **for** $k = 1, \ldots, n$ **do**
3: $\quad \mathbf{v}_i = \mathbf{v}_i + \frac{\delta t}{2m_i}\mathbf{F_s}_i$
4: $\quad \mathbf{r}_i = \mathbf{r}_i + \delta t \mathbf{v}_i$
5: $\quad \mathbf{F_s}_i = \mathrm{update\_F_s}(\mathbf{r}_i)$
6: $\quad \mathbf{v}_i = \mathbf{v}_i + \frac{\delta t}{2m_i}\mathbf{F_s}_i$
7: $\mathbf{F_l}_i = \mathrm{update\_F_l}(\mathbf{r}_i)$
8: $\mathbf{v}_i = \mathbf{v}_i + \frac{\Delta t}{2m_i}\mathbf{F_l}_i$

Този интегратор може да се използва в комбинация със замразяване на вибрационните степени на свобода, което позволява $\delta t = 2$fs. Основният му недостатък е, че води до нелинейни нестабилности [53].

Вместо него може да се използва надстройка на този алгоритъм, MOLLY (Mollified impulse method) [54]. 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: $\mathbf{v}_i = \mathbf{v}_i + \frac{\Delta t}{2m_i}\mathbf{F_l}_i$
2: **for** $k = 1, \ldots, n$ **do**
3: $\quad \mathbf{v}_i = \mathbf{v}_i + \frac{\delta t}{2m_i}\mathbf{F_s}_i$
4: $\quad \mathbf{r}_i = \mathbf{r}_i + \delta t \mathbf{v}_i$
5: $\quad \mathbf{F_s}_i = \mathrm{update\_F_s}(\mathbf{r}_i)$
6: $\quad \mathbf{v}_i = \mathbf{v}_i + \frac{\delta t}{2m_i}\mathbf{F_s}_i$
7: $\overline{\mathbf{r}_i} = A\left(\mathbf{r}_i^k\right)$
8: $\mathbf{F_l}_i = \mathrm{update\_F_l}(\overline{\mathbf{r}_i})$
9: $\mathbf{v}_i = \mathbf{v}_i + \frac{\Delta t}{2m_i}\mathbf{F_l}_i$

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

2.1.10 Поддържане на постоянна температура

МД симулациите запазват пълната енергия на системата, т.е. генерират микроканоничен NVE ансамбъл. Понякога, обаче, е необходимо симулациите да се провеждат в каноничен NVT ансамбъл, например за да се симулира дадена система по-реалистично (повечето белтъци се денатурират при високи температури) или за да могат да се сравняват свойства, получени чрез симулации, с експериментални стойности (обикновено реалните експерименти се провеждат при постоянна температура). За симулация на канонична система е необходим механизъм, термостат, който да позволява добавянето и отделянето на енергия в и от системата, за да се поддържа постоянна температура и да се осигурят правилните флуктуации на енергията.

Един от най-популярните методи за контрол на температурата е термостатът на Берендсен [55], при който системата е слабо свързана с топлинен резервоар с желаната температура T_0. Тя се достига експоненциално, като на всяка стъпка скоростите се скалират, така че скоростта на изменение на температурата да е пропорционална на разликата в зададената и моментната температури:

$$\frac{dT}{dt}(t) = \frac{1}{\tau_T}\big(T_0 - T(t)\big) \tag{49}$$

където τ_T е параметър на свързване, който определя колко силно е свързана системата с топлинния резервоар – колкото по-малка е стойността на τ_T, толкова по-бързо се достига зададената температура. Коефициентът на скалиране на скоростите е:

$$\lambda^2 = 1 + \frac{\Delta t}{\tau_T}\left(\frac{T_0}{T} - 1\right) \tag{50}$$

Термсотатът на Берендсен е много ефективен и стабилен, но недостатъкът му е, че не генерира траектории, съответстващи на каноничен ансамбъл. При достатъчно големи системи, обаче, отклоненията от каноничния ансамбъл са пренебрежими.

Друг начин да се контролира температурата е да се използва термостатът на Ланжевен [56], който се основава на теорията на Брауновото движение. Системата се разглежда като потопена в "разтворител" с малки частици, който налага желаната температура, като влияе на разтворените частици чрез два допълнителни члена в уравненията на Нютон – сила на триене и случайна сила. Уравнението на движение на i-тата частица от системата с температура T е:

$$m_i\ddot{\mathbf{r}}_i = -\nabla_i U - \mathbf{F}_i^{fr}(t) - \mathbf{R}_i(t) \tag{51}$$

където силата на триене е:

$$\mathbf{F}_i^{fr}(t) = -\gamma\dot{\mathbf{r}}_i = -\frac{1}{6}\frac{\pi k_B T}{\hbar}\dot{\mathbf{r}}_i \tag{52}$$

Случайната сила $\mathbf{R}_i(t)$ е с Гаусово разпределение с нулева средна стойност и дисперсия:

$$\sigma^2 = 2m_i\gamma k_B T \tag{53}$$

Този избор на $\mathbf{F}_i^{fr}(t)$ и $\mathbf{R}_i(t)$ гарантира, че термостатът възпроизвежда каноничен ансамбъл.

Комбинация от горните два метода е термостатът *v-rescale* [57], при който скоростите се рескалират с подходящо изран случаен коефициент, който гарантира правилното възпроизвеждане на канонично разпределение. Моментната кинетична енергия се изменя по следния начин:

$$dK = (K_0 - K)\frac{dt}{\tau_T} + 2\sqrt{\frac{KK_0}{N_f}}\frac{dW}{\sqrt{\tau_T}} \tag{54}$$

където $K_0 = N_f k_B T_0/2$ е средната кинетична енергия при зададената температура, N_f е броят на степените на свобода, dW е Винеров процес, а τ_T е параметърът на свързване към термостата. Рескалиращите коефициенти са:

$$\lambda^2 = \mathrm{e}^{-\frac{\Delta t}{\tau_T}} + \frac{K_0}{N_f K}\left(1 - \mathrm{e}^{-\frac{\Delta t}{\tau_T}}\right)\left(R_1^2 + \sum_{i=2}^{N_f} R_i^2\right) + 2\mathrm{e}^{-\frac{\Delta t}{\tau_T}}\sqrt{\frac{K_0}{N_f K}\left(1 - \mathrm{e}^{-\frac{\Delta t}{\tau_T}}\right)} R_1 \tag{55}$$

където R_i са случайни числа от Гаусово разпределение.

2.1.11 Поддържане на постоянно налягане

Подобно на свързването към топлинен резервоар, за симулиране на изотермно-изобарен ансамбъл системата може да се свърже и към резервоар, който да поддържа на постоянно налягане. Един от най-популярните баростати е баростатът на Берендсен, който е аналогичен на термостатът на Берендсен. Налягането достига зададената стойност експоненциално:

$$\frac{dP}{dt} = \frac{1}{\tau_P}\left(P_0 - P\right) \tag{56}$$

където τ_P е параметърът на свързване към резервоара На всяка стъпка размерите на симулационната кутия и атомните координати се скалират с коефициент:

$$\mu^3 = 1 - \frac{\beta \Delta t}{\tau_P}\left(P_0 - P\right) \tag{57}$$

където β е изотермната свиваемост на разтворителя. Недостатъкът на този метод е, че не възпроизвежда точно NPT ансамбъл.

Алтернативен метод за поддържане на постоянно налягане е баростатът на Паринело-Раман [58]. В този случай свързването към резервоара се осъществява чрез добавяне на нови степени на свобода в Лагранжана на системата, които да описват динамиката на векторите на кутията. Уравнението на движение на векторите на кутията е:

$$\frac{d^2\mathbf{b}}{dt^2} = V\mathbf{W}^{-1}\mathbf{b}^{T-1}\left(\mathbf{P} - \mathbf{P}_0\right) \tag{58}$$

където V е обемът на кутията, а $\mathbf{W}$ е матрица, която контролира степента на свързване с баростата, с елементи $\left(\mathbf{W}^{-1}\right)_{ij} = \frac{4\pi^2\beta_{ij}}{2\tau_P^2 L}$, където L е най-дългата страна на кутията. Уравненията на движение на частиците стават:

$$\begin{aligned} \ddot{\mathbf{r}}_i &= \frac{\mathbf{F}_i}{m_i} - \mathbf{M}\dot{\mathbf{r}}_i \\ \mathbf{M} &= \mathbf{b}^{-1}\left[\mathbf{b}\frac{d\mathbf{b}^T}{dt} + \frac{d\mathbf{b}}{dt}\mathbf{b}^T\right]\mathbf{b}^{T-1} \end{aligned} \tag{59}$$

2.2 Метадинамика

Въпреки успешното приложение на стандартната МД в множество области – от материалознанието до молекулярната биология и фармакологията, методът страда от множество ограничения. Един от най-сериозните му недостатъци е ограничената времева скала, която днешните компютърни технологии и алгоритми за семплиране позволяват да се изследва. В момента стандартните МД симулации достигат продължителност от десетки микросекунди за средни и големи системи до няколко милисекунди за по-малки системи [59]. За съжаление повечето от представляващите интерес явления, например структурни фазови преходи, нагъването на протеини, взаимодействието между протеини и макромолекулното разпознаване, преминаване на вещества през мембрани, активиране на мембранни канали и др., настъпват за времена, които са на порядъци по-дълги.

Друг проблем е, че пълното разбиране на повечето химични процеси обикновено изисква да се познава обуславящата ги повърхнина на свободната енергия (ПСЕ), т.е. трябва да се обходят всички достъпни състояния на системата. За съжаление обикновено фазовото пространство има множество локални минимуми, разделени от относително високи бариери. Това затруднява адекватното му обхождане в разумно време и не позволява директното пресмятане на разлики в свободните енергии от равновесни траектории.

Не на последно място може да се добави необходимостта резултатите от симулацията да се обобщят в малък брой лесни за разбиране и интерпретация параметри, описващи изследвания феномен. По тези причини активно се разработват на методи на статистическата механика за намаляване на размерността на задачата и оценка на свободната енергия по малко на брой променливи [60, 61], наречени параметри на подреждане, колективни променливи (КП) или реакционни координати. Свободната енергия, пресметната в ограничено фазово пространство, дефинирано от една или повече реакционни координати се нарича "потенциал на средното поле" (ПСП) по съответните параметри.

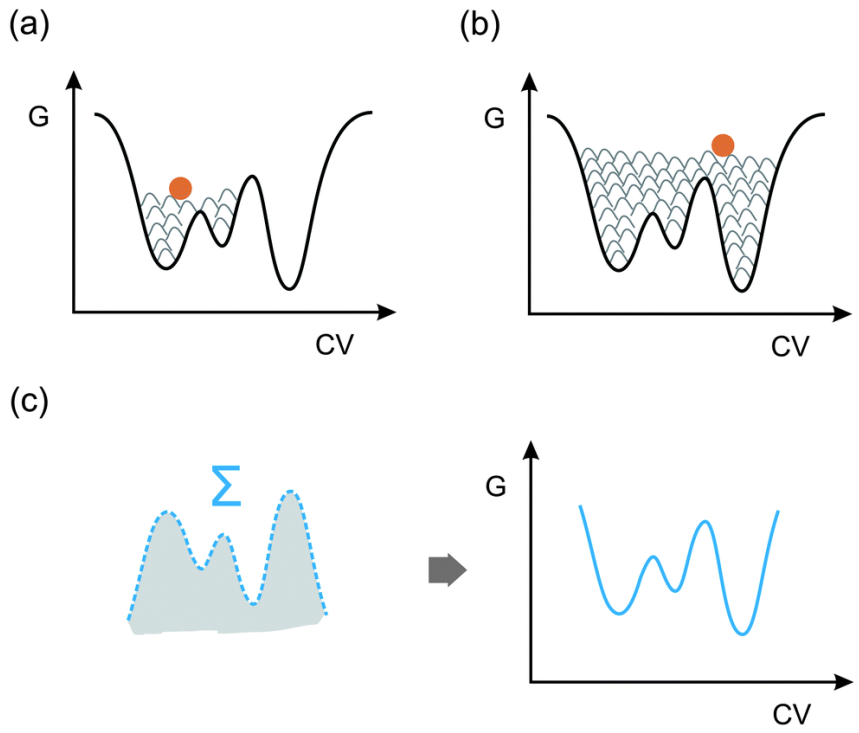
Сред по-популярните методи за пресмятане на ПСП са [1] обхождане по метода на чадърите (umbrella sampling) [62], адаптивно повлияване на силите (adaptive biasing force) [63], насочена молекулна динамика (steered molecular dynamics) [64] и метадинамика (metadynamics) [65, 66].

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 [66]. 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 минимуми по път с ниска свободна енергия и да изследва съседни минимуми в повърхнината на свободната енергия. Сумата от гаусианите се използва за итеративно реконструиране на профила на свободната енергия в пространството на избраните КП (фиг. 13).

2.2.1 Време-зависим метадинамичен потенциал

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 $s_\alpha(\mathbf{r})$, $\alpha = 1, \ldots, d$, където d е

[1] Тъй като няма приета стандартна терминология на български език, сме използвали свободен превод по преценка на автора.



Фигура 13: Схематична илюстрация на метода на метадинамиката. (a) Системата изследва неизвестна повърхнина на свободната енергия. Когато бариерите, разделящи локалните минимуми в повърхнината на свободната енергия, са по-високи от няколко пъти $k_B T$, вероятността системата да излезе от минимума и да започне да изследва други области е много малка. Метадинамиката ускорява излизането от локалния минимум и обхождането на съседни региони, като повлиява динамиката на системата чрез добавяне на гаусиани към потенциала. (b) Щом повърхнината на свободната енергия е запълнена, траекторията на колективната променлива става дифузна и (c) профилът на свободната енергия се реконструира посредством сумата от добавените гаусиани. Фигурата е от [67]

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.

Равновесното поведение на КП се определя напълно от вероятностното ѝ разпределение:

$$P(s) = \frac{\mathrm{e}^{-\frac{1}{k_B T} F(s)}}{\int ds\, \mathrm{e}^{-\frac{1}{k_B T} F(s)}} \tag{60}$$

където свободната енергия $F(s)$ е:

$$F(s) = -k_B T \ln \left(\int d\mathbf{r}\, \mathrm{e}^{-\frac{1}{k_B T} V(\mathbf{r})} \delta\big(s - s(\mathbf{r})\big) \right) \tag{61}$$

Ако може да се пресметне много дълга траектория $\mathbf{r}(\mathrm{t})$ при температура T, $P(s)$ може да се получи от хистограмата на КП по тази траектория, т.е. в момент t, $P(s) \sim \frac{1}{t} \int_0^t dt' \delta\big(s(\mathbf{r}(t')) - s\big)$. Ако системата проявява метастабилност, движението ѝ е ограничено в някакъв локален минимум на повърхнината на свободната енергия $F(s)$ и тя ще излезе от него с много малка вероятност.

Метадинамиката решава този проблем, като добавя външен потенциал, действащ на системата в момента t:

$$V_G\left(s\left(\mathbf{r}\right),t\right)=\omega\sum_{\substack{t'=\tau_G,2\tau_G,\ldots\\ t'<t}}\exp\left\{-\frac{\left(s\left(\mathbf{r}\right)-s\left(t'\right)\right)^2}{2\delta s^2}\right\} \tag{62}$$

където $s(t) = S(\mathbf{r}(t))$ е стойността на 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. Ако едновременно се използват d на брой КП, метадинамичният потенциал е:

$$V_G\left(s\left(\mathbf{r}\right),t\right)=\omega\sum_{\substack{t'=\tau_G,2\tau_G,\ldots\\ t'<t}}\exp\left\{-\sum_{\alpha=1}^{d}\frac{\left(s_\alpha\left(\mathbf{r}\right)-s_\alpha\left(t'\right)\right)^2}{2\delta s_\alpha^2}\right\} \tag{63}$$

След достатъчно дълъг период от време $V_G(s,t)$ от (62) може да се използва за статистическа оценка на повърхнината на свободната енергия в пространството на КП:

$$\lim_{t\to\infty}V_G\left(s,t\right)=-F\left(s\right)+C \tag{64}$$

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. 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 [68].

2.2.2 Темперирана метадинамика

Влиянието на избора на височината на гаусианите върху точността на реконструкция на ПСЕ може да се елиминира, ако се използва т.нар. "темперирана" метадинамика (well-tempered metadynamics) [69]. В този случай височината на добавяните гаусиани не е постоянна по време на симулацията, а се изменя по следния начин:

$$W=\omega_0\,\mathrm{e}^{-\frac{V_G(s,t)}{k_BT\Delta T}} \tag{65}$$

където ω_0 е началната височина на гаусианите, ΔT е температура, а $V_G\left(s,t\right)$ е метадинамичният потенциал, натрупан за време t.

За разлика от обикновената метадинамика, при темперираната версия на метода външният потенциал не компенсира напълно ПСЕ, но схожда по следния начин:

$$V_G\left(s,t\to\infty\right)=-\frac{\Delta T}{T+\Delta T}F\left(s\right)+C \tag{66}$$

където T е температурата на системата.

Следователно КП се обхождат при по-висока, фиктивна температура $T + \Delta T$:

$$P\left(s, t \to \infty\right) \propto \mathrm{e}^{-\frac{F(s)}{k_B(T+\Delta T)}} \tag{67}$$

ΔT контролира степента на обхождане на ПСЕ: $\Delta T \to 0$ съответства на стандартна МД, а $\Delta T \to \infty$ – на стандартна метадинамика.

2.2.3 Избор на колективни променливи

Подобно на другите методи за реконструкция на свободната енергия в набор от обобщени координати, успехът и надеждността на метадинамиката зависят силно от избора на КП. Времето, необходимо на системата да излезе от локалния минимум на повърхнината на свободната енергия, се определя от броя на гаусианите, необходим за запълване на ямата. Той е пропорционален на $(1/\delta s)^d$, където d е броят на КП, т.е. ефективността на метода намалява експоненциално с броя на КП. Ако той е голям, единственият начин за ефективно излизане от минимума е да се използват гаусиани с ширина, сравнима с ширината на минимума, което води до ниска точност на реконструираната ПСЕ.

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.

Очевидно второто и третото условия са взаимно изключващи се и в много случаи е много трудно да се намери "добър" набор от КП. Ако не е възможно да се намери набор от две или три променливи, които да са в състояние да опишат всички най-бавни степени на свобода, метадинамиката може да се използва в комбинация с паралелно темпериране[2] (parallel tempering metadynamics) [70] или обмяна на метадинамичния потенциал (bias exchange metadynamics) [71]. Друга възможност е да се използват **path** КП [72].

Един от основните недостатъци на метадинамиката е, че не съществува универсално правило за намиране на подходящ набор от КП. В повечето случаи се налага да се процедира по метода на пробите и грешките и да се пробват различни комбинации от КП. Това, което може да се направи, е да се провери *a posteriori* дали изборът на КП предоставя достатъчно коректно описание на изследвания процес. За целта могат да се използват техники за семплиране на реакционния път [2] (transition path sampling techniques) [73, 74].

[2] Тъй като няма приета стандартна терминология на български език, сме използвали свободен превод по преценка на автора.

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.

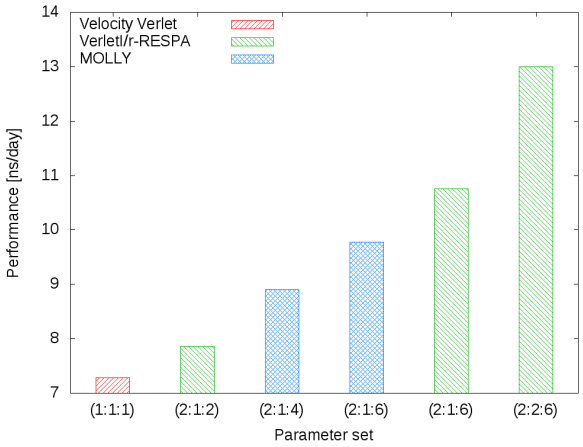
Всички представени симулации са направени на суперкомпютъра IBM BlueGene/P към Националния център за суперкомпютърни приложения [75] на различен брой изчислителни ядра. Молекулно-динамичните пакети, чиято производителност беше изследвана и профилирана, са GROMACS 4.5.3 [76] и NAMD [77] версия CVS 2011-02-19.

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 [52]. Тъй като при него се проявяват линейни и нелинейни нестабилности [53], в NAMD е имплементирана и надстройка на този алгоритъм, MOLLY (Mollified impulse method) [54]. 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} късодействащите 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

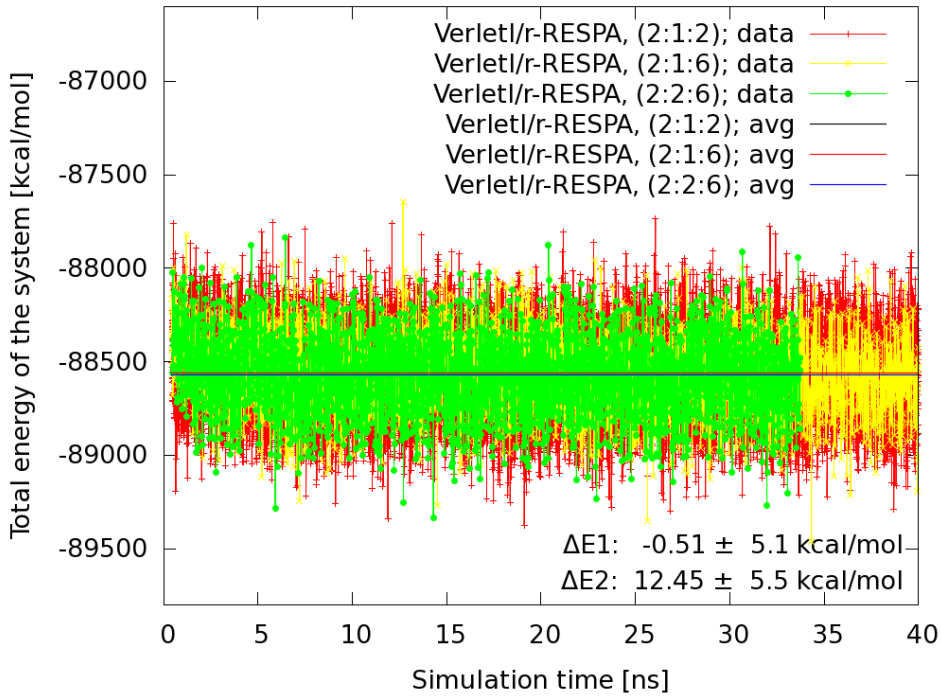
система с около 35 хиляди атома (човешки гама-интерферон във вода) от броя на използваните изчислителни ядра. Получените резултати са представени на фиг. 14. За получаването на първата точка е използван стандартния алгоритъм на Верле със скорости с интеграционна стъпка от 1 fs, а за втората – алгоритъм 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 2.1.8), 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 14: Производителност на NAMD при използването на различни алгоритми за интегриране и различни параметри.

Направено беше сравнение на получените с различни параметри на алгоритъма с променлива стъпка Verlet-I/r-RESPA макро-характеристики на системата, в частност нейната енергия. Референтната симулация е с интегратора Verlet-I/r-RESPA със стандартни параметри 2,1,2 и е с продължителност 2 000 000 стъпки (40 ns). На фиг. 15 е показана общата енергия на системата като функция на симулационното времето. Представени са и средните стойности. Усредняването е проведено без да се отчитат стойностите на енергията от първите 200 ps.

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.



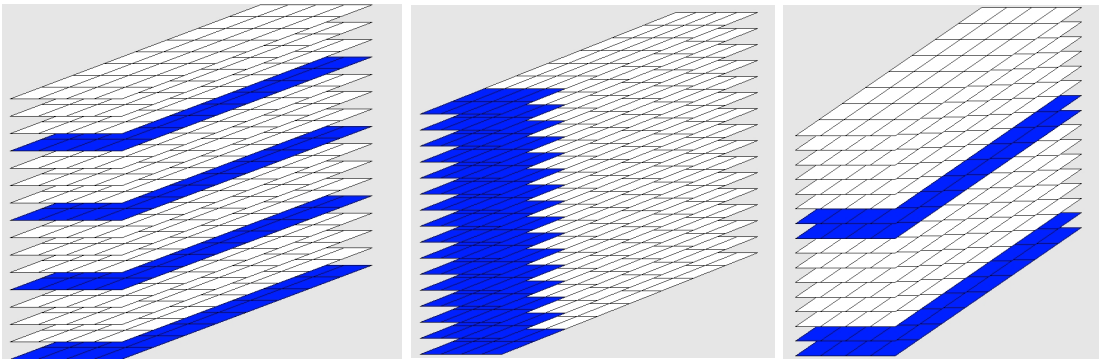
Фигура 15: Зависимост на пълната енергия от параметрите на интегратора Verlet-I/r-RESPA. $\Delta E1 = \overline{E}(2:1:2) - \overline{E}(2:1:6)$ и $\Delta E2 = \overline{E}(2:1:2) - \overline{E}(2:2:6)$.

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) [76].

Разделянето по частици е най-простият метод за разделяне на системата. При него в началото на симулацията на всеки процесор се назначават определени частици от системата. След това се разпределят и силите, действащи между частиците, така че натоварването на процесорите да е равномерно разпределено. Този алгоритъм изисква всеки процесор да познава координатите на поне половината от частиците на системата. При голям брой изчислителни ядра N трябва да се обменят $N \times N/2$ съобщения с координати. Заради тази квадратична зависимост разделянето по частици не се скалира добре и се използва само в специфични случаи, когато в системата има далекодействащи ковалентни взаимодействия.

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 процесорите. В общия случай на клиногонална 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



(а) `interleave` (б) `pp_pme` (в) `cartesian`

Фигура 16: Режими на подреждане на pp/pme изчислителните възли в GROMACS. PP възлите са представени в бяло, а pme – в синьо.

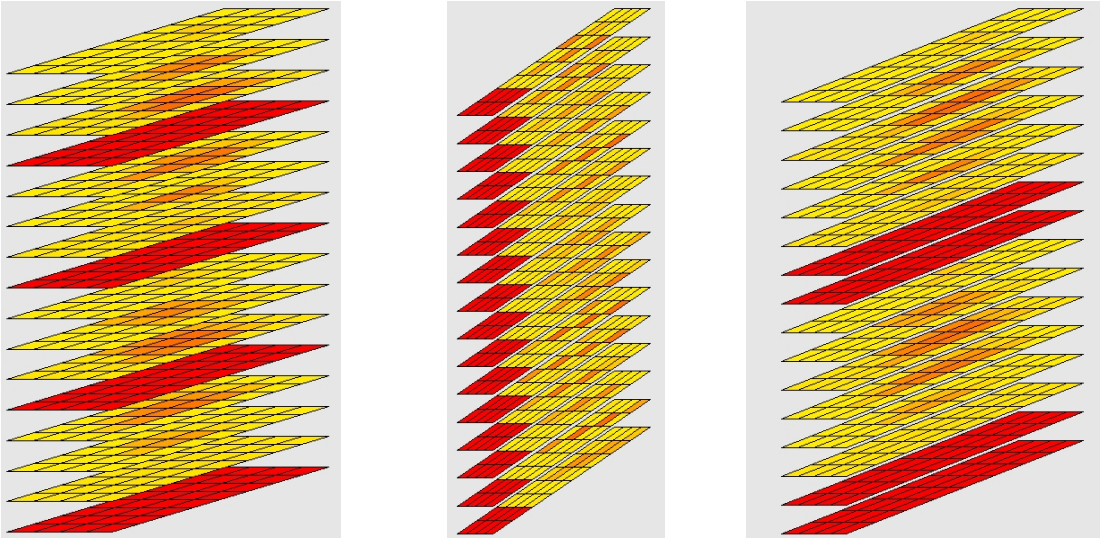
частиците (обикновено на 10 времеви стъпки).

Електростатичното взаимодействие е далекодействащо, а не локално. Затова за пресмятането му се използват специални алгоритми. В GROMACS обикновено се използва алгоритъмът PME [48], 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 [76].

По подразбиране GROMACS разделя pp:pme изчислителните ядра в съотношение 3:1 и ги подрежда в режим `interleave` (фиг. 16а). 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` (фиг. 16б) първо се нареждат pp ядрата и накрая следват pme ядрата. Режимът `cartesian` (фиг. 16в) 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. Scalasca [78] е софтуерен пакет за профилиране на производителността на паралелни програми. Целта му е да подпомага оптимизацията на паралелните програми чрез измерване и анализиране на поведението им по време на изпълнение на програмата. В резултат могат да се идентифицират проблемните части от програмата, които забавят нейното изпълнение, особено такива, свързани с неравномерно разпределение на натоварването, комуникация и синхронизация. Тестовата система се състои от 103079 атома, симулацията е с продължителност 10000 стъпки (20ps, времевата 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 (Фиг 17). 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



(а) `interleave` (б) `pp_pme` (в) `cartesian`

Фигура 17: Разпределение на комуникацията при различните режими на подреждане на pp/pme изчислителните ядра в GROMACS. Червено означава по-интензивна комуникация, а жълто – по-слабо интензивна. Както се вижда във всички режими комуникацията с pme ядрата е най-интензивна.

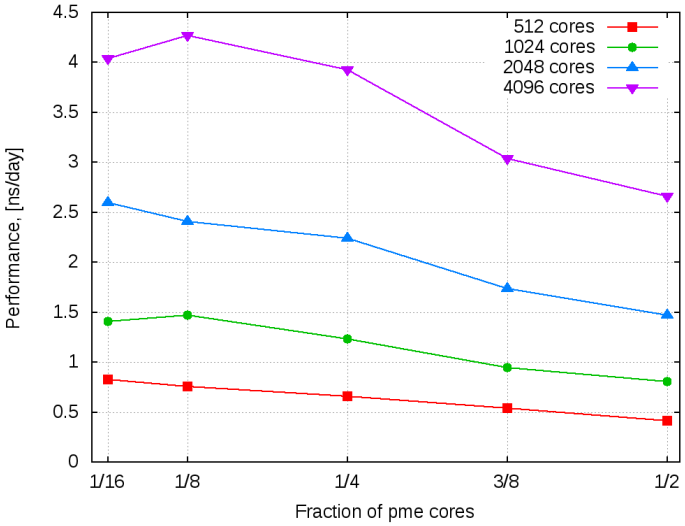
да изпратят на pp ядрата силите, действащи на частиците.

Резултатите за производителността (табл. 2) 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`.

Таблица 2: 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

Изпълнение на 512 ядра е отнело средно по 4668 s на ядро. Pme ядрата, които са 64 на брой, са смятали средно по 5888 s, докато pp ядрата – средно по 4494 s. Основната част от времето за изчисления отнема процедурата `do_md`, която извършва фактическата молекулна динамика – около 70% от общото време. Останалите проценти са разпределени главно за пресмятане на далекодействащите електростатични сили и за разделянето на системата на домейни. Броят на комуникациите отново е разпределен главно между процедурата `do_md` – около 58% и пресмятането на далекодействащите електростатични сили – 21%, както и инициализацията на паралелната среда – 20%. Тук обаче, средният брой комуникации на pme ядрата е 10453, докато на pp ядрата едва 4174. Интензивната комуникация на pme ядрата обяснява, защо те отнемат средно повече време на програмата от pp ядрата. По-



Фигура 18: Производителност на GROMACS в зависимост от дела на pme ядра при различен брой изчислителни ядра.

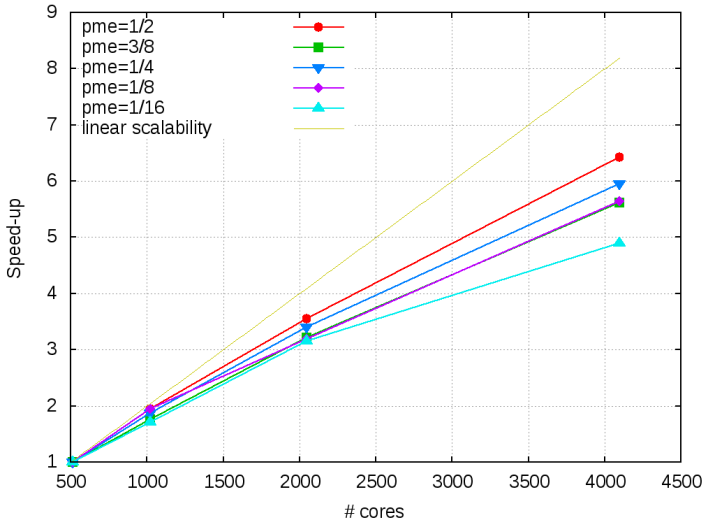
добно поведение се наблюдава и при изчислението направено на 1024 изчислителни възела. Отново най-много време е отнела процедурата `do_md` – около 66%. Средното време за изпълнение на процес е 4912 s, като средното време само за pme ядрата е 6829 s, а за pp ядрата – 4637 s. Общо са обменени 7.1×10^6 съобщения, 60% от които са в процедурата `do_md`, 22% за пресмятането на далекодействащите електростатични сили и приблизително 18% – при инициализацията на паралелната среда. Средният брой комуникации на pp изчислителните възли е около 6100, а на pme ядрата двойно повече 13600.

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 пресмятания. 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 инструментиращия пакет, изследването беше проведено на базата на симулации с неинструментиран GROMACS 4.5.3, като данните за производителността за взети направо от изходните log-файлове.

На фиг. 18 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.

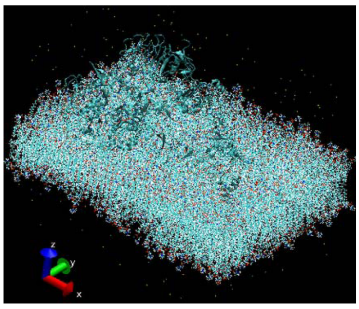
На фиг. 19 е показано ускорението на GROMACS, нормирано към производителността за 512 изчислителни ядра, при различен брой pme ядра. При по-малко pme ядра скалируемостта спада, като следствие от увеличаването на комуникациите (при намаляване на броя на pme ядрата комуникацията им нараства). До 4096 ядра производителността нараства с намаляване на броя на pme ядрата, но скалируемостта леко се влошава.



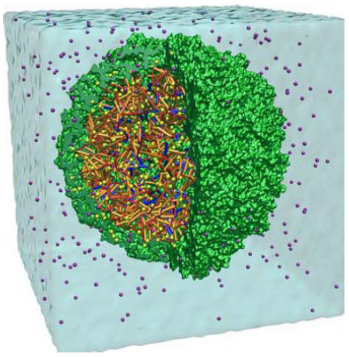
Фигура 19: Ускорение на GROMACS за различен брой изчислителни ядра и дял pme ядра (нормирано към 512 изчислителни ядра).

3.3 Зависимост на производителността от размера на симулираната системата

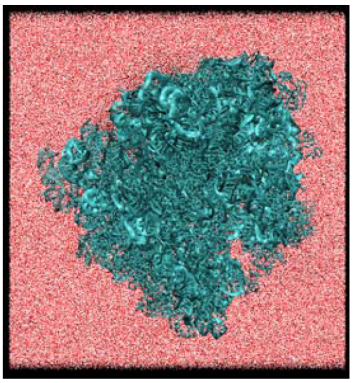
За изследване на производителността в зависимост от големината на симулираната система бяха използвани следните три тестови системи:



(а)



(б)



(в)

Фигура 20: Тестови системи.

а Рецептор на епидермалния растежен фактор (EGFR; ErbB-1; HER1 при хората). Това е мембранен рецептор, димер, инкорпориран в двоен липиден слой. Симулационният обем е запълнен с водни молекули, като общият брой на атомите е 465 399 ($\sim 5 \times 10^5$). Тази система беше предоставена от д-р Илиян Тодоров от Отдела за изчислителни науки и инжинеринг към CCLRC Daresbury Laboratory, Обединеното кралство.

б Сателитен вирус на тютюнева мозайка. Структурата е взета от базата данни PDB с код 1A34. Системата се състои от вируса, разтворен във вода, общо 1 007 930 ($\sim 1 \times 10^6$) атома.

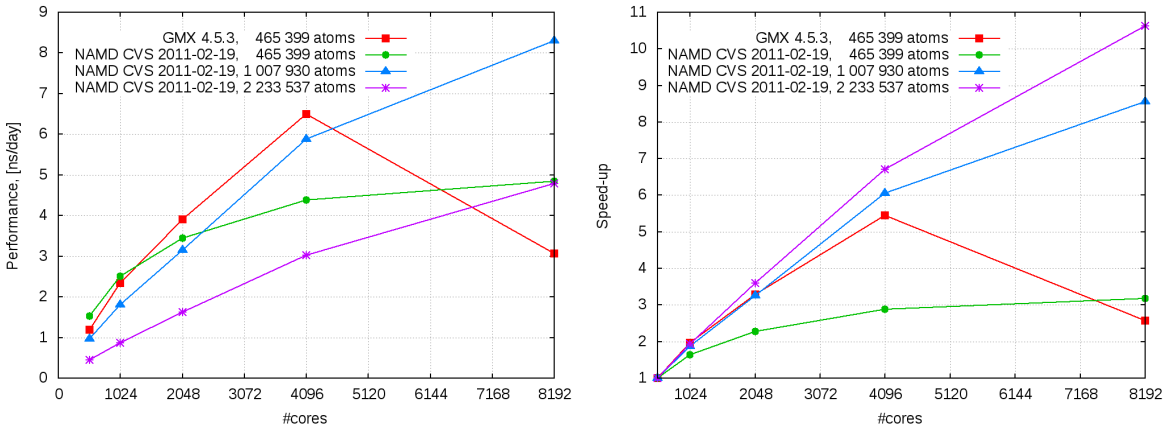
в Рибозома, съдържаща малката и голямата поединици (PDB ID 3FIK и 3HIH). Разтворената и електронеутрализирана система съдържа 2 233 537 атома ($\sim 2 \times 10^6$).

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 A).

Резултатите са обобщени в табл. 3 и са представени на фиг. 21а. 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. 21б). 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 под 100, защото времето за осъществяване на комуникациите вече доминира времето за изчисление. Такова поведение не се забелязва при пакета NAMD (фиг. 22).

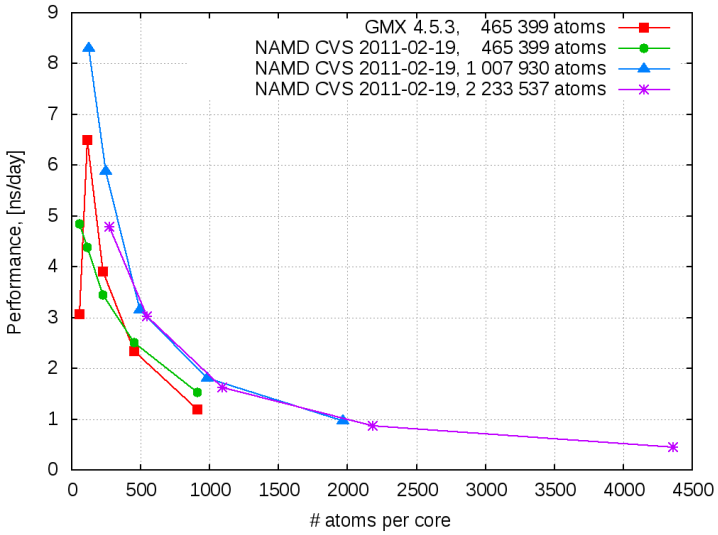
Таблица 3: Производителност в [ns/day] на пакетите GROMACS 4.5.3 и NAMD CVS 2011-02-19 на архитектурата IBM BlueGene/P.

# изчислителни	GROMACS 4.5.3	NAMD CVS 2011-02-19		
ядра	465 399 атома	465 399 атома	1 007 930 атома	2 233 537 атома
512	1.19	1.52	0.97	0.45
1024	2.34	2.50	1.81	0.87
2048	3.90	3.45	3.15	1.62
4096	6.49	4.38	5.88	3.02
8192	3.06	4.84	8.30	4.78



(а) (б)

Фигура 21: Производителност (а) и ускорение (б) на пакетите GROMACS 4.5.3 и NAMD CVS 2011-02-19 на архитектурата IBM BlueGene/P.



Фигура 22: 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

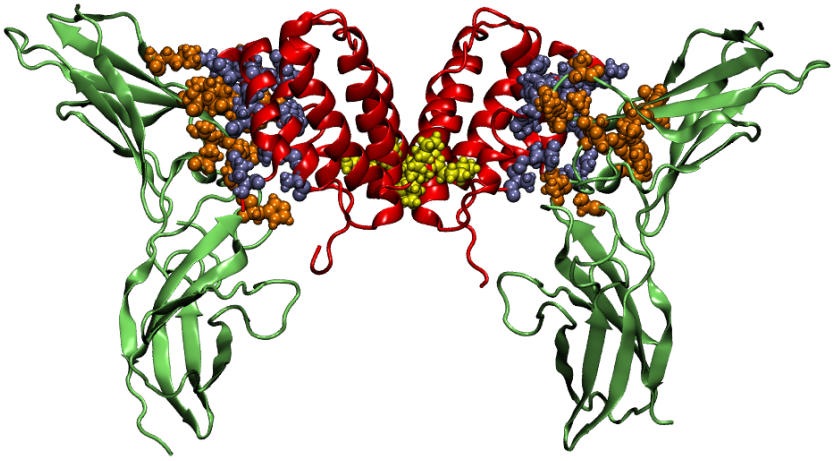
Както стана ясно в секция 1.1 анормалната експресия на IFN-γ се свързва с множество автовъзпалителни и автоимунни заболявания. Основната идея за контролиране на такива състояния е да се блокира или молекулата на цитокина, или на неговия рецептор. Това може да се постигне чрез намирането на подходяща мутагенна форма на hIFN-γ , която да е способна да се свързва с рецептора, но да не притежава биологична активност. Както беше споменато в секция 1.1.2 (фиг. 4, табл. 1), счита се, че IFN-γ се свързва с hIFNγRα 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 Lys^{86} – $Lys - Lys^{88}$ (фиг. 23), 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 спирали E. Списъкът с всички 100 мутагенни форми на hIFN-γ е представен в първите четири колони на табл. 4.

4.2 Анализ на вторичната структура на белтъци

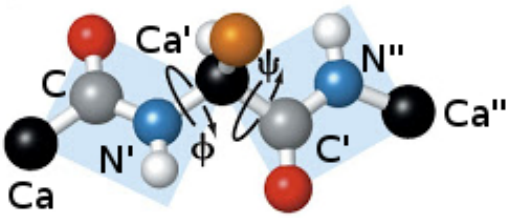
4.2.1 Ъгли на усукване на гръбнака

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 центъра на ротация (фиг. 24). Това са ъгълът ϕ, включващ карбоксилния въглерод C на предишния а.к. остатък,



Фигура 23: Място на мутациите - аминокиселинни остатъци $\mathrm{Lys}^{86} - \mathrm{Lys} - \mathrm{Lys}^{88}$ са изобразени в жълти сфери. 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 IFNγRα. Както се вижда, остатъци 86–88 по никакъв начин не участват във взаимодействието с hIFNγRα .

азотът от амидната група N', алфа въглеродния атом C'_α и карбоксилния въглерод C'; и ψ, включващ атоми $\mathrm{N}' - \mathrm{C}'_\alpha - \mathrm{C}' - \mathrm{N}''$.



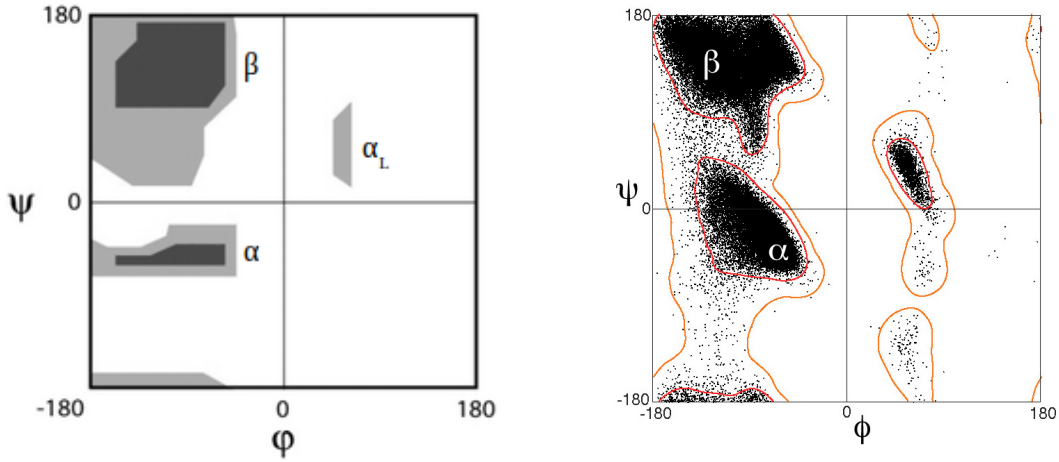
Фигура 24: Ъгли на усукване на гръбнака на протеин: ϕ, който описва усукването на средната връзка между атоми $\mathrm{C} - \mathrm{N}' - \mathrm{C}'_\alpha - \mathrm{C}'$, и ψ, дефиниран от атоми $\mathrm{N}' - \mathrm{C}'_\alpha - \mathrm{C}' - \mathrm{N}''$.

4.2.2 Диаграма на Рамачандран

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 [79, 80]. Свободна ротация (дължаща се на топлинното движение) около C-C връзката е възможна само при липсата на каквито и да е стерични ограничения. 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 $+180°$. През 1963 г. Рамачандран и съавтори [79] използват компютърен модел на малки пептиди и варират ъглите ϕ и ψ като търсят стабилни конформации. Атомите се третират като твърди сфери с радиус равен на Ван дер Ваалсовия. Така получават карта, със забранени области,

в които атомите си пречат стерично и са енергетично неизгодни. Двойките ъгли ϕ и ψ се ступват 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 (фиг. 25а).

През 2003 г. Ловел и съавтори правят селекция в базата данни PDB и избират най-добрите 500 нехомоложни белтъчни структури, които са получени с резолюция, по-добра от 1.8 $\AA$ и имат ниски температурни фактори. Така получават разпределението на всички реално наблюдавани ъгли на Рамачандран от тези структури и дефинират нови контури на разрешените области в диаграмата (фиг. 25б). Диаграмата на Рамачандран е прост и много бърз метод за изследване на поведението и стабилността на полипептидна верига.



(а) (б)

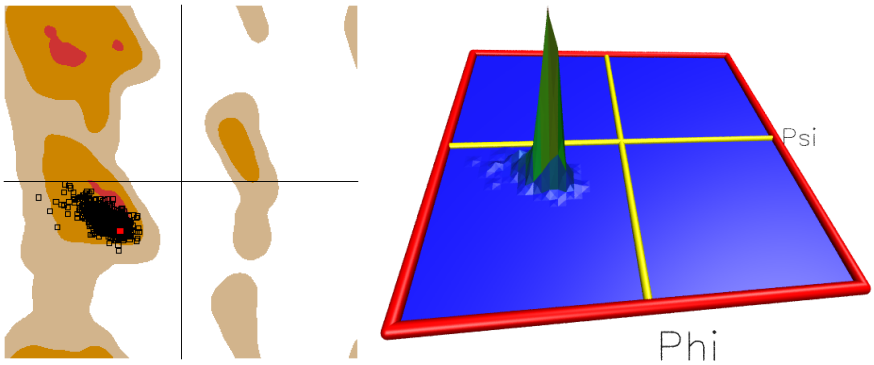
Фигура 25: Диаграма на Рамачандран. (а) Оригиналната карта на Рамачандран [79]. Тъмно сивите области (т.нар. благоприятни) са получени при използване на сфери с Ван дер Ваалсови радиуси, а светло сивите (частично благоприятни) – с редуцирани Ван дер Ваалсови радиуси; (б) Диаграма на Рамачандран с реалното разпределение на ъглите на усукване на гръбнака. Червеният (вътрешен) контур обхваща области, в които са струпани 98% от точките, а оранжевият (външен) – 99,95%. Фигурата е от [81].

Диаграмата на Рамачандран на аминокиселинен остатък Lys^{86} на нативен IFN-γ от референтната МД симулация е представена на фиг. 26.

4.2.3 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) [82]. Алгоритъмът се базира основно на откриването на мотиви от водородни връзки. Дефиницията на структури, базирани на H-връзки, е йерархична. Първо се дефинират H-връзки, а въз основа на тях:

- "n-врътки", които имат H-връзка между C=O групата на аминокиселинен остатък i и N-H групата на остатък i+n, където n=3,4,5;
- "мостове", при които два непоследователни във веригата остатъка са свързани с H-връзка.



Фигура 26: Диаграма на Рамачандран на Lys^{86} на нативен hIFN-γ , получена при референтната МД симулация: остатъците не напускат областта на α-спиралите.

с

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 компенсиране на разклонени Н-връзки и за експерименталните грешки в координатите, прага за обрязване по енергии е по-висок. Приема се, че Н-връзка между карбонилната (C=0) група на остатък i и амино 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 – $(-q_2, +q_2)$:

$$E = q_1 q_1 \left[\frac{1}{r\,(ON)} + \frac{1}{r\,(CH)} - \frac{1}{r\,(OH)} - \frac{1}{r\,(CN)} \right] f, \tag{68}$$

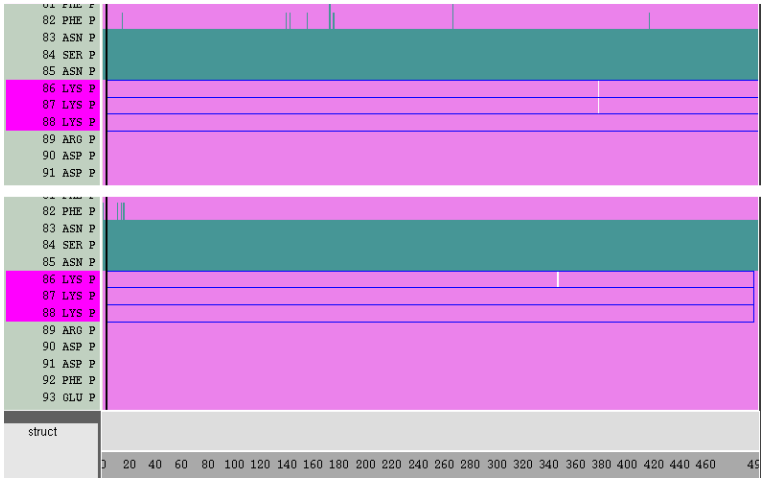
където $q_1 = 0.42e$ и $q_2 = 0.20e$, e е елементарния заряд на електрона, $r\,(AB)$ е разстоянието между атоми A и B в $\mathring{A}$, $f = 332$ е безразмерен коефициент и E е в kcal/mol.

Вторичната структура на нативен hIFN-γ , определена по метода DSSP за всеки кадър от референтната МД симулация е показана на фиг. 27. Както се вижда, аминокиселини $Lys^{86} - Lys - Lys^{88}$ и в двата мономера на hIFN-γ участват стабилно в спирала E по време на цялата симулация.

4.3 Подготовка

Структурите на мутантите бяха създадени с помощта на пакета за визуализация и манипулация на молекули PyMOL [83]. Страничните вериги на трите лизина бяха заменени със страничните вериги на трите заместващи аминокиселини, като гръбнака 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 [10]. Симулациите бяха направени с програмния пакет NAMD 2.6 [77], като беше използвано силовото поле CHARMM 22 с модифициран TIP3P модел на водата [84]. Молекулите на протеините



Фигура 27: Вторична структура на на нативен hIFN-γ , получена при референтната МД симулация, определена по метода DSSP. По вертикалната ос на фигурата са дадени поредните номера на аминокиселините и техните имена. В розово са маркирани трите лизина, които ще бъдат мутирани. По хоризонталната ос за нанесени кадрите от МД симулацията, като 10 кадъра съответстват на 0.2 ns. α-спиралите са изобразени във виолетово, бримките - в зелено, неструктурираните части в бяло.

бяха разтворени в кубични кутии с размери $80 \times 80 \times 80$ $\mathring{A}$ (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 [56] с коефициент на триене от 5/ps. За пресмятане на електростатиката беше използван методът PME с радиус на обрязване на правата сума от 12 $\mathring{A}$ и минимална константа на решетката за реципрочната сума 0.625 $\mathring{A}^{-1}$. За Ван дер Ваалсовите взаимодействия беше използвана изглаждаща функция, която се включва на радиус от 10 $\mathring{A}$ и плавно ги нулира на радуса на обрязване от 12 $\mathring{A}$. Списъците с най-близки съседи се конструираха на всеки 10 стъпки с радиус на обрязване от 13.5 $\mathring{A}$. Използван беше интеграторa с променлива времева стъпка Verlet-I/r-RESPA с времева стъпка от 2fs и пресмятане на далечните електростатични взаимодействия на всеки 2 стъпки. Първо беше направена кратка минимизация на енергията от 1000 стъпи по метода на спрегнатите градиенти, последвана от същинската МД симулация с продължителност от 10 ns. Изчисленията бяха направени на 512 изчислителни ядра на системата IBM BlueGene/P към Националния център за суперкомпютърни приложения [75].

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 (фиг. 24). 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 [77]. Бяха използвани същото силово поле, симулационна кутия, параметри на термостата и методи и радиуси на обрязване за нековалентните взаимодействието както в МД симулациите. За метадинамичните симулации е използван е 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 [75].

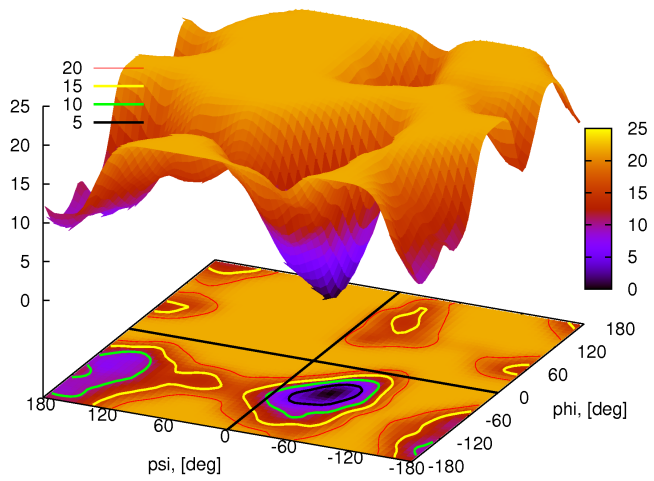
4.4.1 Анализ на данните от метадинамични симулации

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. 28) и съответния мутант в областта на α-спиралите ($\phi \in [0; 30], \psi \in [-180; 0]$):

$$S = \sqrt{\frac{\sum_i^n \left(E\left(\phi_i, \psi_i\right) - E\left(\phi_{ref}, \psi_{ref}\right)\right)^2}{n}} \tag{69}$$

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 бариер 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 28: Повърхнина на свободната енергия в пространството на ъглите на усукване на гръбнака на аминокиселина Lys^{86} на нативен hIFN-γ .

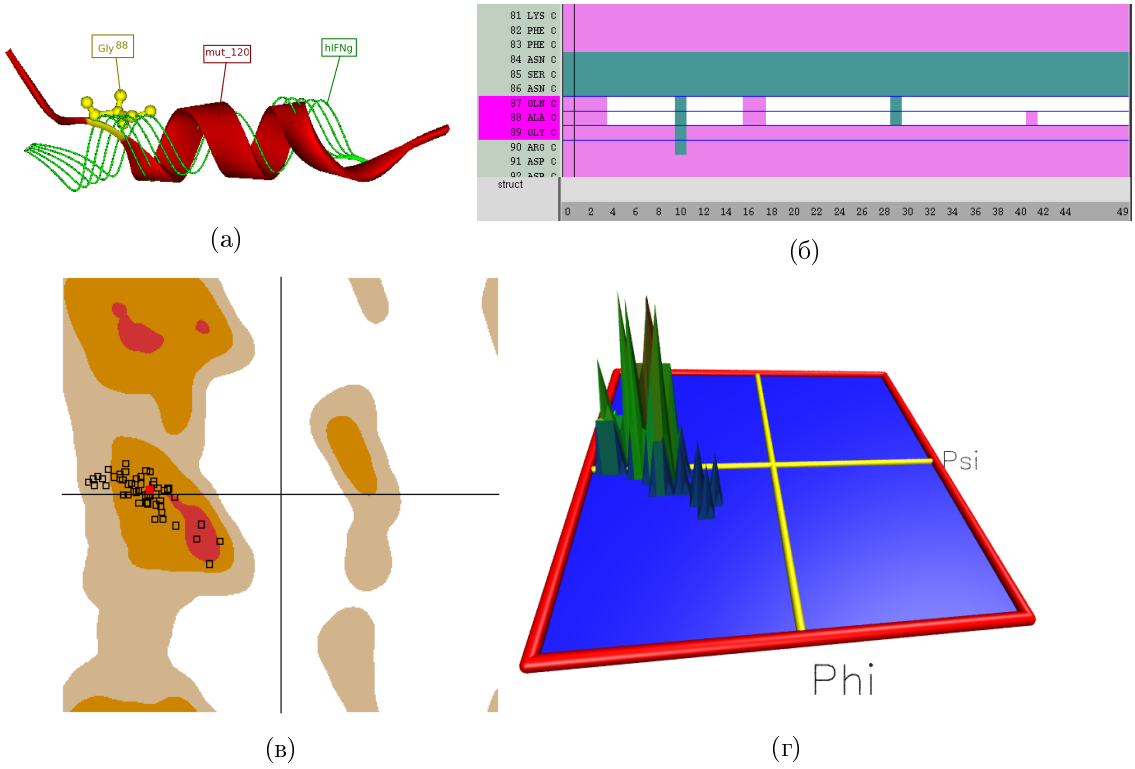
4.5 Резултати

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.

Резултатите от симулациите са представени в пета колона на Таблица 4. Мутациите, които водят до изменения в стабилността и дължината на α-спиралата са означени с 1, а тези, които запазват локалната структура непроменена - с 0.

Наблюдаваха се два случая на прекъсване на спиралата. В първия случай аминокиселинен остатък Lys^{88} (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.

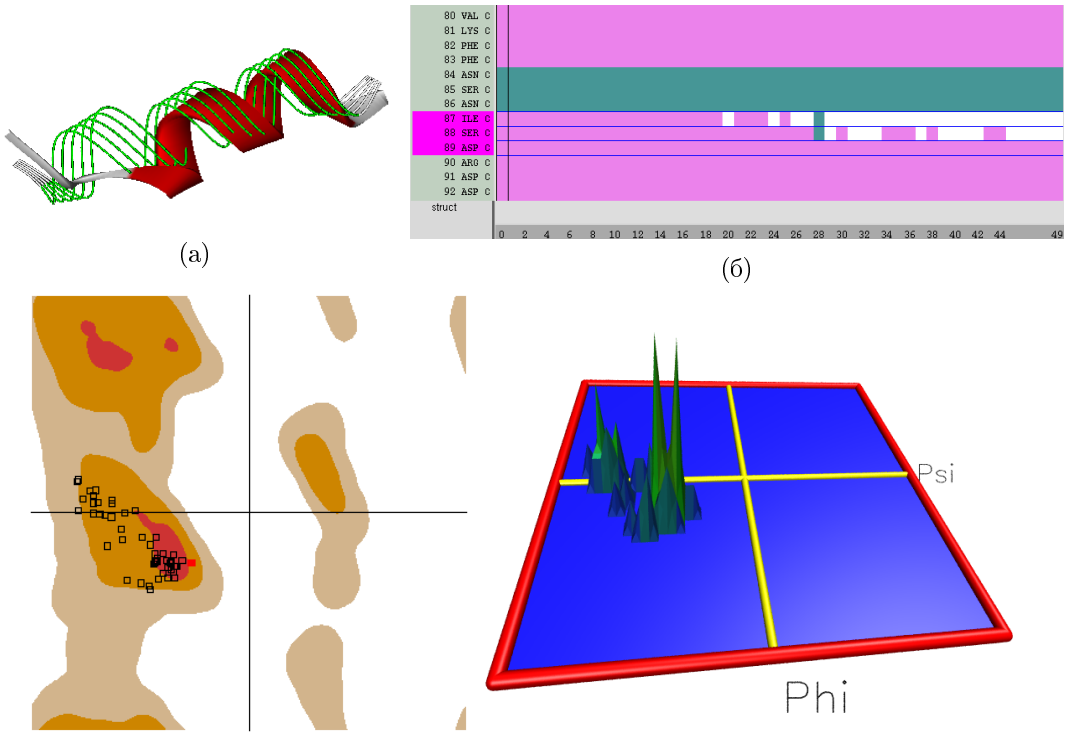


Фигура 29: (а) Крайна конфигурация на спирала E на нативната форма на IFN-γ (в зелено) с насложена тази на мутант №120 (в червено), Gly^{88} (в жълто); (б) Вторична структура на спирала E в мутант №120, определена по метода DSSP. По вертикалната ос са дадени поредните номера на аминокиселините и техните имена. В розово са маркирани трите мутирани аминокиселини. По хоризонталната ос за нанесени кадрите от МД симулацията, като 1 кадър съответства на 0.2 ns. α-спиралите са изобразени във виолетово, бримките – в зелено, а неструктурираните части – в бяло; (в) Диаграма на Рамачандран на Gln^{86}; (г) Тримерна хистограма на диаграмата на Рамачандран на Gln^{86}.

Например при мутагенна форма №120 (Gln^{86} – Ala – Gly^{88}) на фиг. 29а се вижда как глицинът прекъсва спиралата. Изменението на структурата в този случай настъпва още

в самото начало на симулацията, краят на спиралата се разплита и мутираните аминокиселини приемат неструктурирана конформация, както се вижда на фиг. 29б. Въпреки че конформацията на 86-тата аминокиселина остава в разширената α-спирална област на диаграмата (фиг. 29в), това не е достатъчно условие тя да участва в α-спирала. В случая тази конформация съответства на бримка.

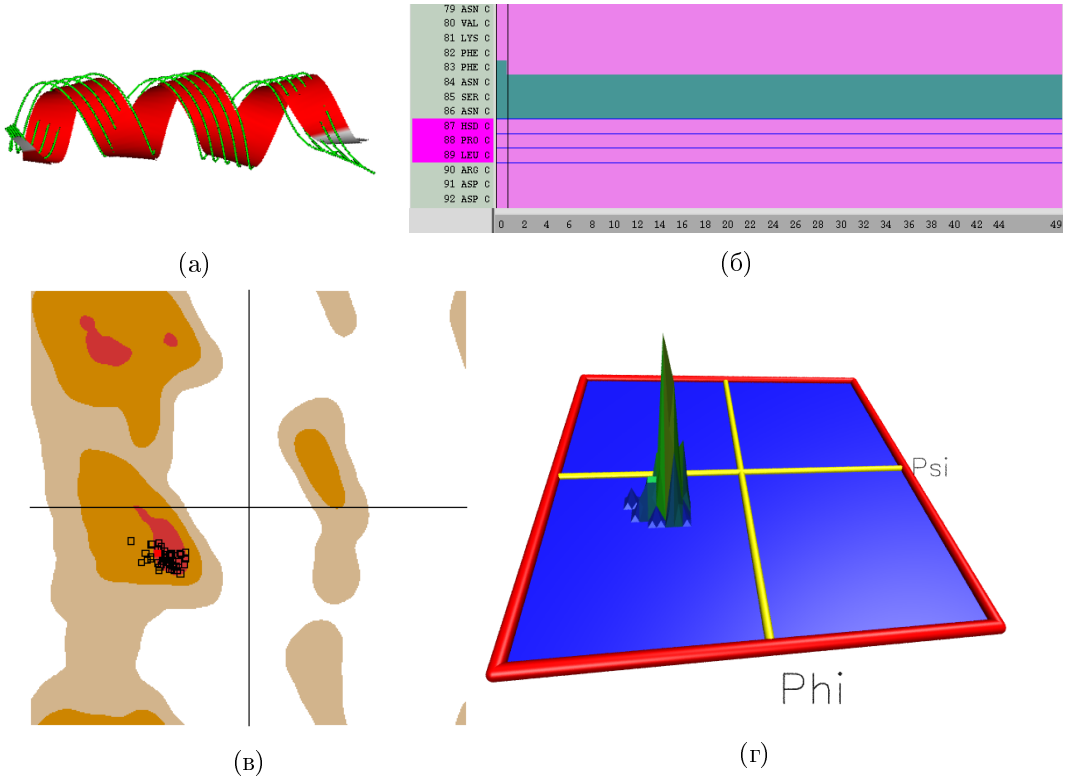
Във втория случай на позиция 88 няма "терминатор", но въпреки това спиралата се прекъсва. Например при мутагенна форма №155 (Ile^{86} – Ser – Asp^{88}) на фиг. 30а спиралата се скъсва при Asp^{88}. На графиката с вторичната структура (фиг. 30б) се вижда, че след средата на симулацията (след 6-тата ns) краят на спиралата се разплита, мутираните аминокиселини приемат неструктурирана конформация и спиралата се прекъсва.



(а) (б) (в) (г)

Фигура 30: (а) Крайна конфигурация на спирала E на нативната форма на IFN-γ (в зелено) с насложена тази на мутант №155 (в червено); (б) Вторична структура на спирала E в мутант №155, определена по метода DSSP. По вертикалната ос са дадени поредните номера на аминокиселините и техните имена. В розово са маркирани трите мутирани аминокиселини. По хоризонталната ос за нанесени кадрите от МД симулацията, като 1 кадър съответства на 0.2 ns. α-спиралите са изобразени във виолетово, бримките – в зелено, а неструктурираните части – в бяло; (в) Диаграма на Рамачандран на Ile^{86}; (г) Тримерна хистограма на диаграмата на Рамачандран на Ile^{86}.

При мутациите, които не предизвикват изменения във вторичната структура, аминокиселинни остатъци 86-88 остават в по-строго дефинираната зона на α-спиралите в диаграмата на Рамачандран през цялата траектория, например при мутагенна форма №61 (His^{86} – Pro – Leu^{88}) (фиг. 31в,31г). Стабилността на спиралата, в която участват мутираните остатъци, се вижда и на фиг. 31б, като по време на цялата симулация вторичната структура не се изменя.



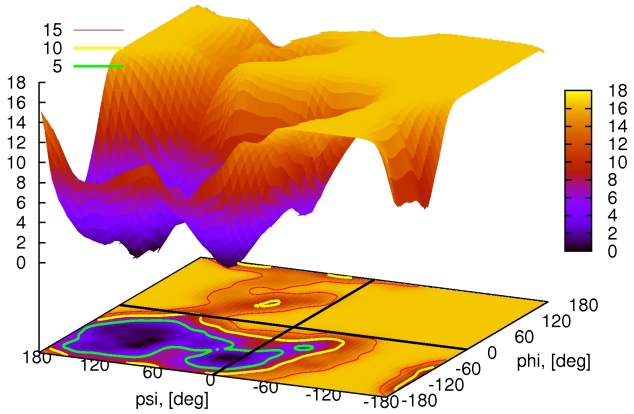
Фигура 31: (а) Крайна конфигурация на спирала E на нативната форма на IFN-γ (в зелено) с насложена тази на мутант №61 (в червено); (б) Вторична структура на спирала E в мутант №120, определена по метода DSSP. По вертикалната ос са дадени поредните номера на аминокиселините и техните имена. В розово са маркирани трите мутирани аминокиселини. По хоризонталната ос за нанесени кадрите от МД симулацията, като 1 кадър съответства на 0.2 ns. α-спиралите са изобразени във виолетово, бримките – в зелено, а неструктурираните части – в бяло; (в) Диаграма на Рамачандран на His^{86}; (г) Тримерна хистограма на диаграмата на Рамачандран на His^{86}.

4.5.2 Метадинамични симулации

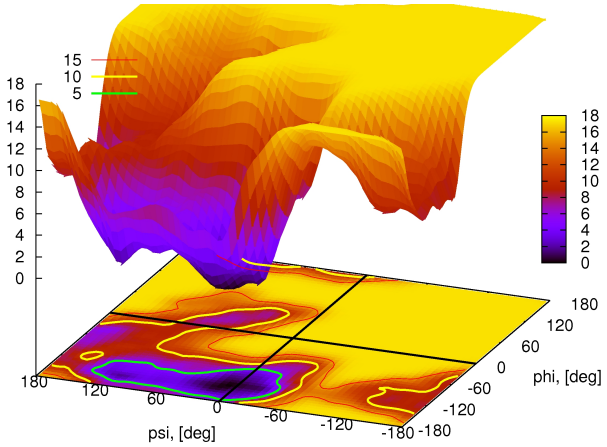
Целта на метадинамичните симулации беше от 100-те изследвани мутанта да се отсеят само няколко на брой, които не само да запазват локалната вторична структура, но и тя да остава с подобна или по-добра стабилност от тази на нативната форма на hIFN-γ . Метадинамиката не само дава допълнителна информация,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

Нека разгледаме повърхнините на свободната енергия в пространството на ъглите на усукване на гръбнака на аминокиселинния остатък на позиция 86 на мутагенните форми на hIFN-γ , дадени като пример в секция 4.5.1. На фиг. 32 е представен профилът на свободната енергия за мутант №120. Вижда се, че Gln^{86} изобщо не участва в α-спирала, защото е излязъл от тази област на диаграмата на Рамачандран още по време на МД симулацията. Подобен профил се наблюдава при мутант №155 на фиг. 33, Ile^{86} не участва в α-спирала, защото е приел разгъната конформация 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, профилът на повърхнината на свободната енергия

на мутагенна форма №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 $\pm$ 0.2 kcal/mol) от тази на нативната форма (16.3 $\pm$ 0.2 kcal/mol). Такива мутагенни форми имат ниски стойности на S ($<$ 0.17 kcal/mol).

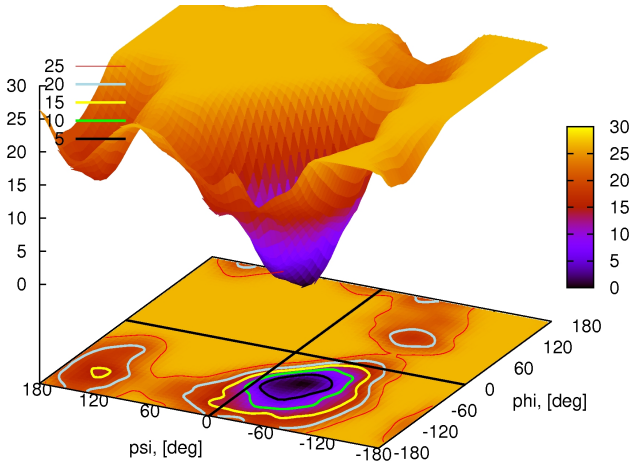


Фигура 32: Повърхнина на свободната енергия в пространството на ъглите на усукване на гръбнака на аминокиселина Gln^{86} на мутант №120.

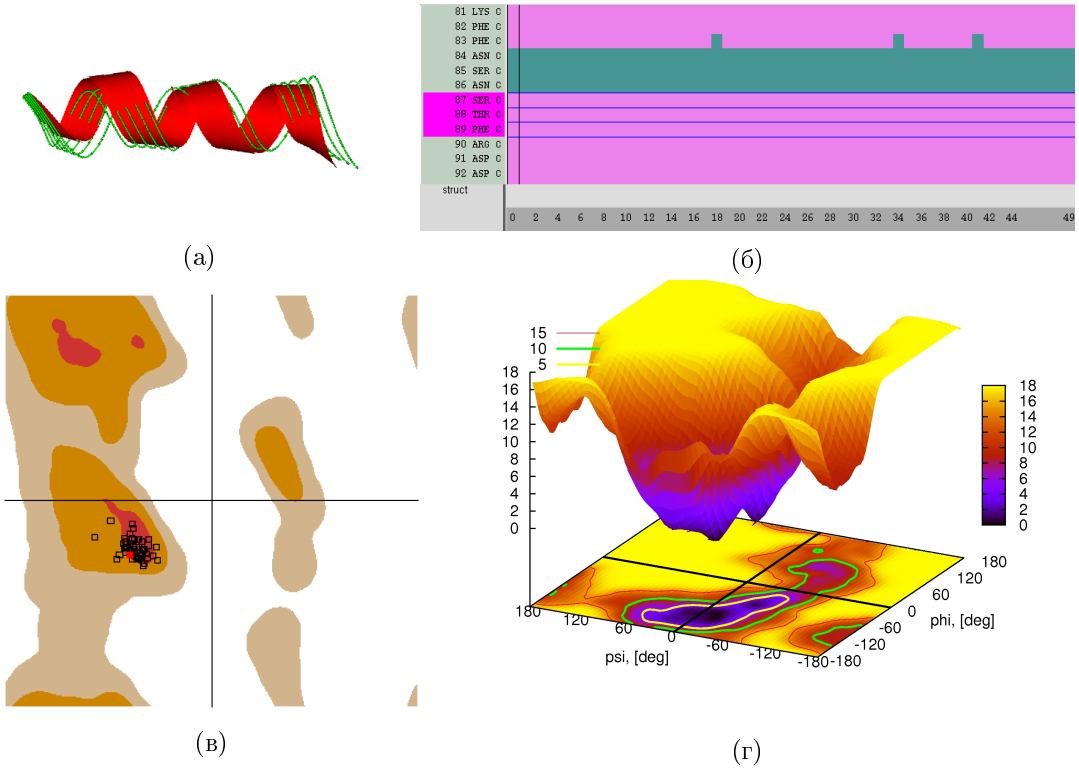


Фигура 33: Повърхнина на свободната енергия в пространството на ъглите на усукване на гръбнака на аминокиселина Ile^{86} на мутант №155.

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. Например при мутант №19 на фиг. 35а, 35б и 35в се вижда, че по време на стандартната МД симулация спирала E остава стабилна. Профилът на свободната енергия (фиг. 35г) обаче показва, че ямата е доста широка и излиза от областта на α-спиралите. Ето защо, въпреки че по време на МД симулацията мутагенната форма изглежда стабилна и запазваща локалната структура на спирала E, метадинамичната симулация показва, че мутацията може да предизвика конформационни изменения.

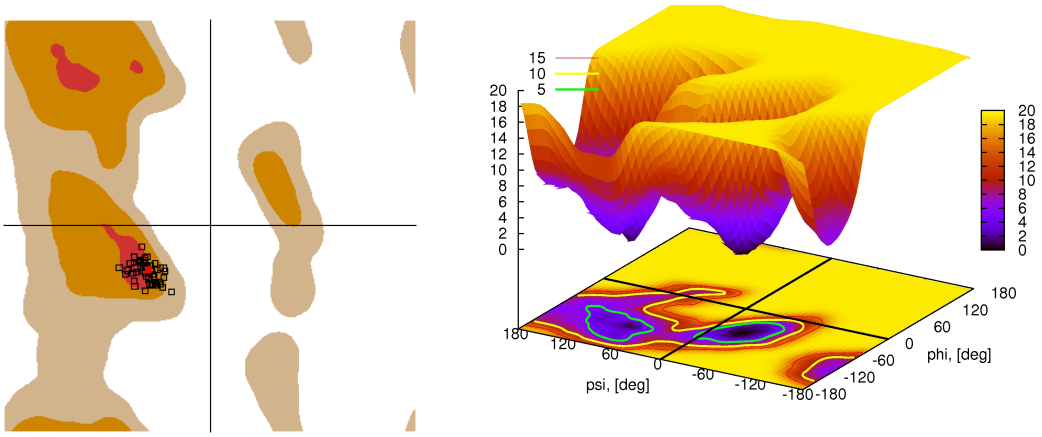


Фигура 34: Повърхнина на свободната енергия в пространството на ъглите на усукване на гръбнака на аминокиселина His^{86} на мутант №61.



Фигура 35: (а) Крайната конформация от МД симулация на спирала E на нативен hIFN-γ (в зелено) с насложена тази на мутант №19 (в червено); (б) Вторична структура на спирала E в мутант №19, определена по метода DSSP. По вертикалната ос са дадени поредните номера на аминокиселините и техните имена. В розово са маркирани трите мутирани аминокиселини. По хоризонталната ос за нанесени кадрите от МД симулацията, като 1 кадър съответства на 0.2 ns. α-спиралите са изобразени във виолетово, бримките – в зелено, неструктурираните части – в бяло; (в) Диаграма на Рамачандран на аминокиселинен остатък Ser^{86} на мутант №19; (г) Повърхнина на свободната енергия в пространството на ъглите на усукване на гръбнака на аминокиселина Ser^{86} на мутант №19.

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. Пример за такъв мутант е мутант №33. Диаграмата на Рамачандран на аминокиселинен остатък Pro^{86} от МД симулацията е дадена на 36а. Ъглите на усукване на Pro^{86} остават в областта на α-спиралите по време на цялата симулация. Профилът на повърхнината на свободната енергия в пространството на ъглите на усукване на гръбнака на аминокиселина Pro^{86} е представен на фиг. 36б. S от метадинамичната симулация на тази мутагенна форма е доста нисък (0.15457 kcal/mol), но бариерът, разделяща нагънато от разгънато състояние ΔG е само 7.8 kcal/mol. Причината спиралата да остава стабилна по време на МД симулацията е краткото симулационно време. За 10 ns системата не е успяла да преодолее тези близо 8 kcal/mol и да приеме нова конформация.



(а) (б)

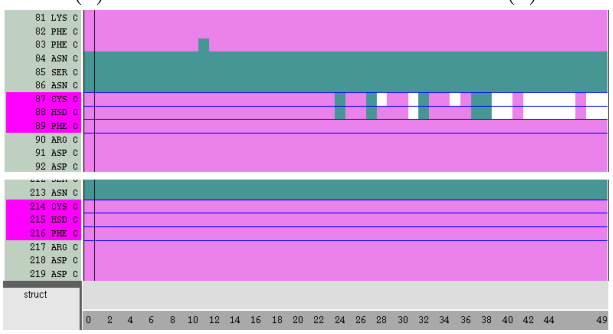
Фигура 36: (а) Диаграма на Рамачандран на аминокиселинен остатък Pro^{86} на мутант №33; (б) Повърхнина на свободната енергия в пространството на ъглите на усукване на гръбнака на аминокиселина Pro^{86} на мутант №33.

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. Такъв е случаят на мутант №25. Диаграмите на Рамачандран за двата остатъка Cys^{86} във едната верига (фиг. 37а) и другата верига (фиг. 37б) се различават значително. Настъпването на конформационни изменения в първата верига и липсата на такива във втората се вижда и на фигурата с вторичната структура 37в. Реконструираният профил на свободната енергия (фиг. 37г) дава обяснение на това поведение. Както се вижда, 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 Cys^{86} в едната подединица да навлезе в областта на разгънатите конформации, а в другата подединица да остане в спиралата.

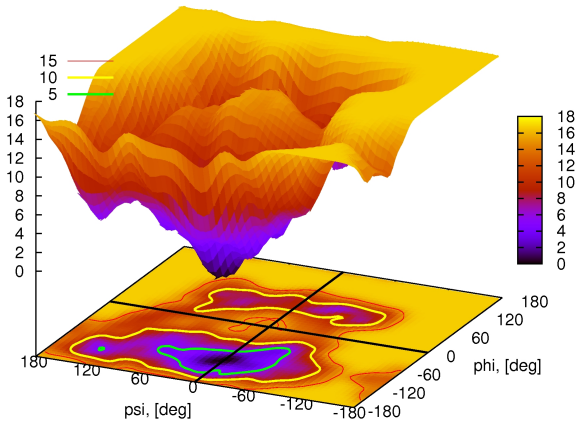
Обобщение на резултатите от молекулно-динамичните и метадинамичните симулации за всички мутагенни форми на hIFN-γ е дадено в таблица 4. В последната колона на таблицата като запазващи локалната структура на спирала E и като достатъчно стабилни са избрани тези мутагенни форми, които отговарят на следните 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;

(а) (б)



(в)



(г)

Фигура 37: (а) Диаграма на Рамачандран на аминокиселинен остатък Cys^{86} в първата подединица на мутант №25; (б) Диаграма на Рамачандран на аминокиселинен остатък Cys^{86} във втората подединица на мутант №25; (в) Вторична структура на спирала E в мутант №25, определена по метода DSSP. По вертикалната ос са дадени поредните номера на аминокиселините и техните имена. В розово са маркирани трите мутирани аминокиселини. По хоризонталната ос за нанесени кадрите от МД симулацията, като 1 кадър съответства на 0.2 ns. α-спиралите са изобразени във виолетово, извивките - в зелено, неструктурираните части в бяло; (г) Повърхнина на свободната енергия в пространството на ъглите на усукване на гръбнака на аминокиселина Cys^{86} на мутант №25.

- 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.

Таблица 4: Резултати от молекулно-динамичните и метадинамичните симулации на мутагенните форми на hIFN-γ . В първата колона са идентификационните номера на мутантите, в следващите три колони са аминокиселините, с които са заместени съответно Lys^{86}, Lys^{87} и Lys^{88}. В пета колона с 0 са означени мутантите, при които локалната вторична структура се запазва непроменена по време на МД симулациите, а с 1 – тези, при които структурата се изменя. В шеста колона са дадени стойностите на параметъра S (у-ние 69). В седма колона са представени стойностите в kcal/mol на бариера, разделящ областите на α-спиралите и разгънатите конформации в диаграмата на Рамачандран, като символът "-" означава, че аминокиселинния остатък на позиция 86 е приел разгъната конформация още по време на МД симулациите и не се наблюдава минимум в α-спиралната област. В последната колона със "-" са означени мутагенните форми, за които S $>$ 0.155 kCal/mol и/или $\Delta G <$ 13 kCal/mol. С "+" са означени мутантите, които при избраните радиуси на обрязване се считат за запазващи стабилността на локалната вторична структура на спирала E.

м.ф. №	а.к. 86	а.к. 87	а.к. 88	МД	S [kcal/mol]	ΔG [kcal/mol]	МетаД
2	Pro	Tyr	Leu	0	0.18741	14.7	-
3	Pro	Asn	Tyr	0	0.20645	15.5	-
4	Trp	Ser	Ser	1	0.28884	-	-
5-3	Val	Ser	Arg	0	0.15756	14.1	-
5-4	Pro	Leu	Ser	1	0.13350	13.0	+
6	His	Val	Cys	0	0.19064	>25.	-
8	Pro	Tyr	Val	0	0.16867	17.5	-
9-1	Arg	Ser	Ser	1	0.29564	5.5	-
9-2	Phe	Ser	Arg	1	0.29070	-	-
10	Pro	Cys	Cys	0	0.21008	12.4	-
11	Pro	Ser	Val	0	0.18656	14.3	-
12	Thr	Phe	Trp	0	0.16218	17.3	-
14	Leu	Pro	Phe	0	0.12005	24.4	+
15	Asp	Leu	Leu	0	0.24653	-	-
16	Ala	His	Leu	0	0.22780	>20.	-
17	Thr	Val	Leu	1	0.21244	-	-
18	Cys	Phe	Pro	1	0.15658	6.7	-
19	Ser	Thr	Phe	0	0.27390	1.9	-
21	Pro	Ser	Pro	1	0.25876	-	-
22	Ser	Ser	Leu	0	0.15407	20.0	+
23	Val	Ser	Gly	1	0.21778	9.7	-
24	Thr	Pro	Thr	0	0.26041	>20.	-

Продължава на следващата страница

Таблица 4 – продължение

м.ф. №	а.к. 86	а.к. 87	а.к. 88	MD	S [kcal/mol]	ΔG [kcal/mol]	Meta
25	Cys	His	Phe	1	0.27552	-	-
26	Ser	Val	Ser	1	0.27289	-	-
27	Glu	Met	Pro	1	0.20379	7.1	-
28	Leu	Thr	Pro	1	0.28413	9.8	-
32	Leu	Pro	Pro	1	0.18654	>20.	-
33	Pro	Pro	Thr	0	0.15457	7.8	-
34	Phe	Ser	Leu	0	0.17665	>20.	-
35	Phe	Phe	Pro	1	0.26996	-	-
36	Leu	Cys	Pro	1	0.29187	-	-
39-11	Pro	Ser	Ala	0	0.21507	-	-
39-12	Asp	Leu	Leu	0	0.15614	>20	-
41	Ala	Phe	Phe	0	0.24639	13.6	-
45	Leu	Leu	His	1	0.25132	-	-
46-1	Thr	Leu	Leu	0	0.21630	15.6	-
54	Phe	Thr	Ala	1	0.26521	-	-
61	His	Pro	Leu	0	0.10548	24.3	+
62	Phe	Thr	Arg	0	0.13945	21.1	+
63	Arg	Leu	Arg	0	0.16698	23.2	-
66	Pro	Leu	Ala	0	0.13594	12.9	-
71	Phe	Cys	Arg	1	0.18528	17.3	-
72	His	Ser	Arg	0	0.17536	>20.	-
73	Pro	Tyr	Pro	1	0.21464	20.7	-
74	Ser	Leu	Leu	0	0.23474	11.4	-
75	Trp	Ser	Ala	1	0.27041	-	-
77	Ala	Ile	Pro	1	0.29079	11.8	-
81	Arg	Pro	Val	0	0.17622	11.3	-
82-34	Phe	Cys	Arg	0	0.17497	-	-
84	Pro	Phe	Ala	0	0.17603	-	-
85	Arg	Arg	Ser	0	0.13516	12.3	-
87	Arg	Pro	Ser	0	0.15533	18.1	+
88	Arg	Ser	Cys	1	0.25147	4.8	-
92	Pro	Phe	Leu	0	0.16049	22.6	-
93	Leu	Tyr	Pro	1	0.22813	-	-
94	Pro	Val	Phe	0	0.17632	23.1	-
96	Pro	Met	Phe	0	0.17323	23.9	-
97	Ser	Phe	Phe	0	0.17648	>20.	-
103	His	Ala	Ala	0	0.24031	8.0	-
104	Pro	Phe	Ser	1	0.18752	>20.	-
105-2	Ala	Thr	Ala	1	0.27478	15.7	-
106-1	Leu	Phe	Ser	0	0.18761	17.4	-
106-2	Leu	Val	Ser	1	0.24389	-	-
108	Phe	Leu	Val	0	0.13894	>20.	+
109	Pro	Arg	Ser	0	0.20172	19.2	-
110	Pro	Arg	Ser	0	0.20695	13.5	-
111	Pro	Arg	Ser	0	0.18081	-	-

Продължава на следващата страница

Таблица 4 – продължение

м.ф. №	а.к. 86	а.к. 87	а.к. 88	MD	S [kcal/mol]	ΔG [kcal/mol]	Meta
112	Phe	Ser	Arg	1	0.27937	-	-
113	Leu	Tyr	Phe	0	0.17196	>20.	-
114-1	Arg	Ser	Ala	0	0.21459	>20.	-
115	Gln	Phe	His	0	0.23558	>20.	-
116	Val	Leu	Leu	0	0.12167	>20.	+
117	Val	Leu	Pro	1	0.19199	15.0	-
118	Val	Ser	Ala	1	0.26378	-	-
119	Thr	Leu	Val	0	0.25228	>20.	-
120	Gln	Ala	Gly	1	0.26074	-	-
121	Leu	Ser	Val	0	0.19477	22.1	-
123	Ser	Leu	Phe	0	0.19794	20.2	-
126	Tyr	Ala	Phe	1	0.20386	>20.	-
127-1	His	Tyr	Pro	1	0.26517	16.0	-
129	Ala	Ser	Leu	0	0.24421	>20.	-
131	Phe	Pro	Leu	0	0.22081	10.3	-
133	Pro	Pro	Ser	0	0.15484	17.6	+
134	Thr	Asn	Gly	1	0.25833	15.9	-
135	Val	Ser	Pro	1	0.14989	14.8	+
136	Ser	Pro	Pro	1	0.19614	21.6	-
140	Phe	Pro	Ser	0	0.16430	18.6	-
143	Cys	Ser	Pro	1	0.26660	-	-
144	Cys	Ala	Pro	1	0.25912	9.2	-
145	Ser	Phe	Cys	0	0.27522	>20.	-
146	Leu	Phe	Glu	0	0.25208	>20.	-
148	Phe	Thr	Pro	1	0.26454	-	-
149-1	His	Gln	Arg	0	0.13683	>20.	+
150-1	Leu	Ser	Ser	1	0.24400	-	-
150-2	Trp	Leu	Ser	1	0.27979	-	-
151	Leu	Thr	Ala	1	0.25945	-	-
153	Ser	Phe	Cys	0	0.14239	>20.	+
155	Ile	Ser	Asp	1	0.29210	-	-
157	Phe	Tyr	Thr	0	0.24915	-	-
158	Pro	Leu	Phe	0	0.19886	>20.	-

Глава 5

Молекулно моделиране на комплекса hIFNI-γ–hIFNγR–производни на хепарина олигозахариди

Както стана ясно в секция 1.4.4, 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 (виж секции 1.4.1 и 1.4.2 и фиг. 6e). 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

5.1.1 Избор на силово поле

Изборът на силово поле при параметризацията на нови молекули е до голяма степен ограничен от начина на параметризацията на самото силово поле. Някои от популярните биомолекулни полета като GROMOS [85] и OPLS [86] се параметризират изключително чрез фитиране към експериментални данни като плътност и специфична топлина на изпарение на течности. Теоретично такива емпирични полета могат да бъдат допълвани с нови параметри, но на практика това се случва само от екипите, които разработват самото поле, защото обикновено параметризационната процедура е изключително неясна и зле документирана.

Полу-емпиричните силови полета като AMBER [87] и CHARMM [84] са значително по-подходящи за допълване с параметри за нови молекули. При тях също се използват експериментални данни за определяне на Ван дер Ваалсовите параметри на различните типове атоми, но останалите параметри, например равновесни дължини на химичните връзки, силови константи и частични заряди и всички останали параметри в (4) се параметризират с помощта на *ab initio* квантово-механични симулации. И двете полета разполагат с разнообразен набор от вече налични параметри за различни класове съединения – протеини, нуклеинови киселини, липиди, въглехидрати. Освен това и AMBER, и CHARMM имат т. нар. общи силови полета, чиято цел е именно да се използват за параметризация на нови биологично активни съединения, наречени лиганди. Тези полета предлагат Ван дер Ваалсови

параметри за определен набор от типове атоми, често срещани в малките биологично активни молекули. При AMBER този параметричен набор се нарича GAFF (General AMBER Force Field) [88], а при CHARMM – CGenFF (CHARMM General Force Field) [89]. И двете полета са добре документирани с подробна информация за параметризационната процедура.

За параметризация на N-сулфатиран глюкозамин сулфат избрахме да използваме полето CHARMM 36 [90]. Файловете с различните набори топологии и параметри могат да се свалят свободно от сайта на групата на Маккерълъ [91]. Полето за въглехидрати [92] съдържа набор от параметри за множество различни захариди, гликозидни връзки и функционални групи.

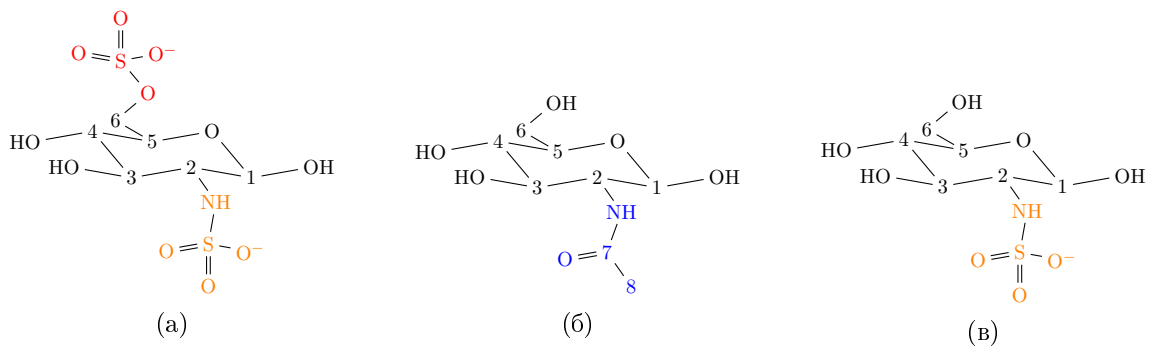
5.1.2 Топология и параметри за L-IdoA(2S)

CHARMM36 полето за въглехидрати съдържа топология и параметри за α-L-идуронова киселнина (`RESI AIDOA` във файла `top_all36_carb.rtf`) [92]. Освен това в полето има параметри и за сулфатната група при атом C_2 (`PRES SUF2A`) [93].

Пълната топология на 2-O-сулфатирана-α-L-идуронова киселина е дадена със запис "`RESI IDS`" в Приложение A.

5.1.3 Начални параметри за D-Glc2NS

Въглехидратното поле CHARMM36 предлага топология и параметри за 2-ацетил-2-дезокси-α-D-глюкозамин (`RESI AGLCNA` във файла `top_all36_carb.rtf`, фиг. 38б) [92]. Освен това има и параметри за O-сулфатната група при C_6 (`PRES SUF6`) [93].

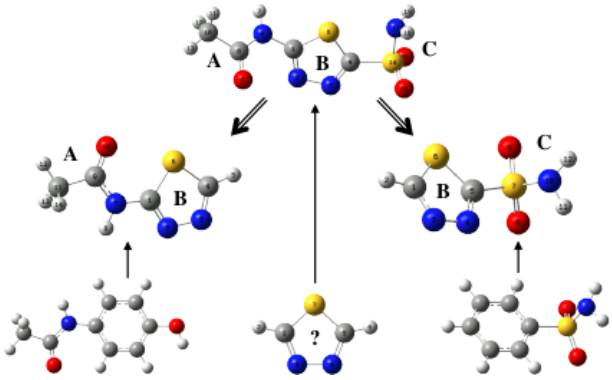


Фигура 38: Различни производни на D-глюкозамин. Ацетилната група е оцветена в синьо, сулфатната – в червено и N-сулфатните – в оранжево. (а) 2-N-сулфатиран-D-глюкозамин-6-сулфат. Това е монозахарида, който участва в основната дизахаридна последователност, която изгражда хепарин и NS-областите на ХС; (б) 2-ацетил-2-дезокси-D-глюкозамин. Този монозахарид участва в NA-областите на ХС. 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 сулфаматната анионна група NSO_3^-.

В литературата няма разработени параметри за сулфаматната анионна група NSO_3^- при C_2 (представена в оранжево на фиг.38а). За параметризацията на тази група беше използвано моделно съединение – 2-N-сулфатиран глюкозамин (фиг.38в).

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 (фиг.39). 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`.



Фигура 39: Модулната идеология при параметризацията на нови молекули в CHARMM. Вместо да се параметризира наведнъж цялата сложна молекула на най-горния ред, съединението се разделя на малки моделни съединения, съдържащи съответните функционални групи (последния ред), като след това се параметризират само връзките между отделните функционални групи (средния ред). Графиката е от [94].

За дефиниране на начални параметри за новата функционална група беше използван сървърът Paramchem [95–97]. На него върви програмата CGenFF от пакета за МД CHARMM [98] 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. Наказателни оценки до 10 се считат за много ниски и са показател за много добра аналогия с други съединения. Оценки между 10 и 50 се считат за добри, но изискват допълнителна валидация на параметрите от страна на потребителя. Оценки над 50 се дават на параметри, за които не е намерена достатъчно добра аналогия с други молекули в полето и изискват допълнителна оптимизация.

Изходният файл, който Paramchem генерира е т.нар. “stream” (`.str`) файл, който съдържа едновременно начална топология и начални предположения за параметрите в CHARMM формат. Началният stream-файл, описващ N-сулфатиран глюкозамин а представен на фиг. 40. Той е комбинация от непозната топология и параметри на новата функционална група, получени от Paramchem и вече известните параметри на останалите атоми от `RESI AGLCNA`. В секциите `BONDS`, `ANGLES` и `DIHEDRALS` са изброени само параметрите, които не фигурират във файловете `par_all36_carb.prm` и `par_all36_cgen.prm`

model.str

```
* Toppar stream file generated by
* CHARMM General Force Field (CGenFF) program version 0.9.7.1 beta
* For use with CGenFF version 2b8

read rtf card append
* Topologies generated by
* CHARMM General Force Field (CGenFF) program version 0.9.7.1 beta
*
36 1
```

```
! "penalty" is the highest penalty score of the associated parameters.
! Penalties lower than 10 indicate the analogy is fair; penalties between 10
! and 50 mean some basic validation is recommended; penalties higher than
! 50 indicate poor analogy and mandate extensive validation/optimization.

RESI /scrat      -1.000 ! param penalty= 201.500 ; charge penalty= 143.077
GROUP            ! CHARGE   CH_PENALTY
ATOM C1     CC3162  0.338 !   14.091
ATOM H1     HCA1    0.090 !    0.634
ATOM O1     OC311  -0.645 !    8.019
ATOM HO1    HCP1    0.408 !    0.925
ATOM C2     CC3161  0.004 !   18.456
ATOM H2     HCA1    0.090 !    1.940
ATOM N      NG2S3  -0.889 !  143.077! typer WARNING: sulfamate anion not explicitly supported
ATOM HN     HCP1    0.327 !    8.873
ATOM S1     SG3O1   1.316 !  142.557! typer WARNING: sulfamate anion not explicitly supported
ATOM O1S    OG2P1  -0.614 !    4.218
ATOM O2S    OG2P1  -0.614 !    4.218
ATOM O3S    OG2P1  -0.614 !    4.218
ATOM C3     CC3161  0.141 !    6.777
ATOM H3     HCA1    0.090 !    0.450
ATOM O3     OC311  -0.653 !    3.195
ATOM HO3    HCP1    0.419 !    0.000
ATOM C4     CC3161  0.149 !    3.978
ATOM H4     HCA1    0.090 !    0.600
ATOM O4     OC311  -0.655 !    0.732
ATOM HO4    HCP1    0.419 !    0.000
ATOM C5     CC3163  0.098 !    7.116
ATOM H5     HCA1    0.090 !    0.446
ATOM C6     CC321   0.055 !    2.798
ATOM H61    HCA2    0.090 !    0.300
ATOM H62    HCA2    0.090 !    0.300
ATOM O6     OC311  -0.651 !    0.626
ATOM HO6    HCP1    0.421 !    0.000
ATOM O5     OC3C61 -0.390 !   13.136

BOND O3S  S1  HO3  O3  HN   N   O2S  S1  S1   N
BOND S1   O1S O3   C3  N    C2  H3   C3  C3   C2
BOND C3   C4  HO4  O4  C2   H2  C2   C1  O4   C4
BOND C4   H4  C4   C5  HO1  O1  H1   C1  C1   O1
BOND C1   O5  H5   C5  C5   O5  C5   C6  HO6  O6
BOND H62  C6  C6   O6  C6   H61

END

read param card flex append
* Parameters generated by analogy by
* CHARMM General Force Field (CGenFF) program version 0.9.7.1 beta

! Penalties lower than 10 indicate the analogy is fair; penalties between 10
! and 50 mean some basic validation is recommended; penalties higher than
! 50 indicate poor analogy and mandate extensive validation/optimization.

BONDS
CC3161 NG2S3   261.00     1.4900 ! /scrat , from CG331 NG2S3, penalty= 10
NG2S3  SG3O1   180.00     1.7920 ! /scrat , from NG2S3 PG1, penalty= 200

ANGLES
CC3161 CC3161 NG2S3    70.00    113.50 ! /scrat , from CG311 CG311 NG2S1, penalty= 9
CC3162 CC3161 NG2S3    70.00    113.50 ! /scrat , from CG311 CG311 NG2S1, penalty= 9
```

```
NG2S3  CC3161 HCA1     48.00    108.00 ! /scrat , from NG2S1 CG311 HGA1, penalty= 9
CC3161 NG2S3  SG3O1   110.00    118.30   35.00   2.33000 ! /scrat , from CG331 NG2S3 PG1,
                                                                          penalty= 17.5
CC3161 NG2S3  HCP1     35.00    109.00 ! /scrat , from CG331 NG2S3 HGP1, penalty= 1.5
SG3O1  NG2S3  HCP1     30.00    123.60   40.00   2.35000 ! /scrat , from PG1 NG2S3 HGP1,
                                                                            penalty= 16
NG2S3  SG3O1  OG2P1   130.00    109.47   35.00   2.45000 ! /scrat , from OG2P1 SG3O1 OG2P1,
                                                                            penalty= 46

DIHEDRALS
CC3161 CC3161 CC3161 NG2S3      0.2000  3     0.00 ! /scrat , from CG321 CG311 CG311 NG2S1,
                                                                           penalty= 9.6
NG2S3  CC3161 CC3161 OC311      0.2000  3     0.00 ! /scrat , from NG2S1 CG311 CG311 OG311,
                                                                             penalty= 9
NG2S3  CC3161 CC3162 OC3O1      0.2000  3     0.00 ! /scrat , from NG2S1 CG311 CG311 OG311,
                                                                             penalty= 9
NG2S3  CC3161 CC3162 OC311      0.2000  3     0.00 ! /scrat , from NG2S1 CG311 CG311 OG311,
                                                                             penalty= 9
NG2S3  CC3161 CC3162 OC3C61     0.2000  3     0.00 ! /scrat , from NG2S1 CG311 CG311 OG311,
                                                                            penalty= 24
NG2S3  CC3161 CC3161 HCA1       0.2000  3     0.00 ! /scrat , from NG2S1 CG311 CG311 HGA1,
                                                                             penalty= 9
NG2S3  CC3161 CC3162 HCA1       0.2000  3     0.00 ! /scrat , from NG2S1 CG311 CG311 HGA1,
                                                                             penalty= 9
CC3161 CC3161 NG2S3  SG3O1      1.5000  1   180.00 ! /scrat , from CG3C52 CG3C51 NG2S3 PG1,
                                                                           penalty= 131
CC3161 CC3161 NG2S3  SG3O1      0.6500  2   180.00 ! /scrat , from CG3C52 CG3C51 NG2S3 PG1,
                                                                           penalty= 131
CC3161 CC3161 NG2S3  SG3O1      1.0000  3     0.00 ! /scrat , from CG3C52 CG3C51 NG2S3 PG1,
                                                                           penalty= 131
CC3162 CC3161 NG2S3  SG3O1      1.5000  1   180.00 ! /scrat , from CG3C52 CG3C51 NG2S3 PG1,
                                                                           penalty= 131
CC3162 CC3161 NG2S3  SG3O1      0.6500  2   180.00 ! /scrat , from CG3C52 CG3C51 NG2S3 PG1,
                                                                           penalty= 131
CC3162 CC3161 NG2S3  SG3O1      1.0000  3     0.00 ! /scrat , from CG3C52 CG3C51 NG2S3 PG1,
                                                                           penalty= 131
CC3161 CC3161 NG2S3  HCP1       0.0000  1     0.00 ! /scrat , from CG311 CG311 NG2S1 HGP1,
                                                                            penalty= 40
CC3162 CC3161 NG2S3  HCP1       0.0000  1     0.00 ! /scrat , from CG311 CG311 NG2S1 HGP1,
                                                                            penalty= 40
HCA1   CC3161 NG2S3  SG3O1      0.1500  3     0.00 ! /scrat , from HGA3 CG331 NG2S3 PG1,
                                                                            penalty= 27
HCA1   CC3161 NG2S3  HCP1       0.0100  3     0.00 ! /scrat , from HGA3 CG331 NG2S3 HGP1,
                                                                            penalty= 11
CC3161 NG2S3  SG3O1  OG2P1      0.4000  3     0.00 ! /scrat , from CG331 NG2S3 PG1 OG2P1,
                                                                         penalty= 201.5
CC3161 NG2S3  SG3O1  OG2P1      0.5000  4     0.00 ! /scrat , from CG331 NG2S3 PG1 OG2P1,
                                                                         penalty= 201.5
HCP1   NG2S3  SG3O1  OG2P1      0.0500  3     0.00 ! /scrat , from HGP1 NG2S3 PG1 OG2P1,
                                                                           penalty= 200

IMPROPERS

END
RETURN
```

Фигура 40: Началният stream-файл, описващ N-сулфатиран глюкозамин.

5.1.4 Параметризация на D-Glc2NS

За оптимизиране на параметрите с висока наказателна оценка (> 12) от Paramchem беше използван плъгина Force Field Toolkit [99] (FFTK) на пакета за визуализация и манипулация на молекули VMD [100]. Той съдържа няколко инструмента, които улесняват потребителя в разработването на съвместими със силовото поле CHARMM параметри, включително заряди, връзки, ъгли и ъгли на усукване. Инструментите са достъпни чрез графичен интерфейс, който съществено опростява подготовката и анализа на съответните симулации. Основното предимство на FFToolkit е, че в него е интегрирана цялата параметризационна процедура на CHARMM.

Типове атоми и Ван дер Ваалсови параметри по аналогия

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-ацетилиран 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 (фиг. 41). Новите типове атоми, аналогичните им атоми от CGenFF и съответните Ван дер Ваалсови параметри са представени в табл. 5.

(а) (б)

Фигура 41: Съединения, използвани за дефиниране на новите типове атоми и назначаване на Ван дер Ваалсови параметри по аналогия. (а) `RESI TPHC`, използван за дефиниране на новите типове атоми за азота в сулфаматната анионна група `NSO3` и полярния водород, закачен за него, `HNSO3`; (б) `RESI MSNA`, използван за дефиниране на новите типове атоми за сярата `SNSO3` и трите кислородни атома `ONSO3`, закачени за нея.

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

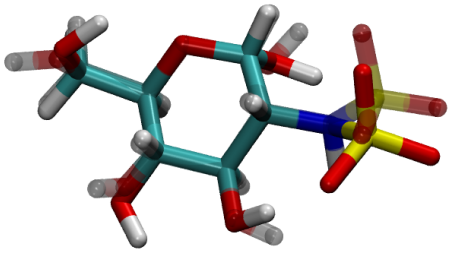
Таблица 5: Типове атоми в сулфаматната анионна група. В първата колона са дадени новите типове атоми, във втората – аналогичните им типове в полето CGenFF, в трета колона е дадена дълбочината на ямата на Ленард-Джоунс потенциала, а в последната колона – половината от разстоянието до минимума на Ленард-Джоунс потенциала.

Оптимизация на геометрията

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 свали от базата данни PDB с идентификационен код 1HPN [101]. 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 [102]. За симулацията беше използван методът на Молер-Плесет от 2-ри ред [103] (MP2) с набор от базисни функции 6-31G* (който описва вътрешните електрони с една слейтърова орбитала, съставена от 6 примитивни гаусови орбитали, а валентните електрони - с 2 слейтърови 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).

Началната и оптимизираните геометрии са насложени една върху друга на фиг. 42.



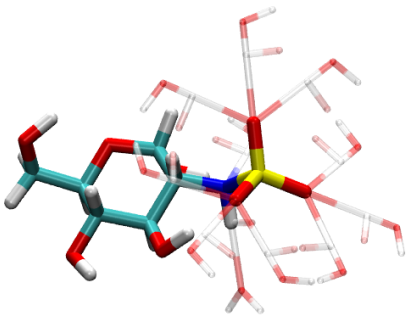
Фигура 42: Начална (полу-прозрачни пръчици) и оптимизирана (плътни пръчици) геометрия на моделното съединение D-Glc2NS.

Оптимизация на зарядите

Силовите полета се различават съществено по начина, по който се определят частичните заряди на атомите. Например при AMBER зарядите се получават от фитиране на електростатичния потенциал около молекулата [87], докато при OPLS зарядите се получават чрез фитиране към експериментални данни за свойства на моделни съединения в кондензирана фаза [86].

В СП CHARMM се набляга на възпроизвеждане на квантово-механични взаимодействия с TIP3P водни молекули [89]. 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. Генерираните комплекси с водни молекули около атомите, които ще се параметризират в новата функционална група са представени на фиг.43. 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 параме-



Фигура 43: Комплекси на моделното съединение и водни молекули за определяне на частичните заряди на атомите от новата функционална група. Водните молекули в различни ориентации около местата на взаимодействие са представени с тънки полу-прозрачни пръчици.

търа – 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*.

FFTK извлича от изходните файлове от GAUSSIAN оптимизираните разстояния между водните молекули и местата на взаимодействие, както и оптималните енергии на взаимодействие, които се пресмятат като разликата между пълната енергия на оптимизирания комплекс и енергиите на двете молекули поотделно. Освен това за фитирането се използва и диполният момент на съединението.

Резултатите от фитиращата процедура за различните типове атоми и места на взаимодействие са представени на фиг. 44. Крайните заряди на новата функционална група са дадени в табл. 6. Пълният заряд на групата е -1e.

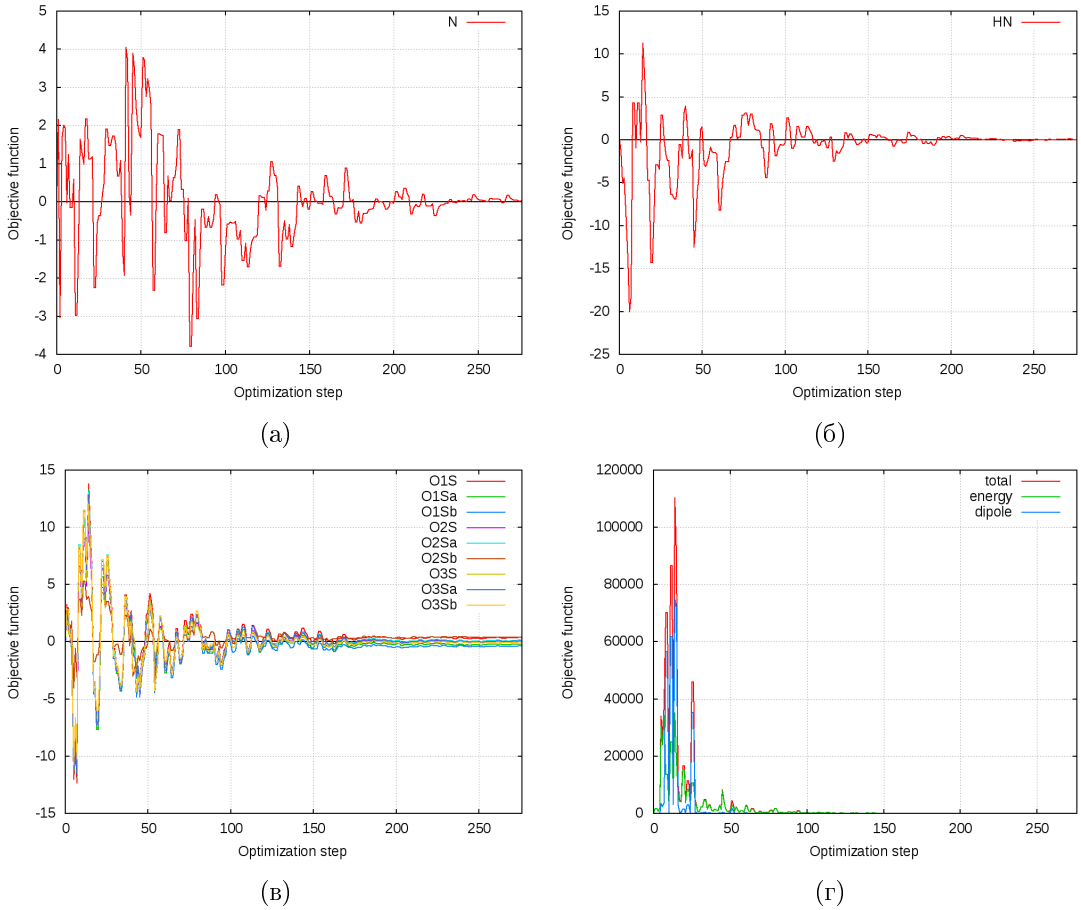
Атом	Частичен заряд
C_2	-0.02
H_2	0.09
N	-0.33
HN	0.26
S	0.47
$O1S$	-0.49
$O2S$	-0.49
$O3S$	-0.49

Таблица 6: Оптимизирани заряди на новата функционална група (части от заряда на електрона).

Оптимизация на химичните връзки и валентните ъгли

Следващата стъпка е да се определят ковалентните параметри, т.е. равновесни стойности и силови константи за всички химични връзки и валентни ъгли, които трябва да се параметризира.

В тази стъпка сравнението между квантово-механични и молекулно-механични вели-



Фигура 44: Сходимост на различни параметри по време на фитиращата процедура. (а) Поведение на целевата функция при вариране на параметрите на взаимодействието на водни молекули с атоми (а) N, (б) HN, (в) $O1S$, $O2S$ и $O3S$. (г) Сходимост на пълната целева функция (в червено), компонентата на енергията (в зелено) и компонентата на диполния момент (в синьо).

чини не е толкова директно, заради различните координатни системи, които се използват за описание на вибрационните спектри и за ковалентните взаимодействия в силовото поле. Вибрационните спектри се описват най-добре чрез вибрации по нормалните моди, а взаимодействието на атомите в СП се описват основно чрез хармонични потенциали по вътрешните координати, дефинирани от връзките, ъглите, двустенните ъгли и неправилните двустенни ъгли. Случва се някои нормални моди да дават принос към динамиката на само една координата в СП, например една химична връзка, но обикновено повечето моди рефлектират върху множество вътрешни координати.

FFTK изчислява 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 пресмятат с помощта на плъгина NAMD Energy [104] 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:

$$\Delta E = \sum_j \frac{1}{2} \frac{\partial^2 E}{\partial q_i \partial q_j} \delta q_i \delta q_j \tag{70}$$

където q_i е прицелната координата, сумата е по всички q_j координати, а производната е съответния елемент в хесиана. В съответствие с 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 [105]. 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. 7.

	Начални параметри		Оптимизирани параметри	
Връзка	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 7: Начални и оптимизирани параметри за химичните връзки и валентните ъгли на новата функционална група.

Оптимизация на ъглите на усукване

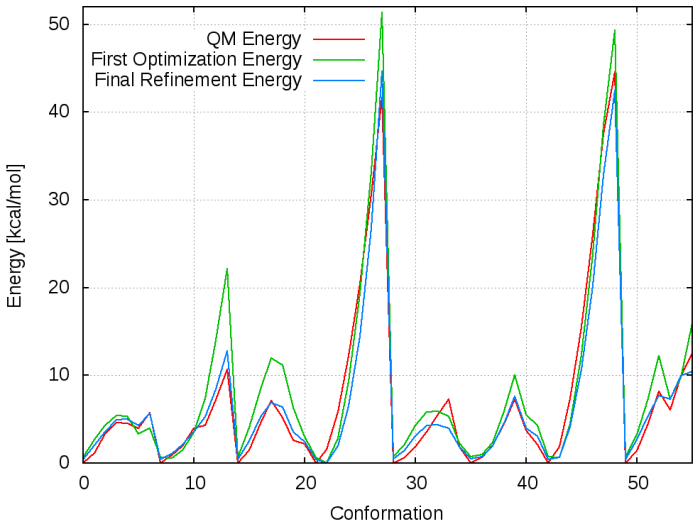
Параметризационната процедура завършва с оптимизация на параметрите на ъглите на усукване, които описват ротациите около химични връзки. Параметрите се получават от КвМ ППЕ, генерирани чрез явно сканиране на двустенните ъгли за параметризация с метода MP2/6-31G*. При сканирането се пресмята енергията 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°. ППЕ от квантово-механични симулации, ППЕ от ММ параметри след първата оптимизация и след последната оптимизация са представени на фиг.45. 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.8.



Фигура 45: Профили на потенциалната енергия от квантово-механични симулации (в червено) и от молекулно-механичните параметри след първата (в зелено) и последната (в синьо) оптимизации.

Ъгъл на усукване				K_χ [kcal/mol]	n	δ [°]
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

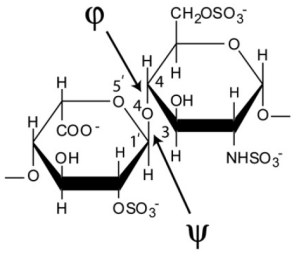
Таблица 8: Оптимизираните параметри на подлежащите на параметризиране ъгли на усукване .

5.1.5 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 производен на хепарина додеказахарид бяха получени от магнитно-резонансната структура с код 1HPN [101] 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 съседи 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. 46.



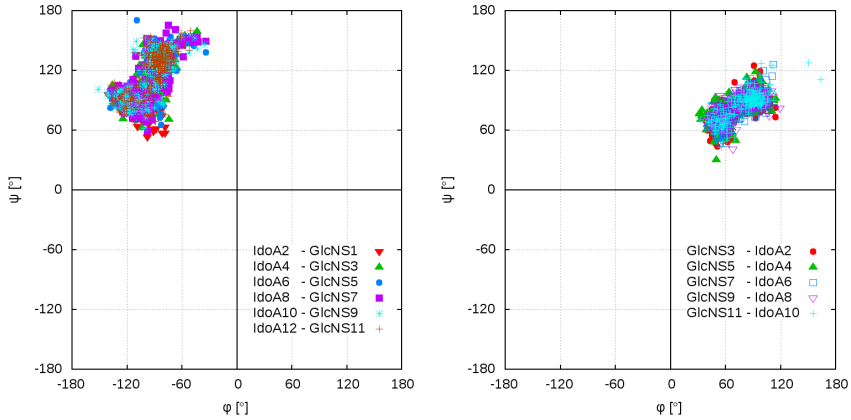
Фигура 46: Дефиниране на ъглите на усукване на гликозидната връзка между два монозахарида. Ъгълът ϕ се дефинира от атоми O_5'-C_1' от първия монозахарид и O_4-C_4 от втория захарид. Ъгъл ψ се дефинира от атоми C_1'-O_4-C_4-C_3. Фигурата е от [106].

Получените от симулацията ϕ, ψ-диаграми са представени на фиг. 47. Те бяха сравнени с разпределението на тези ъгли, получено от Кан и съавтори [106] (фиг. 48). 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. 9.

	$\phi_{IdoA-GlcNS}$ [°]			$\psi_{IdoA-GlcNS}$ [°]			$\phi_{GlcNS-IdoA}$ [°]			$\psi_{GlcNS-IdoA}$ [°]		
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

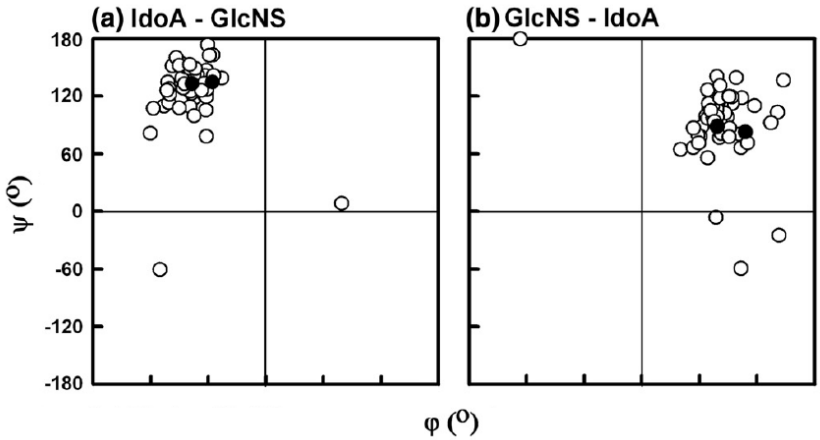
Таблица 9: Средни стойности на ъглите ϕ и ψ за всички връзки IdoA-GlcNS и GlcNS-IdoA: експерименталните стойности и получените от симулация с новите параметри за NSO_3^- групата се съгласуват в рамките на стандартното отклонение.

Освен ефекта върху гликозидните връзки, беше проверено и какво е влиянието на новата група върху конформацията на пръстена на глюкозамина. За целта беше измерен ъгълът на усукване θ, дефиниран от атоми O_5-C_1-C_2-C_3 (фиг. 49a). Той измерва усукването на



(а) IdoA-GlcNS (б) GlcNS-IdoA

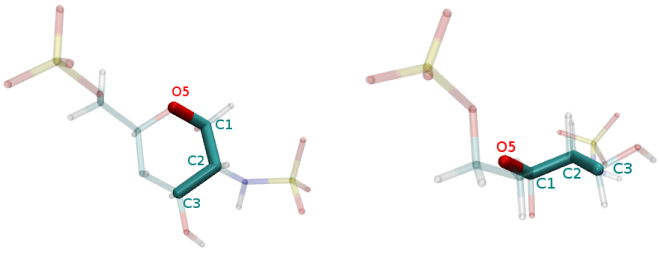
Фигура 47: Разпределение на ъглите ϕ, ψ по време на симулациите за различните гликозидни връзки.



Фигура 48: 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, 1E0O, 1EO3, 1FQ9, 1GMN, 1GMO, 1G5N, 1RID, 1QQP, 1XMN, 2AXM, 2GD4, 2HYU, 2HYV, 3DY0, и 3EVJ). С плътни кръгчета са изобразени ψ и ψ на додеказахарида от структурата с PDB код 1HPN. Фигурата е от [106].

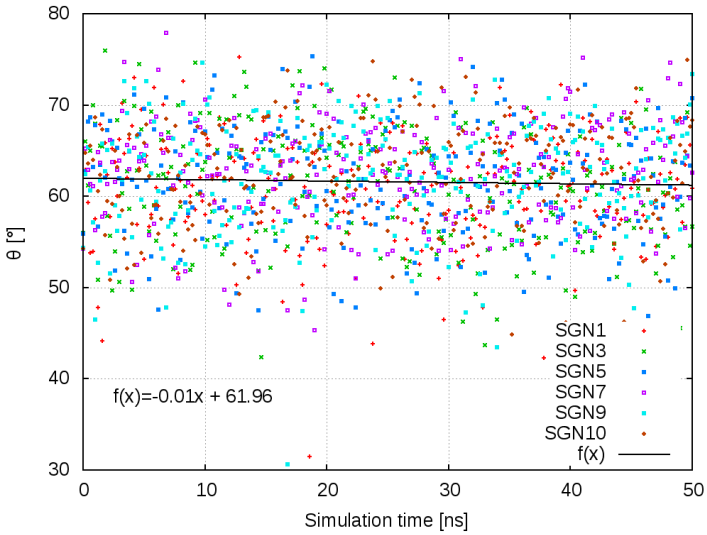
пръстена (фиг. 49б). В експерименталната структура 1HPN този ъгъл има стойност $\theta^{exp} = 55.9°$. Както се вижда от фиг. 50, θ не се изменя съществено по време на симулацията и остава стабилен за всички 6 глюкозамина в додеказахарида. Средната за стойност на ъгъла от симулацията е $\theta^{sim} = 62.2 \pm 5.8°$.

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 ,IFNγRα и производни на хепарина олигозахариди с различна степен на полимеризация.



(а) (б)

Фигура 49: Дефиниция на ъгъла на усукване на пръстена на GlcNS. С полу-прозрачни пръчици е представен целия GlcNS, а с плътни дебели линии – атоми O_5-C_1-C_2-C_3, които дефинират ъгъл θ.



Фигура 50: Изменение на θ за различните GlcNS монозахариди по време на симулацията. Както се вижда, ъгълът остава стабилен.

5.2 Изследване на взаимодействието на IFN-γ и IFNγRα с МД симулации

Основната цел на това изследване беше да се създаде модел на взаимодействието на IFN-γ и двата му извънклетъчни рецептора IFNγRα , който да обясни някои специфики в свързването. Както беше отбелязано в секция 1.1.3, С-краищата на цитокина имат модулират афинитета му към рецепторите. В кристалографската структура на комплекса hIFN-γ–hIFNGγRα (PDB ID 1FG9 [10]) присъстват само координатите на атомите на първите 126 аминокиселини на цитокина. Координатите на 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.

5.2.1 Нагъване на С-краищата на IFN-γ

Експерименталните данни сочат, че С-краищата на hIFN-γ са много подвижни и не притежават конкретна твърда конформация [15]. Затова първоначално липсващите във всеки от мономерите 18 аминокиселини бяха изградени в напълно разгъната конформация с прог-

рамата за манипулация и визуализация на молекули PyMOL [83]. 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 (фиг. 51a).

МД симулациите бяха направени с пакета GROMACS 4.6.7 [76]. За подобряване на обхождането на фазовото пространсво 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 ns (т.е. 450 ns общо симулационно време) и бяха проведени при периодични гранични условия в симулационна кутия с размери 100 $\times$ 120 $\times$ 130 Å^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 14Å. Симулациите бяха направени на клъстера BIOSIM[1], както и на клъстера PHYSON [107] 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 параметърът на сходство ρ [108] за конформацията на гръбнака на C-опашките (а.к. 123-144). Клъстерите бяха формирани по метода за клъстеризиране GROMOS [109] при радиус на обрязване по параметъра *rho* 1. Центроидите на трите най-големи клъстери са представени на фиг. 51.

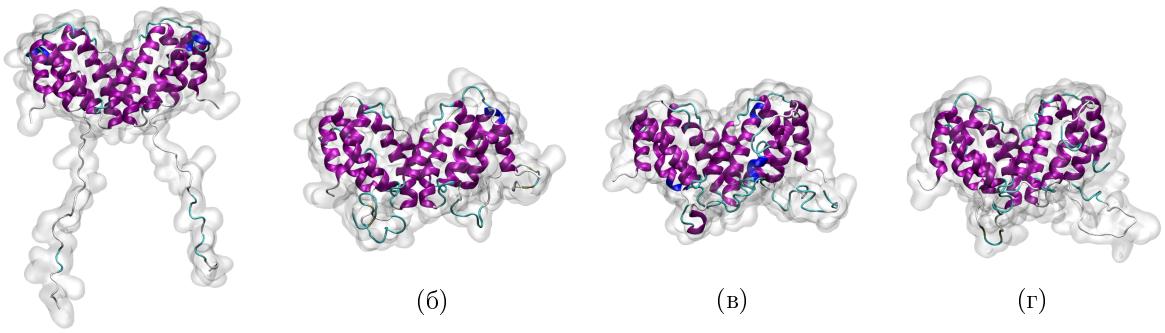
Анализът показва, че C-опашките на hIFN-γ не остават в разгънато състояние. Във всички симулации те веднага започват да се нагъват и след около 15-20 ns вече са доста близо до глобуларната част на белтъка, като така се намалява общата повърхнина на молекулата, достъпна за разтворителя, както и радиусът 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 IFN-γ и IFNγRα

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-–hIFNγRα [10], за да се получи информация за средното разстояние между центровете

[1] Локален Linux клъстер, който разполага с 24 процесора Intel Xeon E55402.53GHz и 8 процесора Intel Xeon E5-2620v22.10GHz с общо 288 нишки и 544 GB RAM, както и с 4 графични ускорителя Tesla K20m, 8 TB дисково пространство.



(а) (б) (в) (г)

Фигура 51: Конформация на С-краищата на hIFN-γ . Във виолетово са изобразени α-спиралите, в тъмно синьо – 3_{10}-спиралите, в бяло и светло синьо – нетсруктурирани бримки. Ван дер Ваалсовата повърхнина на структурите е представена в полупрозрачни светло сиви мехурчета. (а) Начална разгъната конформация на изградените опашки; (б) Центроидът на най-големия клъстер от симулациите по нагъване; Центроидите на втория (в) и третия (г) по големина клъстери.

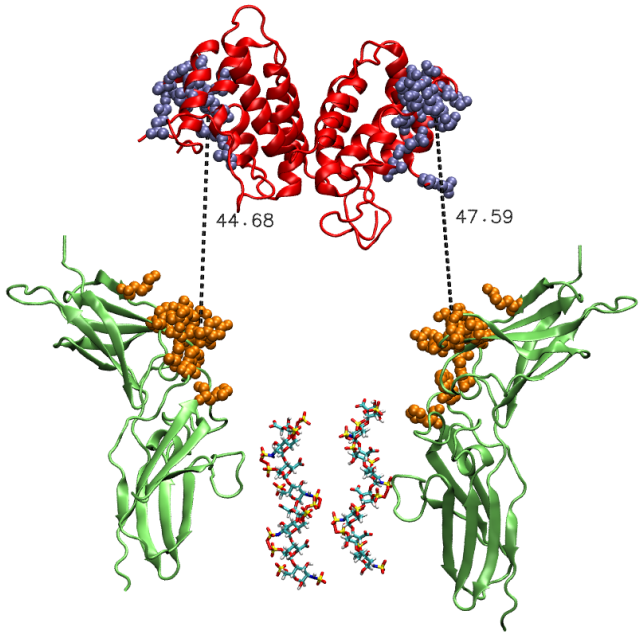
на свързване в цитокина и двата рецептора.

Конформацията на напълно изградената и нагъната молекула на IFN-γ получена от центроида на най-посещавания клъстер по време на симулациите по нагъване, отдалечена на 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–hIFNγRα. Началната конформация на системата цитокин–рецептори–олигозахариди е представена на фиг. 52).

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 Å^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 Å. За Ван дер Ваалсовите взаимодействия беше използвана изглаждаща функция, която започваше да действа на разстояние 10 Å с радиус на обрязване 12 Å. Симулациите бяха на направени на локалния клъстер BIOSIM, както и на клъстера PHYSON [107] към Физически факултет на Софийски университет “Св. Климент Охридски”. Тъй като всички симулирани системи бяха доста големи, симулациите са с продължителност от само 50 ns.

Разстоянията $d1$ и $d2$ 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. 53. Средното разстояние между центровете на свързване в молекулата на интерферона и рецепторите в комплекса от референтната симулация е 6.5 $\pm$ 0.4 Å.

В симулацията само с IFN-γ и IFNγRα силно положително заредените опашки на цитокина активно си взаимодестват с отрицателно заредени домейни в молекулите на всеки от рецепторите. Това взаимодействие пречи на глобуларната част на IFN-γ , където се намират рецептор-свързващите места, да се доближи достатъчно до рецепторите. Така опашките на



Фигура 52: Начална конфигурация за симулациите по свързване на hIFN-γ и hIFNγRα . Цитокинът е представен с червени ленти, рецепторите – със зелени ленти, а олигозахарридите – с цветни пръчици. Със сини и оранжеви сфери съответно са представени аминокиселинните остатъци формиращи контактния интерфейс в IFN-γ и IFNγRα. С черни пунктирани линии са представени разстоянията $d1 = 47.6$Å и $d2 = 44.7$Å между центровете на свързване от двете страни на молекулата на цитокина и тези в двата рецептора.

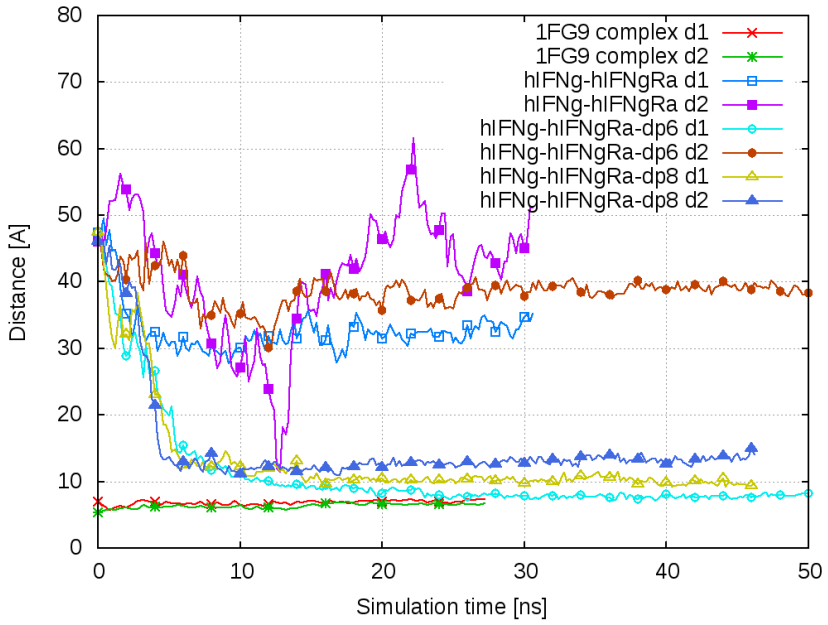
цитокина всъщност пречат на правилното формиране на комплекс. Това поведение беше наблюдавано и в поредица други симулации с различни начални разстояния и ортиентации между трите белтъка, както и с различни конформации на C-краищата на IFN-γ . Въпреки това C-краищата са важни за взаимодействието с извънклетъчните домейни на рецепторите. Резултати от предварителни симулации, в които молекулата на цитокина с липсващи по 18 ак. от опашките (hIFN-γ от кристалографската структура) е отдалечен по аналогичен начин от рецепторите, показват, че цитокинът изобщо не се 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 C-краищата на IFN-γ – $+18e$. 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 накара 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. Както стана ясно в секция 1.4, ГАГ 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-γ с рецепторите му.

Експерименталните данни сочат, че цитокинът взаимодейства най-силно с хекса- и окта-захариди (секция 1.4.4). По тази причина бяха направени две симулации с тези два олигозахарида. Както се вижда от фиг. 53, 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 C-краищата на IFN-γ , 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 dp6 се оказва, че хексазахаридите не успяват да привлекат достатъчно силно и C-краищата и на двата мономера на IFN-γ . 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.



Фигура 53: Разстояния $d1$ и $d2$ между центровете на масите на центровете на свързване в молекулата на hIFN-γ и в рецепторите IFNγRα .

Тези резултати дават качествено потвърждение на хипотезата, че формирането на комплекс между IFN-γ и IFNγRα е бавен и несигурен процес, който изисква участието на допълнителни молекули. Както се вижда от фиг. 53 производните на хепарина олигозахариди подпомагат приближаването на цитокиновата молекула към рецепторите заради силното електростатично привличане между отрицателно заредените олигозахариди и положително заредените C-краища на IFN-γ . Освен това беше установено, че октазахаридите оказват по-силно благорпиятно 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-γ и IFNγRα са необходими допълнителни симулации, с по-голяма продължителност, набор от няколко различни начални конфигурации на участниците и вероятно използване на методи за ускорено обхождане на фазовото пространсво като, например, метадинамика.

Заключение

Интерферон-γ (IFN-γ) е плейотропен цитокин, който играе основна роля в регулацията на защитата на организма и имунопатологичните процеси. Анормалната експресия на цитокина е свързани с патогенната му роля в автоимунни и автовъзпалителни заболявания, включително възпалителни заболявания на червата, множествена склероза, захарен диабет и оказват негативно влияние върху някои видове тумори. Това налага търсенето на начини за контролиране на биологичната активност на IFN-γ . Целта на дисертацията е да се изгради модел на цитокина с метода на молекулната динамика, който да обясни взаимодействието му с рецепторите и с други лиганди и да подпомогне разработването на стратегии за инхибиране активността на белтъка.

За постигането на тази цел беше разработен структурен модел на цялостната молекула на цитокина. Последващите симулации на изцяло изградена и нагъната молекула на hIFN-γ и двата му извънклетъчни рецептора hIFNγRα показаха, че електростатичното привличане между цитокина и двата рецептора не е достатъчен фактор за бързо и правилно формиране на комплекс. Това се дължи на активното взаимодействие между силно положително заредените опашки на цитокина и отрицателно заредени домейни в молекулите на всеки от рецепторите, което пречи на глобуларната част на IFN-γ , където се намират рецептор-свързващите места, да се доближи достатъчно до рецепторите. Това наблюдение ни накара да издигнем хипотезата, че бързото и правилно формиране на цитокин - рецепторен комплекс зависи от други помощни молекули – протеогликаните, съдържащи хепарин и хепаран сулфат. За проверката на тази хипотеза беше разработено силово поле за сулфаматната анионна група NSO_3^-, присъстваща в монозахариди, изграждащи гликозаминогликаните хепарин и NS-хепаран сулфат. Получените с новите параметри и топология свойства на захаридите се съгласуват добре с оскъдните експериментални данни. Те позволяват да се изследва взаимодействието на тези захариди с множество протеини, включително hIFN-γ. Направените симулации на макромолекулната система hIFN-γ – hIFNγRα – производни на хепарина олигозахариди потвърждават качествено предложената хипотеза. Олигозахаридите привличат по-силно от рецепторите C-краищата на цитокина и по този начин придърпват цялата молекула на hIFN-γ между рецепторите, позволявайки на свързващите центрове да се приближат значително и да си взаимодействат. Получените в последната глава резултати потвърждават качествено издигнатата от нас хипотеза и едновременно с това обосновават необходимостта от по-нататъшни изследвания за количествено характеризиране на влиянието на производните на хепарина олигозахариди върху взаимодействието на IFN-γ и IFNγRα .s

Контролирането на активността на IFN-γ може да се постигне чрез блокиране или на самия цитокин, или на неговия рецептор. Последното 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. Следвайки тази стратегия, беше направено мащабно компютърно изследване на 100 мутагенни форми на hIFN-γ с цел да се идентифицират тези молекули, в които локалната вторична структура на мястото на мутациите да се

запазва. Създаден беше метадинамичен модел за изследване на стабилността на локалната структура на спирала E, съдържаща мутираните три аминокиселини $Lys^{86} - Lys - Lys^{88}$. Бяха разработени критерии и протокол за оценка на стабилността на локалната α-спирална структура чрез анализ на повърхнината на свободната енергия в пространството на ъглите на усукване на гръбнака на аминокиселинния остатък на позиция 86. Въз основа на въведените количествени критерии за оценка а стабилността на структурата са селектирани 12 от 100 мутагенни форми на човешки интерферон гама, които запазват локалната структура на спирала E и имат близка или по-голяма стабилност от нативната форма на протеина.

Симулациите на биологични макромолекулни системи включват огромен брой атоми и изискват сериозни изчислителни ресурси като брой процесори и процесорно време. Основният проблем в МД симулациите на големи системи е пресмятането на електростатичните взаимодействия, защото те са далекодействащи и включват огромен брой двучастични взаимодействия. Комплексът hIFN-γ–hIFNγRα, разтворен във вода, съдържа над 240 000 атома. В процеса на изследване на свойствата на IFN-γ и моделирането на взаимодействието му с неговите рецептори възникнаха множество методологични въпроси за избора на различни параметри, които повлияват производителността на симулациите и съответно ефективното оползотворяване на изчислителните ресурси. С цел оптимизиране на симулационните протоколи беше проведено систематично изследване на зависимостта на двата най-популярни пакета за молекулно-динамични симулации GROMACS и NAMD от различни параметри на интегратора и алгоритъмът за пресмятане на далечните електростатични взаимодействия. Беше установено, че използването на интегратори с променлива времева стъпка в пакета NAMD може значително да увеличи производителността, но за сметка на стабилността на симулацията. Беше показано, че оптимални резултати се постигат при двукратно намаляване на честотата на пресмятане на далечните взаимодействия (на 4 вместо на 2 стъпки), при което стабилността на симулацията се запазва, а производителността се увеличава с над 10%. В пакета GROMACS намаляването на отношението на изчислителните ярда, пресмятащи само далечните електростатични взаимодействия и ядрата, пресмятащи всичко останало води до повишаване на абсолютната производителност, макар и да влошава скалируемостта. По подразбиране това отношение е 1/4. Намаляването му наполовина увеличава производителността с между 8.9% (за 4096 изчислителни ядра) и 25% (за 512 изчислителни ядра). Подреждането на различните типове изчислителни ядра също има значение, особено при голям общ брой ядра. При 4096 ядра промяната на режима на подреждане по подразбиране (`interleave`) на `cartesian` вдига производителността на симулацията с над 12%. Тези резултати демонстрират необходимостта от оптимизиране на методологията на симулация.

В заключение ще изложим накратко плановете за използването на получените в дисертацията резултати за бъдещи изследвания. Те ще се фокусират върху разработването на модел за количествена оценка на влиянието на олигозахариди с различна степен на полимеризация върху формирането на комплекса hIFN-γ –hIFNγRα . По детайлното изследване на този процес изисква симулации с по-голяма продължителност, набор от няколко различни начални конфигурации на участниците и вероятно използване на методи за ускорено обхождане на фазовото пространство като. Вече работим по модел на основата на метадинамика с обмяна на метадинамичния потенциал и/или използване на специални колективни променливи за описване на пътя между две състояния в профила на свободната енергия. Чрез него се надяваме да успеем количествено да характеризираме разликите в свободната енергия за формирането на комплекс със и без участието на олигозахариди.

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 големи биомолекулни системи на голям брой изчислителни ядра;
- Изследвани са различни възможности за оптимизиране на далеко-действащите 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 влошеа 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, hIFNI-γ and its extracellular receptors, submitted to *AIP Conference Proceedings*.

Включените в дисертацията резултати са представени на следните международни научни 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 Confe-rence "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., *Internatio-nal 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 hIFNI-γ–hIFNγR–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, hIFNI-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. дбн 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"", финансиран по Оперативна програма "Развитие на човешките ресурси", съфинансирана от Европейския социален фонд на Европейския съюз.

Приложение А

Топология на монозахаридите, изграждащи хепарин и NS-областите на хепаран сулфат в CHARMM формат

heparin.top

```
*>>>>>> CHARMM topology file <<<<<<
36 1

! Hydrogens
MASS   185 HCA1     1.00800  H ! aliphatic proton, CH
MASS   186 HCA2     1.00800  H ! aliphatic proton, CH2
MASS   188 HCP1     1.00800  H ! polar H
MASS   400 HNSO3    1.00800  h ! polar H of from the NSO3- group of SGN
! Carbons
MASS   174 CC3161  12.01100  C ! C2, C3, C4 CH bound to OH
MASS   175 CC3162  12.01100  C ! C1 (anomeric) CH bound to OH
MASS   176 CC3163  12.01100  C ! C5 CH bound to exocylic CH2OH
MASS   177 CC321   12.01100  C ! generic acyclic CH2 carbon (hexopyranose C6)
MASS   182 CC2O2   12.01100  C ! sp2 carbon in carboxylates
! Nitrogen
MASS   401 NNSO3   14.00700  N ! N in the NSO3- group of SGN
! Oxygens
MASS   191 OC311   15.99940  O ! hydroxyl oxygen
MASS   192 OC3C61  15.99940  O ! ether in six membered ring
MASS   193 OC301   15.99940  O ! generic linear ether
MASS   196 OC2D2   15.99940  O ! sp2 oxygen in carboxylates
MASS   222 OC3OP   15.99940  O ! ester O in PO3H (phosphate) || OSL in top_all27_lipid.rtf
MASS   223 OC2DP   15.99940  O ! =O in PO3H (phosphate) || O2L in top_all27_lipid.rtf
MASS   402 ONSO3   15.99940  O ! O in the NSO3-        group of SGN
! Sulfers
MASS   225 SC      32.060000 S ! Sulfate sulfur
MASS   403 SNSO3   32.060000 S ! Ssulfur in the NSO3- group of SGN

AUTOGENERATE angles dihedrals
! DEFAults for patching FIRSt and LAST residues
DEFA FIRS NONE LAST NONE

RESI IDS           -2.000  !     2-O-sulfo-alpha-L-idopyranuronic acid
                           !
GROU                       !
```

```
ATOM C1   CC3162    0.340  !                  H5
ATOM H1   HCA1      0.090  !                  |
ATOM O1   OC311    -0.650  !                  C5---O5
ATOM HO1  HCP1      0.420  !            H4   /|      \    O1-HO1
ATOM C5   CC3163    0.110  !              \ / C6 HO3  \  /
ATOM H5   HCA1      0.090  !               C4   /      C1
ATOM O5   OC3C61   -0.400  !              / \ O3   H2 /  \
GROU             ! -1.     !        HO4-O4   \|    | /    H1
ATOM C2   CC3161   -0.190  !                  C3---C2
ATOM H2   HCA1      0.090  !                  |    |
ATOM O2   OC30P    -0.280  !                  H3   O2
ATOM S2   SC        1.330  !                       |
ATOM OS22 OC2DP    -0.650  !                OS22---S2---OS24
ATOM OS23 OC2DP    -0.650  !                       |
ATOM OS24 OC2DP    -0.650  !                     OS23(-)
GROU                       !
ATOM C3   CC3161    0.140  !        n.b.: O61 and O62 are attached to C6
ATOM H3   HCA1      0.090  !               |
ATOM O3   OC311    -0.650  !          O62=C6
ATOM HO3  HCP1      0.420  !               |
GROU                       !          O61(-)
ATOM C4   CC3161    0.140
ATOM H4   HCA1      0.090
ATOM O4   OC311    -0.650
ATOM HO4  HCP1      0.420
GROU             ! -1.
ATOM C6   CC2O2     0.520  !
ATOM O61  OC2D2    -0.760  !
ATOM O62  OC2D2    -0.760  !
BOND C1   O1        C1   H1        O1   HO1       C1   O5        C1   C2
BOND C2   H2        C2   O2        C2   C3        C3   H3        C3   O3
BOND O3   HO3       C3   C4        C4   H4        C4   O4        O4   HO4
BOND C4   C5        C5   H5        C5   C6        C6   O61       C6   O62
BOND C5   O5        O2   S2        OS22 S2        OS23 S2        OS24 S2
!    I    J    K    L      R(IK)   T(IKJ)    PHI   T(JKL)   R(KL)
IC   O1   C2  *C1   H1     1.4115  105.82  120.26  110.81   1.0905
IC   O1   O5  *C1   C2     1.4115  113.00  118.92  111.32   1.5218
IC   O2   C3  *C2   H2     1.4190  108.02 -120.00  108.77   1.0892
IC   O2   C1  *C2   C3     1.4190  111.74 -120.00  110.04   1.5167
IC   O3   C4  *C3   H3     1.4198  108.29  120.00  108.28   1.0957
IC   O3   C2  *C3   C4     1.4198  111.19  120.00  110.08   1.5102
IC   O4   C5  *C4   H4     1.4163  108.82 -120.00  109.07   1.0972
IC   O4   C3  *C4   C5     1.4163  111.43 -120.00  108.83   1.5171
IC   C6   O5  *C5   H5     1.5099  105.42 -117.88  109.66   1.0926
IC   C6   C4  *C5   O5     1.5099  112.84 -117.26  109.77   1.4384
IC   O61  C5  *C6  O62     1.4163  111.44  180.00  108.72   1.0943
IC   O5   C1   C2   C3     1.4059  111.32  -53.56  110.04   1.5167
IC   C1   C2   C3   C4     1.5218  110.04   54.55  110.08   1.5102
IC   C2   C3   C4   C5     1.5167  110.08  -57.15  108.83   1.5171
IC   C3   C4   C5   O5     1.5102  108.83   58.25  109.77   1.4384
IC   C4   C5   O5   C1     1.5171  109.77  -59.84  114.27   1.4059
IC   C5   O5   C1   C2     1.4384  114.27   57.30  111.32   1.5218
IC   C4   C5   C6  O61     1.5171  112.84  120.00  111.44   1.4163
IC   O5   C1   O1  HO1     1.4059  113.00  -60.00  108.04   0.9634
IC   C2   C3   O3  HO3     1.5167  111.19   60.00  106.26   0.9641
IC   C3   C4   O4  HO4     1.5102  111.43  -60.00  105.89   0.9645
IC   O5   C1   C2   O2     1.4305  111.50 -170.93  108.22   1.4352
IC   C1   O2  *C2   H2     1.5613  108.22 -116.76  114.09   1.1193
IC   C1   C2   O2   S2     1.5613  108.22  131.56  117.57   1.5527
IC   C2   O2   S2  OS22    1.4352  117.57  177.25  104.11   1.4504
IC   O2  OS22 *S2  OS23    1.5527  104.11  112.40  113.72   1.4529
```

```
IC   O2  OS22 *S2  OS24    1.5527  104.11 -114.18  114.03   1.4500
PATC  FIRS NONE LAST NONE

RESI SGN           -2.000  ! N,O6-disulfo-glucosamine
                           !
!Charges from sgn_1.psf
GROU                       !
ATOM C1   CC3162    0.340  !                OS63(-)
ATOM H1   HCA1      0.090  !                  |
ATOM O1   OC311    -0.650  !           OS62---S6---OS64
ATOM HO1  HCP1      0.420  !                  |
ATOM C5   CC3163    0.110  !                  O6
ATOM H5   HCA1      0.090  !                  |
ATOM O5   OC3C61   -0.400  !              H61-C6-H62
GROU             ! -1.     !                  |
ATOM C2   CC3161   -0.020  !               H5-C5---O5
ATOM H2   HCA1      0.090  !            H4   /       \    H1
ATOM N    NNSO3    -0.330  !              \ / HO3     \  /
ATOM HN   HNSO3     0.260  !               C4 |        C1
ATOM S2   SNSO3     0.470  !              / \ O3   H2 /  \
ATOM OS22 ONSO3    -0.490  !        HO4-O4   \|    | /    O1-HO1
ATOM OS23 ONSO3    -0.490  !                  C3---C2
ATOM OS24 ONSO3    -0.490  !                  |    |
GROU                       !                  H3   N-HN
ATOM C3   CC3161    0.140  !                       |
ATOM H3   HCA1      0.090  !                OS22---S2---OS24
ATOM O3   OC311    -0.650  !                       |
ATOM HO3  HCP1      0.420  !                       OS23(-)
GROU                       !
ATOM C4   CC3161    0.140  !
ATOM H4   HCA1      0.090
ATOM O4   OC311    -0.650
ATOM HO4  HCP1      0.420
GROU             ! -1.
ATOM C6   CC321    -0.28
ATOM H61  HCA2      0.09
ATOM H62  HCA2      0.09
ATOM O6   OC3OP    -0.28
ATOM S6   SC        1.33
ATOM OS62 OC2DP    -0.65
ATOM OS63 OC2DP    -0.65
ATOM OS64 OC2DP    -0.65
!
BOND C1   O1        C1   H1        O1   HO1       C1   O5        C1   C2
BOND C2   H2        C2   N         N    HN        C2   C3        C3   H3
BOND C3   O3        O3   HO3       C3   C4        C4   H4        C4   O4
BOND O4   HO4       C4   C5        C5   H5        C5   C6        C6   H61
BOND C6   H62       C6   O6        C5   O5        N    S2        S2   OS22
BOND S2   OS23      S2   OS24      O6   S6        OS62 S6        OS63 S6
BOND OS64 S6
!    I    J    K    L      R(IK)   T(IKJ)    PHI   T(JKL)   R(KL)
IC   O1   C2  *C1   H1     1.4115  105.82 -120.26  110.81   1.0905
IC   O1   O5  *C1   C2     1.4115  113.00 -118.92  111.32   1.5218
IC   N    C3  *C2   H2     1.4190  110.99 -116.97  108.77   1.0892
IC   N    C1  *C2   C3     1.4190  111.74 -123.77  110.04   1.5167
IC   O3   C4  *C3   H3     1.4198  107.51  119.38  108.28   1.0957
IC   O3   C2  *C3   C4     1.4198  111.19  119.03  110.08   1.5102
IC   O4   C5  *C4   H4     1.4163  107.94 -120.61  109.07   1.0972
IC   O4   C3  *C4   C5     1.4163  111.43 -118.91  108.83   1.5171
IC   C6   O5  *C5   H5     1.5099  105.42  117.88  109.66   1.0926
IC   C6   C4  *C5   O5     1.5099  112.84  117.26  109.77   1.4384
```

```
IC   O6  H62  *C6  H61     1.4163  111.41 -118.10  108.62   1.0873
IC   O6   C5  *C6  H62     1.4163  111.44 -123.18  108.72   1.0943
IC   O5   C1   C2   C3     1.4059  111.32   53.56  110.04   1.5167
IC   C1   C2   C3   C4     1.5218  110.04  -54.55  110.08   1.5102
IC   C2   C3   C4   C5     1.5167  110.08   57.15  108.83   1.5171
IC   C3   C4   C5   O5     1.5102  108.83  -58.25  109.77   1.4384
IC   C4   C5   O5   C1     1.5171  109.77   59.83  114.27   1.4059
IC   C5   O5   C1   C2     1.4384  114.27  -57.30  111.32   1.5218
IC   C4   C5   C6   O6     1.5171  112.84 -179.74  111.44   1.4163
IC   O5   C1   O1  HO1     1.4059  113.00   67.18  108.04   0.9634
IC   C1   C2   N   HN      1.5218  111.74  -44.11  105.37   0.9665
IC   C2   C3   O3  HO3     1.5167  111.19  -51.80  106.26   0.9641
IC   C3   C4   O4  HO4     1.5102  111.43   47.58  105.89   0.9645
!
IC   S2   N    C2   H2     1.749   111.44   11.33  109.77   1.097
IC   S2   N    C2   C3     1.749   111.44 -109.19  111.39   1.517
IC   S2   N    C2   C1     1.749   111.44  128.97  108.23   1.531
IC  OS23  S2   N    HN     1.502   103.43  -65.92  104.39   1.025
IC  OS24  S2   N    HN     1.472   102.44   53.02  104.39   1.025
IC  OS22  S2   N    HN     1.474   103.22  175.70  104.39   1.025
IC  OS23  S2   N    C2     1.502   103.43   52.40  111.44   1.472
IC  OS24  S2   N    C2     1.472   102.44  171.35  111.44   1.472
IC  OS22  S2   N    C2     1.474   103.22  -65.97  111.44   1.472
!
IC   O5   C5   C6   O6     1.4383  108.81   76.01  109.30   1.4275
IC   C5   O6  *C6  H61     1.5222  109.30  120.33  111.18   1.1156
IC   C5   O6  *C6  H62     1.5222  109.30 -118.94  111.25   1.1168
IC   C5   C6   O6  S6      1.5222  109.30  169.79  115.83   1.5647
IC   C6   O6   S6  OS62    1.4275  115.83  177.39  103.52   1.4539
IC   O6 OS62  *S6  OS63    1.5647  103.52  114.97  113.22   1.4527
IC   O6 OS62  *S6  OS64    1.5647  103.52 -113.66  112.94   1.4539
PATC FIRS NONE LAST NONE

! axial-axial 1->4 linkage
! use when both units are galactose
! CITSIH10
PRES 14aa           0.02 ! (i)1->4(i-1) axial at C1 and axial at C4
dele atom 1HO4
dele atom 2HO1
dele atom 2O1
ATOM 1C4  CC3161    0.09 !
ATOM 1O4  OC301    -0.36 !
ATOM 2C1  CC3162    0.29 !
BOND 1O4  2C1
!    I    J    K    L      R(IK)   T(IKJ)    PHI   T(JKL)   R(KL)
IC   1C3  1C4  1O4  2C1    1.5075  110.40  -86.29  119.44   1.4036  ! psi
IC   1C4  1O4  2C1  2O5    1.4716  119.44  133.57  109.48   1.4490  ! phi
IC   2O5  1O4 *2C1  2C2    1.4490  109.48  124.57  110.72   1.5065
IC   2O5  1O4 *2C1  2H1    1.4490  109.48 -116.61  107.55   1.1057


! axial-equatorial 1->4 linkage
! use for MALTOS11
PRES 14ab           0.02 ! (i)1->4(i-1) axial at C1 and equatorial at C4
dele atom 1HO4
dele atom 2HO1
dele atom 2O1
ATOM 1C4  CC3161    0.09 !
ATOM 1O4  OC301    -0.36 !
ATOM 2C1  CC3162    0.29 !
BOND 1O4  2C1
```

```
!    I    J    K    L      R(IK)   T(IKJ)    PHI   T(JKL)   R(KL)
IC   1C3  1C4  1O4  2C1    1.5010  117.73   72.71  118.66   1.4036  ! psi
IC   1C4  1O4  2C1  2O5    1.4919  118.66   48.64  109.48   1.4570  ! phi
IC   2O5  1O4 *2C1  2C2    1.4570  109.48  124.57  110.72   1.5065
IC   2O5  1O4 *2C1  2H1    1.4570  109.48 -116.61  107.55   1.1057
```

Приложение Б

Параметри за монозахаридите, изграждащи хепарин и NS-областите на хепаран сулфат в CHARMM формат

heparin.par

```
*>>>>>> CHARMM topology file <<<<<<

! Hydrogens
!MASS   185 HCA1     1.00800  H ! aliphatic proton, CH
!MASS   186 HCA2     1.00800  H ! aliphatic proton, CH2
!MASS   188 HCP1     1.00800  H ! polar H
!MASS   400 HNSO3    1.00800  h ! polar H of from the NSO3- group of SGN
! Carbons
!MASS   174 CC3161  12.01100  C ! C2, C3, C4 CH bound to OH
!MASS   175 CC3162  12.01100  C ! C1 (anomeric) CH bound to OH
!MASS   176 CC3163  12.01100  C ! C5 CH bound to exocylic CH2OH
!MASS   177 CC321   12.01100  C ! generic acyclic CH2 carbon (hexopyranose C6)
!MASS   182 CC2O2   12.01100  C ! sp2 carbon in carboxylates
! Nitrogen
!MASS   401 NNSO3   14.00700  N ! N in the NSO3- group of SGN
! Oxygens
!MASS   191 OC311   15.99940  O ! hydroxyl oxygen
!MASS   192 OC3C61  15.99940  O ! ether in six membered ring
!MASS   193 OC301   15.99940  O ! generic linear ether
!MASS   196 OC2D2   15.99940  O ! sp2 oxygen in carboxylates
!MASS   222 OC30P   15.99940  O ! ester O in PO3H (phosphate) || OSL in top_all27_lipid.rtf
!MASS   223 OC2DP   15.99940  O ! =O in PO3H (phosphate) || O2L in top_all27_lipid.rtf
!MASS   402 ONSO3   15.99940  O ! O in the NSO3- group of SGN
! Sulfers
!MASS   225 SC      32.060000 S ! Sulfate sulfur
!MASS   403 SNSO3   32.060000 S ! Ssulfur in the NSO3- group of SGN

!***************** BONDS *********************
BONDS
!V(bond) = Kb(b - b0)**2
!Kb: kcal/mole/A**2
!b0: A
!
! atom types     Kb      b0
!
```

```
! C-H bonds
CC3161  HCA1     309.00  1.111   ! par22 HA  CT1
CC3162  HCA1     309.00  1.111   ! par22 HA  CT1
CC3163  HCA1     309.00  1.111   ! par22 HA  CT1
CC321   HCA2     309.00  1.111   ! par22 HA  CT2
!
! C-C bonds
CC3161  CC3161   222.50  1.480   ! sng qm and crystal analysis
CC3161  CC3162   222.50  1.480   ! sng qm and crystal analysis
CC3161  CC3163   222.50  1.500   ! par22 CT1 CT1
CC3163  CC202    200.00  1.480   ! og xtal
CC321   CC3163   222.50  1.490   ! sng qm and crystal analysis
CC321   CC321    222.50  1.530   ! par22 CT2 CT2
CC321   CC202    200.00  1.522   ! par22 CT2 CC
!
! C-O bonds
CC3161  OC311    410.00  1.410   ! og crystal analysis
CC3161  OC30P    360.00  1.425   !
CC3161  OC301    360.00  1.415   ! og/rmv disac crystals
CC3162  OC311    428.00  1.400   ! og crystal analysis
CC3162  OC3C61   360.00  1.425   ! og crystal analysis
CC3162  OC301    360.00  1.395   ! og disac model compounds 2 3 og/rmv crystals
CC3162  OC30P    360.00  1.425   !
CC3163  OC3C61   360.00  1.425   ! sng qm and crystal analysis
CC321   OC311    428.00  1.420   ! par22 OH1 CT2
CC321   OC30P    360.00  1.452   !
CC321   OC301    360.00  1.415   ! par35 CC32A OC30A
CC202   OC2D2    525.00  1.260   ! par22 OC  CC
!
! O-H bonds
OC311   HCP1     545.00  0.960   ! par22 OH1 H
!
! S-O bonds
SC      OC30P    250.0   1.610   ! methylsulfate
SC      OC2DP    540.0   1.448   ! methylsulfate
SNSO3   ONSO3    540.0   1.448   ! from carb i cgen
!
! N-H bonds
NSO3    HNSO3    432.500 1.025   ! NABAKB phosphoramidates
!
! C-N bonds
CC3161  NNSO3    261.000 1.490   ! PARAMCHEM: from CG331 NG2S3, PENALTY= 10
!
! N-S bonds
NSO3    SNSO3    204.000 1.657   !!! TOVA E NASHE, model, new_params

!***************** ANGLES ************
ANGLES
!V(angle) = Ktheta(Theta - Theta0)**2
!V(Urey-Bradley) = Kub(S - S0)**2
!Ktheta: kcal/mole/rad**2
!Theta0: degrees
!Kub: kcal/mole/A**2 (Urey-Bradley)
!S0: A
!
! atom types             Ktheta  Theta0   Kub    S0
!
! C-C-C angles
CC3161  CC3161  CC3161   53.35   111.00   8.00  2.561  ! par22 CT1 CT1 CT1
CC3161  CC3161  CC3163   53.35   111.00   8.00  2.561  ! par22 CT1 CT1 CT1
CC3162  CC3161  CC3161   53.35   111.00   8.00  2.561  ! par22 CT1 CT1 CT1
```

```
CC3161  CC3163  CC321    58.35   113.50  11.16  2.561  ! sng
CC2O2   CC3163  CC3161   52.00   108.00                ! par22 CT1  CT1  CC
CC321   CC321   CC321    58.35   113.60  11.16  2.561  ! par22 CT2 CT2 CT2
!
! C-C-H angles
CC3161  CC3161  HCA1     34.50   110.10  22.53  2.179  ! par22 HA  CT1 CT1
CC3161  CC3162  HCA1     34.50   110.10  22.53  2.179  ! par22 HA  CT1 CT1
CC3162  CC3161  HCA1     34.50   110.10  22.53  2.179  ! par22 HA  CT1 CT1
CC3163  CC3161  HCA1     34.50   110.10  22.53  2.179  ! par22 HA  CT1 CT1
CC3161  CC3163  HCA1     34.50   110.10  22.53  2.179  ! par22 HA  CT1 CT1
CC321   CC3163  HCA1     34.50   110.10  22.53  2.179  ! par22 HA  CT1 CT2
CC3163  CC321   HCA2     33.43   110.10  22.53  2.179  ! par22 HA  CT2 CT1
CC2O2   CC3163  HCA1     50.00   109.50                ! par 22 HB   CT1  CC
CC2O2   CC321   HCA2     33.00   109.50  30.00  2.163  ! par22 HA   CT2  CC
HCA2    CC321   CC321    26.50   110.10  22.53  2.179  ! par22 HA   CT2  CT2
!
! C-C-O angles
CC3161  CC3163  OC3C61   45.00   110.00                ! sng mod (qm and crystal data)
CC3161  CC3162  OC3C61   45.00   106.00                ! og crystal analysis
CC3161  CC3162  OC311    75.70   107.00                ! sng mod (qm and crystal data) III
CC3161  CC3161  OC311    75.70   110.10                ! par22 OH1  CT1  CT1
CC3162  CC3161  OC311    75.70   110.10                ! par22 OH1  CT1  CT1
CC3163  CC3161  OC311    75.70   110.10                ! par22 OH1  CT1  CT1
CC3163  CC321   OC311    75.70   110.10                ! par22 OH1  CT2  CT1
CC3163  CC2O2   OC2D2    40.00   114.00  50.00  2.388  ! og amop amol and xtals
CC321   CC3163  OC3C61   45.00   109.00                ! sng mod (qm and crystal data) II
CC321   CC2O2   OC2D2    40.00   114.00  50.00  2.388  ! og amop amol and xtals
CC2O2   CC3163  OC3C61   45.00   103.00                ! og amop mp2/631gd + xtals
OC301   CC3162  CC3161   45.00   105.00                 ! og/rmv disac crystal analysis
OC301   CC3161  CC3161   45.00   109.00                ! og OMeTHP qm compounds 2 and 3
OC301   CC3161  CC3163   45.00   109.00                ! og OMeTHP qm compounds 2 and 3
OC301   CC3161  CC3162   45.00   109.00                ! og OMeTHP qm compounds 2 and 3
OC301   CC321   CC3163   45.00   106.00                ! og fit compounds 11 and 12
OC30P   CC3161  CC3161   45.00   107.00                ! og OMeTHP qm compounds 2 and 3
OC30P   CC3161  CC3163   45.00   107.00                ! og OMeTHP qm compounds
OC30P   CC3162  CC3161   45.00   109.00
OC30P   CC3161  CC3162   45.00   109.00
OC30P   CC321   CC3163   75.700  107.50                !
OC301   CC321   CC321    45.00   111.50                ! par35 OC30A CC32A  CC32A
!
! C-O-C angles
CC3161  OC301   CC3162   50.00   109.20                ! og QM theta vs phi/psi comp 7 to 10
CC3162  OC301   CC321    95.00   109.70                ! par35 CC33A  OC30A CC32A
CC3163  OC3C61  CC3162   95.00   109.70                ! par35 CC326A OC30A CC326A
!
! C-O-H angles
CC3161  OC311   HCP1     50.00   109.00                ! sng mod (qm and crystal data) II
CC3162  OC311   HCP1     50.00   109.00                ! sng mod (qm and crystal data) II
CC321   OC311   HCP1     50.00   109.00                ! sng mod (qm and crystal data)
!
! H-C-O angles
HCA1    CC3161  OC311    55.00   108.89                ! og 1/06 EtOH NIST IR fit
HCA1    CC3162  OC311    55.00   108.89                ! og 1/06 EtOH NIST IR fit
HCA1    CC3163  OC311    55.00   108.89                ! og 1/06 EtOH NIST IR fit
HCA1    CC3161  OC301    60.00   109.50                ! par35 HCA2 CC32A OC30A
HCA1    CC3162  OC301    60.00   109.50                ! par35 HCA2 CC32A OC30A
HCA1    CC3162  OC3C61   45.00   109.50                ! sng from thp
HCA1    CC3163  OC3C61   45.00   109.50                ! sng from thp
HCA2    CC321   OC311    55.00   108.89                ! og 1/06 EtOH NIST IR fit
OC30P   CC3161  HCA1     60.00   109.50                ! par35 HCA2 CC32A OC30A
OC30P   CC3162  HCA1     60.00   109.50                ! par35 HCA2 CC32A OC30A
```

```
OC30P   CC321   HCA2     60.0    109.5                  !
OC301   CC321   HCA2     60.00   109.50                 ! par34 HCA2 CC32A OC30A
!
! H-C-H
HCA2    CC321   HCA2     35.50   109.00   5.40  1.802  ! par22 HA  CT2 HA
!
! O-C-O angles
OC30P   CC3162  OC3C61   90.00   110.00
OC311   CC3162  OC3C61   45.00   116.50                 ! og crystal analysis
OC301   CC3162  OC3C61   90.00   112.00                 ! og/rmv disac crystal analysis
OC2D2   CC202   OC2D2   100.00   132.00  70.00  2.225  ! og amop amol and xtals
!
! C-O-S angles
CC3161  OC30P   SC       15.0    109.00  27.00  1.90   ! methylsulfate
CC3162  OC30P   SC       15.00   109.00  27.00  1.90   ! sairam CC3161   OC30P   SC     15.0     109.00
                                                                                            27.00   1.90
CC321   OC30P   SC       15.0    110.10  27.00  1.90   ! methylsulfate
!
! O-S-O angles
OC30P   SC      OC2DP    85.0     98.0                  ! methylsulfate
OC2DP   SC      OC2DP   130.0    109.47  35.0   2.45   ! methylsulfate
ONSO3   SNSO3   ONSO3   130.000  109.470 35.00  2.45   ! from cgen i carg, OC2DP   SC      OC2DP     130.0
                                                                  109.47   35.0    2.45   ! methylsulfate
!
! H-C-N angles
HCA1    CC3161  NNSO3    48.000  108.000                ! PARAMCHEM: from NG2S1 CG311 HGA1, PENALTY= 9
!
! C-C-N angles
CC3161  CC3161  NNSO3    70.000  113.500                ! PARAMCHEM: from CG311 CG311 NG2S1, PENALTY= 9
CC3162  CC3161  NNSO3    70.000  113.500                ! PARAMCHEM: from CG311 CG311 NG2S1, PENALTY= 9
!
! C-N-H angles
CC3161  NNSO3   HNSO3    35.000  109.000                ! PARAMCHEM: from CG331 NG2S3 HGP1, PENALTY= 1.5
!
! C-N-S angles
CC3161  NNSO3   SNSO3   126.700  106.200                !!! TOVA E NASHE, model, new_params
!
! N-S-O angles
NSO3    SNSO3   ONSO3   116.100  112.400                !!! TOVA E NASHE, model, new_params
!
! H-N-S angles
HNSO3   NNSO3   SNSO3   112.800  109.000                !!! TOVA E NASHE, model, new_params

!***************** DIHEDRALS **********************
DIHEDRALS
!V(dihedral) = Kchi(1 + cos(n(chi) - delta))
!
! c-c-c-c
CC3161  CC3161  CC3161  CC3162   0.19    3  180.0 ! og/sng cyclohexane
CC3161  CC3161  CC3161  CC3163   0.19    3  180.0 ! og/sng cyclohexane
CC3161  CC3161  CC3163  CC321    0.20    3    0.0 ! par27 X CTL1 CTL1 X
CC3161  CC3161  CC3163  CC202    0.20    3    0.0 ! par27 X CTL1 CTL1 X
CC3161  CC3161  CC3161  CC3161   2.31    3  180.0 ! erh fit inositol RIMP2/cc-pVTZ//MP2/631gd C1-C2-C3-C4
CC3162  CC3163  CC3161  CC3161   0.19    3  180.0 ! og/sng cyclohexane
!
! c-c-c-o
CC3161  CC3161  CC3162  OC3C61   0.31    3  180.0 ! og/sng thp
CC3161  CC3161  CC3163  OC3C61   0.31    3  180.0 ! og/sng thp
CC3161  CC3161  CC3161  OC311    0.20    3    0.0 ! par27 X CTL1 CTL1 X
CC3161  CC3161  CC3162  OC311    0.20    3    0.0 ! par27 X CTL1 CTL1 X
CC3161  CC3163  CC321   OC311    0.20    3    0.0 ! par27 X CTL1 CTL2 X
```

```
CC3162  CC3161  CC3161  OC311    0.20    3    0.0 ! par27 X CTL1 CTL1 X
CC3163  CC3161  CC3161  OC311    0.20    3    0.0 ! par27 X CTL1 CTL1 X
CC321   CC3163  CC3161  OC311    0.20    3    0.0 ! par27 X CTL1 CTL1 X
CC2O2   CC3163  CC3161  OC311    0.20    3    0.0 ! par27 X CTL1 CTL1 X
CC2O2   CC3163  CC3161  OC301    0.20    3    0.0 ! par27 X CTL1 CTL1 X
CC321   CC3163  CC3161  OC301    0.20    3    0.0 ! par27 X CTL1 CTL1 X
CC3161  CC3161  CC3161  OC301    0.20    3    0.0 ! par27 X CTL1 CTL1 X
CC3161  CC3161  CC3162  OC301    0.20    3    0.0 ! par27 X CTL1 CTL1 X
CC3162  CC3161  CC3161  OC301    0.20    3    0.0 ! par27 X CTL1 CTL1 X
CC3163  CC3161  CC3161  OC301    0.20    3    0.0 ! par27 X CTL1 CTL1 X
CC3161  CC3163  CC2O2   OC2D2    0.05    6  180.0 ! par22 X CT1 CC X
CC3161  CC3163  CC321   OC301    0.50    1  180.0 ! dpatel, modified
CC3161  CC3163  CC321   OC301    0.37    2  180.0 ! LS fit
CC3161  CC3163  CC321   OC301    0.15    3    0.0 !
CC3161  CC3163  CC321   OC30P    0.43    3    0.0
CC3161  CC3161  CC3162  OC30P    1.76    1    0.0
CC3161  CC3161  CC3162  OC30P    2.56    2    0.0
CC3161  CC3161  CC3162  OC30P    0.01    3    0.0
CC3161  CC3161  CC3161  OC30P    1.76    1    0.0
CC3161  CC3161  CC3161  OC30P    2.56    2    0.0
CC3161  CC3161  CC3161  OC30P    0.01    3    0.0
CC3162  CC3161  CC3161  OC30P    1.76    1    0.0
CC3162  CC3161  CC3161  OC30P    2.56    2    0.0
CC3162  CC3161  CC3161  OC30P    0.01    3    0.0
CC3163  CC3161  CC3161  OC30P    1.76    1    0.0
CC3163  CC3161  CC3161  OC30P    2.56    2    0.0
CC3163  CC3161  CC3161  OC30P    0.01    3    0.0
CC321   CC3163  CC3161  OC30P    0.20    3    0.0 ! par27 X CTL1 CTL1 X
CC321   CC321   CC321   OC301    0.19    3    0.0 ! par27 X CTL2 CTL2 X
!
! c-c-c-h
CC321   CC321   CC321   HCA2     0.19    3    0.0 ! par27 X CTL2 CTL2 X
!
! o-c-c-o
OC311   CC3161  CC3161  OC311    2.65    1  180.0 ! og ethylene glycol combo
OC311   CC3161  CC3161  OC311    0.00    2    0.0 ! "
OC311   CC3161  CC3161  OC311    0.13    3  180.0 ! "
OC311   CC3162  CC3161  OC311    2.65    1  180.0 ! og ethylene glycol combo
OC311   CC3162  CC3161  OC311    0.00    2    0.0 ! "
OC311   CC3162  CC3161  OC311    0.13    3  180.0 ! "
OC301   CC3162  CC3161  OC311    2.65    1  180.0 ! og ethylene glycol combo
OC301   CC3162  CC3161  OC311    0.00    2    0.0 ! "
OC301   CC3162  CC3161  OC311    0.13    3  180.0 ! "
OC301   CC3161  CC3161  OC311    2.65    1  180.0 ! og ethylene glycol combo
OC301   CC3161  CC3161  OC311    0.00    2    0.0 ! "
OC301   CC3161  CC3161  OC311    0.13    3  180.0 ! "
OC301   CC3161  CC3162  OC311    2.65    1  180.0 ! og ethylene glycol combo
OC301   CC3161  CC3162  OC311    0.00    2    0.0 ! "
OC301   CC3161  CC3162  OC311    0.13    3  180.0 ! "
OC301   CC3161  CC3162  OC301    0.59    1  180.0 ! par35 OC30A CC32A CC32A OC30A
OC301   CC3161  CC3162  OC301    1.16    2    0.0 ! " og/rmv RESMOR and FABYOW10
OC301   CC3161  CC3162  OC3C61   2.75    1  180.0 ! from
OC301   CC3161  CC3162  OC3C61   0.26    2  180.0 ! " OC311 CC3161 CC3162 OC3C61
OC301   CC3161  CC3162  OC3C61   0.10    3    0.0 ! "
OC301   CC3161  CC3163  OC3C61   1.36    1  180.0 ! from
OC301   CC3161  CC3163  OC3C61   0.16    2    0.0 ! " OC311 CC3161 CC3163 OC3C61
OC301   CC3161  CC3163  OC3C61   1.01    3    0.0 ! "
OC311   CC3161  CC3162  OC3C61   2.75    1  180.0
OC311   CC3161  CC3162  OC3C61   0.26    2  180.0
OC311   CC3161  CC3162  OC3C61   0.10    3    0.0
OC311   CC3161  CC3163  OC3C61   1.36    1  180.0
```

```
OC311   CC3161  CC3163  OC3C61   0.16    2    0.0
OC311   CC3161  CC3163  OC3C61   1.01    3    0.0
OC301   CC3161  CC3161  OC301    0.59    1  180.0 !from OC301 CC3161 CC3162 OC301
OC301   CC3161  CC3161  OC301    1.16    2    0.0 ! "
OC30P   CC3161  CC3161  OC311    2.50    1  180.0
OC30P   CC3161  CC3161  OC311    2.00    2    0.0
OC30P   CC3161  CC3161  OC311    2.50    3    0.0
OC30P   CC3162  CC3161  OC311    2.50    1  180.0
OC30P   CC3162  CC3161  OC311    2.00    2    0.0
OC30P   CC3162  CC3161  OC311    2.50    3    0.0
OC30P   CC3161  CC3162  OC301    2.50    1  180.0
OC30P   CC3161  CC3162  OC301    2.00    2    0.0
OC30P   CC3161  CC3162  OC301    2.50    3    0.0
OC30P   CC3161  CC3161  OC301    2.50    1  180.0
OC30P   CC3161  CC3161  OC301    2.00    2    0.0
OC30P   CC3161  CC3161  OC301    2.50    3    0.0
OC30P   CC3161  CC3161  OC30P    1.27    3    0.0
OC3C61  CC3162  CC3161  OC30P    0.61    3    0.0
OC3C61  CC3163  CC3161  OC30P    0.61    3    0.0
OC311   CC3162  CC3161  OC30P    2.50    1  180.0 ! OC30P    CC3162    CC3161    OC311    2.50    1   180.0
OC311   CC3162  CC3161  OC30P    2.00    2    0.0 ! OC30P    CC3162    CC3161    OC311    2.00    2     0.0
OC311   CC3162  CC3161  OC30P    2.50    3    0.0 ! OC30P    CC3162    CC3161    OC311    2.50    3     0.0
OC3C61  CC3163  CC2O2   OC2D2    0.64    2  180.0 ! og amop mp2/ccpvtz
OC301   CC321   CC3163  OC3C61   1.4906  1  180.0 !dpatel modified by LS fitting
OC301   CC321   CC3163  OC3C61   0.3759  2    0.0 !    "
OC301   CC321   CC3163  OC3C61   0.5916  3    0.0 !    "
OC311   CC321   CC3163  OC3C61   0.75    1  180.0
OC311   CC321   CC3163  OC3C61   0.17    2    0.0
OC311   CC321   CC3163  OC3C61   0.25    3  180.0  ! og empirical lower barrier by 1.5 kcal
OC3C61  CC3163  CC321   OC30P    0.06    1    0.0
OC3C61  CC3163  CC321   OC30P    0.78    2    0.0
OC3C61  CC3163  CC321   OC30P    0.74    3  180.0
!
! h-c-c-c
HCA1    CC3161  CC3161  CC3161   0.20    3    0.0 ! par27 X CTL1 CTL1 X
HCA1    CC3162  CC3161  CC3161   0.20    3    0.0 ! par27 X CTL1 CTL1 X
HCA1    CC3161  CC3161  CC3163   0.20    3    0.0 ! par27 X CTL1 CTL1 X
HCA1    CC3161  CC3161  CC3162   0.20    3    0.0 ! par27 X CTL1 CTL1 X
HCA1    CC3163  CC3161  CC3161   0.20    3    0.0 ! par27 X CTL1 CTL1 X
HCA1    CC3161  CC3163  CC321    0.20    3    0.0 ! par27 X CTL1 CTL1 X
HCA1    CC3161  CC3163  CC2O2    0.20    3    0.0 ! par27 X CTL1 CTL1 X
HCA2    CC321   CC3163  CC3161   0.20    3    0.0 ! par27 X CTL1 CTL2 X
!
! h-c-c-o
HCA1    CC3161  CC3161  OC311    0.14    3    0.0 ! og ethanol
HCA1    CC3161  CC3162  OC311    0.14    3    0.0 ! og ethanol
HCA1    CC3162  CC3161  OC311    0.14    3    0.0 ! og ethanol
HCA1    CC3163  CC3161  OC311    0.14    3    0.0 ! og ethanol
HCA1    CC3161  CC3163  OC3C61   0.20    3    0.0 ! par27 X CTL1 CTL1 X
HCA1    CC3161  CC3162  OC3C61   0.20    3    0.0 ! par27 X CTL1 CTL1 X
HCA1    CC3161  CC3162  OC301    0.20    3    0.0 ! par27 X CTL1 CTL1 X
HCA1    CC3161  CC3162  OC30P    0.20    3    0.0 ! par27 X CTL1 CTL2 X
HCA1    CC3162  CC3161  OC30P    0.20    3    0.0 ! par27 X CTL1 CTL2 X
HCA1    CC3163  CC3161  OC30P    0.20    3    0.0 ! par27 X CTL1 CTL2 X
HCA1    CC3161  CC3161  OC30P    0.20    3    0.0 ! par27 X CTL1 CTL2 X
HCA1    CC3163  CC321   OC311    0.14    3    0.0 ! og ethanol
HCA2    CC321   CC3163  OC3C61   0.20    3    0.0 ! par27 X CTL1 CTL2 X
HCA1    CC3163  CC2O2   OC2D2    0.05    6  180.0 ! par22 X CT1  CC   X
HCA1    CC3163  CC321   OC30P    0.00    6  180.0
HCA2    CC321   CC321   OC301    0.19    3    0.0 ! par27 X CTL2 CTL2 X
HCA2    CC321   CC2O2   OC2D2    0.05    6  180.0 ! par22 X CT2  CC   X
```

```
!
! h-c-c-h
HCA1    CC3161  CC3161  HCA1     0.20    3    0.0 ! par27 X CTL1 CTL1 X
HCA1    CC3162  CC3161  HCA1     0.20    3    0.0 ! par27 X CTL1 CTL1 X
HCA1    CC3161  CC3163  HCA1     0.20    3    0.0 ! par27 X CTL1 CTL1 X
HCA1    CC3163  CC321   HCA2     0.20    3    0.0 ! par27 X CTL1 CTL2 X
HCA2    CC321   CC321   HCA2     0.19    3    0.0 ! par27 X CTL2 CTL2 X
!
! o-c-c-h
OC301   CC3161  CC3161  HCA1     0.20    3    0.0 ! par27 X CTL1 CTL1 X
OC301   CC3161  CC3162  HCA1     0.20    3    0.0 ! par27 X CTL1 CTL1 X
OC301   CC3161  CC3163  HCA1     0.20    3    0.0 ! par27 X CTL1 CTL1 X
OC301   CC321   CC3163  HCA1     0.20    3    0.0 ! par27 X CTL1 CTL2 X
!
! c-c-o-h
CC3161  CC3162  OC311   HCP1     0.29    1    0.0
CC3161  CC3162  OC311   HCP1     0.62    2    0.0
CC3161  CC3162  OC311   HCP1     0.05    3    0.0
CC3162  CC3161  OC311   HCP1     0.29    1    0.0
CC3162  CC3161  OC311   HCP1     0.62    2    0.0
CC3162  CC3161  OC311   HCP1     0.05    3    0.0
CC3161  CC3161  OC311   HCP1     0.29    1    0.0
CC3161  CC3161  OC311   HCP1     0.62    2    0.0
CC3161  CC3161  OC311   HCP1     0.05    3    0.0
CC3163  CC3161  OC311   HCP1     0.29    1    0.0
CC3163  CC3161  OC311   HCP1     0.62    2    0.0
CC3163  CC3161  OC311   HCP1     0.05    3    0.0
CC3163  CC321   OC311   HCP1     0.29    1    0.0
CC3163  CC321   OC311   HCP1     0.62    2    0.0
CC3163  CC321   OC311   HCP1     0.05    3    0.0
!
! c-c-o-c
CC3161  CC3162  OC3C61  CC3163   0.20    3    0.0 ! og/sng thp
CC3161  CC3163  OC3C61  CC3162   0.20    3    0.0 ! og/sng thp
CC3161  CC3263  OC3C61  CC3162   0.20    3    0.0 ! og/sng thp
CC3161  CC3162  OC301   CC321    0.22    1    0.0 ! og OMeTHP compounds 2 and 3
CC3161  CC3162  OC301   CC321    0.28    2  180.0 ! "  MCSA fit
CC3161  CC3162  OC301   CC321    0.89    3    0.0 ! "  MP2/cc-pVTZ//MP2/6-31G*
CC3161  CC3162  OC301   CC3161   0.41    1  180.0 ! og THP-1-O-2-CYHX
CC3161  CC3162  OC301   CC3161   0.66    2  180.0 ! "  MCSA fit
CC3161  CC3162  OC301   CC3161   1.60    3    0.0 ! "  MP2/cc-pVTZ//MP2/6-31G*
CC3162  CC3161  OC301   CC3162   0.13    1  180.0 ! og THP-1-O-2-CYHX
CC3162  CC3161  OC301   CC3162   0.25    2  180.0 ! "  MCSA fit
CC3162  CC3161  OC301   CC3162   0.06    3  180.0 ! "  MP2/cc-pVTZ//MP2/6-31G*
CC3161  CC3161  OC301   CC3162   0.13    1  180.0 ! og THP-1-O-2-CYHX
CC3161  CC3161  OC301   CC3162   0.25    2  180.0 ! "  MCSA fit
CC3161  CC3161  OC301   CC3162   0.06    3  180.0 ! "  MP2/cc-pVTZ//MP2/6-31G*
CC3163  CC3161  OC301   CC3162   0.13    1  180.0 ! og THP-1-O-2-CYHX
CC3163  CC3161  OC301   CC3162   0.25    2  180.0 ! "  MCSA fit
CC3163  CC3161  OC301   CC3162   0.06    3  180.0 ! "  MP2/cc-pVTZ//MP2/6-31G*
CC321   CC3163  OC3C61  CC3162   0.20    3    0.0 ! og/sng thp
CC2O2   CC3163  OC3C61  CC3162   0.20    3    0.0 ! og/sng thp
CC3163  CC321   OC301   CC3162   0.64    1  180.0 ! og compounds 11 and 12
CC3163  CC321   OC301   CC3162   0.03    2  180.0 ! "  MCSA fit
CC3163  CC321   OC301   CC3162   0.61    3    0.0 ! "  MP2/cc-pVTZ//MP2/6-31G*
!
! h-c-o-c
HCA1    CC3162  OC3C61  CC3163   0.284   3    0.0 ! par35 HCA2 CC32A OC30A CC32A
HCA1    CC3163  OC3C61  CC3162   0.284   3    0.0 ! par35 HCA2 CC32A OC30A CC32A
!
! h-o-c-h
```

```
HCP1    OC311   CC3161  HCA1     0.18    3    0.0 ! og methanol
HCP1    OC311   CC3162  HCA1     0.18    3    0.0 ! og methanol
!
! h-o-c-o
HCP1    OC311   CC301   OC301    1.55    1    0.0 !from HCP1 OC311 CC3162 OC3C61
HCP1    OC311   CC3162  OC3C61   1.55    1    0.0
HCP1    OC311   CC3162  OC3C61   1.17    2    0.0
HCP1    OC311   CC3162  OC3C61   1.07    3    0.0
!
! c-o-c-h
CC3161  OC301   CC3162  HCA1     0.284   3    0.0 ! par35 HCA2 CC32A OC30A CC32A
CC3162  OC301   CC3161  HCA1     0.284   3    0.0 ! par35 HCA2 CC32A OC30A CC32A
CC321   OC301   CC3162  HCA1     0.284   3    0.0 ! par35 HCA2 CC32A OC30A CC32A
CC3162  OC301   CC321   HCA2     0.284   3    0.0 ! par35 HCA2 CC32A OC30A CC32A
!
! c-o-c-o
CC3161  OC301   CC3162  OC3C61   0.05    1  180.0 ! og THP-1-O-2-CYHX
CC3161  OC301   CC3162  OC3C61   0.91    2    0.0 ! "  MCSA fit
CC3161  OC301   CC3162  OC3C61   1.27    3  180.0 ! "  MP2/cc-pVTZ//MP2/6-31G*
CC3163  OC3C61  CC3162  OC301    0.41    1  180.0 ! og from
CC3163  OC3C61  CC3162  OC301    0.89    2    0.0 ! " CC3163 OC3C61 CC3162 OC311
CC3163  OC3C61  CC3162  OC301    0.05    3    0.0 ! "
CC321   OC301   CC3162  OC3C61   0.14    1    0.0 ! og OMeTHP compounds 2 and 3
CC321   OC301   CC3162  OC3C61   0.97    2    0.0 ! "  MCSA fit
CC321   OC301   CC3162  OC3C61   0.11    3  180.0 ! "  MP2/cc-pVTZ//MP2/6-31G*
CC3163  OC3C61  CC3162  OC311    0.41    1  180.0
CC3163  OC3C61  CC3162  OC311    0.89    2    0.0
CC3163  OC3C61  CC3162  OC311    0.05    3    0.0
CC3163  OC3C61  CC3162  OC30P    2.44    3  180.0
!
! c-o-c-c
CC3162  OC301   CC321   CC321    0.40    1    0.0 ! par35
CC3162  OC301   CC321   CC321    0.49    3    0.0 !  " CC33A CC32A OC30A CC32A
!
! c-c-o-s
CC3162  CC3161  OC30P   SC       0.21    3    0.0
CC3163  CC3161  OC30P   SC       0.21    3    0.0
CC3161  CC3161  OC30P   SC       0.21    3    0.0
CC3163  CC321   OC30P   SC       1.41    1  180.0
CC3163  CC321   OC30P   SC       0.40    2    0.0
CC3163  CC321   OC30P   SC       0.21    3  180.0
!
! h-c-o-s
HCA1    CC3161  OC30P   SC       0.00    3    0.0 ! methylsulfate
HCA2    CC321   OC30P   SC       0.284   3    0.0 !
!
! s-o-c-c
SC      OC30P   CC3162  CC3161   0.21    3    0.0 ! CC3162   CC3161    OC30P     SC       0.21    3     0.0
!
! s-o-c-h
SC      OC30P   CC3162  HCA1     0.00    3    0.0 ! HCA1     CC3161    OC30P     SC       0.00    3     0.0
SC      OC30P   CC3162  HCA1     0.00    3    0.0 ! HCA1     CC3161    OC30P     SC       0.00    3     0.0
!
! s-o-c-o
SC      OC30P   CC3162  OC3C61   0.76    1    0.0 ! OC3C61   CC311D    OC30P     SC       0.76    1     0.0
SC      OC30P   CC3162  OC3C61   0.68    2    0.0 ! OC3C61   CC311D    OC30P     SC       0.68    2     0.0
SC      OC30P   CC3162  OC3C61   0.26    3  180.0 ! OC3C61   CC311D    OC30P     SC       0.26    3   180.0
!
! c-o-s-o
CC3161  OC30P   SC      OC2DP    0.17    3    0.0
CC321   OC30P   SC      OC2DP    0.17    3    0.0
```

```
CC3162  OC30P   SC      OC2DP    0.17    3    0.0 ! CC3161   OC30P     SC   OC2DP    0.17    3     0.0
!
!
!n-c-c-c
NSO3    CC3161  CC3161  CC3161   0.2000  3    0.0 ! PARAMCHEM: from CG321 CG311 CG311 NG2S1,
                                                                                          PENALTY= 9.6
!
!n-c-c-h
NSO3    CC3161  CC3161  HCA1     0.2000  2    0.0 ! PARAMCHEM: NG2S1 CG311 CG311 HGA1, PENALTY= 9
NSO3    CC3161  CC3162  HCA1     0.2000  2    0.0 ! PARAMCHEM: NG2S1 CG311 CG311 HGA1, PENALTY= 9
!
!h-c-n-h
HCA1    CC3161  NNSO3   HNSO3    0.0100  3    0.0 ! PARAMCHEM: from HGA3 CG331 NG2S3 HGP1,
                                                                                        PENALTY= 11
!
!n-c-c-o
NSO3    CC3161  CC3161  OC311    0.2000  3    0.0 ! PARAMCHEM: from NG2S1 CG311 CG311 OG311,
                                                                                          PENALTY= 9
NSO3    CC3161  CC3162  OC301    0.2000  3    0.0 ! PARAMCHEM: from NG2S1 CG311 CG311 OG311,
                                                                                          PENALTY= 9
NSO3    CC3161  CC3162  OC311    0.2000  3    0.0 ! PARAMCHEM: from NG2S1 CG311 CG311 OG311,
                                                                                          PENALTY= 9
NSO3    CC3161  CC3162  OC3C61   2.9950  3  180.0 !!! TOVA E NASHE, model, new_params
!
!h-n-s-o
HNSO3   NNSO3   SNSO3   ONSO3    0.0910  3  180.0 !!! TOVA E NASHE, model, new_params
!
!c-n-s-o
CC3161  NNSO3   SNSO3   ONSO3    0.6310  3    0.0 !!! TOVA E NASHE, model, new_params
CC3161  NNSO3   SNSO3   ONSO3    0.7460  4    0.0 !!! TOVA E NASHE, model, new_params
!
!c-c-n-s
CC3161  CC3161  NNSO3   SNSO3    2.0300  1  180.0 !!! TOVA E NASHE, model, new_params
CC3161  CC3161  NNSO3   SNSO3    1.0250  2  180.0 !!! TOVA E NASHE, model, new_params
CC3161  CC3161  NNSO3   SNSO3    0.3260  3    0.0 !!! TOVA E NASHE, model, new_params
CC3162  CC3161  NNSO3   SNSO3    2.2610  1    0.0 !!! TOVA E NASHE, model, new_params
CC3162  CC3161  NNSO3   SNSO3    1.1800  2  180.0 !!! TOVA E NASHE, model, new_params
CC3162  CC3161  NNSO3   SNSO3    0.2630  3    0.0 !!! TOVA E NASHE, model, new_params
!
!c-c-n-h
CC3161  CC3161  NNSO3   HNSO3    0.0400  1  180.0 !!! TOVA E NASHE, model, new_params
CC3162  CC3161  NNSO3   HNSO3    0.5690  1  180.0 !!! TOVA E NASHE, model, new_params
!
!h-c-n-s
HCA1    CC3161  NNSO3   SNSO3    0.2420  3    0.0 !!! TOVA E NASHE, model, new_params

!************ IMPROPERS ****************************
IMPROPER
!
!V(improper) = Kpsi(psi - psi0)**2
!
!Kpsi: kcal/mole/rad**2
!psi0: degrees
!
!atom types                      Kpsi         psi0
CC2O2   CC3163  OC2D2   OC2D2    96.00   0    0.0 ! par22 OC   X    X    CC
CC2O2   CC321   OC2D2   OC2D2    96.00   0    0.0 ! par22 OC   X    X    CC

!************ VAN DER WAALS PARAMETERS **************
NONBONDED nbxmod  5 atom cdiel shift vatom vdistance vswitch -
cutnb 14.0 ctofnb 12.0 ctonnb 10.0 eps 1.0 e14fac 1.0 wmin 1.5
```

```
!
!V(Lennard-Jones) = Eps,i,j[(Rmin,i,j/ri,j)**12 - 2(Rmin,i,j/ri,j)**6]
!
!epsilon: kcal/mole, Eps,i,j = sqrt(eps,i * eps,j)
!Rmin/2: A, Rmin,i,j = Rmin/2,i + Rmin/2,j
!
!atom  ignored    epsilon      Rmin/2   ignored   eps,1-4       Rmin/2,1-4
!
HCA1      0.0  -0.0450   1.340               ! alkane,isobutane 1/5/05 igor
HCA2      0.0  -0.0350   1.340               ! alkane, igor, 6/05
HCP1      0.0  -0.0460   0.2245              ! CHARMM polar H
HNSO3     0.0  -0.0460   0.2245              ! Analogy from HGP1 in cgen
!
CC3161    0.0  -0.0320   2.000 0.0 -0.01 1.9 ! alkane,isobutane 1/5/05 viv
CC3162    0.0  -0.0320   2.000 0.0 -0.01 1.9 ! alkane,isobutane 1/5/05 viv
CC3163    0.0  -0.0320   2.000 0.0 -0.01 1.9 ! alkane,isobutane 1/5/05 viv
CC321     0.0  -0.0560   2.010 0.0 -0.01 1.9 ! par22, CT2x
CC2O2     0.0  -0.0700   2.000               ! par22, CC
!
NSO3      0.0  -0.2000   1.850               ! Analogy from NG2S3 in cgen
!
OC311     0.0  -0.1921   1.765                     ! og MeOH and EtOH 1/06
OC3C61    0.0  -0.1000   1.650               ! sng thp 1/06
OC301     0.0  -0.1000   1.6500              ! tetrahydropyran sng, all34_ethers_1a OC30A
OC2D2     0.0  -0.1200   1.700               ! par22, OC
OC30P     0.0  -0.1521   1.77
OC2DP     0.0  -0.12     1.70
ONSO3     0.0  -0.1200   1.700               ! Analogy from OG2P1 in cgen
!
SC        0.0  -0.47     2.1                 ! methylsulfate
SNSO3     0.0  -0.4700   2.100               ! Analogy from SG301 in cgen

END
```

Библиография

[1] **Farrar M. A., and Schreiber R. D.**, The molecular cell biology of interferon-γ and its receptor, *Annual Review of Immunology*, 1993, **11**, 571-611.

[2] **Schroder K., Hertzog P. J., Ravasi T., and Hume D. A**, Interferon-γ: an overview of signals, mechanisms and functions, *Journal of Leukocyte Biology*, 2004, **75**, 163-189.

[3] **Schoenborn J. R., and Wilson C. B.**, Regulation of Interferon-γ During Innate and Adaptive Immune Responses, In: Alt F. W., Editor(s), *Advances in Immunology*, 2007, Academic Press, **96**, 41-101.

[4] **Hertzog P., Forster S., and Samarajiwa S.**, Systems Biology of Interferon Responses, *Journal of Interferon & Cytokine Research*, 2011, **31**(1), 5-11.

[5] **Valente G., Ozmen L., Novelli F., Geuna M., Palestro G., Forni G., and Garotta G.** Distribution of interferon-γ receptor in human tissues, *European Journal of Immunology*, 1992, **22**, 2403-2412.

[6] **Pollard K. M., Cauvi D. M., Toomey C. B., Morris K. V., and Kono K. V.**, Interferon-γ and Systemic Autoimmunity, *Discovery Medicine*, 2013, **16**(87), 123-131.

[7] **Zaidi M. R. and Merlino G.**, The Two Faces of Interferon-γ in Cancer, *Clinical Cancer Research*, 2011, **17**(19), 6118-6124.

[8] **Ealick S.E., Cook W. J., Vijay-Kumar S., Carson M., Nagabhushan T. L., Trotta P. P., and Bugg C. E.**, Three-dimensional structure of recombinant human interferon-γ, *Science*, 1991, **252**, 698-702.

[9] **Walter M. V., Windsor W. T., Nagabhushan T. L., Lundell D. J., Lunn C. A., Zauodny P. J., and Narula S. K.**, Crystal structure of a complex between interferon-γ and its soluble high-affinity receptor, *Nature*, 1995, **376**, 230-235.

[10] **Thiel D. J., le Du M.-H., Walter R. L., D'Arcy A., Chène C., Fountoulakis M., Garotta G., Winkler F. K., and Ealick S. E.**, Observation of an unexpected third receptor molecule in the crystal structure of human interferon-γ receptor complex, *Structure*, 2000, **8**(9), 927-936.

[11] **Altobelli G., Nacheva G., Todorova K., Ivanov I., and Karshikoff A.**, Role of the C-Terminal Chain in Human Interferon-γ Stability: An Electrostatic Study, *PROTEINS: Structure, Function, and Genetics*, 2001, **43**, 125–133.

[12] **Jarpe M. A., and Johnson H. M.**, Structure of an epitope in an immunodominant region of the interferon-γ molecule that is involved in receptor interaction, *Journal of Interferon Research*, 1990, **10**, 243–252.

[13] **Russell J. K., Hayes M. P., Carter J. M., Torres B. A., Dunn B. M., Russell S. W., and Johnson H. M.**, Epitope and functional specificity of monoclonal antibodies to mouse interferon-gamma: the synthetic peptide approach, *The Journal of Immunology*, 1986, **136**, 3324–3328.

[14] **Hogrefe H. H., McPhie P., Bekisz J. B., Enterline J. C., Dyer D., Webb D. S. A., Gerrard T. L., and Zoon K. C.**, Amino terminus is essential to the structural integrity of recombinant hIFN-γ, *The Journal of Biological Chemistry*, 1989, **264**(21), 12179-12186.

[15] **Grzesiek S., Döbeli H., Gentz R., Garotta G., Labhardt A. M., and Bax A.**, ^{1}H, ^{13}C, and ^{15}N NMR backbone assignments and secondary structure of human interferon-γ, *Biochemistry*, 1992, **31**, 8180-8190.

[16] **Rinderknecht E., O'Connor B. H., and Rodriguez H.**, Natural human interferon-gamma. Complete amino acid sequence and determination of sites of glycosylation, *The Journal of Biological Chemistry*, 1984, **259**, 6790-6797.

[17] **Pan Y.-C. E., Stern A. S., Familletti P. C., Khan F. R. and Chizzonite R.**, Structural characterization of human interferon γ Heterogeneity of the carboxyl terminus, *European Journal of Biochemistry*, 1987, **166**, 145-149.

[18] **Haelewyn J., Michaels L., Verhaert P., Hoylaertsce M. F., Witters R., and Deley M.**, Interaction of truncated hIFN-γ variants with the IFNGR : crucial importance of 129-Arg, *Biochemical Journal*, 1997, **324**, 591-595.

[19] **Slodowski O., Böhm J., Schöne B., Otto B.**, Carboxy-terminal truncated rhIFN-γ with a substitution of Gln^{133} or Ser^{132} to Leucine leads to higher biological-activity than in the wild type, *FEBS Letters*, 1991, **202**, 1133–1140.

[20] **Lundell D., Lunn C., Dalgarno D., Fossetta J., Greenberg R., Reim R., Grace M., Narula S.**, The carboxyl-terminal region of human interferon gamma is important for biological activity: mutagenic and NMR analysis, *Protein Engineering*, 1991, **4**, 335-341.

[21] **Subramaniam P. S., Mujtaba M. G., Paddy M. R., and Johnson H. M.**, The Carboxyl Terminus of Interferon-γ Contains a Functional Polybasic Nuclear Localization Sequence, *The Journal of Biological Chemistry*, 1999, **274**(1), 403-407.

[22] **Johnson H. M., Torres B. A., Green M. M., Szente B. E., Siler K. I., Larkin III J., and Subramaniam P. S.**, Cytokine-Receptor Complexes as Chaperones for Nuclear Translocation of Signal Transducers. *Biochemical and Biophysical Research Communications*, 1998, **244**, 607–614.

[23] **Celada A., and Schreiber, R. D.**, Internalization and degradation of receptor-bound interferon-gamma by murine macrophages. Demonstration of receptor recycling, *The Journal of Biological Chemistry*, 1987, **139**(1), 147-153.

[24] **Zu X., Jay F. T.**, The E1 functional epitope of the human interferon gamma is a nuclear targeting signal-like element. Mapping of the E1 epitope., *The Journal of Biological Chemistry*, 1991, **266**, 6023-6026.

[25] **Larkin III J., Subramaniam P. S., Torres B. A., and Johnson H. M.**, Differential Properties of Two Putative Nuclear Localization Sequences Found in the Carboxyl-Terminus of Human IFN-γ, *Journal of Interferon & Cytokine Research*, 2001, **21**(6), 341-348.

[26] **Petrov S., Boyanova M., Berzal-Herranz A., Karshikoff A., Nacheva G., and Ivanov I.**, Significance of the putative upstream polybasic nuclear localisation sequence for the biological activity of human interferon-gamma, *Biotechnology & Biotechnological Equipment*, 2009, **23**(1), 1058-1062.

[27] **Marsters S. A., Pennica D., Bach E., Schreiber R. D.**, and Ashkenazi A, Interferon γ signals via a high-affinity multisubunit receptor complex that contains two types of polypeptide chain, *Proceedings of the National Academy of Sciences USA*, 1995, **92**, 5401-5405.

[28] **Bach E. A., Tanner J. W., Marsters S., Ashkenazi A., Aguet M., Shaw A. S., and Schreiber R. D.**, Ligand-Induced Assembly and Activation of the Gamma Interferon Receptor in Intact Cells, *Molecular And Cellular Biology*, 1996, **16**(6), 3214-3221.

[29] **Pestka S., Kotenko S. V., Muthukumaran G., Cook J. R., and Garotta G.**, The Interferon Gamma (IFN-γ) Receptor: a Paradigm for the Multichain Cytokine Receptor, *Cytokine & Growth Factor Reviews*, 1997, **8**(3), 189-206.

[30] **Stüber D., Friedlein A., Fountoulakis M., Lahm H. W., Garotta G.**, Alignment of disulfide bonds of the extracellular domain of the interferon-γ receptor and investigation of their role in biological activity, *Biochemistry*, 1993, **32**, 2423-1430.

[31] http://www.pdb.org

[32] **Johnson H. M., Noon-Song E. N., Dabelic R., and Ahmed C. M.**, IFN signaling: how a non-canonical model led to the development of IFN mimetics, *Frontiers in Immunology*, 2013, **4**, 202.

[33] **Alberts B., Johnson A., Lewis J., Raff M., Roberts K., and Walter P.**, The Extracellular Matrix of Animals, *in Molecular Biology of the Cell*, 4th edition, 2002, New York: Garland Science.

[34] **Hileman R. E., Fromm J. R., Weiler J. M., and Linhardt R. J.**, Glycosaminoglycan-protein interactions: definition of consensus sites in glycosaminoglycan binding proteins, *BioEssays*, 1998, **20**, 156-167.

[35] **Bernfield M., Gotte M., Park P. W., Reizes O., Fitzgerald M. L., Lincecum J., and Zako M.**, Functions of cell surface heparan sulfate proteoglycans, *Annual Review of Biochemistry*, 1999, **68**, 729-777.

[36] **Xu D., and Esko J. D.**, Demystifying Heparan Sulfate–Protein Interactions, *Annual Review of Biochemistry*, 2014, **83**, 129-157.

[37] **Camejo E. H., Rosengren B., Camejo G,, Sartipy P., Fager G., and Bondjers G.**, Interferon Gamma Binds to Extracellular Matrix Chondroitin-Sulfate Proteoglycans, Thus Enhancing Its Cellular Response, *Arteriosclerosis, Thrombosis, and Vascular Biology*, 1995, **15**, 1456-1465.

[38] **Lortat-Jacob H., Turnbull J. K., and Grimaud J.-A.**, Molecular organization of the interferon y-binding domain in heparan sulphate, *Biochemical Journal*, 1995, **310**, 497-505.

[39] **Gallagher J. T., Lyon M., and Steward W. P.**, Structure and function of heparan sulphate proteoglycans, *Biochemical Journal*, 1986, **236**, 313-325.

[40] **Mulloy B., and Forster M. J.**, Conformation and dynamics of heparin and heparan sulfate, *Glycobiology*, 2000, **10**(11), 1147-1156.

[41] **Lortat-Jacob H.**, Interferon and heparan sulphate, **Biochemical Society Transactions**, 2006, **34**(Pt 3), 461-464.

[42] **Sanchez H. P., Tatarenko K., Nigen M., Pavlov G., Imberty A., Lortat-Jacob H., De la Torre J. G., and Ebel C.**, Organization of Human Interferon γ–Heparin Complexes from Solution Properties and Hydrodynamics, *Biochemistry*, 2006, **45**, 13227-13238.

[43] **Saesen E., Sarrazin S., Laguri C., Sadir R., Maurin D., Thomas A., Imberty A., and Lortat-Jacob H.**, Insights into the Mechanism by Which Interferon-γ Basic Amino Acid Clusters Mediate Protein Binding to Heparan Sulfate, *Journal of the American Chemical Society*, 2013, **135**, 9384-9390.

[44] **Frenkel D., and Smit B.**, *Understanding Molecular Simulation*, 2nd Ed., 2002, Academic Press, San Diego.

[45] **Leach, A.**, *Molecular Modelling: Principles and Applications*, 2nd Ed., 2001, Prentice-Hall.

[46] **Beauchamp K. A., Lin Y.-S., Das R., and Pande V. S.**, Are Protein Force Fields Getting Better? A Systematic Benchmark on 524 Diverse NMR Measurements, *Journal of Chemical Theory and Computation*, 2011, **8**, 1409-1414.

[47] **van der Spoel D., and van Maaren P. J.**, The Origin of Layer Structure Artifacts in Simulations of Liquid Water, *Journal of Chemical Theory and Computation*, 2006, **2**, 1-11.

[48] **Darden T., York D., and Pederson L.**, Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems, *Journal of Chemical Physics*, 1993, **98**, 10089-10092.

[49] **Abraham M. J., van der Spoel D., Lindahl E., Hess B., and the GROMACS development team**, GROMACS User Manual version 5.0.4, www.gromacs.org, 2014.

[50] **Ryckaert J.-P., Ciccotti G., Berendsen H. J. C.**, Numerical integration of the cartesian equations of motion of a system with constraints; molecular dynamics of n-alkanes, *Journal of Computational Physics*, 1977, **23**, 327-341.

[51] **Andersen H. C.**, Rattle: A "Velocity" Version of the Shake Algorithm for Molecular Dynamics Calculations, *Journal of Computational Physics*, 1983, **52**, 24-34.

[52] **Tuckerman M., Berne B. J., and Martyna G. J.**, Reversible multiple time scale molecular dynamics, *The Journal of Chemical Physics*, 1992, **97**(3), 1990-2001.

[53] **Ma Q., Izaguirre J. A., Skeel R. D.**, Verlet-I/R-RESPA/IMPULSE is limited by nonlinear instabilities, *SIAM Journal on Scientific Computing*, 2003, **24**, 1951-1973.

[54] **Garcia-Archilla B., Sanz-Serna J. M., and Skeel R. D.**, Long-time-step methods for oscillatory differential equations, *SIAM Journal on Scientific Computing*, 1998, **20**(3), 930-963.

[55] **Berendsen H. J. C., Postma J. P. M., van Gunsteren W. F., DiNola A., Haak J. R.**, Molecular-Dynamics with Coupling to an External Bath, *Journal of Chemical Physics*, 1984, **81**(8), 3684-3690.

[56] **Grest G. S., and Kremer K.**, Molecular-dynamics simulation for polymers in the presence of a heat bath, *Physical Review A*, 1986, **33**(5), 3628-3631.

[57] **Bussi G., Donadio D., and Parrinello M**, Canonical sampling through velocity rescaling, *The Journal of Chemical Physics*, 2007, **126**, 014101.

[58] **Parrinello M., and Rahman A.**, Polymorphic transitions in single crystals: A new molecular dynamics method, *Journal of Applied Physics*, 1981, **52**, 7182-7190.

[59] **Klepeis J. L., Lindorff-Larsen K., Dror R. O., Shaw D. E.**, Long-timescale molecular dynamics simulations of protein structure and function, *Current Opinion in Structural Biology*, 2009, **19**, 120-127.

[60] **Chipot C., and Pohorille A., Eds.** *Free Energy Calculations: Theory and Applications in Chemistry and Biology*, 2007, Springer Series in Chemical Physics, Volume 86.

[61] **Berendsen H. J. C.**, *Simulating the Physical World, Hierarchical Modeling from Quantum Mechanics to Fluid Dynamics*, 2007, Cambridge University Press.

[62] **Torrie G. M., and Valleau J. P.**, Nonphysical sampling distributions in monte carlo free energy estimation: Umbrella sampling, *Journal of Computational Physics*, 1977, **23**, 187-199.

[63] **Darve E., and Pohorille A.**, Calculating free energies using average force, *The Journal of Chemical Physics*, 2001, **115**, 9169-9183.

[64] **Gullingsrud J., Braun R., and Schulten K.**, Reconstructing potentials of mean force through time series analysis of steered molecular dynamics simulations, *Journal of Computational Physics*, 1999, **151**, 190-211.

[65] **Laio A., and Parrinello M.**, Escaping free-energy minima, *Proceedings of the National Academy of Sciences*, 2002, **99**(20), 12562-12566.

[66] **Laio A., and Gervasio F. L.**, Metadynamics: a method to simulate rare events and reconstruct the free energy in biophysics, chemistry and material science, *Reports on Progress in Physics*, 2008, **71**, 126601.

[67] **Van Speybroeck V., De Wispelaere K., Van der Mynsbrugge J., Vandichel M., Hemelsoet K., and Waroquier M.**, First principle chemical kinetics in zeolites: the methanol-to-olefin process as a case study, *Chemical Society Reviews*, 2014, **43**(21), 7326-7357.

[68] **Ensing B., Laio A., Parrinello M., and Klein M. L.**, A Recipe for the Computation of the Free Energy Barrier and the Lowest Free Energy Path of Concerted Reactions, *The Journal of Physical Chemistry B*, 2005, **109**, 6676-6687.

[69] **Barducci A., Bussi G., and Parrinello M.**, Well-Tempered Metadynamics: A Smoothly Converging and Tunable Free-Energy Method, *Physical Review Letters*, 2008, **100**, 020603.

[70] **Bussi G., Gervasio F. L., Laio A., and Parrinello M.**, Free-Energy Landscape for β-Hairpin Folding from Combined Parallel Tempering and Metadynamics, *Journal of the American Chemical Society*, 2006, **128**(41), 13435-13441.

[71] **Piana S., and Laio A.**, A bias-exchange approach to protein folding, *The Journal of Physical Chemistry B*, 2007, **111**(17), 4553-4559.

[72] **Branduardi D., Gervasio F. L., and Parrinello M.**, From A to B in free energy space, *The Journal of Chemical Physics*, 2007, **126**(5), 054103.

[73] **Dellago C., Bolhuis P. G., Csajka F. S., and Chandler D.**, Transition path sampling and the calculation of rate constants, *The Journal of Chemical Physics*, 1998, **108**(5), 1964-1977.

[74] **Bolhuis P. G., Chandler D., Dellago C, and Geissler P. L.**, Transition Path Sampling: Throwing Ropes Over Rough Mountain Passes, in the Dark, *Annual Review of Physical Chemistry*, 2002, **53**, 291-318.

[75] http://scc.acad.bg/ncsa/index.php/bg/

[76] **Hess B., Kutzner C., van der Spoel D., and Lindahl E.**, GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation, *Journal of Chemical Theory and Computation*, 2008, **4**(3), 435-447.

[77] **Phillips J. C., Braun R., Wang W., Gumbart J., Tajkhorshid E., Villa E., Chipot C., Skeel R. D., Kale L., and Schulten K.**, Scalable molecular dynamics with NAMD, *Journal of Computational Chemistry*, 2005, **26**, 1781-1802.

[78] **Geimer M., Wolf F., Wylie B. J. N., Ábrahám E., Becker D., and Mohr B.**, The Scalasca performance toolset architecture, *Concurrency and Computation:Practice and Experience*, 2010, **22**(6), 702-719.

[79] **Ramachandran G. N., Ramakrishnan C., and Sasisekharan V.**, Stereochemistry of polypeptide chain configurations, *Journal of Molecular Biology* , 1963, **7**, 95-99.

[80] **Lovell S. C., Davis I. W., Arendall W. B. 3rd, de Bakker P. I., Word J. M., Prisant M. G., Richardson J. S., Richardson D. C**, Structure validation by C_α geometry: ϕ, ψ and C_β deviation, *Proteins*, 2003, **50**(3), 437-450.

[81] http://proteopedia.org/wiki/index.php/Ramachandran_Plots

[82] **Kabsch W., and Sander C.**, Dictionary of Protein Secondary Structure: Pattern Recognition of Hydrogen-Bonded and Geometrical Features, *Biopolymers*, 1983, **22**, 2577-2637.

[83] The PyMOL Molecular Graphics System, Version 1.7.4 Schrödinger, LLC., https://www.pymol.org

[84] **MacKerell Jr., A. D., Bashford D., Bellott M., Dunbrack Jr., R. L., Evanseck J. D., Field M. J., Fischer S., Gao J., Guo H., Ha S., Joseph-McCarthy D., Kuchnir L., Kuczera K., Lau F. T. K., Mattos C., Michnick S., Ngo T., Nguyen D. T., Prodhom B., Reiher III W. E., Roux B., Schlenkrich M., Smith J. C., Stote R., Straub J., Watanabe M., Wiorkiewicz-Kuczera J., Yin D., and Karplus M.**, All-atom empirical potential for molecular modeling and dynamics studies of proteins, *Journal of Physical Chemistry B*, 1998, **102**, 3586-3616.

[85] **Schmid N., Eichenberger A. P., Choutko A., Riniker S., Winger M., Mark A. E., van Gunsteren W. F.**, Definition and testing of the GROMOS force-field versions 54A7 and 54B7, *European Biophysics Journal*, 2011, **40**(7), 843-856.

[86] **Jorgensen W. L., Maxwell D. S., Tirado-Rives J.**, Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids, *Journal of the American Chemical Society*, 1996, **118**(45), 11225-11236.

[87] **Wang J. M., Cieplak P., Kollman P. A.**, How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules?, *Journal of Computational Chemistry* , 2000, **21**, 1049-1074.

[88] **Wang J., Wolf R. M., Caldwell J. W., Kollman P. A., Case D. A.**, Development and testing of a general AMBER force field, *Journal of Computational Chemistry*, 2004, **25**, 1157-1174.

[89] **Vanommeslaeghe K., Hatcher E., Acharya C., Kundu S., Zhong S., Shim J., Darian E., Guvench O., Lopes P., Vorobyov I., MacKerell Jr. A. D.**, CHARMM General Force Field (CGenFF): A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields, *Journal of Computational Chemistry*, 2010, **31**, 671-690.

[90] **Best R. B., Zhu X., Shim J., Lopes P. E. M., Mittal J., Feig M., and MacKerell Jr. A. D.**, Optimization of the Additive CHARMM All-Atom Protein Force Field Targeting Improved Sampling of the Backbone ϕ, ψ and Side-Chain χ_1 and χ_2 Dihedral Angles, *Journal of Chemical Theory and Computation*, 2012, **8**(9), 3257-3273.

[91] http://mackerell.umaryland.edu/charmm_ff.shtml

[92] **Guvench O., Mallajosyula S. S., Raman E. P., Hatcher E. R., Vanommeslaeghe K., Foster T. J., Jamison F. W., and MacKerell Jr. A. D.**, CHARMM additive all-atom force field for carbohydrate derivatives and its utility in polysaccharide and carbohydrate-protein modeling, *Journal of Chemical Theory and Computation*, 2011, **7**(10), 3162-3180.

[93] **Mallajosyula S. S., Guvench O, Hatcher E. R., and MacKerell Jr. A. D.**, CHARMM Additive All-Atom Force Field for Phosphate and Sulfate Linked to Carbohydrates, *Journal of Chemical Theory and Computation*, 2012, **8**(2), 759-776.

[94] http://mackerell.umaryland.edu/~kenno/cgenff/download.shtml

[95] **Vanommeslaeghe K., MacKerell Jr. A. D.**, Automation of the CHARMM General Force Field (CGenFF) I: bond perception and atom typing, *Journal of Chemical Information and Modeling*, 2012, **52**, 3144-3154.

[96] **Vanommeslaeghe K., Raman E. P., MacKerell Jr. A. D.**, Automation of the CHARMM General Force Field (CGenFF) II: Assignment of bonded parameters and partial atomic charges, *Journal of Chemical Information and Modeling*, 2012, **52**, 3155-3168.

[97] https://cgenff.paramchem.org

[98] **Brooks B. R., Brooks III C. L., MacKerell Jr. A. D., Nilsson L., Petrella R. J., Roux B., Won Y., Archontis G., Bartels C., Boresch S., Caflisch A., Caves L., Cui Q., Dinner A. R., Feig M., Fischer S., Gao J., Hodoscek M., Im W., Kuczera K., Lazaridis T., Ma J., Ovchinnikov V., Paci E., Pastor R. W., Post C. B., Pu J. Z., Schaefer M., Tidor B., Venable R. M., Woodcock H. L., Wu X., Yang W., York D. M., and Karplus M.**, CHARMM: The Biomolecular Simulation Program, J*ournal of computational chemistry*, 2009, **30**(10), 1545-1614.

[99] **Mayne C. G., Saam J., Schulten K., Tajkhorshid E., Gumbart J. C.**, Rapid parameterization of small molecules using the force field toolkit, *Journal of Computational Chemistry*, 2013, **34**(32), 2757–2770.

[100] **Humphrey W., Dalke A., and Schulten K.**, VMD – Visual Molecular Dynamics, *Journal of Molecular Graphics*, 1996, **14**, 33-38.

[101] **Mulloy B., Forster M. J., Jones C., Davies D. B.**, N.m.r. and molecular-modelling studies of the solution conformation of heparin, *Biochemical Journal*, 1993, **293**, 849-858.

[102] **Frisch M. J., Trucks G. W., Schlegel H. B., Scuseria G. E., Robb M. A., Cheeseman J. R., Scalmani G., Barone V., Mennucci B., Petersson G. A., Nakatsuji H., Caricato M., Li X., Hratchian H. P., Izmaylov A. F., Bloino J., Zheng G., Sonnenberg J. L., Hada M., Ehara M., Toyota K., Fukuda R., Hasegawa J., Ishida M., Nakajima T., Honda Y., Kitao O., Nakai H., Vreven T., Montgomery J. A. Jr., Peralta J. E., Ogliaro F., Bearpark M., Heyd J. J., Brothers E., Kudin K. N., Staroverov V. N., Kobayashi R., Normand J., Raghavachari K., Rendell A., Burant J. C., Iyengar S. S., Tomasi J., Cossi M., Rega N., Millam J. M., Klene M., Knox J. E., Cross J. B., Bakken V., Adamo C., Jaramillo J., Gomperts R., Stratmann R. E., Yazyev O., Austin A. J., Cammi R., Pomelli C., Ochterski J. W., Martin R. L., Morokuma K., Zakrzewski V. G., Voth G. A., Salvador P., Dannenberg J. J., Dapprich S., Daniels A. D., Farkas Ö., Foresman J. B., Ortiz J. V., Cioslowski J., Fox D. J.**, Gaussian 09, Revision A.02, Gaussian, Inc., Wallingford CT, 2009.

[103] **Møller C., and Plesset M. S.**, Note on an approximation treatment for many-electron systems, **Physical Review**, 1934, **46**, 0618-22.

[104] http://www.ks.uiuc.edu/Research/vmd/plugins/namdenergy/

[105] http://www.charmm.org/ubbthreads/ubbthreads.php?ubb=showflat&Number=33420

[106] **Khan S., Gor J., Mulloy B., and Perkins S. J.**, Semi-Rigid Solution Structures of Heparin by Constrained X-ray Scattering Modelling: New Insight into Heparin–Protein Complexes, *Journal of Molecular Biology*, 2010, **395**, 504-521.

[107] http://physon.phys.uni-sofia.bg/

[108] **Maiorov V. N., and Crippen G. M.**, Size-Independent Comparison of Protein Three-Dimensional Structures, *PROTEINS: Structure, Function, and Genetics*, 1995, **22**, 273-283.

[109] **Daura X., Gademann K., Jaun B., Seebach D., van Gunsteren W. F., and Mark A. E.**, Peptide Folding: When Simulation Meets Experiment, *Angewandte Chemie International Edition*, 1999, **38**, 236-240.