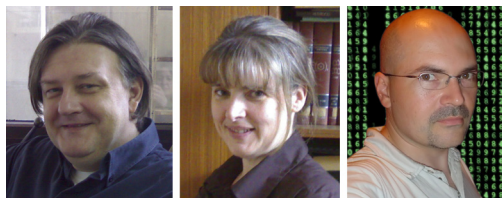




## ЧЛАНЦИ



Марио ЗЛАТОВИЋ, Хемијски факултет, [mario@chem.bg.ac.yu](mailto:mario@chem.bg.ac.yu)  
Слађана КОСТИЋ-РАЈАЧИЋ, ИХТМ Центар за хемију,  
Универзитет у Београду  
Владимир ШУКАЛОВИЋ, ИХТМ Центар за хемију,  
Универзитет у Београду

### ХОМОЛОГО МОДЕЛОВАЊЕ СТРУКТУРЕ ПРОТЕИНА

Познавање тродимензионалне структуре протеина је од непроцењиве важности при утврђивању понашања и механизма дејства рецептора и ензима у интеракцији са другим молекулима. Упркос растућем броју разрешених 3Д структура протеина, број утврђених протеинских секвенци увелико прелази број дефинисаних просторних структура протеина. Осим тога, базе података секвенци расту много брже него базе 3Д структура протеина. У случају када су кристалне структуре или подаци добијени NMR-ом недоступни, морамо прибећи следећем „најбољем” решењу, а то је моделовање просторне структуре. Један од начина који се може применити да би се добили употребљиви модели протеина је хомолого моделовање. Ова метода се базира на дизајну тродимензионалне структуре непознатог протеина на основу њему хомологог протеина чија је структура позната. Да би се стекла слика о методи, у овом тексту ћемо изложити неколико аспеката хомологог моделовања.

При утврђивању понашања и механизма дејства рецептора и ензима у интеракцији са другим молекулима од непроцењиве важности је, свакако, познавање тродимензионалне структуре протеина. Ту структуру је могуће добити кристалографским или NMR мерењима<sup>[1]</sup>. После пречишћавања, ове структуре се могу користити за предвиђање и квантификацију различитих особина протеина, као и за испитивање интеракција лиганата и протеина. Подаци о 3Д структури до сада одређених кристалних структура протеина су доступни преко Protein Data Bank (PDB)<sup>[2]</sup> базе података (<http://www.rcsb.org/>). У овој бази података су скупљене атомске координате протеина и ДНК и до 2. децембра 2008. године она је садржавала 54559 структура. Брзи раст броја структура у овој бази можемо видети из податка да је крајем 2006. она садржавала око 41000 структура.

Упркос растућем броју разрешених 3Д структура протеина, број утврђених протеинских секвенци увелико прелази број дефинисаних просторних структура протеина, а осим тога, базе података секвенци расту много брже него базе 3Д структура протеина. Уско грло у проблему утврђивања 3Д структуре протеина је свакако компликована и дугачка процедура добијања квалитетног кристала. Осим тога, велики број протеина (преко 30%), као што су то трансмембрански рецептори, јонски канали и сл., није ни могуће издвојити у нативном облику да би им се кристалографски утврдила структура. Извесно побољшање у овој области је донео напредак NMR техника, али је разрешавање секвенци нуклеинских киселина које кодирају протеине далеко брже него што то експерименти могу да прате.

Да би кристалографски подаци били употребљиви, пожељно је да резолуција на којој је рађена кристалографска анализа буде што мања. Оквирна горња граница употребљивости је резолуција од 2,5 Å. С друге стране, одређивање структуре NMR-ом је ограничено донекле и величином протеина. У случају када су кристалне структуре или подаци добијени NMR-ом недоступни, морамо прибећи следећем „најбољем” решењу, а то је моделовање просторне структуре. Знајући колико је структура протеина сложена и од колико фактора може да зависи, поставља се логично питање колико су уопште модели у овој области употребљиви.

Један од начина који се може применити да би се добили употребљиви модели протеина је хомолого моделовање, понекад названо и компаративно моделовање. Ова метода се базира на дизајну тродимензионалне структуре протеина из секвенце на основу структуре њему хомологог протеина. О овом методу је написано неколико ревијских чланака<sup>[3,4]</sup> па ће овде бити дато само неколико аспеката хомологог моделовања, да би се стекла слика о методи.

<sup>1</sup> Rule, G.R., Hitchens, T.R., *Fundamentals of Protein NMR Spectroscopy*, Springer, 2006, ISBN 1402034997.

<sup>2</sup> Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., Bourne, P.E., *Nuc. Acids Res.*, **28** (2000) 235.

<sup>3</sup> Šali, A., Overington, J.P., Johnson, M.S., Blundell, T.L., *Trends Biochem. Sci.*, **15** (1990) 235.

<sup>4</sup> Johnson, M.S., Srinivasan, N., Sowdhamini, R., Blundell, T.L., *Crit. Rev. Biochem. Mol. Biol.*, **29** (1994) 1.

При употреби термина „хомолого моделовање” морамо имати на уму да термин хомологија не значи обавезно и сличност у структури. Прецизније значење термина хомологија је „са истим еволутивним йореклом”<sup>[1,2]</sup>. Стога је хомологија у ствари квалитативан опис односа два или више протеина, а сличности у секвенци или тродимензионалној структури могу бити додатни податак за подржавање тезе о хомологији, помоћу којих се може квантитативно описати њихов однос.

Основна поставка оваквог моделовања лежи у запаженој чињеници да је за групу протеина који су хомологи њихова тродимензионална структура више конзервирана него примарна. Због тога можда и није потпуно исправно да се називи компаративно и хомолого моделовање сматрају синонимима. Компаративно моделовање полази од премисе да ће подударност у примарној структури довести до идентичног савијања у простору па стога и до сличних тродимензионалних структура и функције, док хомолого моделовање полази од становишта да се током молекуларне еволуције протеина међу хомологим протеинима лакше задржавала просторна структура од секвенце<sup>[3]</sup>. На основу овога је било могуће добити просторне моделе протеина чак и у случајевима када је сличност секвенци била релативно ниска.

Процес хомологог моделовања протеина непознате структуре чију секвенцу имамо се састоји од неколико корака:

- Идентификација протеина познате тродимензионалне структуре хомолого са нашом секвенцом и утврђивање степена сличности секвенци. Што је проценат сличности већи, то је и моделовање лакше.
- Поравнавање секвенци протеина.
- Идентификовање структурно конзервираних (SCR, *Structure Conserved Region*) и структурно варијабилних (SVR, *Structure Variable Region*) региона.
- Генерисање координата за SCR протеина непознате просторне структуре на основу познате 3Д структуре.
- Генерисање конформација за петље (SVR) у моделованој структури.
- Генерисање конформација бочних остатака.
- Рафинисање и евалуација добијене структуре.

#### Идентификација хомолога

Постоји неколико рачунарских метода за испомоћ при идентификацији хомолога. У већини случајева ми

имамо секвенцу протеина непознате тродимензионалне структуре. Потом претражујемо базе података протеина чија је структура већ позната и покушавамо да нађемо одређен степен сличности у секвенци. Неке од метода које су данас у употреби су FASTA и BLAST.<sup>[4]</sup> Обе ове методе су имплементирани у многим комерцијалним пакетима, као што је и MODELLER.<sup>[5]</sup> Одабрани хомолог познате структуре се користи као предлог за дизајн непознате 3Д структуре. У идеалном случају као предлог за развој непознате структуре ћемо имати више хомолога, али је моделовање могуће извести и са једном структуром.

Поређење секвенци се такође може употребити и за идентификацију функције неокarakterисаног протеина за који је примарна структура одређена из секвенци ДНК.

У ређим случајевима потребно је да идентификујемо хомологију код протеина чија је тродимензионална структура позната. Тада ћемо, због чињенице да је просторна структура протеина конзервирана у већем степену од секвенце, потрагу за хомолозима извести радије преко просторних координата а нађени хомолози могу да покажу релативно малу сличност у секвенци.

#### Поравнавање секвенци йрошеина

Поравнавање секвенци непознате и познате 3Д структуре хомолога је критичан корак у моделовању. У основи, две или више секвенци се поравнавају једна изнад друге трудећи се да се што више аминокиселина из једне секвенце по свом положају поклапа са истим или сродним аминокиселинама из друге секвенце. Да би се постигло што боље поравнавање, некада је потребно у секвенцу унети празне просторе (*gap*) или изоставити делове секвенце (*deletion*). За ову процедуру су развијени многи методи и некада је кључна одлука који метод применити. Да би могао правилно да интерпретира резултате, истраживач мора да разуме опције које су примењене у овим процедурама поравнавања. При извођењу поравнавања треба одлучити који алгоритам за поравнавање одабрати, који метод вредновања применити и како третирали празне просторе. Алгоритми за поравнавање секвенци се генерално базирају на Нидлман-Вуншовом (Needleman-Wunsch) динамичком алгоритму<sup>[6]</sup> и данас постоји више програмских решења као што су CLUSTALW, KALIGN<sup>[7]</sup> или CLUSTALX<sup>[8]</sup>.

За одеђивање квалитета поравнавања користе се типичне 20×20 матрице вредновања у којима је бројем вреднована могућност да се једна аминокиселина у секвенци нађе на месту где се налази друга у секвенца-

<sup>1</sup> Lewin, R., *Science*, 237 (1987) 1570.

<sup>2</sup> Reeck, G.R., de Haën, C., Teller, D.C., Doolittle, R.F., Fitch, W.M., Dickerson, R.E., Chambon, P., McLachlan, A.D., Margoliash, E., Jukes, T.H., Zuckerkandl, E., *Cell*, 50 (1987) 667.

<sup>3</sup> Vriend, G., Sander, C., *Proteins*, 11 (1991) 52.

<sup>4</sup> Pearson, W.R., *Meth. Enzym.*, 183 (1990) 63; Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., *J. Mol. Biol.*, 215 (1990) 403.

<sup>5</sup> Sali, A., Blundell, T.J., *J. Mol. Biol.*, 234 (1993) 779.

<sup>6</sup> Needleman, S.B. and Wunsch, C.D., *J. Mol. Biol.*, 48 (1970) 442.

<sup>7</sup> Devereux, J., Haerberli, P., Smithies, O., *Nuc. Acids Res.*, 12 (1984) 387; Lassmann, T., Sonnhammer, E.L.L., *Nuc. Acids Res.*, 34 (2006) W596.

<sup>8</sup> Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., Higgins, D.G., *Nuc. Acids Res.*, 25 (1997) 4876.

	A	R	N	D	C	Q	E	G	H	I	L	K	M	F	P	S	T	W	Y	V	B	Z	X	
A	2	-2	0	0	-2	0	0	1	-1	-1	-2	-1	-1	-3	1	1	1	-6	-3	0	0	0	0	-8
R	-2	6	0	-1	-4	1	-1	-3	2	-2	-3	3	0	-4	0	0	-1	2	-4	-2	-1	0	-1	-8
N	0	0	2	2	-4	1	1	0	2	-2	-3	1	-2	-3	0	1	0	-4	-2	-2	2	1	0	-8
D	0	-1	2	4	-5	2	3	1	1	-2	-4	0	-3	-6	-1	0	0	-7	-4	-2	3	3	-1	-8
C	-2	-4	-4	-5	12	-5	-5	-3	-3	-2	-6	-5	-5	-4	-3	0	-2	-8	0	-2	-4	-5	-3	-8
Q	0	1	1	2	-5	4	2	-1	3	-2	-2	1	-1	-5	0	-1	-1	-5	-4	-2	1	3	-1	-8
E	0	-1	1	3	-5	2	4	0	1	-2	-3	0	-2	-5	-1	0	0	-7	-4	-2	3	3	-1	-8
G	1	-3	0	1	-3	-1	0	5	-2	-3	-4	-2	-3	-5	0	1	0	-7	-5	-1	0	0	-1	-8
H	-1	2	2	1	-3	3	1	-2	6	-2	-2	0	-2	-2	0	-1	-1	-3	0	-2	1	2	-1	-8
I	-1	-2	-2	-2	-2	-2	-2	-3	-2	5	2	-2	2	1	-2	-1	0	-5	-1	4	-2	-2	-1	-8
L	-2	-3	-3	-4	-6	-2	-3	-4	-2	2	6	-3	4	2	-3	-3	-2	-2	-1	2	-3	-3	-1	-8
K	-1	3	1	0	-5	1	0	-2	0	-2	-3	5	0	-5	-1	0	0	-3	-4	-2	1	0	-1	-8
M	-1	0	-2	-3	-5	-1	-2	-3	-2	2	4	0	6	0	-2	-2	-1	-4	-2	2	-2	-2	-1	-8
F	-3	-4	-3	-6	-4	-5	-5	-5	-2	1	2	-5	0	9	-5	-3	-3	0	7	-1	-4	-5	-2	-8
P	1	0	0	-1	-3	0	-1	0	0	-2	-3	-1	-2	-5	6	1	0	-6	-5	-1	-1	0	-1	-8
S	1	0	1	0	0	-1	0	1	-1	-1	-3	0	-2	-3	1	2	1	-2	-3	-1	0	0	0	-8
T	1	-1	0	0	-2	-1	0	0	-1	0	-2	0	-1	-3	0	1	3	-5	-3	0	0	-1	0	-8
W	-6	2	-4	-7	-8	-5	-7	-7	-3	-5	-2	-3	-4	0	-6	-2	-5	17	0	-6	-5	-6	-4	-8
Y	-3	-4	-2	-4	0	-4	-4	-5	0	-1	-1	-4	-2	7	-5	-3	-3	0	10	-2	-3	-4	-2	-8
V	0	-2	-2	-2	-2	-2	-2	-1	-2	4	2	-2	2	-1	-1	-1	0	-6	-2	4	-2	-2	-1	-8
B	0	-1	2	3	-4	1	3	0	1	-2	-3	1	-2	-4	-1	0	0	-5	-3	-2	3	2	-1	-8
Z	0	0	1	3	-5	3	3	0	2	-2	-3	0	-2	-5	0	0	-1	-6	-4	-2	2	3	-1	-8
X	0	-1	0	-1	-3	-1	-1	-1	-	-1	-1	-1	-1	-2	-1	0	0	-4	-2	-1	-1	-1	-1	-8
*	-8	-8	-8	-8	-8	-8	-8	-8	-8	-8	-8	-8	-8	-8	-8	-8	-8	-8	-8	-8	-8	-8	-8	1

Слика 1. Пример PAM250 матрице.

ма које користимо као предложак. Што је већи број у матрици то је замена једне аминокиселине другом повољнија. Ове матрице се могу формирати на основу следећих параметара:

- Матрица идентитета: ово је најједноставнија матрица где је идентичним паровима дата вредност 1, док остали добијају вредност 0.
- Матрица сличности физичких особина: ова матрица се формира на основу сличности неких физичких особина аминокиселина као што су хидрофобност, поларизабилност, тенденција да се формира хеликоидална структура...
- Матрица кодонске супституције: у овој матрици се посматра број потребних промена база у нуклеотидном триплету да би се променио кодон за аминокиселину. Идентичне аминокиселине имају вредност 9, једна потребна мутација даје вредност 3, две дају вредност 1.
- Мутациона матрица (Dayhoff) у обзир узима фреквенцију примећених мутација у поравнатим секвенцама.

Једну од првих мутационих матрица су развили Дејхоф (Dayhoff) и сарадници<sup>[1]</sup> анализирајући еволуцију протеина. Ове матрице се називају PAM матрице (Percentage of Acceptable Point Mutations) и означавају се бројевима, нпр. PAM40, PAM250 и сл. PAM матрица нижег броја ће фаворизовати кратке поравнате секвен-

це високе сличности док ће PAM матрица вишег броја препознавати дуже али мање сличне локално поравнате секвенце (слика 1). Нпр. употребом PAM250 матрице, 20% аминокиселинске секвенце мора да буде идентично у поравнатим секвенцама да би се оне сматрале повезаним.

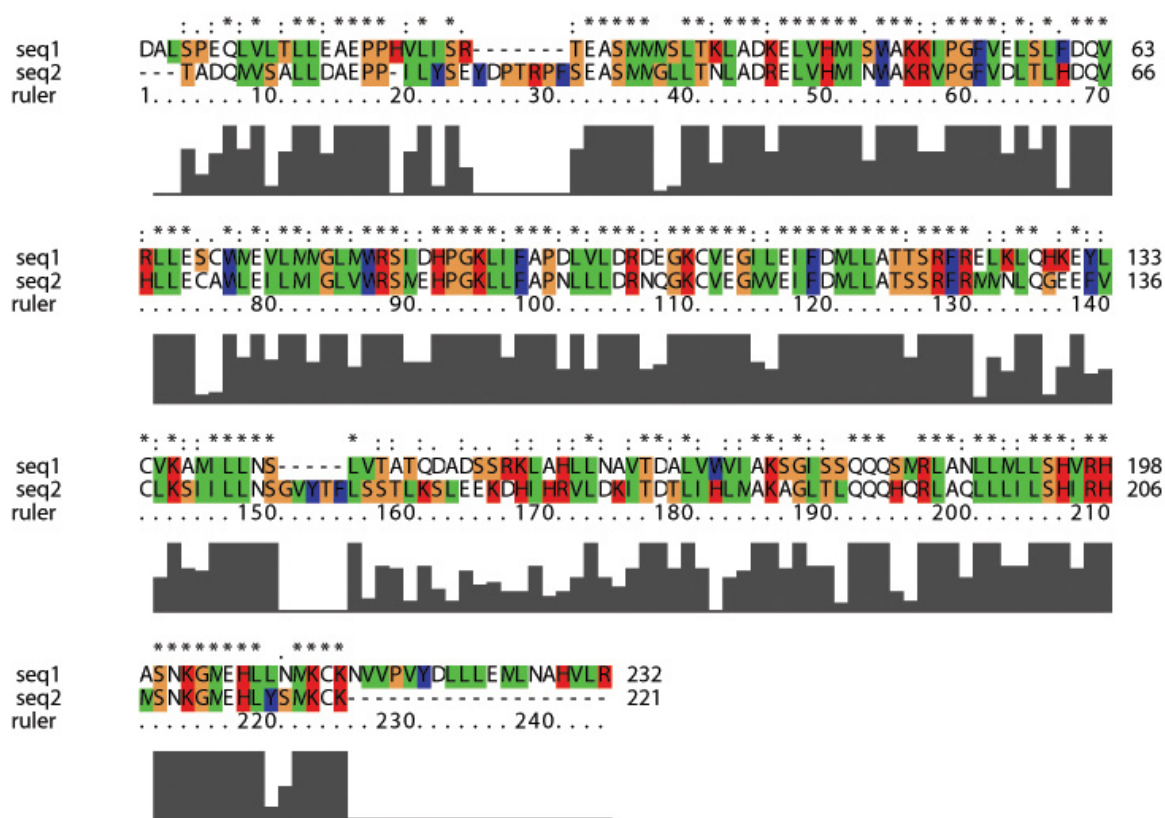
#### Идентификација структурно конзервираних (SCR) и структурно варијабилних региона (SVR)

Пошто су познате структуре поравнате, испитују се да би се пронашли структурно конзервирани региони помоћу којих је могуће моделовати структуру за те зоне. Варијабилни региони се могу разликовати у конформацији, али је и њих потребно уочити, јер је за њих потребно применити посебну тактику моделовања (слика 2).

SCR је теже идентификовати уз помоћ само једне структуре протеина за хомолого моделовање. Анализом хомолога се уочило да се SCR углавном налазе на елементима дефинисане секундарне структуре,  $\alpha$ -хеликсима и  $\beta$ -набраној структури као и местима везивања лигананда. Ови региони се углавном налазе на местима у протеину где би промена просторне организације ланца пептида довела до значајне промене у терцијарној структури протеина<sup>[2]</sup> и по правилу показују много већу хомологију секвенце од SVR. Такође, примећено је да секундарне структурне јединице,  $\alpha$ -хеликси и  $\beta$ -набране структуре, теже да у целој фамилији

<sup>1</sup>. Dayhoff, M.O., Barker, W.C., Hunt, L.T., *Meth. Enzymol.*, 91 (1983) 524.

<sup>2</sup>. Perutz, M.F., Kendrew, J.C., Watson, H.C., *J. Mol. Biol.*, 13 (1965) 669.



Слика 2. Поравнавање две протеинске секвенце програмом ClustalX.

Исте аминокиселине су означене са \*, парови који показују високу вредност у мутационим матрицама са : . Поједине групе аминокиселина означене су различитим бојама.

протеина заузимају исте или сличне релативне позиције. С друге стране, SVR се обично налазе на површини протеина и граде петље у којима главни ланци мењају правац.

Да би се препознали SCR, протеини се суперпонирају један у односу на други, обично методом најмањих квадрата. Проблем представља то што у почетку није познато који атоми се суперпонирају. Стога се у првој апроксимацији полази од преклапања  $C^{\alpha}$  атома<sup>[1]</sup>. Иницијална структура може даље бити пречишћена и само помоћу подударних тачака секундарних структурних елемената које се подударају. За аутоматско решавање овог проблема развијено је неколико метода.<sup>[2,3]</sup> Посебан случај моделовања је када су познати SCR протеина који служи као предлог. У овом случају потребно је лоцирати зоне на моделованој секвенци које најбоље одговарају овим регионима. Проблем настаје у томе што у SCR по дефиницији не сме да буде уметања и брисања, тако да стандардни Нидлман-Вуншов алгоритам не може да буде примењен. Због овога су развијене методе које сваки SCR обрађују засебно.<sup>[4]</sup>

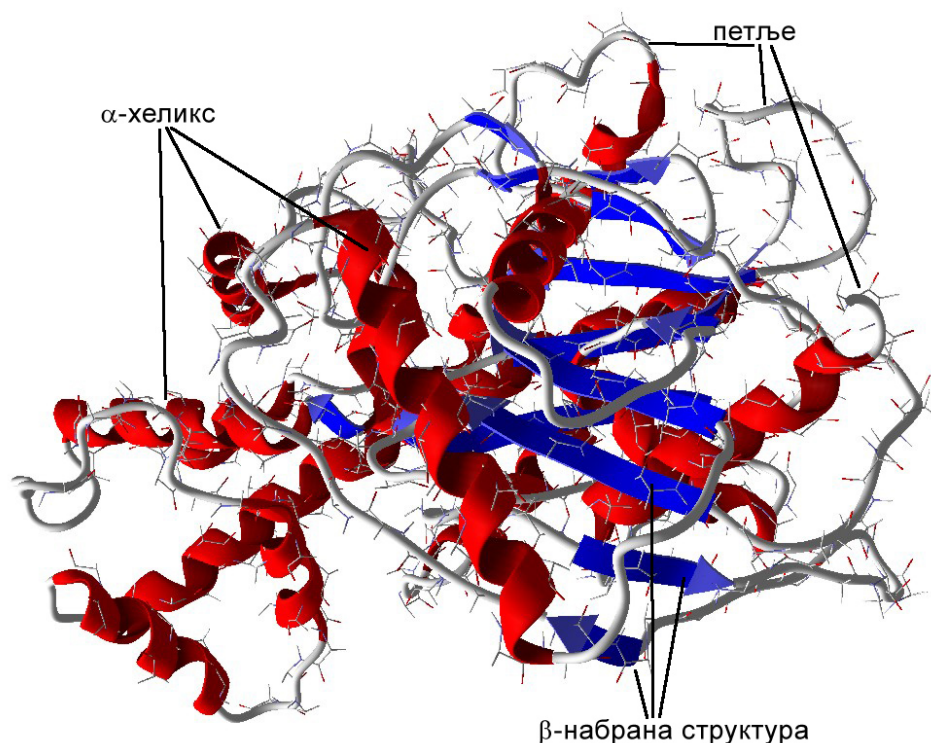
Када се утврди сличност између референтног и моделованог протеина, могуће је доделити координате

за SCR. За ово додељивање координата се као полазне тачке користе координате протеина познате структуре. У сегментима са идентичном структуром преносе се све координате аминокиселина, док се у деловима који се разликују преносе само координате основног ланца. Координате аминокиселинских остатака ће бити додате пошто се генерише комплетан основни ланац протеина (SCR и SVR).

Моделовање SVR је далеко већи проблем. Овде је углавном реч о петљама недефинисане структуре (слика 3), где се ситуација додатно искомпликовала због бројних уметања и брисања до којих овде обично долази. Развијено је неколико добро описаних и познатих метода за генерисање петљи.<sup>[3]</sup> Један од начина за генерисање координата SVR је додела координата помоћу сегмената исте дужине и аминокиселинског карактера у хомологим протеинима. Ако овај приступ није могућ, координате се могу генерисати помоћу координата пептида који се налазе у базама података а који се уклапају у просторни распоред модела или је могуће генерисати петље *de novo*.

Сви SVR су обично далеко од оптималне геометрије, стога се увек њихова геометрија пречишћава ми-

<sup>1</sup> Greer, J., *J. Mol. Biol.*, **153** (1981) 1027.  
<sup>2</sup> Vriend, G., Sander, C., *Prot. Struct. Func. Gen.*, **11** (1991) 52.  
<sup>3</sup> Matthews, B.W., Rossmann, M.G., *Methods Enzymol.*, **115** (1985) 397.  
<sup>4</sup> Šali, A., *Curr. Opin. Biotechnology*, **6** (1995) 437.



**Слика 3.** Схематски приказ протеина са назначеним структурним елементима. Протеин приказан на слици је ензим хумана глукокиназа (идентификациони број у PDB бази података 1V4S, ознака у ензимској номенклатури EC 2.7.1.1)

нимизацијом енергије, да би се избегле стерне сметње и релаксирале конформације петљи.

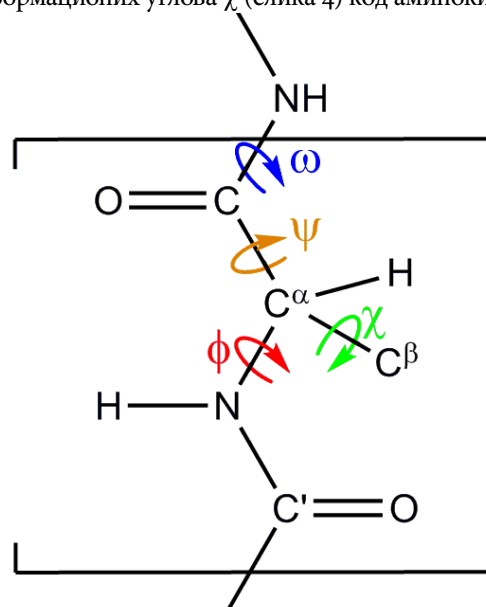
#### Моделовање бочних остатака

Следећи корак после генерисања основног ланца протеина је додавање бочних остатака. Ова операција је далеко сложенија од одређивања координата основног ланца протеина. Многи бочни остаци имају један или више степена слободне и стога у простору могу заузети многе енергетски допуштене конформације.

Сматра се да у хомологим протеинима идентични остаци заузимају сличне конформације. Када у поравнатим секвенцама протеина са високом хомологијом постоји случај мутације аминокиселине чији пар има високу вредност у Дејхоф-матрици (нпр. пар Ile-Val или Gln-Glu), претпоставља се да бочни остаци заузимају сличну оријентацију као и у предлошку протеина.<sup>[1]</sup>

Ситуација постаје компликованија ако парови аминокиселина у поравнатим секвенцама нису мутационо повезани. Када је бочни ланац аминокиселине у протеину који моделујемо дужи или структурно различит од одговарајућег у хомологом протеину, он мора бити постављен у такву конформацију која онемогућује неповољне интеракције са суседним бочним остатцима. Алтернативни метод за одабир ове конформације је коришћење конформације која одговара енергетском минимуму одговарајућег дипептида.

Много поузданија процедура је развијена на основу примењене статистичке правилности у распореду конформационих углова  $\chi$  (слика 4) код аминокисели-



**Слика 4.** Означавање торзионих углова у протеинима.

на у природним протеинима. На основу ове статистичке расподеле добијене су библиотеке ротамера бочних

<sup>1</sup> Feldmann, R.J., Bing, D.H., Potter, M., Mainhart, C., Furie, B., Furie, B.C., Caporale, L.H., *Ann. N.Y. Acad. Sci.*, 439 (1985) 12.

остатака.<sup>[1]</sup> Једна од најчешће примењиваних библиотека ротамера садржи 67 ротамера за 17 аминокиселина. Овакав приступ моделовању бочних остатака може донекле да занемари информацију о конформацији коју даје одговарајући бочни остатак у структури хомолог протеина који служи као предложак. Осим тога, бочни остаци се често налазе веома густо упаковани у протеин, тако да њихова конформација зависи и од околине, секундарне и терцијарне структуре протеина, па је потребно у одабир укључити и утицај локалног окружења бочних остатака.

Данас постоји више метода за предвиђање конформације и координата бочних ланаца у моделу. Сви ма њима је заједничко да координате основног ланца протеина држе непромењеним. Без обзира на све ове методе, некада је неопходно да положије бочних ланаца подесимо чак и ручно, као што је то у случајевима где се јавља нека интеракција која може утицати на конформацију - формирање јонског пара, унутрашња водонична веза и слично.

Без обзира на то који метод и стратегију дизајнирања модела протеина применимо, најчешће је потребно пречистити добијену структуру. Спојени SCR и SVR у првом моделу углавном показују приличан стерни напон, а неки бочни ланци могу се наћи у положајима који доводе до неповољних ван дер Валсових (van der Waals) контаката. Потпуна оптимизација свих бочних остатака би вероватно довела до губљења важних водоничних веза и унела конформационе промене у структурно сачуване регионе. Да би се уклониле ове стерне сметње, потребно је на моделу извести конформациону претрагу да би се идентификовали бочни низови са неповољним ван дер Валсовим интеракцијама. Да би се пронашла конформациона стања са минимумима енергије (локални минимуми), најуобичајеније је да се модел у овој фази подвргне контролисаној енергетској минимизацији и/или молекулској динамици, уз пажљиво постављена ограничења која ће сачувати основну идеју и структуру модела.

## ОПТИМИЗАЦИЈА МОДЕЛА

### Методе поља сила за моделовање протеина

Без обзира да ли је модел протеина изведен из кристалне структуре или добијен компаративним моделовањем, он захтева даље пречишћавање и обраду. При генерисању модела протеина конформације петљи и бочних остатака углавном не одговарају енергетски разумним структурама. Такође, кристалне структуре је не-

опходно донекле релаксирати да би се уклонио напон који се јавља због сила кристалног паковања, или да се отклони блиски контакт водоникових атома или аминокиселинских остатака који су додати кристалним координатама после утврђивања структуре.

Пошто се протеини састоје од стотина или хиљада атома, једини разуман метод за рачунарску обраду овако великих система је примена молекулско-механичких израчунавања. Треба имати на уму да се параметри поља сила (*force field*) за примену у моделовању протеина разликују од оних које се примењују за мале молекуле. Осим неких специфичних параметризација за протеине и ДНК, ове методе често садрже додатна поједностављења. У неким пољима сила неполярни водоникови атоми нису експлицитно представљени већ су укључени у опис тежих атома за које су везани. Насупрот овоме, поларни атоми водоника, који могу бити укључени у водонично везивање се третирају експлицитно. Овакав приступ се назива модел уједињеног атома. У AMBER пољу сила<sup>[2,3]</sup> могу се применити и модел уједињеног атома и приказ свих атома, док GROMOS поље сила<sup>[4]</sup> нуди само модел уједињеног атома. Могу се увести и додатна поједностављења као на пример ограничење полупречника интеракција атома,<sup>[5]</sup> да би се смањило временски захтевно израчунавање невезивних интеракција између атома удаљених више од дефинисане граничне вредности.

Постоје друге, бројне модификације поља сила која се користе за калкулације везане за протеине<sup>[6]</sup>. Треба стално имати на уму да свако поједностављење доводи до смањења тачности. Одлука о томе које поље сила треба применити зависи од проблема који се проучава. Правило је да се увек користи најпрецизније поље сила које је применљиво на целокупно изучавање, јер треба избегавати употребу различитих метода у разним фазама молекулског моделовања. Нека од најчешће коришћених поља сила су AMBER,<sup>[2,22]</sup> CVFF,<sup>[7]</sup> CHARM<sup>[8]</sup> и GROMOS.<sup>[23]</sup>

### Оптимизација геометрије

Код протеина добијених из кристалних координата, понекад је потребно додатно оптимизовати геометрију протеина, да би атоми из кристалних координата подесили свој положај тако да се позиционирају (релаксирају) на нове, блиске положаје који су енергетски повољнији и који боље одговарају његовим нативним. Алгоритми који се користе за процедуру минимизације код протеина су исти као и они коришћени код малих молекула. Алгоритми за минимизацију енергије

1. Ponder, J., Richards, F.M., *J. Mol. Biol.*, **193** (1987).

2. Weiner, S.J., Kollman, P.A., Case, D.A., Chandra Singh, U., Ghio, C., Alagona, G., Profeta, S., Weiner, P., *J. Am. Chem. Soc.*, **106** (1984) 765.

3. Weiner, S.J., Kollman, P.A., Nguyen, D.T., Case, D.A., *J. Comp. Chem.*, **7** (1986) 230.

4. van Gunsteren, Berendsen, H.J.C., **Molecular Dynamics and Protein Structure**, Ed. Hermans, J., Polycrystal Books Service, Western Springs, **1985**, ISBN 9996204634

5. Brooks, C.L. III, Pettitt, B.M., Karplus, M., *J. Chem. Phys.*, **83** (1985) 5897.

6. van Gunsteren, W.F., Bakowies, D., Baron, R., Chandrasekhar, I., Christen, M., Daura, X., Gee, P., Geerke, D.P., Glättli, A., Hünenberger, P.H., Kastenholz, M.A., Oostenbrink, C., Schenk, M., Trzesniak, D., van der Vegt, N.F.A., Yu, H.B., *Ang. Chem. Int. Ed.*, **45** (2006) 4064.

7. Dauber-Osguthorpe, P., Roberts, V.A., Osguthorpe, D.J., Wolff, J., Genest, M., Hagler, A.T., *Prot. Struct. Funct. Gen.*, **4** (1988) 31.

8. Brooks, B.R., Brucoleri, R.E., Olafson, B.D., States, D.J., Swaminathan, S., Karplus, M., *J. Comp. Chem.*, **4** (1983) 187.

који се примењују за оптимизацију геометрије углавном налазе локални минимум потенцијалне енергије који је најближи почетним координатама. Код кристалних структура високе резолуције тај процес ће обично дати једну енергетски повољну конформацију. Међутим, кристалне координате некада могу да садрже неколико неповољних атомских интеракција. Ови поремећени атомски положаји могу да резултују велике почетне енергије, које могу да потичу од неповољних стерних интеракција или торзионог напона, што ће при процесу минимизације довести до вештачких покрета далеко од првобитне структуре. Да би се избегли овако велико одступање, углавном се релаксација протеина врши постепено.

Далеко темељитије решење је примена ограничавајућих сила на све тешке атоме у кристалној структури у првој фази минимизације. Ограничавајућа (*tethering*) константа је сила која се користи да би се атомске координате учврстиле на предефинисаним позицијама. У првом кораку, са ограничавањем покрета тешких атома, само атоми водоника и растварача се могу померати да би се минимизовала укупна потенцијална енергија. Потом се ограничавају само добро дефинисани атоми главног ланца. Сада бочни остаци могу да се померају и подешавају своју оријентацију. На крају се уклањају рестрикције и у последњем кораку се минимизација изводи на целој релаксираној конформацији. У прва два дела подесан алгоритам за минимизацију је тзв. *steepest descent*, док је за последњи корак погодан метод коњугованог градијента.

Код модела добијених хомологим моделовањем, конформације петљи и бочних остатака захтевају додатна пречишћавања. Потребно је пажљиво анализирати њихове конформације и проверити потенцијалне енергије других могућих нискоенергетских конформера. За ове потребе је корисно употребити молекулску динамику, а као полазни молекул може се употребити релаксирана геометрија протеина добијена процедуром минимизације.

Додатни проблем при генерисању што тачнијих 3Д структура представља и то што конформациона мобилност протеина, нарочито на њиховој површини и у петљама, веома зависи од окружења. Тачност резултата минимизација и молекулске динамике расте укључењем молекула растварача у симулацију. На жалост, ово представља још увек нерешен проблем. Један од начина је подражавање присуства растварача употребом тзв. метода имплицитне солватације. Код ових метода (понека познатих и као методе континуалне солватације) се растварач не посматра као систем састављен

од многобројних молекула, већ као континуални медијум. Значајно побољшање у квалитету модела, наравно на штрб рачунарског времена, нам могу донети израчунавања која се изводе тако што се модел протеина посматра у окружењу молекула растварача, обично воде (метод експлицитне солватације). Важно је направити разлику између структурне воде и молекула воде као растварача. Структурни молекули воде су они који су значајни за функцију протеина и они грађењем водоничних веза са аминокиселинама могу утицати на конформацију чак и у самом центру протеина. Развој рачунара у последњих десетак година је донео значајни помак у симулацијама са употребом експлицитно дефинисаних молекула растварача, тако да је све више и више симулација које се изводе са присутним молекулима растварача.<sup>[1,2]</sup>

### Валидација модела

Квалитет 3Д структуре моделованог протеина највише зависи од тачности кристалне структуре која је узета као предлог за моделовање.<sup>[3]</sup> Модел изведен из кристалне структуре никако не може бити тачнији од саме структуре. Али, протеинске структуре добијене рендгенском дифракцијом могу имати грешке, како експерименталне тако и грешке у интерпретацији резултата.<sup>[30,4,5]</sup>

Када је модел протеина направљен употребом хомологог моделовања и његова геометрија оптимизирана молекулском механиком или молекулском динамиком, важно је да се провери његов квалитет и поузданост. Овај задатак је веома комплексан, јер квалитет хомолого моделованих протеина зависи од великог броја особина. Три основна аспекта су стереохемијска тачност, квалитет паковања и позданост просторног савијања.

### Стереохемијска тачност

Да би се проверила стереохемијска тачност модела потребно је проверити параметре као што су дужине и углови веза, торзиони углови, хиралност аминокиселина, планарност пептидне везе, планарност бочних низова. Пошто би мануална инспекција свих стереохемијских параметара била преобиман задатак, развијени су програми који аутоматски проверавају стереохемијске особине, као што су PROCHECK,<sup>[6]</sup> WHAT-CHECK<sup>[7]</sup> или VADAR.<sup>[8]</sup> Ове провере се заснивају на чињеници да стереохемијски параметри код протеина показују одређену правилност у дистрибуцији и углавном су константни код свих познатих протеина. Они представљају осетљиву меру квалитета модела протеи-

<sup>1</sup> Schlegel, B., Sippl, W., Höltje, H.-D., *J. Mol. Med.*, **12** (2005) 49.

<sup>2</sup> Jöhren, K., Höltje, H.-D., *Arch. Pharm.*, **338** (2005) 260.

<sup>3</sup> Bränden, C.J., Jones, T.A., *Nature*, **343** (1990) 687.

<sup>4</sup> Jones, T.A., Zou, J.Y., Cowan, S.W., *Acta Crystallographica*, **A47** (1991) 110.

<sup>5</sup> Engh, R.A., Huber, R., *Acta Crystallographica*, **A47** (1991) 392.

<sup>6</sup> Lakowski, R.A., MacArthur, M.W., Moss, D.S., Thornton, J.M., *J. Appl. Cryst.*, **26** (1993) 283.

<sup>7</sup> Hoof, R.W.W., Vriend, F., Sander, C., Abola, E.E., *Nature*, **381** (1996) 272.

<sup>8</sup> Willard, L., Ranjan, A., Zhang, H., Monzavi, H., Boyko, R.F., Sykes, B.D., Wishart D.S., *Nuc. Acids Res.*, **31** (2003) 3316.

на, па их треба пажљиво проверити, јер су у неким случајевима извесна одступања дозвољена.

#### Квалијетет паковања

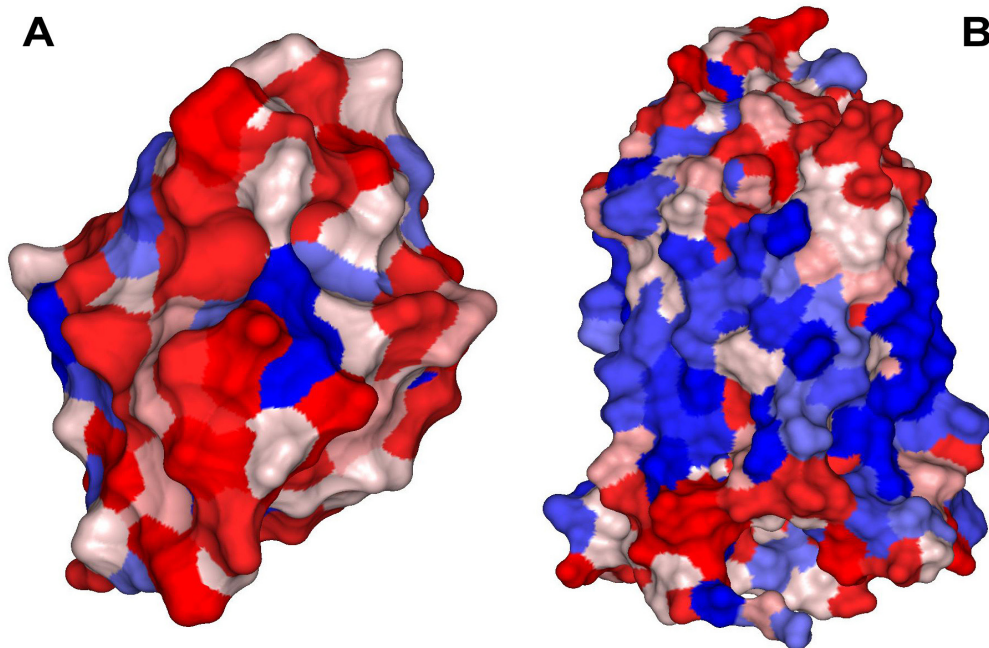
Претпоставља се да специфичне интеракције при паковању унутар протеина имају значајну функцију у структурној специфичности протеина.<sup>[1]</sup> Унутрашње паковање, нарочито код глобуларних протеина, веома доприноси стабилности укупне конформације, стога се квалитет паковања модела протеина може користити да би се проценила његова поузданост.

Први корак у тој провери је провера ван дер Валсових контаката у моделу. Проверавају се сва интерактомска растојања, да би се видело да ли леже у оним областима које су уочене у добро дефинисаним кристалним структурама. Овај задатак се углавном поверава програмима као што су већ поменути PROCHECK и/или WHATCHECK.

Потом је неопходно проверити да ли је модел задржао секундарне структурне елементе који одговарају предлошку, пошто су они најконзервиранији региони у високо хомологим протеинима. Овај задатак могу успешно обавити програми као што су DSSP<sup>[2]</sup> или STRIDE.<sup>[3]</sup>

Постоје различити методи, који углавном користе велики број информација изведених из кристалних структура, којима се може утврдити квалитет паковања у моделу протеина.<sup>[4,5,6]</sup>

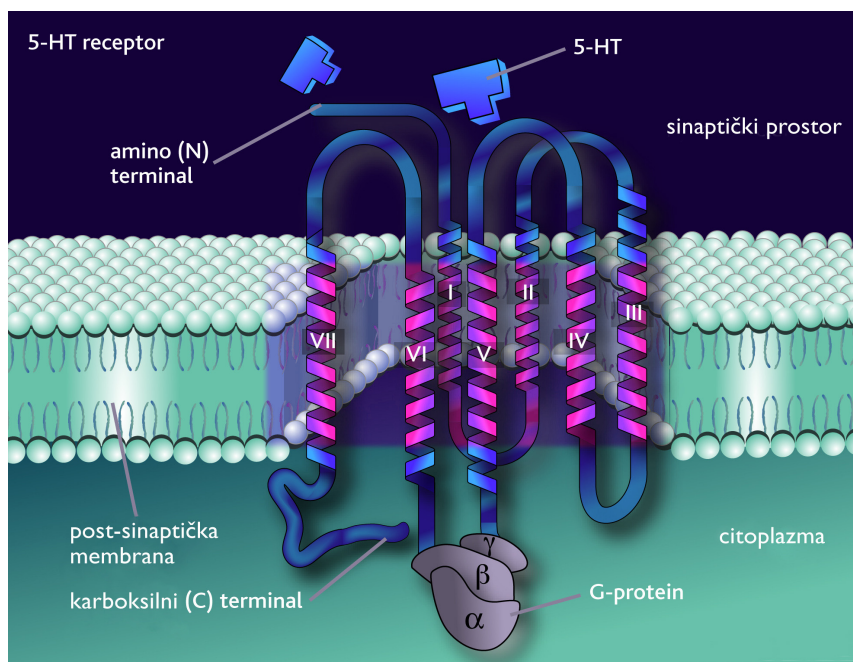
Утврђено је да је расподела поларних и неполарних аминокиселинских остатака између унутрашњости и површине протеина генерални принцип изградње глобуларних протеина. Укратко, нађено је да се глобуларни протеини у принципу састоје од хидрофобне унутрашњости окружене хидрофилном спољашњом површином која интерагује са молекулима растварача. Унутрашњост протеина је густо пакована, без великих празних простора и хидрофобна је (слика 5). Неполарни аминокиселински остаци, Val, Leu, Ile, Phe, Ala и Gly, доминирају у унутрашњости протеина и чине 63% унутрашњих аминокиселина. Јонизовани парови киселих и базних група се ретко јављају у унутрашњости. На површини протеина, лако доступне молекулима растварача, се налазе наелектрисане и поларне групе, Asp, Glu, Lys, Arg. Оне чине око 27% спољашњих аминокиселина, а само 4% унутрашњих. Мембрански протеини се разликују по паковању од глобуларних првенствено по томе што имају изузетно неполарне површине, које су у контакту са хидрофобним средњим слојем фосфолипидног дела мембрана (слика 6).



**Слика 5.** Хидрофобне (плаво) и хидрофилне (црвено) аминокиселине на површини глобуларног (А) и трансмембранског (В) протеина. Приметна је већа присутност хидрофобних аминокиселина у средњој зони трансмембранског протеина. Глобуларни протеин је ланац А лизозима из псећег млека (PDB код 1EL1), трансмембрански протеин је говеђи родопсин (PDB код 1F88). Протеини нису приказани у размери.

1. Chothia, C., *Ann. Rev. Biochem.*, **53** (1984) 537.
2. Kabsch, W., Sander, C., *Biopolymers*, **22** (1983) 2577.
3. Frishman, D., Argos, P., *Prot. Struct. Funct. Gen.*, **23** (1995) 566.
4. Hoof, R.W., Sander, C., Vriend, G., *Comp. App. Biosci.*, **13** (1997) 425.
5. Laskowski, R.A., Thornton, J.M., Hubert, C., Singh, J., *J. Mol. Biol.*, **259** (1996) 175.
6. Hunt, N.G., Gregoret, L.M., Cohen, F.E., *J. Mol. Biol.*, **241** (1994) 214.





Слика 6. Схематски приказ положаја трансмембранског рецептора у мембрани. Истакнута је зона у којој се на површини трансмембранских хеликса налазе хидрофобне киселине које су у контакту са липидима мембране.

### Поузданост и просторно савијање

Протеини који имају хомологе аминокиселинске секвенце би требало да имају и исто савијање у простору. Стога би укупна 3Д структура протеина требало да одговара структури његовог предлошка. Истоветност конформација би требала да се покаже нарочито у структурно конзервираним регионима. Сличност 3Д структура се мери одређујући одступања од атомских координата  $C^{\alpha}$ -атома или атома главног ланца модела и предлошка. Мера одступања је дата као RMSD (*root mean square deviation*). Обично се количина структурне дивергенције два хомолога протеина мери суперпозицијом конзервираних региона и израчунавањем RMSD. Мања вредност RMSD указује на већу сличност координата. За одређивање поузданости просторног савијања постоји неколико приступа.<sup>[1,2,3]</sup>

Извесно је да ће моделовање структуре протеина још дуго остати важна техника која ће помагати не само да боље и брже схватимо улогу и функцију протеина у метаболизму, већ и да ефикасније дизајнирамо нове фармаколошки активне препарате и модификујемо њихову структуру и дејство на протеине. Наравно, модели никада нису савршени, али као што је рекао Хенри А. Бент (Henry A. Bent): „Модел мора бити погрешан у некој мери, јер би иначе представљао саму ствар. Трик је у томе да препознамо где је исправан.“<sup>[4]</sup>

- <sup>1</sup> Fischer, D., Eisenberg, D., *Curr. Op. Struct. Biol.*, 9 (1999) 208.
- <sup>2</sup> PROFILES-3D, *User Guide*, Accelrys, San Diego, <http://www.accelrys.com>.
- <sup>3</sup> Watson, J.D., Laskowski, R.A., Thorthon, J.M., *Curr. Op. Struct. Biol.*, 15 (2005) 275.
- <sup>4</sup> Bent, H.A., *J. Chem. Ed.* 61 (1984) 774.

шан у некој мери, јер би иначе представљао саму ствар. Трик је у томе да препознамо где је исправан.”<sup>[4]</sup>

### Abstract

#### HOMOLOGY MODELING of PROTEIN STRUCTURE

Mario Zlatović, *Faculty of Chemistry*,  
mario@chem.bg.ac.yu

Sladana Kostić-Rajačić, *ITHM Department of Chemistry*,  
University of Belgrade

Vladimir Šukalović, *ITHM Department of Chemistry*,  
University of Belgrade

Knowledge of three-dimensional protein structure is of the paramount importance for determination of receptor and enzyme action mechanism in interaction with other molecules. In spite of growing number of solved 3D protein structures, number of determined protein sequences exceeds by far the number of defined spatial protein structures. Besides, sequence databases grow more rapidly than databases of 3D protein structures. When crystal protein structures or data obtained by NMR are not available, we have to reach for next „best“ possible solution - modeling of the protein spatial structure. One of the aspects that can be applied to obtain usable protein model is homology modeling. This method is based on the design of the three-dimensional structure of the unknown protein based on layout of the homolog protein of known 3D structure. To get a big picture of this method, in this text, we explained few aspects of homology modeling.