

მალაქია ფირცხალავა

ბიოინფორმატიკის საფუძვლები

მონაცემთა ბაზები, თანამიმდევრობათა ანალიზი,
სტრუქტურული ბიოინფორმატიკა

თბილისი

2008

წიგნი წარმოადგენს სახელმძღვანელოს ბიოინფორმატიკაში. მას საფუძვლად დაედო ლექციათა კურსი, რომელიც იკითხებოდა თბილისის ივანე ჯავახიშვილის სახელობის სახელმწიფო უნივერსიტეტის ზუსტ და საბუნებისმეტყველო მეცნიერებათა ფაკულტეტის მაგისტრატურის განყოფილების სტუდენტებისათვის. ბიოინფორმატიკა ინტერდისციპლინარული დარგია და წიგნისადმი ინტერესის არსებობა შეიძლება ვივარაუდოთ გამოყენებითი მათემატიკისა და ინფორმატიკის, ბიოფიზიკისა და კოდენსირებული გარემოს ფიზიკის, ორგანული ქიმიის, პოლიმერების ქიმიისა, ქიმიურ ნაერთთა პროექტირებისა და სინთეზის, ფარმაკოლოგიის, ანუ საბუნებისმეტყველო მეცნიერებათა და მედიცინის სრულიად განსხვავებული განყოფილებების სტუდენტებისათვის. შესაბამისად წინამდებარე სახელმძღვანელო შეიძლება რეკომენდირებულ იქნეს სამედიცინო, ბიოლოგიის, მათემატიკის, ფიზიკისა და ქიმიის განყოფილებების სტუდენტებისა და ახალგაზრდა მკვლევარებისათვის.

© მალაქია ფირცხალავა, 2008

ISBN 978-9941-0-0840-5

ემღვნება დედაჩემის, ოლიკო იოსელიანის ხსოვნას

შინაარსი

შინაარსი	5
ავტორისაგან	7
თავი 1. ბიოინფორმატიკა და სტრუქტურული	11
ბიოინფორმატიკა. არსებობისა და განვითარების საფუძვლები	11
1. ბიოინფორმატიკის არსებობისა და განვითარების.....	11
საფუძვლები	11
2. ბიოინფორმატიკის როლი გენომური პროექტების რეალიზების პროცესში.....	19
3. ბიოინფორმატიკა თანამედროვე ბიოლოგიისა და მედიცინის განვითარების ჭრილში.	25
თავი 2. ბიომაკრომოლეკულათა სივრცული მოწყობა.....	35
4. ნუკლეინის მჟავათა სტრუქტურული ორგანიზაციის ზოგადი პრინციპები.....	35
5. ცილათა სტრუქტურული ორგანიზაციის ზოგადი პრინციპები	49
6. პოლიპეპტიდის ცილად დახვევის გზები და მექანიზმები.	69
თავი 3. მონაცემთა ბაზები.....	83
7. ნუკლეოტიდურ თანამიმდევრობათა შესახებ მონაცემთა ბაზები.	83
8. ამინმჟავურ თანამიმდევრობათა შესახებ მონაცემთა ბაზა UniProt.....	95
9. ბიომაკრომოლეკულათა სივრცული სტრუქტურების შესახებ მონაცემთა ბაზა - PDB	106
თავი 4. თანამიმდევრობათა შედარება	149
10.ორი თანამიმდევრობის შედარება. გლობალური.....	149
ურთიერთსწორება.....	149
11. ორი თანამიმდევრობის შედარება. ნახევრადგლობალური და ლოკალური ურთიერთსწორება.....	159

12. ქულათა დარიცხვის სქემები. ჩანაცვლებათა (მსგავსების) მატრიცები.	169
13. მსაგავსება და დისტანცია. ფილოგენეტიკური ხე	183
14. მონაცემთა ბაზების სწრაფი ანალიზი.....	202
თავი 5. სტრუქტურათა შედარება და კლასიფიკაცია	213
15. ცილებისა და რნმ-ების სტრუქტურული ურთიერთსწორება.....	213
16. ცილათა სტრუქტურული კლასიფიკაცია	221
თავი 6. სივრცული სტრუქტურისა და ფუნქციის წინასწარმეტყველება.....	229
17. მეორეული სტრუქტურის წინასწარმეტყველება	229
18. ცილის სივრცული მოწყობის წინასწარმეტყველება	245
19. მრავალწევრა ერთდროული ურთიერთსწორება.....	261
თავი 7. მოლეკულათშორისი ურთიერთქმედებების წინასწარმეტყველება.....	271
20. ცილა - ცილა ურთიერთქმედების წინასწარმეტყველება.....	271
21. დოკინგი. ლიგანდთა (წამალთა) პროექტირება.....	284
თავი 8. სტრუქტურულ ბიოლოგიური კვლევების დაგეგმარების სტრატეგია	297
22 . სტრუქტურული გენომიკა	297
საგნობრივი სარჩევი	306

ავტორისაგან

თქვენს წინაშეა პირველი ქართულენოვანი სახელმძღვანელო ბიოინფორმატიკაში. ამ სახელმძღვანელოს საფუძვლად დაედო ლექციათა კურსი, რომელიც იკითხებოდა თბილისის ივანე ჯავახიშვილის სახელობის სახელმწიფო უნივერსიტეტის ზუსტ და საბუნებისმეტყველო მეცნიერებათა ფაკულტეტის მაგისტრატურის განყოფილების სტუდენტებისათვის. ბიოინფორმატიკა ინტერდისციპლინარული დარგია. მისი მიზნები თანამედროვე ბიოლოგიისა და ბიომედიცინის პრობლემების გადაწყვეტით შემოიფარგლება. ამისათვის იგი იყენებს მათემატიკაში, კომპიუტერულ მეცნიერებებში, ფიზიკაში, ქიმიაში და ბუნებრივია, ბიოლოგიაში დაგროვილ ცოდნასა და საშუალებებს. ბიოინფორმატიკაში დაგროვილი ცოდნის გაცნობა, მის ფარგლებში შემუშავებულ მდგომარეობებზე წარმოდგენის შექმნა და განვითარების პერსპექტივებზე ინფორმაციის მიღება უდავოდ დიდ სარგებელს მოუტანს თანამედროვე ბიოლოგიისა და ბიომედიცინის საკითხებით დაინტერესებულ ნებისმიერ სტუდენტსა და ახალგაზრდა მკვლევარს. ამასთან ბიოინფორმატიკისადმი ინტერესის არსებობა შეიძლება ვივარაუდოთ გამოყენებითი მათემატიკისა და ინფორმატიკის, ბიოფიზიკისა და კოდენსირებული გარემოს ფიზიკის, ორგანული ქიმიის, პოლიმერების ქიმიისა, ქიმიურ ნაერთთა პროექტირებისა და სინთეზის, ფარმაცოლოგიის, ანუ საბუნებისმეტყველო მეცნიერებათა და მედიცინის სრულიად განსხვავებული განყოფილებების სტუდენტებისათვის. შესაბამისად წინამდებარე სახელმძღვანელო შეიძლება რეკომენდირებულ იქნეს სამედიცინო, ბიოლოგიის, მათემატიკის, ფიზიკისა და ქიმიის განყოფილებების სტუდენტებისა და ახალგაზრდა მკვლევარებისათვის.

თანამედროვე ბიოლოგია ადამიანის მოღვაწეობის ის სფეროა, რომელიც გამოირჩევა განსაკუთრებით დიდი ინტენსივობის ინფორმაციულ ნაკადთა არსებობით. ბიოინფორმატიკის ჩამოყალიბებისა და ინტენსიურად განვითარების საწინდარს სწორედ ამ ინფორმაციების მიღების, შენახვისა და ანალიზის აუცილებლობა წარმოადგენს. წიგნის პირველი თავი მკითხველს წარმოადგენს უქმნის იმაზე, თუ რა როლი შეიძლება ითამაშოს ბიოინფორმატიკულმა საშუალებებმა ინფორმაციების მიღების, შენახვისა და ანალიზის საქმეში და ბიოინფორმატიკის განვითარების გარკვეული დონის მიღწევის შემთხვევაში, როგორი პერსპექტივა იწვლება ადამიანის მოღვაწეობის მნიშვნელოვანი სფეროების: მედიცინის, სოფლის მეურნეობის, ეკოლოგიური უსაფრთხოების, თვისობრივად ახალ სიმაღლეებზე აყვანის თვალსაზრისით.

სტრუქტურული ბიოინფორმატიკა ბიოინფორმატიკის განუყოფელი და მნიშვნელოვანი ნაწილია. შეიძლება ითქვას, რომ პირველი კომპიუტერული

მიდგომები ბიოლოგიაში სწორედ სტრუქტურული ბიოლოგიის ამოცანების გადაწყვეტის მიზნით შეიქმნა. წინამდებარე სახელმძღვანელოს დიდი ნაწილი ეხება სტრუქტურული ბიოინფორმატიკის საკითხების მიმოხილვას და ამდენად, მოცემული წიგნის სახელი „სტრუქტურული ბიოინფორმატიკის საფუძველში“ შეიძლება ყოფილიყო. ამ სპეციფიკიდან გამომდინარე, მეორე თავი ითვალისწინებს სტრუქტურული ბიოლოგიის საფუძვლებთან, მის ტერმინოლოგიასა და ცნებებთან გაცნობას. ამით, ჩვენი აზრით, სტრუქტურულ ბიოლოგიაში გაუთვითნებობიერებელ მკითხველს ვუადვილებთ სტრუქტურულ ბიოინფორმატიკასთან დაკავშირებული თავების გაცნობას. შესაბამისად, სტრუქტურული ბიოლოგიის ცოდნის მქონე მკითხველს შეუძლია წიგნის გაცნობისას გამოტოვოს მეორე თავი.

მესამე თავი ეხება ინფორმაციის შენახვის საკითხებს. კერძოდ, განხილულია ამინმჟავურ და ნუკლეოტიდურ თანამიმდევრობათა შესახებ მონაცემთა ძირითადი ბაზები. ასევე ცილების სივრცული მოწყობების შესახებ მონაცემთა ძირითადი ბაზა.

მეოთხე და მეხუთე თავში მიმოხილულია მონაცემების ანალიზის საშუალებები. მეოთხე თავი შეეხება თანამიმდევრობათა შედარებისა და ევოლუციური ხის აგების საშუალებებს, მეხუთე კი, სტრუქტურათა კლასიფიკატორებსა და ცილათა სტრუქტურების შედარების მეთოდებს.

დღეს ჩვენ ვცხოვრობთ სტრუქტურული გენომიკის ერაში, როცა დღის წესრიგში მთელი სიმწვავეით დგას ამოცანა - თითოეული გენი უზრუნველყოთ ინფორმაციით მისი ექსპრესიის პროდუქტის სივრცული მოწყობისა და ფუნქციის შესახებ. რადგან სტრუქტურული გენომიკის პრობლემების გადასაწყვეტად გამოყენებული ექსპერიმენტული მეთოდები ძვირია და დიდ დროს მოითხოვს, მათ გადასაწყვეტად იაფი კომპიუტერული მიდგომების ინტენსიურად ჩართვის აუცილებლობა გაჩნდა. შესაბამისად, შემუშავებულ იქნა სივრცული სტრუქტურის წინასწარმეტყველების მეთოდების მთელი რიგი. წიგნის მეექვსე თავი სწორედ ცილის სტრუქტურული მოწყობისა და ფუნქციის წინასწარმეტყველების მეთოდებს ეძღვნება.

ბიოლოგიური სისტემების ფუნქციონირების მექანიზმებში გარკვევა და ამ მექანიზმებზე ზემოქმედების საშუალებების შემუშავება თანამედროვე ბიომედიცინის უმთავრეს ამოცანას წარმოადგენს. ამ ამოცანის წარმატებით გადაწყვეტა დიდწილად დაამოკიდებული მოლეკულათმორისი ურთიერთქმედებათა ქსელების, ამ ურთიერთქმედებათა თავისებურებების შესახებ ამომწურავი ინფორმაციის არსებობაზე. მეშვიდე თავში მკითხველი გაეცნობა ცილა - ცილა და ცილა - ლიგანდის ტიპის ურთიერთქმედებათა შემსწავლელ კომპიუტერულ მიდგომებსა და შესაბამის პროგრამულ პაკეტებს. აქვე მკითხველი წარმოადგენას შეიქმნის

ბიოლოგიურად აქტიურ ნაერთთა (წამალთა) პროექტირების პროცესებსა და ამ პროცესებში გამოყენებულ კომპიუტერულ მიდგომებზე.

ბოლო, მერვე თავში მიმოხილულია თანამედროვე ბიომედიცინის ერთ-ერთი მნიშვნელოვანი მიმართულების, სტრუქტურული გენომიკის განვითარების გზები და პერსპექტივები.

ავტორი დიდ მადლობას უხდის: მათა გრიგოლავეს, თენგიზ სამსონიძეს და თენგიზ ზაალიშვილს ხელნაწერის დახვეწის პროცესში გაწეული დახმარებისათვის.

მალაქია ფირცხალავა

თავი 1

ბიოინფორმატიკა და სტრუქტურული ბიოინფორმატიკა. არსებობისა და განვითარების საფუძვლები

1. ბიოინფორმატიკის არსებობისა და განვითარების საფუძვლები

ჩვენ ვცხოვრობთ მძლავრი ინფორმაციული ნაკადების ეპოქაში. ამ ინფორმაციების შენახვა და გადამუშავება ახალი ცოდნის მიღების მიზნით, კაცობრიობისათვის ერთ-ერთ უმნიშვნელოვანეს ამოცანას წარმოადგენს.

თანამედროვე ბიოლოგია ადამიანის მოღვაწეობის ის სფეროა, რომელიც გამოიჩინა განსაკუთრებით დიდი ინტენსივობის ინფორმაციულ ნაკადთა არსებობით. ეს აიხსნება იმ უდიდესი როლით, რომელიც აკისრია თანამედროვე ბიოლოგიას ნებისმიერი ქვეყნის სრულფასოვანი განვითარების საქმეში. სოფლის მეურნეობის (მათ შორის კვების მრეწველობის), მედიცინის (მათ შორის ფარმაცოლოგიის), ეკოლოგიური და სხვა მნიშვნელოვანი პრობლემების გადაწყვეტა თანამედროვე ტექნოლოგიების გამოყენებით, ნებისმიერი სახელმწიფოს უმთავრესი ამოცანებია. თუ თვალს მივადევნებთ თანამედროვე ბიოლოგიური კვლევების განვითარების გზას, ადვილად შევამჩნევთ ამ კვლევათა გაიაფების ტენდენციას. ასე მაგალითად, ჯერ კიდევ ორიოდ ათეული წლის წინ დაგვემილი, “ადამიანის გენომის” სახელწოდებით ცნობილი პროექტის რეალიზაცია დაკავშირებული იყო კოლოსალურ დანახარჯებთან. ასეთივე მასშტაბისა

და ბუნების პროექტები, რომლებშიც ინტენსიურად არიან ჩართულნი სხვადასხვა ქვეყნის მკვლევარები, მიმდინარეობს დღესაც და იგეგმება მომავლისთვისაც. დღეისათვის უკვე შეიძლება ითქვას, რომ მათზე მოთხოვნილი დანახარჯები გაცილებით დაბალია. ექსპერიმენტულ მეთოდთა დახვეწისა და კომპიუტერულ მიდგომათა წილის გაზრდასთან ერთად მოსალოდნელია ასეთ პროექტთა თვითღირებულების შემდგომი მნიშვნელოვანი შემცირება და ეს უახლოესი მომავლის პერსპექტივად მოიაზრება.

ბიოლოგიურ პრობლემათა კვლევების მიზანია შეიქმნას წინაპირობა ცოცხალი უჯრედის, თუ მრავალუჯრედიანი ორგანიზმის ნორმალური ფუნქციონირებისთვის საჭირო ბიოლოგიური პროცესების საფუძველში მყოფი მოლეკულური მექანიზმების სრულად წარმოჩენისათვის, ანომალიათა მიზეზების დადგენისათვის. სწორედ ამ მიზნით გაიშიფრა სხვადასხვა ორგანიზმთა გენომების ნუკლეოტიდური თანამიმდევრობები. იქმნება ამ იმფორმაციის შენახვისა და გადამუშავების უზრუნველყოფი, კომპიუტერულ მეცნიერებათა თანამედროვე მიღწევებზე დაფუძნებული ახალი საშუალებები. ეს პროცესი ალბათ კიდევ დიდხანს გაგრძელდება. სწორედ ამან განაპირობა ახალი ინტერდისციპლინარული დარგის, ბიოინფორმატიკის ჩამოყალიბება. სულ რაღაც ათიოდ წელიწადში ამ დარგმა მიაღწია განვითარების მაღალ დონეს. თამამად შეიძლება ითქვას, რომ არ ასებობს პრობლემა ფუნქციური გენომიკის, პროტეომიკის, სტრუქტურული ბიოლოგიის, სისტემური ბიოლოგიის ფარგლებში, ანუ თანამედროვე ბიოლოგიის ძირითადი მიმართულებების ფარგლებში, რომლის გადარწვევისათვის ბიოინფორმატიკულ კვლევებში დაგროვილმა ცოდნამ თავისი დადებითი წვლილის შეტანა ვერ შეძლოს.

ბიო-სამედიცინო პრობლემები და ბიოინფორმატიკა

გენური პროექტები, როგორც ინფორმაციის ძირითადი წყარო.

მოდით ახლა შევეცადოთ უფრო ამომწურავად ვუპასუხოთ კითხვას - რამ განაპირობა ბიოლოგიაში მძლავრი იმფორმაციული ნაკადებისა და შესაბამისად ბიოინფორმატიკის წარმოშობა და რას ემსახურება იგი?

ცნობილია, რომ გენომური დნმ-ს ნუკლეოტიდური თანამიმდევრობა მატარებელია ყველა იმ ინფორმაციისა, რომელიც აუცილებელია ცოცხალი ორგანიზმის არსებობისათვის საჭირო მოლეკულური პროცესების სწორად განსახორცი-

ელებლად. აქედან გამომდინარე, გენომში თავმოყრილი სრული ინფორმაციის მოძიება და ამ ინფორმაციის დროსა და სივრცეში სწორად რეალიზების საშუალებების წარმოჩენა წარმოადგენს თანამედროვე ბიოლოგიის უპირველეს ამოცანას. რამდენიმე ათეული წლის განმავლობაში მრავალი ქვეყნის მკვლევართა ძალისხმევამ განაპირობა ზემოთხსენებულ მექანიზმთა შესახებ ზღვა ინფორმაციის დაგროვება. მიუხედავად ამისა, დღეისათვის შესაძლებელი გახდა მხოლოდ გარკვეული საფუძვლების შექმნა და თანამედროვე ბიოლოგიის ძირითადი მიზნისკენ სწორი მიმართულებით მოძრაობის უზრუნველსაყოფად აქცენტი გაკეთდა ურთიერთკავშირში მყოფ, მაგრამ მაინც გარკვეულწილად ცალკეულ მიმართულებებზე. ამათგან შეიძლება გამოვყოთ ფუნქციური გენომიკა, პროტეომიკა, სტრუქტურული ბიოლოგია, სისტემური ბიოლოგია. აქ უპირიანია მოკლედ მიმოვიხილოთ ზემოთ ხსენებული მიმართულებები, წარმოვაჩინოთ მიმართულებათა ფარგლებში ჩასატარებელ კვლევათა ამოცანები.

კიდევ ერთხელ აღვნიშნავ, რომ ცოცხალი ორგანიზმის არსებობისათვის აუცილებელი მთელი ინფორმაცია თავმოყრილია დნმ-ს ნუკლეოტიდურ თანამიმდევრობაში და ამან განაპირობა სწორედ სხვადასხვა ორგანიზმების გენომთა ნუკლეოტიდურ თანამიმდევრობების შესახებ ინფორმაციების სულ უფრო მზარდი ტემპებით დაგროვების ტენდენცია. ასეთი ტენდენციის შენარჩუნების საფუძველს იძლევა არსებული სეკვენირების მეთოდების შემდგომი დახვეწა ან პრინციპულად ახალ მიდგომებზე გადასვლის შესაძლებლობები. ამის შესახებ ქვემოთ, ახლა კი ხაზი გავუსვათ იმას, რომ გენომის ნუკლეოტიდური თანამიმდევრობის ცოდნა თავისთავად ვერ უზრუნველგვყოფს სრული ინფორმაციით, მაგალითად ინფორმაციით გენეტიკური აპარატის ფუნქციონირების საფუძველში მყოფ მოლეკულური მექანიზმების შესახებ, თუ არაა ცნობილი ღია წასაკითხი ჩარჩოების ადგილმდებარეობა, რეგულატორული უბნების სტრუქტურა, მათი ურთიერთ და წასაკითხ ჩარჩოებთან მიმართება და სხვა ანალოგიური მონაცემები. სწორედ ამ უკანასკნელი საკითხების გარკვევის მიზნით ჩატარებული კვლევები უზრუნველგვყოფს ინფორმაციით, რომელიც საფუძველად უდევს თანამედროვე ბიოლოგიის ერთერთ მნიშვნელოვან მიმართულებას - ფუნქციურ გენომიკას.

პროტეომიკა მოიცავს ინფორმაციას მთელი გენომის მასშტაბით გენთა, ანუ ექსპრესირებად უბანთა ფუნქციური დატვირთვის შესახებ და მას ერთ-ერთი ცენტრალური ადგილი უკავია ფუნქციონალურ გენომიკაში. ფუნქციონალური გენომიკა პროტეომიკის გარდა გულისხმობს აგრეთვე გენომის არაექსპრესირებადი ნაწილების კვლევას, მათ შორის რეგულატორული უბნებისა და ე.ი. რეგულაციის მექანიზმებისა.

ფუნქციასა და სტრუქტურას შორის ურთიერთმიმართების საკითხების გარკვევა თანამედროვე ბიოლოგიის ერთ-ერთი მნიშვნელოვანი ამოცანაა. სტრუქტურული ბიოლოგია სწორედ ფუნქცია-სტრუქტურას შორის ურთიერთმიმართების

გარკვევის მიზნით ჩატარებული კვლევების პროცესში ცოდნის დაგროვებას გულისხმობს. ასეთი ცოდნა საშუალებას იძლევა გადაწყვეტით თანამედროვე ბიოლოგიისთვის ისეთი მნიშვნელოვანი ამოცანები, როგორცაა ბიომაკრო-მოლეკულების სივრცული მოწყობის კანონზომიერებათა დადგენა (მაგალითად, მხოლოდ ღია წასაკითხი ჩარჩოს ნუკლეოტიდური თანამიმდევრობის ცოდნის საფუძველზე ცილის სივრცული მოწყობისა და შესაბამისი ფუნქციური დატვირთვის დადგენა), ბიოლოგიური ფუნქციონირებისათვის საჭირო მოლეკულური მექანიზმების სტრუქტურულ საფუძვლებში გარკვევა, ბიომაკრომოლეკულებს შორის ურთიერთქმედების უნარის შეფასება, და აქედან გამომდინარე გენეტიკური აპარატის ფუნქციონირების მოლეკულური მექანიზმების ახსნა. აღსანიშნავია, რომ მოლეკულური მექანიზმების სტრუქტურული საფუძვლების ცოდნა თავის მხრივ უზრუნველგვყოფს ამ მექანიზმებზე ზემოქმედების შესაძლებლობებით. აქვე შეგვიძლია შევნიშნოთ, რომ ზემოთ მოცემულ სფეროთა შორის გამიჯვნა არამკვეთრია და მათში თავმოყრილ ცოდნათა შორის ადგილი აქვს გადაფარვას.

საერთოდ, ბიოლოგიურ კვლევათა მომავალი უფრო ინტეგრაციული ხასიათისაა და ამით განსხვავდება წარსული რეალობისაგან, როცა ძირითადად სპეციალიზაციაზე კეთდებოდა აქცენტი. წარსულში ძირითადად შეისწავლებოდა ცალკეული ბიოლოგიური ერთეულების სტრუქტურული მოწყობა და ბიოლოგიური ფუნქცია, დღეს კი, კვლევები სულ უფრო მეტად მიმართულია ამ ერთეულებს შორის ურთიერთქმედების გარკვევაზე. ასეთი „სისტემური მიდგომების“ განხორციელების აუცილებლობა გახდა საწინდარი სწორედ ე.წ. სისტემური ბიოლოგიის განვითარებისა.

აქ ხაზი უნდა გაესვას იმ გარემოებას, რომ თანამედროვე ბიოლოგიის უმნიშვნელოვანესი პრობლემის გადაწყვეტის პროცესში დაგროვილი ცოდნა მე-20 საუკუნის 80-ანი წლების ბოლომდე ძირითადად ექსპერიმენტული კვლევების შედეგად მოიპოვებოდა, კომპიუტერული მიდგომების გამოყენებით დაგროვილი ცოდნის წილი კი, მცირე იყო. მდგომარეობა ძირეულად შეიცვალა წინა საუკუნის 90-ანი წლების ბოლოსა და ამ საუკუნის დასაწყისიდან, როცა ერთის მხრივ მოხდა რიგ ორგანიზმთა გენომების სრული ნუკლეოტიდური თანამიმდევრობების დადგენის მიზნით განხორციელებული პროექტების, მათ შორის ადამიანის გენომის პროექტის წარმატებული დასრულება და მეორეს მხრივ შეიქმნა ტექნოლოგიური საფუძველი სისტემური მიდგომების რეალიზებისათვის.

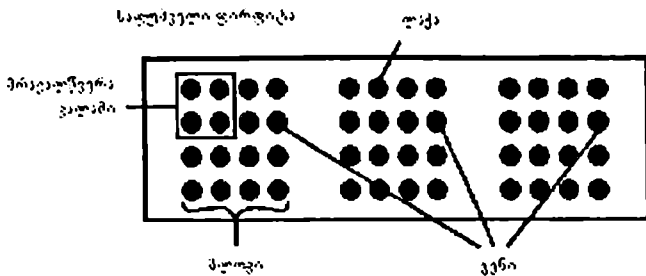
სკვენირების მეთოდების დახვეწის კვალდაკვალ გაიზარდა იმ ორგანიზმთა რიცხვი, რომელთა გენომების ნუკლეოტიდური თანამიმდევრობები უკვე გაიშიფრა ან გაშიფრის პროცესშია. დღევანდელი პროგნოზით ასეთ ტენდენციას კიდევ დიდხანს ექნება ადგილი. შესაბამისად, დღის წესრიგში მეტად მტკივნულად დადგა საკითხი - გენომის ნუკლეოტიდურ ტექსტში შენახული ყველა ინფორმაციის წარმოჩენისა და ამ პროცესის გენომების ნუკლეოტიდური თანამიმდევრობის შესახებ ინფორმაციის მოპოვების სიჩქარის შესაბამისი

სიჭქართ განხორციელების. ცხადია, ექსპერიმენტული მიდგომების ძვირადღირებულებისა და შრომატევადობის გამო, უკანასკნელი ამოცანის მხოლოდ ექსპერიმენტულ მიდგომებზე დაფუძნებით განხორციელება არარეალურია. და აქ გაჩნდა კომპიუტერული მიდგომები, რომლებიც ეფუძნებოდა რა მათემატიკურ და კომპიუტერულ მეცნიერებათა მიღწევებს, წარმოადგენდა ნუკლეოტიდურ ტექსტებში შენახული ინფორმაციის მცირე დანახარჯებითა და მოკლე დროში ანალიზის საშუალებებს. შეიძლება ითქვას, რომ თანამედროვე ბიოლოგიის ნებისმიერი მიმართულების განვითარების უზრუნველყოფის მიზნით შემუშავებულია შესაბამისი სპექტრი მონაცემთა ბაზებისა და ამ ბაზებთან სამუშაოდ განკუთვნილი პროგრამათა პაკეტებისა. ცხადია, ინფორმაციის დაგროვებასთან ერთად, მონაცემთა ბაზების, კომპიუტერული მიდგომებისა და შესაბამისი პროგრამათა პაკეტების რიცხვი გაიზრდება. შედეგად გაფართოვდება ბიოლოგიურ პრობლემათა სპექტრი, რომელთა წარმატებით გადაწყვეტა დიდწილად დამოკიდებული იქნება შესაბამისი პროგრამული უზრუნველყოფის არსებობაზე. ასეთ პრობლემათა თვალსაჩინო მაგალითად გამოდგება თანამედროვე ბიოლოგიის უახლოესი მიმართულების ე.წ. „სისტემური ბიოლოგიის“ განვითარების გზაზე მდგომი პრობლემები.

დნმ მიკრო-მატრიცებზე დაფუძნებულ კვლევები და ბიოინფორ-მატიკა.

როგორც ზემოთ ითქვა, ბიოლოგიურ კვლევათა მომავალი უფრო ინტეგრაციული ხასიათისა და ამით განსხვავდება წარსული რეალობისაგან, როცა ძირითადად სპეციალიზაციაზე კეთდებოდა აქცენტი. წარსულში ძირითადად შეისწავლებოდა ცალკეული ბიოლოგიური ერთეულების სტრუქტურული მოწყობა და ბიოლოგიური ფუნქცია. დღეს კი, კვლევები სულ უფრო მეტად მიმართულია ამ ერთეულებს შორის ურთიერთქმედების გარკვევაზე. ასეთი „სისტემური მიდგომების“ განხორციელება კი, შესაბამისი მათემატიკური აპარატისა და პროგრამული უზრუნველყოფის გარეშე შეუძლებელია. ასეთი სისტემური მიდგომის მაგალითებად გამოდგება გენთა ექსპრესიის ფართო, მასშტაბური სურათის ანალიზი, რომელიც ეფუძნება ჰიბრიდიზაციის ექსპერიმენტებს და ხორციელდება დნმ მიკრომატრიცებზე (იხ. სურათი 1.1)

ასეთი ექსპერიმენტების აუცილებლობა კიდევ უფრო აძლიერებს ბიოინფორმატიკულ მიდგომებზე მოთხოვნილებას. შეიძლება თამამად ითქვას, რომ სულ უფრო იზრდება მოთხოვნა გენთა ფუნქციის წინასწარმეტყველების



სურათი 1.1 დნმ მიკრო-მატრიცის სქემატური წარმოდგენა. სურათზე მოცემულია საფუბელი-ფირფიტის (slide) მიკრო-მატრიცად მოწყობის მაკეტი. მიკრო-მატრიცის შემადგენელი ძირითადი ერთეულებია უმცირესი უბნები, ე.წ. „ლაქები“ (wells). ლაქებს კენის მინის საფუბელთან მყარად მიმაგრებული ცდნმ კლონები. ლაქათა ზომა, ფორმა და მათში განთავსებული დნმ-ს კონცენტრაცია ცვალებადია. ლაქების გენეირებისათვის შექმნილია სპეციალური რობოტიზირებული მექანიზმები, რომლებიც უზრუნველყოფილია მრავალწევრიანი (მრავალკალმანი, plurisoup) დასატანი მოწყობილობით. დატანა შესაძლებელია წარმოებდეს ბლოკების სახით (subarrays), ანუ არაერთდროულად დატანილი უბნების სახით. ერთი და იგივე გენი (კლონი) შეიძლება რამდენჯერ იქნეს დატანილი ერთსა და იმავე საფუბელ-ფირფიტაზე. ლაქების სახით დატანილი ზონდების (საკვლევი დნმ კლონების) რიცხვი შეიძლება აჭარბებდეს 40 000-ს. ასეთი რაოდენობა ელემენტებისაგან შემდგარ მატრიცაზე ჰიბრიდიზაციის ექსპერიმენტის შედეგების ინტერპრეტირება სპეციალურად შემუშავებული პროგრამული უზრუნველყოფის გარეშე პრაქტიკულად შეუძლებელია. ზოგადად ნათელია, რომ რთული ბიოლოგიური სისტემების ფართომასშტაბური ექსპერიმენტებით კვლევის შესაძლებლობები მძლავრი ბიოინფორმატიკული ზაზის არ ქონის პირობებში შეზღუდულია.

მეთოდებზე. გენთა ბიოლოგიური ფუნქციის ცოდნა როგორც ბიოლოგიური, ისე კლინიკური ინტერესების სფეროა და შესაბამისად, გვეხმარება როგორც შემდგომი ბიოლოგიური კვლევების შესახებ გადაწყვეტილებების მიღებაში, ასევე თერაპიის სამიზნეთა შერჩევაში. ასე მაგალითად, მიკრომატრიცული ექსპერიმენტების სამართავად შექმნილ პროგრამათა პაკეტების უმრავლესობა ინფორმაციას გენთა ფუნქციის შესახებ იღებს საზოგადო მოხმარების მოზარდებთან ბაზებიდან, რომლებიც მოიცავს ინფორმაციას გენომში დაფიქსირებულ გენთა მხოლოდ დაახლოებით ნახევრის შესახებ. ამ ხარვეზის აღმოფხვრა კი, დიდწილად სტრუქტურული ბიოინფორმატიკის პრეროგატივაა.

სტრუქტურული ბიოლოგია და ბიოინფორმატიკა

უჯრედში ინფორმაცია შენახულია გენომის სტრუქტურული ორგანიზაციისა და გენთა ნუკლეოტიდური თანამიმდევრობების ფორმით. ამ ინფორმაციის გარკვეულ ბიოლოგიურ ფუნქციად რეალიზების გზა სქემატურად ასე შეიძლება წარმოვადგინოთ:

გენის ნუკლეოტიდური თანამიმდევრობა - ცილის ამინმჟავური თანამიმდევრობა - ცილის სივრცული სტრუქტურა - ბიოლოგიური ფუნქცია

ეს გზა ნათლად მიუთითებს სტრუქტურული ბიოლოგიის როლზე თანამედროვე ბიოლოგიის ძირითადი ამოცანების გადაწყვეტის პროცესში. სტრუქტურული ბიოლოგია გულისხმობს ცოდნას ბიომაკრომოლეკულათა სტრუქტურის, პოლიპეპტიდთა ცილად დახვევის, ცილის, დნმ-სა და რნმ-ს სივრცული მოწყობის, მათი ურთიერთქმედების კანონზომიერებათა, ასევე იმ მეთოდთა შესახებ, რომლებიც გამოიყენება ამ კანონზომიერებათა დასადგენად. ბიომაკრომოლეკულათა ფუნქცია ძირითად განპირობებულია მათი სივრცული მოწყობის თავისებურებებით. მოლეკულური მექანიზმების ახსნა კი მოითხოვს ბიომაკრომოლეკულათა ფუნქციონირების შესახებ და ე.ი სივრცული სტრუქტურის შესახებ ამომწურავი ცოდნის ფლობას. აქედან გამომდინარე, სტრუქტურული ბიოლოგიის მიღწევები დიდწილად განაპირობებს თანამედროვე ბიოლოგიის ძირითადი მიზნის, ცოცხალი უჯრედისა თუ მრავალუჯრედიანი ორგანიზმის ნორმალური ცხოველმყოფელობის უზრუნველმყოფი მოლეკულური მექანიზმების შესახებ სრული ცოდნის დროულად მიღების შესაძლებლობას.

სტრუქტურული ბიოლოგიის საკვანძო როლი განსაკუთრებით თვალშისაცემია დღეს, როცა თავი იჩინა დისბალანსმა გენომთა ნუკლეოტიდური თანამიმდევრობების შესახებ დაგროვილი ინფორმაციის მოცულობას, ამ ინფორმაციული ნაკადის სიჩქარეს, ერთის მხრივ, და სტრუქტურული ინფორმაციის მოცულობას, მისი ექსპერიმენტული მეთოდებით მოპოვების სიჩქარეს შორის, მეორეს მხრივ. სტრუქტურული ინფორმაციის ექსპერიმენტული მეთოდებით მიღების პროცედურა საკმაოდ რთული, ძვირადღირებული და დროში გაწელილია. ამის მაგალითებად შეიძლება მოვიყვანოთ რენტგენოსტრუქტურული ანალიზისა და ბირთვული მაგნიტური რეზონანსის მეთოდები. სტრუქტურული ბიოლოგიის პრობლემების სწრაფად და მცირე დანახარჯებით გადაწყვეტისათვის ბუნებრივია, დღის წესრიგში დადგა ალტერნატიული, კომპიუტერულ მიდგომებზე დაფუძნებული მეთოდის შემუშავების აუცილებლობა. ასეთი მეთოდების შემუშავება-განვითარების მცდელობას ადგილი ჰქონდა უკვე კომპიუტერული ტექნიკით სარგებლობის საშუალების გაჩენისთანავე. ამიტომ შეიძლება თამამად ითქვას, რომ ისტორიულად ბიოინფორმატიკას საფუძველი ჩაეყარა სწორედ სტრუქტურული ბიოლოგიის პრობლემების კომპიუტერული მიდგომებით გადაწყვეტის მცდელობის მომენტიდან და შესაბამისად, სტრუქტურული

ბიოლოგიის პრობლემების გადასაწყვეტად შექმნილი ადრეული კომპიუტერული მიდგომები შეიძლება განხილულ იქნეს, როგორც ბიოინფორმატიკულ მეთოდთა პირველი ნიმუშები.

შემდგომში, სტრუქტურული ბიოლოგიის ამოცანათა აქტუალობის ზრდის პირობებში, შესაბამის კომპიუტერულ მიდგომათა მეტად მრავალფეროვანი და დახვეწილი სპექტრის ჩამოყალიბებამ და მისი შემდგომი განვითარების მწვავე აუცილებლობამ დღის წესრიგში დააყენა სტრუქტურული ბიოინფორმატიკის ბიოინფორმატიკიდან გამოყოფის და მისი დამოუკიდებელ მიმართულებად ჩამოყალიბების საკითხი. თუმცა მართებულია სტრუქტურული ბიოინფორმატიკის ბიოინფორმატიკის მნიშვნელოვან, განუყოფელ ნაწილად განხილვაც.

სტრუქტურული ბიოინფორმატიკა დღეს მეტად სწრაფად განვითარებადი დარგია. მისი მიღწევები უკვე თვალში საცემია. სტრუქტურულ ბიოინფორმატიკაში დაგროვილმა ცოდნამ და მეთოდებმა შეუცვლელი სამსახური შეუძლია გაუწიოს ბიოლოგიური პროცესების საფუძველში მყოფი მოლეკულური მექანიზმების შემსწავლელ მკვლევარს.

2. ბიოინფორმატიკის როლი გენომური პროექტების რეალიზების პროცესში

დღეს უკვე შეიძლება თამამად ითქვას, რომ ბიოლოგიაში სრულფასოვანი კვლევების ჩასატარებლად აუცილებელი ხდება ბიოინფორმატიკის, კომპიუტერული ბიოლოგიის მეთოდთა საფუძვლების ცოდნა. უკანასკნელი მტკიცების დამაჯერებლობის საჩვენებლად შესაძლებელია წარმოვადგინოთ კომპიუტერული მიდგომებით თანამედროვე ბიოლოგიის სრულიად განსხვავებულ პრობლემათა გადაწყვეტის მაგალითები. რადგანაც ჩვენ გენომური პროექტების ერაში ვცხოვრობთ, სადემონსტრაციოდ გამოვიყენოთ თანამედროვე ბიოლოგიის უმნიშვნელოვანესი ამოცანის, გენომის ნუკლეოტიდური თანამიმდევრობის დადგენის ამოცანის გადასაწყვეტად არსებული საშუალებები და ვნახოთ, რა როლს თამაშობს კომპიუტერული მიდგომები მოცემული ამოცანის წარმატებულად გადაწყვეტის საქმეში.

გენომური პროექტის ამოცანები და მათი გადაწყვეტის საშუალებები

გენომური პროექტები ითვალისწინებს ორგანიზმთა (ცხოველები, მცენარეები, სოკოები, ბაქტერიები, არქიბაქტერიები, ერთუჯრედიანი ორგანიზმები, ვირუსები) გენომების სრული ნუკლეოტიდური თანამიმდევრობების დადგენას. ორგანიზმის გენომური თანამიმდევრობის დადგენა ნიშნავს თითოეული ქრომოსომის დნმ-ს ნუკლეოტიდური თანამიმდევრობის დადგენას. ბაქტერიებისათვის, რომელთა უჯრედებში, როგორც წესი, ერთი ქრომოსომაა, გენომური პროექტი მიზნად ისახავს სწორედ ამ ქრომოსომის თანამიმდევრობის დადგენას. ადამიანის შემთხვევაში კი, გენომის შესახებ სრული ინფორმაციის ცოდნა ითვალისწინებს 24 წყვილი ქრომოსომის ნუკლეოტიდური თანამიმდევრობის დადგენას.

ადამიანის გენომის პროექტი შეიძლება ეპოქალურ პროექტად ჩაითვალოს. გენომიკის ერის დაწყება კი თამამად შეიძლება იქნეს მიჩნეული კაცობრიობის ერთ-ერთ უმნიშვნელოვანეს მიღწევად. აქ დაწვრილებით არ შეეჭვრებოდა ადამიანის გენომის პროექტის თავისებურებებზე. თქვენს ყურადღებას შევაჩერებთ მხოლოდ გენომური პროექტის რეალიზების პროცესში გადასაწყვეტ ორ მნიშვნელოვან ამოცანაზე. ესენია ასამბლირების (აწყობის) და ანოტირების ამოცანები.

ასამბლირების პრობლემის წარმოშობა უკავშირდება დიდი სიგრძის დნმ-ს სეკვენირებისათვის შემუშავებული მიდგომის, დაფანტვით სეკვენირების (Shotgun sequencing) სპეციფიკას.

დაფანტვით სეკვენირება

დიდი სიგრძის დნმ-ს სეკვენირების პროცესში, როგორც წესი, გამოიყენება ე.წ. დაფანტვით სეკვენირების მიდგომა (Shotgun sequencing. Shotgun - არის საფანტის თოფი და მასთან ანალოგია უკავშირდება იმ ფაქტს, რომ ხდება დნმ-ს შემთხვევითად, უწესრიგოდ, მცირე ნაწილებად დანაწევრება, დაფანტვა). ასეთი მიდგომის გამოყენება განპირობებულია იმ ფაქტით, რომ ჯაჭვის ტერმინირებაზე დაფუძნებული მეთოდი (სენჯერის), რომელიც დნმ-ს მხოლოდ შედარებით მოკლე ჯაჭვების სეკვენირების საშუალებას იძლევა, იძულებულს გვხდის დავანაწევროთ დნმ-ს გრძელი ჯაჭვი მცირე ფრაგმენტებად, მოვახდინოთ მათი ცალ-ცალკე სეკვენირება და შემდგომ, სეკვენირებული ფრაგმენტების რეასამბლირებით აღვადგინოთ საწყისი (საკვლევი) დნმ-ს თანამიმდევრობა. ზოგადად, სეკვენირებისას, გამოიყენებოდა ორი მიდგომა: ერთი, ე.წ. ქრომოსომაზე გასეირნება (chromosome walking), ან პრაიმერთ გასეირნება (primer walking). ასეთ მიდგომით შესაძლებელია მხოლოდ 1.7-7 ათას ფუძის (kilobase) სიგრძის დნმ-ს ფრაგმენტების სეკვენირება და ამ დროს დნმ-ს თანამიმდევრობის გავლა ხდება თანამიმდევრულად ნაწილ-ნაწილ; მეორე - დაფანტვით სეკვენირება, რომელიც წარმოადგენს უფრო სწრაფ, მაგრამ ამასთანავე კომპლექსურ პროცესს, გამომდინარე იქიდან, რომ საქმე გვაქვს შემთხვევით ფრაგმენტებთან, მოითხოვს შესაბამისი მათემატიკური საფუძვლების შემუშავებასა და პროგრამულ უზრუნველყოფას.

როგორც უკვე ითქვა, დაფანტვით სეკვენირების პროცესში ხდება დნმ-ს შემთხვევითი დანაწევრება მთელ რიგ მცირე სიგრძის ფრაგმენტებად, რომელთა ჯაჭვის ტერმინირების მეთოდით სეკვენირების შედეგად მივიღებთ წანაკითხებს (დნმ-ის სრული ტექსტის მოკლე ფრაგმენტებს). იმისათვის, რომ მიღებულ იქნეს წანაკითხთა ურთიერთგადაფარვადი წყვილების დიდი რაოდენობა, ტარდება ფრაგმენტაციისა და სეკვენირების რამდენიმე რაუნდი. ამის შემდგომ სპეციალური კომპიუტერული პროგრამა, წანაკითხთა გადაფარვად ბლოკების შესახებ ინფორმაციაზე დაყრდნობით, ახდენს მათ ასამბლირებას ერთიან უწყვეტ თანამიმდევრობად.

ასამბლირება

ასამბლირება გულისხმობს შემთხვევითი გზით მიღებული მოკლე დნმ ფრაგმენტების სეკვენირების შედეგად წარმოჩენილ მოკლე დნმ თანამიმდევრობათაგან ქრომოსომის სრული ნუკლეოტიდური თანამიმდევრობის აწყობის

პროცესს (ანუ შემთხვევითი დანაწევრებით მიღებულ მოკლე თანამიმდევრობათა სწორი ურთიერთგანლაგების პროცესს).

მაგალითისათვის მოვიყვანოთ ორრაუნდიანი დაფანტვის შედეგად წარმოებული წანაკითხები (სურათი 1.2). ამ უმარტივეს მაგალითში ვხვდებით ოთხ წანაკითხს, რომელთა ასამბლირება საწყის თანამიმდევრობად წარმოებს მათი

საკვლევი დნმ-ს ჯაჭვი : XXXAGCATGCTGCAGTCATGCTTAGGCTAXXXX

პირველი დანაწევრების შედეგები: XXXAGCATGCTGCAG
TCATGCTTAGGCTAXXXX

მეორედ დანაწევრების შედეგები : XXXAGCATGCTGCAGTCATGC
TTAGGCTAXXXX

რეკონსტრუირებული ჯაჭვი : XXXAGCATGCTGCAGTCATGCTTAGGCTAXXXX

სურათი 1.2. ასამბლირების მარტივი მაგალითი

ბოლოების გადაფარვების შესახებ ინფორმაციაზე დაყრდნობით. რეალობაში საქმე გვაქვს დიდ რაოდენობა ინფორმაციასთან, რომელიც გამომდინარე იქიდან, რომ სეკვენირების პროცესში ადგილი აქვს შეცდომებს, შეიძლება იყოს არაგალსახა. სრული გენომის ასამბლირებას კიდევ უფრო ართულებს განმეორებები, ანუ მსგავსი მოკლე წანაკითხები, რომელთა გენომში განლაგების ადგილი განსხვავებულია. ამრიგად, გენომის ასამბლირება თავისთავად საკმაოდ რთული კომპიუტერული პროცედურაა და მას კიდევ უფრო ართულებს ის ფაქტი, რომ გენომური თანამიმდევრობა, როგორც წესი, მდიდარია განმეორებადი ფრაგმენტებით. ასეთი განმეორებადი თანამიმდევრობები შეიძლება წარმოადგენდეს ათასამდე ნუკლეოტიდის სიგრძის ფრაგმენტებს და ერთი და იგივე თანამიმდევრობის მქონე ფრაგმენტს შეიძლება შევხვდეთ გენომის ათასამდე სხვადასხვა ადგილას. განმეორებები განსაკუთრებით დამახასიათებელია მცენარეთა და ცხოველთა გენომებისათვის.

გადაფარვად წანაკითხთა დიდი სიმრავლის პირობებში შესაძლებელია ზემოთ მითითებულ სირთულეთა დაძლევა და ზუსტი ასამბლირების მიღწევა. ასე მაგალითად, ადამიანის გენომის პროექტის რეალიზების პროცესში უმრავლესობა გენომებისა სეკვენირებულ იქნა 12X (12-ჯერადი) ან უფრო დიდი რიგის გადაფარვებით, რაც იმას ნიშნავს, რომ სრული თანამიმდევრობის თითოეული ფუძე, საშუალოდ შედიოდა 12 სხვადასხვა წანაკითხში. მიუხედავად ამისა,

მეთოდმა მაინც ვერ შეძლო ადამიანის გენომის დაახლოებით 1%-ის ცალსახად ასამბლირება (ცალსახად განლაგება).

ნაგლეჯთა ასამბლირებისათვის შექმნილია უამრავი, სხვადასხვა ალგორითმზე დაფუძნებული პროგრამათა პაკეტი. გენომური (დიდი სიგრძის) დნმ-ს სეკვენირებით დაკავებული ნებისმიერი კვლევითი ჯგუფი აუცილებლად იყენებს რომელიმე მათგანს.

ანოტირება

ანოტირება გულისხმობს თანამიმდევრობისათვის ბიოლოგიური დანიშნულების მოძიებისა და მინიჭების პროცესს. ეს პროცესი მიმდინარეობს ორ ეტაპად:

1. გენომის შემადგენელი ელემენტების იდენტიფიცირება, ანუ გენთა მოძიების ეტაპი.
2. მოძიებული ელემენტებისათვის ბიოლოგიური დანიშნულების განსაზღვრა.

ავტომატური ანოტირებისათვის განკუთვნილი პროგრამული საშუალებანი ამ ყველაფრის რეალიზებას ახდენს კომპიუტერული ანალიზის საფუძველზე. ანოტაცია შესაძლებელია კომპიუტერის გარეშე (ხელით). რეალურად ეს ორივე მიდგომა თანაარსებობს ანოტაციის რეალიზების პროცესში.

ანოტაციის ძირითადი ეტაპი შეიძლება განხორციელდეს იმ პროგრამათა პაკეტის გამოყენებით, რომელიც წარმოაჩენს თანამიმდევრობათა შორის მსაგავსებას და სწორედ მსგავსების შეფასებებზე დაფუძნებით მოხდება ანოტაცია. დღეს-დღეობით სულ უფრო მეტი დამატებითი ინფორმაციით ივსება საანოტაციო ბაზები. ეს დამატებითი ინფორმაცია საშუალებას აძლევს ანოტატორს წარმოაჩინოს განსხვავებები გენებში, რომლებსაც ადრე გააჩნდათ ერთნაირი ანოტაცია. მონაცემთა ბაზებში გამოყენებად ინფორმაციათა სახეების მაგალითებად შეიძლება მოვიყვანოთ: გენომური კონტექსტები (ელემენტები), მსგავსებათა დამახასიათებელი ქულები, ექსპერიმენტული მონაცემები და სხვა რესურსებთან (ბაზებთან) ინტეგრირების შესაძლებლობები. ასეთი ქვესისტემური მიდგომით შესაძლებელია ზედმიწევნით ზუსტი გენომური ანოტაციის ფორმირება.

არჩვენ ანოტაციის სტრუქტურულ და ფუნქციურ ნაწილს

სტრუქტურული ანოტაცია გულისხმობს გენომური ელემენტების ინდენტიფიცირებას:

- ORF (open reading frame) -ებს და მათ ლოკალიზაციას
- გენთა სტრუქტურის აღწერას
- მაკოდირებელი უბნების იდენტიფიცირებას
- რეგულატორული უბნების იდენტიფიცირებასა და ლოკალიზაციას

ფუნქციონალური ანოტაცია გულისხმობს გენომური ელემენტებისათვის ბიოლოგიური დანიშნულების განსაზღვრას:

- ბიოქიმიური ფუნქციის
- ბიოლოგიური ფუნქციის
- ჩართულობას რეგულაციასა და ურთიერთქმედებებში
- ექსპრესიის მახასიათებლებს (დონეებს და მისი რეგულაციის საშუალებებს)

ფუნქციური ანოტაციისას ეყრდნობიან როგორც ბიოლოგიური ექსპერიმენტის მონაცემებს, ასევე *in silico* (კომპიუტერულ) ანალიზს.

შექმნილა უამრავი მონაცემთა ბაზა და მათი მომსახურე პროგრამათა პაკეტი და მათი საშუალებით ფართო სამეცნიერო საზოგადოებისათვის ხელმისაწვდომი ხდება ანოტირებულ გენომთა შესახებ ინფორმაცია.

გენომური პროექტების განვითარების ტენდენციები და შედეგები

ისტორიულად, როცა ხდებოდა ერთ-ერთი პირველი ეუკარიოტული გენომის (*Caenorhabditis elegans*) სეკვენირება, პირველ ეტაპზე მოხდა გენომის გასწვრივ მთელი რიგი ორიენტირების დაფიქსირება და ამით გენომის საერთო რუკის წარმოდგენა. შესაბამისად, გენომის სეკვენირება წარმოებდა არა ერთი გავლით, არამედ ნაწილ-ნაწილ (იგოდნენ რა ამ ნაწილების დაახლოებითი მდებარეობა ქრომოსომაში). ახალმა ტექნოლოგიებმა, და კერძოდ თანამედროვე გამომთვლელ მანქანათა გაზრდილმა შესაძლებლობებმა მეცნიერებს უზიბა გამოეყენებინათ მოკლე, შემთხვევითი, დნმ ფრაგმენტებით სეკვენირების (shotgun sequencing) მეთოდი, რომელიც გულისხმობს გენომის ერთჯერზე გავლას.

დნმ-ს სეკვენირების ტექნოლოგიათა დახვეწა ნიშნავს სეკვენირების გაიაფებას (ერთი ფუძე წყვილის დადგენაზე გათვლილი დანახარჯების შემცირებას), სეკვენირების პროცესის დაჩქარებას. კვლევების დაგეგმვისას პრიორიტეტი

ენიკება იმ ორგანიზმების გენომების სეკვენირებას, რომლებიც მნიშვნელოვანია ადამიანის ჯანმრთელობისათვის (მაგალითად ბაქტერიები ან ავადმყოფობის გამომწვევი სხვა ორგანიზმები), ან მნიშვნელოვანია კომერციული თვალსაზრისით. ნაკლებად პრიორიტეტულია, თუმც მათგან ხდება იმ ორგანიზმების გენომთა თანამიმდევრობის დადგენა, რომლებმაც შეიძლება წვლილი შეიტანონ მოლეკულური ევოლუციის საკითხების გარკვევაში .

სავარაუდოდ, მომავალში კიდევ უფრო სწრაფად და იაფად მოხდება გენომის სეკვენირების ამოცანის გადაწყვეტა. ეს საშუალებას მოგვცემს უკვე ერთი და იგივე სახეობის სხვადასხვა ინდივიდისთვის მოვახდინოთ გენომის სრული თანამიმდევრობის დადგენა. ადამიანის შემთხვევაში ეს უსაზღვრო, ფანტასტიკური პერსპექტივების საწინდარია.

ასეთი პერსპექტივის ერთ-ერთ მაგალითად შეიძლება მოვიყვანოთ ფარმაკოგენომიკის (ან ტოქსიკოგენომიკის) განვითარება. ამ პროცესში ბიონფორმატიკის როლის მნიშვნელობაზე შემდეგ პარაგრაფში ვისაუბრებთ. იქვე შევეცდებით სეკვენირების მეთოდიკის დახვეწის გზების განხილვის მაგალითზე ვაჩვენოთ, რომ გენომური თანამიმდევრობის განსაზღვრის (სეკვენირების) პროცედურის გაიაფება შესაძლებელია მიღწეულ იქნეს უკანასკნელ პროცედურაში კომპიუტერული მიდგომების წვლილის გაზრდით.

3. ბიოინფორმატიკა თანამედროვე ბიოლოგიისა და მედიცინის განვითარების კრილში.

არსებობს მოლოდინი, რომ მომავალში კიდევ უფრო სწრაფად და იაფად მოხდება გენომის სეკვენირების ამოცანის გადაწყვეტა. ამ პარაგრაფში შევეცდებით ვაჩვენოთ თუ რა არის ამ მოლოდინის არსებობის საფუძველი. ასევე შევეცდებით ვისაუბროთ იმ ფანტასტიკურ პერსპექტივებზე მედიცინაში, რომლებიც ამ მოლოდინის რეალიზების შემთხვევაში ისახება.

დნმ-ს სეკვენირების მეთოდები და მათი განვითარება-დახვეწის დინამიკა

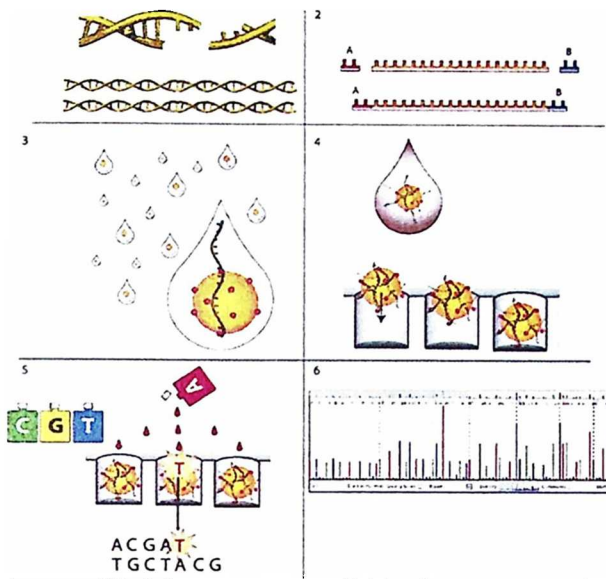
დნმ-ს სეკვენირება წარმოადგენს ნუკლეოტიდური თანამიმდევრობის განსაზღვრის პროცესს. ახლო წარსულში, როგორც წესი, სეკვენირებისას გამოიყენებოდა და რიგ შემთხვევებში დღესაც გამოიყენება ჯაჭვის ტერმინირების მეთოდი. ეს მეთოდი, შემუშავებული ფრედერიკ სენჯერის ჯგუფის მიერ ჯერ კიდევ 1975 წელს, იყენებს მოდიფიცირებულ ნუკლეოტიდურ სუბსტრატებს, რათა მოახდინოს დნმ-ს სინთეზის თანამიმდევრობა-სპეციფიკური ტერმინაცია (იხ. სურათი 1.3). დღეისათვის სეკვენირების ბაზარზე გამოჩნდა ახალი, პიროსეკვენირებაზე დაფუძნებული ტექნოლოგია, რომელიც სულ უფრო მეტად იპყრობს ამ ბაზარს. პიროსეკვენირების იდეა და შესასბამისი მეთოდი შემუშავებული და შემოთავაზებული იქნა მოსტაფა რონაჰის (Mostafa Ronaghi) მიერ. დაბალი წარმადობის სეკვენატორების შექმნით ამ მეთოდის კომერციალიზაცია მოახდინა „ბიოტაგმა“ (Biotage), ხოლო მაღალი წარმადობის სეკვენატორების შექმნა შეძლო „454 ლაიფ საიენსმა“ (454 Life Sciences). უკანასკნელის მიერ შემუშავებული, მატრიცაზე დაფუძნებული მეთოდით (იხ. სურათი 1.4) მომუშავე სეკვენატორის (454 სეკვენატორის სახელწოდებით ცნობილი) წარმადობა წარმოადგენს დაახლოებით 100 მეგაფუძეს 7 საათში. ამ ეტაპზე გენომური თანამიმდევრობების დიდი ნაწილი სწორედ პიროსეკვენირების დახმარებით იზიფრება, რადგანაც იგი გაცილებით სწრაფი პროცედურაა. ასე მაგალითად, ბაქტერიული გენომის გავლა პიროსეკვენირებით შეიძლება ერთ ჯერზე. სწორედ ამ ტექნიკით მოხდა მოლეკულური ბიოლოგიის ერთ-ერთი ფუძემდებლის, ჯეიმს უოტსონის გენომის სეკვენირება. როგორც ჯაჭვის ტერმინირების, ისე პიროსეკვენირების მეთოდი გულისხმობს ასამბლირების პროცედურას, მაგრამ პიროსეკვენირების შემთხვევაში უკანასკნელი პროცედურის წილი უფრო მნიშვნელოვანია.



სურათი 1.3 ჯაჭვის ტერმინირების მეთოდით სეკვენირების (ანუ სენჯერის მეთოდით სეკვენირების) პროცესში მიღებული ელექტროფორეზის ტიპური სურათი. სურათზე ჩანს ოთხი პარალელური სვეტი (ბილიკი), რომლებზე შეესაბამება ოთხ სხვადასხვა ნუკლეოტიდს. ბილიკის თითოეული ზოლი, რომელიც შეესაბამება საკვლევი დნმ-ს მოცემული ტიპის ნუკლეოტიდით ტერმინირებულ კონკრეტული სიგრძის ფრაგმენტს, განსაზღვრავს ამ ნუკლეოტიდის პოზიციას საკვლევი დნმ-ში. ჯაჭვის ტერმინირების მეთოდით სეკვენირების იდეა მარტივია: სეკვენირება უნდა დაიწყოს საკვლევი დნმ-ს სპეციფიკურ უბანთან კომპ -

ლემენტარული მოკლე „პრაიმერის“ მიერთებით. შემდგომ დნმ-პოლიმერაზას დამხარებით უნდა მოხდეს ამ პრაიმერის დაგრძელება, ანუ წარმართოს პოლიმერაზაციის რეაქცია. ამ რეაქციის რეალიზება მოითხოვს დნმ-ს ძირითადი საშენი მასალის, ოთხი სხვადასხვა დეზოქსინუკლეოტიდის დამატებას. პოლიმერაზაციის პროცესი უნდა წარმართოს ოთხ სხვადასხვა სინჯარაში, რომლებიც ერთმანეთისაგან განსხვავდება იმით, რომ თითოეულ სინჯარას მცირე რაოდენობით უმატებენ ჯაჭვის ზრდის მატერმინირებელ კონკრეტულ (ოთხიდან ერთს) ნუკლეოტიდს (როგორც წესი, მატერმინირებლად გამოიყენება di-დეზოქსინუკლეოტიდები). ჯაჭვის მატერმინირებელი ნუკლეოტიდის მცირე დოზებით გამოყენების შედეგად თითოეულ სინჯარაში დაგროვდება სხვადასხვა სიგრძის ფრაგმენტები. ამასთან თითოეული მათგანი ტერმინირებული იქნება კონკრეტული, მოცემული კურტ-ლისათვის დამახასიათებელი მატერმინირებელი ნუკლეოტიდით. ამის შემდეგ უნდა მოხდეს ოთხი სხვადასხვა სინჯარის შიგთავსის ელექტროფორეზის აპარატის ოთხ სხვადასხვა ბილიკზე დატანა და ელექტროფორეზის პროცედურის წარმართვა. რაც მივიყვანს ფრაგმენტების სიგრძეების მიხედვით განცალკავებამდე და ცალკეულ ზოლებად გამოყოფამდე, ანუ ზემოთ მოცემულ სურათამდე.

ახლახანს შემოთავაზებულ იქნა დნმ-ს სეკვენირების ახალი მეთოდი, დამყარებული ჰიბრიდიზაციაზე (ე.წ. SBH მეთოდი). იდეა მეტად მარტივია: იქმნება ორგანოზომილებიანი მატრიცა k-სიგრძის ფრაგმენტებისა (ზონდებისა), ანუ მატრიცის თითოეულ (i,j) უჯრასთან (ლაქასთან, იხ სურ. 1.1) დაკავშირებულია k სიგრძის ზონდი (k-ური ზონდი). ასეთ, მატრიცის სახით ორგანიზებულ k-ზონდთა ერთობას უწოდებენ სეკვენირების ჩიპს.



სურათი 1.4 454-სეკვენირების ცალკეული ეტაპების სქემატური წარმოდგენა. (1) მოკლე ამპლიკონები, ანუ დნმ-ს ან ცდნმ-ს ფრაგმენტები, რომლებიც შემდგომი ეტაპების საწყის ნიშნებს წარმოადგენენ. ფრაგმენტები მიიღება სხვადასხვაგვარი ნუკლეოზარესტრიქტაზული დამუშავების შედეგად. (2) ხდება ერთჯაჭვა ფრაგმენტების 5'- და 3' ბოლოების ლიგირება სპეციფიკური ადაპტორებით და შესაბამისად იქმნება ბიბლიოთეკა შემდგომი PCR ამპლიფიცირებისათვის. (3) ფრაგმენტები, ერთ-ერთ ადაპტორთან მი-

ერთებული ბიოტინური კუდის მეშვეობით იმობილიზირდება სტრეპტავიდინით დაფარულ პატარა ბურთულეებზე. თითოეულ ბურთულაზე ერთი ტიპის ფრაგმენტია მიერთებული. შემდგომ ასეთნაირად მომზადებული ბურთულეები თავსდება წყალ-ზეთოვან ნარევი (ემულსიაში). (4) ზეთის თითოეული წვეთი მოიცავს PCR-სთვის საჭირო ყველა ინგრედიენტს და ამრიგად წარმოადგენს მიკრორეაქტორს ამპლიფიცირებისათვის. შემდგომ ხდება ამპლიფიცირებული ფრაგმენტებით დახუნძლული ბურთულეების ზეთისაგან გამოთავისუფლება და მათი ჩატვირთვა ოპტიკური ბოჭკოსაგან დამზადებულ ჩიპში, პიკოლიტრატორული შესაძლებლობების მეორე ფირფიტისმაგვარ ჭურჭელში შემდგომი სეკვენირების განხორციელების მიზნით. ფირფიტის თითოეული = 44 μm დიამეტრის ღრმული ივსება ერთადერთი ბურთულით (შევსება მიიღწევა ცენტრიფუგირებით). (5) ჩიპზე (ფირფიტაზე) სხვადასხვა რეაგენტების თანამიმდევრული დამატების პირობებში მიმდინარეობს პიროსეკვენირების პროცესი. თუ დამატებული ნუკლეოტიდი კომპლემენტარულია გენერირდება სინათლის სიგნალი. (6) თითოეული ნიშნის პიროსეკვენირების შედეგი მოცემულია პიროგრამის სახით. თითოეული სიგნალის სიმაღლე პროპორციულია მეზობლად განლაგებულ, იდენტურ ნუკლეოტიდთა რიგებისა. თუ დამატებული ნუკლეოტიდი (ოთხი სხვადასხვა ნუკლეოტიდის თანამიმდევრული დამატების პროცესში, რომელიც ქმნის სეკვენირების ერთ ციკლს) არკომპლემენტარულია, სინათლის სიგნალი არ გენერირდება. (7) აღწერს ქიმიურ რეაქციათა იმ ერთობლიობას, რომელსაც მიეყვართ სინათლის სიგნალამდე.

ა) ATAAGCT

ბ) AAA AAT AAC AAG
ACA ACT ACC ACG
ATA ATT ATC ATG
AGA AGT AGC AGG
TAA TAT TAC TAG
TTA TTT TTC TTG
TCA TCT TCC TCG
TGA TGT TGC TGG
CGA CGT CGC CGG

სურათი 1.5 ჰიბრიდიზაციით სეკენირების პროცესში განხორციელებული ექსპერიმენტის ზოგადი პრინციპის წარმოდგენა. ა) საკვლევი დნმ-ის თანამიმდევრობა; ბ) $k=3$ ზონდებით დალაქულ მიკრომატრიცაზე (სულ $4^3=64$ ლაქა) საკვლევი ნიმუშით ჰიბრიდიზაციის ექსპერიმენტის შედეგების სქემატური წარმოდგენა. წითლად აღნიშნულია ის ლაქები, რომლებთანაც მოხდა ჰიბრიდიზაცია, ანუ რომელთა კომპლემენტარული ტრინუკლეოტიდები გვხვდება საკვლევი დნმ-ში .

საკვლელ ერთჯაჭვიან დნმ-ს, რომელიც მონიშნულია რადიოაქტიული ან ფლუორესცირებადი ნივთიერებით, დაიტანენ უკანასკნელ მატრიცაზე. დნმ-ს k სიგრძის უბნები, მონახავენ რა მატრიცაზე შესაბამის (კომპლემენტარულ) ზონდებს, ქმნიან მათთან ჰიბრიდულ წყვილებს. თუ ჩვენ არაჰიბრიზირებულ საკვლევ დნმ-ს მატრიციდან მოვხსნით, დარჩენილი, მონიშნული დნმ-ს საშუალებით გამოვლინდება ჰიბრიდიზაციაში მონაწილე k -ზონდთა ტიპები (თანამიმდევრობები). 1.5 სურათზე მარტივი მაგალითის სახით მოცემულია მატრიცა $k=3$ შემთხვევისათვის და ასეთ მატრიცაზე დასასეკენირებელი ATAAGCT თანამიმდევრობის ჰიბრიდიზაციის ექსპერიმენტის შედეგები.

ასეთი მიდგომის შემთხვევაში უკიდურესად გამარტივებული და გაიაფებულია ექსპერიმენტული ნაწილი და დატვირთვის დიდი წილი მათემატიკურ უზრუნველყოფაზე მოდის. მეთოდი განვითარების სტადიაშია და ცხადია დასაძლევია მთელი რიგი სიმწელებებისა, როგორც ექსპერიმენტული ნაწილის, ასევე მათემატიკური აპარატის სრულყოფის თვალსაზრისით. ექსპერიმენტული სიმწელებები დაკავშირებულია ჰიბრიდიზაციის პროცესის სპეციფიკასთან. რიგი ზონდებისა, მაგალითად G-C-ით მდიდარი, ჰიბრიდიზირდება უფრო მყარად, ვიდრე სხვები. ე.ი შეიძლება ადგილი ჰქონდეს არასრულყოფილ ჰიბრიდიზაციას და ყოველივე ამან შეიძლება გამოიწვიოს შეცდომები. სხვა სიმწელებები შეიძლება იყოს დაკავშირებული საკვლევი დნმ-ს თანამიმდევრობაში k-ზონდის კომპლემენტარული უბნის მრავალჯერად არსებობასთან.

დღესდღეობით მათემატიკური პრობლემები ძირითადად უკავშირდება ექსპერიმენტულ შეზღუდვებს, კერძოდ, გამოსაყენებლად შესაძლებელი ზონდის სიგრძეზე შეზღუდვებს. ასე მაგალითად, k-ზონდთა სრული მატრიცა უნდა მოიცავდეს 4^k ლაქას (ელემენტს), რაც იწვევს სიგრძეზე ზღვრის არსებობას. ამასთან მცირე სიგრძის ზონდთა შემთხვევაში შეიძლება არც არსებობდეს ცალსახა მათემატიკური ამონახსნი, ასე მაგალითად, $k=2$ ზონდის შემთხვევაში პრაქტიკულად შეუძლებელია თანამიმდევრობის განსაზღვრა. დღეისათვის $k=8$ სიგრძის ზონდთაგან შემდგარი მატრიცა უკვე კონსტრუირებულია და მიღების პროცესშია $k=10$ ზონდიანი მატრიცა. ამრიგად, მიმდინარეობს მეთოდის დახვეწის პროცესი.

ყოველივე ზემოთქმულის საფუძველზე შეიძლება დავასკვნათ, რომ სეკვენირების მეთოდიკის დახვეწა-განვითარების დინამიკა აშკარად მიუთითებს მეთოდიკის ექსპერიმენტული ნაწილის გამარტივების და შესაბამისად, უკანასკნელ ნაწილთან მიმართებაში ასამბლირების, ანუ კომპიუტიზირებული პროცედურების წილის ზრდის ტენდენციაზე.

ფარმაკოგენომიკა და ბიოინფორმატიკა.

რას წარმოადგენს ფარმაკოგენომიკა და როგორია მისი ამოცანა? ფარმაკოგენომიკის ამოცანაა ინდივიდის გენეტიკური თავისებურებების შესახებ მონაცემებზე დაფუძნებით კონკრეტულ წამალზე ორგანიზმის მოსალოდნელი საპასუხო რეაქციის წინასწარმეტყველება. ფარმაკოგენომიკის განვითარება იძლევა იმედს, რომ მომავალში შესაძლებელი გახდება პერსონალურად ინდივიდისათვის წამლის შემუშავებისა თუ დამზადების პრაქტიკის დანერგვა. ანუ წამალი ადაპტირებული იქნება ინდივიდის გენეტიკურ მონაცემებთან. მიუხედავად იმისა, რომ გარემო, დიეტა, ასაკი, ცხოვრების წესი, თუ ჯარმთელობის მდგომარეობა გავლენას ახდენს წამლის მიმართ ინდივიდის

პასუხზე, თამამად შეიძლება ითქვას, რომ უსაფრთხო და ეფექტური წამლის შემუშავების წინაპირობას უდავოდ წარმოადგენს ინდივიდის გენეტიკური თავისებურებების გათვალისწინება. ფარმაკოგენომიკა, ტრადიციულ ფარმაკოლოგიურ, ვთქვათ ბიოქიმიურ მიდგომებთან ერთად, იყენებს გენების, ცილებისა და ერთეულოვან-ნუკლეოტიდური პოლიმორფიზმების შესახებ ანოტირებულ ცოდნას.

დღეს ფარმაკოგენომიკა იშვიათად, მაგრამ მაინც გამოიყენება. ამის მაგალითად გამოდგება ფერმენტი TPMT (თიოპურინ მეთილტრანსფერაზა), რომელიც მეტად მნიშვნელოვან როლს თამაშობს ბავშვთა ლეიკემიის ქიმიოთერაპიული მეთოდით მკურნალობის პროცესში. იგი მონაწილეობს თერაპიული მიზნით გამოყენებული ნაერთების, თიოპურინების მეტაბოლიზმში. კავასიელთა მცირე პროცენტში გვხვდება შესაბამის გენთა ისეთი ვარიანტები, რომლებიც ვერ აწარმოებენ საჭირო ცილის აქტიურ ფორმას. ამის შედეგად, ასეთ პაციენტებზე ქიმიოთერაპიული ზემოქმედებისას თიოპურინები აჭარბებენ ტოქსიკურ ზღვარს. დღეისათვის ექიმი ვალდებულია გამოიყენოს გენეტიკური ტესტი, რათა გამოავლინოს პაციენტში შესაბამისი გენის ნაკლებობა, ანდა მოახდინოს TPMT-ს აქტივობის მონიტორინგი და ამის საფუძველზე შეარჩიოს ქიმიოთერაპიული საშუალებების (ამ შემთხვევაში თიოპურინების) საჭირო დოზები.

მკურნალმა უნდა იცოდეს, რომ ადამიანებს საკმაოდ განსხვავებული რაექცია შეიძლება გააჩნდეთ ისეთ წამლებზე როგორებიცაა ტკივილგამაყუჩებლები, ანტიდეპრესანტები, წნევის მარეგულირებლები, ან ასთმის საწინააღმდეგო საშუალებები. ეს განსხვავებები, შეიძლება უკავშირდებოდეს მცირეოდენ ცვლილებებს დნმ-ს თანამიმდევრობაში. ხშირად ამ ცვლილებების ცოდნა გვეხმარება დოზათა სწორ შერჩევაში. რიც შემთხვევებში გარკვეული წამლის გამოყენება შეიძლება ერთი პაციენტის სამკურნალოდ და დაუმეგებელია მეორისათვის. გენეტიკური თავისებურებების განსხვავებიდან გამომდინარე წამლებზე განსხვავებული რეაქციის მონიტორინგი აწარმოეს აშშ-ში და შედეგები ასეთი იყო: გამოყენებულ მედიკამენტთა მიმართ უარყოფითი რეაქციის დიაგნოზით სამ წელწინადში ჰოსპიტალიზირებულ იქნა 1 მილიონი ადამიანი, რომელთაგანაც 100 000 გარდაიცვალა.

ადამიანის გენეტიკური აპარატის ფუნქციონირების თავისებურებების შესახებ ცოდნის ზრდასთან ერთად ექიმებს ექმნებათ შესაძლებლობა შეიმუშავონ ინდივიდის პერსონალურ გენეტიკაზე დაფუძნებული მკურნალობა. ცოდნის დაგროვება, რომელიც შექმნის საფუძველს მკურნალობის პროცესში პერსონალური დანიშნულების შემუშავებისათვის, წარმოადგენს „ფარმაკოგენომიკის“ ძირითად ამოცანას.

როგორ შეიძლება მოახდინოს გენმა წამლის მიმართ ადამიანის რეაქციაზე გავლენა? გამომდინარე იქიდან, რომ მთელი ინსტრუქცია თითოეული ცილის

მოქმედების შესახებ კონცენტრირებულია შესაბამის გენში, მცირედმა ცვლილებამ გენის ნუკლეოტიდურ თანამიმდევრობაში შეიძლება იქონიოს მნიშვნელოვანი გავლენა ცილის ფორმასა და ფუნქციაზე. წამლის პროცესინგზე (გადამუშავებაზე) პასუხისმგებელი ცილის შესაბამის გენში ცვლილების შედეგად შეიძლება მივიღოთ ისეთი ცილა, რომელსაც დაკარგული აქვს მოცემული წამლის აქტივაციის უნარი, ან არასაკმარისად ააქტიურებს მას. სხვა ცილებში ცვლილებებმა შეიძლება გამოიწვიოს წამლის დეაქტივაცია ან მისი დაშლა მანამდე, სანამ ის თავისი ფუნქციის შესრულებას მოასწრებდეს, ან პირიქით გამოიწვიოს დაშლის უნარის დაკარგვა და ამით ხელი შეუწყოს წამლის სისხლში დარჩენასა და არასასურველი გვერდითი ეფექტების წარმოქმნას. აქედან გამომდინარე ნათელია, რომ სწორედ შესაბამისი ცილების შესაბამის გენებში მიმდინარე ცვლილებები განაპირობებს ორგანიზმის წამალზე პასუხის სახეს და სწორედ ამ ცვლილებათა ანალიზის საფუძველზე შეიძლება ავხსნათ მოცემული წამლის მიმართ ერთი ინდივიდის საპასუხო რეაქციის განსხვავება მეორე ინდივიდის რეაქციისაგან. ამიტომ მომავალში, როდესაც ექიმს ექნება შესაძლებლობა იწინასწარმეტყველოს ორგანიზმის რეაქცია სხვადასხვა წამალზე, თითოეული პაციენტისთვის პერსონალურად შესაძლებელი იქნება წამლისა და დოზის შერჩევა. ასე მაგალითად, თუ თქვენმა ექიმმა იცის, რომ თქვენს ორგანიზმს არ შეუძლია საჭიროებისამებრ დაამუშავოს კოდეინი, მაშინ იგი ეცდება შეგიჩიოთ სხვა, თქვენთვის უფრო გამოსადეგი ტკივილგამაყუჩებელი. ნათელია, რომ უახლოეს მომავალში თითოეული ინდივიდის სამედიცინო ფურცელში მოცემული იქნება პერსონალური გენეტიკური ინფორმაცია, რაც საშუალებას მისცემს ექიმს სწორად და სწრაფად შეარჩიოს მოცემული პაციენტისათვის ეფექტური წამალი.

მოვიყვანოთ მაგალითი. ღვიძლის ფერმენტთა, ციტოქრომ P450 ოჯახის წევრები ქმნიან გენების ჯგუფს, რომლებიც მნიშვნელოვან როლს თამაშობენ ღვიძლში წამლებისა და ტოქსინების ეფექტურად გადამუშავების საქმეში. ერთ-ერთი P450, რომელიც ცნობილია CYP2D6 სახელით, არსებობს 2 ფორმით. 7-დან 10%-მდე კავკასიელებში გავრცელებულია ისეთი ფორმა CYP2D6-სა, რომელიც ეფექტურად ვერ ამუშავებს მთელ რიგს წამლებისას, მათ შორის კოდეინს, ზოგიერთ ანტიდეპრესანტს, ანტიფსიქოტროპულ და გულის საშუალებებს. არსებობს დნმ-ტესტი, რომელიც განსაზღვრავს CYP2D6 გენის ფორმას და შესაბამისად თუ რა წამალი და რა დოზით იქნება ეფექტური მოცემული პაციენტისათვის. დღეს დანიშნულების განსაზღვრისას ექიმები იზვიათად ეყრდნობიან პაციენტისათვის CYP2D6-ზე ჩატარებული გენეტიკური ტესტის შედეგებს. უფრო ხშირად იყენებენ მარტივ, „გამოცდისა და შეცდომის უკუგდების“ მიდგომას, ანუ არასასურველი შედეგის შემთხვევაში ცვლიან დანიშნულებას.

რა უშლის ხელს ფარმაკოგენომიკის განვითარებას? 1) წამალზე პასუხის განმსაზღვრელ გენთა და მათი ვარიანტების სიმრავლე, რაც განპირობებულია მათ შორის ერთეულოვან-ნუკლეოტიდური პოლიმორფიზმით (SNP). ერთეულოვან-ნუკლეოტიდური პოლიმორფიზმი გულისხმობს ცვლილებებს,

რომელთა დროსაც გენომური თანამიმდევრობის მხოლოდ ცალკეული, ერთეულოვანი ნუკლეოტიდია შეცვლილი. SNP გვხვდება ყოველ 100-დან 300-მდე ფუძე-წყვილის სიგრძის ინტერვალში. ადამიანის გენომში, რომელიც 3 მილიარდი ფუძე წყვილის რიგისაა, იდენტიფიცირდება მილიონობით SNP. საჭიროა შეფასდეს თითოეული მათგანის უნარი, იყოს მიზეზი წამალზე ორგანიზმის პასუხისა. შეფასებების გაკეთება შესაძლებელია ძირითადად ბიოინფორმატიკული მიდგომებით.

2) არასაკმარისი ცოდნა იმის შესახებ, თუ რომელი გენებია პასუხისმგებელი მოცემულ წამალზე. ამასთან, ასეთ გენთა რაოდენობა შეიძლება იყოს დიდი და მათში დაფიქსირებულ ვარიაციათა საერთო სურათი მეტად ძნელად გასაანალიზებელი. აქაც უდავოა ბიოინფორმატიკული მეთოდების გამოყენების საჭიროება.

3) ალტერნატიულ წამალთა მცირე სპექტრი, რამაც შეიძლება მიგვიყვანოს შემდეგ სიტუაციამდე: პაციენტს დაუფიქსირდა გენთა ისეთი ვარიაცია, რომელიც ზღუდავს მოცემული წამლის გამოყენებას, მაგრამ არ არსებობს სხვა ალტერნატიული საშუალება. გამოსავალი სტრუქტურული ბიოინფორმატიკის მიდგომებზე დაყრდნობით ახალი საშუალებათა დაპროექტებაა.

4) შეზღუდვები, რომლებიც ამუხრუჭებენ ფარმაცოლოგიური კომპანიების მიერ სხვადასხვა ალტერნატიული ფარმაცოლოგიური საშუალებების წარმოებას. ფარმაცოლოგიური კომპანიების უმრავლესობა მოღვაწეობს ცალკეულ ფარმაცოლოგიურ სფეროში და ამ სფეროში წარმოებულ მედიკამენტთა ბაზარზე დამკვიდრებისათვის გაღებული მილიონიანი კაპიტალდაბანდებების გამო, მათ ეკარგებათ სურვილი იზრუნონ ალტერნატიურ, მოსახლეობის მხოლოდ მცირე ნაწილისათვის საჭირო მედიკამენტთა შემუშავებაზე.

5) ალტერნატიულ ფარმაცოლოგიურ საშუალებათა ფართო სპექტრის გამოყენება ართულებს მკურნალობის კურსის შერჩევის პროცედურას. იმის შესაფასებლად, თუ რომელი წამალია მისაღები მოცემული პაციენტისათვის, ექიმს მოუწევს დამატებითი პროცედურების ჩატარება. იმისათვის, რომ თითოეული პაციენტისათვის მკურნალობის სუკეთესო კურსი იქნეს შერჩეული, ნებისმიერი სპეციალიზაციის ექიმი კარგად უნდა ერკვეოდეს გენეტიკაში და უნდა ჰქონდეს შესაძლებლობა და უნარი მძლავრი და გამოსაყენებლად მოსახერხებელი მონაცემთა ბაზებითა და შესაბამისი პროგრამული უზრუნველყოფით სარგებლობისა.

რა სარგებელს უნდა ველოდოთ ფარმაცოლოგიისათვის? 1) მძლავრ ფარმაცეპტულ კომპანიებს შეეძლებათ გენეტიკური და პათოლოგიური თავისებურებებიდან გამომდინარე აწარმოონ წამლები ცილოვან, ფერმენტულ და რნმ საფუძვლებზე. ეს ხელს შეუწყობს ისეთი ახალი წამლების წარმოებას,

რომლებიც იქნება უფრო მიზანმიმართული კონკრეტული დაავადების მიმართ. ასეთი მიზანმიმართულება არა მხოლოდ გააძლიერებს წამლის თერაპიულ ეფექტს სამიზნე უჯრედებზე, არამედ შეამცირებს არასასურველ, დამანგრეველ ეფექტებს ჯანსაღი უჯრედების მიმართაც.

2) ძველი, „გამოცდისა და შეცდომის უკუგდებაზე“ დაფუძნებული მიდგომის მაგივრად, ექიმები შეძლებენ გამოიყენონ გენეტიკურ ანალიზზე დაფუძნებული მიზანმიმართული თერაპია. გენეტიკური მიდგომის საფუძველზე მოინახება სწორი და ეფექტური სამკურნალო საშუალება, დაჩქარდება გამოჯანმრთელების პროცესი და გაიზრდება უსაფრთხოება, რადგანაც უარყოფითი რეაქცია წამალზე იქნება მოხსნილი.

3) შემუშავებულ იქნება წამალთა უფრო ზუსტი დოზის შესარჩევი მეთოდი. ანუ არსებული მეთოდი, რომელიც დოზის განსაზღვრისას ეყრდნობა ინფორმაციას წონისა და ასაკის შესახებ, შეიცვლება მეთოდით, რომელიც დაეფუძნება ინფორმაციას პაციენტის გენეტიკური თავისებურებების შესახებ - რამდენად აქვს უნარი ინდივიდის ორგანიზმს სრულად დაამუშავოს წამალი და რა დროში მოხდება მოცემულ ორგანიზმში წამლის მეტაბოლიზმი. ამით ცხადია გაიზრდება თერაპიის ეფექტურობა და შემცირდება გადამეტებული დოზირების ალბათობა.

4) დაავადების გენეტიკური კოდის ცოდნა ინდივიდს საშუალებას მისცემს შეარჩიოს ადეკვატური ცხოვრების წესი და გარემო პირობები უკვე ადრეული ასაკიდან და ამით აიცილოს ან შეამციროს რისკი ზოგიერთი გენეტიკური დაავადებისა. კონკრეტული დაავადების შესახებ ცოდნა საშუალებას მოგვცემს შეირჩეს ის დრო, როცა მკურნალობის დაწყება ყველაზე დიდ ეფექტს მოგვცემს.

5) შეიქმნება და გამოიცდება ახალი პრეპარატები. ფარმაცეპტული კომპანიები შეიმუშავენ სამკურნალო საშუალებებს გენომურ სამიზნეებზე დაყრდნობით. წამალთა გამოცდის პროცესი გააძვირდება, რადგანაც გამოცდა მოხდება მიზანმიმართულად, შესაბამის პოპულაციურ ჯგუფებზე. კლინიკური გამოცდების ფასი და რისკი შემცირდება. გამოცდა ჩატარდება მხოლოდ იმ ინდივიდებში, რომელთაც არ გააჩნიათ უკუზვენება მოცემულ წამალზე.

6) შემცირდება ჯანდაცვის ხარჯები. შემცირდება უკუზვენებები, გამოსაცდელ წამალთა რიცხვი, დრო რომელიც საჭიროა წამლის გამოსაცდელად, დრო რომლის განმავლობაშიც ავადმყოფმა უნდა მიიღოს წამალი, წამალთა რიცხვი, რომელიც უნდა გამოსცადოს ავადმყოფმა, რათა შეირჩეს ოპტიმალური საშუალება, ავადმყოფობის ხანგრძლივობა (ადრეული გამოვლენის ხარჯზე). გაიზრდება გამოსაყენებლად შესაძლო სამკურნალო საშუალებათა სპექტრი. ყველაფერი ეს, საერთო ჯამში, უდავოდ გამოიწვევს ჯანდაცვის ხარჯების შემცირებას.

ამრიგად წარმოვადგინეთ რა ფარმაკოგენომიკის განვითარების გზაზე არსებულ პრობლემათა ნუსხა და განვიხილეთ ფარმაკოგენომიკის როლი ეფექტური და იაფი ჯანდაცვის სისტემის ჩამოყალიბების საქმეში, კიდევ უფრო თვალნათელივ გამოჩნდა სეკენირების სწრაფი და იაფი პროცედურების, მძლავრი და გამოსაყენებლად მოსახერხებელი მონაცემთა ბაზების, ეფექტური ბიონფორმატიკული მეთოდების ფლობის მწვავე აუცილებლობა. უდავოა, რომ თითოეული პაციენტისათვის ინდივიდუალურ მიდგომებზე დაფუძნებით ზუსტი დიაგნოზის დასმა და მკურნალობის ოპტიმალური კურსის შერჩევა მძლავრი ბიონფორმატიკული ბაზისის არ არსებობის პირობებში შეუძლებელია.

წასაკითხად რეკომენდირებული ლიტერატურა

Gary A. Churchill, (2002) „Fundamentals of experimental design for cDNA microarrays“ Nature Publishing Group <http://www.nature.com/naturegenetics>

Sanger F, Nicklen S, Coulson AR., (1977) „DNA sequencing with chain-terminating inhibitors“, Proc Natl Acad Sci U S A., 74(12):5463-7

Richard E. Green et. al, (2006) „Analysis of one million base pairs of Neanderthal DNA“ *Nature* 444, 330-336, 16 .

Wheeler, D. et al. (2008) “The complete genome of an individual by massively parallel DNA sequencing.” *Nature* 452.

Phillips KA, Veenstra DL, Oren E, Lee JK, Sadee W. (2001) „Potential role of pharmacogenomics in reducing adverse drug reactions: a systematic review.“, *JAMA* 286:2270-2279.

Weinshilboum R., Inheritance and Drug Response., (2003), *New Engl J Med* 348:529-537.

Abbott A., (2003) “ With your genes? Take one of these, three times a day.” *Nature* 425:760-762.

თავი 2

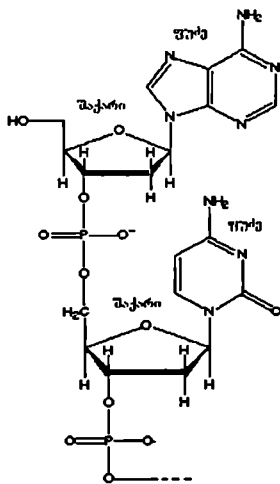
ბიომაკრომოლეკულათა სივრცული მოწყობა.

4. ნუკლეინის მჟავათა სტრუქტურული ორგანიზაციის ზოგადი პრინციპები.

დნმ -ს სტრუქტურა

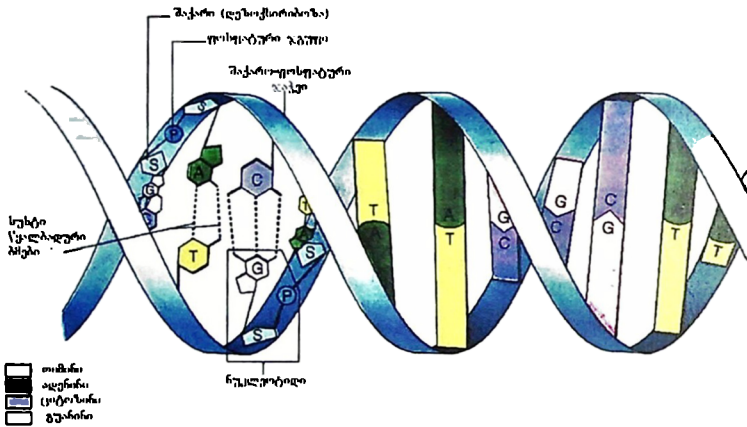
ქიმიური სტრუქტურა

დნმ წარმოადგენს გრძელ პოლიმერს, რომლის განმეორებად ერთეულს ეწოდება ნუკლეოტიდი. დნმ-ს სივანე 20–26 Å-ს (2.0-2.6 ნმ)) რიგისაა, ხოლო ნუკლეოტიდის სპირალის ღერძზე პროექციის სიგრძე მხოლოდ 3.4 Å (0.34 ნმ) რიგის. მაგრამ გამომდინარე იქიდან, რომ დნმ პოლიმერი მოიცავს მილიონობით ნუკლეოტიდს, იგი გიგანტური სიგრძის მოლეკულას წარმოადგენს. ადამიანის უდიდესი ქრომოსომა შედგება 220 მილიონი ფუძეწყვილისაგან. ცოცხალ ორგანიზმში დნმ გვხვდება არა ცალკეული ჯაჭვის, არამედ ერთმანეთთან მჭიდროდ დაკავშირებული ჯაჭვთა წყვილის სახით. ორი გრძელი ჯაჭვი ერთმანეთს ეხვევა და ქმნის ორმაგ სპირალს. თითოეული მოლეკულის შაქრის და ფოსფატურ ნაშთთაგან შემდგარი ნაწილი ქმნის პოლიმერის



სურათი 4.1 დნმ-ს ქიმიური სტრუქტურა

ძირითად ჯაჭვს. სხვადასხვა ჯაჭვის ფუძეები ურთიერთქმედებენ ერთმანეთთან. შაქართან დაკავშირებულ ფუძეს ეწოდება ნუკლეოზიდი, ხოლო ფუძეს, მასთან დაკავშირებულ შაქარსა და ერთ ან რამდენიმე ფოსფატს, ნუკლეოტიდი. როცა მრავალი ნუკლეოტიდი უკავშირდება ერთმანეთს, ისე როგორც ეს ხდება დნმ-ში, მიიღება პოლინუკლეოტიდი. როგორც უკვე აღვნიშნეთ დნმ-ს ცალკეული ჯაჭვი იქმნება ფოსფატთა და შაქართა ნაშთთაგან. დნმ-ში შაქარს წარმოადგენს პენტოზა (ზუთ ნახშირბადიანი შაქარი), 2-დეოქსირიბოზა. შაქრები ერთმანეთს უკავშირდება ფოსფატური ჯგუფებით. მეზობელ შაქართა მესამე და მეხუთე ნახშირბადებს შორის იქმნება ფოსფორიეთერული ბმები. სწორედ უკანასკნელი ბმის ასიმეტრიულობა განაპირობებს იმას, რომ დნმ-ს ჯაჭვს ახასიათებს მიმართულება.



სურათი 4.2 დნმ-ის ორმაგი სპირალის სტრუქტურული წარმოდგენა

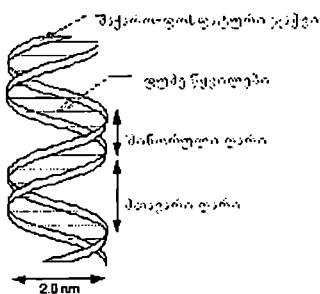
ორმაგ სპირალში, ერთ ჯაჭვში ნუკლეოტიდთა განლაგების მიმართულება საწინააღმდეგოა მეორე ჯაჭვის მიმართულებისა. ჯაჭვთა ასეთ მოწყობას უწოდებენ ანტიპარალელურს, დნმ-ს ჯაჭვთა ასიმეტრიულ ბოლოებს კი, 5'-ს (ზუთ შტრიხს) და 3'-ს (სამ შტრიხს). დნმ და რნმ ერთმანეთისაგან განსხვავდება შაქრებით. კერძოდ, რნმ-ში 2-დეოქსირიბოზის ნაცევლად ვხვდებით ასევე პენტოზურ შაქარს - რიბოზას.

დნმ-ს ორმაგი სპირალი სტაბილიზირებულია სხვადასხვა ჯაჭვის მოპირდაპირე ფუძეებს შორის არსებული წყალბადური ბმებით. დნმ-ში ვხვდებით ოთხი სახის ფუძეს. ესენია ადენინი (შეოკლებით A), ციტოზინი (C), გუანინი (G) და თიმიინი

(T) (იხ. სურათი 4.4). ფუძეები ორი ტიპისაა: ადენინი და გუანინი ხუთ- და ექვსწევრა რგოლების შერწყმის შედეგად წარმოქმნილი ჰეტეროციკლებია და ეწოდებათ პურინები, ციტოზინი და თიმინი კი ექვსწევრა რგოლებია და ეწოდებათ პირიმიდინები. მეხუთე პირიმიდინი, სახელად ურაცილი, როგორც წესი, გვხვდება რნმ-ში თიმინის ნაცვლად. მას თიმინისაგან განსხვავებით რგოლში მეთილის ჯგუფი არ აქვს. ჩვეულებრივ ურაცილი არ გვხვდება დნმ-ში, მაგრამ არსებობს იშვიათი გამონაკლისებიც. მაგალითად, ბაქტერიული ვირუსის PBS1-ის დნმ შეიცავს ურაცილს. ზოგიერთი რნმ-ს მოლეკულის სინთეზის პროცესში ადგილი აქვს ზოგიერთ ურაცილთა თიმინად გარდაქმნას მეთილის ჯგუფის დამატების გზით. ასეთ მოვლენებს ძირითადად ვხვდებით სტრუქტურულ და ფერმენტულ რნმ-ებში (როგორებიცაა, მაგალითად, რიბოსომული და ტრანსპორტული რნმ).

მთავარი და მინორული ღარი

დნმ-ს ორმაგი სპირალი წარმოადგენს მარჯვენა სპირალს. ჯაჭვების ურთიერთ დახვევისა და ორმაგი სპირალის ფორმირებისას ჯაჭვებს შორის წარმოიქმნება გარკვეული ღრმულები (ღარები), რომელთა საშუალებითაც შესაძლებელია



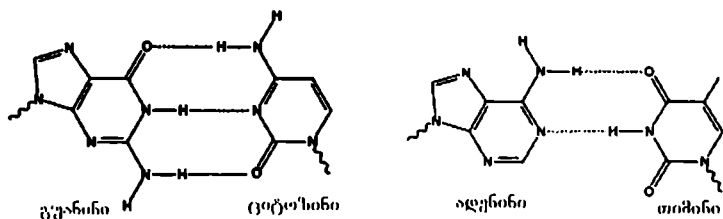
სურათი 4.3 ორმაგი სპირალის მთავარი და მინორული ღარი

სპირალის სპეციფიკურ ფრაგმენტებს უკავშირდება მთავარი ღარის მხრიდან რეალიზებული კონტაქტების საშუალებით.

ფუძეებამდე შეღწევა. არჩევენ ორ ასეთ ღარს, რომლებიც მიუყვება ორმაგი სპირალის მთელ ზედაპირს. ღარს, რომლის სიგანე დაახლოებით 22 Å-ა ეწოდება მთავარი ღარი, ხოლო მეორეს, რომლის სიგანე დაახლოებით 12 Å-ა, მინორული ღარი. მინორული ღარის სივიწროვე მიანიშნებს იმაზე, რომ ფუძეები მთავარი ღარის მხრიდან უფრო მისაწვდომია. შესაბამისად, დნმ-თან დაკავშირებული ცილები, მაგალითად ტრანსკრიფციის ფაქტორები, ორმაგი

ფუძეთა შორის კავშირები

ფუძე წყვილთა ტიპები განსხვავდება შესაძლო წყალბადური კავშირების რაოდენობით. A და T ქმნის ორ წყალბადურ ბმას, ხოლო G და C სამს. შესაბამისად, GC ფუძე წყვილი უფრო სტაბილურია და სწორედ GC წყვილთა რაოდენობა განაპირობებს ორმაგი სპირალის სტაბილურობას. ამრიგად, ერთი ჯაჭვის კონკრეტული ფუძე ქმნის გარკვეულ ბმათა ქსელს მეორე ჯაჭვის მხოლოდ კონკრეტული ტიპის ფუძესთან, რასაც უწოდებენ კომპლემენტარულ დაწყვილებას. ჩვეულებრივ ამ დაწყვილების პროცესში პურინები ქმნიან წყალბადურ ბმებს პირიმიდინებთან. ამასთან A აუცილებლად უწყვილდება T-ს და C კი G-ს (იხ. სურათი 4.4). ორმაგი სპირალის მასტაბილიზირებელი ფაქტორებია აგრეთვე ჰიდროფობული ეფექტი და π -სტეკინგი, რომლებიც არაა დამოკიდებული დნმ-ს კონკრეტულ ნუკლეოტიდურ თანამიმდევრობაზე. წყალბადური ბმა არაკოვალენტური კავშირია, ამიტომ ის ადვილად წყდება და კვლავ წარმოიქმნება.



სურათი 4.4 კომპლემენტარულ ფუძეთა შორის შესაძლო წყალბადური ბმები

აქედან გამომდინარე, მექანიკური თუ ტემპერატურული ზემოქმედების საშუალებით ადვილად შესაძლებელი "ელვას" ტიპის ორმაგი სპირალის დაშლა ცალკეულ ჯაჭვებად. კომპლემენტარულობიდან გამომდინარე, ინფორმაცია თავმოყრილი ერთ ჯაჭვში სრულად დუბლირებულია მეორე ჯაჭვით, რაც მეტად მნიშვნელოვანია რეპლიკაციის პროცესის განხორციელებისათვის. კომპლემენტარულობით განპირობებული ფუძე წყვილთა აღდგენის უნარი განპირობებული მათი ურთიერთქმედების სპეციფიკურობით დნმ-ს ფუნქციონირების საფუძველია.

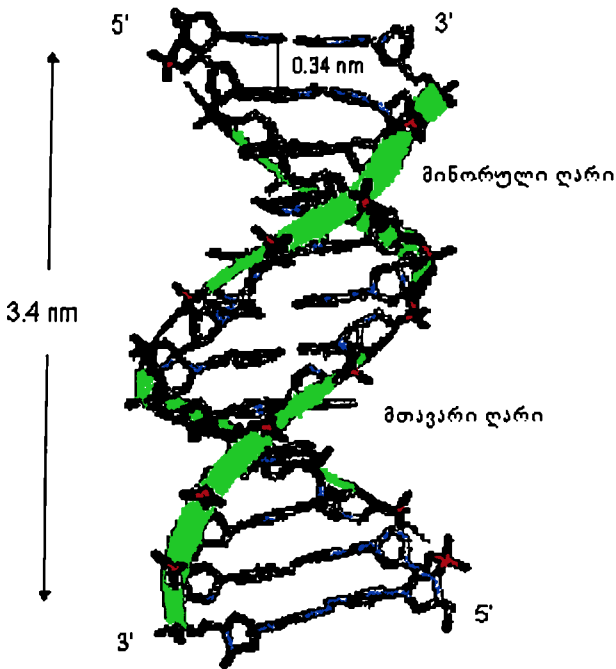
GC ფუძე წყვილებით მდიდარ დნმ სპირალს ახასიათებს ჯაჭვებს შორის ძლიერი ურთიერთქმედება, მაშინ, როცა AT ფუძე წყვილთა დიდი შემცველობა არ მიუთითებს სპირალის მაღალ სტაბილურობაზე. ამიტომ, რომ დნმ-ს ის უზანი, რომელიც საჭიროა რომ ადვილად განცალკავდეს ცალკეულ ჯაჭვებად (მაგალითად ბაქტერიული პრომოტორული უზნისათვის დამახასიათებელი TATAAT პრიზნოვის ბოქსი), ხასიათდება AT ფუძე წყვილების დიდი კონცენტრაციით. ლაბორატორიული კვლევებისას, ჯაჭვებს შორის ურთიერთქმედების ძალის შესაფასებლად ახდენენ ორმაგი სპირალის ტემპერატურულ ლღობას და აფასებენ ლღობის ტემპერატურასა და ენტალპიას. როცა ყველა კომპლემენტარული ფუძე წყვილის დაშორება მოხდება, ანუ დნმ სრულად გალღვება, ჯაჭვები ერთმანეთს ცილდება და ხსნარში არსებობენ დამოუკიდებელი მოლეკულების სახით. ასეთ, ერთ-ჯაჭვა დნმ მოლეკულებს არ გააჩნია ერთი კონკრეტული სივრცული სტრუქტურა.

ალტერნატიული ორმაგ-სპირალური სტრუქტურები

დნმ შესაძლებელია იმყოფებოდეს რამდენიმე განსხვავებულ კონფორმაციაში. კონფორმაციები რომელთა იდენტიფიკაცია დღემდე მოხერხდა შემდეგია: A-DNA, B-DNA, C-DNA, D-DNA, E-DNA, H-DNA, L-DNA, P-DNA და Z-DNA. ამასთან, ბუნებრივ ბიოლოგიურ სისტემებში აღმოჩენილ იქნა მხოლოდ სამი მათგანი: A-DNA, B-DNA, და Z-DNA. კონფორმაციის სახეს განაპირობებს დნმ-ს ნუკლეოტიდური თანამიმდევრობა, სუპერსპირალიზების ხარისხი, ფუძეების ქიმიური მოდიფიკაციები, გარე პირობები, მაგალითად, მეტალთა იონების და პოლიამინების კონცენტრაცია. სამი ბუნებრივი კონფორმციიდან ცოცხალი უჯრედისათვის დამახასიათებელ პირობებში ყველაზე სტაბილურია B- ფორმა. ორმაგ-სპირალური ფორმები ერთმანეთისაგან განსხვავდება გეომეტრიითა და ზომებით.

B ფორმის შემთხვევაში (იხ. სურათი 4.5) სპირალის ბიჯი ანუ ერთ სრულ ხვიაზე მოსული წანაცვლება სპირალური ღერძის გასწვრივ უდრის 3.4 ნმ, ხოლო ორ მეზობელ ფუძეს შორის დაშორება 0.34 ნმ.. ერთ ხვიაზე მოსული ფუძეების რიცხვი 10-ს ტოლია. როგორც უკვე აღვნიშნეთ ერთმანეთთან სპირალურად დახვეული ორი ჯაჭვი წარმნოქმნის ორ ღარს - ძირითადს და მინორულს, რომლებიც განაპირობებს ცილებთან ურთიერთქმედების შესაძლებლობას.

მარილის კონცენტრაციის გაზრდის, ან სპირტის დამატების შემთხვევაში დნმ-ს სტრუქტურა იცვლება და B ფორმა გადადის A-ში. A ფორმა ასევე მარჯვენა სპირალია, მაგრამ ამ შემთხვევაში ბიჯი უტოლდება 2.3 ნმ, ხოლო ერთ ხვიაში განთავსებულ ფუძეთა რიცხვი კი 11-ს.

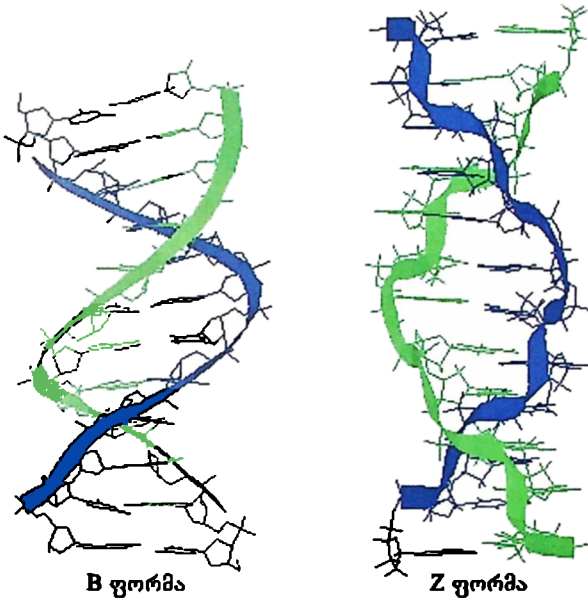


სურათი 4.5 B ფორმის დნმ-ს სივრცული სურათი

A ფორმა ყველაზე განიერი მარჯვენა სპირალია, განიერი მინორული ღარით და შედარებით ვიწრო მთავარი ღარით. A ფორმა ძირითადად წარმოიქმნება არაფიზიოლოგიურ პირობებში. მაგალითად, დნმ-ს დეჰიდრატაციის შემთხვევაში. თუმცა იგი შეიძლება შეგვხვდეს უჯრედშიც, მაგალითად, დნმ-რნმ-ს ჰიბრიდულ ჯაჭვთა წყვილში, ან ფერმენტ-დნმ-ს კომპლექსში.

მეთილირების გზით ფუძეების ქიმიური მოდიფიცირების შემთხვევაში დნმ-ს სუბსტრუქტურული განივების მნიშვნელოვან კონფორმაციულ ცვლილებას და წარმოიქმნება Z სტრუქტურული ფორმა (იხ. სურათი 4.6). ამ შემთხვევაში ცალკეული ჯაჭვი ორმაგი სპირალის ღერძის ირგვლივ იხვევა მარცხენა სპირალის ფორმით, ანუ B

ფორმის საპირისპირო მიმართულებით. ასეთ უზვეულო სტრუქტურას ცნობს სპეციალური, Z-DNA-თან დაკავშირებადი ცილა და არსებობს მოსაზრება, რომ ეს პროცესი მონაწილეობას იღებს ტრანსკრიპციის რეგულირებაში. Z ფორმა წარმოადგენს მარცხენა სპირალს, რომლის ბიჯია 4.6 ნმ და ერთ ხვიაში ფუძეთა რიცხვი - 12. სპირტის ან მარილის მაღალი კონცენტრაციის პირობებში G-C დინუკლეოტიდების განმეორების შედეგად მიღებული დნმ-ს მოლეკულის სეგმენტი ასევე გადადის Z ფორმაში.



სურათი 4.6 B და Z ფორმის დნმ-თა სივრცული სურათები

კვადრუპოლური სტრუქტურა

სახეობების უმრავლესობისათვის გენომური თანამიმდევრობის მხოლოდ მცირე ნაწილი აკოდირებს ცილებს. მაგალითად, ადამიანის გენომის მხოლოდ 1.5% წარმოადგენს ცილის მკოდირებელ ნაწილს, მაშინ, როცა არამკოდირებელი, განმეორებების შემცველი, უზნების წილი გენომური დნმ-ს 50 %-ს შეადგენს. კითხვა, რატომ ვხვდებით ეუკარიოტულ გენომებში ასეთი დიდი რაოდენობით

არამაკოდირებელ უბნებს და სხვადასხვა ორგანიზმთა გენომების ზომებს შორის რატომაც ასეთი დიდი განსხვავება, წარმოადგენს „C-თავსატეხის“ სახელწოდებით ცნობილ, ჯერ ბოლომდე გადაუქრულ ამოცანას. თუმცა ნაწილობრივი პასუხები ამ კითხვაზე არსებობს. მაგალითად, შეიძლება აღვნიშნოთ, რომ დნმ-ს უბანი შეიძლება არ აკოდირებდეს ცილას, მაგრამ ატარებდეს გარკვეულ ფუნქციურ დატვირთვას. მაგალითად, აკოდირებდეს გენურ ექსპრესიაში არ მონაწილე (არამესენჯერულ) რნმ-ს. გარდა ამისა, ადამიანის გენომში არსებობს ე.წ ფსევდოგენები, რომლებიც წარმოადგენს არამაკოდირებელი უბნების ყველაზე გავრცელებულ ფორმას. ფსევდოგენები ასლებია იმ გენებისა, რომლებიც მუტაციურმა ცვლილებებმა ფუნქციურად მიუღებელი გახადა. ასეთი „სამარაგო“ უბნები გარკვეულ პირობებში შესაძლებელია გამოყენებულ იქნეს, როგორც ნახევარფაბრიკატები ახალ გენთა წარმოქმნისათვის.

ცალკე აღსანიშნავია ის არამაკოდირებელი უბნები, რომლებიც თამაშობენ სტრუქტურულ როლს. მაგალითად, ტელომერები და ცენტრომერები, რომლებშიც განთავსებულია მხოლოდ რამდენიმე (მცირე რაოდენობით) გენი და მათი ძირითადი დანიშნულებაა ქრომოსომის სტაბილიზაცია და გენომის სწორი ფუნქციონირებისათვის ხელშეწყობა. ტელომერები ხაზოვანი ქრომოსომების კიდურა, სპეციალიზირებული უბნებია. ტელომერთა ძირითად ფუნქციას



სურათი 4.7 კვადრუპოლური სტრუქტურა

წარმოადგენს ფერმენტ ტელომერაზას დახმარებით ქრომოსომას კიდურა უბნის რეპლიკაციისათვის ხელის შეწყობა, რადგან დნმ-ს მხოლოდ ტრადიციული რეპლიკაციური ფერმენტებით ქრომოსომას კიდურა 3' ბოლოს რეპლიკაცია შეუძლებელია. ქრომოსომას ეს დნმ-სპეციალური უბანი აგრეთვე იცავს დნმ-ს ბოლოებს ენდონუკლეაზებისაგან. ადამიანის უჯრედებში ტელომერი წარმოადგენს მარტივი თანამიმდევრობის, TTAGGG-ს, რამდენიმე ასულჯერადი განმეორების შედეგად წარმოებულ ერთჯაჭვა დნმ-ს მონაკვეთს. ეს გუანინით მდიდარი თანამიმდევრობა არ ქმნის კომპლემენტარულ ფუძე წყვილებს მოპირდაპირე ჯაჭვის ფუძეებთან. იგი წარმოქმნის უჩვეულო, ოთხი ფუძის სტეკინგის საფუძველზე ფორმირებულ სტრუქტურას და ამით ახდენს ქრომოსომას სტაბილიზაციას. ამ უკანასკნელ სტრუქტურაში ოთხი გუანინი განლაგებულია თითქმის ერთ სიბრტყეში. ასეთი განმეორებადი სიბრტყეების სტეკინგით ფორმირდება სტაბილური G-კვადრუპოლური სტრუქტურა (იხ. სურათი 4.7). ამ სტრუქტურის სტაბილიზაციას ახდენს ფუძეთა წიბოებსა და მეტალის იონს შორის წარმოქმნილი წყალბადური ბმები. მეტალის იონი განლაგდება ოთხ ფუძეს შორის, ცენტრში.

რნმ-ს სტრუქტურა

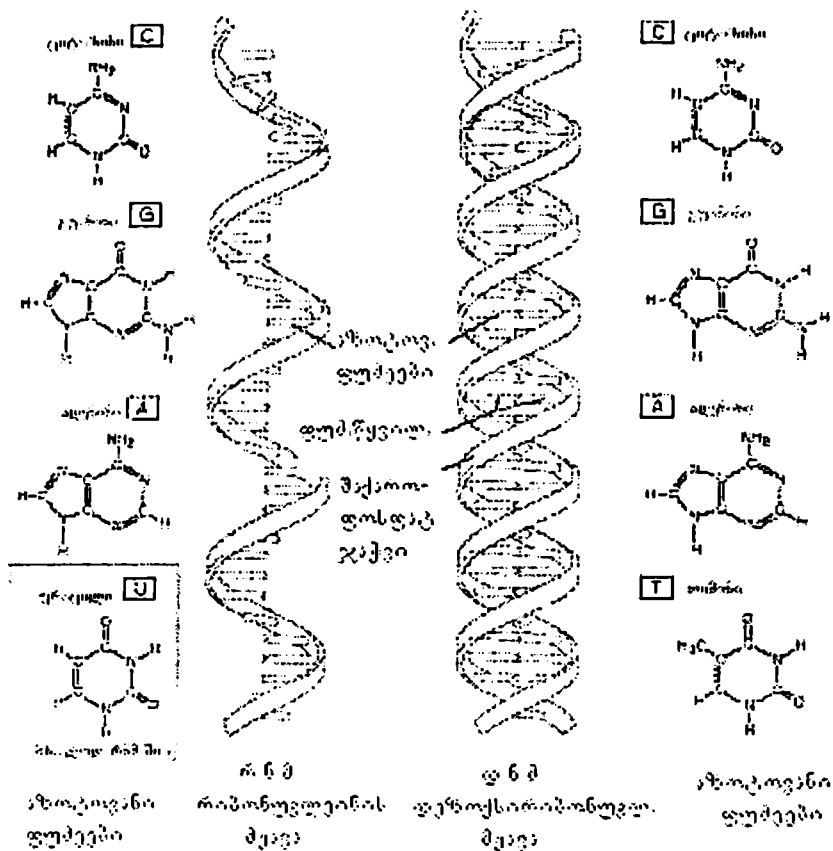
რნმ -ს ქიმიური სტრუქტურა

რნმ-ს დნმ-საგან გამოარჩევს სამი ძირითადი სტრუქტურული მახასიათებელი. უპირველეს ყოვლისა უნდა აღინიშნოს, რომ რნმ თავის ბიოლოგიურ ფუნქციას ძირითადად ანხორციელებს ერთ-ჯაჭვა მოლეკულის სახით. მორე, დნმ-საგან განსხვავებით, რომლის შემადგენლობაშიც შედის დეზოქსირიბოზა, რნმ შეიცავს რიბოზას (ანუ რნმ-ს პენტოზურ რგოლში გვხვდება ორი ჰიდროქსილის ჯგუფი, მაშინ, როცა დნმ-ს პენტოზური რგოლის 2' პოზიციაში ჰიდროქსილი არაა). შესაბამისად, სწორედ ჰიდროქსილის ზედმეტი ჯგუფი განაპირობებს რნმ-ს მოლეკულის ჰიდროლიზის სიადვილეს. და ბოლოს, მესამე განსხვავება ეხება ფუძე შემადგენლობას. როგორც აღვნიშნეთ დნმ-ში ადენინის კომპლემენტარულ ფუძეს წარმოადგენს თიმიინი, რნმ-ში კი - ურაცილი, რომელიც თიმიინის არამეთილირებულ ფორმას წარმოადგენს.

არამაკოდირებელი რნმ-ს სივრცული სტრუქტურა.

ერთჯაჭვა რნმ-ს ფუნქციური ფორმა ხშირ შემთხვევაში (ცილების მსგავსად) საჭიროებს სპეციფიკური მესამეული სტრუქტურის არსებობას, რაც განსაკუთრებით აქტუალურია არამაკოდირებელი რნმ-ების შემთხვევაში. მაკოდირებელი, ანუ ინფორმაციული რნმ, თავისი ფუნქციონალური თავისებურებებიდან გამომდინარე ძირითადად არ საჭიროებს სტაბილურ მესამეულ სტრუქტურას (ინფორმაციულ რნმ-ში მხოლოდ შინაგანი ანუ rho-დამოუკიდებელი ტერმინაციის დროს ხდება გარკვეული ტიპის მესამეული სტრუქტურის ფორმირება), მაშინ, როცა არამაკოდირებელ რნმ-თა ფუნქცია, როგორც წესი, განპირობებულია მათი სივრცული მოწყობის ტიპით.

რნმ გენები წარმოადგენს ცილად არატრანსლირებად, რნმ-თა მაკოდირებელ უბნებს. ასეთ რნმ-თა თვალსაჩინო წარმომადგენლებად შეგვიძლია მოვიყვანოთ ტრანსპორტული (tRNA) და რიბოსომული (rRNA) რნმ-ები. არამაკოდირებლებს



სურათი 4.8 რნმ მოლეკულის დნმ-გან განმასხვავებელი ნიშნები

მიკუთვნებენ ფართო სპექტრს მცირე ბირთვული რნმ-ებისას (snRNA) და ასევე მიკრო-რნმ-ებს (microRNA), რომლებიც ე.წ რნმ-ინტერფერირების მექანიზმით არეგულირებენ გენთა ექსპრესიის პროცესს. არამაკოდირებელ რნმ-თა მაგალითად გამოდგება სპლაისოსომის შემადგენლობაში მყოფი რნმ-ები. სპლაისოსომას ფუნქციაში შედის პრე-ინფ.რნმ-დან (pre-mRNA) ინტრონების ამოჭრა.

მართალია რნმ ძირითადად ერთჯაჭვა მოლეკულაა, მაგრამ უმრავლესობა ბიოლოგიურად აქტიური რნმ-ების (მათ შორის tRNA, rRNA, snRNAs და სხვა არამაკოდირებელი ფორმები) სივრცულ სტრუქტურებში ვხვდებით ორმაგ-

სპირალურ ფრაგმენტებს, რომლებიც წარმოიქმნება ერთი ჯაჭვის სხვადასხვა სეგმენტებში განთავსებულ კომპლემენტარულ ფუძეთა სივრცეში დაახლოებით და მათი დაწყვილებით (მათ შორის წყალბადურ ბმათა ქსელის წარმოქმნით). რნმ-თა სტრუქტურების შესწავლის შედეგად დადგინდა, რომ ისინი არ წარმოადგენენ უბრალო ხაზოვან მოლეკულებს, არამედ გააჩნიათ მაღალსტრუქტურირებული სივრცული მოწყობა. დნმ-ს ორმაგ-სპირალური, მარტივი სივრცული მოწყობისგან განსხვავებით, რნმ მოლეკულებში არის მოკლე ორმაგი სპირალები, რომელთა სივრცეში ურთიერთდაახლოების საფუძველზე წარმოიქმნება საკმაოდ რთული სივრცული სტრუქტურა. ამრიგად, სივრცული მოწყობის თავისებურებების თვალსაზრისით, რნმ-ები გვევლინება შუალედურ რგოლად დნმ-სა და ცილებს შორის. შესაბამისად, ბუნებრივია ის ფაქტი, რომ ფერმენტების მსგავსად ზოგიერთ რნმ-საც გააჩნია ქიმიური რეაქციის კატალიზირების უნარი. ასე მაგალითად, აღმოჩნდა, რომ რიბოსომათა მნიშვნელოვანი კომპონენტის, პეპტიდურ ბმათა ფორმირების მკატალიზირებელი ფერმენტის აქტიური ცენტრი მთლიანად რნმ-ს მოლეკულას წარმოადგენს (ამიტომ, რომ რიბოსომას უწოდებენ „რიბოზიმს“ - რნმ-ზე დაფუძნებულ ფერმენტს).

რნმ-ს სწორედ ასეთი შუალედური ბუნება დაედო საფუძველად ცოცხალი სამყაროს წარმოქმნის რნმ-ზე დაფუძნებულ ჰიპოტეზას. ამ ჰიპოტეზის თანახმად სიცოცხლის ადრეული ფორმები, როგორც გენეტიკური ინფორმაციის შესანახად, ისე ბიოქიმიურ რეაქციათა კატალიზატორებად იყენებდნენ რნმ მოლეკულებს. ევოლუციის პროცესში კი, მოხდა ამ ფუნქციათა გამოიჯვნა და ერთი გადაეცა დნმ-ს, მეორე კი, ცილებს.

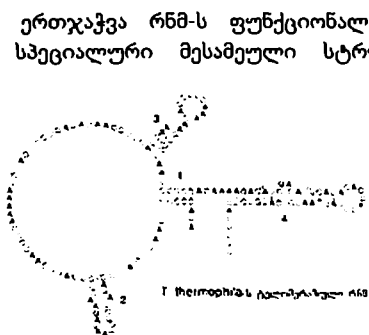
ორმაგ-სპირალური (ორჯაჭვა) რნმ-ები

მიუხედავად იმისა, რომ რნმ უმრავლეს შემთხვევაში ერთჯაჭვა მოლეკულაა, ვხვდებით ორჯაჭვა რნმ-ებსაც. ორჯაჭვა რნმ (dsRNA), რომელიც დნმ-ს მსგავსად ფორმირებულია ორი კომპლემენტარული ჯაჭვის საფუძველზე, ძირითადად გვხვდება მაღალგანვითარებულ ორგანიზმთა უჯრედებში. თუმცა ისიც უნდა აღინიშნოს, რომ ზოგიერთ ვირუსში გენეტიკური ინფორმაცია ინახება ორჯაჭვა რნმ-ს სახით.

ეუკარიოტებში ორჯაჭვა რნმ-ს (siRNA) მთავარ ფუნქციას წარმოადგენს რნმ-ინტერფერირების პროცესის ინიციაცია და მცირე ინტერფერატორი რნმ-ების წარმოქმნის პროცესში იგი გვევლინება შუალედურ პროდუქტად. ასეთი ფუნქციის გამო ხშირად ხდება siRNA და microRNA-ების არევა. უნდა გვახსოვდეს, რომ siRNA წარმოადგენს ორჯაჭვას, მაშინ, როცა microRNA (ან miRNA) ერთჯაჭვაა. თუმცა ეს უკანასკნელი თავის pre-miRNA მდგომარეობაში

ქმნის სარკის მსგავს სპირალურ სტრუქტურას. ახლახან გაირკვა, რომ dsRNA შეუძლია გენთა ექსპრესიის ინიციაცია. ამ მოვლენას უწოდებენ „მცირე რნმ-ით ინდუცირებად გენთა აქტივაციას“, ასეთ dsRNA-ებს კი, „აქტივაციის გამომწვევ მცირე რნმ-ებს (saRNA)“

რნმ-ს მეორეული სტრუქტურა



სურათი 4.9 რნმ-ს მეორეული სტრუქტურა

ერთჯაჭვია რნმ-ს ფუნქციონალური ფორმა ცილების მსგავსად მოითხოვს სპეციალური მესამეული სტრუქტურის არსებობას. ასეთი სტრუქტურის საფუძველს ქმნის მეორეული სტრუქტურა, რომელიც კომპლემენტარულ ფუძეებს შორის შიდამოლეკულური წყალბადური ბმების წარმოქმნის შედეგია.

შიდამოლეკულური წყალბადური ბმების წარმოქმნის შედეგად მიიღება ისეთი მკვეთრად გამოკვეთილი მეორეული სტრუქტურული ელემენტები, როგორებიცაა „სარკისებრი ყულფი“ (hairpin loops or stem-loop), ამოზნექილობა (bulges) ან „შიდა ყულფი“ (internal loops) (იხ. სურათი 4.9). სარკისებრი ყულფის ფორმირება განპი-

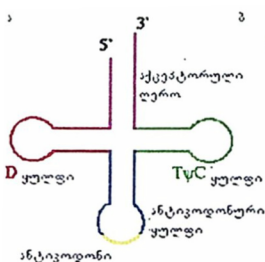
რობებულია სპირალური და ყულფის ნაწილების სტაბილურობით. სპირალის შექმნის წინაპირობაა ისეთი თანამიმდევრობის არსებობა, რომელიც მოხრისა და კიდურა სეგმენტების ერთმანეთთან ანტიპარალელურად შეთავსებისას წარმოქმნის კომპლემენტარულად შეწყვილებულ ორმაგ სპირალურ ფრაგმენტს. ასეთი სპირალური უბნის სტაბილურობა განისაზღვრება მისი სიგრძით, არაკომპლემენტარული წყვილების, ანუ ამოზნექილობების რიცხვით (გრძელ სპირალურ ფრაგმენტისათვის ასეთი ამოზნექილობების მცირე რიცხვი დასაშვებია) და სპირალური უბნის ფუძოვანი შემადგენლობით. ასე მაგალითად, გუანინისა და ციტოზინის შეწყვილება გვაძლევს სამ წყალბადურ ბმას და ე.ი. ამ წყვილს უფრო დიდი წვლილი შეაქვს სტაბილურობაში, ვიდრე ორ წყალბადურ ბმიან ადენინ-ურაცილის წყვილს. ფუძეთა შორის სტეკინგს, რომელიც შედეგია ფუძეთა არომატული რგოლების pi-ორბიტალების ხელსაყრელი ორიენტაციისა, ასევე შეაქვს თავისი წვლილი სპირალის სტაბილიზაციაში.

სარკისებრი ყულფის ჩამოყალიბებაში მნიშვნელოვანია ყულფის სტაბილურობაც. სამი ფუძისაგან შემდგარი „ყულფი“ სტერიულად მიუღებელია და ამიტომ არ გვხვდება. დიდი ყულფები, რომელთა შემადგენლობაში შემავალი ფუძეები საერთოდ არ მონაწილეობს მეორეული სტრუქტურის ჩამოყალიბებაში, ასევე

არასტაბილურია. ყულფის დასაშვები ოპტიმალური სიგრძე მერყეობს 4-სა და 8 ფუძეს შორის. ფუძეთა თანამიმდევრობა - TUCG ერთ-ერთ ყველაზე სტაბილურ და გავრცელებულ ყულფს ქმნის. მას უწოდებენ „ტეტრაყულფს“ და მისი სტაბილურობა ძირითადად განპირობებულია შემადგენელ ფუძეებს შორის სტეკინგით.

რნმ-ს სივრცული სტრუქტურა.

სარკისებრი ყულფები გვხვდება არამაკოდირებელ რნმ-ში, მათ შორის ტრანსპორტულ რნმ-ში. ტრანსპორტულ რნმ-ს სივრცული მოწყობა, მისი თანამიმდევრობის სიმოკლის გამო განსაკუთრებით კარგადაა შესაწავლილი. მასში შემავალი სამი ჭეშმარიტი სარკისებრი ყულფი და ერთი სპირალური ფრაგმენტი ქმნის სამწვერა ფოთლის მსგავს სტრუქტურას. კოდონის გამომცნობი ანტიკოდონი განთავსებულია ერთერთ სარკისებრ ყულფში, რომელიც არაა შეწყვილებული სხვა ყულფთან. ორი დანარჩენი შეწყვილებულია ერთმანეთთან



სურათი 4.10 ტრანსპორტული რნმ-ს ა) მოწყობის სქემა და ბ) მესამეული სტრუქტურა

„ფსევდო მარყუჟის“ წარმოქმნით, ანუ ერთ-ერთის ყულფის ნაწილი მონაწილეობს მეორის სპირალური ნაწილის ფორმირებაში. რიბოზიმების უმრავლესობას ახასიათებთ სარკისებრი ყულფების არსებობა. მათი არსებობა აუცილებელია სწორი სივრცული სტრუქტურის ჩამოსაყალიბებლად და ფერმენტული ფუნქციის შესასრულებლად.

მიუხედავად იმისა, რომ ინფორმაციულ რნმ-ს თავისი ფუნქციიდან გამომდინარე ზოგადად არ უნდა ახასიათებდეს რაიმე სტაბილური მეორეული ან მესამეული სტრუქტურა, არსებობს გამონაკლისებიც. მაგალითად, როცა ტრანსკრიპციის პროცესში საჭიროა შინაგანი ანუ rho- დამოუკიდებელი ტერმინაცია, რნმ-ში წარმოიქმნება ყულფი და იწვევს დნმ-დან რნმ-პოლიმერაზას ჩამოცილებას. ინფორმაციული რნმ-ს იმ თანამიმდევრობას, რომელსაც შეუძლია ასეთი ყულფის წარმოქმნა ეწოდება „ტერმინატორული“ თანამიმდევრობა. rho

წარმოადგენს ცილოვან ფაქტორს, რომელიც „ტერმინატორული“ თანამიმდევრობის“ არ არსებობის პირობებში უზრუნველყოფს დნმ-დან რნმ-ს ჩამოცილებას.

სარქისმაგვარ ყულფთა თანამიმდევრობების მაგალითად გამოდგება ე.წ. პოლინდრომული თანამიმდევრობა ---CCTGCXXXXXXXXXGCAGG---

ასეთ თანამიმდევრობას შესწევს უნარი ჩამოაყალიბოს შემდეგი სარქისმაგვარი ყულფი

```
---C G---
  C G
  T A
  G C
  C G
  X X
X   X
  X X
  X
  X
```

სხვა მაგალითად გამოდგება თანამიმდევრობა, რომელსაც ასევე გააჩნია უნარი სარქისებრი ყულფის ჩამოყალიბებისა. :

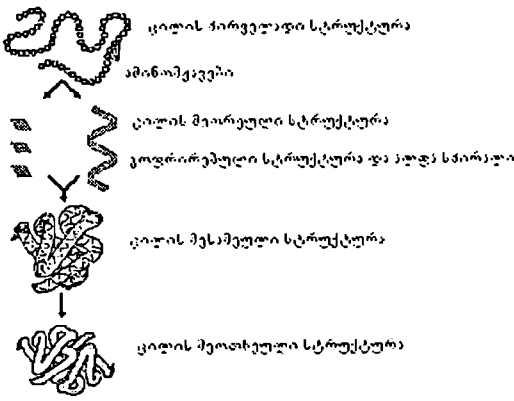
GCCGCGGGCCGAAAAAACCCCCCGGCCCGCGGC

ამრიგად შეიძლება ითქვას, რომ დნმ და რნმ-ს მოლეკულების კონფორმაციული შესაძლებლობები საკმაოდ მრავალფეროვანია. ჩვენ ავღწერეთ მხოლოდ ყველაზე გავრცელებული და ცოცხალი უჯრედისათვის მნიშვნელოვანი სივრცული სტრუქტურები. მაგრამ აქვე უნდა აღვნიშნოთ, რომ პოლიპეპტიდური ჯაჭვის და შესაბამისად, ცილის მოლეკულის კონფორმაციული შესაძლებლობები გაცილებით უფრო ფართოა, ვიდრე დნმ-სა და რნმ-ს. ეს ბუნებრივიცაა. დნმ და რნმ ხომ მხოლოდ ოთხი მონომერის, ხოლო ცილა (პოლიპეპტიდი) ოცი სხვადასხვა მონომერის საფუძველზე წარმოებული პოლიმერია. ალბათ სწორედ ეს წარმოადგენს იმის მიზეზს, რომ ცილების სტრუქტურული მრავალფეროვნება, მათზე მოსული ფუნქციონალური დატვირთვა ძალზედ დიდია. შესაბამისად გასაგები ხდება ის ფაქტი, რომ სტრუქტურულ-ბიოლოგიური კვლევების უმნიშვნელოვანეს ობიექტს წარმოადგენს სწორედ ცილები და მათი სივრცული ორგანიზების კანონზომიერებები.

5. ცილათა სტრუქტურული ორგანიზაციის ზოგადი პრინციპები

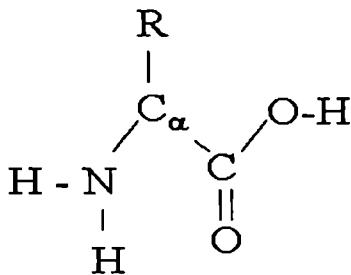
საშენი მასალა - ამინმჟავები

ცილები წარმოადგენს პოლიმერულ ჯაჭვებს, სადაც მონომერები ამინმჟავებია. ცილების ყველა სტრუქტურული თუ ფუნქციონალური თავისებურება განისაზღვრება პოლიპეპტიდური ჯაჭვის ქიმიური თავისებურებებით. განასხვავებენ ცილის სტრუქტურული ორგანიზაციის ოთხ დონეს: პირველადი, მეორეული, მესამეული და მეოთხეული (იხ. სურათი 5.1). პირველად სტრუქტურას განსაზღვრავს პოლიპეპტიდურ ჯაჭვში ამინმჟავათა თანამიმდევრობა. მეორეულ სტრუქტურაში იგულისხმება პოლიპეპტიდური ჯაჭვის მიერ ფორმირებული რეგულარული (გარკვეული სპირალური სიმეტრიის მქონე) გეომეტრიული ფიგურები, რომლებიც ახლო ურთიერთქმედებების შედეგია (იგულისხმება ჯაჭვის გასწვრივ ახლო). მესამეული სტრუქტურა შედეგია პოლიპეპტიდური ჯაჭვის შიგნით წარმოქმნილი შორეული კონტაქტებისა, მეოთხეული კი, ცილის სუბერთეულებს, ანუ ორ ან მეტ ჯაჭვს შორის ურთიერთქმედებისა.



სურათი 5.1 ცილის სტრუქტურული ორგანიზაციის ოთხი დონე

ამინმევა წარმოადგენს ცილის ძირითად ქიმიურ კომპონენტს. მისთვის

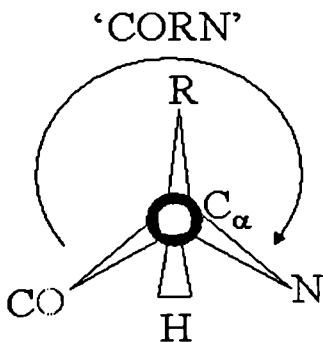


სურათი 5.2 α-ამინმევა

დამახასიათებელი ცენტრალური α-ნახშირბადი მიუთითებს პრიორიტეტულ პოზიციასზე, რომლის მიმართაც აღიწერება სხვა დამახასიათებელი ჯგუფების განლაგება. α-ნახშირბადს უკავშირდება ოთხი ჯგუფი (იხ. სურათი 5.2). ესენია: α-პროტონი (წყალბადი) H, გვერდითი ჯაჭვი R, რომელიც განსაზღვრავს ამინმევას ქიმიურ მრავალფეროვნებას, კარბოქსილური (მეავური) ფუნქციონალური ჯგუფი (COOH), და ასევე ფუნქ-

ციონალური ამინო-ჯგუფი (-NH₂). α-ნახშირბადი ოცივე ამინმევაში, გლიცინის გარდა, წარმოადგენს მოლეკულის ასიმეტრიულ ცენტრს. გლიცინის გვერდით ჯაჭვს, სხვა ამინმევაებისაგან განსხვავებით, წარმოადგენს წყალბადი.

რიბოსომაზე სინთეზისას აუცილებელი პირობაა, რომ α-ნახშირბადოვანი ცენტრის კონფიგურაცია შეესაბამებოდეს L-იზომერს (იხ. სურათი 5.3). ამის მიზეზი შეიძლება იყოს ადრეულ ეტაპზე განხორციელებული ქიმიური ევოლუციის პროცესი, როცა რაღაც მიზეზების გამო პრებიოტური სისტემისათვის სწორედ ეს კონფიგურაცია იყო უპირატესი, და შემდგომ ეს არჩევანი დაფიქსირდა და შენარჩუნდა მთელი ევოლუციური პროცესის განმავლობაში.



სურათი 5.3. ამინმევის იზომერული ფორმის განმსაზღვრელი CO-R-N წესი. იგი გულისხმობს, რომ თუ C_α-ს ვუყურებთ H ატომის მხრიდან მაშინ L-ფორმის ამინმევას შემთხვევაში CO-დან N-კენ საათის ისრის მიმართულებით მოძრაობისას გავივლით R-ზე.

გვერდითი ჯაჭვები გამოირჩევა ქიმიური მრავალფეროვნებით (იხ.სურათი 5.4), რაც ძალზედ მნიშვნელოვანია ცილათა ფუნქციური მრავალფეროვნების უზრუნველსაყოფად. ამასთან, შესაძლებელია გვერდით ჯაჭვთა სამ კატეგორიად დაჯგუფება. ესენია: არაპოლარული, პოლარული დაუმუხტავი და პოლარული დამუხტული. გლიცინი წარმოადგენს უმარტივეს ამინმჟავას. ალანინი, ვალინი, ლეიცინი, იზოლეიცინი და პროლინი არის ამინმჟავები, რომელთა გვერდითი ჯაჭვი მთლიანად ალიფატურია. მათ შორის გამორჩეულია პროლინი, რომლის გვერდითი ჯაჭვი იმიდური კავშირით, კოვალენტურად უკავშირდება α -ამინო ჯგუფის აზოტის ატომს და შესაბამისად იქმნება ხუთწევრა ჩაკეტილი რგოლი.

არაპოლარული გვერდითი ჯაჭვები ცუდად იხსნება წყალში, რადგან წყლის მოლეკულებთან მათ შეუძლიათ ურთიერთქმედება მხოლოდ ვან-დერ-ვაალსის ძალებით, მაშინ როცა სხვა ამინმჟავათათვის მრავალგვარი სხვა ურთიერთ-ქმედებაა შესაძლებელი.

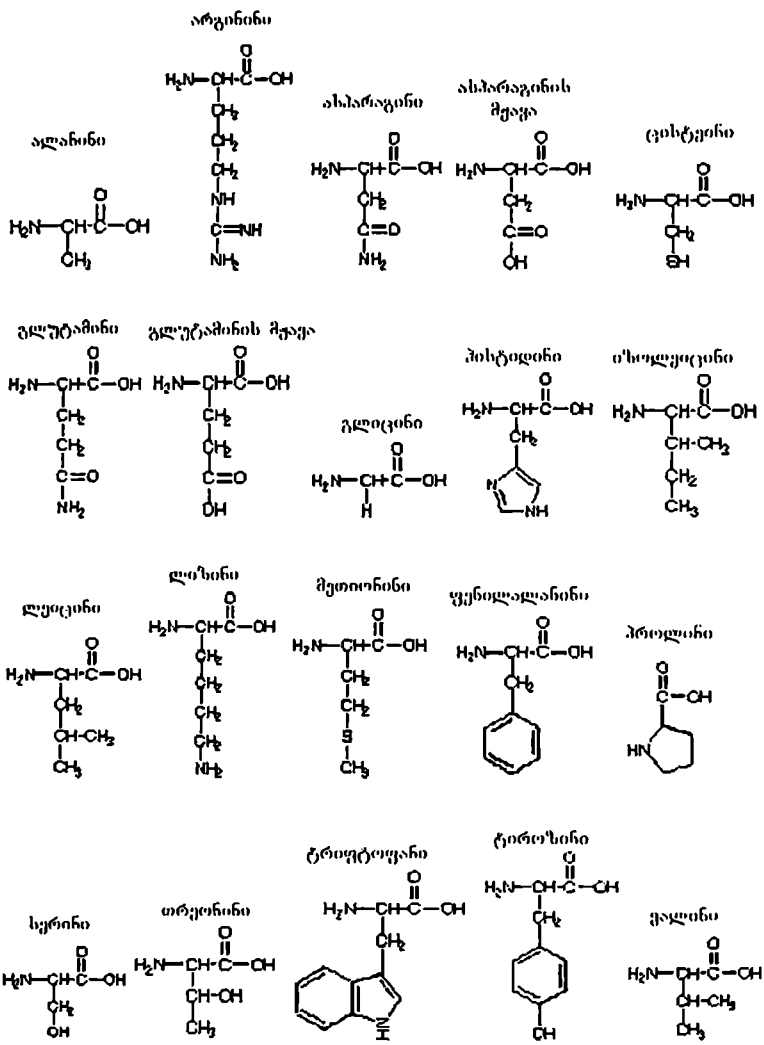
პოლარულ, დაუმუხტავ ამინმჟავათა ჯგუფის წევრებია: სერინი, ტრეონინი, ასპარაგინი, გლუტამინი, ტიროზინი და ტრიფტოფანი. სერინი, ტრეონინი და ტიროზინი შეიცავს ჰიდროქსილის ჯგუფს და შესაბამისად, წყალბადური ბმის წარმოქმნის პროცესში მათ შეუძლიათ ითამამონ როგორც დონორების, ისე აქცეპტორების როლი. ასპარაგინისა და გლუტამინის გვერდითი ჯაჭვები შედარებით პოლარულია და წყალბადურ ბმაში ისინი ასევე შეიძლება ასრულებდნენ დონორის ან აქცეპტორის როლს. ტრიფტოფანის ინდოლურ რგოლში გაერთიანებულ აზოტისა და წყალბადის ატომებსაც გააჩნია უნარი წყალბადური ბმების წარმოქმნისა. პოლარული ჯგუფის სხვა ამინმჟავებს შეუძლიათ დაიმუხტონ. დამუხტული მდგომარეობის წარმოქმნა განისაზღვრება გარემოს pH-ით. ამასთან მათი pK ისეთია, რომ სწორედ ნეიტრალური pH-ის (უჯრედის ბუნებრივი pH) პირობებში ვიღებთ დამუხტულ ფორმას. ასეთ ამინმჟავებს მიეკუთვნება ლიზინი, არგინინი, ჰისტიდინი, ასპარაგინისა და გლუტამინის მჟავები. ლიზინი და არგინინი ფუძე ბუნების ამინმჟავებია, რომელთა გვერდით ჯაჭვებს შეუძლია დადებითად დამუხტვა. ლიზინის α -ამინო ჯგუფის pK 10-ს ტოლია, მაშინ როცა არგინინის გუანიდინური ჯგუფის pK უტოლდება 12-ს. ჰისტიდინი წარმოადგენს კიდევ ერთ ფუძე ბუნების ამინმჟავას, მისი pK 6-ა. ასპარაგინისა და გლუტამინის მჟავები ერთმანეთისაგან განსხვავდება ერთი მეთილის ჯგუფით. მათი კარბოქსილის ჯგუფი ექსტრემალურად პოლარულია და შეუძლია წყალბადური ბმის წარმოქმნა. მათი pK დაახლოებით 4.5-ა.

ფენილალანინი, ტრიფტოფანი და ტიროზინი ქმნის არომატული გვერდითი ჯაჭვის მქონე ამინმჟავების ჯგუფს. ფენილალანინის არომატული რგოლი ბენ -

ცხრილი 5.1 ძირითადი ამინომჟავების ასოითი კოდები და ფიზიკო-ქიმიური მახასიათებლები.

დასახელება	3- ასოიანი კოდი	1- ასოიანი კოდი	შებვედრის საშ.სიხშირე (%).	მ. წ.	pK	VdW მოცულობა (Å³)	C-დამუხტ. P-პოლარული H-ჰიდროფობ. N- ნეიტრალ.
ინალანი	ALA	A	13.0	71		67	H
არგინინი	ARG	R	5.3	157	12.5	148	C+
ასპარეგინი	ASN	N	9.9	114		96	P
ასპარაგინის მჟ.	ASP	D	9.9	114	3.9	91	C-
ცისტეინი	CYS	C	1.8	103		86	P
გლუტამინის მჟ.	GLU	E	10.8	128	4.3	109	C-
გლუტამინი	GLN	Q	10.8	128		114	P
გლიცინი	GLY	G	7.8	57		48	N
ჰისტიდინი	HIS	H	0.7	137	6.0	118	P,C+
იზოლეიციანი	ILE	I	4.4	113		124	H

ლეიცინი	LEU	L	7.8	113		124	H
ლიზინი	LYS	K	7.0	129	10.5	135	C+
მეთიონინი	MET	M	3.8	131		124	H
ფენილალანინი	PHE	F	3.3	147		135	H
პროლინი	PRO	P	4.6	97		90	H
სერინი	SER	S	6.0	87		73	P
თრეონინი	THR	T	4.6	101		93	P
ტრიფტოფანი	TRP	W	1.0	186		163	P
ტიროზინი	TYR	Y	2.2	163	10.1	141	P
ვალინი	VAL	V	6.0	99		105	H



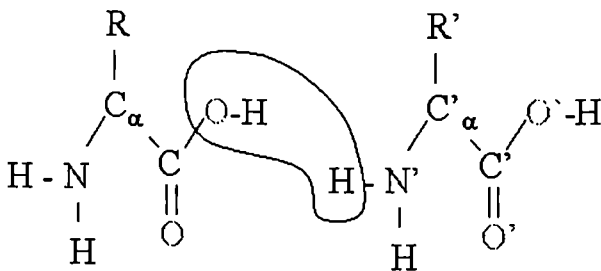
სურათი 5.4 ოცი ძირითადი ამინომჟავის ქიმიური სტრუქტურა

ზოლის რგოლის ანალოგიურია. ის ძალზედ ჰიდროფობულია და ქიმიურად რეაქტიული მხოლოდ განსაკუთრებულ პირობებში ხდება. არომატულად განიხილება ჰისტიდინის გვერდითი ჯაჭვიც.

და ბოლოს, უნდა აღინიშნოს, რომ შესაძლებელია ამინმჟავას α -ამინო და α -კარბოქსილის ჯგუფების იონიზაცია. მათი pK, დათვლილი ალიფატური ამინმჟავების შემთხვევაში, უტოლდება 6.8-7.9-ს და 3.5-4.3-ს შესაბამისად. დამუხტული გვერდითი ჯაჭვი, ცხადია, გავლენას ახდენს უკანასკნელ pK-თა მნიშვნელობაზე. საჭიროა შევნიშნოთ, რომ ამინმჟავა, როგორც წესი, ჩართულია პოლიპეპტიდურ ჯაჭვში და შესაბამისად მხოლოდ ამინო და კარბოქსი ბოლოებში განლაგებულ ამინმჟავებს გააჩნიათ თავსუფალი α -ამინო და α -კარბოქსი ჯგუფები.

პოლიპეპტიდური ჯაჭვი

პოლიპეპტიდური ჯაჭვის სინთეზი ხდება დეჰიდრატაციის საფუძველზე. ასეთი რეაქცია გულისხმობს მოცემული ამინმჟავას კარბოქსილის ჯგუფისა და შემდეგი ამინმჟავას ამინო ჯგუფს შორის C-N პეპტიდური ბმის წარმოქმნას წყლის გამოთავისუფლებით. პოლიმერიზაციის ეს რეაქცია არაა სპონტანური, მისი ენერგეტიკული უზრუნველყოფა ხდება რიბოსომაზე. რიბოსომა წარმოადგენს ცილებისა და სპეციალური რნმ-ს რთულ კომპლექსს, რომლის დანიშნულებაცაა მოახდინოს mRNA-ში ასახული გენის თანამიმდევრობის ტრანსლაცია ცილის თანამიმდევრობაში. ზემოთ წარმოდგენილი ოცივე ამინმჟავას გააჩნია თავისი გენეტიკური კოდი და ცილის სინთეზის პროცესში მათ სწორ, კოდის შესაბამისად ჩართვას უზრუნველყოფს რიბოსომული სისტემა.



სურათი 5.5. ორი ამინმჟავას შერთებისას დეჰიდრატაციის სქემატური წარმოდგენა

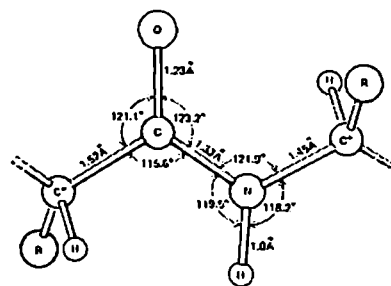
სინთეზის შეზღუდული პროცესი, პეპტიდური ბმის ჰიდროლიზიც არ წარმოადგენს სპონტანურ პროცესს, ის შესაძლებელია მხოლოდ განსაკუთრებულ, ექსტრემალურ პირობებში. ასე მაგალითად, პეპტიდური ბმის ჰიდროლიზს ადგილი შეიძლება ჰქონდეს ან ძალიან მკაფე არეში (1 მოლარულ HCl-ში) ან 100°C -ზე. ამრიგად, შეიძლება ითქვას, რომ ნორმალურ პირობებში ჰიდროლიზის პროცესი ნაკლებად ალბათურია და თუ არ ექნება ადგილი მიზანმიმართულ დეპოლიმერიზაციას, ცილა ქიმიურად და ბიოლოგიურად სტაბილური ნაერთია. პოლიპეპტიდური ჯაჭვის ამინომიმუჯვებად დაშლა, როგორც წესი, ხორციელდება სპეციალური, მაჰიდროლიზირებელი ფერმენტების საშუალებით.

უმრავლესობა ცილებისა ჰეტეროპოლიმერებია (ანუ თავის ჯაჭვში მოიცავს უმრავლესობა ან ყველა ბუნებრივ ამინომუჯვას). ზოგჯერ, თუმცა იშვიათად, ცილის გარკვეული უბანი წარმოადგენს მხოლოდ რამდენიმე ან სულაც ერთი კონკრეტული ამინომუჯვასაგან შემდგარ თანამიმდევრობას. შეიძლება ითქვას, რომ ტიპური ცილის ნებისმიერი უბანი ხასიათდება დიდი ქიმიური ჰეტეროგენურობით. სწორედ ეს მაღალი ჰეტეროგენურობა განაპირობებს ცილის სივრცული მოწყობის მრავალფეროვნებას.

პეპტიდური ბმა

ამინომუჯვათშორისი პეპტიდური ბმა წარმოადგენს ორი (თანამიმდევრული) α -ნახშირბადის ატომებით ფლანკირებულ (შემოსაზღვრულ) ამიდური ბმის

სპეციალურ სახეს. (C-N) ბმის სიგრძეა $\sim 1.33 \text{ \AA}$ (იხ. სურათი 5.6 და ცხრილი 5.2). იგი შედარებით მოკლეა მეზობელ N-C ბმაზე, რომლის სიგრძეა $\sim 1.45 \text{ \AA}$ და გრძელია C=O ბმაზე, რომლის სიგრძეა $\sim 1.23 \text{ \AA}$. 5.6 სურათზე მოცემული ბმის სიგრძეთა და ვალენტურ კუთხეთა სიდიდეები ასახავს ელექტრონთა ატომთშორის განაწილებებს. ეს განაწილებები განსხვავდება ატომთა პოლარიზაციის ხარისხითა და მათი ბმული ორბიტალებით. ორ უფრო ელექტროუარყოფით ატომს, O-სა და N-



სურათი 5.6. პეპტიდური ჯგუფის

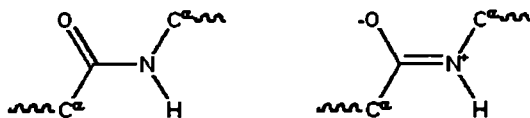
ს ანიჭებენ უარყოფით პარციალურ მუხტს, მაშინ როცა ნაკლებად ელექტროუარ-

ცხრილი 5.2 ზოგიერთ ბმათა სიგრძეების მნიშვნელობები Å-ში

პეპტიდური სიბრტყის ბმები	საშ. სიგრძე	ცალკეული ბმები	საშ. სიგრძე	წყალბადური ბმები	საშ. სიგრძე (±0.3)
Cα - C'	1.52	C - C	1.54	O-H --- O-H	2.80
C' - N	1.33	C - N	1.48	N-H --- O=C	2.90
N - Cα	1.45	C - O	1.43	O-H --- O=C	2.80

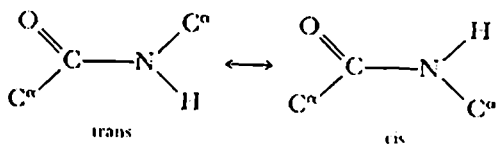
ყოფით ატომებს, C'-სა და H-ს ენიჭებათ დადებითი პარციალური მუხტი. უკანასკნელი ოთხი ატომი ქმნის პეპტიდურ ჯგუფს, რომელიც განიხილება როგორც რეზონანსული სტრუქტურა (იხ.სურათი 5.7). ანუ პეპტიდურ ბმას გააჩნია ორმაგი ბმის ბუნება და ამით აიხსნება მისი სიგრძის მნიშვნელობის შუალედურობა ორმაგ და ერთმაგ ბმებს შორის.

ზოგადად ორმაგი ბმის ირგვლივ და კონკრეტულად, C'-N ბმის ირგვლივ ბრუნვა შეზღუდულია. cis და trans ფორმებს შორის ბარიერი უტოლდება 3კკალ/მოლზე. cis იზომერი (კონფორმაცია) შეესაბამება მდგომარეობას, როცა მეზობელი (i,i+1) α- ნახშირბადები განლაგებულია პეპტიდური ბმის ერთ მხარეს, trans კი - მოპირდაპირე მხარეს (იხ.სურათი 5.8). ანუ trans კონფიგურაცია მეზობელ გვერდით ჯაჭვებს ისე განლაგებს, რომ ამცირებს სტერიულ წინააღმდეგობებს პეპტიდური ბმის ირგვლივ მცირე კუთხით ბრუნვის პროცესში. უმრავლეს ამინმჟავებში პეპტიდური ბმის cis / trans კონფიგურაციათა თანაფარდობა დაახლოებით 90/10 რიგისაა, მაშინ როცა პროლინისათვის ეს შეფარდება 70/30-ს უდრის. ორმაგი ბმის ბუნებიდან გამომდინარე, პეპტიდური ბმის კონფორმაციის განმსაზღვრელი ატომები, გარკვეულ მიახლოებაში, ერთ სიბრტყეშია განლაგებული. შეგვიძლია ჩავთვალოთ, რომ მეზობელ α-ნახშირ-



სურათი 5.7. პეპტიდური ჯგუფის სტრუქტურის რეზონანსულობა

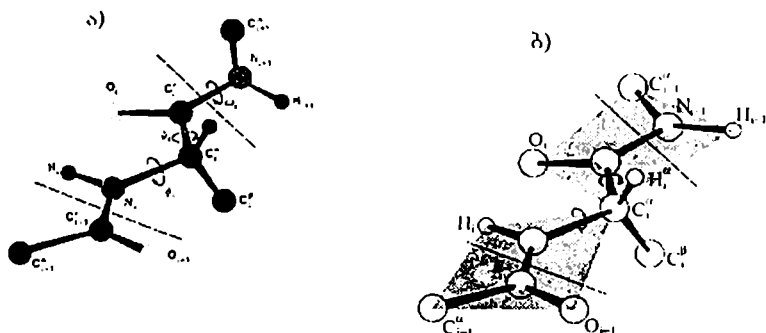
ბადებს შორის (მათი ჩათვლით) განლაგებული ატომები ერთ სიბრტყეშია (იხ. სურათი 5.9). პეპტიდური ბმის trans კონფორმაციის, ატომთა ერთ სიბრტყეში განლაგებისა და მათი პოლარულობის გამო პეპტიდური ბმის გასწვრივ წარმოიქმნება დიპოლი შესაბამისი დიპოლური მომენტი. დიპოლის უარყოფითი ბოლო ემთხვევა კარბონილის ჯგუფს (იხ. სურათი 5.13).



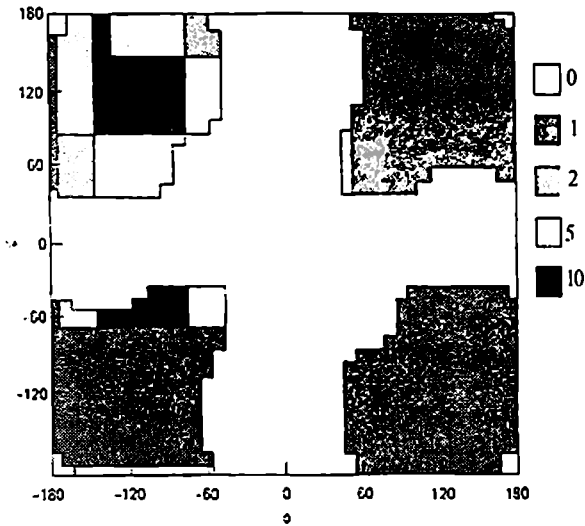
სურათი 5.8. ტრანს და ცის კონფიგურაციები

შეზღუდვები ერთმაგი ბმის ირგვლივ ბრუნვისას

პეპტიდური ბმის ირგვლივ ბრუნვა შეზღუდულია. შედარებით თავისუფალია α -ნახშირბადთან არსებული ოთხი ბმის ირგვლივ ბრუნვა. ამასთან, ამ ოთხიდან ორი ბმის ირგვლივ ბრუნვა განსაზღვრავს პოლიპეპტიდის ძირითადი ჯაჭვის კონფორმაციას. ეს ბმებია: α -ნახშირბადსა და კარბონილურ ნახშირბადს შორის, რომლის ირგვლივ ბრუნვის ორწახნაგა კუთხეს აღნიშნავენ ψ -თი; ამინო ჯგუფსა და α -ნახშირბადს შორის, რომლის ირგვლივ ბრუნვის კუთხეს აღნიშნავენ ϕ -თი (იხ. სურათი 5.9).



სურათი 5.9 ა) ძირითადი ჯაჭვის ბმების ირგვლივ ბრუნვის აღმწერი ორწახნაგა კუთხეები; ბ) პეპტიდური სიბრტყეები



სურათი 5.10 არადრეკადი სფეროების მიახლოებაში განსაზღვრული პოტენციალით ატებული რამაჩანდრანის რუკა ამინმჟავა გლიცინისათვის.

გამომდინარე იქიდან, რომ α -ნახშირბადი წარმოადგენს პეპტიდური სიბრტყის ერთ-ერთ ატომს, ϕ ან ψ კუთხით ბრუნვა განაპირობებს მომიჯნავე პეპტიდური სიბრტყეების ერთმანეთის მიმართ ბრუნვას. ამასთან, ϕ და ψ -ს გარკვეული მნიშვნელობები იწვევს გარკვეულ სტერიულ დამაბულობებს. ცხადია, ϕ -სა და ψ -ს სტერიულად დასაშვებ მნიშვნელობათა არეები განისაზღვრება α -ნახშირბადის გვერდით ჯაჭვთა ბუნებიდანაც. მიუხედავად იმისა, რომ გლიცინს ყველაზე მცირე გვერდითი ჯაჭვი (წყალბადი) აქვს, მისთვისაც არსებობს ორწახნაგა კუთხეების დაუშვებელ მნიშვნელობათა არეები (იხ. სურათი 5.10).

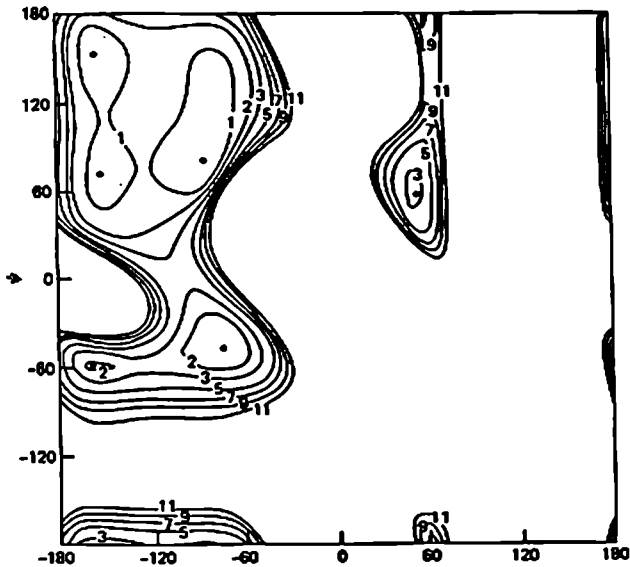
თითოეული ამინმჟავისათვის ϕ , ψ სიბრტყეზე დასაშვები უბნები მოიცემა რამაჩანდრანის რუკებით (ინდოელმა მეცნიერმა რამაჩანდრანმა პირველმა შემოგვთავაზა ამინმჟავათა კონფორმაციული შესაძლებლობების რუკა). დასაშვები უბნები შემოსაზღვრულია დაბალენერგეტიკული კონტურებით (იხ. სურათი 5.11), ან არადრეკად სფეროთა (ატომთა) არაგადაფარვადი უბნების შემომსაზღვრელი წრფეებით (სურათი 5.10). შეზღუდვები β -ნახშირბადის მომცველ ამინმჟავათათვის საკმაოდ მკაცრია და შესაბამისად ϕ , ψ სიბრტყის მხოლოდ მცირე ნაწილია სტერიულად დასაშვები. მაგალითად ვალინისა და იზოლუეოცინისათვის ϕ , ψ სიბრტყის მხოლოდ 5%-ა დასაშვები. თუმცა

ამინმეავათა უმრავლესობას აქვს საერთო დასაშვები უბნები, რომლებიც მდებარეობს ϕ , ψ სიბრტყის მარცხენა ზედა და ქვედა ზონებში. როგორც ქვემოთ ვნახავთ ამ ზონებში განლაგებულია ϕ , ψ -ს ისეთი კომბინაციები, რომლებიც შეესაბამება ფართოდ გავრცელებულ რეგულარულ მეორეულ სტრუქტურებს, კერძოდ კი α -სპირალსა და β -სტრუქტურას.

მიუხედავად იმისა, რომ ამინმეავათა უმრავლესობა შეიძლება იყოს როგორც ერთ, ისე მეორე კონფორმაციულ ზონაში, α -სპირალურ და β -სტრუქტურულ უბნებს შორის არის ენერგეტიკული ბარიერი. ანუ α -სპირალის პირდაპირი და ადგილი გარდაქმნა β -სტრუქტურაში გამწვანებულია. აქ ხაზი უნდა გაესვას ორ მნიშვნელოვან გარემოებას, რომელიც ცილათა და პეპტიდთა სტრუქტურების ახლახან ჩატარებული ანალიზის შედეგად გამოვლინდა. პირველი გარემოება მდგომარეობს იმაში, რომ გასაოცრად ხშირია შემთხვევები, როცა ადგილი აქვს პეპტიდური ბმის მცირე, მაგრამ შესამჩნევ გადახრებს სიბრტყიდან. შესაძლოა, პეპტიდური ბმის ირგვლივ ბრუნვის კუთხის დასაშვებ მნიშვნელობათა ზონის გაზრდამ გამოიწვიოს ϕ , ψ მნიშვნელობათათვის დასაშვებ უბანთა გაფართოება და ამის შედეგად α/β ბარიერის შემცირება. მეორე გარემოება უკავშირდება იმ ფაქტს, რომ ცილოვან სტრუქტურებში ადგილი აქვს ზოგიერთი ამინმეავას (განსაკუთრებით მცირე ზომის პოლარული ამინმეავების) კონფორმაციის რამაჩანდრანის რუკის არადასაშვებ უბნებში მოსალოდნელზე ხშირ მოხვედრას. ამ ფაქტის ახსნა შესაძლებელია პოლარულ ამინმეავათათვის დამახასიათებელი, ლოკალურ ურთიერთქმედებებში შესვლის უნარით. ეს ალბათ სწორედ ის რესურსია, რომელსაც შეუძლია აანაზღაუროს სტერიულად არახელსაყრელ კონფორმაციებში მოხვედრით გამოწვეული ენერგეტიკული დანაკარგები.

მეორეული სტრუქტურა

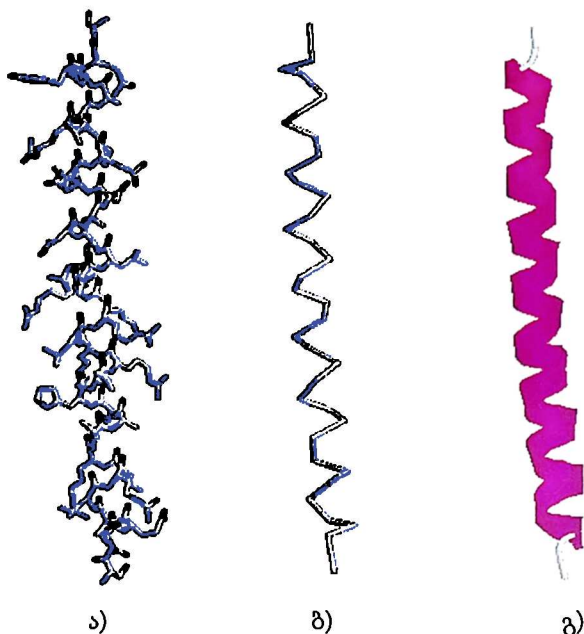
გვერდით ჯაჭვსა და ძირითად ჯაჭვს შორის წარმოქმნილი სტერიული დამაბულობები იწვევს კონფორმაციულ შეზღუდვებს და აქედან გამომდინარე მეორეული სტრუქტურის განსაკუთრებული, α და β ტიპების არსებობას. სტერიული დამაბულობები რამაჩანდრანის რუკის α და β ზონებში მნიშვნელოვნად არაა თანამიმდევრობაზე დამოკიდებული, რადგან ისინი შედეგია არა ამინმეავათშორისი, არამედ ამინმეავათშიდა ურთიერთქმედებებისა.



სურათი 5.11 რამაჩანდრანის რუკა ამინმყავა ალანინისთვის, აგებული დრეკად (ატომთშორის მანძილების ცვლილებისას პოტენციალური ფუნქცია უწყვეტად იცვლება) პოტენციალზე დაყრდნობით.

თუ თანამიმდევრობა შედგება ისეთი ამინმყავებისაგან, რომელთაც დასაშვები ϕ, ψ უბნები ერთნაირი აქვს, მაშინ ფორმირდება ჯაჭვის რეგულარული კონფორმაციის მქონე სპირალური სეგმენტი. რამაჩანდრანის რუკიდან ჩანს, რომ ალბათურია α ან β რეგულარული სტრუქტურის შესაბამისი კონფორმაცია. ანუ ცილის მეორეული სტრუქტურის ფორმირების საფუძველი თვითონ პოლიპეპტიდურ ჯაჭვშია ჩადებული და მეორეული სტრუქტურის ზემოთხსენებული ელემენტები შინაგანად მომგებიანია ნებისმიერი ჯაჭვისთვის, ნებისმიერ პირობებში.

სპირალური სტრუქტურა მოგვაგონებს ზამბარას. ყველაზე მეტად გავრცელებულია მარჯვენა α -სპირალი (იხ. სურათი 5.12), რომლის ხვიაზეც მოდის 3.6 ამინმყავა და ბიჯი შეადგენს 5.4 Å-ს.

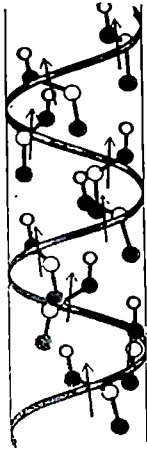


სურათი 5.12. α -სპირალთა წარმოდგენის ფორმები ა) ლეროვან-ბურთულოვანი ბ) ძირითადი ჯაჭვის გ) ზაფთის

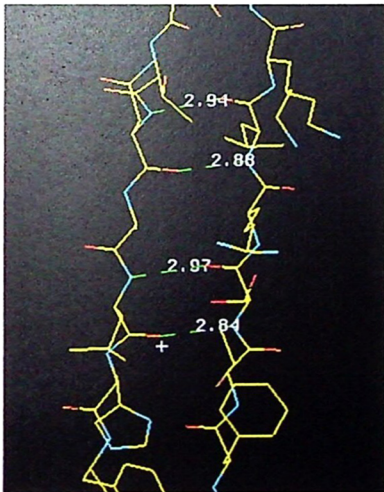
ასეთი პარამეტრების პირობებში ცხადია, რომ i -ურ ამინმჟავასთან სივრცულად უახლოესი ამინმჟავები განლაგებულია ჯაჭვის $i+3$, $i-3$, $i+4$, $i-4$ პოზიციებში (იხ. სურათი 5.13). სპირალური ლერძის მიმართ პეპტიდური ბმის დახრა და სპირალის სხვა პარამეტრები უზრუნველყოფს პეპტიდური დიპოლური მომენტის სპირალური ლერძის პარალელურ მდგენელთა აჯამვასა და მართობულ მდგენელთა ურთიერთკომპენსაციას, რასაც მივყავართ მნიშვნელოვან ჯამურ დიპოლურ მომენტამდე.

მარჯვენა α -სპირალის მნიშვნელოვან თავისებურებას წარმოადგენს i -ური ნაშთის კარბონილის ჟანგბადის $i+3$ ან $i+4$ ნაშთის ამიდურ წყალბადთან სივრცულ სიახლოვეში მოსვლა, რაც განაპირობებს წყალბადური ბმის წარმოქმნას.

ცილებში რეგულარული სტრუქტურის მეორე ტიპს წარმოადგენს β სტრუქტურა (იხ. სურათი 5.14). ჯაჭვი ამ შემთხვევაში იღებს გაჭიმულ კონფორმაციას და პეპტიდთა დიპოლური მომენტების აჯამვისას ჯამური დიპოლური მომენტი



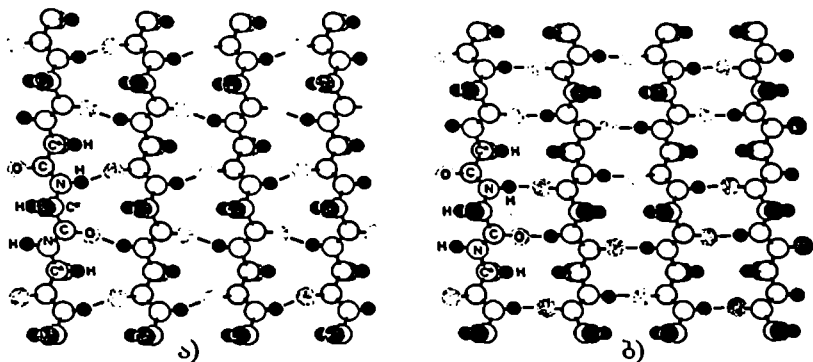
სურათი 5.13. მარჯვენა α -სპირალის სქემატური წარმოდგენა. სურათზე წრეები შეესაბამება პეპტიდური ჯგუფის ატომებს, ხოლო ისარი, პეპტიდური დიპოლის მიმართულებას.



სურათი 5.14. წყალბადურ ბმათა ქსელი ორ-სეგმენტა ანტიპარალელური გოფრირებული ფურცლის შემთხვევაში. გვერდითი ჯაჭვები სურათის სიბრტყიდან მიმართულია ზემოთ ან ქვემოთ.

უმნიშვნელოა. სტერიულად დაუბაბავ ϕ , ψ უბნებს უფრო დიდ ფართობი უკავია რამაჩანდრანის რუკაზე. ასეთ გაკვირვებულ კონფორმაციაში მყოფი ჯაჭვის რამდენიმე ფრაგმენტი (სეგმენტი) ურთიერთქმედებს სეგმენტზორისი წყალბადური ბმებით, ელექტროსტატიკურად და შედეგად ვიღებთ β გოფირებული ფურცლის მსგავს სტრუქტურას (β -სტრუქტურას). ამრიგად β -ს α -გან განასხვავებს ის, რომ წყალბადური ბმები სეგმენტებს შორისაა და არა სეგმენტს შიგნით, როგორც ეს α -ს შემთხვევაშია და აქედან გამომდინარე, β -ს ცალკეულ ზოლს (სეგმენტს) არ გააჩნია ცალკე არსებობის რესურსი. ასევე, α -გან განასხვავებით β სტრუქტურის შემადგენელი სეგმენტები შესაძლებელია პოლიპეპტიდური ჯაჭვის გასწვრივ საკმაოდ დაშორებული იყოს. β -სტრუქტურის შემადგენელი სეგმენტები შეიძლება იყოს ურთიერთ პარალელური ან ანტიპარალელური (მიმართულება განისაზღვრება N-დან C ბოლოსაკენ). შესაბამისად, არჩევენ პარალელურ და ანტიპარალელურ β -სტრუქტურებს (იხ. სურათი 5.15). იმ შემთხვევაშიც კი, როცა ჯაჭვის გასწვრივ მეზობელი სეგმენტები ქმნის თმის სამაგრის (სარქის) ტიპის ანტიპარალელურ β -სტრუქტურულ ფრაგმენტს, ამ ფრაგმენტის პოლიპეპტიდური ჯაჭვის გასწვრივ საკმაოდ დაშორებული N და C ბოლოების ნაშთები საკმაოდ უახლოვდება ერთმანეთს სივრცეში. შესაბამისად, მიუხედავად იმისა, რომ β სტრუქტურას, მისი რეგულარობიდან გამომდინარე, მიაკუთვნებენ მეორეული სტრუქტურის ელემენტს, არალოკალური ბუნების გამო იგი მესამეული სტრუქტურის თვისებების მატარებელია. ამრიგად, შეიძლება ითქვას, რომ მეორეულ და მესამეულ სტრუქტურებს შორის მკაფიო თვისობრივი ზღვრის გავლება ძალიან ძნელია.

ე.წ. „მოსახვევები“ ასევე კლასიფიცირდება როგორც მეორეული სტრუქტურის ელემენტები, მაგრამ სპირალებისა და გოფირებული ფურცლებისაგან (β სტრუქტურა) განსხვავებით არ ახასიათებთ განმეორებადი, რეგულარული გეომეტრია. მოსახვევები ხასიათდება ამინომოყვავათა გარკვეული, გამოკვეთილი ურთიერთგანლაგებით, რასაც განაპირობებს ϕ და ψ -ს კონკრეტული მნიშვნელობები. შესაბამისად, მოსახვევის არსებობას განაპირობებს კონკრეტული თანამიმდევრობები, ამინმყავათა სპეციალური ტიპები და წყალბადური ბმების წარმოქმნის კონკრეტული შესაძლებლობები. მოსახვევების უმრავლესობა შექმნილია ამინმყავურ ნაშთთა მცირე რიცხვისაგან და ლოკალური ურთიერთქმედებების შედეგია. მაგრამ არსებობს მრავალი ამინმყავის შემცველი Ω ყულფებიც, რომლის შემადგენელი ამინმყავები ზოგადად არ ქმნიან ჯაჭვის განსაზღვრულ (კონკრეტულ) გეომეტრიას. ამ შემთხვევაში მოსახვევი ამინმყავური ჯაჭვის მოცემული ფრაგმენტის არარეგულარი კონფორმაციის შედეგია და სახელი, Ω მოსახვევი, კავშირშია ჯაჭვის დამახასიათებელ ტოპოლოგიასთან (ფორმასთან). მოსახვევის უზრუნველმყოფი თანამიმდევრობათა ფრაგმენტების არსებობა აუცილებელია პოლიპეპტიდური ჯაჭვის დახვევის პროცესისათვის და ნატივური



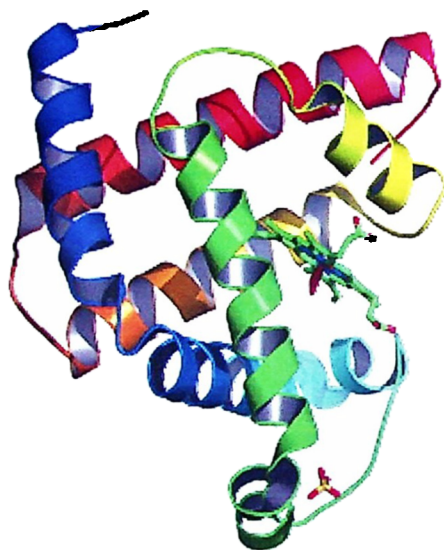
სურათი 5.15. β გოფირებული ფურცლის ტიპის სტრუქტურები. ა) პარალელური და ბ) ანტიპარალელური.

მესამეული სტრუქტურის დამახასიათებელ ურთიერთქმედებათა უზრუნველყოფისათვის. უკანასკნელი ურთიერთქმედებები, როგორც წესი, შორი ურთიერთქმედებებია და მათი საშუალებით ცილა იღებს კომპაქტურ, გლობულარულ სტრუქტურას. ჩვეულებრივ, მოსახვევები განლაგდება გლობულის ზედაპირთან ახლოს, ხოლო სპირელები და/ან გოფირებული ფურცლები ქმნის ცილის სტაბილურ ბირთვს (ჩონჩხს). დინამიკურობის თვალსაზრისით, ცილის ზედაპირზე განლაგებული მოსახვევები განსხვავდებიან. ვხვდებით საკმაოდ მობილურ მოსახვევებს, რომლებიც ცილის დანარჩენ ზედაპირთან მხოლოდ რამდენიმე ურთიერთქმედებით შემოიფარგლებიან. არსებობს ფიქსირებული სტრუქტურის მქონე მოსახვევები, რაც დიდი რაოდენობა კონტაქტების არსებობის შედეგია. ასე რომ, მოსახვევის მეორეულ სტრუქტურად კლასიფიცირება პირობითია.

მესამეული სტრუქტურა

α -სპირალისა და β -სტრუქტურის ზედაპირზე განლაგებული გვერდითი ნაშთები ურთიერთქმედებენ ერთმანეთთან ჰიდროფობული კონტაქტებითა და სხვა ბუნების ძალებით და შედეგად ხდება მესამეული სტრუქტურის ფორმირება (იხ. სურათი 5.16).

α-სპირალურ სტრუქტურაში გვერდითი ჯაჭვები მიმართულია რადიალურად, სპირალის ღერმიდან გარეთ, β-სტრუქტურაში კი, გოფირებული ფურცლის ზედაპირიდან ზემოთ და ქვემოთ. მეორეული სტრუქტურის წარმოქმნა გარკვეულწილად ახსნილ იქნა φ, ψ სივრცეში არსებული სტერიული შეზღუდვებით. მესამეული სტრუქტურის წარმოქმნის საფუძვლების წარმოჩენა უფრო რთული ამოცანაა. მაღალორგანიზებული მესამეული სტრუქტურის მქონე ცილებს ახასიათებთ კარგად გამოკვეთილი, ძირითადად ჰიდროფობულ ამინმჟავათაგან შემდგარი ბირთვი, რომლის შექმნაშიც თავისი წვლილი შეაქვს მეორეული სტრუქტურის თითქმის ყველა არსებულ ელემენტს. ამრიგად, არსებობს მჭიდრო კავშირი მეორეულ და მესამეულ სტრუქტურებს შორის.



სურათი 5.16. მიოგლობინის მესამეული სტრუქტურა. ალფა სპირალები წარმოდგენილია სხვადასხვა ფერის ბაფთებით. მიოგლობინი პირველი ცილაა, რომლის სივრცული სტრუქტურა დადგენილ იქნა რენტგენოკრისტალოგრაფიული მეთოდით

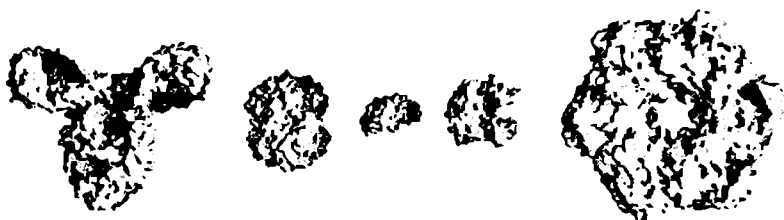
ბირთვში შემავალი ამინმჟავები მჭიდრო ურთიერთკონტაქტშია და ქმნის დიდი სიმკვრივის სტაბილურ წარმონაქმნს. წყლის გარემოში, სწორედ ჰიდროფობული ეფექტი განაპირობებს ცილის კომპაქტიზაციას და იცავს მას არასასურველი სოლვატაციისაგან. მართალია, თითქოს კომპაქტიზაცია და შინაგანი ორგანიზაცია ენტროპულად არახელსაყრელია, რადგან მცირდება ჯაჭვის თავისუფლების ხარისხები, მაგრამ მთელი სისტემისათვის (პოლიპეპტიდი + გამხსნელი) ამას მიყვავართ ენტროპიის ზრდამდე. კომპაქტიზაციის პროცესში,

გამხსნელთან ურთიერთქმედება ჩანაცვლდება მოლეკულის შიგნით სეგმენტშორისი ურთიერთქმედებით, რასაც დამატებითი ენერგეტიკული წვლილი შეაქვს სტაბილიზაციაში. მაგალითად, მცირდება დიელექტრიკული კონსტანტა ახლად ჩამოყალიბებული სტრუქტურის შიგნით. სეგმენტშორისი წყალბადური ბმები მომგებიანია მეორეული სტრუქტურის ელემენტის წარმოქმნისათვისაც. მათ ჩამოყალიბებას ხელს უწყობს სტერიული შეზღუდვები და აქედან გამომდინარე, ბმის წარმოქმნა შესაძლებელია დამატებითი ენტროპული დანახარჯების გარეშე. ამრიგად, ცილის მეორეული და მესამეული სტრუქტურები არაა დამოუკიდებელი. ისინი მჭიდრო ურთიერთდამოკიდებულებაში იმყოფებიან. შეიძლება დავუშვათ, რომ სწორედ ეს ურთიერთდამოკიდებულება ქმნის ცილის სტრუქტურული სტაბილურობისათვის დამახასიათებელი კოოპერატიულობის მოლეკულურ საფუძველს. კოოპერატიულობა გამოიხატება იმაში, რომ არახელსაყრელ პირობებში (მაღალი ტემპერატურა ან დეტერგენტი) დახვეული (კომპაქტური, ნატივური) მდგომარეობისათვის დამახასიათებელი მეორეული და მესამეული სტრუქტურის რღვევა მიმდინარეობს პარალელურად (თითქმის ერთდროულად) „ყველაფერი ან არაფერი“-ს პრინციპით.

მეოთხეული სტრუქტურა

ცილის სტრუქტურული ორგანიზაციის შემდეგი დონეა მეოთხეული სტრუქტურა. ასოცირებული სუბერთეულები შეიძლება იყოს იდენტური, ან არა. შეიძლება სიმეტრიულად იყოს სივრცეში ორგანიზებული, ან არა. ზოგადად, მეოთხეული სტრუქტურა წარმოადგენს დამოუკიდებელი მესამეული სტრუქტურების ზედაპირული კონტაქტების საშუალებით ასოცირების შედეგს. მაგალითად გამოდგება ჰემოგლობინის ტეტრამერი, რომელიც მიოგლობინის მსგავს მონომერთა ასოცირების შედეგია. მაგრამ ამასთანავე არსებობს მრავალი მაგალითი, რომლებიც გვიჩვენებს, რომ მესამეული სტრუქტურის ჩამოყალიბება მიმდინარეობს მეოთხეულის ჩამოყალიბებასთან ერთად. ამის ნათელი მაგალითია ტრიპტოფანის რეპრესორი.

ამრიგად, შეიძლება ითქვას, რომ მესამეულ და მეოთხეულ სტრუქტურათა ურთიერთდამოკიდებულება მსგავსია მეორეულ და მესამეულ სტრუქტურებს შორის ურთიერთდამოკიდებულებისა. შესაბამისად, ცილის სტრუქტურული ორგანიზაციის ოთხ დონედ დაყოფის სურათი არაა ცალსახად მკვეთრი. უფრო მეტიც, ზედაპირულად წარმოდგენილმა დაყოფამ შეიძლება მოგვეცეს პოლიპეპტიდური ჯაჭვის ცილად დახვევის კანონზომიერებების, მისი სტაბილობის განმსაზღვრელ ურთიერთქმედებათა შორის ურთიერთდამოკიდებულების მდგარი სურათი.



სურათი 5.17. ცილათა ზომის, ფორმისა და ფუნქციური მრავალფეროვნების დემონსტრირება რამდენიმე ცილის მაგალითზე. მარცხნიდან მარჯვნივ წარმოდგენილია : ანტისხეული (IgG), ჰემოგლობინი, ინსულინი (ჰორმონი), ადენილატკინაზა (ფერმენტი) და გლუტამინ სინთეტაზა (ფერმენტი).

ცილების მესამეულ სტრუქტურათა სიმრავლიდან, სხვადასხვა კომბინაციით მეოთხეულ სტრუქტურათა წარმოქმნის შესაძლებლობა, კიდევ უფრო აფართოებს ცილოვანი მოლეკულების შესაძლო გეომეტრიულ ფორმათა და შესაძლო ფუნქციათა სპექტრს. მაგალითისათვის იხილეთ სურათი 5.17

6. პოლიპეპტიდის ცილად დახვევის გზები და მექანიზმები.

ცილის შესაძლო კონფორმაციათა რიცხვი გაშლილ მდგომარეობაში უსაზღვროდ დიდია, რაც ძნელად წარმოსადგენს ხდის იმ ფაქტს, რომ მცირე ცილების უმრავლესობის მიერ უნიკალური ნატივური სტრუქტურის მიღება წამის, ზოგიერთების კი, მილიწამების რიგის დროის ინტერვალში ხდება. ევოლუციის პროცესში მოიძებნა დახვევის ეფექტური გზები და მექანიზმები. ადამიანებსა და ცხოველებში ამ მექანიზმების დაკარგვა ხშირად მომაკვდინებელი ავადმყოფობების საფუძველი ხდება.

გაშლილ, სტერიულად დასაშვებ კონფორმაციათა გიგანტური რიცხვის მთავარი მიზეზი შეიძლება იყოს სხვადასხვა როტამერთა (იხ. რამაჩანდრანის რუკა) ენერგიებს შორის სხვაობის სიმცირე, რასაც მიეყვართ იქამდე, რომ ძირითადი ჯაჭვისა თუ გვერდითი ჯაჭვის მიერ სხვადასხვა ორიენტაციათა მიღების ალბათობები მნიშვნელოვნად არ განსხვავდება ერთმანეთისაგან. შეიძლება დავუშვათ, რომ ამინმჟავაზე მოსული დასაშვებ კონფორმაციათა საშუალო რიცხვია ათი. ლევინტალმა ჯერ კიდევ 1968 წელს შენიშნა, რომ დახვევა არ შეიძლება კონფორმაციათა შემთხვევითი შერჩევის გზით მიმდინარეობდეს. თუ დავუშვებთ, რომ ცალკეული კონფორმაციის შემთხვევით მიღების დრო 1 ნანოსეკუნდია, ნატივურ სტრუქტურად დახვევას არარეალურად დიდი (სამყაროს ასაკზე გაცილებით დიდი) დროის ინტერვალი დასჭირდებოდა. სინამდვილეში ბევრი ცილის დახვევის დრო მილიწამის რიგისაა. ლევინტალის ეს შენიშვნა ცნობილია როგორც „ლევინტალის პარადოქსი“. აქედან გამომდინარე, ლოგიკურია დავუშვათ, რომ დახვევა მიმდინარეობს არა შემთხვევითი, არამედ გაცილებით ეფექტური გზით.

უნდა ითქვას, რომ ცილად დახვევის გზებსა და მექანიზმებში გარკვევის ამოცანა სულ უფრო მეტ აქტუალობას იძენს. ამის დამადასტურებელი მრავალი არგუმენტი არსებობს :

1. პრიონის (Prion), ჰანტიგტონის (Huntington) და ალცჰეიმერის (Alzheimer) დაავადებებს სწორედ ცილათა გაშლა, აგრეგაცია და ფიბრინოგენები იწვევს. ბოლო ხანებში იზრდება ახლად აღმოჩენილ პათოლოგიათა რიცხვი, რომლებსაც მიზეზად ცილათა არასწორად დახვევა უდევს საფუძვლად.

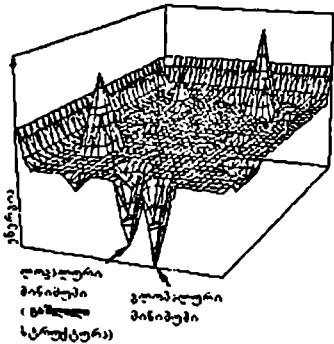
2. როგორც წარმოებაში, ისე მეცნიერებაში დიდია მოთხოვნა სწორად დახვეულ (ანუ ბიოლოგიურად აქტიურ) რეკომბინანტულ ცილათა დიდ რაოდენობებზე. ახალი ცილის დაპროექტების პროცესში აქცენტი უნდა გაკეთდეს სწრაფი

დახვევის უნარზე, რადგანაც ამ შემთხვევაში უზრუნველყოფილია აგრეგაციის დაბალი დონე, დაცულობა პროტეოლიზური დეგრადაციისგან და მაღალი აქტივობა.

3. იზრდება ისეთი ცილების მიღების აუცილებლობა, რომლებსაც გარკვეულ ექსტრემალურ პირობებში შენარჩუნებული აქვთ ფერმენტული აქტივობა. შესაბამისად, აქტუალური ხდება მაღალსტრუქტურირებული, სტაბილური და ექსტრემალურ პირობებში მომუშავე ფერმენტების პროექტირების ამოცანა.

4. მრავალი ცილის სივრცული მოწყობის წინასწარმეტყველების მეთოდების წარმატება დამოკიდებულია მრავალგანზომილებიანი პოტენციური ფუნქციის გლობალური მინიმუმის პოვნის მეტად რთული მათემატიკური ამოცანის გადაწყვეტაზე. დახვევის მექანიზმებში გარკვევა ხელს შეუწყობს გლობალური მინიმუმის მოძიების პრობლემის გადაწყვეტას.

კომპიუტერული მეთოდების გამოყენების პროცესში გაჩნდა ვარაუდი, რომ ენერგეტიკული ზედაპირი, რომელზედაც გადის დახვევის გზა არც ისე ერთგვაროვანია და ადგილი აქვს მრავალ სტაბილურ თუ არასტაბილურ ინტერმედიანტზე გავლას (იხ. სურათი 6.1). რთული, მრავალ ადრეულ ინტერმედიანტზე გამავალი გზა განსაკუთრებით დამახასიათებელია იმ ცილათათვის, რომელთაც არ ახასიათებთ ენერჯის ცალსახად გამოხატული გლობალური მინიმუმი. განსაკუთრებით დიდ ინტერესს იწვევს ე.წ. მოლტენ გლობულის ტიპის ინტერმედიანტები.



სურათი 6.1 ცილის ენერგეტიკული ზედაპირი. წარმოდგენილია სამგანზომილებიანი ენერგეტიკული ზედაპირი. რეალურად ენერგეტიკული ფუნქცია მ კანზობილუმიანია იმა კანზიდერულია მისი კანზობილუმიან მიკარზეფაირზე. მ მოლეკულის კონფორმაციული თავისუფლების ხარისხია.

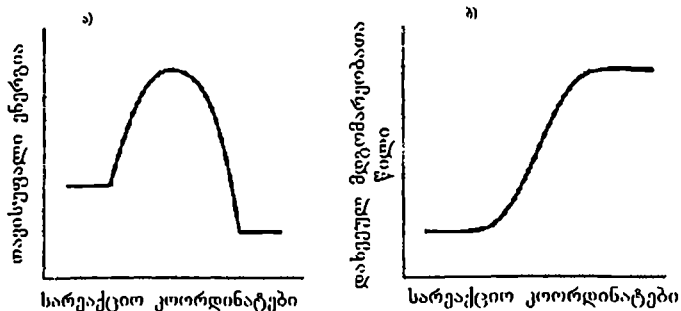
ინტერმედიანტების არსებობის შესაძლებლობა ასევე იკვეთება კოჟაქტორების არსებობის შემთხვევაში. კოჟაქტორებს მნიშვნელოვანი წვლილი შეაქვს ცილის სტაბილიზაციაში

დახვევის მექანიზმები და მოდელები

მიუხედავად მნიშვნელოვანი ძალისხმევისა, დიდი ხნის განმავლობაში დახვევის მექანიზმებზე ძალზედ ბუნდოვანი წარმოდგენა არსებობდა. მხოლოდ უახლოეს წარსულში მოხერხდა რეალური წარმოდგენის შექმნა იმაზე, თუ როგორ იღებს დიდი კონფორმაციული თავისუფლებების მქონე პოლიპეპტიდური ჯაჭვი უნიკალურ ნატიურ სტრუქტურას ასე სწრაფად და ეფექტურად. უფრო ღრმად რომ შევძლოთ ამ საკითხებში გარკვევა, საჭიროა დახვევის გზის კინეტიკურ და სტრუქტურულ თავისებურებებზე ნათელი წარმოდგენის შექმნა. შემთხვევითი შერჩევის მოდელზე დაფუძნებით შეფასებულ სიჩქარესთან შედარებით მაღალი დახვევის სიჩქარის არსებობის ასახსნელად შეიქმნა დახვევის უამრავი სხვადასხვა მოდელი.

მცირე, ერთდომენიანი ცილის, სტატისტიკური გორგლის მდგომარეობიდან ნატიურ სტრუქტურაში გადასვლის პროცესის აღმწერ მოდელებს ახასიათებთ ორი რამ - დახვევის პროცესის ორმდგომარეობიანობა და კოოპერატიულობა. ეს თვისებები არაა ურთიერთგამომრიცხავი. სწორედ კოოპერატიულობის დიდი ხარისხი, რომელიც გამოარჩევს მცირე ზომის ცილების დახვევის პროცესს, წარმოადგენს ორმდგომარეობით დახვევის საფუძველს. ორმდგომარეობიანობა გულისხმობს წონასწორობას კონკრეტულ, დახვეულ და გაშლილ მდგომარეობებს შორის და გადასვლას მხოლოდ ამ ორ მდგომარეობას შორის, სტაბილური შუალედური კონფორმაციული ფორმების (ინტერმედიანტების) გარეშე. რეაქციის კოორდინატებში ასეთი პროცესის აღწერა გვაძლევს სურათს ორი ენერგეტიკული მინიმუმიდან და ერთი ენერგეტიკული მაქსიმუმით (იხ. სურათი 6.2ა). მცირე ცილების უმრავლესობისათვის გამართლებულია დახვევის ამ მიახლოებაში წარმოდგენა. მაგრამ დიდი ცილების შემთხვევაში შეიძლება დავუშვათ, რომ დახვევა მიმდინარეობს სტაბილურ ინტერმედიანტებზე გავლით.

დახვევის მოდელების მეორე ზოგადი, სახასიათო თვისებაა კოოპერატიულობა, რაც გამოიხატება გადასვლის მრუდის სიგმოიდური ფორმით (იხ. სურათი 6.2ბ). კოოპერატიულობა ექსპერიმენტულად დადასტურებული მოვლენაა. ცილის სტრუქტურის ფორმირებასა და შენარჩუნებაში ძირითადად მონაწილეობენ სუსტი ძალები: წყალბადური ბმა, ელექტროსტატიკური ურთიერთქმედება, ვან-დერ-ვაალსის და ჰიდროფობული ძალები. მართალია, ცალკეული ურთიერთქმედების წილი ცილის სტაბილიზაციაში მცირეა, მაგრამ მრავალი



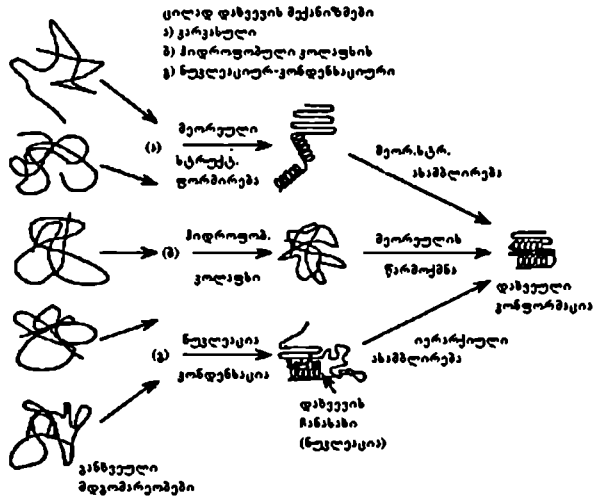
სურათი 6.2 ა) თავისუფალი ენერგიის სარეაქციო კოორდინატებზე დამოკიდებულება ორმდგომარეობით დახვევის შემთხვევაში; ბ) კოოპერატიულობისათვის დამახასიათებელი გადასვლის სიგმიოიდური მრუდი

ასეთი ურთიერთქმედება, მოქმედებს რა კოოპერატიულად, ხდება მნიშვნელოვანი მასტაბილიზირებელი ფაქტორი ცალკეული კონფორმაციებისათვის. კოოპერატიულობა გულისხმობს, რომ დახვეული მდგომარეობის სტაბილურობის განმსაზღვრელი ენერგია აღემატება ცალკეულ სუსტ ურთიერთქმედებათა ენერგიების უბრალო ჯამს.

პირველი მოდელი, რომელიც აღწერდა დახვევის კოოპერატიულობას იყო სპირალ-გორგალის გადასვლის თეორია. ეს თეორია გულისხმობდა კოოპერატიულობას პოლიპეპტიდურ ჯაჭვში მეზობლებს შორის და ითვლებოდა, რომ ამინმჟავას მიდრეკილება სპირალური კონფორმაციისაკენ იზრდება, თუ მისი მეზობელი უკვე იმყოფება სპირალურ კონფორმაციაში. ანუ სპირალის განვრცობა ხდება სწრაფად და კოოპერატიულად, თუ შემთხვევითი ნუკლეაციის პროცესში რამდენიმე მეზობელი ამინმჟავასათვის მოხერხდება სპირალური კონფორმაციის მიღება. მიუხედავად გარკვეული შეზღუდვებისა, ამ თეორიამ მაინც შეძლო შეექმნა საფუძველი შემდგომ შემოთავაზებული მოდელებისათვის. მათ შორის გამოვყოფთ: კარკასულ (ჩონჩხურ) მოდელს (იხ. სურათი 6.3ა), რომელიც გულისხმობს ან, ნუკლეაციის ზრდის მექანიზმს, ან დიფუზია-შეჯახებების მექანიზმს; ჰიდროფობული კოლაფსის მოდელს (იხ. სურათი 6.3ბ) და დახვევის კინეტიკის აღმწერ ძაბრულ მოდელს (იხ. სურათი 6.4) .

6.3 სურათზე სქემატურად წარმოდგენილია დახვევის მოდელები და შესაბამისი მექანიზმები.

ა) შეესაბამება კარკასულ მოდელს, რომლის მიხედვითაც ცილად დახვევა იწყება მეორეული სტრუქტურის ელემენტების ფორმირებით. ფორმირებას ადგილი აქვს მესამეული სტრუქტურის (კონტაქტების) ჩამოყალიბებამდე. შემდგომ, მეორეული სტრუქტურის ეს ელემენტები ერთიანდებიან ნატივურ სტრუქტურად ან დიფუზიისა და შეჯახებების გზით, ან სტრუქტურის თანამიმდევრული შენების

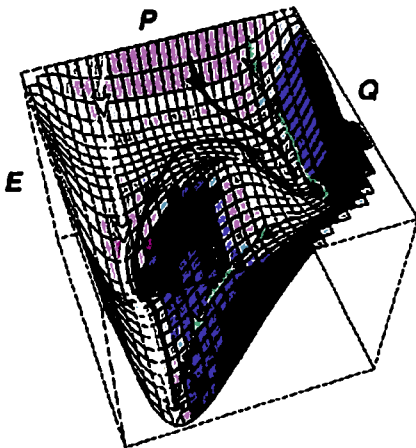


სურათი 6.3. დახვევის მოდელთა და შესაბამის მექანიზმთა სქემატური წარმოდგენა

გზით (ნუკლეაციის ზრდის მექანიზმი). ბ) შეესაბამება ჰიდროფობული კოლაფსის მოდელს, რომლის მიხედვითაც დახვევის პირველ ეტაპზე ადგილი აქვს ძირითადად ჰიდროფობული ეფექტით გამოწვეულ ჯაჭვის კოლაფსს. მეორეული სტრუქტურის ფორმირება იწყება კოლაფსური მდგომარეობის (გარკვეული მესამეული კონტაქტების) არსებობის პირობებში. გ) შეესაბამება ნუკლეაციურ კონდენსაციურ მექანიზმს, რომლის მიხედვითაც დახვევის ადრეულ ეტაპზე ფორმირებული დიფუზური ბირთვი აკატალიზებს და წარმართავს დახვევის შემდგომ ეტაპებს. ბირთვი თავიდან აერთიანებს რამდენიმე ამინომჟავას, რომელთა კონფორმაცია განაპირობებს მთელ რიგ სწორ მეორეულ სტრუქტურათა ურთიერთქმედებას, და ეს ნუკლეაცია (ჩონჩხი) ძირითადად სტაბილიზირდება თითქმის სწორი (ნატივური) მესამეული ურთიერთქმედებებით (კონტაქტებით). დიდი ცილების დახვევის პროცესში საწყის

ეტაზე შეიძლება ჩამოყალიბდეს რამდენიმე ნუკლეაცია.

ძაბრული მოდელი შემოთავაზებულ იქნა დახვევის კინეტიკის აღსაწერად. ამ მოდელის მიხედვით, საწყის ეტაპზე დგილი აქვს კონფორმაციულ თავისუფლებათა ხარისხების შემცირების პროცესს და ეს პროცესი დახვევის ნელ, სიჩქარის განმსაზღვრელ ფაზას შეესაბამება. ძაბრული მოდელის შესაბამისად ცილის დახვევის გზა გადის ენერგეტიკულ ზედაპირზე, რომელსაც ძაბრის ფორმა აქვს. ამ ძაბრის განიერი ნაწილი (ფუძე) შეესაბამება გაშლილ კონფორმაციათა დიდ სიმრავლეს, წვრილი ნაწილი (ცხვირი) კი, დახვეულ ნატივურ კონფორმაციას. ამ თეორიის მიხედვით გაშლილიდან დახვეული მდგომარეობისაკენ მოძრაობა შესაძლებელია მრავალი განსხვავებული გზით. ცილის მოლეკულა ირჩევს ან განსაკუთრებით ციცაბო (სწრაფ) ბილიკს, ან უფრო მიხვეულ-მოხვეულ, რამდენიმე ლოკალურ მინიმუმსა და მაქსიმიმზე (ენერგეტიკულ ბარიერზე) გამავალ (ნელ) გზას. სურათი 6.4 გვიჩვენებს ლიზოციმის დახვევის შესაბამის ძაბრს და ნატივური კონფორმაციისაკენ მიმავალ სწრაფ (ყვითელი) და ნელ (მწვანე და წითელი) ბილიკებს.



სურათი 6.4 ლიზოციმის დახვევის შესაბამისი ძაბრი (იხ. ტექსტი)

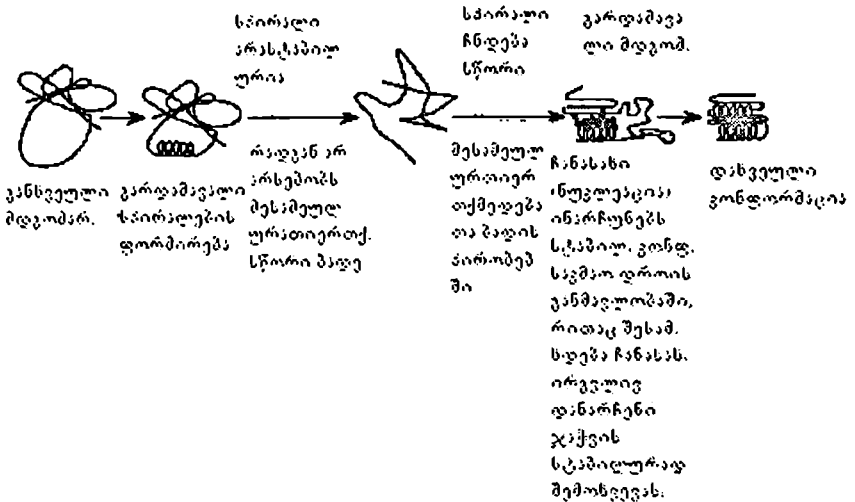
ყველა აქ წარმოდგენილი მოდელი შეიძლება განვრცობილ იქნეს ცილებზე, რომელთა დახვევასაც ახასიათებს მრავალი შუალედური, გარდამავალი მდგომარეობა. ცილად დახვევის ბოლო ეტაპი გულისხმობს გვერდითი ჯაჭვების

მკიდრო სიახლოვეში მოსვლასა და დიდი სიმკვირივის სტაბილური წარმონაქმნების ფორმირებას. ახლახან ჩატარებულმა ექსპერიმენტებმა აჩვენა, რომ მცირე ცილების დახვევა კარგად აღიწერება ნუკლეაცია-კონდენსირების მექანიზმით, რომელიც წარმოადგენს ნუკლეაციის ზრდისა და ჰიდროფობული კოლაფსის მოდელის შერწყმის შედეგს. განსხვავებით ნუკლეაციის ზრდის მექანიზმისა, სადაც დახვევის საწყის ეტაპზე გათვალისწინებულია ჩამოყალიბებული სტრუქტურული ბირთვის არსებობა, ნუკლეაციურ-კონდენსაციური მექანიზმის შემთხვევაში იგულისხმება, რომ დახვევის ბირთვი დიფუზურია და დასაწყისში მოიცავს რამდენიმე ამინოჟავას. დახვევის ასეთი ბირთვის (ჩონჩხის) კონფორმაცია სტაბილიზირდება ჯაჭვის გასწვრივ დაშორებულ ამინოჟავათა ურთიერთქმედების ხარჯზე. კარკასული მოდელის მსგავსად ნუკლეაციურ-კონდენსაციური მექანიზმის შემთხვევაშიც მეორეული სტრუქტურის ფორმირება დახვევის წარმართველი ფაქტორია, მაგრამ კარკასული მოდელისაგან განსხვავებით, ამ მექანიზმის მნიშვნელოვან ასპექტს წარმოადგენს მეორეულ და მესამეულ სტრუქტურათა ფორმირების კონკურენტულობა (პარალელურობა).

ამრიგად, დღეს ხელთ არსებულ ექსპერიმენტულ მონაცემებზე დაყრდნობით გამართლებულია ნუკლეაცია-კონდენსირების მექანიზმისა და ჰიდროფობული კოლაფსის მოდელის შერწყმა, და შედეგად, დახვევის აღწერა განზოგადებული ნუკლეაცია-კონდენსირების მექანიზმით, რომელიც მისაღებია მცირე ცილების უმრავლესობისათვის, და რომელიც დახვევის პროცესს წარმოგვიდგენს შემდგი ეტაპების სახით:

1. დახვევა იწყება კოლაფსით, ანუ ადგილი აქვს რამდენიმე ნუკლეაციის (ჩანსახოვანი სტრუქტურის) ჩამოყალიბებისა და ზრდის პროცესს. სწორედ ეს ნუკლეაციები წარმართავს დახვევის შემდეგომ ეტაპებს.
2. გარდამავალი მდგომარეობის შესაბამისი სტრუქტურა გულისხმობს ამინოჟავათა ერთი ან რამდენიმე ისეთი კლასტერის არსებობას, რომლებიც სტაბილიზირებულია მესამეული და მეორეული სტრუქტურული ურთიერთქმედებებით, და ამასთან ასეთ კლასტერებში, რეგულარულ მეორეულ სტრუქტურებში განთავსებულ ამინოჟავათა წილი გაცილებით დიდია, ვიდრე მოლეკულის დანარჩენ ნაწილში. ეს კლასტერები ან იდენტურებია დახვევის ბირთვებისა (ნუკლეაციებისა), ან იწარმოება სხვა სტრუქტურებთან მათი შერწყმის მეშვეობით. დიდი ცილების შემთხვევაში, შესაძლებელია ადგილი ჰქონდეს დახვევის რამდენიმე ბირთვის (ნუკლეაციის) გაერთიანებას .
3. არსებობს გარდამავალი მდგომარეობა, გარკვეულ კონფორმაციათა ერთობა, რომლის ფორმირებაც განაპირობებს დახვევის პროცესის სიჩქარის მალიმიტირებელ ეტაპს. სწორედ ამ ეტაპს მოყვება სწრაფი გადასვლა ნატივური სტრუქტურისაკენ.

სურათ 6.5-ზე სქემატურად გამოსახულია მცირე ცილების ნუკლეაცია-კონდენსირების მექანიზმით დახვევის პროცესის ამსახველი ჰიპოტეტური სურათი.



სურათი 6.5. ნუკლეაცია-კონდენსირების მექანიზმით დახვევისას მოვლენათა ჯაჭვის სქემატური წარმოდგენა.

როგორც სურათიდან ჩანს ადრეულ ეტაპზე, ჯერ კიდევ გაშლილი ჯაჭვის (სწორი მესამეული კონტაქტების ჯერ კიდევ არ ჩამოყალიბების) პირობებში წარმოქმნილი სპირალური სტრუქტურები განიცდიან ფლუქტუაციას. ანუ სპირალური სტრუქტურები ჯერ კიდევ არამყარადაა სტაბილური. ასეთი, ახლადჩამოყალიბებული, არასტაბილური სპირალის ირგვლივ მესამეულ სტრუქტურაში მიმდინარე ფლუქტუაციები აიძულებს მეორეულს (სპირალს) გაიშალოს. მრავალი ასეთი გაშლისა და კვლავ დახვევის, ანუ მრავალი მგდელობის შემდეგ, პოტენციური სპირალის გარშემო ბოლოსდაბოლოს მიიღწევა სწორი მესამეულ კონტაქტები (სტრუქტურა), რასაც მივყავართ სპირალის სტაბილურობის მნიშვნელოვან გაზრდამდე და ფლუქტუაციის დროის მნიშვნელოვან შემცირებამდე. ამრიგად, სწორი მეორეული და მესამეული ურთიერთქმედებებით გამყარებული, ნუკლეაციის (დახვევის ბირთვის) არსებობის დროის ზრდის პირობებში შესაძლებელი ხდება მის ირგვლივ

სტრუქტურის შემდგომი განვრცობა-გამკვირვება (კონდენსირება), ანუ ზრდა. ასე რომ, დახვევის სწორი ბირთვის ჩამოყალიბება აკატალიზებს დახვევის შემდგომ ეტაპებს.

პოლიპეპტიდური ჯაჭვის ცილად დახვევის პროცესში მონაწილე სხვა საშუალებები. შაპერონები

თანამედროვე ბიოლოგიის ერთ-ერთ მნიშვნელოვან ამოცანას ხაზოვანი ჰეტეროპოლიმერის, კერძოდ, პოლიპეპტიდური ჯაჭვის ცილად დახვევის მექანიზმებში გარკვევა წარმოადგენს. ასეთი ამოცანის გადაწყვეტის აუცილებლობა, როგორც ზემოთ იყო აღნიშნული, უკავშირდება ბევრი სხვადასხვა ბიოსამედიცინო პრობლემის გადაჭრის მოთხოვნას, მათ შორის იმ სიმპტომების დაძლევის მცდელობას, რომლებიც ხშირად წარმოიშობა *in vitro* სისტემებში გენურ-ინჟინრული მეთოდების გამოყენებით სინთეზირებული მრავალდომენიანი ან წინამორბედების სახით სინთეზირებული ცილების შესაბამისი პოლიპეპტიდური ჯაჭვების სწორ ნატივურ სტრუქტურად დახვევის პროცესში. უახლოეს წარსულში ითვლებოდა, რომ მხოლოდ ამინმჟავურ თანამიმდევრობაში თავმოყრილი ინფორმაცია საკმარისია ნატივური სივრცული სტრუქტურის მოსაწყობად, და შესაბამისად, მის წინასწარმეტყველებისათვის აღმოჩნდა, რომ საბოლოო, ბიოლოგიურად აქტიური სივრცული მოწყობის სახეზე გავლენას ახდენს მრავალი სხვა ფაქტორიც. ასე მაგალითად, გენეტიკური კოდის გადაგვარება, გარემო (რიბოსომული, თუ სხვა ცილების გარემოცვა), პოსტრანსლაციური მოდიფიკაციები და სხვა, რომლებიც განაპირობებს დახვევის უამრავ შესაძლებლობათა შორის იმ ერთადერთი სწორი გზის შერჩევას, რომელსაც მივყავართ ბიოლოგიურად აქტიურ სტრუქტურამდე. დახვევის მექანიზმების სრულად გაგებამდე მართალია, ჯერ კიდევ შორს ვართ, მაგრამ წარმატებები ამ მექანიზმების შესწავლის საქმეში სახეზეა.

მიუხედავად იმისა რომ მოკლე (დაახლოებით 100-150 ამინმჟავას მომცველი) პოლიპეპტიდური ჯაჭვების დახვევა სწორ ნატივურ სტრუქტურად სინჯარაში წარმოებს სპონტანურად, უჯრედში დახვევის წარმატებული პროცესი დამოკიდებულია სპეციალიზირებულ შაპერონულ ცილებზე. სულ ცოტა ხნის წინ შაპერონი მიიჩნეოდა კონკრეტული ფუნქციის მატარებელ ცილად, რომელიც აკონტროლებს პოლიპეპტიდური ჯაჭვების შემადგენლობაში მყოფ ამინმჟავათა ატომების სივრცეში დაახლოებისა და ცილოვან სტრუქტურად გაერთიანების პროცესს ან ამ სტრუქტურის დაშლის პროცესს. გაერთიანება ზოგადად გულისხმობს არაკოვალენტური კონტაქტების წარმოქმნას, ხოლო დისულფიდური ბმების წარმოქმნის აუცილებლობის შემთხვევაში კი, კოვალენტურისაც.

ამასთან, შაპერონი არ მოიაზრება საბოლოოდ ჩამოყალიბებული ნატივეური ცილის სტრუქტურის შემადგენელ ნაწილად. ჯერ კიდევ ბოლომდე არაა გარკვეული მექანიზმები, რომელთა საშუალებითაც შაპერონი ახორციელებს თავის ფუნქციას. ორი ყველაზე კარგად შესწავლილი შაპერონული ოჯახის, შაპერონინებისა და მადალტემპერატურული შოკის (Hsp70) ცილების ფუნქციონირების მექანიზმების შესწავლამ გამოავლინა მათი არაპირდაპირი და არასტერიული როლი დახვევის პროცესში. შაპერონების სხვა ოჯახს წარმოადგენს წინამორბედი ფრაგმენტები და პროტეაზები, რომელთა გავლენა დახვევის პროცესზე უკვე აქტიურია და სტერიული ინფორმაციით უზრუნველყოფას გულისხმობს.

ამრიგად, შაპერონები ცილებია, რომლებიც მონაწილეობენ მთელი რიგი პოლიპეპტიდური ჯაჭვების ცილად დახვევის პროცესში და ე.ი. გარკვეულწილად თავისი წვლილი შეაქვთ სწორი ნატივეური სტრუქტურის ფორმირებაში. უნდა აღინიშნოს, რომ შაპერონების ოჯახს ასევე მიეკუთვნება ცილები, რომლებიც სინთეზირდება ექსტრემალური გარემო პირობების წარმოქმნის საპასუხოდ და ხელს უშლის ნაწილობრივ დენატურირებული ცილის აგრეგირების პროცესს (მაგალითად, ტემპერატურული შოკის ცილები).

დღეისათვის, შაპერონული ცილების სტრუქტურისა და ფუნქციონირების თავისებურებათა შესასწავლად ჩატარებული კვლევების შედეგად დაგროვილი ინფორმაცია იძლევა საშუალებას, რომ შაპერონული ცილები დახვევის პროცესზე ზემოქმედების ხასიათის მიხედვით დაიყოს ორ ჯგუფად :

1) პირველ ჯგუფში შედის ცილები, რომლებიც არ უზრუნველყოფს დახვევის პროცესს სტერიული ინფორმაციით და აქედან გამომდინარე, აქვთ მხოლოდ არაპირდაპირი, პასიური გავლენა ამ პროცესზე, რაც გულისხმობს მხოლოდ აგრეგაციისაგან დაცვის ფუნქციას. დაცვას საჭიროებს ან ახლად სინთეზირებული ჯაჭვები, ან ნაწილობრივ დენატურირებული ცილები. თავის მხრივ ეს ჯგუფი იყოფა ორ ქვეჯგუფად: ა) ახლად სინთეზირებულ ჯაჭვთან მიერთების უნარის მქონე ცილები. მაგალითად Hsp 70 (heat shock protein 70) და მსგავსი სისტემები და ბ) ცილინდრული შაპერონინები - ცილოვან მოლეკულათა ისეთი აგრეგატები, რომლებიც ქმნიან გარკვეული ზომის სპეციალურ შიგა სივრცეს ახლადსინთეზირებული ან ნაწილობრივ დენატურირებული ცილის ჯაჭვის სწორად დახვევის ხელშეწყობის მიზნით. დღევანდელი მონაცემებით ასეთი ტიპის შაპერონების მონაწილეობა აუცილებელია ძირითადად პროკარიოტული წარმოშობის ზოგიერთი პოლიპეპტიდური ჯაჭვის პოსტტრანსლაციურად დახვევის პროცესში.

2) მეორე ჯგუფი აერთიანებს წინამორბედი ჯაჭვების სახით სინთეზირებული ცილების ფრაგმენტებს, რომელთა პოსტტრანსლაციური მოცილების შემდეგ ხდება ცილის ნატივეური სტრუქტურის ჩამოყალიბება. ამ შემთხვევაში წინამორბედი ფრაგმენტი დახვევის პროცესს გარკვეული სტერიული ინფორმაციით უზრუნველყოფს. ანუ ფრაგმენტის გავლენა დახვევის პროცესზე მკაცრად სპეციფიკურია და სწორედ მოცემული თანამიმდევრობის მქონე წინამორბედი ჯაჭვია საჭირო ცილის სწორი ნატივეური სტრუქტურის ჩამოსაყალიბებლად.

შაპერონების, როგორც დახვევის პროცესში რედაქტორული ფუნქციის მატარებელი ცილების არსებობის შესახებ ინფორმაცია სულ უახლოესი წარსულის მონაპოვარია. მანამდე მიიჩნეოდნენ, რომ ტრანსკრიფციისა და ტრანსლაციის პროცესში წარმოშობილი შედომების აცილება, და დახვევის პროცესების სწორად მიმდინარეობა კორექტირდება მხოლოდ ენერგეტიკული მექანიზმებით. თუ არ განვიხილავთ მოკლე ჯაჭვებს, შეიძლება ითქვას, რომ პოლიპეპტიდური ჯაჭვი ცილად დახვევის პროცესში გადის მრავალ შუალედურ სტრუქტურაზე, ე.წ. "გამლვადი გლობულების" ტიპის სტრუქტურებზე

(molten globules). უკანასკნელი ტიპის მდგომარეობას ახასიათებს მეტნაკლებად ნატივი მსგავსი მეორეული სტრუქტურა, მაგრამ მეორეული სტრუქტურის ელემენტები ჯერ კიდევ არ არიან გაერთიანებული სწორ, ნატიურ მესამეულ სტრუქტურაში. ასეთი, არანატივი სივრცული სტრუქტურის მქონე პოლიპეტიდის მთელი რიგი ჰიდროფობული ამინმჟავებისა ექსპონირებულება გამხსნელისკენ, მაშინ, როცა ნატიურ სტრუქტურაში ისინი დამალულია. ანუ "გამლვადი გლობული" ტიპის სტრუქტურა შეიძლება ქმნიდეს ჯაჭვთა აგრეგაციისათვის ხელსაყრელ პირობებს. ასეთი პირობები განსაკუთრებით საშიში ხდება *in vivo*, რამეთუ მაღალია ლოკალური კონცენტრაცია დაუხვეველი პოლიპეტიდებისა პოლირიბოსომაზე და ზოგადად მაკრომოლეკულებისა ციტოზოლში. აქედან ნათელია, რომ უჯრედში დახვევის პროცესში მყოფი პოლიპეტიდების აგრეგაციის აღმართმა უფრო დიდა, ვიდრე *in vitro*. შაპერონი უერთდება ექსპონირებულ ჰიდროფობულ ამინმჟავებს და იცავს ცილის ჯაჭვებს აგრეგაციისაგან. ამრიგად, ახლად სინთეზირებულ პოლიპეტიდურ ჯაჭვებს (არასწორად დახვეულსაც კი) უნარჩუნდებით სწორად დახვევის შესაძლებლობა.

უკანასკნელი მონაცემების თანახმად ციტოზოლში მოქმედებს როგორც შაპერონებით მართული, ისე უშაპერონო დახვევის მექანიზმები. თუ ტრანსლირებული ჯაჭვებისათვის პირველი ჯგუფის პირველი ქვეჯგუფის შაპერონების დახმარებით სწორად დახვევის უნარის შენარჩუნება ფართოდ გავრცელებული მოვლენაა, ცილინდრული ანუ მეორე ქვეჯგუფის შაპერონები მონაწილეობენ მხოლოდ მცირე რაოდენობა, აგრეგაციისადმი ძალზედ მგმობიარე ცილების დახვევაში.

საერთოდ მეორე ქვეჯგუფის შაპერონების (ისინი ძირითადად ბაქტერიალური ცილებია. მაგალითად GroEL ოჯახის) ზომები ზღუდავს დასახვევი ცილის ზომებს 55 kDa-მდე, ანუ 500 ამინმჟავამდე. *E. coli* შეიცავს დაახლოებით 4300 სახეობის ცილას, რომელთაგან დაახლოებით 60% (2600) ციტოპლაზმაშია. ამ ცილების საშუალო ზომა 35kDa-ია (317 ამინმჟავა) და მხოლოდ 13%-ის ზომა აღემატება 55 kDa. რადგანაც დიდი ცილების უმრავლესობა მემბრანული ცილებია, ციტოპლაზმაში გვაქვს მხოლოდ 300-მდე 55 kDa-ზე მეტი წონის ხსნადი ცილა. ამრიგად, *E. coli*-ის ცილების უმრავლესობა შესაბამისობაშია შაპერონების (GroEL-ების) დამცავი უზნების ზომებთან.

ევკარიოტული უჯრედი შეიცავს მულტიდომენური ცილების უფრო დიდ რაოდენობას, ვიდრე ბაქტერიული. ასე მაგალითად, საფუარის უჯრედის ცილათა საშუალო ზომაა 53 kDa (496 ამინმჟავა) და ამასთან თითქმის 38% 55kDa-ზე დიდია. რომელთაგან 1445 ცილა ხსნადია. ცილის დომენთა ზომების განაწილება ერთგვაროვანი მთელი ცილების სამყაროსათვის და 50-300 ამინმჟავას ფარგლებშია. ამრიგად, გენომის მიერ გრძელი პოლიპეტიდური ჯაჭვების კოდირება ნიშნავს მრავალდომენიანი ცილის კოდირებას. *in vitro* ხშირად გამწვანებულია ასეთი პოლიპეტიდების ცილად სწორი დახვევა, რამაც წარმოშვა მოსაზრება, რომ ევკარიოტული უჯრედი უზრუნველყოფილია საკმაო რაოდენობა სპეციალური (რადაც ზოგადი მექანიზმით მომუშავე), სწორი დახვევის უზრუნველმყოფი შაპერონებით. თუმცა, მიუხედავად დიდი მდებლობისა, ასეთი შაპერონების აღმოჩენა ვერ მოხერხდა. მაგალითად, დღეისათვის ცნობილი ციტოზოლური შაპერონი, Tric, უმრავლეს ევკარიოტულ უჯრედებში გვხვდება ძალზე მცირე რაოდენობით და დღეისათვის ცნობილ მის სუბსტრატებს წარმოადგენს მხოლოდ აქტინი და ტუბულინი. ამრიგად ალტერნატიული შაპერონული სისტემა, რომელსაც გააჩნია რაიმე ზოგადი მოქმედების მექანიზმი ევკარიოტულ ციტოზოლში ჯერ არ აღმოჩენილა.

დომენთა კოტრანსლაციური დახვევა - როგორ შეიძლება დამლეულ იქნეს პოლიპეპტიდების ზომის პრობლემა დახვევის პროცესში

მოდულურ (მრავალდომენიან) ცილებში დომენები ერთმანეთს უერთდება მოქნილი ლინკურებით. პრინციპში ასეთი ცილების დახვევას შეიძლება ქონდეს კოტრანსლაციური ხასიათი. ანუ დახვევა შეიძლება მიმდინარეობდეს სინთეზის პროცესში (პარალელურად), რიბოსომიდან დომენის გამოთავისუფლებისთანავე. შეიძლება სწორედ კოტრანსლაციური დახვევა წარმოადგენს იმ მექანიზმს, რომელიც საშუალებას იძლევა პოლიპეპტიდის ცილად დახვევის პროცესი წარმართოს შაპერონინების გარეშე. ანუ გრძელი პოლიპეპტიდის დახვევა შეიძლება დაყვანილ იქნეს დამოუკიდებელი მოდულების ანუ დომენების სპონტანურ დახვევაზე. ამრიგად, კოტრანსლაციური, დომენური სახით დახვევის არსებობა საშუალებას გვაძლევს ავხსნათ, თუ რატომ არ საჭიროებს დახვევის პროცესში შაპერონინებს ეუკარიოტული უჯრედის ციტოზოლური ცილების უმრავლესობა.

კოტრანსლაციური დახვევის განხორციელებისათვის აუცილებელი შესაბამისი სტრუქტურული უზრუნველყოფა.

ტერმინი კოტრანსლაციური დახვევა ანუ დახვევა ტრანსლაციის პროცესში გულისხმობს სტაბილური, ნატივურის მსგავსი მესამეული კონტაქტების ფორმირებას ტრანსლაციის (ჯაჭვის სინთეზის) პროცესში. ტრანსლაციის პროცესში გარკვეული რაოდენობით შეორეული სტრუქტურის ფორმირებასა და ჯაჭვის გარკვეულ კომპაქტიზაციას ადგილი აქვს ერთდომენიანი ცილების შემთხვევაშიც, მაგრამ წარმატებული, ბოლომდე მიყვანილი დახვევა ამ შემთხვევაში მოითხოვს მთელ რიგ პოსტტრანსლაციურ პროცედურებს. ამრიგად, კოტრანსლაციურ გარემოში ცალკეული დომენის ნატივური სტრუქტურის ფორმირებისათვის შესაძლოა საჭირო იყოს რიბოსომიდან დომენის სრულ თანამიმდევრობასთან ერთად მისი C-ბოლოს შემდგომი, საკმაო სიგრძის ფრაგმენტის ჩამოსვლა. C-ბოლოს შემდგომი ფრაგმენტი შეძლებს დაიცვას დომენის დახვევის პროცესი რიბოსომული ზემოქმედებისაგან და ამასთან უზრუნველყოფს დახვევას თავისუფლების საჭირო ხარისხებით. ასეთი დამეხების პირობებში გასაგები ხდება თუ ერთდომენიანი ცილების დახვევა რატომ ატარებს პოსტტრანსლაციურ ხასიათს. ერთდომენიანი ცილის ტრანსლაციის პროცესში არ ვართ უზრუნველყოფილი თავისუფალი C-ბოლო ფრაგმენტით. შესაბამისად, სწორი დახვევა მოითხოვს სრულად სინთეზირებული და რიბოსომიდან გამოთავისუფლებული ჯაჭვის არსებობას.

იმ მულტიდომენური ცილების დახვევისას, რომელთა დომენები კონსტრუირებულია პოლიპეპტიდური ჯაჭვის არათანმიმდევრულად განლაგებული სეგმენტებით, შეიძლება არსებობდეს დამატებითი პრობლემები. აქ უწყვეტი სინთეზის პროცესში წარმოიქმნება დომენის ნაწილი, შემდეგ წყდება ამ დომენის შემადგენელი უზნის სინთეზი და იწყება სხვა დომენის ან მისი ნაწილის სინთეზი, შემდეგ კვლავ წინა დომენისა და ა.შ. ასეთი ცილისათვის ნატივური სტრუქტურის კოტრანსლაციურად ფორმირება ან შეუძლებელია, ან შესაძლებელი მხოლოდ პოლიპეპტიდის დიდი ნაწილის სინთეზირების შემდეგ. ასევე შესაძლებელია, რომ ასეთი ცილის კოტრანსლაციური სინთეზი დამთავრდეს კოლაფსით, როცა რიბოსომაზე ნატივურისაგან განსხვავებული დომენური სტრუქტურა წარმოიქმნება. ასეთი ახლად სინთეზირებული ჯაჭვები ხასიათდება გამიშვებულ ჰიდროფობული ზედაპირით, რომელსაც შესაძლოა ცნობს შაპერონინი TRiC. მსგავს პროცესებს შეიძლება

ადგილი ქონდეს ისეთი ცილების შემთხვევაშიც, სადაც თანამიმდევრული დომენები ერთმანეთს ასტაბილიზირებენ ჰიდროფობული ინტერფეისების მეშვეობით. მართლაც ნაჩვენებია, რომ TRiC მონაწილეობს აქტინის დახვევაში, ანუ ცილისა, რომელსაც არ შეუძლია დაეხვეს კოტრანსლაციურად. ამრიგად TRiC შესაძლოა უზრუნველყოფდეს გაკვივის გასწვრივ არათანმიმდევრულად განლაგებული ფრაგმენტებისაგან კონსტრუირებული დომენების შემცველი ცილების დახვევას.

"დახვევის უნარის გადაცემის" ჰიპოტეზა

სრულიად მოულოდნელად აღმოჩნდა, რომ ბაქტერიალური ტრანსლაციური სისტემა ვერ უზრუნველყოფს ორდომენიანი, ჰიბრიდული ცილის კოტრანსლაციურად სწორ დახვევას. ალბათ იგივე მიზეზის გამო E.coli-ში ადგილი აქვს მრავალდომენიანი (მოდულური) ეუკარიოტული პოლიპეპტიდების არასწორად დახვევის მოვლენას. E.coli-ის ორდომენიანი ცილის, OmpR-ის დახვევა ბაქტერიალურ უჯრედში პოსტტრანსლაციურია, ეუკარიოტულ უჯრედში კი, კოტრანსლაციური. თუ რამდენად შეიძლება უკანასკნელი მონაცემების განზოგადება სხვა ცილებზე ჯერ კიდევ გასარკვევია. შეიძლება მხოლოდ იმის თქმა, რომ ბაქტერიებში დახვევის პოსტტრანსლაციური ბუნების სასარგებლოდ ლაპარაკობს მრავალი მონაცემი. მაგალითად, E.coli-ის ახლადსინთეზირებული ცილების უმრავლესობას შეუძლია GroEL-თან ურთიერთქმედება. დღეისათვის არ შეიძლება იმის გამოჩვენება, რომ მთელი რიგი ცილები, განსაკუთრებით მრავალდომენიანი, როგორცაა მაგალითად β-გალაქტოზიდაზა, ბაქტერიებში განიცდის ისეთ კოტრანსლაციურ დახვევას, რომელიც შეიძლება განსხვავდებოდეს ეუკარიოტული, თანამიმდევრულ-დომენური დახვევისაგან. ბაქტერიალურ და ეუკარიოტულ ტრანსლაცია-დახვევის სისტემებს შორის განსხვავებების გარკვევა უახლოესი მომავლის საქმეა. ამჯერად კი, ჩვენს ხელთ არსებული მონაცემები საშუალებას გვაძლევს ვთქვათ შემდეგი. ბაქტერიებში ელონგაციის სიჩქარე (20 ამინოჰაქს/წმ.) გაცილებით აღემატება ეუკარიოტულს, და შეიძლება მნიშვნელოვნად ზღუდავდეს კოტრანსლაციური დახვევის შესაძლებლობას.

მიჩნეულია, რომ მოდულური პოლიპეპტიდების ევოლუციის მექანიზმებს საფუძვლად უდევს გენთა შემთხვევითი შერწყმა - დუბლირება, ე.ი, შესაძლებელია "დახვევის უნარის გადაცემა". თუ ბაქტერიებში კოტრანსლაციური დახვევა ნაწილობრივ მაინც შეზღუდულია, მაშინ გენთა შემთხვევითი არევით გამოწვეული მრავალდომენიანი ცილების ევოლუციური წარმოქმნა აქ ნაკლებად მოსალოდნელია, ვიდრე იმ ორგანიზმებში, სადაც დომენთა თანამიმდევრული დახვევა უზრუნველყოფილია. კონკრეტული დახვევის უნარი ახალ კომპოზიტურ ცილას საშუალებას აძლევს იყოს შემთხვევითი შერჩევის ობიექტი. ამრიგად, კოტრანსლაციური დახვევის დაბალი ეფექტურობით შეიძლება აიხსნას მრავალდომენიანი ცილების სიმცირე პროკარიოტებში, და ასევე ის ფაქტი, თუ რატომ ვერ ხერხდება მთელი რიგი მრავალდომენიანი ეუკარიოტული ცილების სწორი დახვევა პროკარიოტულ უჯრედში ექსპრესიისას. უფრო მეტიც, ბაქტერიალური უჯრედისათვის დამახასიათებელი რეპროდუცირებისა და სინთეზის შედარებით მაღალი სიჩქარეები ალბათ კარგ შესაბამისობაშია პოსტტრანსლაციური დახვევის მექანიზმებთან. შეიძლება ითქვას, რომ ბაქტერიალური მრავალდომენიანი ცილები (და ასევე მათ შეზღუდვებულადაა მყოფი დომენები) შედეგაა პოსტტრანსლაციური დახვევის უნარის მიხედვით ეფექტური შერჩევისა. ამ დასკვნასთან კარგ შესაბამისობაშია ბაქტერიალური მრავალდომენიანი ცილების უფრო დიდი უნარი (იგივე სიგრძის მრავალდომენიან ეუკარიოტულ ცილებთან შედარებით) სპონტანური კვლავდახვევისა (შესაძლოა შაპერონებით მართული). ამრიგად,

წასაკითხად რაკომენდირებული ლიტერატურა

- Alberts, Bruce; Alexander Johnson, Julian Lewis, Martin Raff, Keith Roberts, and Peter Walters (2002). *Molecular Biology of the Cell; Fourth Edition*. New York and London: Garland Science
- Ghosh A, Bansal M (2003). "A glossary of DNA structures from A to Z". *Acta Crystallogr D Biol Crystallogr* 59 (Pt 4): 620–6.
- Barciszewski J, Frederic B, Clark C (1999). *RNA biochemistry and biotechnology*. Springer, 73–87.
- Lee JC, Gutell RR (2004). "Diversity of base-pair conformations and their occurrence in rRNA structure and RNA structural motifs". *J. Mol. Biol.* 344 (5): 1225–49
- Branden C, Tooze J. (1999). *Introduction to Protein Structure* 2nd ed. Garland Publishing: New York, NY
- Alberts, Bruce; Alexander Johnson, Julian Lewis, Martin Raff, Keith Roberts, and Peter Walters (2002). "The Shape and Structure of Proteins", *Molecular Biology of the Cell; Fourth Edition*. New York and London: Garland Science.
- Lee S, Tsai F (2005). "Molecular chaperones in protein quality control". *J. Biochem. Mol. Biol.* 38 (3): 259-65.
- Rose G, Fleming P, Banavar J, Maritan A (2006). "A backbone-based theory of protein folding". *Proc. Natl. Acad. Sci. U.S.A.* 103 (45): 16623-33.

თავი 3

მონაცემთა ბაზები

7. ნუკლეოტიდურ თანამიმდევრობათა შესახებ მონაცემთა ბაზები.

ჩვენ ვცხოვრობთ პოსტგენომურ ერაში, როდესაც დადგენილია ასობით ორგანიზმის გენომური თანამიმდევრობა და გაცილებით მეტი დადგენის პროცესშია. ორგანიზმის . გენომური თანამიმდევრობების ცოდნა ნიშნავს მილიონობით მაკრომოლეკულისა თუ მისი ფრაგმენტის ნუკლეოტიდური თუ ამინმჟავური თანამიმდევრობის ცოდნას. შესაბამისად შეიქმნა აუცილებლობა ჩამოყალიბებულიყო თანამიმდევრობათა შესახებ მონაცემთა ბაზები, რომლებსაც უნდა ეკისრათ არა მარტო უკანასკნელი თანამიმდევრობების (ასოთა რიგების სახით) , არამედ მრავალი სხვა მათთან დაკავშირებული ინფორმაციის საცავის როლი. ამ ბაზებში გაერთიანებულ თანამიმდევრობათა რიცხვი ექსპონენციალურად იზრდება. ასევე ვითარდება კომპიუტერული ტექნიკა, რაც საშუალებას იძლევა სწრაფად მოხდეს ამ ბაზებიდან მონაცემთა მოძიება და ანალიზი. ამ პარაგრაფში შევეცდებით მოკლედ ავღწეროთ ნუკლეოტიდურ თანამიმდევრობათა შესახებ მონაცემთა საზოგადოდ აღიარებული ბაზები.

GenBank

GenBank® წარმოადგენს გენურ თანამიმდევრობათა შესახებ National Institutes of health(NIH) -ის მონაცემთა ბაზას, ანუ დნმ თანამიმდევრობათა ანოტირებულ საჯარო კოლექციას. 2006 წლის აგვისტოს მონაცემებით GenBank® -ის ტრადიციული განყოფილებები აერთიანებს 65,369,091,950 ფუძეს, რომლებიც ქმნის 61,132,599 ანოტირებულ თანამიმდევრობას, ხოლო WGS (Whole Genome Shotgun) განყოფილება კი, შესაბამისად, 80,369,977,826 ფუძესა და 17,960,667 თანამიმდევრობას.

GenBank -ის მიმდინარე გამოშვება და მისი დახასიათება შეგიძლიათ იხილოთ National Center for Biotechnology Information (NCBI)-ის ვებ-გვერდზე, რომელიც ყოველთვიურად ახლდება. GenBank არის იაპონური ბანკის (DNA Data Bank of Japan (DDBJ)), ევროპის მოლეკულური ბიოლოგიის ლაბორატორიის (EMBL) ნეკლეოტიდურ თანამიმდევრობათა ბაზის და NCBI-ის ნუკლეოტიდურ თანამიმდევრობათა ბაზას შორის ყოველდღიური თანამშრომლობის შედეგი.

GenBank-ში თითოეული ანოტირებული თანამიმდევრობა წარმოდგენილია ცალკეული ჩანაწერის (ფაილის) სახით. ფაილში თავმოყრილი ინფორმაცია შეიძლება დავყოთ ძირითად ინფორმაციად, რომელიც აუცილებელია და გვხვდება ნებისმიერი თანამიმდევრობის შესაბამის ჩანაწერში, და დამხმარე ინფორმაციად, რომელიც არაა აუცილებელი და ჩანაწერებში შეიძლება გვხვდებოდეს ან არა. ქვევით წარმოვადგენთ აუცილებელ ინფორმაციას ჩამონათვალსა და ფორმატს. შემდეგ კი, კონკრეტულ მაგალითზე განვიხილავთ GenBank-ში ანოტირებულ თანამიმდევრობისათვის გაკეთებული ჩანაწერის (ფაილის) სრულ სახეს. ჩანაწერში თანამიმდევრობის შესახებ ინფორმაციები წარმოდგენილია ცალკეული ველების ანუ სტრიქონთა ჯგუფების სახით.

აუცილებელ ინფორმაციულ ველთა და შესაბამის სტრიქონთა ფორმატი და ჩამონათვალი შემდეგია:

ა) ველ LOCUS-ს განსაზღვრავს სტრიქონი, რომლის პირველ 5 პოზიციამდე განლაგებულია სიტყვა LOCUS. თანამიმდევრობის სახელი ან იდენტიფიკატორი მოცემულია სტრიქონის 13-25 პოზიციებში. თანამიმდევრობის სიგრძე ამავე სტრიქონის 23-29 პოზიციებში.

ბ) ველ DEFINITION-ს განსაზღვრავს სტრიქონი, რომლის პირველ 10 პოზიციამდე განლაგებულია სიტყვა DEFINITION. ამ სტრიქონის 13-80 პოზიციას მოიცავს თავისუფალ ტექსტს, რომელიც ახდენს თანამიმდევრობის იდენტიფიცირებას.

გ) ველ ACCESSION-ს განსაზღვრავს სტრიქონი, რომლის პირველ 9 პოზიციამდე განლაგებულია სიტყვა ACCESSION. 13-18 პოზიციას მოიცავს პირველად საინდეტიფიკაციო ნომერს.

დ) ველ ORIGIN-ს განსაზღვრავს სტრიქონი, რომლის პირველ 6 პოზიციამდე განლაგებულია სიტყვა ORIGIN, რაც მიუთითებს, რომ შემდეგი სტრიქონიდან იწყება უშუალოდ ნუკლეოტიდური თანამიმდევრობის წარმოდგენა. სხვა რაიმე ჩანაწერი ამ სტრიქონში არ გვხვდება.

ე) თანამიმდევრობის მომცველი სტრიქონები, როგორც აღვნიშნეთ, იწყება უშუალოდ ORIGIN სტრიქონის შემდეგ. თანამიმდევრობის შესაბამის სტრიქონებში ჩანაწერები იწყება მე-11 პოზიციიდან. პირველი 9 ან თავისუფალია ან მოიცავს

სტრიქონში განლაგებული პირველი ნუკლეოტიდის რიგით ნომერს. თითოეულ სტრიქონში თანამიმდევრობა წარმოდგენლია 10-ასოიანი ბლოკების სახით. ბლოკები ერთმანეთისაგან გამოყოფილია გამოტოვებებით. ამრიგად, თანამიმდევრობების შესაბამის სტრიქონში პირველი ასო (ნუკლეოტიდი) გვხვდება მე-11 პოზიციაში, ხოლო ბოლო ასო 75-ე პოზიციაში (ცხადია თუ სტრიქონი არაა ბოლო და არასრულად შევსებული).

ე) ჩანაწერის ბოლო სტრიქონი, რომლის პირველ 2 პოზიციაში განლაგებული ნიშანი //, მიგვითითებს თანამიმდევრობის დასრულებაზე.

ახლა მაგალითის სახით წარმოგიდგენთ GenBank-დან *Saccharomyces cerevisiae* -ის გენომის ფრაგმენტის შესაბამის ჩანაწერს (ფაილს) და მასში მოცემულ ინფორმაცთა განხილვის საფუძველზე შევეცდებით შეგიქმნათ წარმოდგენა GenBank-ის ჩანაწერებში თავმოყრილი ინფორმაციის შესახებ.

LOCUS SCU49845 5028 bp DNA PLN 21-JUN-1999

DEFINITION *Saccharomyces cerevisiae* TCP1-beta gene, partial cds, and Axl2p (AXL2) and Rev7p (REV7) genes, complete cds.

ACCESSION U49845

VERSION U49845.1 GI:1293613

KEYWORDS .

SOURCE *Saccharomyces cerevisiae* (baker's yeast)

ORGANISM *Saccharomyces cerevisiae*

Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Saccharomyces.

REFERENCE 1 (bases 1 to 5028)

AUTHORS Torpey,L.E., Gibbs,P.E., Nelson,J. and Lawrence,C.W.

TITLE Cloning and sequence of REV7, a gene whose function is required for
DNA damage-induced mutagenesis in *Saccharomyces cerevisiae*

JOURNAL *Yeast* 10 (11), 1503-1509 (1994)

PUBMED 7871890

REFERENCE 2 (bases 1 to 5028)

AUTHORS Roemer,T., Madden,K., Chang,J. and Snyder,M.

TITLE Selection of axial growth sites in yeast requires Axl2p, a novel
plasma membrane glycoprotein

JOURNAL *Genes Dev.* 10 (7), 777-793 (1996)

PUBMED 8846915

REFERENCE 3 (bases 1 to 5028)

AUTHORS Roemer,T.

TITLE Direct Submission

JOURNAL Submitted (22-FEB-1996) Terry Roemer, Biology, Yale University, New
Haven, CT, USA

FEATURES Location/Qualifiers

source 1..5028

/organism="Saccharomyces cerevisiae"

/db_xref="taxon:4932"

/chromosome="IX"

/map="9"

CDS <1..206

/codon_start=3

/product="TCP1-beta"

/protein_id="AAA98665.1"

/db_xref="GI:1293614"

/translation="SSIYNGISTSGLDLNNGTIADMRQLGIVESYKLRVAVSSASEA
AEVLLRVDNIIRARPRTANRQHM"

gene 687..3158

/gene="AXL2"

CDS 687..3158

/gene="AXL2"

/note="plasma membrane glycoprotein"

/codon_start=1

/function="required for axial budding pattern of S.
cerevisiae"

/product="Axl2p"

/protein_id="AAA98666.1"

/db_xref="GI:1293615"

/translation="MTQIQISLLLTATISLHLVAVATPYEAYPIGKQYPPVARVNESF

TFQISNDTYKSSVDKTAQITYNCFDLPSWLSFSSSRTFSGEPSSDLLSDANTTLYFN
VILEGTDASDSTSLNNTYQFVVTNRPSISLSDFNLLALLKNYGYTNGKNAKLPNE
VFNVTFDRSMFTNEESIVSYGRSQLYNAPLPNWLFDSGELKFTGTAPVINSALAPE
TSYSFVILATDIEGFAVEFEFELVIGAHQLTTSIQNSLIINVTDTGNSVSYDLPLNYV
YLDLDDPISSDKLGSINLLDAPDWVALDNATISGSVPDELLGKNSNPANFSVSIYDTYG
DVIYFNFEVVSITDLFAISSLPNINATRGWFSYFLPSQFTDYVNTNVSLEFTNSSQ
DHDWVKFQSSNLTLAGVFPKNFDKLSLGLKANQGSQSQELFYNIIGMDSKITHSNHSA
NATSTRSSHSTSTSSYTSSTYAKISSTSAATSSAPAA LPAANKTSSHNKKAVALA
CGVAIPLGVILVALICFLIFWRRRRENPDENLPHAISGPLDNNPANKPNQENATPLN
NPFDDDASSYDDTSIARRLAA LNTLKLDNHSATESDISSVDEKRDSLSGMNTYNDQFQ
SQSKEELLAKPVPQPESPFFDPQNRSSSVYMDSEPAVNKSWRYTGNLSPVSDIVRDS
YGSQKTVDTEKLFDELEAPEKEKRTSRDVTMSSLDPWNSNISPSVVRKSVTPSPYNTK
HRNRHLQNIQDSQSGKNGITPTTMTSSSDDFVVPKDGENFCWVHSMEPDRRPSKKRL
VDFS NKS NVNVGQVKDIHGRIPLEML"

gene complement(3300..4037)

/gene="REV7"

CDS complement(3300..4037)

/gene="REV7"

/codon_start=1

/product="Rev7p"

/protein_id="AAA98667.1"

/db_xref="GI:1293616"

/translation="MNRWVEKWLRVYLKCYINLILFYRNVPYPPQSFYDITYYQSFNLPQ

FVPINRHPALIDYIEELILDVLSKLTHTVYRFSICINKNKNDLCIEKYVLDSELQHVD

KDDQITETEVFDEFRRSSLSLIMHLEKLPKVNDTITFEA VINAIELELGHKLDNR
RVDSLEEKAEIERDSNWVKQEDENLPDNGFQPPKIKLTSLVGSDVGPLIIHQFSEK
LISGDDKILNGVYSQYEEGESIFGSLF

ORIGIN

1 gatctccat atacaacggt atctccacct cagggttaga tctcaacaac ggaaccattg
61 ccgacatgag acagttaggt atcgtcgaga gttacaagct aaaacagcga gtagtcagct
121 ctgcatctga agccgctgaa gttctactaa ggggtggataa catcatcctg gcaagaccaa
181 gaaccgcca tagacaacat atgtaacata tttaggatat acctcgaata taataaaccc
241 ccacactgtc attattataa ttagaacag aacgcaaaaa ttatccacta tataattcaa
301 agacgcgaaa aaaaaaagac aacgcgctcat agaacttttg gcaattcgcg tcacaataa
361 attttggcaa cttatgttc ctcttcgagc agtactcgag cctctctca agaagtgaat
421 aaatcccatc gtaggtatgg ttaaaatag catctccaca acctcaaac tccttgcgca
481 gagtgcacct cctttgtcga gtaattttca ctttccat gagaactat tttctattc
541 ttactctca catcctgtag tgattgacac tgcaacgcc accatccacta gaagaacaga
.....
4921 ttttcaggtg tagattgctc taattctttg agctgttctc tcagctctc atattttct
4981 tgccatgact cagattctaa ttttaagcta ttcaattct ctttgatc

//

კომენტარები

ველი LOCUS, როგორც ზემოთ აღვნიშნეთ, მოიცავს ლოკუსის სახელს, თანამიმდევრობის სიგრძეს, მოლეკულის ტიპს, GenBank-ის შესაბამის განყოფილების შემოკლებულ დასახელებას და მოდიფიცირების თარიღს.

მოცემული მაგალითის შემთხვევაში

ლოკუსის სახელია SCU49845. პირველი ასოები შეესაბამება ორგანიზმის დასახელებას. შესაძლებელია ლოკუსის სახელის სხვა წესით ფორმირებაც, მაგრამ ამას აქ არ შეეხებით.

თანამიმდევრობის სიგრძეა 5028 bp. სიგრძის ზედა ზღვრით შეზღუდვას არა აქვს ადგილი. შესაძლებელია ცალკე ჩანაწერის სახით სრული გენომური თანამიმდევრობის რეგისტრირებაც. რეგისტრაციისთვის შეზღუდვა მხოლოდ მინიმალურ სიგრძეზეა, რომელიც შეადგენს 50 bp. თუმცა გამონაკლისი დასაშვებია.

გენბანკის განყოფილება, რომელსაც ეს ჩანაწერი ეკუთვნის არის PLN.

გენბანკის მოცემული ვერსია შედგება 18 განყოფილებებისაგან. ესენია:

1. PRI - primate sequences
2. ROD - rodent sequences
3. MAM - other mammalian sequences
4. VRT - other vertebrate sequences
5. INV - invertebrate sequences
6. PLN - plant, fungal, and algal sequences
7. BCT - bacterial sequences
8. VRL - viral sequences
9. PHG - bacteriophage sequences

10. SYN - synthetic sequences
11. UNA - unannotated sequences
12. EST - EST sequences (expressed sequence tags)
13. PAT - patent sequences
14. STS - STS sequences (sequence tagged sites)
15. GSS - GSS sequences (genome survey sequences)
16. HTG - HTG sequences (high-throughput genomic sequences)
17. HTC - unfinished high-throughput cDNA sequencing
18. ENV - environmental sampling sequences

ძირითადად განყოფილებებში თანამიმდევრობათა გაერთიანება ხდება ორგანიზმთა სპეციფიკური ჯგუფებისადმი კუთვნილების მიხედვით, თუმცა ზოგიერთ განყოფილებაში (EST, GSS, HTG, და ა.შ.) - სეკვენირების მეთოდის მიხედვით.

მოდულირების დრო გვიჩვენებს ბოლო შესწორების შეტანის თარიღს. ჩვენ შემთხვევაში იგი შეესაბამება 21-JUN-1999.

ველი DEFINITION იძლევა ინფორმაციას იმის შესახებ, თუ რომელი ორგანიზმიდანაა თანამიმდევრობა, როგორია შესაბამისი გენისა თუ ცილის სახელი, მოკლედ აღწერს თანამიმდევრობის ფუნქციონალურ დატვირთვას (თუ თანამიმდევრობა შეესაბამება არამაკოდირებელ უბანს). თუ თანამიმდევრობა მოიცავს კოდირებად უბანს (CDS), სრულყოფილი აღწერისათვის მითითებულია სპეციფიკატორი, მაგალითად „complete cds“.

ველი ACCESSION მოიცავს საინდეტიფიკაციო ნომერს, რომელიც წარმოადგენს ჩანაწერის (თანამიმდევრობის) ზუსტ იდენტიფიკატორს. ნომერი წარმოადგენს ასოებისა და რიცხვების გარკვეულ კომბინაციას, ვთქვათ 1 ასოსა და 5 რიცხვისაგან (ა.მ. U12345), ან 2 ასოსა და 6 რიცხვისაგან (ა.მ. AF123456) შემდგარ ნომერს. გვხვდება უფრო გრძელი საინდეტიფიკაციო ნომრებიც. წარმოდგენილი მაგალითში საინდეტიფიკაციო ნომერია U49845.

ველი VERSION ასევე ემსახურება ჩანაწერის იდენტიფიცირებას. თუ ჩანაწერში შეტანილია რაიმე უმნიშვნელო ცვლილებაც კი, სავერსიო ნომერში ხდება წერტილის შემდგომი რიცხვის გაზრდა ერთით. ამ დროს კი ძირითადი საიდენტიფიკაციო ნომერი (ანუ ველი ACCESSION) უცვლელი რჩება. ჩვენი მაგალითში ვერსიის ნომერია U49845.1. ACCESSION-ის სავერსიო სისტემის პარალელურად GI ("GenInfo Identifier" sequence identification number) იდენტიფიკატორთა სისტემაშიც ხდება ერთით გაზრდილი, ახალი GI რიცხვის ფორმირება. განხილულ მაგალითში GI არის 1293613.

ველი KEYWORD მოიცავს სიტყვას ან ფრაზას, რომელიც ახასიათებს თანამიმდევრობას. როგორც წესი, გვხვდება ძველ ჩანაწერებში. მოცემულ მაგალითში (ახალ ჩანაწერი) ეს ველი თავისუფალია.

ველი SOURCE არის ორგანიზმის დასახელების აღმწერი ჩანაწერი, რომელიც შეიძლება აღწერდეს მოლეკულის ტიპსაც. ამ ველს ეკუთვნის ქვეველი **Organism**, ის ემყარება NCBI ტაქსონომიური ჯგუფების შესახებ მონაცემთა ბაზის ფილოგენეტიკურ კლასიფიკაციას, მიუთითებს წყარო ორგანიზმის სახელს (ოჯახს, გვარს და სახეობას, სადაც შესაძლებელია) და მის წარმომავლობას. თუ სრული წარმომავლობის აღწერა გრძელდება, მაშინ მოცემულია შემოკლებული წარმომავლობა, ხოლო სრულის ნახვა შესაძლებელია ტაქსონომიური ჯგუფების შესახებ მონაცემთა ბაზაში.

ველი REFERENCE გვაწვდის ინფორმაციას იმ ნაშრომების შესახებ, რომლებიც ეხება მოცემულ თანამიმდევრობას. ინფორმაცია მოიცავს ავტორთა გვარებს, ნაშრომთა სათაურს, ჟურნალის დასახელებას, გვერდებს და ა.შ. ასევე სხვადასხვა ვებ-საიტებზე მათი განთავსების შესახებ ინფორმაციას.

ველი FEATURES ახასიათებს როგორც თანამიმდევრობაში შემავალ ცილის მაკო-დირებელ გენებსა და მათ პროდუქტებს, ასევე სხვა ბიოლოგიურად მნიშვნელოვან უბნებს. მაგალითად უბნებს, რომლებიც აკოდირებენ არაინფორმაციული რნმ-ს მოლეკულებს, ასევე ტრანსკრიპციის რეგულირებისათვის მნიშვნელოვან უბნებს (მაგალითად ტატა ბოქსები და სხვა) და ა.შ. Features - ების სრული ჩამონათვალი შეიძლება ვიხილოთ GenBank -ის ვებ-გვერდზე, აქ მათი სრული განხილვა შორს წაგვიყვანს. ჩვენ შევეცდებით მხოლოდ მოცემულ მაგალითში გამოყენებული features-ების საშუალებით შევქმნათ წარმოდგენა მოცემულ ველში წარმოდგენილი ინფორმაციის შესახებ.

როგორც მაგალითიდან შეგვიძლია დავასკვნათ თითოეულ Features ელემენტის (რომელიც შეესაბამება რაღაც უბანს) ლოკალიზაცია განსაზღვრულია ცალსახად. თუ იგი ეკუთვნის კომპლემენტარულ ჯაჭვს, მაშინ ამ უბნის აღმწერი სტრიქონი იწყება სიტყვა "complement" -ით. თუ ნიშანი < წინ უძღვის ფუძეთა უბნის ლოკალიზაციის განმსაზღვრელ რიცხვთა წყვილს (წყვილს ქმნის პირველი ფუძისა და ბოლო ფუძის ნომრები. ნომერი აითვლება 3' - 5' მიმართულებით), მაშინ ეს ნიშნავს, რომ საქმე გვაქვს თანამიმდევრობის 5' ნაწილთან. თუ კი ნიშანი > მოყვება ფუძეთა უბნის ლოკალიზაციის განმსაზღვრელ რიცხვთა წყვილს, მაშინ ეს ნიშნავს, რომ საქმე გვაქვს თანამიმდევრობის 3' ნაწილთან. ამრიგად features-ში აღწერილი უბანი შეიძლება წარმოადგენდეს ან სრულ უბანს, ან უბნის 3' ან 5' ნაწილს, თანაც ის შეიძლება განლაგებული იყოს ან ძირითად (წაკითხულ) ჯაჭვში ან კომპლემენტარულში.

მაგალითის სახით წარმოგიდგენთ რამდენიმე Features -ის მახასიათებელ სიტყვას (Key -სა) და შესაბამის განმარტებას.

Feature -თა მაგალითები:

მახასიათებელი სიტყვა	განმარტება
CDS	ცილის მაკოდირებელი თანამიმდევრობა
gene	გენი
RBS	რიბოსომასთან მიმართებული უზანი (საიტი)
rep_origin	რეპლიკაციის დასაწყისი
protein_bind	დნმ-ში ცილის მისაერთებელი უზანი
rRNA	ტრანსკორტული რნმ
source	წყარო ორგანიზმი

features-თა სრული ჩამონათვალი შესაძლებელია მოვიძიოთ შემდეგ ვებ-გვერდზე <http://www.ncbi.nlm.nih.gov/projects/collab/FT/index.html#7.3>

ზემოთ მოყვანილ მაგალითში თანამიმდევრობის დასახასიათებლად გამოყენებულია მხოლოდ source, CDS, gene Feature-ული სიტყვები. მოცემულია შესაბამისი უზნების ლოკალიზაცია/სპეციფიკაცია. გენური უზნებისათვის კი, დამატებით მოცემულია ტრანსლირებული ამინმჟავური თანამიმდევრობა, რომელიც იწყება ნიშნით /translation= .

ველი ORIGIN მოიცავს თვით თანამიმდევრობას, რომელიც იწყება უშუალოდ ORIGIN-ის შემდგომი სტრიქონიდან. თანამიმდევრობა წარმოდგენილია FASTA ფორმატით (იხ. პარაგრაფი 14).

თანამიმდევრობის წარმოდგენის ზემოთ აღწერილ ფორმატს ბევრი აქვს საერთო DDBJ-სა და EMBL -ის ჩანაწერთა ფორმატებთან. ზემოთ აღწერილი ფორმატი წარმოადგენს GenBank-ს, EMBL-სა და DDBJ-ს შორის საერთაშორისო თანამშრომლობის შედეგს და აადვილებს ამ სამ ორგანიზაციას შორის მონაცემთა ყოველდღიურ გაცვლას.

ჩანაწერთა ფორმატების შესადარებლად ქვევით მოცემულია ერთი და იგივე თანამიმდევრობის ჩანაწერები, გაკეთებული GenBank-ისათვის, EMBL-ისათვის და DDBJ -თვის

1. EMBL ფორმატი

ID X64011; SV 1; linear; genomic DNA; STD; PRO; 756 BP.
XX
AC X64011; S78972;
XX
SV X64011.1
XX
DT 28-APR-1992 (Rel. 31, Created)
DT 30-JUN-1993 (Rel. 36, Last updated, Version 6)
XX
DE *Listeria ivanovii* sod gene for superoxide dismutase
XX
KW sod gene; superoxide dismutase.
XX
OS *Listeria ivanovii*
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; *Listeria*.
XX
RN [1]
RX MEDLINE; 92140371.
RA Haas A., Goebel W.;
RT "Cloning of a superoxide dismutase gene from *Listeria ivanovii* by
RT functional complementation in *Escherichia coli* and characterization of the
RT gene product.";
RL Mol. Gen. Genet. 231:313-322(1992).
XX
RN [2]
RP 1-756
RA Kreft J.;
RT ;
RL Submitted (21-APR-1992) to the EMBL/GenBank/DBJ databases.
RL J. Kreft, Institut f. Mikrobiologie, Universitaet Wuerzburg, Biozentrum Am
RL Hubland, 8700 Wuerzburg, FRG
XX
DR SWISS-PROT; P28763; SODM_LISIV.
XX
FH Key Location/Qualifiers
FH
FT source 1..756
FT /db_xref="taxon:1638"
FT /organism="Listeria ivanovii"
FT /strain="ATCC 19119"
FT /mol_type="genomic DNA"
FT RBS 95..100

```

FT          /gene="sod"
FT terminator 723..746
FT          /gene="sod"
FT CDS       109..717
FT          /db_xref="SWISS-PROT:P28763"
FT          /transl_table=11
FT          /gene="sod"
FT          /EC_number="1.15.1.1"
FT          /product="superoxide dismutase"
FT          /protein_id="CAA45406.1"
FT          /translation="MTYELPKLPYTYDALEPNFDKETMEIHVTKHHNTYVTKLNEAVSG
FT
HAELASKPGEELVANLDSVPEIRGAVRNHGGGHANHTLFWSSLSPNGGGAPTGNLKAA
FT
IESEFGTFDEFKEKFNAAAAARFGSGWAWLVVNNGKLEIVSTANQDSPLSEGKTPVLGL
FT          DVWEHAYYLKFNRRPEYIDTFWNVINWDERNKRFDAAK"
XX
SQ Sequence 756 BP; 247 A; 136 C; 151 G; 222 T; 0 other;
cggtatttaa ggtgttacat agttctatgg aaatagggtc tatacccttc gccttacaac      60
gtaattctt .....
//

```

2. GenBank ფორმატი

```

LOCUS LISOD 756 bp DNA linear BCT 30-JUN-1993
DEFINITION Listeria ivanovii sod gene for superoxide dismutase.
ACCESSION X64011 S78972
VERSION X64011.1 GI:44010
KEYWORDS sod gene; superoxide dismutase.
SOURCE Listeria ivanovii
ORGANISM Listeria ivanovii
Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
REFERENCE 1 (bases 1 to 756)
AUTHORS Haas,A. and Goebel,W.
TITLE Cloning of a superoxide dismutase gene from Listeria ivanovii by
functional complementation in Escherichia coli and characterization
of the gene product
JOURNAL Mol. Gen. Genet. 231 (2), 313-322 (1992)
MEDLINE 92140371
REFERENCE 2 (bases 1 to 756)
AUTHORS Kreft,J.
TITLE Direct Submission
JOURNAL Submitted (21-APR-1992) J. Kreft, Institut f. Mikrobiologie,
Universitaet Wuerzburg, Biozentrum Am Hubland, 8700 Wuerzburg, FRG

```

FEATURES	Location/Qualifiers
source	1..756 /organism="Listeria ivanovii" /strain="ATCC 19119" /db_xref="taxon:1638" /mol_type="genomic DNA"
RBS	95..100 /gene="sod"
gene	95..746 /gene="sod"
CDS	109..717 /gene="sod" /EC_number="1.15.1.1" /codon_start=1 /transl_table=11 /product="superoxide dismutase" /db_xref="GI:44011" /protein_id="CAA45406.1" /db_xref="SWISS-PROT:P28763" /translation="MTYELPKLPYTYDALEPNFDKETMEIHVYTKHHNTYVTKLNEAVS GHAELASKPGEELVANLDSVPEEIRGAVRNHGGGHANHTLFWSSLSPPNGGGAPTGNLK AAIESEFGTFDEFKEKFNAAAAAARFGSGWAWLVVNGKLEIVSTANQDPSLSEGKTPV LGLDVWEHAYYLKFKQNRPEYIDTFWNVINWDERNKRFDAAK"
terminator	723..746 /gene="sod"

ORIGIN

```

1 cggtatttaa ggtgttacat agttctatgg aaatagggtc tataccttc gcctacaat
61 gtaatttctt .....
//

```

3. DDBJ ցորմատ

LOCUS LISOD 756 bp DNA linear BCT 30-JUN-1993
DEFINITION Listeria ivanovii sod gene for superoxide dismutase.
ACCESSION X64011 S78972
VERSION X64011.1 GI:44010
KEYWORDS sod gene; superoxide dismutase.
SOURCE Listeria ivanovii
ORGANISM Listeria ivanovii
 Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
REFERENCE 1 (bases 1 to 756)
AUTHORS Haas,A. and Goebel,W.

TITLE Cloning of a superoxide dismutase gene from *Listeria ivanovii* by functional complementation in *Escherichia coli* and characterization of the gene product

JOURNAL Mol. Gen. Genet. 231 (2), 313-322 (1992)

MEDLINE 92140371

REFERENCE 2 (bases 1 to 756)

AUTHORS Krefst,J.

TITLE Direct Submission

JOURNAL Submitted (21-APR-1992) J. Krefst, Institut f. Mikrobiologie, Universitaet Wuerzburg, Biozentrum Am Hubland, 8700 Wuerzburg, FRG

FEATURES Location/Qualifiers

source 1..756

/organism="Listeria ivanovii"

/strain="ATCC 19119"

/db_xref="taxon:1638"

/mol_type="genomic DNA"

RBS 95..100

/gene="sod"

gene 95..746

/gene="sod"

CDS 109..717

/gene="sod"

/EC_number="1.15.1.1"

/codon_start=1

/transl_table=11

/product="superoxide dismutase"

/db_xref="GI:44011"

/protein_id="CAA45406.1"

/db_xref="SWISS-PROT:P28763"

/translation="MTYELPKLPYTYDALEPNFDKETMEIHYYTKHHNIYVTKLNEAVS
GHAELASKPGEELVANLDSVPEEIRGAVRNHGGGHANHTLFWSSLSPNGGGAPTGNLKAIESE
FGTFDEFKEKFNAAAAARFGSGWAWLWVNNNGKLEIVSTANQDPSLSEKTPVLGLDVEHAYY
LKFQNRPRPEYIDTFWNVINWDERNKRFDAAK"

terminator 723..746

/gene="sod"

BASE COUNT 247 a 136 c 151 g 222 t

ORIGIN

1 cgttatattaa ggtgttacat agttctatgg aatatgggtc tataaccttc gccttacaat
61 gtaatttctt

//

8. ამინმჟავურ თანამიმდევრობათა შესახებ მონაცემთა ბაზა UniProt.

სულ ცოტა ხნის წინ, EBI/SIB Swiss-Prot + TrEMBL და PIR (Protein Sequence Database (PIR-PSD)) არსებობდნენ როგორც ცილების თანამიმდევრობათა შესახებ მონაცემთა დამოუკიდებელი ბაზები, თანამიმდევრობების წარმოდგენის განსხვავებული ფორმატითა და ანოტირების განსხვავებული ფორმებით. 2002 წლიდან EBI (The European Bioinformatics Institute), SIB (Swiss Institute of Bioinformatics), და PIR (at the Georgetown University Medical Center and National Biomedical Research Foundation) გაერთიანდნენ და შექმნეს UniProt კონსორციუმი. კონსორციუმის უმთავრესი ამოცანაა ბიოლოგიური კვლევების მაღალხარისხიანი მონაცემთა ბაზით უზრუნველყოფა, რომელშიც თავმოყრილი იქნება ზუსტად ანოტირებულ ცილათა თანამიმდევრობების შესახებ არსებული მონაცემები. ამასთან ბაზა იქნება სტაბილური, სრულყოფილად კლასიფიცირებული, უზრუნველყოფილი სხვა მონაცემთა ბაზებთან ინფორმაციის გაცვლის საშუალებებით, საურთიერთოდ მოსახერხებელი და ადვილი.

UniProt Knowledgebase (UniProtKB) წარმოადგენს სწორედ ცილათა თანამიმდევრობების შესახებ ძირითად მონაცემთა ბაზას, რომელიც უზრუნველგვყოფს თანამიმდევრობებისა და ფუნქციების ზუსტი, განახლებადი, მრავალმხრივი, ანოტაციით.

UniProtKB შედგება ორი ნაწილისაგან: Swiss-Prot -საგან, რომელიც წარმოადგენს ხელით ანოტირებულ ჩანაწერთა განყოფილებას. ინფორმაცია აღებულია ლიტერატურიდან და ამავე დროს გამოიყენება UniProtKB-ს კურატორების მიერ განხორციელებული კომპიუტერული ანალიზი; და TrEMBL -სგან, რომელიც წარმოადგენს მხოლოდ კომპიუტერული ანალიზის შედეგად წარმოებულ ჩანაწერთა განყოფილებას. ცალ-ცალკე, მოკლედ მიმოვიხილოთ ეს განყოფილებები.

Swiss-Prot Protein Knowledgebase

Swiss-Prot წარმოადგენს ანოტირებულ ამინმჟავურ თანამიმდევრობათა შესახებ მონაცემთა ბაზას. ის შეიქმნა the Department of Medical Biochemistry of the University of Geneva ახლა the Swiss Institute of Bioinformatics (SIB) -ის Amos Bairoch ჯგუფისა და the EMBL Data Library (now the EMBL Outstation - The European Bioinformatics Institute (EBI)) -ის თანამშრომლობის საფუძველზე 1986 წელს. Swiss-Prot Protein Knowledgebase შედგება ცილათა და მათი ამინმჟავური თანამიმდევრობების შესახებ ჩანაწერებისაგან. თითოეული ჩანაწერი წარმოად-

გენს სხვადასხვა ტიპის სტრიქონთა ერთობას. თითოეული სტრიქონი მისთვის დამახასიათებელი ფორმატის მიხედვითაა ჩაწერილი. სტანდარტიზირების მოთხოვნიდან გამომდინარე ხორციელდება Swiss-Prot ფორმატის შეძლებისდაგვარად EMBL ფორმატთან მიახლოება.

თანამიმდევრობათა სხვა მონაცემთა ბაზებთან მიმართებით Swiss-Prot-ის უპირატესობებად შეიძლება აღვნიშნოთ:

1. ანოტაცია

Swiss-Prot-ში, ისევე როგორც თანამიმდევრობათა სხვა მრავალ მონაცემთა ბაზაში, მონაცემები შეიძლება დავყოთ ძირითად მონაცემებად და ანოტაციად. ჩანაწერში ძირითად მონაცემებად მოიაზრებიან:

- უშუალოდ თანამიმდევრობის შესახებ მონაცემები
- ციტირების შესახებ ინფორმაცია (ბიბლიოგრაფიული მონაცემები)
- ტაქსონომიური მონაცემები (ცილის ბიოლოგიური წყაროს აღწერა)

ანოტაცია გულისხმობს მთელ რიგ საკითხებთან დაკავშირებულ დამატებით ინფორმაციას. ეს საკითხები შეიძლება იყოს:

- ცილის ფუნქცია;
- პოსტრანსლაციური მოდიფიკაციები. მაგალითად ფოსფორილირება, აცეტილირება და ა.შ. ;
- დომენები და უბნები. მაგალითად: კალციუმის დამაკავშირებელი უბანი, ATP-ს დამაკავშირებელი უბანი, ცინკის თითაკი და ა.შ. ;
- მეორეული სტრუქტურა. მაგალითად: ალფა სპირალი, ბეტა სტრუქტურა და ა.შ. ;
- მეოთხეული სტრუქტურა. მაგალითად: ჰომოდიმერი, ჰეტეროდიმერი და ა.შ. ;
- მსგავსება სხვა ცილებთან;
- დაავადებები, რომლებიც დაკავშირებულია ცილის შესაძლო დეფექტებთან;
- თანამიმდევრობასთან დაკავშირებული გაურკვეველობები, ვარიანტები და ა.შ.

ანუ ანოტაციაში თავმოყრილია შეძლებისამებრ მაქსიმალური დამატებითი ინფორმაცია. ამ ინფორმაციის მოძიებისას, ახალი თანამიმდევრობების შესახებ პუბლიკაციების გარდა მუშავდება მიმოხილვითი ნაშრომებიც. ასევე ხდება გარე ექსპერტების მოსაზრებების გათვალისწინება. მოცემული მონაცემთა ბაზის უპირატესობა სწორედ იმაში მდგომარეობს, რომ გამოყენებულია საკმაოდ მძლავრი წყაროები, და არა მარტო ძირითადი ინფორმაციის შესავსებად.

Swiss-Prot-ის ჩანაწერში ანოტაციები გვხვდება კომენტარების სტრიქონებში (CC), თვისებათა ცხრილებში (FT) და დამახასიათებელ სიტყვათა სტრიქონებში (KW).

2. ინფორმაციის მინიმალური დუბლირება

თანამიმდევრობათა უმრავლეს მონაცემთა ბაზებში კონკრეტული ცილის თანამიმდევრობის შესახებ რამდენიმე განსხვავებული პუბლიკაციის არსებობის შემთხვევაში, რამდენიმე (პუბლიკაციების შესაბამისად) დამოუკიდებელი ჩანაწერი არსებობს. Swiss-Prot-ში კი, ყველა დუბლირებულად არსებული მონაცემი შეჯერებულია, ანუ წარმოდგენილი ინფორმაცია ოპტიმიზირებულია.

3. სხვა მონაცემთა ბაზებთან ინტეგრაცია.

ძალზედ მნიშვნელოვანია ის, რომ ბიომაკრომოლეკულათა შესახებ მონაცემთა ბაზის მომხმარებლები უზრუნველყოფილნი იყვნენ როგორც მონაცემთა ძირითად ბაზებთან კავშირით (ნუკლეინის მჟავათა თანამიმდევრობების, ცილათა თანამიმდევრობების და ცილათა მესამეული სტრუქტურების), ისე ცალკე ხარისხის ინტეგრაციით სპეციალიზირებულ მონაცემთა ბაზებთან. ამ ეტაპზე Swiss-Prot აწარმოებს ინფორმაციის გაცვლას 50 სხვადასხვა ბაზასთან და ამ ქსელის გაფართოების პროცესი გრძელდება.

4. დოკუმენტაცია

Swiss-Prot უზრუნველყოფილია დიდი რაოდენობა ინდექს-ფაილებით და სპეციალიზირებული დოკუმენტაციით, რითაც სრულად უზრუნველყოფილია მომხმარებლის ინტერესები. სადოკუმენტაციო ფაილები შესაძლებელია მოიძიოთ შემდეგ ვებ-გვერდზე: http://www.expasy.org/sprot/userman.html#list_of_documents

კომპიუტერულად ანოტირებული დამატება -TrEMBL

TrEMBL წარმოადგენს UniProt Knowledgebase-ის კომპიუტერულად ანოტირებულ განყოფილებას. ის მოიცავს ნუკლეოტიდურ მონაცემთა DDBJ/EMBL/GenBank ბაზებიდან მოძიებული ყველა მაკოდირებელი უბნის ტრანსლაციის შედეგსა და ლიტერატურიდან მოპოვებულ ან UniProtKB-ში წარდგენილ ყველა იმ ცილის თანამიმდევრობას, რომლებიც ჯერ კიდევ არაა ინტეგრირებული Swiss-Prot-ში.

TrEMBL ჩანაწერში წარმოდგენილი ინფორმაცია უშუალოდ იწარმოება DDBJ/EMBL/GenBank -ის ჩანაწერიდან და მონაცემთა ხარისხი პირდაპირაა დამოკიდებული ნუკლეოტიდური ჩანაწერის წარდგენის ხარისხზე. ეს ინფორმაცია შესაძლებელია შემდგომში გაზრდილ იქნეს ანოტირების ავტომატური პრო-

ცესის საშუალებით. მაგრამ თუ ეს ვერ მოხერხდა, ის რჩება წარმომდგენის მიერ შემოთავაზებული ჩანაწერის სახით მანამ, სანამ არ მოხდება ხელით ანოტირება და Swiss-Prot-ში წარდგენა.

TrEMBL ჩანაწერის ფორმირების შემდეგ, მომხმარებლისთვის შესათავაზებელ მონაცემთა ხარისხის გაუმჯობესების მიზნით ხორციელდება შემდეგი ოპერაციები:

1. ავტომატური ანოტაცია.

ხელით ანოტაციის მოლოდინში მყოფი TrEMBL-ს ჩანაწერის შევსება ხდება ავტომატური ანოტაციით. ცილათა ოჯახების, დომენებისა და ფუნქციური უბნების შესახებ მონაცემთა ბაზის სახელით ცნობილ UniProt-ის და ასევე Swiss-Prot-ის ანოტირებული ჩანაწერებიდან ხდება ინფორმაციის გადაქაჩვა TrEMBL-ის არაანოტირებულ ჩანაწერში. ამ პროცესის შედეგად TrEMBL-ს ანოტაციის სტანდარტი უახლოვდება Swiss-Prot-ისას და ამრიგად მომხმარებლისთვის მონაცემთა ხარისხი მაღლდება.

2. დუბლირებისაგან თავის დაცვა

ხდება იმ თანამიმდევრობების გაერთიანება ერთ ჩანაწერში, რომელთა სრული სიგრძეები დადგენილია, ეკუთვნის ერთსა და იმავე ორგანიზმს და ავლენენ 100% იდენტურობას.

მონაცემთა ბაზის ზოგადი სტრუქტურა

UniProt Knowledgebase შედგება ამინმჟავურ თანამიმდევრობათა შესახებ ჩანაწერებისაგან. თითოეული ჩანაწერი შეესაბამება კონკრეტულ უწყვეტ თანამიმდევრობას, რომელიც წარდგენილია უშუალოდ ბანკში ან გამოქვეყნებულია სამეცნიერო ლიტერატურაში. რიგ შემთხვევებში ჩანაწერი სხვადასხვა ნაშრომებში წარმოდგენილ გადაფარვად თანამიმდევრობათა უბნების შესახებ ინფორმაციის გაერთიანების შედეგია. არსებობს საწინააღმდეგო შემთხვევებიც, როცა ერთი ნაშრომი საფუძველად უდევს რამდენიმე ჩანაწერს. ამ უკანასკნელის მაგალითად გამოდგება ნაშრომი, სადაც გამოქვეყნებულია სხვადასხვა ორგანიზმის მონათესავე თანამიმდევრობები.

პოზიციების დანომვრვა თანამიმდევრობაში იწყება 1-დან, რომელიც შეესაბამება N- ბოლო პოზიციას.

თანამიმდევრობა შეესაბამება ცილის პრეკურსორულ ფორმას, ანუ მდგომარეობას პოსტრანსლაციურ მოდიფიკაციებამდე და პროცესინგამდე.

სტატუსი

ერთმანეთისაგან სხვადასხვა განყოფილების ჩანაწერების, ანუ Swiss-Prot-ის სრულად ანოტირებული და TrEMBL-ის ავტომატურად ანოტირებული ჩანაწერების განვასხვავება შესაძლებელია ჩანაწერის (ID) სტრიქონში მითითებული ჩანაწერის "სტატუსით" ("status"). ვხვდებით ორი სხვადასხვა სტატუსის ჩანაწერს:

Reviewed	ჩანაწერები, რომელთა რეცენზირებასა და ანოტირებას უშუალოდ UniProtKB-ს კურატორები ახდენენ (UniProt Knowledgebase Swiss-Prot განყოფილება).
Unreviewed	კომპიუტერულად ანოტირებული ჩანაწერები, რომელთა რეცენზირება არ ხდება UniProtKB-ს კურატორების მიერ (UniProt Knowledgebase-ის TrEMBL განყოფილება).

ჩანაწერის სტრუქტურა

ჩანაწერის სტრუქტურა ისეთი უნდა იყოს, რომ მისი გამოყენება გაუადვილდეს როგორც მომხმარებელს, ისე კომპიუტერულ პროგრამას. განმარტებები, აღწერები, კლასიფიკაციები და სხვა მოცემულია მარტივი ინგლისურით. სადაც ეს შესაძლებელია, გამოიყენება ბიოქიმიკოსებისათვის და მოლეკულურ ბიოლოგებისათვის მისაღები ტერმინები და სიმბოლოები. ჩანაწერი შედგება სტრიქონებისაგან. სხვადასხვა მონაცემთა წარმოდგენა ხდება სხვადასხვა ფორმატის მქონე სტრიქონებითა თუ სტრიქონთა ჯგუფებით. მაგალითად, ქვევით წარმოდგენილია ერთ-ერთი თანამიმდევრობის შესაბამისი ჩანაწერი.

ID GRAA_HUMAN Reviewed; 262 AA.
 AC P12544; Q6IB36;
 DT 01-OCT-1989, integrated into UniProtKB/Swiss-Prot.
 DT 01-OCT-1989, sequence version 1.
 DT 12-JUN-2007, entry version 95.
 DE Granzyme A precursor (EC 3.4.21.78) (Cytotoxic T-lymphocyte proteinase
 DE 1) (Hanukkah factor) (H factor) (HF) (Granzyme-1) (CTL tryptase)
 DE (Fragmentin-1).

GN Name=GZMA; Synonyms=CTLA3, HFSP;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini;
OC Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE [MRNA].
RC TISSUE=T-cell;
RX MEDLINE=88125000; PubMed=3257574;
RA Gershenfeld H.K., Hersberger R.J., Shows T.B., Weissman I.L.;
RT "Cloning and chromosomal assignment of a human cDNA encoding a T cell-
RT and natural killer cell-specific trypsin-like serine protease.";
RL Proc. Natl. Acad. Sci. U.S.A. 85:1184-1188(1988).
RN [2]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA].
RA Ebert L., Schick M., Neubert P., Schatten R., Henze S., Korn B.;
RT "Cloning of human full open reading frames in Gateway(TM) system entry
RT vector (pDONR201).";
RL Submitted (JUN-2004) to the EMBL/GenBank/DDBJ databases.
RN [3]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA].
RC TISSUE=Blood;
RX PubMed=15489334; DOI=10.1101/gr.2596504;
RG The MGC Project Team;
RT "The status, quality, and expansion of the NIH full-length cDNA
RT project: the Mammalian Gene Collection (MGC).";
RL Genome Res. 14:2121-2127(2004).
RN [4]
RP NUCLEOTIDE SEQUENCE [GENOMIC DNA] OF 1-23.
RA Goralski T.J., Krensky A.M.;
RT "The upstream region of the human granzyme A locus contains both
RT positive and negative transcriptional regulatory elements.";
RL Submitted (NOV-1995) to the EMBL/GenBank/DDBJ databases.
RN [5]
RP PROTEIN SEQUENCE OF 29-53.
RX MEDLINE=88330824; PubMed=3047119;
RA Poe M., Bennett C.D., Biddison W.E., Blake J.T., Norton G.P.,
RA Rodkey J.A., Sigal N.H., Turner R.V., Wu J.K., Zweierink H.J.;
RT "Human cytotoxic lymphocyte tryptase. Its purification from granules
RT and the characterization of inhibitor and substrate specificity.";
RL J. Biol. Chem. 263:13215-13222(1988).
RN [6]
RP PROTEIN SEQUENCE OF 29-40, AND CHARACTERIZATION.
RX MEDLINE=89009866; PubMed=3262682;
RA Hameed A., Lowrey D.M., Lichtenheld M., Podack E.R.;
RT "Characterization of three serine esterases isolated from human IL-2

RT activated killer cells.*;
RL J. Immunol. 141:3142-3147(1988).
RN [7]
RP PROTEIN SEQUENCE OF 29-39, AND CHARACTERIZATION.
RX MEDLINE=89035468; PubMed=3263427;
RA Kraehenbuhl O., Rey C., Jenne D.E., Lanzavecchia A., Groscurth P.,
RA Carrel S., Tschopp J.;
RT "Characterization of granzymes A and B isolated from granules of
RT cloned human cytotoxic T lymphocytes.*";
RL J. Immunol. 141:3471-3477(1988).
RN [8]
RP 3D-STRUCTURE MODELING OF 29-262.
RX MEDLINE=89184501; PubMed=3237717; DOI=10.1002/prot.340040306;
RA Murphy M.E.P., Moulton J., Bleackley R.C., Gershenfeld H.,
RA Weissman I.L., James M.N.G.;
RT "Comparative molecular model building of two serine proteinases from
RT cytotoxic T lymphocytes.*";
RL Proteins 4:190-204(1988).
RN [9]
RP X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS) OF 29-262 IN COMPLEX WITH A
RP TRIPEPTIDE CMK INHIBITOR.
RX MEDLINE=22708839; PubMed=12819769; DOI=10.1038/nsb944;
RA Bell J.K., Goetz D.H., Mahrus S., Harris J.L., Fletterick R.J.,
RA Craik C.S.;
RT "The oligomeric structure of human granzyme A is a determinant of its
RT extended substrate specificity.*";
RL Nat. Struct. Biol. 10:527-534(2003).
RN [10]
RP X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS) OF 29-262 IN COMPLEX WITH
RP SUBSTRATE.
RX MEDLINE=22708840; PubMed=12819770; DOI=10.1038/nsb945;
RA Hink-Schauer C., Estebanez-Perpina E., Kurschus F.C., Bode W.,
RA Jenne D.E.;
RT "Crystal structure of the apoptosis-inducing human granzyme A dimer.*";
RL Nat. Struct. Biol. 10:535-540(2003).
CC -|- FUNCTION: This enzyme is necessary for target cell lysis in cell-
CC mediated immune responses. It cleaves after Lys or Arg. May be
CC involved in apoptosis.
CC -|- CATALYTIC ACTIVITY: Hydrolysis of proteins, including fibronectin,
CC type IV collagen and nucleolin. Preferential cleavage: -Arg-|-
CC Xaa-, -Lys-|-Xaa- >> -Phe-|-Xaa- in small molecule substrates.
CC -|- SUBUNIT: Homodimer; disulfide-linked.
CC -|- INTERACTION:
CC Self; NbExp=1; IntAct=EBI-519800, EBI-519800;
CC -|- SUBCELLULAR LOCATION: Secreted. Cytoplasmic granule.
CC -|- SIMILARITY: Belongs to the peptidase S1 family. Granzyme
CC subfamily.

CC -!- SIMILARITY: Contains 1 peptidase S1 domain.
 CC -----
 CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>
 CC Distributed under the Creative Commons Attribution-NoDerivs License
 CC -----
 DR EMBL; M18737; AAA52647.1; -; mRNA.
 DR EMBL; CR456968; CAG33249.1; -; mRNA.
 DR EMBL; BC015739; AAH15739.1; -; mRNA.
 DR EMBL; U40006; AAD00009.1; -; Genomic_DNA.
 DR PIR; A31372; A31372.
 DR UniGene; Hs.90708; -.
 DR PDB; 1HF1; Model; @=29-262.
 DR PDB; 1OP8; X-ray; A/B/C/D/E/F=29-262.
 DR PDB; 1ORF; X-ray; A=29-262.
 DR IntAct; P12544; -.
 DR MEROPS; S01.135; -.
 DR Ensembl; ENSG00000145649; Homo sapiens.
 DR KEGG; hsa:3001; -.
 DR H-InvDB; HIX0004862; -.
 DR HGNC; HGNC:4708; GZMA.
 DR MIM; 140050; gene.
 DR PharmGKB; PA29086; -.
 DR LinkHub; P12544; -.
 DR ArrayExpress; P12544; -.
 DR GermOnline; ENSG00000145649; Homo sapiens.
 DR GO; GO:0001772; C:immunological synapse; TAS:UniProtKB.
 DR GO; GO:0005634; C:nucleus; TAS:UniProtKB.
 DR GO; GO:0004277; F:granzyme A activity; IDA:UniProtKB.
 DR GO; GO:0042803; F:protein homodimerization activity; IDA:UniProtKB.
 DR GO; GO:0006922; P:cleavage of lamin; IDA:UniProtKB.
 DR GO; GO:0006955; P:immune response; TAS:UniProtKB.
 DR InterPro; IPR001254; Peptidase_S1_S6.
 DR InterPro; IPR001314; Peptidase_S1A.
 DR Pfam; PF00089; Trypsin; 1.
 DR PRINTS; PR00722; CHYMOTRYPSIN.
 DR SMART; SM00020; Tryp_SpC; 1.
 DR PROSITE; PS50240; TRYPSIN_DOM; 1.
 DR PROSITE; PS00134; TRYPSIN_HIS; 1.
 DR PROSITE; PS00135; TRYPSIN_SER; 1.
 PE 1: Evidence at protein level;
 KW 3D-structure; Apoptosis; Cytolysis; Direct protein sequencing;
 KW Glycoprotein; Hydrolase; Polymorphism; Protease; Serine protease;
 KW Signal; Zymogen.
 FT SIGNAL 1 26
 FT PROPEP 27 28 Activation peptide.
 FT /FTId=PRO_0000027393.
 FT CHAIN 29 262 Granzyme A.

FT /FTId=PRO_0000027394.
 FT DOMAIN 29 259 Peptidase S1.
 FT ACT_SITE 69 69 Charge relay system.
 FT ACT_SITE 114 114 Charge relay system.
 FT ACT_SITE 212 212 Charge relay system.
 FT CARBOHYD 170 170 N-linked (GlcNAc...) (Potential).
 FT DISULFID 54 70
 FT DISULFID 148 218
 FT DISULFID 179 197
 FT DISULFID 208 234
 FT VARIANT 121 121 T -> M (in dbSNP:rs3104233).

FT /FTId=VAR_024291.

FT STRAND 43 47
 FT STRAND 49 51
 FT STRAND 53 60
 FT STRAND 63 66
 FT STRAND 77 81
 FT STRAND 83 87
 FT STRAND 93 95
 FT STRAND 97 102
 FT TURN 108 110
 FT STRAND 116 122
 FT STRAND 127 130
 FT STRAND 147 154
 FT STRAND 156 160
 FT STRAND 167 174
 FT HELIX 176 179
 FT TURN 186 189
 FT STRAND 195 199
 FT STRAND 215 218
 FT STRAND 221 228
 FT STRAND 241 245
 FT HELIX 251 258

..
 ** ##### INTERNAL SECTION #####

SQ SEQUENCE 262 AA; 28969 MW; DA87363A0D92BAF4 CRC64;
 MRNSYRFLAS SISVVVSLLL IPEDVCEKII GGNEVTPHSR PYMVLLSLDR KTICAGALIA
 KDWVLTAABC NLNKRISQVIL GAHSITREEP TKQIMLVKKE FPYPYCDPAT REGDLKLLQL
 TEKAKINKYV TILHLPKKGD DVKPGTMCQV AGWGRTHNSA SWSDTLREVN ITIIDRKVCN
 DRNHYNFNPV IGMNMVCAGS LRGGRDSCNG DSGSPLLCEG VFRGVTSTFGL ENKCGDPRGP
 GVYILLSKKH LNWIIMTKG AV

//

ცხრილი 8.1 UniProt Knowledgebase ჩანაწერის სტრიქონთა კოდები

სტრიქ.კოდი	მონაცემთა სახე	ჩანაწერში სტრიქ. შეხვედრის რაოდენობა
ID	იდენტიფიცირება	ერთხელ; ჩანაწერის საწყისი სტრიქონი
AC	საინდეტიფიკაციო ნომერი	ერთხელ ან მეტჯერ.
DT	თარიღი	სამჯერ
DE	აღწერა	ერთხელ ან მეტჯერ
GN	გენის(თა) დასახელება	არაა აუცილებელი
OS	ორგანიზმი, სახეობა	ერთხელ
OG	ორგანო	არაა აუცილებელი
OC	ორგანიზმის კლასიფიკაცია	ერთხელ ან მეტჯერ
OX	სხვა ბაზების ტაქსონომიური მონაცემები	ერთხელ
RN	მითითებული ლიტ. ნომერი	ერთხელ ან მეტჯერ
RP	ლიტ.-ში მითითებული ობიექტის ტიპი	ერთხელ ან მეტჯერ
RC	კომენტარი მითითებულ ლიტ.-ზე	არაა აუცილებელი
RX	მითითებულ ლიტ.-ზე მითითება	არაა აუცილებელი
RG	ჯგუფზე მითითება	ერთხელ ან მეტჯერ (არაა აუცილებელი თუ RA არის)
RA	მითითებული ლიტ. ავტორები	ერთხელ ან მეტჯერ (არაა აუცილებელი თუ RG არის)
RT	მითითებული ლიტ. სათაური	არაა აუცილებელი
RL	მითითებული ლიტ. განთ. ადგ.	ერთხელ ან მეტჯერ
CC	კომენტარები ან შენიშვნები	არაა აუცილებელი
DR	მითითება სხვა ბაზაზე	არაა აუცილებელი
KW	დამახასიათებელი სიტყვები	არაა აუცილებელი
FT	თვისებათა ცხრლ. მონაცემები	არაა აუცილებელი
SQ	თანამიმდევრობის თავსტრიქონი	ერთხელ
(blanks)	თვითონ თანამიმდევრობა	ერთხელ ან მეტჯერ
//	ბოლო სტრიქონი	ერთხელ; ჩანაწერის ბოლო სტრიქონი.

ანალოგიურია TrEMBL განყოფილების ჩანაწერის ფორმატი (თუ არ მივიღებთ მხედველობაში უმნიშვნელო განსხვავებებს). როგორც მაგალითიდან ჩანს თითოეული სტრიქონი იწყება 2-ასოიანი კოდით, რომელიც აღწერს სტრიქონში განთავსებულ მონაცემთა ხასიათს. სტრიქონთა ტიპები, შესაბამისი კოდები, და

ჩანაწერში სტრიქონის შეხვედრის რაოდენობა მოცემულია 8.1 ცხრილში. როგორც ცხრილიდან ჩანს, რიგი სტრიქონებისა გვხვდება ყველა ჩანაწერში, სხვები არაა აუცილებელი. ზოგი სტრიქონი შეიძლება ბევრჯერ შეგვხვდეს ერთსა და იმავე ჩანაწერში. თითოეული ჩანაწერი იწყება საინდეტიფიკაციო სტრიქონით (ID) და მთავრდება // სტრიქონით.

უნდა აღინიშნოს, რომ GN-ის გამოკლებით, ყველა აქ მოხსენიებულ სტრიქონთა ტიპი გვხვდება EMBL ნუკლეოტიდურ მონაცემთა ბაზაში, თუმცა არსებობს მცირე განსხვავებებიც.

სტრიქონთა ტიპის განმსაზღვრელ 2-ასოიან კოდს ყოველთვის მოყვება 3 ცარიელი პოზიცია (გამოტოვება), ასე რომ, შესაბამისი ინფორმაციის განთავსება იწყება მე-6 პოზიციიდან. როგორც წესი, განთავსებული ინფორმაცია არ გადის 75 პოზიციის გარეთ, თუმცა არის გამონაკლისებიც, სადაც სტრიქონი უფრო გრძელია (მაგალითად OG და CC სტრიქონები)

არაექსპერიმენტულობის სპეციფიკატორები

საკომენტარო სტრიქონებსა და თვისებათა ცხრილებში ვხვდებით სამ სპეციფიკატორს, რომლებიც გვიჩვენებენ, რომ ინფორმაცია არაა ექსპერიმენტულად დადასტურებული. ესენია:

- Potential
- Probable
- By similarity

ტერმინი "Potential" გვიჩვენებს, რომ არსებობს გარკვეული ლოგიკური საფუძველი, რომ გამოყენებულ იქნეს მოცემული ანოტაცია. როგორც წესი, ამის საფუძველს იძლევა ცილის თანამიმდევრობის კომპიუტერული ანალიზის სხვადასხვა მეთოდები.

ტერმინი "Probable" გულისხმობს, რომ არსებობს რაღაც ექსპერიმენტული საფუძველი იმისა, რომ ასეთი ინფორმაცია შესაძლებელია რეალობას შეესაბამებოდეს.

"By similarity" იხმარება იმ ფაქტის აღწერისას, რომელიც დამტკიცებულია ტაქსონომურ ჯგუფში გაერთიანებული ერთი ცილის ან ცილათა ნაწილისთვის და შემდეგ გამოყენებულია იმავე ჯგუფის სხვა ცილისათვის .

9. ბიომაკრომოლეკულათა სივრცული სტრუქტურების შესახებ მონაცემთა ბაზა - PDB

ცოცხალ უჯრედში შენახული ინფორმაციის რეალიზების გზების თავისებურებებიდან გამომდინარე ბიომაკრომოლეკულათა მხოლოდ ხაზოვანი, ქიმიური მოწყობის (ბიოლოგიური ტექსტები) შესახებ ცოდნა ვერ უზრუნველგვყოფს ყველა იმ ინფორმაციით, რომლითაც შესაძლებელია ცხოველმყოფელობის მექანიზმებში გარკვევა. ეს მტკიცება განსაკუთრებით სამართლიანია ცილოვანი მოლეკულების შემთხვევაში. ცილები მოიაზრება როგორც კონკრეტული ბიოლოგიური ფუნქციის მატარებელი მანქანები. მანქანის მიერ კი, შესაბამისი ფუნქციის შესრულების უნარი განპირობებულია მისი სივრცული აღნაგობით, სამგანზომილებიანი სტრუქტურით. აქედან გამომდინარე დღის წესრიგში დადგა ცილათა სივრცული მოწყობის შესახებ, პირველად და მესამეულ სტრუქტურებს შორის ინფორმაციული კავშირების შესახებ ცოდნის დაგროვების მოთხოვნილება. ამ მოთხოვნების საპასუხოდ შემუშავდა და დაიხვეწა მრავალი მიდგომა, როგორც ექსპერიმენტული, ისე კომპიუტერული. ექსპერიმენტულთა შორის გამოვყოფთ ცილათა კრისტალების რენტგენოგრაფულ ანალიზსა და ხსნარში ბირთვული მაგნიტური რეზონანსის მეთოდს. ამ მეთოდთა ძვირადღირებულობისა და დროში გაწელილობის გამო დიდი განვითარება ჰპოვა ცილის სივრცული მოწყობის დადგენის კომპიუტერულმა მიდგომებმა. სწორედ ექსპერიმენტულ და კომპიუტერულ მეთოდთა დამსახურებაა, რომ დღეისათვის მრავალი ცილოვანი მოლეკულის, ნუკლეინის მჟავას, მათი ერთმანეთთან, თუ სხვადასხვა დაბალმოლეკულურ ნაერთებთან ურთიერთქმედების შედეგად ჩამოყალიბებული კომპლექსის სამგანზომილებიანი მოწყობის შესახებ სრული ინფორმაცია მოგვეპოვება. ბიომაკრომოლეკულათა და მათი სხვადასხვა კომპლექსთა სივრცული მოწყობის შესახებ ინფორმაციის მოპოვება გრძელდება და ამ საქმეში ჩართულია ასობით ქვეყნის მკვლევარი. მათ მიერ მოპოვებული ზღვა ინფორმაციის დეპონირება და შენახვა ხდება სივრცული მოწყობის შესახებ მონაცემთა ბაზაში, PDB-ში (Protein Data Bank). PDB შეიქმნა და განვითარდა ბრუკჰავენის ლაბორატორიის (Brookhaven National Laboratory) ბაზაზე. დღეისათვის ამ ბაზის მომსახურებას უზრუნველყოფს გაერთიანება - Research Collaboratory for Structural Bioinformatics (RCSB) (გაერთიანებაში შედის სამი ოჯგუფი: 1) RCSB-Rutgers (Department of Chemistry and Chemical Biology and the Biomaps Institute for Quantitative Biology at Rutgers, The State University of New Jersey); 2) RCSB-SDSC (San Diego Supercomputer Center (SDSC) and the Skaggs School of Pharmacy and Pharmaceutical Sciences (SSPPS) at the University of California, San Diego) და 3) RCSB-BMRB, (Department of Biochemistry at the University of Wisconsin-Madison.). PDB-ის 2007 წლის ვერსია აერთიანებს 45744 ერთეულს (ჩანაწერს, ფაილს). თითოეული აღწერს კონკრეტული ცილის, ნუკლეინის მჟავას, მათი ერთმანეთთან ან სხვა

დაბალმოლეკულურ ნაერთთან კომპლექსის სივრცულ მოწყობას. შეიძლება თამამად ითქვას, რომ PDB-ში თავმოყრილი ინფორმაცია ფასდაუდებელ სამსახურს უწევს, ზოგადად თანამედროვე ბიოლოგიისა და კონკრეტულად სტრუქტურული ბიოლოგიის განვითარების საქმეს, უმნიშვნელოვანეს როლს თამაშობს თანამედროვე ბიოლოგიის მთავარი ამოცანის, ცოცხალი უჯრედის ცხოველყოფილობის მექანიზმებში გარკვევის ამოცანის გადაწყვეტის საქმეში. აქედან გამომდინარე, მიმოვიხილოთ PDB, კერძოდ კი მისი ერთეულის მოწყობის თავისებურებანი, რათა შეგვეძლოს ამ მეტად ღირებული ინფორმაციით სარგებლობა.

PDB ფორმატი

თითოეული სტრუქტურის შესახებ ინფორმაცია ინახება ცალკეული PDB ფაილის სახით. PDB ფაილი ორგანიზებულია განყოფილებებად. PDB მოიცავს შემდეგ განყოფილებებს: სათაო, პირველადი სტრუქტურის, მეორეული სტრუქტურის, ჰეტეროგენულ ნაშთთა, კავშირების ანოტირების, კრისტალოგრაფიული და კოორდინატთა ტრანსფორმირების, კოორდინატების, კავშირების და ფაილშიდა აღრიცხვიანობის. ქვემოთ მიმოხილული იქნება თითოეული განყოფილების ფორმატი.

სათაო განყოფილება

განყოფილება აერთიანებს იმ ჩანაწერებს, რომლებიც აღწერს მოცემული ფაილის შესაბამის ბიოლოგიურ მაკრომოლეკულებსა და ექსპერიმენტებს, რომელთა გამოყენებითაც ამ მოლეკულათა სრულტურის დადგენა მოხდა. სათაო განყოფილებაში გაერთიანებულ ჩანაწერთა ჩამონათვალი შემდეგია: HEADER, OBSLTE, TITLE, CAVEAT, COMPND, SOURCE, KEYWDS, EXPDTA, AUTHOR, REVDAT, SPRSDE, JRNL, და REMARK. თითოეული ჩანაწერი შედგება მონაცემთა რამდენიმე, განსხვავებული ველისაგან.

ჩანაწერი HEADER

ჩანაწერის მიმოხილვა

ჩანაწერი HEADER idCode ველით ახდენს PDB ფაილის ცალსახა იდენტიფიცირებას. ამ ჩანაწერით ასევე ხდება ფაილის შესაბამისი მოლეკულის კლასიფიკაცია. აქვეა მოცემული PDB-ში კოორდინატების შენახვის (დეპოზირების) თარიღი.

ჩანაწერის ზოგადი ფორმატი ასეთია

პოზიციები	მონაცემის ტიპი	ველი	განსაზღვრება
1 - 6	ჩანაწერის სახელი	'HEADER'	
11 - 50	ასოთა რიგი(40)	კლასიფიცირების	ახდენს მოლეკულის კლასიფიცირებას
51 - 59	რიცხვი	დებ.თარიღი	PDB -ში დეპონირების თარიღი,
63 - 66	ID კოდი	id კოდი	PDB ფაილის ცალსახა იდენტიფიკატორი

დეტალები:

- მაკროკოლეკულათა კომპლექსების შემთხვევაში კლასიფიცირების ველმა უნდა აღწეროს თითოეული მოლეკულის კლასი. იმის გამო, რომ კლასიფიცირების ველის სიგრძე შეზღუდულია 40 ასო-ნიშნით, ხშირად მიმართავენ შემოკლებებს. ამ შემთხვევაში სრული დასახელება მოიცემა KEYWDS ჩანაწერში.

- კლასიფიცირებისას ეყარებიან მონაცემებს ფუნქციის, მეტაბოლური როლის, მოლეკულის ტიპის, უჯრედული ლოკალიზაციის და სხვათა შესახებ. როცა მოლეკულას ორმაგი ფუნქცია გააჩნია, ორივე მოიცემა.

კოორდინატების მომცველი PDB ჩანაწერის ID კოდი არ იწყება 0-დან, რადგანაც 0 იდენტიფიცირებას უკეთებს იმ NOC ("no coordinates) ფაილს, რომელიც მოიცავს მხოლოდ ბიბლიოგრაფიულ მონაცემებს და არ მოიცავს სტრუქტურულ ინფორმაციას.

დამოკიდებულება სხვა ჩანაწერებთან

HEADER-ში წარმოდგენილი კლასიფიკაციური ინფორმაცია დუბლირებულია KEYWDS-ში შეუმოკლებლად და მკაცრი რიგითობის დაუცველად.

მაგალითი

```
1 2 3 4 5 6 7
123456789012345678901234567890123456789012345678901234567890
HEADER MUSCLE PROTEIN 02-JUN-93 1MYS

HEADER HYDROLASE (CARBOXYLIC ESTER) 08-APR-93 2PHI

HEADER COMPLEX (LECTIN/TRANSFERRIN) 07-JAN-94 1LGB
```

ჩანაწერი OBSLTE

ჩანაწერის მიმოხილვა

OBSLTE მხოლოდ იმ ფაილებში გვხვდება, რომლებიც ამოღებულია ხმარებიდან.

იგი მიუთითებს სხვა ფაილების არსებობაზე, რომლებითაც მოცემულის ჩანაცვლება მოხდა.

ჩანაწერის ზოგადი ფორმატი ასეთია

პოზიციები მონაცემის ტიპი ველი განსაზღვრება

პოზიციები	მონაცემის ტიპი	ველი	განსაზღვრება
1 - 6	ჩანაწერის სახელი	"OBSLTE"	
9 - 10	გაგრძელება	გაგრძელება	მიუთითებს, რომ სტრიქონი წინას გაგრძელება
12 - 20	თარიღი	ჩანაცვ.თარიღი	ამ ფაილით წინას ჩანაცვლების თარიღი
22 - 25	IDკოდი	idკოდი	ფაილის საწყისი IDკოდი .
32 - 35	IDკოდი	ridკოდი	წინა ფაილის ჩამანაცვლებელი ფაილის ID კოდი
37 - 40	IDკოდი	ridკოდი	წინა ფაილის ჩამანაცვლებელი ფაილის ID კოდი
42 - 45	IDკოდი	ridკოდი	წინა ფაილის ჩამანაცვლებელი ფაილის ID კოდი
47 - 50	IDკოდი	ridკოდი	წინა ფაილის ჩამანაცვლებელი ფაილის ID კოდი
52 - 55	IDკოდი	ridკოდი	წინა ფაილის ჩამანაცვლებელი ფაილის ID კოდი
57 - 60	IDკოდი	ridკოდი	წინა ფაილის ჩამანაცვლებელი ფაილის ID კოდი
62 - 65	IDკოდი	ridკოდი	წინა ფაილის ჩამანაცვლებელი ფაილის ID კოდი
67 - 70	IDკოდი	ridკოდი	წინა ფაილის ჩამანაცვლებელი ფაილის ID კოდი

მაგალითი

```

1      2      3      4      5      6      7
123456789012345678901234567890123456789012345678901234567890
OBSLTE 31-JAN-94 IMBP 2MBP
  
```

ჩანაწერი TITLE

ჩანაწერის მიმოხილვა

TITLE ჩანაწერი იძლევა იმ ექსპერიმენტებისა და ანალიზის საშუალებების დასახელებებს, რომლებიც გამოიყენებოდნენ ფაილში წარმოდგენილი სტრუქტურის შესახებ ინფორმაციის მიღებისას. ის ფაქტიურად წარმოადგენს PDB ფაილის სათაურს.

ჩანაწერის ზოგადი ფორმატი ასეთია

პოზიციები მონაცემის ტიპი ველი განსაზღვრება

პოზიციები	მონაცემის ტიპი	ველი	განსაზღვრება
1 - 6	ჩანაწერის სახელი	"TITLE "	
9 - 10	გაგრძელება	გაგრძელება	მიუთითებს წინა სტრიქონის გაგრძელებაზე
11 - 70	ასოთა რიგი	სახელი	ექსპერიმენტის დასახელება

დეტალები

- ფაილის სათაური წარმოადგენს თავისუფალ ტექსტს და აღწერს ფაილში განთავსებულ მოლეკულას და იმ პროცედურებსა თუ პირობებს, რომლებიც მოცემულ ფაილს განასხვავებს სხვა მსგავსი ფაილებისაგან.

TITLE შესაძლებელია აღწერდეს :

- ექსპერიმენტის ტიპს

- მუტაციებს

- იმას, ვთქვათ, რომ ფაილში წარმოდგენილია მხოლოდ ალფა ნახშირბადების კოორდი ნატები.

დამოკიდებულება სხვა ჩანაწერებთან

COMPND, SOURCE, EXPDTA, და REMARK-ში ვხვდებით ინფორმაციას, რომელიც ასევე შეიძლება შეგვხვდეს TITLE-ში.

მაგალითი

```
1 2 3 4 5 6 7
123456789012345678901234567890123456789012345678901234567890
TITLE RHIZOPUSPEPSIN COMPLEXED WITH REDUCED PEPTIDE INHIBITOR
TITLE BETA-GLUCOSYLTRANSFERASE, ALPHA CARBON COORDINATES ONLY
TITLE NMR STUDY OF OXIDIZED THIOREDOXIN MUTANT (C62A,C69A,C73A)
TITLE 2 MINIMIZED AVERAGE STRUCTURE
```

ჩანაწერი CAVEAT

ჩანაწერის მიმოხილვა

ჩანაწერი CAVEAT გაფრთხილებთ შეცდომების შესახებ.

ჩანაწერის ზოგადი ფორმატი ასეთია

პოზიციები მონაცემის ტიპი ველი განსაზღვრება

1 - 6	ჩანაწერის სახელი	"CAVEAT"
9 - 10	გაგრძელება	გაგრძელება მიუთითებს წინა სტრიქონის გაგრძელებაზე
12 - 15	IDკოდი	IDკოდი PDB ფაილის ID კოდი.
20 - 70	ასოთა რიგი	კომენტარი აღწერს გაფრთხილების მიზეზს.

დეტალები

- CAVEAT გვხვდება იმ ფაილში, რომელიც ითვლება რაღაც შეცდომის მომცველად. შეცდომა შემოწმების პროცესში ვლინდება. შემოწმება ფაილის წარდგენისთანავე წარმოებს.

ჩანაწერი COMPND

ჩანაწერის მიმოხილვა

ჩანაწერი COMPND გვძლევს ფაილში განთავსებულ მოლეკულათა ჩამონათვალს. თითოეული მოლეკულა აღიწერება მთელი რიგი მახასიათებლებით. გამომდინარე იქიდან, რომ მოლეკულათა ზუსტი დახასიათება საკმაოდ რთულია, მახასიათებელთა ზუსტი შერჩევის მიზნით წარმოებს დამატებითი კონსულტაციები დეპოზიტორებთან.

ჩანაწერის ზოგადი ფორმატი ასეთია

პოზიციები მონაცემის ტიპი ველი განსაზღვრება

1 - 6	ჩანაწერის სახელი	"COMPND"	
9 - 10	გაგრძელება	გაგრძელება	მოუთითებს წინა სტრიქონის გაგრძელებაზე
11 - 70	სპეციფიკაც. ჩამონათვლ.	ქიმ. ნაერთი	მოლეკულური კომპონენტების აღწერა

დეტალები

- ჩანაწერი compound წარმოადგენს მახასიათებელთა ჩამონათვალს, რომლებიც მოცემულია ცხრილის სახით:

მახასიათებელი გვიჩვენებს

MOL_ID	თითოეული კომპონენტის ნომერს; გამოიყენება ასევე SOURCE-ში
MOLECULE	მაკრომოლეკულის სახელს.
CHAIN	ჯაჭვთა იდენტიფიკატორების ჩამონათვალს.
FRAGMENT	თუ მოლეკულის დომენთან ან რაღაც უბანთან გვაქვს საქმე.
SYNONYM	MOLECULE-ის სინონიმთა ჩამონათვალს.
EC	მოლეკულის დამახასიათებელი EC (Enzyme Commission) რიცხვს. თუ ამ რიცხვთა რაოდენობა ერთზე მეტია, ისინი გამოიყოფიან მძიძით.
ENGINEERED	მოლეკულა მიღებულია რეკომბინანტული ტექნოლოგიებითა, თუ ქიმიური სინთეზის შედეგია.
MUTATION	მუტაციების არსებობას.
OTHER_DETAILS	დამატებით ინფორმაციას.

დამოკიდებულება სხვა ჩანაწერებთან

მუტაციების შემთხვევაში ჩანაწერი SEQADV ასევე წარმოგვიდგენს ცვლილებებს აღნიშნულ მოლეკულაში.

მაგალითი

1 2 3 4 5 6 7
1234567890123456789012345678901234567890123456789012345678901234567890
COMPND MOL_ID: 1;
COMPND 2 MOLECULE: HEMOGLOBIN;
COMPND 3 CHAIN: A, B, C, D;
COMPND 4 ENGINEERED: YES;
COMPND 5 MUTATION: YES
COMPND 6 OTHER_DETAILS: DEOXY FORM

COMPND MOL_ID: 1;
COMPND 2 MOLECULE: COWPEA CHLOROTIC MOTTLE VIRUS;
COMPND 3 CHAIN: A, B, C;
COMPND 4 SYNONYM: CCMV;
COMPND 5 MOL_ID: 2;
COMPND 6 MOLECULE: RNA (5'-("AP"UP"AP"U)-3');
COMPND 7 CHAIN: D, F;
COMPND 8 ENGINEERED: YES;
COMPND 9 MOL_ID: 3;
COMPND 10 MOLECULE: RNA (5'-("AP"U)-3');
COMPND 11 CHAIN: E;
COMPND 12 ENGINEERED: YES

COMPND MOL_ID: 1;
COMPND 2 MOLECULE: HEVAMINE A;
COMPND 3 CHAIN: A;
COMPND 4 EC: 3.2.1.14, 3.2.1.17;
COMPND 5 OTHER_DETAILS: PLANT ENDOCHITINASE/LYSOZYME

ჩანაწერი SOURCE

ჩანაწერის მიმოხილვა

ჩანაწერი SOURCE განსაზღვრავს ფაილში განთავსებული თითოეული მოლეკულის ბიოლოგიურ და/ან ქიმიურ პირველწყაროს. პირველწყარო აღიწერება როგორც საზოგადო სახელით ისე მეცნიერული სახელით, ვთქვათ გვარზე და სახეობაზე დაფუძნებით. საკვლევი ბიოლოგიური ერთეულის ცალსახა იდენტიფიცირების ხელის შეწყობის მიზნით უჯრედთა ხაზისათვის მოიცემა ასევე შტამი და/ან უჯრედოვანი ხაზის დასახელება.

ჩანაწერის ზოგადი ფორმატი ასეთია

პოზიციები მონაცემის ტიპი ველი განსაზღვრება

1 - 6	ჩანაწერის სახელი	"SOURCE"	
9 - 10	გაგრძელება	გაგრძელება	მოუთითებს წინა სტრიქონის გაგრძელებაზე
11 - 70	სპეციფიკაცია	წყაროს სახელი	მაკრომოლეკულის წყაროს იდენტიფიცირება

დეტალები

„წყაროს სახელის“ ველში გამოყენებადი
მახასიათებელი ნიშნები (Token)

გვიჩვენებს

MOL_ID	თითოეული მოლეკულის ნომერს . იგივე რაც COMPND
SYNTHETIC	წყარო ქიმიური სინთეზია
FRAGMENT	საქმე გვაქვს დომენტთან ან ფრაგმენტთან
ORGANISM_SCIENTIFIC	ორგანიზმის სამეცნიერო სახელს .
ORGANISM_COMMON	ორგანიზმის საზოგადოებრივ სახელს
STRAIN	შტამს .
VARIANT	ჯიშს (variant).
CELL_LINE	ექსპერიმენტში გამოყენებულ უჯრედის სპეციფიკურ ხაზს
ATCC	ქსოვილური კულტურის ნომერს, განსაზღვრულს კულტურათა კოლექციის ამერიკული ტიპირების შესაბამისად
ORGAN	ქსოვილთა ორგანულ ჯგუფს, რომელიც პასუხისმგებელია სპეციალიზირებული ფუნქციის განხორციელებაზე
TISSUE	უჯრედთა ორგანულ ჯგუფს, რომელიც ხასიათდება საერთო ფუნქციითა და სტრუქტურით
CELL	უჯრედის ტიპს
ORGANELLE	მემბრანით შემოსაზღვრულ უჯრედშიდა სტრუქტურას
SECRETION	სეკრეციებს (მაგალითად ნერწყვი, შარდი, შხამი ან სხვა) საიდანაც მოხდა მოლეკულის გამოყოფა
CELLULAR_LOCATION	უჯრედშიდა ლოკალიზაციას
PLASMID	გენის შემცველ პლაზმიდას
GENE	გენს
EXPRESSION_SYSTEM	სისტემას, რომელიც გამოყენებული იყო რეკომბინანტული მოლეკულის ექსპრესიის პროცესში
EXPRESSION_SYSTEM_STRAIN	შტამს, რომელშიც განხორციელდა ექსპრესია
EXPRESSION_SYSTEM_VARIANT	ორგანიზმის ჯიშს, რომელშიც განხორციელდა ექსპრესია
EXPRESSION_SYSTEM_CELL_LINE	იმ უჯრედულ ხაზს, რომელშიც განხორციელდა ექსპრესია
EXPRESSION_SYSTEM_ATCC_NUMBER	ექსპრესიის სისტემის ATCC რიცხვს
EXPRESSION_SYSTEM_ORGAN	იმ სპეციალურ ორგანოს, რომელშიც განხორციელდა ექსპრესია
EXPRESSION_SYSTEM_TISSUE	იმ ქსოვილს, რომელშიც განხორციელდა ექსპრესია
EXPRESSION_SYSTEM_VECTOR	ექსპრესიის ვექტორს .
EXPRESSION_SYSTEM_PLASMID	რეკომბინაციისას გამოყენებულ პლაზმიდას
EXPRESSION_SYSTEM_GENE	რეკომბინაციისას გამოყენებულ გენის სახელს
OTHER_DETAILS	სხვა ინფორმაციას.

დამოკიდებულება სხვა ჩანაწერებთან

COMPND-ში მოცემულ თითოეულ მოლეკულას უნდა გააჩნდეს თავისი პირველწყარო.

მაგალითი

```
1 2 3 4 5 6 7
123456789012345678901234567890123456789012345678901234567890
SOURCE MOL_ID: 1;
SOURCE 2 ORGANISM_SCIENTIFIC: AVIAN SARCOMA VIRUS;
SOURCE 3 STRAIN: SCHMIDT-RUPPIN B;
SOURCE 4 EXPRESSION_SYSTEM: ESCHERICHIA COLI;
SOURCE 5 EXPRESSION_SYSTEM_PLASMID: PRC23IN

SOURCE MOL_ID: 1;
SOURCE 2 ORGANISM_SCIENTIFIC: GALLUS GALLUS;
SOURCE 3 ORGANISM_COMMON: CHICKEN;
SOURCE 4 ORGAN: HEART;
SOURCE 5 TISSUE: MUSCLE

SOURCE MOL_ID: 1;
SOURCE 2 EXPRESSION_SYSTEM: ESCHERICHIA COLI;
SOURCE 3 EXPRESSION_SYSTEM_STRAIN: BE167;
SOURCE 4 FRAGMENT: RESIDUES 1-16;
SOURCE 5 ORGANISM_SCIENTIFIC: BACILLUS AMYLOLIQUEFACIENS;
SOURCE 6 EXPRESSION_SYSTEM: ESCHERICHIA COLI;
SOURCE 7 FRAGMENT: RESIDUES 17-214;
SOURCE 8 ORGANISM_SCIENTIFIC: BACILLUS MACERANS
```

ჩანაწერი KEYWDS

ჩანაწერის მიმოხილვა

ჩანაწერი KEYWDS მოიცავს ფაილისათვის დამახასიათებელ სიტყვათა (ტერმინთა) ერთობას.

ჩანაწერის ზოგადი ფორმატი ასეთია

პოზიციები მონაცემის ტიპი ველი განსაზღვრება

1 - 6	ჩანაწერის სახელი	"KEYWDS"	
9 - 10	გაგრძელება	გაგრძელება	მოუთითებს წინა სტრიქონის გაგრძელებაზე
11 - 70	ჩამონათვალი	დამახ.სიტყვები	გვიქმნის ზოგად წარმოდგენას მოლეკულაზე

დამოკიდებულება სხვა ჩანაწერთან

ჩანაწერი HEADER მოიცავს იმ განმსაზღვრელ ტერმინებს, რომლებიც ასევე არის KEYWDS-ში.

მაგალიტი

```
1 2 3 4 5 6 7
123456789012345678901234567890123456789012345678901234567890
KEYWDS LYASE, TRICARBOXYLIC ACID CYCLE, MITOCHONDRION, OXIDATIVE
KEYWDS 2 METABOLISM
```

ჩანაწერი EXPDTA

ჩანაწერის მიმოხილვა

ჩანაწერი EXPDTA წარმოგვიდგენს ინფორმაციას ექსპერიმენტის შესახებ.

ინფორმაცია მოიცემა შემდეგი ტიპის ჩანაწერების სახით:

```
ELECTRON DIFFRACTION
ELECTRON MICROSCOPY
CRYO-ELECTRON MICROSCOPY
SOLUTION SCATTERING, THEORETICAL MODEL
FIBER DIFFRACTION
FLUORESCENCE TRANSFER
NEUTRON DIFFRACTION
NMR (may have a qualifier e.g. number of models see examples below)
SOLUTION SCATTERING
THEORETICAL MODEL*
X-RAY DIFFRACTION
```

ჩანაწერის ზოგადი ფორმატი ასეთია

პოზიციები მონაცემის ტიპი ველი განსაზღვრება

```
-----
1 - 6 ჩანაწერის სახელი "EXPDTA"
9 - 10 გაგრძელება გაგრძელება მოუთითებს წინა სტრიქონის გაგრძელებაზე
11 - 70 ჩამონათვალი მეთოდები აღწერს გამოყენებულ ექსპერიმენტულ
მიდგომებს
```

დამოკიდებულება სხვა ჩანაწერებთან

თუ ექსპერიმენტი არის NMR, ფიბრილათა ანალიზი ან თეორიული მოდელირება, მაშინ ამის შესახებ ინფორმაცია იქნება TITLE-ში და შესაძლებელია EXPDTA და REMARK-შიც

მაგალითი

```
1 2 3 4 5 6 7
123456789012345678901234567890123456789012345678901234567890
EXPDTA X-RAY DIFFRACTION
EXPDTA NEUTRON DIFFRACTION; X-RAY DIFFRACTION
EXPDTA NMR, 32 STRUCTURES
EXPDTA NMR, REGULARIZED MEAN STRUCTURE
EXPDTA FIBER DIFFRACTION
```

ჩანაწერი AUTHOR

ჩანაწერის მიმოხილვა

ჩანაწერი AUTHOR წარმოგვიდგენს იმ პიროვნებათა სახელებს, რომლებიც პასუხისმგებელი არიან ფაილში განთავსებულ მონაცემებზე.

ჩანაწერის ზოგადი ფორმატი ასეთია

პოზიციები მონაცემის ტიპი ველი განსაზღვრება

1 - 6	ჩანაწერის სახელი	"AUTHOR"		
9 - 10	გაგრძელება	გაგრძელება	მიუთითებს წინა სტრიქონის გაგრძელებაზე	
11 - 70	ჩამონათვალი	ავტორები	გვამლევს მძიმეებით გამოყოფილ ავტორთა სახელებს	

დამოკიდებულება სხვა ჩანაწერთან

სახელთა ფორმატი AUTHOR-ში იგივეა, რაც JRNL და REMARK -ში.

მაგალითი

1	2	3	4	5	6	7
123456789012345678901234567890123456789012345678901234567890						
AUTHOR	M. B. BERRY, B. MEADOR, T. BILDERBACK, P. LIANG, M. GLASER,					
AUTHOR	2 G. N. PHILLIPS JUNIOR, T. L. ST. STEVENS					

ჩანაწერი REVDAT

ჩანაწერის მიმოხილვა

ჩანაწერი REVDAT წარმოგვიდგენს მოდიფიკაციათა ისტორიას.

ჩანაწერის ზოგადი ფორმატი ასეთია

პოზიციები მონაცემის ტიპი ველი განსაზღვრება

1 - 6	ჩანაწერის სახელი	"REVDAT"		
8 - 10	მთელი	მოდიფ. ნომერი	მოდიფიკაციის ნომერი	
11 - 12	გაგრძელება	გაგრძელება	მიუთითებს წინა სტრიქონის გაგრძელებაზე	
14 - 22	თარიღი	მოდიფ. თარიღი	მოდიფიკაციის თარიღი	
24 - 28	ასოთა რიგი (5)	მოდიფ. Id	ახდენს კონკრეტული მოდიფიციების იდენტიფიცირებას	
32	მთელი	მოდიფ. ტიპი	მთელი რიცხვი ახასიათებს კონკრეტულ მოდიფიკაციას	
40 - 45	ასოთა რიგი (6)	დახასიათება	მოდიფიკაციის ტიპის სახელი	

დეტალები

- მოდიფიკაციის ტიპის შესაბამისი ცვლადი შემდეგ მთელ მნიშვნელობებს იღებს:

0	შეესაბამება პირველ (საწყის) წარდგენილ ფაილს.
1	შეესაბამება სხვადასხვა ხასიათის, ძირითადად ტიპოგრაფულ შესწორებებს
2	შეესაბამება შესწორებას CONECT ჩანაწერში.
3	შეესაბამება შესწორებას კოორდინატებში ან ტრანსფორმაციის მატრიცებში.

მაგალითი

	1	2	3	4	5	6	7
123456789012345678901234567890123456789012345678901234567890							
REVDAT	3	15-OCT-89	1PRC	1	REMARK		
REVDAT	2	19-APR-89	1PRC	2	CONECT		
REVDAT	1	09-JAN-89	1PRC	0			

ჩანაწერი SPRSDE

ჩანაწერის მიმოხილვა

ჩანაწერი SPRSDE გვაძლევს იმ ფაილების ID კოდთა ჩამონათვალს, რომლებიც შეცვლილია და შესაბამისად არ ფიგურირებს დღევანდელ PDB-ის საერთო ჩამონათვალში. ერთმა ფაილმა შეიძლება შეცვალოს რამდენიმე.

ჩანაწერის ზოგადი ფორმატი ასეთია

პოზიციები მონაცემის ტიპი ველი განსაზღვრება

1 - 6	ჩანაწერის სახელი	"SPRSDE"	
9 - 10	გაგრძელება	გაგრძელება	მიუთითებს წინა სტრიქონის გაგრძელებაზე
12 - 20	თარიღი	ჩანაცვ.თარიღ.	მიუთითებს თარიღზე როცა მოცემულმა ფაილმა ჩანაცვლა წინა
22 - 25	IDკოდი	idკოდი	არსებული ვარიანტის IDკოდი.
32 - 35	IDკოდი	sIDკოდი	ჩანაცვლებულის ID კოდი.
37 - 40	IDკოდი	sIDკოდი	ჩანაცვლებულის ID კოდი.
42 - 45	IDკოდი	sIDკოდი	ჩანაცვლებულის ID კოდი.
47 - 50	IDკოდი	sIDკოდი	ჩანაცვლებულის ID კოდი.
52 - 55	IDკოდი	sIDკოდი	ჩანაცვლებულის ID კოდი.
57 - 60	IDკოდი	sIDკოდი	ჩანაცვლებულის ID კოდი.
62 - 65	IDკოდი	sIDკოდი	ჩანაცვლებულის ID კოდი.
67 - 70	IDკოდი	sIDკოდი	ჩანაცვლებულის ID კოდი.

მაგალითი

	1	2	3	4	5	6	7
123456789012345678901234567890123456789012345678901234567890							
SPRSDE	17-JUL-84	4HHB	1HHB				
SPRSDE	27-FEB-95	1GDJ	1LH4	2LH4			

ჩანაწერი JRNL

ჩანაწერის მიმოხილვა

ჩანაწერი JRNL მოიცავს ლიტერატურის ციტირებას.

ჩანაწერის ზოგადი ფორმატი ასეთია

პოზიციები	მონაცემის ტიპი	ველი	განსაზღვრება
1 - 6	ჩანაწერის სახელი	"JRNL "	
13 - 70	ასოთა რიგი	ტექსტი	დეტალები იხილეთ მაგალითში

დამოკიდებულება სხვა ჩანაწერებთან

JRNL -ში მითითებული პუბლიკაციები შესაძლებელია არ იქნეს განმეორებული REMARK 1-ში.

მაგალითი

	1	2	3	4	5	6	7
123456789012345678901234567890123456789012345678901234567890							
JRNL	AUTH	G. FERMI, M. F. PERUTZ, B. SHAANAN, R. FOURME					
JRNL	TITL	THE CRYSTAL STRUCTURE OF HUMAN DEOXYHAEMOGLOBIN AT					
JRNL	TITL 2	1.74 Å RESOLUTION					
JRNL	REF	J. MOL. BIOL.			V. 175	159	1984
JRNL	REFN	ASTM JMOBAK	UK	ISSN	0022-2836		

პირველადი სტრუქტურის განყოფილება

ფაილში პირველადი სტრუქტურის განყოფილება წარამოგვიდგენს მაკრომოლეკულის თითოეული ჯაჭვის ამინმჟავურ თანამიმდევრობას. ამ ჩანაწერებში ასევე მოიცემა ჯაჭვების იდენტიფიკატორები და ჯაჭვების რიცხვი.

ჩანაწერი DBREF

ჩანაწერის მიმოხილვა

ჩანაწერი DBREF აკავშირებს PDB-ში განთავსებულ თანამიმდევრობებს სხვა თანამიმდევრობათა შესახებ მონაცემთა ბაზების შესაბამის ფაილებთან .

ჩანაწერის ზოგადი ფორმატი ასეთია

პოზიციები მონაცემის ტიპი	ველი	განსაზღვრება
1 - 6	ჩანაწერის სახელი	"DBREF "
8 - 11	ID კოდი	1dკოდი
13	ასო ნიშ.	ჯაჭვის ID
15 - 18	მთელი რიცხ.	თანამიმდ. დასაწყ.
19	ასო ნიშ.	ჩამატ. დასაწყისში
21 - 24	მთელი რიცხ.	თანამიმდ. ბოლო
25	ასო ნიშ.	ჩამატ. ბოლოში
27 - 32	ასოთა რიგი	მონაცემთ. ბაზა
34 - 41	ასოთა რიგი	იდენტიფიკატორი
43 - 54	ასოთა რიგი	ბაზაში 1dკოდი
56 - 60	მთელი რიცხ.	თანამიმდ. დასწყ.
61	ასო ნიშ.	ჩამატ. დასაწყისში
63 - 67	მთელი რიცხ.	თანამიმდ. ბოლო
68	ასო ნიშ.	ჩამატ. ბოლოში

არსებულის (PDB) ID კოდი.
 ჯაჭვის იდენტიფიკატორი.
 თანამიმდევრობის საწყისის ნომერი
 თანამიმდევრობის დასაწყისში ჩამატ. კოდი
 თანამიმდევრობის ბოლოს ნომერი
 თანამიმდევრობის ბოლოში ჩამატების კოდი
 თანამიმდევრობათა მონაცემთ. ბაზის სახელი.
 ბაზაში თანამიმდევრობის იდენტიფიკატორი.
 ბაზაში თანამიმდ. საიდენტ. კოდი
 ბაზის თანამიმდევრობის საწყისის ნომერი.
 ბაზის თანამიმდევრობის დასაწყისში ჩამატ. კოდი, თუ მითითება ხდება PDB-ზე
 ბაზის თანამიმდევრობის ბოლოს ნომერი
 ბაზის თანამიმდევრობის ბოლოში ჩამატ. კოდი, თუ მითითება ხდება PDB-ზე

დეტალები

- DBREF ჩანაწერში გამოყენებული მონაცემთა ბაზების სახელები და მათი შემოკლებები

მონაცემთა ბაზის სახელი მონაცემთა ბაზ. შემოკლ. სახელი
 (განთავსდება 27 - 32 პოზიციებში)

GenBank	GB
Protein Data Bank	PDB
Protein Identification Resource	PIR
SWISS-PROT	SWS
TREMBL	TREMBL
UNIPROT	UNP

დამოკიდებულება სხვა ჩანაწერებთან

DBREF წარმოადგენს თანამიმდევრობას, რომელიც განთავსებულია SEQRES ჩანაწერში.

მაგალითი

	1	2	3	4	5	6	7
123456789012345678901234567890123456789012345678901234567890							
DBREF	2J83	A	61	322	UNP	Q8TL28	Q8TL28_METAC
DBREF	2J83	B	61	322	UNP	Q8TL28	Q8TL28_METAC
DBREF	1ABC	B	1B	36	PDB	1ABC	1ABC
DBREF	3AKY		3	220	SWS	P07170	KAD1_YEAST
DBREF	1HAN		2	288	GB	397884	X66122
DBREF	3HSV	A	1	92	SWS	P22121	HSF_KLULA
DBREF	3HSV	B	1	92	SWS	P22121	HSF_KLULA
DBREF	1ARL		1	307	SWS	P00730	CBPA_BOVIN

ჩანაწერი SEQADV

ჩანაწერის მიმოხილვა

ჩანაწერი SEQADV მიუთითებს კონფლიქტს, რომელიც შეიძლება არსებობდეს SEQRES-ში განთავსებულ თანამიმდევრობათა შესახებ ინფორმაციასა და DBREF მითითებულ კონკრეტული თანამიმდ. ბაზის თანამიმდევრობის შესაბამის ელემენტს შორის. უნდა შევნიშნოთ, რომ ჩანაწერი გვიჩვენებს განსხვავებებს და არა შეცდომებს. PDB შიძლება შეიცავდეს შენიშვნას, რომლის მიხედვითაც ჩვენთვის ცნობილი გახდება დეპოზიტორის აზრი იმის თაობაზე, თუ რომელ ბაზაში განთავსებული თანამიმდევრობაა კორექტული.

ჩანაწერის ზოგადი ფორმატი ასეთია

პოზიციები	მონაცემის ტიპი	ველი	განსაზღვრება
1- 6	ჩანაწერის სახელი	"SEQADV"	
8-11	IDკოდი	idკოდი	PDB ფაილის ID კოდი.
13-15	ამინომე.სახელი	ამინომე.სახელი	კონფლიქტთან დაკავშირებული ამინომე.სახელი
17	ასო ნიშ.	ჯაჭვის ID	ჯაჭვის იდენტიფიკატორი.
19-22	მთელი რიგი.	ნომერი	ამინომე.სახელი თანამიმდევრობაში
23	ასო ნიშ.	idკოდი	ჩამატების კოდი.
25-28	ასოთა რიგი	მონაც.ბაზა	
30-38	ასოთა რიგი	ბაზაში IDკოდი	თანამიმდ.ბაზაში საიდენტ. კოდი
40-42	ამინომე.სახელი	ბაზაში ამ. სახ.	თანამიმდ. ბაზაში ამინომე.სახელი.
44-48	მთელი რიგი.	ბაზის თანამიმდ.	თანამიმდ.ბაზის თანამ.-ში ნომერი
50-70	ასოთა რიგი	კონფლიქტი	კონფლიქტის კომენტირება.

დეტალები

• შესაძლო კონფლიქტთა კომენტირებისათვის გამოყენებად ტერმინთა მაგალითებად მოვიყვანთ:

Cloning artifact
Conflict
Engineered
Disordered
Variant
Insertion
Deletion
Microheterogeneity
D-configuration

• თუ კონფლიქტი არ კლასიფიცირდება მოცემული კომენტარების შესაბამისად, მაშინ მოიცემა მითითება ლიტერატურაზე ან შენიშვნა.

დამოკიდებულება სხვა ჩანაწერებთან

SEQADV მიგვითითებს თანამიმდევრობაზე SEQRES-დან და მონაცემთა ბაზაზე DBREF-დან.

მაგალითი

	1	2	3	4	5	6	7
123456789012345678901234567890123456789012345678901234567890							
SEQADV 2J83 ALA A	269	UNP	Q8TL28	CYS	269	ENGINEERED MUTATION	
SEQADV 2J83 ALA B	269	UNP	Q8TL28	CYS	269	ENGINEERED MUTATION	
SEQADV 3ABC MET A	-1	SWS	P10725			CLONING ARTIFACT	
SEQADV 3ABC GLY A	50	SWS	P10725	VAL	50	ENGINEERED	

ჩანაწერი SEQRES

ჩანაწერის მიმოხილვა

ჩანაწერი SEQRES მოიცავს ფაილში განთავსებული მოლეკულის თითოეული ჯაჭვის ამინომჟავურ ან ნუკლეოტიდურ თანამიმდევრობას.

ჩანაწერის ზოგადი ფორმატი ასეთია

პოზიციები მონაცემის ტიპი	ველი	განსაზღვრება
1- 6	ჩანაწერის სახელი	" SEQRES "
9-10	მთელი რიცხვ.	რიგითი ნომერი
12	სპო ნომერი (გამოტოვ.)	IDკოდი
14-17	მთელი რიცხვ	ამინომჟ.რიცხვი
20-22	ამინომჟ.დასახელება	ამინომჟ.დასახ
24-26	ამინომჟ.დასახელება	ამინომჟ.დასახ
28-30	ამინომჟ.დასახელება	ამინომჟ.დასახ
32-34	ამინომჟ.დასახელება	ამინომჟ.დასახ
36-38	ამინომჟ.დასახელება	ამინომჟ.დასახ
40-42	ამინომჟ.დასახელება	ამინომჟ.დასახ
44-46	ამინომჟ.დასახელება	ამინომჟ.დასახ
48-50	ამინომჟ.დასახელება	ამინომჟ.დასახ
52-54	ამინომჟ.დასახელება	ამინომჟ.დასახ
56-58	ამინომჟ.დასახელება	ამინომჟ.დასახ
60-62	ამინომჟ.დასახელება	ამინომჟ.დასახ
64-66	ამინომჟ.დასახელება	ამინომჟ.დასახ
68-70	ამინომჟ.დასახელება	ამინომჟ.დასახ

მაგალითი

	1	2	3	4	5	6	7
123456789012345678901234567890123456789012345678901234567890							
SEQRES 1 A	21	GLY ILE VAL	GLU GLN	CYS CYS	THR SER	ILE CYS	SER LEU
SEQRES 2 A	21	TYR GLN LEU	GLU ASN	TYR CYS	ASN		
SEQRES 1 B	30	PHE VAL ASN	GLN HIS	LEU CYS	GLY SER	HIS LEU	VAL GLU
SEQRES 2 B	30	ALA LEU TYR	LEU VAL	CYS GLY	GLU ARG	GLY PHE	PHE TYR
SEQRES 3 B	30	THR PRO	LYS ALA				
SEQRES 1 C	21	GLY ILE VAL	GLU GLN	CYS CYS	THR SER	ILE CYS	SER LEU
SEQRES 2 C	21	TYR GLN LEU	GLU ASN	TYR CYS	ASN		
SEQRES 1 D	30	PHE VAL ASN	GLN HIS	LEU CYS	GLY SER	HIS LEU	VAL GLU
SEQRES 2 D	30	ALA LEU TYR	LEU VAL	CYS GLY	GLU ARG	GLY PHE	PHE TYR
SEQRES 3 D	30	THR PRO	LYS ALA				

შეზღუდვებთან დაკავშირებული გაურკვევლობანი.

პოლისაქარიდების წარმოდგენა SEQRES-ში არ ხდება

ციკლური პეპტიდების შემთხვევაში PDB ყველა ნაშთს წარმოადგენენ როგორც N-ბოლო ნაშთს.

SEQRES ჩანაწერში ვერ განასხვავებთ რიბო- და დეზოქსირიბო-ნუკლეინის მკავათა თანამიმდევრობებისათვის გაკეთებულ ჩანაწერს.

ჩანაწერი MODRES

ჩანაწერის მიმოხილვა

ჩანაწერი MODRES აღწერს მოდიფიკაციებს (მაგალითად: ქიმიურს, პოსტტრანსლაციურს და სხვა). ასევე იძლევა PDB-ში წარმოდგენილი მოდიფიცირებული ნაშთისა და შესაბამისი სტანდარტული ნაშთის სახელებს.

ჩანაწერის ზოგადი ფორმატი ასეთია

პოზიციები	მონაცემის ტიპი	ველი	განსაზღვრება
1-6	ჩანაწერის სახელი	"MODRES"	
8-11	IDკოდი	idკოდი	ფაილის ID კოდი.
13-15	ამინომე.დასახელება	ამინომე.დასახ	მოდიფ. ამინმეჯის დასახელება.
17	ასო ნიშანი	ჯაჭვის ID	ჯაჭვის იდენტიფიკატორი.
19-22	მთელი რიცხ.	ნომერი თან.	თანამიმდევრობაში ამინომე. ნომერი.
23	ასო ნიშანი	i კოდი	ჩამატების კოდი.
25-27	ამინომე.დასახელება	სტნდ.ამინომე.	სტანდარტული ამინმეჯის სახელი
30-70	ასოთა როგი	კომენტარი	მოდიფიკაციის აღწერა

დეტალები

- მოდიფიკაციათა მაგალითებია:
 - Glycosylation site
 - Post-translational modification
 - Designed chemical modification
 - Phosphorylation site
 - Blocked N-terminus
 - Aminated C-terminus
 - D-configuration
 - Reduced peptide bond

მაგალითი

	1	2	3	4	5	6	7
1234567890123456789012345678901234567890123456789012345678901234567890							
MODRES 1ABC ASN A	22A	ASN	GLYCOSYLATION SITE				
MODRES 2ABC TRQ A	50A	TRP	POST-TRANSLATIONAL MODIFICATION				
MODRES 3ABC DAL A	32	ALA	POST-TRANSLATIONAL MODIFICATION, D-ALANINE				
MODRES 3ABC DAL B	32	ALA	POST-TRANSLATIONAL MODIFICATION, D-ALANINE				

მეორეული სტრუქტურის განყოფილება

მეორეული სტრუქტურის განყოფილება აღწერს ცილის სივრცულ სტრუქტურაში დამზერით სპირალებს, გოფირებულ-ფურცლოვან სტრუქტურებსა და მოსახვევებს.

ჩანაწერი HELIX

ჩანაწერის მიმოხილვა

მოლეკულაში სპირალური ფრაგმენტების პოზიციათა იდენტიფიცირების მიზნით გამოიყენება ჩანაწერი HELIX. სპირალს მინიჭებული აქვს სახელი და ნომერი. მითითებულია როგორც სპირალის სიგრძე, ისე საწყისი და საბოლოო ნაშთის ნომერი.

ჩანაწერის ზოგადი ფორმატი ასეთია

პოზიციები მონაცემის ტიპი	ველი	განსაზღვრება
1–6	ჩანაწერის სახელი	"HELIX "
8–10	მთელი რიცხვ.	რიგითი ნომერი
12–14	ასოთა რიგი (3)	სპირალის ID
16–18	ამინომე. დასახელ.	საწყ.ამინომე.სახ.
20	ასო ნიშანი	ჯაჭვის ID
22–25	მთელი რიცხვ.	საწყ.ნომ.ჯაჭვში
26	ასო ნიშანი	საწყ.ჩამ.კოდი
28–30	ამინომე. დასახელ.	ბოლოამინომე.სახ.
32	ასო ნიშანი	ჯაჭვის ID
34–37	მთელი რიცხვ.	ბოლოს.ნომ.ჯაჭვში
38	ასო ნიშანი	ბოლოს.ჩამ.კოდი
39–40	მთელი რიცხვ.	სპირალის ტიპი
41–70	ასოთა რიგი	კომენტარი
72–76	მთელი რიცხვ.	სიგრძე

დეტალები

- საწყის ნაშთად მიიჩნევა სპირალის N-ბოლოში განთავსებული ნაშთი. გამოიყენება სპირალთა შემდეგი კლასიფიკაცია:

სპირალის კლასი (ტიპი) კლასის ნომერი(39 – 40 პოზიციებში)

Right-handed alpha (default)	1
Right-handed omega	2
Right-handed pi	3
Right-handed gamma	4
Right-handed 3 ₁₀	5
Left-handed alpha	6
Left-handed omega	7
Left-handed gamma	8
27 ribbon/helix	9
Polyproline	10

მაგალითი

	1	2	3	4	5	6	7		
123456789012345678901234567890123456789012345678901234567890123456									
HELIX	1	HA	GLY	A	86	GLY	A	94	1
HELIX	2	HB	GLY	B	86	GLY	B	94	1

ჩანაწერი SHEET

ჩანაწერის მიმოხილვა

ჩანაწერი SHEET გამოიყენება მოლეკულაში β- გოფირებული ფურცლის წარმომქმნელი გაკიშული სტრუქტურის მქონე ფრაგმენტის პოზიციების იდენტიფიცირების მიზნით.

ჩანაწერის ზოგადი ფორმატი ასეთია

პოზიციები მონაცემის ტიპი ველი განსაზღვრება

1- 6	ჩანაწერის სახელი	"SHEET "	
8-10	მთელი რიცხვი	სეგმენტი	სეგმენტის ნომერი.
12-14	ასოთა რიგი (3)	ფურცლის ID	გოფრ.ფურცლის იდენტიფიკატორი .
15-16	მთელი რიცხვი.	სეგმენტთა რიცხვი	სეგმენტთა რიცხვი ფურცელში
18-20	ამინომე. დასახელ.	საწყ.ამინომე.სახ.	საწყისი ამინმეავას სახელი .
22	ასო ნიშანი	ჯაჭვის ID	ჯაჭვის იდენტიფიკატორი
23-26	მთელი რიცხვი.	საწყ.ნომ.ჯაჭვში	სეგმენტის საწყისი ამინმეავას ნომერი
27	ასო ნიშანი	საწყ.ჩამ.კოდი	სეგმენტის საწყისი ამინმეავას ჩამატ. კოდი .
29-31	ამინომე. დასახელ.	ბოლო ამინომე.სახ.	სეგმენტის ბოლო ამინმეავას სახელი
33	ასო ნიშანი	ჯაჭვის ID	ჯაჭვის იდენტიფიკატორი
34-37	მთელი რიცხვი.	ბოლო ამინომე.ნომ.	სეგმენტის ბოლო ამინმეავას ნომერი
38	ასო ნიშანი	ბოლოს.ჩამ.კოდი	სეგმენტის ბოლო ამინმეავას ჩამატ. კოდი .
39-40	მთელი რიცხვი.	ორიენტაცია	ახასიათებს ფურცელში მოცემული სეგმენტის ორიენტაციას წინას მიმართ. 0 - შეესაბამება პირველ ზოლს, 1 - პარალელურს, -1 კი ანტაპარალელურს.
42-45	ატომის სახელი	მოც. ატომი	მოცემული სეგმენტიდან წყ.ბმამი მონაწილე ატომის სახელი
46-48	ამინომე.სახელი	ამინომე.სახ.	მოცემული სეგმენტიდან წყ.ბმამი მონაწილე ამინმეავას სახელი.
50	ასო	ჯაჭვის Id	ჯაჭვის იდენტიფიკატორი.
51-54	მთელი რიცხვი.	ამინომე.ნომერი	მოცემული სეგმენტიდან წყ.ბმამი მონაწილე ამინმეავას ნომერი .
55	ასი ნიშანი	I კოდი	მოცემული სეგმენტიდან წყ.ბმამი მონაწილე ამინმეავას ჩამატების კოდი.
57-60	ატომის სახელი	წინას ატომი	წინა სეგმენტიდან წყ.ბმამი მონაწილე ატომის სახელი .
61-63	ამინომე.სახელი	წინას ამინომე.სახ.	წინა სეგმენტიდან წყ.ბმამი მონაწილე ამინმეავას სახელი.
65	ასო	ჯაჭვის Id	ჯაჭვის იდენტიფიკატორი.
66-69	მთელი რიცხვი.	ამინომე.ნომერი	წინა სეგმენტიდან წყ.ბმამი მონაწილე ამინმეავას ნომერი .
70	ასი ნიშანი	I კოდი	წინა სეგმენტიდან წყ.ბმამი მონაწილე ამინმეავას ჩამატების კოდი.

მაგალითი

	1	2	3	4	5	6	7
1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890
SHEET	1	A 5 THR A	107 ARG A	110 0			
SHEET	2	A 5 ILE A	96 THR A	99 -1	N LYS A	98 O THR A	107
SHEET	3	A 5 ARG A	87 SER A	91 -1	N LEU A	89 O TYR A	97
SHEET	4	A 5 TRP A	71 ASP A	75 -1	N ALA A	74 O ILE A	88
SHEET	5	A 5 GLY A	52 PHE A	56 -1	N PHE A	56 O TRP A	71
SHEET	1	B 5 THR B	107 ARG B	110 0			
SHEET	2	B 5 ILE B	96 THR B	99 -1	N LYS B	98 O THR B	107
SHEET	3	B 5 ARG B	87 SER B	91 -1	N LEU B	89 O TYR B	97
SHEET	4	B 5 TRP B	71 ASP B	75 -1	N ALA B	74 O ILE B	88
SHEET	5	B 5 GLY B	52 ILE B	55 -1	N ASP B	54 O GLU B	73

ქვევით მორიგი მაგალითის სახით მოცემული BS1 ფურცელი წარმოდგენს 9 ზოლიან ბეტა-კასრს (beta-barrel). იგი როგორც ვხედავთ წარმოდგენილია 9-ზოლიანი ფურცლით, რომლის პირველი და ბოლო სეგმენტები იდენტურია.

SHEET	1	BS1 9 VAL	13 ILE	17 0			
SHEET	2	BS1 9 ALA	70 ILE	73 1	O TRP	72 N ILE	17
SHEET	3	BS1 9 LYS	127 PHE	132 1	O ILE	129 N ILE	73
SHEET	4	BS1 9 GLY	221 ASP	225 1	O GLY	221 N ILE	130
SHEET	5	BS1 9 VAL	248 GLU	253 1	O PHE	249 N ILE	222
SHEET	6	BS1 9 LEU	276 ASP	278 1	N LEU	277 O GLY	252
SHEET	7	BS1 9 TYR	310 THR	318 1	O VAL	317 N ASP	278
SHEET	8	BS1 9 VAL	351 TYR	356 1	O VAL	351 N THR	318
SHEET	9	BS1 9 VAL	13 ILE	17 1	N VAL	14 O PRO	352

ფურცლის გაორებული ანუ ბიფურკაციული სტრუქტურის მაგალითად ქვემოთ მოყვანილია ორი ფურცლოვანი სტრუქტურა BS7 და BS8. ამ ფურცელთა მეორე და მესამე სეგმენტები იდენტურია.

SHEET	1	BS7 3 HIS	662 THR	665 0			
SHEET	2	BS7 3 LYS	639 LYS	648 -1	N PHE	643 O HIS	662
SHEET	3	BS7 3 ASN	596 VAL	600 -1	N TYR	598 O ILE	646
SHEET	1	BS8 3 ASN	653 TRP	656 0			
SHEET	2	BS8 3 LYS	639 LYS	648 -1	N LYS	647 O THR	655
SHEET	3	BS8 3 ASN	596 VAL	600 -1	N TYR	598 O ILE	646

ჩანაწერი TURN

ჩანაწერის მიმოხილვა

ჩანაწერი TURN ახდენს იმ მოსახვევთა და სხვა მცირე ზომის ყულვთა იდენტიფიცირებას, რომლებიც აკავშირებენ მეორეული სტრუქტურის სეგმენტებს.

ჩანაწერის ზოგადი ფორმატი ასეთია

პოზიციები	მონაცემის ტიპი	ველი	განსაზღვრება
1- 6	ჩანაწერის სახელი	"TURN "	
8-10	მთელი რიცხ.	რიგითი ნომ.	მოსახვევის რიგითი ნომერი
12-14	ასოთა რიგი (3)	მოსახვ.Id	მოსახვევის იდენტიფიკატორი
16-18	ამინომე. დასახელ.	საწყ.ამინომე.სახ.	მოსახვევის საწყისი ამინმეჯავას სახელი
20	ასო ნიშანი	ჯაჭვის ID	ჯაჭვის იდენტიფიკატორი
21-24	მთელი რიცხ.	საწყ.ამინომე.ნომ.	მოსახვევის საწყისი ამინმეჯავას ნომერი
25	ასო ნიშანი	ჩამ.კოდი	მოსახვევის საწყისი ამინმეჯავას ჩამატ. კოდი .
27-29	ამინომე. დასახელ.	ბოლოამინომე.სახ.	მოსახვევის ბოლო ამინმეჯავას სახელი
31	ასო ნიშანი	ჯაჭვის ID	ჯაჭვის იდენტიფიკატორი
32-35	მთელი რიცხ.	ბოლო. მინომე.ნომ.	მოსახვევის ბოლო ამინმეჯავას ნომერი
36	ასო ნიშანი	ჩამ.კოდი	მოსახვევის ბოლო ამინმეჯავას ჩამატ. კოდი .
41-70	ასოთა რიგი	კომენტარი	კომენტარები .

დეტალები

- ჩანაწერი Turns, როგორც წესი, აღწერს იმ ნაშთებს, რომლებიც ქმნიან β -მოსახვევს, ანუ როცა წყალბადური კავშირი ფორმირებულია i-ს (C-O)-სა და i+3-ის (N-H)-ს შორის. ასევე მოსახვევს რომელიც აკავშირებს i ნაშთს i+2-თან (gamma-bends), სხვა მოსახვევებზე შესაძლებელია შეგვხვდეს..

• საწყისი ნაშთი N-ბოლოსია.

მაგალითი

	1	2	3	4	5	6	7
123456789012345678901234567890123456789012345678901234567890							
TURN	1	S1A GLY A	16	GLN A	18	SURFACE	
TURN	2	FLA ILE A	50	GLY A	52	FLAP	
TURN	3	S2A ILE A	66	HIS A	69	SURFACE	
TURN	4	S1B GLY B	16	GLN B	18	SURFACE	
TURN	5	FLB ILE B	50	GLY B	52	FLAP	
TURN	6	S2B ILE B	66	HIS B	69	SURFACE	

ჰეტეროგენულ ნაშთთა (ჰეტეროგენთა) განყოფილება

ჰეტეროგენთა განყოფილება გვაძლევს მოცემულ ერთეულში (PDB ფაილში) მოხვედრილ არასტანდარტულ ნაშთთა სრულ აღწერას.

ჩანაწერი HET

ჩანაწერის მიმოხილვა

ჩანაწერი HET გამოიყენება ისეთი არასტანდარტული ნაშთების აღსაწერად, როგორც: აცა პრისტეტიკული ჯგუფები, ინჰიბიტორები, გამსსსლის მოლეკულები და ის იონები, რომელთათვისაც მოცემულია კოორდინატები. ნაშთი (ატომთა ჯგუფი) აღიწერება HET-ით, თუ იგი არ წარმოადგენს SEQRES -ში აღწერილი ბიოლოგიური პოლიმერის ნაწილს და წარმოადგენს პოლიმერთან დაკავშირებულ მოლეკულას. ანუ თუ იგი წარმოადგენს პოლიმერის შემადგენლობაში მყოფ ისეთ ქიმიურ ნერთს, რომელიც არ შედის შემდეგ ჩამონათვალში:

- სტანდარტული ამინომჟავები
- ნუკლეინის მჟავის შედამდგენლობაში მყოფი ფუძეები (C, G, A, T, U, და I)
- უცნობი ამინომჟავები ან ფუძეები, თუ მათი დასახელების საჩვენებლად გამოიყენება UNK

ჩანაწერი Het ასევე აღწერს იმ ჰეტეროგენებს, რომელთათვისაც ქიმიური იდენტიფიცირება არ მომხდარა და ამ შემთხვევაში ჯგუფის განსაზღვრა ხდება hetID UNL-ით (Unknown Ligand).

ჩანაწერის ზოგადი ფორმატი ასეთია

პოზიციები	მონაცემის ტიპი	ველი	განსაზღვრება
1 - 6	ჩანაწერის სახელი	"HET"	
8 - 10	ასოთა რიგი (3)	het ID	ჰეტეროს იდენტიფიკატორი
13	ასო ნიშანი	ჯაჭვის ID	ჯაჭვის იდენტიფიკატორი.
14 - 17	მთელი რიგბ.	რიგ-ნომერი	რიგითი ნომერი.
18	ასო ნიშანი	კოდი	ჩამატების კოდი.
21 - 25	მთელი რიგბ.	het.ატომთ.რიგბ.	HETATM რიგები
31 - 70	ასოთა რიგი	ტექსტი	Het ჯგუფის აღწერილი ტექსტი.

დეტალები

- თითოეული HET ჯგუფისათვის hetID განისაზღვრება არაუმეტეს სამი ასო თუ რიგებისაგან შემდგარი სიტყვით
- PDB ერთეულში (ფაილში) წყლის, დეიტერირებული წყლისა თუ მეთანოლისათვის არ ყალიბდება HET ჩანაწერი
- უცნობი ატომები ან იონები წარმოიდგინება როგორც UNX-ები, ქიმიური ფორმულით X1; უცნობი ლიგანდები - როგორც UNL; უცნობი ამინომჟავები კი - UNK-თი.

მაგალითი

	1	2	3	4	5	6	7
123456789012345678901234567890123456789012345678901234567890							
HET	TRS	975	8				
HET	STA	I	4	25	PART_OF:	HIV INHIBITOR;	
HET	FUC	Y	1	10	PART_OF:	NONOATE COMPLEX; L-FUCOSE	
HET	GAL	Y	2	11	PART_OF:	NONOATE COMPLEX	
HET	NAG	Y	3	15			
HET	FUC	Y	4	10			
HET	NON	Y	5	12			
HET	UNX	A	161	1	PSEUDO CARBON ATOM OF UNKNOWN LIGAND		
HET	UNX	A	162	1	PSEUDO CARBON ATOM OF UNKNOWN LIGAND		
HET	UNX	A	163	1	PSEUDO CARBON ATOM OF UNKNOWN LIGAND		

ჩანაწერი HETNAM

ჩანაწერის მიმოხილვა

ჩანაწერი გვაძლევს მოცემული hetID-ის შესაბამისი ნაერთის ქიმიურ დასახელებას.

ჩანაწერის ზოგადი ფორმატი ასეთია

პოზიციები	მონაცემის ტიპი	ველი	განსაზღვრება
1 - 6	ჩანაწერის სახელი	"HETNAM"	
9 - 10	გაგრძელება	გაგრძელება	მიანიწმენებს წინა სტრიქონის გაგრძელებაზე
12 - 14	ასოთა რიგი (3)	ჰეტID	ჰეტეროს იდენტიფიკატორი
16 - 70	ასოთა რიგი	ტექსტი	ქიმიური სახელი.

მაგალითი

	1	2	3	4	5	6	7
123456789012345678901234567890123456789012345678901234567890							
HETNAM	GLC	GLUCOSE					
HETNAM	SAD	BETA-METHYLENE SELENAZOLE-4-CARBOXAMIDE ADENINE					
HETNAM	2	SAD	DINUCLEOTIDE				
HETNAM	UNX	UNKNOWN ATOM OR ION					
HETNAM	UNL	UNKNOWN LIGAND					
HETNAM	CYE	45-(3-AMINOPROPYL)-5, 11, 22, 28, 34-PENTAMETHYL-3, 9, 15,					
HETNAM	2	CYE	20, 26, 32, 38, 43-OCTAOXO-2, 5, 8, 14, 19, 22, 25, 28, 31, 34, 37,				
HETNAM	3	CYE	42, 45, 48-TETRADECAAZA-11-AZONIAHEPTACYCLO[4.2.1.1-4,				
HETNAM	4	CYE	7-.1-10, 13-.1-21, 24-.1-27, 30-.1-33, 36-]DOPENTAONTA-				
HETNAM	5	CYE	1(46), 4(52), 6, 10(51), 12, 21(50), 23, 27(49), 29, 33(48), 35,				
HETNAM	6	CYE	44(47)-DODECAENE				

ჩანაწერი HETSYN

ჩანაწერის მიმოხილვა

ჩანაწერი უზრუნველგვყოფს HETNAM-ში აღწერილ ნაერთთა სინონიმებით, თუ ისინი არსებობენ.

ჩანაწერის ზოგადი ფორმატი ასეთია

პოზიციები	მონაცემის ტიპი	ველი	განსაზღვრება
1 - 6	ჩანაწერის სახელი	"HETSYN"	
9 - 10	გაგრძელება	გაგრძელება	მიანიშნებს წინა სტრიქონის გაგრძელებაზე
12 - 14	ასოთა რიგი (3)	het.ID	ჰეტეროს იდენტიფიკატორი.
16 - 70	ჩამონათვალი	hetსინონიმები	სინონიმების ჩამონათვალი

მაგალითი

	1	2	3	4	5	6	7
1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890
HETSYN	NAD	NICOTINAMIDE	ADENINE	DINUCLEOTIDE			
HETSYN	COA	COA					
HETSYN	CMP	CYCLIC AMP;	CYCLIC ADENOSINE	MONOPHOSPHATE			
HETSYN	TRS	TRIS BUFFER;	TRISAMINE;				
HETSYN	2	TRS TRIS (HYDROXYMETHYL)	AMINOMETHANE;	TRIMETHYLOL			
HETSYN	3	TRS AMINOMETHANE					

ჩანაწერი FORMUL

ჩანაწერის მიმოხილვა

ჩანაწერი FORMUL წარმოგიდგენს არასტანდარტული გვუფის ქიმიურ ფორმულასა და მუხტს.

ჩანაწერის ზოგადი ფორმატი ასეთია

პოზიციები	მონაცემის ტიპი	ველი	განსაზღვრება
1 - 6	ჩანაწერის სახელი	"FORMUL"	
9 - 10	მთელი რიცხ.	კომპონენტ.რიცხ.	კომპონენტთა რიცხვი.
13 - 15	ასოთა რიგი (3)	het.ID	ჰეტეროს იდენტიფიკატორი.
17 - 18	მთელი რიცხ.	გამეორება	გამეორებათა რიცხვი.
19	ასო ნიშანი	ვარსკვლავი	*** წყლის მოლეკულის შემთხვევაში.
20 - 70	ასოთა რიგი	ტექსტი	ქიმიური ფორმულა.

მაგალითი

	1	2	3	4	5	6	7
1234567890123456789012345678901234567890123456789012345678901234567890							
FORMUL	2	SO4	2(O4 S1 2-)				
FORMUL	3	GLC	C6 H12 O6				
FORMUL	3	FOL	2(C19 H17 N7 O6 2-)				
FORMUL	4	CL	2(CL1 1-)				
FORMUL	5	CA	CA1 2+				
FORMUL	1	ACE	C2 H3 O1				
FORMUL	2	ACE	C2 H3 O1				
FORMUL	8	HOH	*463 (H2 O1)				

კავშირების ანოტირების განყოფილება

კავშირების ანოტირების განყოფილება საშუალებას აძლევს დეპოზიტორს დააფიქსიროს დისულფიდური ბმების, თუ სხვა კავშირების არსებობა.

ჩანაწერი SSBOND

ჩანაწერის მიმოხილვა

ჩანაწერი SSBOND იდენტიფიცირებას უკეთებს ცილაში არსებულ თითოეულ დისულფიდურ კავშირს, აფიქსირებს რა თითოეულ ამ კავშირში მონაწილე ნაშთთა წყვილს.

ჩანაწერის ზოგადი ფორმატი ასეთია

პოზიციები	მონაცემის ტიპი	ველი	განსაზღვრება
1 - 6	ჩანაწერის სახელი	"SSBOND"	
8 - 10	მთელი რიცხ.	რიგ.ნომერი	რიგითი ნომერი.
12 - 14	ასოთა რიგი (3)	"CYS"	ამინმეავის სახელი.
16	ასო ნიშანი	ჯაჭვისID	ჯაჭვის იდენტიფიკატორი.
18 - 21	მთელი რიცხ.	ნომ.თანამიმდ1	ნაშთის ნომერი თანამიმდ..
22	ასო ნიშანი	1 კოდი	ჩამატების კოდი.
26 - 28	ასოთა რიგი (3)	"CYS"	ამინმეავის სახელი.
30	ასო ნიშანი	ჯაჭვისID2	ჯაჭვის იდენტიფიკატორი.
32 - 35	მთელი რიცხ.	ნომ.თანამიმდ2	ნაშთის ნომერი თანამიმდ.
36	ასო ნიშანი	1 კოდი2	ჩამატების კოდი.
60 - 65	სიმეტრ.ოპერატ.	sym1	სიმეტრიის ოპერატორი 1 ნაშთისათვის
67 - 72	სიმეტრ.ოპერატ.	sym2	სიმეტრიის ოპერატორი 2 ნაშთისათვის

მაგალითი

	1	2	3	4	5	6	7
123456789012345678901234567890123456789012345678901234567890123456789012							
SSBOND	1	CYS E 48	CYS E 51			2555	
SSBOND	2	CYS E 252	CYS E 285				
SSBOND	1	CYS A 250	CYS A 277			1555	1555
SSBOND	2	CYS B 250	CYS B 277			1555	1555

ჩანაწერი LINK

ჩანაწერის მიმოხილვა

ჩანაწერი LINK განსაზღვრავს იმ ნაშთებს შორის კავშირებს, რომლებიც არ გამომდინარეობს პირველადი სტურქტურის შესახებ ინფორმაციიდან. კავშირის გამოსახვისას მიეთითება ატომთა სახელები. ამ ჩანაწერში მოცემული ინფორმაცია წარმოადგენს დამატებითს ჩანაწერ CONECT-ში არსებული ინფორმაციისა.

ჩანაწერის ზოგადი ფორმატი ასეთია

პოზიციები	მონაცემის ტიპი	ველი	განსაზღვრება
1 - 6	ჩანაწერის სახელი	"LINK "	
13 - 16	ატომი	სახელი1	ატომის სახელი.
17	ასო ნიშანი	ალტერ. ლოკალ.1	ალტერნატიული ლოკალიზ. ინდიკატორი
18 - 20	ნაშთის სახელი	ნაშთ.სახ.1	ნაშთის სახელი.
22	ასო ნიშანი	ჯაჭვისID1	ჯაჭვის იდენტიფიკატორი
23 - 26	მთელი რიცხ.	თანამ.ნომერი1	ნაშთის ნომერი თანამიმდევრობაში.
27	ასო ნიშანი	კოდი1	ჩამატების კოდი.
43 - 46	ატომი	სახელი2	ატომის სახელი.
47	ასო ნიშანი	ალტერ. ლოკალ.2	ალტერნატიული ლოკალიზ. ინდიკატორი
48 - 50	ნაშთის სახელი	ნაშთ.სახ.2	ნაშთის სახელი.
52	ასო ნიშანი	ჯაჭვისID2	ჯაჭვის იდენტიფიკატორი
53 - 56	მთელი რიცხ.	თანამ.ნომერი2	ნაშთის ნომერი თანამიმდევრობაში.
57	ასო ნიშანი	კოდი2	ჩამატების კოდი.
60 - 65	სიმეტრ.ოპერატ.	sym1	სიმეტრიის ოპერატორი 1 ატომისთვის
67 - 72	სიმეტრ.ოპერატ.	sym2	სიმეტრიის ოპერატორი 2 ატომისთვის

მაგალითი

	1	2	3	4	5	6	7
123456789012345678901234567890123456789012345678901234567890123456789012							
LINK	O1	DDA 1		C3	DDL 2		
LINK	MN	MN 391		OE2	GLU 217		2565
LINK	C	LYS A 296		N	CME A 297	1555	1555
LINK	C	CME A 297		N	MET A 298	1555	1555
LINK	CA	CA A 997		O	GLN A 262	1555	1555
LINK	CA	CA A 997		O	TRP A 240	1555	1555
LINK	CA	CA A 997		O	HOH Z 169	1555	1555
LINK	CA	CA A 997		O	PRO A 249	1555	1555

ჩანაწერი CISPEP

ჩანაწერის მიმოხილვა

ჩანაწერი CISPEP განსაზღვრავს cis-კონფორმაციაში მყოფ პეპტიდურ ბმებს.

ჩანაწერის ზოგადი ფორმატი ასეთია

პოზიციები	მონაცემის ტიპი	ველი	განსაზღვრება
1 - 6	ჩანაწერის სახელი	"CISPEP"	
8 - 10	მთელი რიცხ.	რიგ.ნომერი	რიგითი ნომერი .
12 - 14	ასოთა რიგი (3)	პეპტიდ.დასაწყ.	ნაშთის სახელი .
16	ასო ნიშანი	ჯაჭვისID1	ჯაჭვის იდენტიფიკატორი
18 - 21	მთელი რიცხ.	თანამ.ნომერი1	ნაშთის ნომერი თანამიმდევრობაში .
22	ასო ნიშანი	1კოდი1	ჩამატების კოდი .
26 - 28	ასოთა რიგი (3)	პეპტიდ.ბოლო	ნაშთის სახელი .
30	ასო ნიშანი	ჯაჭვისID2	ჯაჭვის იდენტიფიკატორი
32 - 35	მთელი რიცხ.	თანამ.ნომერი2	ნაშთის ნომერი თანამიმდევრობაში .
36	ასო ნიშანი	1კოდი1	ჩამატების კოდი .
44 - 46	მთელი რიცხ.	მოდელის ნომ.	ახდენს სპეც. მოდელის იდენტიფიკაციას.
54 - 59	ნამდვილი რიცხ. (6.2)	სიდიდე	ორჩანაგა კუთხის სიდიდე (გრადუსებში)

მაგალითი

	1	2	3	4	5	6	7
1234567890123456789012345678901234567890123456789012345678901234567890							
CISPEP	1	GLY A 116	GLY A 117		0	18.50	
CISPEP	2	THR D 92	PRO D 93		0	359.80	

კრისტალოგრაფიული და კოორდინატთა ტრანსფორმირების განყოფილება.

კრისტალოგრაფიული განყოფილება აღწერს კრისტალოგრაფიული ექსპერიმენტის გეომეტრიულ მხარეს და იძლევა სხვადასხვა კოორდინატთა სისტემების მიმართ განსაზღვრული კოორდინატების ურთიერთტრანსფორმირების საშუალებებს.

ჩანაწერი CRYST1

ჩანაწერის მიმოხილვა

ჩანაწერი CRYST1 წარმოგვიდგენს კრისტალური უჯრედის პარამეტრებს, სივრცული ჯგუფის სახეს, და Z სიდიდეს. თუ სტრუქტურა არაკრისტალოგრაფიულადაა განსაზღვრული, CRYST1 უბრალოდ მიუთითებს ერთეულოვანი სივრცის წიბოს მქონე კუბზე.

ჩანაწერის ზოგადი ფორმატი ასეთია

პოზიციები მონაცემის ტიპი	ველი	განსაზღვრება	
1 - 6	ჩანაწერის სახელი	"CRYST1"	
7 - 15	ნამდვილი რიცხ. (6.2)	a	a (ანგსტემებში).
16 - 24	ნამდვილი რიცხ. (6.2)	b	b (ანგსტემებში).
25 - 33	ნამდვილი რიცხ. (6.2)	c	c (ანგსტემებში).
34 - 40	ნამდვილი რიცხ. (7.2)	alpha	alpha (გრადუსებში).
41 - 47	ნამდვილი რიცხ. (7.2)	beta	beta (გრადუსებში).
48 - 54	ნამდვილი რიცხ. (7.2)	gamma	gamma (გრადუსებში).
56 - 66	ასო-რიცხ. რიგი	სივრც.ჯგუფი	კრისტალოგრაფ. სივრცული ჯგუფი.
67 - 70	მთელი რიცხვი	z	Z სიდიდე.

დეტალები

- თუ PDB ერთეული აღწერს არაკრისტალოგრაფიულად განსაზღვრულ სტრუქტურას, CRYST1-ში აღწერილი სიდიდეები მიიღებს შემდეგ მნიშვნელობებს $a = b = c = 1.0$, $\alpha = \beta = \gamma = 90^\circ$, სივრცული ჯგუფი = P 1, და $Z = 1$.
- Hermann-Mauguin სივრცული ჯგუფის სიმბოლოები მოიცავს ფრჩხილების გარეშე. მაგალითად: P 43 21 2. უნდა აღინიშნოს, რომ სპირალური დერძის აღნიშვნა ხდება ორი ერთნიშნა რიცხვით
- გამოიყენება საერთაშორისო Hermann-Mauguin-ის სიმბოლოთა სრული ვერსიები. ასე მაგალითად , P 1 21 1 და არა P 21.
- Z სიდიდე გვიჩვენებს ერთეულოვან (ასიმეტრიულ) უჯრედში პოლიმერული ჯაჭვების რაოდენობას. ჰეტეროპოლიმერების შემთხვევაში Z გვიჩვენებს მხოლოდ ყველაზე მრავალრიცხოვან ჯაჭვთა რაოდენობას. მაგალითად, თუ საქმე გვაქვს A და B, სხვადასხვა თანამიმდევრობის მქონე ჯაჭვებთან, და სივრცული ჯგუფი აღიწერება P 2 -ით, რაც ნიშნავს მეორე რიგის სიმეტრიის დერძის არსებობას ერთეულოვან უჯრედში, Z სიდიდეთა სწორი მნიშვნელობების მაგალითები იხილეთ ქვემოთ მოცემულ ცხრილში:

ასიმეტრ. ერთეულის შემადგენლ. Z სიდიდე

A	2
AA	4
AB	2
AAB	4
AABB	4

მაგალითი

	1	2	3	4	5	6	7
1234567890123456789012345678901234567890123456789012345678901234567890							
CRYST1	52.000	58.600	61.900	90.00	90.00	90.00 P 21 21 21	8
CRYST1	1.000	1.000	1.000	90.00	90.00	90.00 P 1	1
CRYST1	42.544	69.085	50.950	90.00	95.55	90.00 P 1 21 1	2

ჩანაწერი ORIGXn

ჩანაწერის მიმოხილვა

ჩანაწერი ORIGXn (n = 1, 2, ან 3) აღწერს PDB ერთეულში (ფაილში) მოცემული ორთოგონალური კოორდინატების წარდგენის (არაორთოგონალურ) კოორდინატებში ტრანსფორმაციას.

ჩანაწერის ზოგადი ფორმატი ასეთია

პოზიციები	მონაცემის ტიპი	ველი	განსაზღვრება
1 - 6	ჩანაწერის სახელი	"ORIGXn"	n=1, 2, ან 3
11 - 20	ნამდვილი რიგბ. (10.6)	o [n] [1]	მობრ. მატრიცის პირველი სტრიქონი
21 - 30	ნამდვილი რიგბ. (10.6)	o [n] [2]	მობრ. მატრიცის მეორე სტრიქონი
31 - 40	ნამდვილი რიგბ. (10.6)	o [n] [3]	მობრ. მატრიცის მესამე სტრიქონი
46 - 55	ნამდვილი რიგბ. (10.5)	t [n]	ტრანსლაციის ვექტორის კომპონენტები

დეტალები

- PDB აფიქსირებს ამ ინფორმაციას იმ შემთხვევაშიც კი, როცა საქმე გვაქვს ერთეულოვან ტრანსფორმაციასთან (ანუ ერთეულოვან მატრიცასთან და ნულოვან ვექტორთან).
- თუ წარდგენისას გამოყენებულ სისტემაში კოორდინატებს აღვნიშნავთ X_{sub}, Y_{sub}, Z_{sub} და PDB ერთეულის შესაბამის ორთოგონალურ სისტემაში კოორდინატებს (გამოსახულს ანგსტრემებში) როგორც X, Y, Z, მაშინ:

$$\begin{aligned} X_{sub} &= O11X + O12Y + O13Z + T1 \\ Y_{sub} &= O21X + O22Y + O23Z + T2 \\ Z_{sub} &= O31X + O32Y + O33Z + T3 \end{aligned}$$

მაგალითი

	1	2	3	4	5	6	7
1234567890123456789012345678901234567890123456789012345678901234567890							
ORIGX1	0.963457	0.136613	0.230424		16.61000		
ORIGX2	-0.158977	0.983924	0.081383		13.72000		
ORIGX3	-0.215598	-0.115048	0.969683		37.65000		

ჩანაწერი SCALEn

ჩანაწერის მიმოხილვა

ჩანაწერი SCALEn (n = 1, 2, ან 3) აღწერს PDB ერთეულის ორთოგონალურ კოორდინატთა სისტემიდან ფრაქციონალურ კრისტალოგრაფულ კოორდინატთა სისტემაში გადასვლის მატრიცას.

ჩანაწერის ზოგადი ფორმატი ასეთია

პოზიციები მონაცემის ტიპი	ველი	განსაზღვრება
1 - 6	ჩანაწერის სახელი	"SCALEn" n=1, 2, ან 3
11 - 20	ნამდვილი რიცხ. (10.6)	s [n] [1] მოზრ. მატრიცის პირველი სტრიქონი
21 - 30	ნამდვილი რიცხ. (10.6)	s [n] [2] მოზრ. მატრიცის პირველი სტრიქონი
31 - 40	ნამდვილი რიცხ. (10.6)	s [n] [3] მოზრ. მატრიცის პირველი სტრიქონი
46 - 55	ნამდვილი რიცხ. (10.5)	u [n] ტრანსლაციის ვექტორის კომპონენტები

დეტალები

- სტანდარტული PDB ერთეულის შესაბამისი ორთოგონალურ კოორდინატთა სისტემა ერთეულოვანი უჯრედის (ჩანაწერ CRYST1-ში აღწერილი) ედრულ სისტემასთან დაკავშირებულია შემდეგნაირად:

თუ ვექტორები a, b, და c აღწერს (შეესაბამება) კრისტალოგრაფიული უჯრედის წიბოებს, და ვექტორები A, B და C წარმოადგენს ისეთ ორთოგონალურ კოორდინატთა სისტემის დერძთა ვექტორებს, რომლისთვისაც ვექტორ A-ს, B-ს და C-ს და ვექტორ a-ს, b-ს და c-ს აქვთ საერთო საწყისი; ვექტორი A პარალელურია ვექტორი a-სი, ვექტორი B პარალელურია ვექტორ A-სა და C-ს ნამრავლის, და ვექტორი C პარალელურია ვექტორი a-სა და b-ს ნამრავლის (ანუ ვექტორ c-სი)

- თუ ორთოგონალურ სისტემაში კოორდინატებია X, Y, Z, და ფრაქციონალური უჯრედული კოორდინატები კი xfrac, yfrac, zfrac, მაშინ:

$$\begin{aligned} xfrac &= S11X + S12Y + S13Z + U1 \\ yfrac &= S21X + S22Y + S23Z + U2 \\ zfrac &= S31X + S32Y + S33Z + U3 \end{aligned}$$

- NMR, ფიბრილათა დიფრაქციისა, თუ თეორიული მოდელირების საფუძველზე წარმოებული PDB ერთეულების შემთხვევაში SCALE მოიცემა ერთეულოვანი მატრიცით ანუ ტრანსფორმაცია არაა აქტუალური.

მაგალითი

	1	2	3	4	5	6	7
1234567890123456789012345678901234567890123456789012345678901234567890							
SCALE1	0.019231	0.000000	0.000000	0.000000	0.000000		
SCALE2	0.000000	0.017065	0.000000		0.000000		
SCALE3	0.000000	0.000000	0.016155		0.000000		

ჩანაწერი MTRIX

ჩანაწერის მიმოხილვა

ჩანაწერი MTRIX ($n = 1, 2, \text{ ან } 3$) აღწერს კრისტალოგრაფიული სიმეტრიის არაშესაბამის ტრანსფორმაციას.

ჩანაწერის ზოგადი ფორმატი ასეთია

პოზიციები მონაცემის ტიპი	ველი	განსაზღვრება
1 - 6	ჩანაწერის სახელი	"MTRIXn"
8 - 10	მთელი რიცხ.	რიგ.ნომერი
11 - 20	ნამდვილი რიცხ. (10.6)	m[n] [1]
21 - 30	ნამდვილი რიცხ. (10.6)	m[n] [2]
31 - 40	ნამდვილი რიცხ. (10.6)	m[n] [3]
46 - 55	ნამდვილი რიცხ. (10.5)	v[n]
60	მთელი რიცხ.	iGiven

$n=1, 2, \text{ ან } 3$
 რიგითი ნომერი.
 მობრ. მატრიცის პირველი სტრიქტონი
 მობრ. მატრიცის პირველი სტრიქტონი
 მობრ. მატრიცის პირველი სტრიქტონი
 ტრანსლაციის ვექტორის კომპონენტები
 უდრის 1-ს თუ ტრანსფორმაციის შესაბამისი კოორდინატები მოცემულია ფაილში, წინააღმდეგ შემთხვევაში ველი ცარიელია

დეტალები

- MTRIX ტრანსფორმაცია იყენებს PDB ერთეულში წარმოდგენილ კოორდინატებს და აწარმოებს მოლეკულის გვერდულ ტრანსფორმაციას იგივე კოორდინატთა სისტემაში. თუ მოცემული ტრანსფორმაციით მიღებული წარმოდგენის კოორდინატები ფიგურირებს ფაილში (PDB ერთეულში) iGiven ველში დგას 1. წინააღმდეგ შემთხვევაში ველი ცარიელია.

- ტრანსფორმაციის დამახასიათებელი შენიშვნის REMARK-ის არსებობა აუცილებელია

მაგალითი

	1	2	3	4	5	6	7
1234567890123456789012345678901234567890123456789012345678901234567890							
MTRIX1	1	-1.000000	0.000000	-0.000000	0.00001	1	
MTRIX2	1	-0.000000	1.000000	0.000000	0.00002	1	
MTRIX3	1	0.000000	-0.000000	-1.000000	0.00002	1	

ჩანაწერი TVECT

ჩანაწერის მიმოხილვა

ჩანაწერი TVECT განსაზღვრავს ტრანსლირების ვექტორს მრავალჯერ განმეორებადი კოვალენტურად დაკავშირებული სტრუქტურის კოორდინატების მისაღებად.

ჩანაწერის ზოგადი ფორმატი ასეთია

პოზიციები	მონაცემის ტიპი	ველი	განსაზღვრება
1 - 6	ჩანაწერის სახელი	"TVECT "	
8 - 10	მთელი რიცხ.	რიგ.ნომერი	რიგითი ნომერი.
11 - 20	ნამდვილი რიცხ. (10.5)	t [1]	ტრანსლაციის ვექტორის კომპონენტა
21 - 30	ნამდვილი რიცხ. (10.5)	t [2]	ტრანსლაციის ვექტორის კომპონენტა
31 - 40	ნამდვილი რიცხ. (10.5)	t [3]	ტრანსლაციის ვექტორის კომპონენტა.
41 - 70	ასოთა რიგი	ტექსტი	კომენტარი.

დეტალები

- მოლეკულის შემადგენლობაში არ მყოფი სტრუქტურის განსასაზღვრავად (მაგალითად მრავალჯერ განმეორებადი პოლისაქარიდული ჯაჭვის სტრუქტურის), PDB ერთეული მოიგავს უკანასკნელი სტრუქტურის ფრაგმენტს, რომლისა და TVECT-ში მოცემული ტრანსლაციის ვექტორის დახმარებით შესაძლებელია მთელი ჯაჭვის სტრუქტურის კოორდინატების მიღება.
- სტრუქტურის აღმწერი შენიშვნის, REMARK-ის არსებობა არაა აუცილებელი.

მაგალითი

	1	2	3	4	5	6	7
1234567890123456789012345678901234567890123456789012345678901234567890							
TVECT	1	0.00000	0.00000	28.30000			

კოორდინატების განყოფილება

კოორდინატების განყოფილება მოიცავს ატომურ კოორდინატებს და ასევე ჩანაწერებს MODEL და ENDMDL .

ჩანაწერი MODEL

ჩანაწერის მიმოხილვა

თუ PDB ერთეულში (ფაილში) წარმოდგენილია კოორდინატთა კომპლექტების ვარიანტები, ისე როგორც ამას ადგილი აქვს NMR-ის შემთხვევაში, ჩანაწერი MODEL მიგვითითებს მოდელის სერიულ ნომერზე.

ჩანაწერის ზოგადი ფორმატი ასეთია

პოზიციები	მონაცემის ტიპი	ველი	განსაზღვრება
1 - 6	ჩანაწერის სახელი	"MODEL "	
11 - 14	მთელი რიცხ.	რიგ.ნომერი	მოდელის რიგითი ნომერი .

დეტალები

- თუ მრავალი მოდელის არსებობის გამო ატომთა (რომელთა კოორდინატებიც უნდა მოიცეს) რიცხვმა გადააჭარბა 99,999 მაშინ უნდა შეიქმნას დამატებითი ერთეული (ფაილი)

დამოკიდებულება სხვა ჩანაწერებთან

თითოეული MODEL-ს გააჩნია შესაბამისი ENDMDL ჩანაწერი.

NMR მიღებულ PDB ერთეულის შემთხვევაში ჩანაწერი EXPDTA მოიცავს მოდელურ სტრუქტურათა რიცხვს.

მაგალითი

	1	2	3	4	5	6	7			
1234567890123456789012345678901234567890123456789012345678901234567										
ATOM	230	N	PRO	15	20.860	29.640	13.460	1.00	12.20	N
SIGATM	230	N	PRO	15	0.040	0.030	0.030	0.00	0.00	N
ATOM	231	CA	PRO	15	22.180	29.010	12.960	1.00	14.70	C
SIGATM	231	CA	PRO	15	0.060	0.040	0.050	0.00	0.00	C
ATOM	232	C	PRO	15	23.170	30.090	12.670	1.00	19.10	C
SIGATM	232	C	PRO	15	0.080	0.070	0.060	0.00	0.00	C
ATOM	233	O	PRO	15	24.360	29.860	12.670	1.00	17.50	O
SIGATM	233	O	PRO	15	0.040	0.030	0.030	0.00	0.00	O
ATOM	234	CB	PRO	15	21.710	28.220	11.640	1.00	17.70	C
SIGATM	234	CB	PRO	15	0.060	0.040	0.050	0.00	0.00	C
ATOM	235	CG	PRO	15	20.470	28.710	11.590	1.00	23.90	C
SIGATM	235	CG	PRO	15	0.080	0.060	0.060	0.00	0.00	C
ATOM	236	CD	PRO	15	19.640	29.320	12.660	1.00	15.50	C
SIGATM	236	CD	PRO	15	0.060	0.040	0.050	0.00	0.00	C
ATOM	237	HA	PRO	15	22.630	28.400	13.620	1.00	14.70	H
ATOM	238	1HB	PRO	15	22.240	28.540	10.860	1.00	17.70	H
ATOM	239	2HB	PRO	15	21.670	27.240	11.840	1.00	17.70	H
ATOM	240	1HG	PRO	15	20.360	29.240	10.740	1.00	23.90	H
ATOM	241	2HG	PRO	15	19.900	28.120	11.020	1.00	23.90	H
ATOM	242	1HD	PRO	15	19.230	30.160	12.320	1.00	15.50	H
ATOM	243	2HD	PRO	15	19.120	28.600	13.120	1.00	15.50	H

ჩანაწერი ANISOU

ჩანაწერის მიმოხილვა

ჩანაწერი ANISOU წარმოგვიდგენს ანიზოტროპულ ტემპერატურულ ფაქტორებს.

ჩანაწერის ზოგადი ფორმატი ასეთია

პოზიციები მონაცემის ტიპი	ველი	განსაზღვრება
1- 6	ჩანაწერის სახელი	"ANISOU"
7-11	მთელი რიცხ.	რიგ.ნომერი
13-16	ატომი	სახელი
17	ასო ნიშანი	ალტ.ლოკალიზ.
18-20	ნაშთის სახელი	ნაშთ.სახელი
22	ასო ნიშანი	ჯაჭვის ID
23-26	მთელი რიცხ.	თანამ.ნომერი
27	ასო ნიშანი	ქ კოდი
29-35	მთელი რიცხ.	u [0] [0]
36-42	მთელი რიცხ.	u [1] [1]
43-49	მთელი რიცხ.	u [2] [2]
50-56	მთელი რიცხ.	u [0] [1]
57-63	მთელი რიცხ.	u [0] [2]
64-70	მთელი რიცხ.	u [1] [2]
77-78	ასო.ნიშ. რიგი (2)	ქიმ.ელემენტი
79-80	ასო ნიშ. რიგი (2)	მუხტი

დეტალები

• ანიზოტროპულ ტემპერატურული ფაქტორები (სეკტები 29 - 70) წარმოდგენილია 10^{*4} -ზე გამრავლებული (ანგსტრემის შემთხვევაში 10^{*2}) სახით ანუ მთვლების სახით.

• ANISOU სიდიდეები მოიცემა დეკოზიტორის მიერ.

მაგალითი

	1	2	3	4	5	6	7	
1234567890123456789012345678901234567890123456789012345678901234567								
ATOM 107 N GLY	13		12.681	37.302	-25.211	1.000	15.56	N
ANISOU 107 N GLY	13	2406	1892	1614	198	519	-328	N
ATOM 108 CA GLY	13		11.982	37.996	-26.241	1.000	16.92	C
ANISOU 108 CA GLY	13	2748	2004	1679	-21	155	-419	C
ATOM 109 C GLY	13		11.678	39.447	-26.008	1.000	15.73	C
ANISOU 109 C GLY	13	2555	1955	1468	87	357	-109	C
ATOM 110 O GLY	13		11.444	40.201	-26.971	1.000	20.93	O
ANISOU 110 O GLY	13	3837	2505	1611	164	-121	189	O
ATOM 111 N ASN	14		11.608	39.863	-24.755	1.000	13.68	N
ANISOU 111 N ASN	14	2059	1674	1462	27	244	-96	N

ჩანაწერი SIGUIJ

ჩანაწერის მიმოხილვა

ჩანაწერი SIGUIJ გვიჩვენებს ანიზოტროპულ ტემპერატურულ ფაქტორთათვის სტანდარტულ გადახრებს. წარმოდგენისას მრავლდება 10^{*4} (ანგსტრემის შემთხვევაში 10^{*2}).

ჩანაწერის ზოგადი ფორმატი ასეთია

პოზიციები მონაცემის ტიპი	ველი	განსაზღვრება
1-6	ჩანაწერის სახელი	"SIGUIJ"
7-11	მთელი რიცხ.	რიგ.ნომერი
13-16	ატომი	სახელი
17	ასო ნიშანი	ალტ.ლოკალიზ.
18-20	ნაშთის სახელი	ნაშთის სახელი.
22	ასო ნიშანი	ჯაჭვის ID
23-26	მთელი რიცხ.	თანამ.ნომერი
27	ასო ნიშანი	i კოდი
29-35	მთელი რიცხ.	sig [1] [1]
36-42	მთელი რიცხ.	sig [2] [2]
43-49	მთელი რიცხ.	sig [3] [3]
50-56	მთელი რიცხ.	sig [1] [2]
57-63	მთელი რიცხ.	sig [1] [3]
64-70	მთელი რიცხ.	sig [2] [3]
77-78	ასო.ნიშ. რიგი (2)	ქიმ.ელემენტი
79-80	ასო ნიშ. რიგი (2)	მუხტი

მაგალითი

	1		2		3		4		5		6		7		
	12345678901	2345678901	2345678901	2345678901	2345678901	2345678901	2345678901	2345678901	2345678901	2345678901	2345678901	2345678901	2345678901	2345678901	
ATOM	107	N	GLY	13			12.681	37.302	-25.211	1.000	15.56				N
ANISOU	107	N	GLY	13			2406	1892	1614	198	519	-328			N
SIGUIJ	107	N	GLY	13			10	10	10	10	10	10			N
ATOM	108	CA	GLY	13			11.982	37.996	-26.241	1.000	16.92				C
ANISOU	108	CA	GLY	13			2748	2004	1679	-21	155	-419			C
SIGUIJ	108	CA	GLY	13			10	10	10	10	10	10			C
ATOM	109	C	GLY	13			11.678	39.447	-26.008	1.000	15.73				C
ANISOU	109	C	GLY	13			2555	1955	1468	87	357	-109			C
SIGUIJ	109	C	GLY	13			10	10	10	10	10	10			C
ATOM	110	O	GLY	13			11.444	40.201	-26.971	1.000	20.93				O
ANISOU	110	O	GLY	13			3837	2505	1611	164	-121	189			O
SIGUIJ	110	O	GLY	13			10	10	10	10	10	10			O
ATOM	111	N	ASN	14			11.608	39.863	-24.755	1.000	13.68				N
ANISOU	111	N	ASN	14			2059	1674	1462	27	244	-96			N
SIGUIJ	111	N	ASN	14			10	10	10	10	10	10			N

ჩანაწერი TER

ჩანაწერის მიმოხილვა

ჩანაწერი TER მიუთითებს ATOM/HETATM ჩანაწერთა ჩამონათვალის დამთავრებაზე.

ჩანაწერის ზოგადი ფორმატი ასეთია

პოზიციები	მონაცემის ტიპი	ველი	განსაზღვრება
1 - 6	ჩანაწერის სახელი	"TER"	
7 - 11	მთელი რიცხ.	რიგ.ნომერი	ატომის რიგითი ნომერი.
18 - 20	ნაშთის სახელი	ნაშთ.სახელი	ნაშთის სახელი.
22	ასო ნიშანი	ჯაჭვის ID	ჯაჭვის იდენტიფიკატორი.
23 - 26	მთელი რიცხ.	თანამ.ნომერი	ნაშთის ნომერი თანამიმდევრობაში.
27	ასო ნიშანი	i კოდი	ჩამატების კოდი.

დეტალები

• ცილების შემთხვევაში კიდურა ჟანგბადის ატომი წარმოიდგინება როგორც OXT, ხოლო ნუკლეინის მჟავების შემთხვევაში, როგორც O5T ან O3T.

• TER ჩანაწერში ამინმჟავური მახასიათებლები განისაზღვრება ბოლო ამინმჟავას მახასიათებლებით, ოღონდ ჯამური ნომერი იღებს ერთით მეტ მნიშვნელობას.

მაგალითი

	1	2	3	4	5	6	7		
123456789012345678901234567890123456789012345678901234567									
ATOM	4150	H	ALA A 431	8.674	16.036	12.858	1.00	0.00	H
TER	4151		ALA A 431						
ATOM	1403	O	PRO P 22	12.701	33.564	15.827	1.09	18.03	O
ATOM	1404	CB	PRO P 22	13.512	32.617	18.642	1.09	9.32	C
ATOM	1405	CG	PRO P 22	12.828	33.382	19.740	1.09	12.23	C
ATOM	1406	CD	PRO P 22	12.324	34.603	18.985	1.09	11.47	C
HETATM	1407	CA	BLE P 1	14.625	32.240	14.151	1.09	16.76	C
HETATM	1408	CB	BLE P 1	15.610	33.091	13.297	1.09	16.56	C
HETATM	1409	CG	BLE P 1	15.558	34.629	13.373	1.09	14.27	C
HETATM	1410	CD1	BLE P 1	16.601	35.208	12.440	1.09	14.75	C
HETATM	1411	CD2	BLE P 1	14.209	35.160	12.930	1.09	15.60	C
HETATM	1412	N	BLE P 1	14.777	32.703	15.531	1.09	14.79	N
HETATM	1413	B	BLE P 1	14.921	30.655	14.194	1.09	15.56	B
HETATM	1414	O1	BLE P 1	14.852	30.178	12.832	1.09	16.10	O
HETATM	1415	O2	BLE P 1	13.775	30.147	14.862	1.09	20.95	O
TER	1416		BLE P 1						

ჩანაწერი HETATM

ჩანაწერის მიმოხილვა

ჩანაწერი HETATM წარმოგიდგენს არასტანდარტული ჯგუფის ატომის კოორდინატებს. ეს ჩანაწერი გამოიყენება წყლის ატომებისათვის და ასევე იმ ატომებისათვის, რომლებიც შედის HET ჯგუფის შემადგენლობაში.

ჩანაწერის ზოგადი ფორმატი ასეთია

პოზიციები მონაცემის ტიპი	ველი	განსაზღვრება
1- 6	ჩანაწერის სახელი	"HETATM"
7-11	მთელი რიცხ.	რიგ.ნომერი
13-16	ატომი	ატომის რიგითი ნომერი
17	ასო ნიშანი	სახელი
18-20	ნაშთის სახელი	ალტ.ლოკალ.
22	ასო ნიშანი	ა.ლოკალიზ. ინდიკატორი
23-26	მთელი რიცხ.	ნაშთ.სახელი
27	ასო ნიშანი	ნაშთის სახელი
31-38	ნამდვილი რიცხ. (8.3)	ჯაჭვისID
39-46	ნამდვილი რიცხ. (8.3)	ჯაჭვისID
47-54	ნამდვილი რიცხ. (8.3)	ჯაჭვის იდენტიფიკატორი
55-60	ნამდვილი რიცხ. (6.2)	თანამ.ნომერი
61-66	ნამდვილი რიცხ. (6.2)	ნაშთის ნომერი
77-78	ასო ნიშ. რიგი (2)	თანამიმდევრობაში
79-80	ასო. ნიშ. რიგი (2)	ჩამატების კოდი
		i კოდი
		x
		y
		z
		occupancy
		ტემპ.ფაქტ.
		კომ.ელემენტო
		მუხტი
		X ორთოგონალური კოორდინატი (ანგსტრ.)
		X ორთოგონალური კოორდინატი (ანგსტრ.)
		X ორთოგონალური კოორდინატი (ანგსტრ.)
		Occupancy.
		ტემპერატურული ფაქტორი.
		ელემენტის სიმბოლო
		ატომის მუხტი.

მაგალითი

	1	2	3	4	5	6	7		
1234567890123456789012345678901234567890123456789012345678901234567890									
HETATM	1357	MG	MG	168	4.669	34.118	19.123	1.00 3.16	MG2+
HETATM	3835	FE	HEM	1	17.140	3.115	15.066	1.00 14.14	FE3+

ჩანაწერი ENDMDL

ჩანაწერის მიმოხილვა

ჩანაწერი ENDMDL იხმარება MODEL-თან ერთად და უზრუნველყოფს ინდივიდუალური სტრუქტურების ერთმანეთისგან გამოყოფას.

ჩანაწერის ზოგადი ფორმატი ასეთია

პოზიციები	მონაცემის ტიპი	ველი	განსაზღვრება
1 - 6	ჩანაწერის სახელი	"ENDMDL"	

დეტალები

- MODEL/ENDMDL იხმარება მხოლოდ მაშინ, თუ PDB ერთეულში ერთზე მეტი სტრუქტურაა წარმოდგენილი. ამას, როგორც წესი, ადგილი აქვს NMR PDB ერთეულში.
- ყველა მოდელი მულტიმოდელურ PDB ერთეულში უნდა წარმოადგენდეს ერთსა და იმავე სტრუქტურას

მაგალითი

	1	2	3	4	5	6	7	8	
1234567890123456789012345678901234567890123456789012345678901234567890									
...									
...									
ATOM	14550	1HG	GLU	122	-14.364	14.787	-14.258	1.00 0.00	H
ATOM	14551	2HG	GLU	122	-13.794	13.738	-12.961	1.00 0.00	H
TER	14552		GLU	122					
ENDMDL									
MODEL	9								
ATOM	14553	N	SER	1	-28.280	1.567	12.004	1.00 0.00	N
ATOM	14554	CA	SER	1	-27.749	0.392	11.256	1.00 0.00	C
...									
...									
ATOM	16369	1HG	GLU	122	-3.757	18.546	-8.439	1.00 0.00	H
ATOM	16370	2HG	GLU	122	-3.066	17.166	-7.584	1.00 0.00	H
TER	16371		GLU	122					
ENDMDL									
MODEL	10								
ATOM	16372	N	SER	1	-22.285	7.041	10.003	1.00 0.00	N
ATOM	16373	CA	SER	1	-23.026	6.872	8.720	1.00 0.00	C
...									
...									
ATOM	18188	1HG	GLU	122	-1.467	18.282	-17.144	1.00 0.00	H
ATOM	18189	2HG	GLU	122	-2.711	18.067	-15.913	1.00 0.00	H
TER	18190		GLU	122					
ENDMDL									

კავშირების განყოფილება

განყოფილება უზრუნველგვყოფს ქიმიური კავშირების შესახებ ინფორმაციით. ეს ინფორმაცია დამატებითაა LINK, HYDBND, SLTBRG, და CISPEP-ში მოცემული ინფორმაციისა.

ჩანაწერი CONECT

ჩანაწერის მიმოხილვა

ჩანაწერი CONECT განსაზღვრავს იმ ატომთა შორის კავშირებს, რომელთა კოორდინატებიც მოცემულია. კავშირების აღწერისას გამოიყენება ატომთა სერიული ნომრები. CONECT აუცილებელია HET ჯგუფისათვის (წყლის გამოკლებით) და ასევე ყველა სხვა კავშირისათვის, რომლებიც არაა განსაზღვრული სტანდარტულ ამინმჟავათა კავშირების ცხრილით.

ჩანაწერის ზოგადი ფორმატი ასეთია

პოზიციები	მონაცემის ტიპი	ველი	განსაზღვრება
1 - 6	ჩანაწერის სახელი	"CONNECT"	
7 - 11	მთელი რიცხვ.	რიგ.ნომერი	ატომის რიგითი ნომერი
12 - 16	მთელი რიცხვ.	რიგ.ნომერი	დაკავშირებული ატომის რიგითი ნომერი
17 - 21	მთელი რიცხვ.	რიგ.ნომერი	დაკავშირებული ატომის რიგითი ნომერი
22 - 26	მთელი რიცხვ.	რიგ.ნომერი	დაკავშირებული ატომის რიგითი ნომერი
27 - 31	მთელი რიცხვ.	რიგ.ნომერი	დაკავშირებული ატომის რიგითი ნომერი

მაგალითი

```
1 2 3 4 5 6 7
12345678901234567890123456789012345678901234567890
CONNECT 1179 746 1184 1195 1203
CONNECT 1179 1211 1222
CONNECT 1021 544 1017 1020 1022 1211 1222 1311
```

ფაილშიდა აღრიცხვის განყოფილება

ფაილშიდა აღრიცხვის განყოფილება უზრუნველგვყოფს ფაილის შესახებ ჯამური ინფორმაციით.

ჩანაწერი MASTER

ჩანაწერის მიმოხილვა

ჩანაწერი MASTER წარმოადგენს მაკონტროლებელ ჩანაწერს და იძლევა მოცემულ PDB ერთეულში (ფაილში) სტრიქონთა რიცხვს.

ჩანაწერის ზოგადი ფორმატი ასეთია

პოზიციები	მონაცემის ტიპი	ველი	განსაზღვრება
1 - 6	ჩანაწერის სახელი	"MASTER"	
11 - 15	მთელი რიცხ.	Remark რიცხ.	REMARK ჩანაწერთა რიცხვი
16 - 20	მთელი რიცხ.	"0"	
21 - 25	მთელი რიცხ.	Het რიცხ.	HET ჩანაწერთა რიცხვი
26 - 30	მთელი რიცხ.	Helix რიცხ.	HELIX ჩანაწერთა რიცხვი
31 - 35	მთელი რიცხ.	Sheet რიცხ.	SHEET ჩანაწერთა რიცხვი
36 - 40	მთელი რიცხ.	Turn რიცხ.	TURN ჩანაწერთა რიცხვი
41 - 45	მთელი რიცხ.	Site რიცხ.	SITE ჩანაწერთა რიცხვი
46 - 50	მთელი რიცხ.	ტრანსფ. რიცხ.	(ORIGX+SCALE+MTRIX) ჩანაწერთა რიცხვი
51 - 55	მთელი რიცხ.	კოორდ. რიცხ.	(ATOM+HETATM) ჩანაწერთა რიცხვი
56 - 60	მთელი რიცხ.	Ter რიცხ.	TER ჩანაწერთა რიცხვი
61 - 65	მთელი რიცხ.	Conect რიცხ.	CONNECT ჩანაწერთა რიცხვი
66 - 70	მთელი რიცხ.	SEQRES რიცხ.	SEQRES ჩანაწერთა რიცხვი

მაგალითი

	1	2	3	4	5	6	7
1234567890123456789012345678901234567890123456789012345678901234567890							
MASTER	40	0	0	0	0	6 2930	2 0 29

ჩანაწერი END

ჩანაწერის მიმოხილვა

ჩანაწერი END მიუთითებს PDB ფაილის დაბოლოებაზე.

ჩანაწერის ზოგადი ფორმატი ასეთია

პოზიციები	მონაცემის ტიპი	ველი	განსაზღვრება
1 - 6	ჩანაწერის სახელი	"END"	

დეტალები

- END ბოლო ჩანაწერია და გვხვდება ნებისმიერ PDB ერთეულში.

მაგალითი

	1	2	3	4	5	6	7
1234567890123456789012345678901234567890123456789012345678901234567890							
END							

(რომელიც ტირეთი აღინიშნება) პირველ თანამიმდევრობის გარკვეულ პოზიციასში ჩამატებამ უზრუნველყო გამოტოვებამდე და გამოტოვების შემდეგ განლაგებულ თანამიმდევრობათა ფრაგმენტების (ასოთა ქვერიგების) დამაკმაყოფილებელი ურთიერთსწორება.

საზოგადოდ საუკეთესო ურთიერთსწორების მოძიება არც ისე ადვილია. საჭირო გახდა ისეთი ეფექტური ალგორითმის შემუშავება, რომელიც შეძლებდა ორ თანამიმდევრობას შორის ურთიერთსწორების საუკეთესო ვარიანტის წარმოჩენას და ამას ოპტიმალური დანახარჯებით განახორციელებდა. სანამ უშუალოდ განვიხილავდეთ ამოცანას, განვმარტოთ, თუ რა იგულისხმება "საუკეთესო" ურთიერთსწორებაში. კერ განვიხილოთ მარტივი შემთხვევები.

დავიწყოთ იმით, თუ რას გულისხმობს ორ თანამიმდევრობას შორის ურთიერთსწორების ცნება. ზოგადად თანამიმდევრობებს აქვთ სხვადასხვა სიგრძე. როგორც ზემოთ მოყვანილი მაგალითიდანაც ჩანს, ურთიერთსწორება გულისხმობს თანამიმდევრობებში გამოტოვების გაჩენას. ამრიგად, ურთიერთსწორება შეიძლება წარმოვიდგინოთ როგორც გამოტოვებების ჩამატება თანამიმდევრობათა რაღაც პოზიციებში ისე, რომ მოხდეს თანამიმდევრობათა სიგრძეების გათანაბრება. თანამიმდევრობების სიგრძეების გათანაბრების შემდეგ შესაძლებელი გახდება მათი ერთმანეთის ქვეშ ისე განთავსება, რომ მოხდეს პირველი თანამიმდევრობის ასოებისა და გამოტოვებების მეორე თანამიმდევრობის ასოებთან და გამოტოვებებთან შესაბამისობაში მოყვანა. ამასთან, არ უნდა მოხდეს სხვადასხვა თანამიმდევრობებში ჩართული გამოტოვებების ერთმანეთის ქვეშ განთავსება. გამოტოვების ჩამატება შეიძლება როგორც თანამიმდევრობის თავში, ისე ბოლოში.

მოვახდენთ რა ორი თანამიმდევრობის ურთიერთსწორებას, ჩვენ შეგვიძლია დავახასიათოთ იგი გარკვეული ქულათა ჯამით. დახასიათებისას შესაძლებელია გამოყენებულ იქნეს ქულათა დარიცხვის, ვთქვათ, ასეთი სქემა: ურთიერთსწორების თითოეულ სვეტს (ერთმანეთის ქვეშ განთავსებულ ასოთა წყვილს) გარკვეული წვლილი შეაქვს ქულათა ჯამში და ამ წვლილს განსაზღვრავს სვეტის შემადგენლობა. ასე მაგალითად, თუ სვეტი მოიცავს ორ იდენტურ ასოს, სვეტზე დასარიცხი ქულა შეიძლება მივიჩნიოთ +1-ის ტოლად (*თანხვედრა*). სვეტში სხვადასხვა ასოების განლაგება კი დავაჯარიმოთ -1-ით (*არათანხვედრა*). და ბოლოს გამოტოვების გამოჩენა სვეტში დავაჯარიმოთ -2-ით; ქულათა სრული ჯამი მიიღება თითოეული სვეტით განსაზღვრულ სიდიდეთა აჯამების შედეგად. საუკეთესო ურთიერთსწორება შეესაბამება მაქსიმალურ ქულათა ჯამს. მაქსიმალური ქულათა ჯამი, რომელსაც აღნიშნავენ $\text{sim}(s,t)$ -თი (s და t თანამიმდევრობებია), ამასთანავე ახასიათებს ორ თანამიმდევრობას შორის მსგავსების ბარისხს. ზოგადად შესაძლებელია მაქსიმალურ ქულიანი მრავალი ურთიერთსწორების არსებობა.

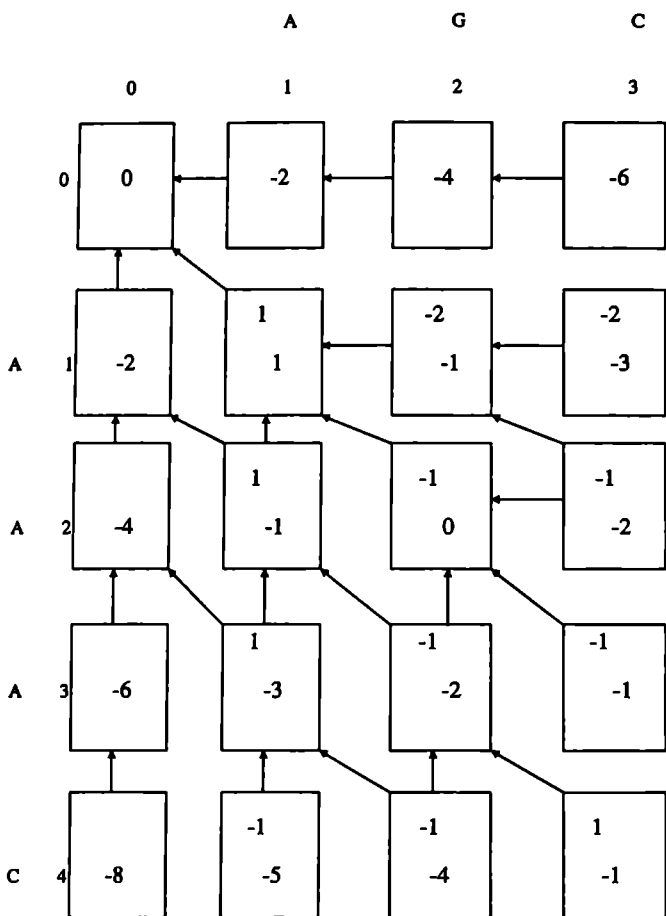
ზემოთ აღწერილ ქულათა დარიცხვის სქემის საილუსტრაციოდ გამოვთვალოთ (10.1) ურთიერთსწორების ქულათა ჯამი. ამ შემთხვევაში საქმე გვაქვს ერთნაირ ასოიან ცხრა სვეტთან, სხვადასხვაასოიან ერთ სვეტთან და გამოტოვების შემცველ ერთ სვეტთან. ამიტომ ქულათა ჯამი იქნება შემდეგი: $9 \times 1 + 1 \times (-1) + 1 \times (-2) = 6$.

რეალურ ამოცანებში შესაძლებელია გამოყენებულ იქნეს ქულათა დარიცხვის უფრო რთული სისტემა, თუმცა დარიცხვის პრინციპი უცვლელი რჩება - ხდება თანხვედრათა წახალისება (დაჯილდოება) და არათანხვედრათა და გამოტოვებათა დაჯარიმება.

ორ თანამიმდევრობას შორის მსგავსების შეფასების პირდაპირი, მარტივი მეთოდი გულისხმობს ყველა შესაძლო ურთიერთსწორების განხილვას და მათ შორის მაქსიმალურქულაიანის ამორჩევას. მაგრამ ორ თანამიმდევრობას შორის შესაძლო ურთიერთსწორებათა რიცხვი მათი სიგრძეების ექსპონენციალური ფუნქციაა (საკმაოდ დიდი რიცხვია), და აქედან გამომდინარე ასეთი მიდგომა მიგვიყვანს ძალზედ ნელა მომუშავე ალგორითმამდე. საბედნიეროდ მოიძებნა გაცილებით სწრაფი ალგორითმი.

სწრაფი ალგორითმი ემყარება *დინამიკური პროგრამირების* სახელით ცნობილ მიდგომას. მიდგომა გულისხმობს გლობალური ამოცანის მზარდი დონის, იგივე სახის ქვეამოცანებად დანაწევრებას და ქვედა დონის ქვეამოცანების გადაწყვეტისას მიღებული შედეგების გამოყენებას შემდგომი (ზედა) დონის ქვეამოცანის გადასაწყვეტად. ვთქვათ შესადარებელია ორი s და t თანამიმდევრობა. იმის მაგივრად, რომ მოვახდინოთ მთლიან თანამიმდევრობათა შედარება, ვახდენთ თანამიმდევრობათა თანდათანობით, სულ უფრო მზარდი სიგრძის *პრეფიქსების* მსგავსების შეფასებას და შეფასების მოცემულ ეტაპზე მოპოვებული ინფორმაციის შემდგომ ეტაპზე გამოყენებას. ამოცანის გადაწყვეტა იწყება მოკლე პრეფიქსების განხილვით და მიღებული ინფორმაციის გათვალისწინებით წარმოებს ანალოგიური ამოცანების გადაწყვეტა სულ უფრო მზარდი სიგრძის პრეფიქსებისათვის.

ვთქვათ s -ის სიგრძეა m , ხოლო t -სი კი n . s -სთვის შეიძლება შეირჩეს $m+1$ შესაძლო პრეფიქსი, t -თვის კი $n+1$, ცხადია ცარიელი (ნულოვანი - გამოტოვებათა შესაბამისი) პრეფიქსის გათვალისწინებით. ჩვენი მიზანია ვაწარმოოთ ისეთი $(m+1) \times (n+1)$ განზომილებიანი მატრიცა, რომლის თითოეული (i, j) ელემენტი ახასიათებს $s[1..i]$ -სა და $t[1..j]$ -ს შორის ოპტიმალურ მსგავსებას (გვაძლევს საუკეთესო ქულათა ჯამს). 10.1 სურათზე წარმოდგენილია მატრიცა $s=AAAC$ და $t=AGC$ თანამიმდევრობების ურთიერთსწორებისათვის. პრეფიქსების ადვილად შედარების მიზნით s განვალაგოთ კიდურა მარცხენა სვეტის სახით, ხოლო t ზედა პირველი სტრიქონის სახით. შევნიშნოთ, რომ პირველი სტრიქონი და პირველი სვეტი აღწერს არანულოვანი პრეფიქსების ნულოვანთან შედარების შედეგებს და აქედან გამომდინარე, შედარების დამახასიათებელი ქულათა ჯამი



სურათი 10.1. ოპტიმალური ურთიერთსწორებისათვის წარმოებული ორგანოზომი-
 ლეზიანი მატრიცა $a(i,j)$. (i,j) უჯრის ზემო მარცხენა კუთხეში განლაგებული
 სიდიდეები მიუთითებენ იმაზე უდრის თუ არა $a[i] \leq a[j]$ -ს. სვეტთა და სტრიქონთა
 შესაბამისი ინდექსების ათვლა წარმოებს 0-დან.

გამოითვლება ფორმულით $-2k$, სადაც k არანულოვანი თანამიმდევრობის (პრეფიქსის) სიგრძეა. ასე რომ, პირველი სტრიქონისა და სვეტის შევსება დიდ სიძნელეს არ წარმოადგენს.

ახლა ვნახოთ, თუ როგორ შეიძლება სხვა სტრიქონთა და სვეტთა შევსება. სტრატეგიის არსი იმაში მდგომარეობს, რომ ჩვენ შეგვიძლია შევაფასოთ (i, j) ელემენტის სიდიდე წინა სამი $(i-1, j)$, $(i-1, j-1)$ და $(i, j-1)$ ელემენტის მნიშვნელობიდან გამომდინარე. ამას საფუძვლად უდევს ის, რომ $s[1..i]$ -ისა და $t[1..j]$ პრეფიქსებს შორის ურთიერთსწორების მიღწევის სამი გზა არსებობს. ეს გზები განისაზღვრება სწორედ ზემოთხსენებული წინა სამი ელემენტის მნიშვნელობებზე დაყრდნობით. კერძოდ, $s[1..i]$ -ისა და $t[1..j]$ შორის ურთიერთსწორებისას გვაქვს სამი არჩევანი:

- გამოვიყენოთ $s[1..i]$ -ის $t[1..j-1]$ -თან ურთიერთსწორება და შევუთავსოთ გამოტოვება $t[j]$ -ს
- გამოვიყენოთ $s[1..i-1]$ -ის $t[1..j-1]$ -თან ურთიერთსწორება და შევუთავსოთ $s[i]$ $t[j]$ -ს
- გამოვიყენოთ $s[1..i-1]$ -ის $t[1..j]$ -თან ურთიერთსწორება და შევუთავსოთ $s[i]$ გამოტოვებას.

გამომდინარე იქიდან, რომ დაუშვებელია გამოტოვებათა ურთიერთშესაბამისობაში მოყვანა (ერთსა და იმავე სვეტში განთავსება), სხვა არჩევანი არ არსებობს. მცირე პრეფიქსებს შორის საუკეთესო ურთიერთსწორების დამახასიათებელი ქულათა ჯამები ინახება მატრიცის ელემენტების სახით, რათა განხორციელდეს მატრიცის თანამიმდევრული შევსება. ამრიგად, მსგავსებისა და შესაბამისი მატრიცის მისაღებად შეიძლება ვისარგებლოთ შემდეგი ფორმულით:

$$\begin{aligned} & \text{sim} (s[1..i], t[1..j-1]) - 2 \\ \text{sim} (s[1..i], t[1..j]) = \max \{ & \text{sim} (s[1..i-1], t[1..j-1]) + p(i,j) \\ & \text{sim} (s[1..i-1], t[1..j]) - 2 \end{aligned} \quad (10.2)$$

სადაც, თუ $s[i]=t[j]$, მაშინ $p(i,j) = 1$ და თუ $s[i] \neq t[j]$, მაშინ $p(i,j) = -1$ -ს. 10.1 სურათზე $p(i,j)$ -ს სიდიდეთა მნიშვნელობები მოცემულია (i,j) ელემენტთა შესაბამისი უჯრედების ზედა მარცხენა კუთხეებში. ამრიგად, უკანასკნელი ტოლობის გათვალისწინებით ჩამოყალიბებული a მატრიცის ელემენტები ასე შეიძლება გამოითვალოს

$$\begin{aligned} & a[i, j-1] - 2 \\ a(i,j) = \max \{ & a[i-1, j-1] + p(i,j) \\ & a[i-1, j] - 2 \end{aligned} \quad (10.3)$$

ამრიგად, გვაქვს გამოთვლების თანამიმდევრულად განხორციელების შესაძლებლობა და ძალზე ადვილია მატრიცის სტრიქონების მარცხნიდან მარჯვნივ და ასევე სვეტების ზემოდან ქვემოთ თანამიმდევრულად შევსება. ანუ $a[i,j]$ -ს გამოთვლა სიძნელეს არ წარმოადგენს, თუ უკვე შევსებული გვაქვს $a[i, j-1]$, $a[i-1, j-1]$ და $a[i-1, j]$ ელემენტები.

და ბოლოს, შევნიშნოთ, რომ 10.1 სურათზე ისრები მიუთითებს იმაზე, თუ რომელი ელემენტი ან ელემენტები (ასეთი შეიძლება ერთზე მეტი იყოს) განაპირობებენ მოცემული ელემენტის მნიშვნელობას ანუ მაქსიმალურს (10.3)-ით განსაზღვრულ შესაძლო სიდიდებს შორის. ასე მაგალითად, $a[1,2]$ ელემენტის მნიშვნელობა განისაზღვრა, როგორც მაქსიმალური შემდეგ სიდიდეებს შორის

$$a[1, 1] - 2 = -1$$

$$a[0, 1] - 1 = -3$$

$$a[0, 2] - 2 = -6$$

ამ შემთხვევაში მაქსიმუმის მიღწევის მხოლოდ ერთი გზა არსებობს (1,1) ელემენტის საშუალებით, და სწორედ ამას გვიჩვენებს ისარი.

10.2 სურათზე მოცემულია ალგორითმი, რომლის მიხედვითაც ხდება მსგავსების მატრიცის შევსება. იგი სტრიქონ-სტრიქონ ადგენს ელემენტთა მნიშვნელობებს. და ეს მნიშვნელობები დამოკიდებულია გამოტოვების დამაჯარიმებელ g -ს სიდიდეზე (როგორც წესი $g < 0$) და ასოთა წყვილებისათვის გამოყენებულ ქულათა დარიცხვის p ფუნქციაზე. 10.1 სურათზე გამოსახული შემთხვევისათვის $g = -2$; $p(a,b) = 1$ თუ $a = b$ და $p(a,b) = -1$ თუ $a \neq b$.

ალგორითმი	მსგავსება
შემავალი სიდიდეები:	s და t თანამიმდევრობანი
გამოსავალი:	s და t - შორის მსგავსება
$m \leftarrow s $	
$n \leftarrow t $	
for $i \leftarrow 0$ to m do	
$a[i,0] \leftarrow i \times g$	
for $j \leftarrow 0$ to n do	
$a[0,j] \leftarrow j \times g$	
for $i \leftarrow 1$ to m do	
for $j \leftarrow 1$ to n do	
$a[i,j] \leftarrow \max (a[i, j-1] + g,$	
$a[i-1, j-1] + p(i,j),$	
$a[i-1, j] + g)$	
return $a[m,n]$	

სურათი 10.2. ორი თანამიმდევრობის შესადარებლად გამოყენებული დინამიკური პროგრამირების ბაზისური ალგორითმი

ოპტიმალური ურთიერთსწორება

ჩვენ განვიხილეთ თუ როგორ შეიძლება ორ თანამიმდევრობას შორის მსგავსების შეფასება. ახლა შევეცდებით წარმოვადგინოთ ოპტიმალური ურთიერთსწორების კონსტრუირების პროცედურა. ამ საქმეში დაგვეხმარება 10.1 სურათზე მოყვანილი ისრები. ერთადერთი მოთხოვნა მდგომარეობს იმაში, რომ $a(m,n)$ ელემენტიდან მივყვეთ ისრებს, ვიდრე არ მივალწვეთ $a(0,0)$ ელემენტს. თითოეული ისარი ურთიერთსწორების გზაზე გადადგმული გარკვეული ნაბიჯია. მაგალითისათვის განვიხილოთ $a(i,j)$ ელემენტიდან გამოსული ისარი. თუ პორიზონტალურია, მაშინ იგი შეესაბამება ურთიერთსწორებულ წყვილს (სვეტს), რომელშიც i/j ასო შესაბამისობაშია გამოტოვებათან a -ში; თუ იგი ვერტიკალურია, მაშინ ერთმანეთის ქვეშ დგება s/i ასო და გამოტოვება t -ში; და ბოლოს, დიაგონალური ისარი მიუთითებს ურთიერთსწორებულ წყვილზე (სვეტზე), რომელშიც ურთიერთსწორებულია s/i და t/j ასოები. როგორც წესი, პირველ თანამიმდევრობას (s) განათავსებენ ვერტიკალურად, სვეტის გასწვრივ. ამრიგად, ოპტიმალური ურთიერთსწორება შესაძლებელია ადვილად კონსტრუირდეს მარჯვნიდან მარცხნივ, თუ ბაზისური ალგორითმის გამოყენებით გამზადებულია a მსგავსების მატრიცა. აქვე უნდა აღინიშნოს, რომ ისრის არჩევისას გაითვალისწინება გარკვეული პირობების შესრულება, რომელთა შესახებაც ვისაუბრებთ ქვემოთ.

10.3 სურათზე წარმოდგენილია ოპტიმალური ურთიერთსწორების რეკურსიული (recursive) ალგორითმი მოცემული s და t თანამიმდევრობებისა და მათი მსგავსების აღმწერი a მატრიცისათვის. *Align* (m,n,len) პროცედურით წარმოებს ოპტიმალური ურთიერთსწორების კონსტრუირება. შედეგი მოიცემა *align-s* და *align-t* ვექტორთა წყვილით, რომლებიც $1... len$ პოზიციებში განალაგებენ შესაბამის ასო-ნიშნებს, ანუ გამოტოვებებსა და თანამიმდევრობათა შემქმნელ სიმბოლოებს. ურთიერთსწორების სიგრძე ასევე განისაზღვრება ალგორითმის საშუალებით და მოიცემა len პარამეტრით. შევნიშნოთ, რომ $max(|s|,|t|) \leq len \leq m+n$.

თანამიმდევრობათა მოცემული წყვილისათვის შესაძლებელია მოძებნილ იქნეს ოპტიმალური ურთიერთსწორების რამდენიმე ვარიანტი. 10.3 სურათზე გამოსახული ალგორითმის რეალზების საფუძველზე შესაძლებელია მიღებულ იქნეს ერთ-ერთი მათგანი, ამასთან (i,j) ელემენტიდან გამოსულ მიმართულებათა შორის უპირატესების განსაზღვრა ამ შემთხვევაში წარმოებს საათის ისრის მიმართულების შესაბამისად (იხ. სურათი 10.4).

ალგორითმი ურთიერთსწორება
 შემაჯავალი სიდიდეები: მსგავსების ალგორითმით მიღებული a
 მატრიცის i, j ელემენტები
 გამოსავალი: ურთიერთსწორება *align-s*, *align-t*-სა და *len*
 სიგრძის სახით.

```

if  $i=0$  and  $j=0$  then
   $len \leftarrow 0$ 
else if  $i > 0$  and  $a[i,j] = a[i-1,j] + g$  then
  Allign ( $i-1, j, len$ )
   $len \leftarrow len + 1$ 
  align-s [ $len$ ]  $\leftarrow s[i]$ 
  align-t [ $len$ ]  $\leftarrow -$ 

else if  $i > 0$  and  $j > 0$  and  $a[i,j] = a[i-1,j-1] + p(i,j)$  then
  Allign ( $i-1, j-1, len$ )
   $len \leftarrow len + 1$ 
  align-s [ $len$ ]  $\leftarrow s[i]$ 
  align-t [ $len$ ]  $\leftarrow t[j]$ 
else // ამ შემთხვევაში  $j > 0$  and  $a[i,j] = a[i,j-1] + g$ 
  Allign ( $i, j-1, len$ )
   $len \leftarrow len + 1$ 
  align-s [ $len$ ]  $\leftarrow -$ 
  align-t [ $len$ ]  $\leftarrow t[j]$ 

```

სურათი 10.3. ოპტიმალური ურთიერთსწორების რეკურსიური ალგორითმი

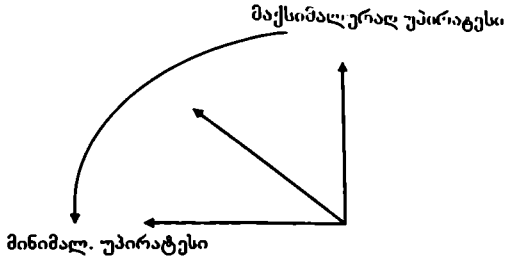
ყოველივე აღნიშნულიდან გამომდინარე, მოცემული ალგორითმით მიღებული ურთიერთსწორებისათვის შეიძლება გამოვეყოთ შემდეგი თავისებურებანი: არჩევნის არსებობის შემთხვევაში გამოტოვებას t -ში უპირატესობა აქვს ორი ასოს დაწყვილებასთან და ეს უკანასკნელი უპირატესობით სარგებლობს გამოტოვებასთან s -ში. ასე მაგალითად, თუ მოვახდენთ $s=ATAT$ -სა და $t=TATA$ -ს ურთიერთსწორებას მივიღებთ

-ATAT
 TATA -

და არა

ATAT -
 -TATA

რომელიც ასევე ამ ორი თანამიმდევრობის ოპტიმალურ ურთიერთსწორებას წარმოადგენს.



სურათი 10.4. ისართა უპირატესობის განმსაზღვრელი მიმართულება

თუ მოვახდენთ $\neq AA$ და $\neq AAAA$ თანამიმდევრობათა ურთიერთსწორებას, მივიღებთ მხოლოდ

$$\begin{aligned} &-- AA \\ &AAAA \end{aligned}$$

მიუხედავად იმისა, რომ არსებობს კიდევ ოთხი სხვა ოპტიმალური ურთიერთსწორება. მოყვანილ ურთიერთსწორებას ხშირად მოიხსენიებენ, როგორც *ზემოუპირატესს*, გამომდინარე იქიდან, რომ უპირატესად გამოიყენება მატრიცის ზემოთ მიმართული ისრები. უპირატესობის შეცვლა ადვილი განსახორციელებელია და მიიღწევა *if* ოპერატორების თანამიმდევრობის შეცვლით ოპერატორთა რიგში. ამრიგად ადვილი მისაღებია *ქვემოუპირატესი* ურთიერთსწორებაც. ორ თანამიმდევრობას შორის ყველა შესაძლო ოპტიმალურ ურთიერთსწორებათა რიცხვი მოიცავს, როგორც *ზემოუპირატესს*, ისე *ქვემოუპირატესს* ურთიერთსწორებებს.

ცხადია, შესაძლებელია ალგორითმის ისე ორგანიზება, რომ მან მოგვეცეს \neq -სა და \neq -ს შორის ყველა შესაძლებელი ოპტიმალური ურთიერთსწორება. ამისათვის საჭიროა იმ წერტილთა სიმრავლის დამახსოვრება, რომელთათვისაც არსებობს არჩევანი და გარკვეული ნაბიჯის გავლის შემდეგ მათ უნდა დავუბრუნდეთ, რათა გამოვავლინოთ ისრების გასწვრივ $(0,0)$ ელემენტის მიღწევის ყველა შესაძლო გზა. შეიძლება არსებობდეს ოპტიმალურ ურთიერთსწორებათა ძალზე დიდი რიცხვი. ასეთ შემთხვევებში, მიზანშეწონილია წარმოვადგინოთ მათი მხოლოდ მოკლე დახასიათება, ან სულაც ამ ურთიერთსწორებათა მხოლოდ ნაწილი, მაგალითად *ზემოუპირატესი* ან *ქვემოუპირატესი* ვარიანტები.

ახლა შევაფასოთ აღწერილი ალგორითმის სირთულე. ბაზისური ალგორითმი (იხ. სურ. 10.2) მოიცავს ოთხ ციკლს. მათგან პირველი ორი გულისხმობს მინიჭების ოპერაციებს და მოითხოვს შესაბამისად $O(m)$ და $O(n)$ დროს. ბოლო ორი ციკლით ხდება მატრიცის ელემენტების შევსება. ოპერაციათა რიცხვი ამ შემთხვევაში ძირითადად დამოკიდებულია თანამიმდევრობათა სიგრძეზე და ე.ი. მატრიცის განზომილებაზე. ამაზე დაიხარჯება $O(mn)$ ანუ mn -ის პროპორციული დრო და იგი ძირითადი ნაწილია ჯამური დროის. საჭირო მეხსიერება ასევე პროპორციულია მატრიცის ზომისა. ამრიგად, ალგორითმის სირთულე როგორც დროის, ისე მეხსიერების თვალსაზრისით შეიძლება დახასიათდეს $O(mn)$ დამოკიდებულებით. როცა თანამიმდევრობებს ერთნაირი ან თითქმის ერთნაირი სიგრძე (მაგ. n) აქვთ, ვიღებთ $O(n^2)$ დამოკიდებულებას. ამიტომაც ეს ალგორითმი ითვლება კვადრატული სირთულის ალგორითმად.

როცა მატრიცა უკვე შევსებულია, ურთიერთსწორების კონსტრუირება მოითხოვს $O(len)$ ანუ len -ის პროპორციულ დროს, სადაც len ურთიერთსწორების სიგრძეა, ანუ მაქსიმუმ $(m+n)$ -ის პროპორციულ დროს. ამიტომ აქ უნდა აღინიშნოს, რომ მოსახმარი სამანქანო დროისა და საჭირო მეხსიერების მოცულობიდან გამომდინარე, დინამიკურ პროგრამირებაზე დაფუძნებული ორი თანამიმდევრობის ოპტიმალური ურთიერთსწორების ალგორითმი, ვერ უზრუნველყოფს დიდი მოცულობის მონაცემთა ბაზებში საკვლევთან მსგავსებული თანამიმდევრობების გამოვლენის ამოცანის რეალურ სამანქანო დროში გადწყვეტას. შესაბამისად, დიდი მოცულობის თანამიმდევრობათა შესახებ მონაცემთა ბაზების ანალიზისათვის შემუშავებულ იქნა თანამიმდევრობათა მსგავსების შემფასებელი სპეციალური, მართალია ნაკლებად ზუსტი, მაგრამ უფრო სწრაფი მეთოდები, რომელთა შესახებაც ვისაუბრებთ შემდეგში. მომდევნო პარაგრაფში კი განვიხილავთ დინამიკური პროგრამირების ბაზური ალგორითმის ვარიაციებს.

11. ორი თანამიმდევრობის შედარება. ნახევრადგლობალური და ლოკალური ურთიერთსწორება

ნახევრადგლობალური ურთიერთსწორება

ნახევრადგლობალური ურთიერთსწორების შემთხვევაში ქულების დარიცხვისას მხედველობაში არ მიიღება თანამიმდევრობათა კიდურა გამოტოვებები. დინამიკური პროგრამირების ბაზისური ალგორითმის ერთ-ერთი საინტერესო თვისება მდგომარეობს იმაში, რომ მატრიცის ფორმირებისას განხორციელებული უმნიშვნელო მოდიფიკაციების საშუალებით შესაძლებელია ვაწარმოოთ კიდურა გამოტოვებების დაჯარიმების კონტროლი.

დავიწყოთ “კიდურა გამოტოვებების” განმარტებით და ვნახოთ გარკვეულ პირობებში რატომაა მიზანშეწონილი მათი იგნორირება. კიდურა გამოტოვება შეესაბამება თანამიმდევრობის პირველ ასომდე ან ბოლო ასოს შემდგომ განლაგებულ უბანს. ასე მაგალითად, ქვემოთ, მაგალითის სახით მოცემულ ურთიერთსწორების შემთხვევაში მეორე თანამიმდევრობის ყველა გამოტოვება კიდურაა, მაშინ, როცა პირველი თანამიმდევრობაში არსებული ერთადერთი გამოტოვება არაკიდურაა.

CAGCA - CTTGGATTCTCGG

(11.1)

--- CAGCGTGG -----

შევნიშნოთ, რომ ამ თანამიმდევრობათა სიგრძეები მნიშვნელოვნად განსხვავდება ერთმანეთისაგან. ერთის სიგრძეა 8, ხოლო მეორე მოიცავს 18 ასოს. როცა სიგრძეები მნიშვნელოვნად განსხვავდება ერთმანეთისაგან და რიგ სხვა შემთხვევებშიც, ურთიერთსწორებისას შეიძლება გაჩნდეს გამოტოვებათა მნიშვნელოვანი რაოდენობა, რაც თავის მხრივ ქულათა ჯამზე აისახება დიდი უარყოფითი წილით. თუ ჩვენ ვახდენთ კიდურა გამოტოვებათა იგნორირებას, მაშინ ურთიერთსწორების (11.1) ვარიანტი საკმაოდ კარგია (6 თანხვედრა, 1 ართანხვედრა და 1 გამოტოვება).

ცხადია, რომ ზემოთ მოყვანილი ურთიერთსწორება არ წარმოადგენს ამ ორი თანამიმდევრობის საუკეთესო გლობალურ ურთიერთსწორებას. ქვემოთ წარმოგიდგენთ სხვა ურთიერთსწორებას (11.2), რომელიც შეესაბამება მოცემულ თანამიმდევრობათა გლობალურ ოპტიმალურ ურთიერთსწორებას და რომელიც ხასიათდება უმაღლესი ქულათა ჯამით (-12, მაშინ როცა წინა ვარიანტი ჯამი -19 ია).

მიუხედავად იმისა, რომ უკანასკნელი ურთიერსწორებისას მიიღწევა ქულათა მაქსიმუმი და ადგილი აქვს მეორე თანამიმდევრობის ყველა ასოს პირველი თანამიმდევრობის იდენტურ ასოებთან შესაბამისობაში მოყვანას, ასეთი ურთიერსწორება მსგავსებული უბნების წარმოჩენის თვალსაზრისით უინტერესოა. იდენტური ასოების მაქსიმალურად შესაბამისობაში მოყვანის მიზნით წარმოებს მეორე თანამიმდევრობის გამოტოვებებით უხეში დანაწევრება. თუ ჩვენ გვინტერესებს გრძელი თანამიმდევრობის ის უბნები, რომლებიც სხვა მოკლე თანამიმდევრობის მსგავსებია, მაშინ ცხადია, რომ პირველი ურთიერსწორება (11.1), უფრო კარგად შეესაბამება ამოცანის ხასიათს. მაგრამ ამ შემთხვევაში ქულათა დარიცხვის პროცესში საჭიროა კიდურა გამოტოვებების იგნორირება (არ დაჯარიმება) და ამის შედეგად ქულათა ჯამი გახდება 3.

ახლა ავღწეროთ ის ცვლილებები ბაზისურ ალგორითმში, რომელთა საშუალებითაც ხდება კიდურა გამოტოვებების იგნორირება. თავიდან განვიხილოთ შემთხვევა, როცა არაა სასურველი საჯარიმო ქულათა დარიცხვა α -ის ბოლო ასოს შემდგომ განთავსებულ გამოტოვებებზე. განვიხილოთ ოპტიმალური ურთიერსწორება ასეთ შემთხვევაში. α -ის ბოლო ასოს შემდგომი გამოტოვებანი შესაბამისობაში მოდის β -ს სუფიქსთან. თუ უკუვაგდებთ ურთიერსწორების ბოლო ნაწილს, დაგვრჩება ურთიერსწორება α -სა და β -ს პრეფიქს შორის, ურთიერსწორების საწყისი ნაწილისათვის დამახასიათებელი ქულათა ჯამით. ასე რომ, α -ის ბოლო ასოს შემდგომ განთავსებულ გამოტოვებებზე საჯარიმო ქულათა დარიცხვის გარეშე, ოპტიმალური ურთიერსწორების მიღწევისათვის საკმარისია მოინახოს საუკეთესო მსგავსება α -სა და β -ს პრეფიქს შორის. ჩვენ უკვე აღვნიშნეთ, რომ a მატრიცის (i, j) ელემენტი ახასიათებს მსგავსებას $s[1..i]$ -ისა და $t[1..j]$ -ის შორის. გამომდინარე აქედან, მიზნის მისაღწევად საკმარისია შეირჩეს a -ს ბოლო სტრიქონში მაქსიმალური ელემენტი და ამრიგად

$$sim(s, t) = \max_{j=1}^n a[m, j] \tag{11.3}$$

შეგნიშნოთ, რომ მოცემულ შემთხვევაში ოდნავ შეიცვალა $sim(s, t)$ არსი. ის უკვე გვიჩვენებს მსგავსებას კიდურა გამოტოვებების იგნორირების პირობებში. ზემოთ მოცემული გამოსახულება განსაზღვრავს საუკეთესო ურთიერსწორების შესაბამის ქულათა ჯამს. რაც შეეხება თვით ურთიერსწორებას, ყოველივე წარმართება ისევე, როგორც ბაზისური ალგორითმის შემთხვევაში, ოღონდ ამ შემთხვევაში ვიწყებთ მოძრაობას (m, k) ელემენტიდან, სადაც k განისაზღვრება

		G	C	A
	0	1	2	3
0	0	-2	-4	-6
A	1	-2	-3	-3
A	2	-4	-3	-2
G	3	-6	-3	-3
C	4	-8	-5	-4

სურათი 11.1. s=AAGC და t=GCA თანამიმდევრობათა შორის მსგავსების მატრიცა, აგებული ბაზისურ ალორითმზე დაყრდნობით. s-ის ბოლო ასოს შემდგომ განთავსებულ გამოტოვებებზე საჯარიმო ქულათა არდარიცხვის პირობებში ოპტიმალური ურთიერთსწორების მიღწევისათვის საკმარისია მოძრაობა დავიწყეთ ქვედა კიდურა სტრიქონის მაქსიმალური ელემენტიდან. ამ შემთხვევაში ოპტიმალური ურთიერთსწორების ქულათა ჯამი = -2.

(11.3) პირობით. 11.1 სურათზე მოცემულია მსგავსების მატრიცა და ისრების მეშვეობით განსაზღვრული ოპტიმალური ურთიერთსწორების შესაბამისი გზა მოცემული შემთხვევისათვის.

ანალოგიურად, როცა არ გვინდა რომ დაჯარიმდეს x -ს ბოლო გამოტოვებები, მაქსიმუმი მოიძებნება x -ს ბოლო სვეტში და იქიდან იწყება მოძრაობა.

შეგვიძლია გავაერთიანოთ ეს ორი მიდგომა და ვიპოვოთ საუკეთესო ურთიერთსწორება იმ შემთხვევისათვის, როცა უნდა მოხდეს ორივე თანამიმდევრობის ბოლო გამოტოვებათა იგნორირება. ამ შემთხვევაში მაქსიმუმის მოძიება ხდება მატრიცის განაპირა ზოლში, რომელიც აერთიანებს კიდურა სვეტსა და კიდურა სტრიქონს და ურთიერთსწორების მისაღწევად მოძრაობა იწყება სწორედ ასეთნაირად შერჩეული მაქსიმალური ელემენტიდან $(0,0)$ ელემენტის მიმართულებით.

ახლა განვიხილოთ საწყისი გამოტოვებების შემთხვევა. დავუშვათ საჭიროა ისეთი საუკეთესო ურთიერთსწორების მიღწევა, რომელიც უნდა მიმდინარეობდეს x -მდე საწყისი გამოტოვებების იგნორირების პირობებში. იგი ეკვივალენტურია x -სა და x -ს სუფიქს შორის საუკეთესო ურთიერთსწორებისა. ასეთი ურთიერთსწორების მისაღწევად, ისე როგორც ბაზურ შემთხვევაში, უნდა გამოვიყენოთ $(x+1) \times (x+1)$ მატრიცა, მაგრამ ოდნავ სახეშეცვლილი ფორმით. მატრიცა ივსება იგივე ფორმულების გამოყენებით, რომლებსაც ვხმარობდით ბაზისური ალგორითმის შემთხვევაში, ანუ (10.3) განტოლებების გამოყენებით. მაგრამ ამ შემთხვევაში უკვე სხვანაირად ხდება მატრიცის ინიციალიზაცია. პირველი სტრიქონის ელემენტებს ენიჭებათ ნულოვანი მნიშვნელობები. 11.3 სურათზე წარმოდგენილია მაგალითი იმის სადემონსტრაციოდ, რომ (10.3) განტოლებები მართლაც მუშაობს ამ შემთხვევაში.

ანალოგიური ხერხს თუ გამოვიყენებთ და უკვე პირველი სვეტის ინიციალიზაციას მოვახდენთ ნულებით, მაშინ მიიღება მატრიცა, რომელიც შესაბამისობაში იქნება x -ს დასაწყისში (პირველი ასოს წინ) გამოტოვებათა იგნორირებასთან. ამ შემთხვევის საილუსტრაციოდ მაგალითი იხილეთ 11.4 სურათზე.

11.1 ცხრილში თავმოყრილია ინფორმაცია ყველა ზემოხსენებულ შემთხვევათა შესახებ. განიხილება ქულათა არდარიცხვის ოთხი შემთხვევა: s -ის საწყისი და ბოლო და t -ს საწყისი და ბოლო. ჩვენ შეგვიძლია მოვახდინოთ ამ შემთხვევათა კომბინირება და ყველა შემთხვევაში ვიპოვოთ მსგავსება თითქმის ანალოგიური გზით. ერთადერთი რაც უნდა შეიცვალოს, არის ინიციალიზაციის სახე და ის თუ საიდან უნდა იქნეს მოძიებული მოძრაობის დასაწყისის განმსაზღვრელი მაქსიმალური ელემენტი. საწყის გამოტოვებათა იგნორირება ხორციელდება გარკვეული

		A	G	C	
		0	1	2	3
C	0	0	-2	-4	-6
C	1	-2	-1	-3	-3
G	2	-4	-3	0	-2
C	3	-6	-5	-2	1
A	4	-8	-5	-4	-1

სურათი 11.2. $s=CGCA$ და $t=AGC$ თანამიმდევრობათა შორის მსგავსების მატრიცა, t -ს ზოლო ასოს შემდგომ განთავსებულ გამოტოვებებზე საჯარიმო ქულათა არდარიცხვის პირობებში ოპტიმალური ურთიერთსწორების მიღწევისათვის საკმარისია მოძრაობა დავიწყოთ მარჯვენა კიდურა სვეტის მაქსიმალური ელემენტიდან. ამ შემთხვევაში ოპტიმალური ურთიერთსწორების ქულათა ჯამი = 1

		A	G	C	
		0	1	2	3
0	0	0	0	0	
G 1	-2	-1	1	-1	
C 2	-4	-3	-1	2	
A 3	-6	-5	-4	0	
A 4	-8	-5	-4	-2	

სურათი 11.3. $s=GCAA$ და $t=AGC$ თანამიმდევრობათა შორის მსგავსების მატრიცა, აგებული ზაზისური ალორითზე დაყრდნობითა და პირველი სტრიქონის წულოვანი ინიციალიზაციით, ანუ s -ის პირველი ასოს წინ განთავსებულ გამოტოვებებზე საჯარიმო ქულათა არდარიცხვის პირობებში. ოპტიმალური ურთიერთსწორების მიღწევისათვის საკმარისია მოპრაობა დავისწყობთ $s(m,t)$ ელემენტიდან. ამ შემთხვევაში ოპტიმალური ურთიერთსწორების ქულათა ჯამი = -2.

		G	C	A
	0	1	2	3
0	0	-2	-4	-6
A 1	0	-1	-3	-3
A 2	0	-1	-2	-2
G 3	0	1	-2	-3
C 4	0	-1	2	0

სურათი 11.4. $s=AAGC$ და $t=GCA$ თანამიმდევრობათა შორის მსგავსების მატრიცა, აგებული ზაზისური ალგორითმზე დაყრდნობითა და პირველი სვეტის წულოვანი ინიციალიზაციით. ანუ t -ს პირველი ასოს წინ განთავსებულ გამოტოვებებზე საჯარიმო ქულათა არდარიცხვის პირობებში. ოპტიმალური ურთიერთსწორების მიღწევისათვის საკმარისა მოძრაობა დავიწყეთ $a(\text{თ},\text{ა})$ ელემენტიდან. ამ შემთხვევაში ოპტიმალური ურთიერთსწორების $\text{ჩოოათა } \text{ჯამი} = 0$.

პოზიციების წულოვანი ინიციალიზაციით. ბოლო გამოტოვებათა იგნორირება მიიღწევა მოძრაობის დასაწყისის ელემენტთა მოძიების წესით. მაგრამ მატრიცის შევსება ყველა შემთხვევაში (10.3) განტოლებებით წარიმართება.

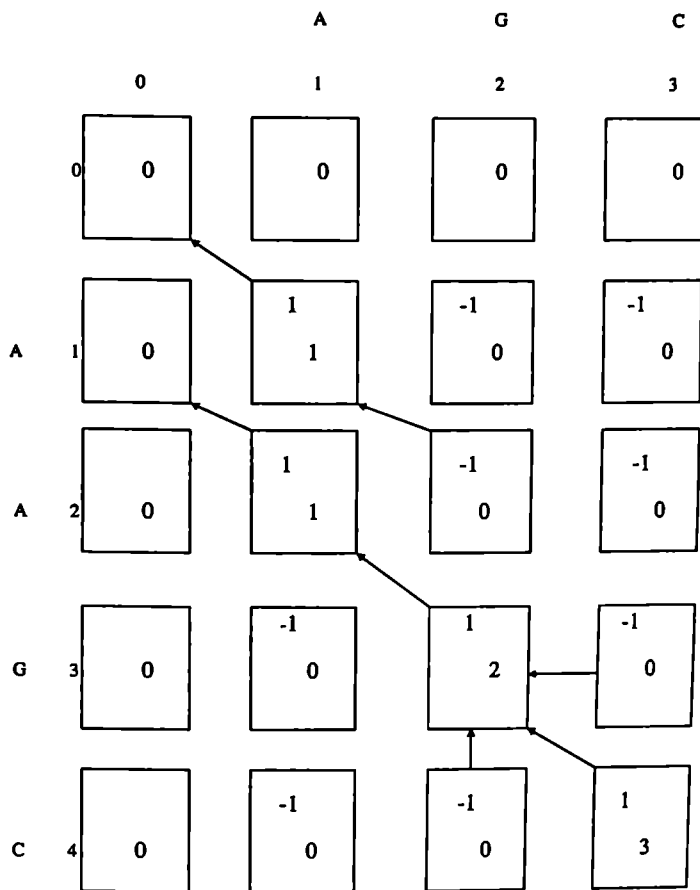
ცხრილი 11.1 გამოტოვებებზე ქულათა არდარიცხვის პირობები.

ადგილი სადაც გამოტოვება არ უნდა დაჯარიმდეს	ქმედება
პირველი თანამიმდევრობის საწყისი	პირველი სტრიქონის წულებით ინიციალიზაცია
პირველი თანამიმდევრობის ბოლო	მაქსიმუმი მოიძიება ბოლო სტრიქონში
მეორე თანამიმდევრობის საწყისი	პირველი სვეტის წულებით ინიციალიზაცია
მეორე თანამიმდევრობის ბოლო	მაქსიმუმი მოიძიება ბოლო სვეტში

ლოკალური ურთიერთსწორება.

i -სა და $i+1$ -ს შორის *ლოკალური ურთიერთსწორება* გულისხმობს i -ის ქვერიგის სწორებას $i+1$ -ს ქვერიგთან. მოცემულ ქვეპარაგრაფში განვიხილავთ ალგორითმს, რომელიც ეძებს ორ თანამიმდევრობას შორის მაქსიმალურ ქულათა ჯამით დახასიათებულ *ლოკალურ ურთიერთსწორებას*. იგი ცნობილია *სმიტ-ვატერმანის ალგორითმის სახელით*.

ეს ალგორითმი წარმოადგენს ბაზისური ალგორითმის ვარიაციას. ამ შემთხვევაშიც, ძირითად მონაცემთა წარმოდგენა ხდება $(m+1) \times (n+1)$ მატრიცის საშუალებით. მაგრამ მატრიცის თითოეული ელემენტი უკვე გვიჩვენებს $s[1..i]$ -ის და $t[1..j]$ -ის შორის სუკეთესო (მაქსიმალურ ქულათა ჯამიანი) *ლოკალური ურთიერთსწორების* ქულებს. პირველი სტრიქონისა და პირველი სვეტის ელემენტები ნულებია. ანუ ყოველი (i,j) -ს შეფასებისას, გამოტოვებების შემცველ



სურათი 11.5. $s=AAGC$ და $t=AGC$ თანამიმდევრობათა შორის ლოკალური მსგავსების მატრიცა, აგებული (11.4) ფორმულით აღწერილი წესით. (i,j) უჯრის ზემო მარცხენა კუთხეში განლაგებული სიდიდეები მიაჩვენებენ უდრის თუ არა $s[i] t[j]$ -ს. სვეტთა და სტრიქონთა შესაბამისი ინდექსების ათვლა წარმოებს 0-დან. ლოკალური ურთიერთსწორების ბილიკის განმსაზღვრელ სასტარტო წერტილად შეიძლება გამოყენებულ იქნეს მატრიცის ნებისმიერი ელემენტი. მათ შორის მაქსიმალურ კულისიანი ურთიერთსწორების გამოსავლენად მაქსიმალური ელემენტი. ურთიერთსწორების ბილიკი მთავრდება ან ელემენტში, საიდანაც ისარი არ გამოდის, ან ნულოვან ელემენტში

სუფიქსებისათვის ხდება ქულათა ჯამის განულება, რათა მათ გავლენა არ იქონიონ ლოკალურად მაქსიმალურ ქულათა ჯამის მქონე ფრაგმენტის მოძიებაზე. აქედან გამომდინარე მატრიცის ელემენტები უნდა იყოს ≥ 0 . გარკვეულწილად ამ ფაქტითაა განპირობებული პირველ სტრიქონსა და სვეტში ნულების განთავსება, ანუ მატრიცის ინიციალიზაციის ნულოვანი ვარიანტი. ინიციალიზაციის შემდგომ, მატრიცის ელემენტთა შევსება მიმდინარეობს ჩვეულებრივი გზით, ანუ $a[i,j]$ -ს სიდიდე დამოკიდებულია წინა ეტაპზე განსაზღვრული სამი ელემენტის მნიშვნელობაზე. საბოლოო სახე დამოკიდებულებისა შეიძლება გამოვსახოთ ასე:

$$a[i,j] = \max \begin{cases} a[i, j-1] + g, \\ a[i-1, j-1] + p(i,j), \\ a[i-1, j] + g, \\ 0 \end{cases} \quad (11.4)$$

რომელიც თითქმის ისეთივეა როგორც ბაზისური ალგორითმის შემთხვევაში. განსხვავება მდგომარეობს მხოლოდ იმაში, რომ არსებობს დამატებით მეოთხე შესაძლებლობა. კერძოდ კი ლოკალურობის განმაპირობებელი, ქულათა ჯამის უარყოფითობისაგან ხელოვნურად დაცვის შესაძლებლობა.

ასეთნაირად შევსებული მატრიცა უზრუნველყოფს მაქსიმალურ ქულათა ჯამიანი ლოკალური ურთიერთსწორების მოძიებას. მატრიცის თითოეული ასეთნაირად მიღებული ელემენტი შეიძლება გამოყენებულ იქნეს, როგორც სასტარტო წერტილი შესაბამისი ურთიერთსწორების მიღწევის პროცესში. ამ წერტილიდან ურთიერთსწორება მიიღება ჩვეულებრივ, უკანა მიმართულებით მოძრაობისას. ურთიერთსწორების ბილიკი მთავრდება ან ელემენტში, სადაც გამომავალი ისარი არ ფიგურირებს, ან ნულოვან ელემენტში.

საზოგადოდ, როცა ვახდენთ ლოკალურ ურთიერთსწორებას, ჩვენ გვაცინტერესებს არა მხოლოდ ოპტიმალური ურთიერთსწორება, არამედ ოპტიმალურთან ახლოს მყოფიც, ანუ რაღაც ზღვრულზე მაღალი ქულათა ჯამით განსაზღვრული.

12. ქულათა დარიცხვის სქემები. ჩანაცვლებათა (მსგავსების) მატრიცები.

ჩვენ გავეცანით დინამიკური პროგრამირების ბაზურ ალგორითმზე დაფუძნებულ ორი ასოთა რიგის (თანამიმდევრობის) შედარებისა და შესაბამისად ურთიერთსწორების თანამედროვე მეთოდებს. მომავალში გავეცნობით სხვა ანალოგიურ მიდგომებს, რომლებიც შემუშავებულია დიდი მოცულობის თანამიმდევრობათა შესახებ მონაცემთა ბაზების ანალიზისათვის. ნათელია, რომ ურთიერთსწორების პროცესში მნიშვნელოვან როლს თამაშობს ქულათა დარიცხვის სისტემა (სქემა). წარმატებული ურთიერთსწორება დიდწილად განპირობებულია ქულათა დარიცხვის სწორი სქემით. მოდით გავეცნოთ ქულათა დარიცხვის თანამედროვე სქემებს და მათი ფორმირების პრინციპებს.

ქულათა დარიცხვისას ცილის შემთხვევაში გამოიყენება ჩანაცვლებათა 20×20 განზომილებიანი მატრიცა, ნუკლეინის მჟავათა შემთხვევაში კი 4×4 მატრიცა. ასეთი მატრიცები უზრუნველგვყოფენ ურთიერთშესაბამისობაში მოყვანილ ყველა შესაძლო ამინმჟავურ თუ ფუძე წყვილზე დასარიცხი ქულის სიდიდით. ცილათშორისი მსგავსების ამოცანის გადაწყვეტისას, მთელი რიგი მოსაზრებებიდან გამომდინარე, მიზანშეწონილია დასკვნა ძირითადად ამინმჟავურ თანამიმდევრობათა შედარებაზე დაყრდნობით გამოვიტანოთ. ეს განპირობებულია იმით, რომ გენეტიკური კოდი გადაგვარებულია, შესაბამისად უფრო დიდი რაოდენობით ინფორმაცია თავმოყრილია ცილის თანამიმდევრობაში; მეორე, სხვადასხვა ორგანოებისა თუ ორგანიზმების ფუძოვანი შემადგენლობები (კომპოზიციები) განსხვავებულია; და ბოლოს, ნუკლეოტიდური თანამიმდევრობის შესახებ ინფორმაციის ნაწილი წარმოებულია ინფორმაციული რნმ-დან, მეორე ნაწილი კი, გენომური დმნ-იდან (თავისი ინტრონებითა და ეგზონებით). სწორედ ასეთი რეალობა განაპირობებს ალბათ ამინმჟავურ ჩანაცვლებათა მატრიცების გაცილებით უფრო ფართო სპექტრის არსებობას. შესაბამისად უპირანია ჩანაცვლებათა მატრიცების ფორმირების პრინციპების განხილვისას ჯერ განვიხილავთ ცილათათვის განკუთვნილი ნიმუშები და შემდეგ შევხვით ნუკლეინის მჟავების შედარებისათვის შემუშავებულ მატრიცებს.

განვიხილოთ ამინმჟავური თანამიმდევრობათა შედარებისათვის განკუთვნილი მატრიცები და აღვნიშნოთ, რომ ამინმჟავათა წყვილის ურთიერთსწორების პროცესში ხშირად გამოიყენება ტერმინი ჩანაცვლება. ეს დაკავშირებულია იმასთან, რომ ქულათა დარიცხვის სქემა ეფუძნება ევოლუციურ მოდელს, რომლის მიხედვითაც ერთი ცილის თანამიმდევრობის მეორეში ტრანსფორმაცია მიმდინარეობს მუტაციათა (რედაქტირების ოპორაციათა - ჩანაცვლება, ჩამატება, გამოტოვება) გარკვეული რიგის რეალიზებით. აქედანაა მატრიცათა დასახე-

ლებაც, „ჩანაცვლებათა ან მსგავსებათა“ მატრიცა. ასეთი მატრიცები, როგორც წესი, სიმეტრიულია, რადგანაც Ala-ის Gly-თან ურთიერთსწორებაში მოყვანას იგივე აზრობრივი დატვირთვა აქვს, როგორც Gly-ის Ala-თან. ქულათა დარიცხვის უმარტივეს წესს გვაძლევს იდენტურობის მატრიცა (იხ. ქვემოთ). იგი ამინმჟავათა თანხვედრის შემთხვევაში არიცხავს 1-ს და არათანხვედრის შემთხვევაში 0-ს. მართალია იდენტურობის მატრიცა სიმარტივის გამო მიზიდავლია, მაგრამ უნდა აღინიშნოს, რომ იგი ყოველთვის ადეკვატურად ვერ აღწერს ცილათა შორის მსგავსებას. შეიძლება ითქვას, რომ ევოლუციის პროცესში ამინმჟავათა ურთიერთჩანაცვლების უნარს დიდწილად განსაზღვრავს მათი ფიზიკურ-ქიმიური თავისებურებანი. მაგალითად, უფრო ალბათურია, რომ ურთიერთჩანაცვლონ ერთი ზომის გვერდითი ნაშთის მქონე ამინმჟავები და არა სხვადასხვა ზომისანი. მუტაციურ ჩანაცვლებათა ალბათობაზე ასევე შეიძლება გავლენა იქონიოს ისეთმა თვისებამ, როგორცაა წყალთან ურთიერთქმედებისაკენ მიდრეკილება. და რადგანაც ცილათა თანამიმდევრობების შედარება ხდება ევოლუციური შეფასებების ჭრილში, იქმნება აუცილებლობა ქულათა დარიცხვის სქემაში შეძლებისდაგვარად გათვალისწინებულ იქნეს ზემოთხსენებული თავისებურებანი. ამრიგად უფრო სრულყოფილ სქემაში გათვალისწინებული უნდა იყოს კონსერვატული ჩანაცვლებების არსებობა. ასე მაგალითად Val-ის Leu-თან ურთიერთსწორება შეიძლება შეფასდეს +4 ქულით, მაშინ როცა Glu-ის Leu-თან, -3 ქულით. უახლოეს წარსულში ასეთი სქემის შესაბამისად შეიქმნა ერთ-ერთ ყველაზე ფართოდ გამოყენებადი მატრიცა, რომელიც ცნობილია PAM (*Percent (probability) of Accepted Mutations*) ან დაიხოფის მატრიცის სახელწოდებით. PAM მატრიცის ფორმირების პირველი ეტაპი გულისხმობდა ცილათა ოჯახებში გაერთიანებულ, მაღალი ჰომოლოგიის მქონე თანამიმდევრობათა ურთიერთსწორებას, ამინმჟავურ ჩანაცვლებათა აღრიცხვას და აღრიცხულ სიდიდეთა ნორმალიზებას, რათა შემდგომში, ევოლუციურად დაშორებულ ცილათათვის ანალოგიურ სიდიდეთა შესაფასებლად შესაძლებელი ყოფილიყო ახლო ნათესაობაში მყოფი ცილების საფუძველზე წარმოებულ სიდიდეთა (ჩანაცვლებების მახასიათებელთა) გამოყენება [Dayhoff *et al.*, 1978]. ევოლუციურ გზაზე დაშორების საზომად გამოიყენებოდა მიღებულ (დაფიქსირებულ) მუტაციათა პროცენტი, ანუ PAM1 (იხ. უფრო დაწვრილებით ქვემოთ) და ყველაზე ფართოდ გამოყენებად მატრიცას კი 250 PAM-დისტანციის შესაბამისი მატრიცა ანუ ე.წ. PAM250 მატრიცა, წარმოადგენდა.

მიუხედავად იმისა, რომ PAM250 მიღებისათვის გამოსადეგ თანამიმდევრობათა რიცხვი შეზღუდული იყო, ეს მატრიცა მაინც სწორად ასახავდა ამინმჟავათა ძირითად ფიზიკურ-ქიმიურ თავისებურებებს. ამ ტიპის მატრიცის შემდგომი გაუმჯობესება მიიღწეოდა თანამიმდევრობათა გაზრდილი რიცხვის გამოყენებით.

BLOSUM-ის სერიის მატრიცები [Henikoff & Henikoff, 1992] ასევე ეფუძნება ცილათა ოჯახის წევრ თანამიმდევრობათა ურთიერთსწორებას და ჩანაცვლებათა შესახებ ინფორმაციას. მაგრამ PAM-საგან განსხვავებით, მატრიცის ფორმირების საწყისი ეტაპი გულისხმობს BLAST ალგორითმით, გამოტოვებათა გარეშე განხორციელებულ ურთიერთსწორებათა სერიას. ურთიერთსწორებაში მონაწილე თანამიმდევრობათათვის დასაშვებია გაცილებით ნაკლები მსგავსების არსებობა, ვიდრე ეს დაიხოფის ალგორითმის შემთხვევაში იყო მიზანშეწონილი. ტესტირებისას BLOSUM სერიის მატრიცებმა გამოამჟღავნეს მნიშვნელოვანი უპირატესობა როგორც PAM-ის, ისე ნებისმიერი სხვა მატრიცასთან მიმართებაში. აქედან გამომდინარე, ცილების ურთიერთსწორების პროცესში ძირითად გამოიყენებოდა BLOSUM მატრიცები. უმრავლეს პოპულარულ პროგრამათა პაკეტებში, რომლებიც იხმარება ცილათა თანამიმდევრობების ურთიერთსწორებისა თუ მონაცემთა ბაზების ანალიზისათვის, ისინი ავტომატურად (არმითითების შემთხვევაშიც) გამოიყენებიან.

ჩანაცვლებათა ანუ მსგავსებათა მატრიცის ფორმირებისას Overington *et al* (1992) იყენებდნენ მხოლოდ ცნობილი სივრცული სტრუქტურის მქონე ცილებს. ამასთან ფორმირების პროცესში ეფუძნებოდნენ არა თანამიმდევრობათა ურთიერთსწორების ტრადიციულ მიდგომებს, არამედ სტრუქტურულ ურთიერთსწორებებს. ანუ ცილების ურთიერთსწორება გულისხმობდა არა თანამიმდევრობების, არამედ სტრუქტურების შესაბამისობაში მოყვანას და სწორედ ასეთი ურთიერთსწორების მონაცემების დამუშავების საფუძველზე მიიღებოდა ჩანაცვლებათა სტატისტიკა. პრინციპში, Overington-ის მატრიცაზე დაყრდნობით მიღებული მონაცემები უფრო დამაჯერებელი უნდა ყოფილიყო, ვიდრე PAM ან BLOSUM -ით მიღებული. თუმცა ცილათა ცნობილი სტრუქტურების სიმცირე ამცირებდა Overington-ის მატრიცების სტატისტიკურ დამაჯერებლობას. შემდგომში Overington-მა და თანამშრომლებმა ცილების სივრცულ სტრუქტურათა შესახებ ინფორმაციის ზრდასთან ერთად გააუმჯობესეს თავიანთი მატრიცა.

ამინმჟავათა ჩანაცვლების უმარტივესი მატრიცები.

ამინმჟავათა ფიზიკურ-ქიმიურ თავისებურებათა არაგამთვალისწინებელი მატრიცის მაგალითად შეიძლება მოვიყვანოთ, ორი უმარტივესი მატრიცა:

1. ერთეულოვანი (იდენტურობის) მატრიცა, სადაც მატრიცის ელემენტი მიიღება

$$M_{ij} = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{otherwise.} \end{cases} \quad (12.1)$$

და

2. გენტიკური კოდის მატრიცა, სადაც M_{ij} გვიჩვენებს ფუძეთა ჩანაცვლების იმ მინიმალურ რიცხვს, რომელთა განხორციელებაც აუცილებელია i -ური ამინმჟავის გენეტიკური კოდის j -ურ ამინმჟავის კოდში გადასაყვანად.

PAM მატრიცები

PAM ერთეული

PAM ერთეული გამოიყენება ორ ამინმჟავურ თანამიმდევრობას შორის ევოლუციური მანძილის (დაშორების) საზომად. ორ თანამიმდევრობას შორის დივერგირების გზა ერთი PAM ერთეულის სიგრძისაა, თუ ამ თანამიმდევრობათა ურთიერთკონვერტირება შესაძლებელია ისეთი რაოდენობა დასაშვები წერტილოვანი მუტაციებით, რომ 100 ამინმჟავაზე მოსული მათი საშუალო რიცხვი უტოლდება ერთს. ტერმინი „დასაშვები“ გულისხმობს ისეთ მუტაციებს, რომლებიც ფიქსირდება ცილაში და გადაეცემა შთამომავლობას. ანუ მუტაციები ან არ ცვლის ცილის ფუნქციას, ან თუ ცვლის, ეს ცვლილებები არაა უსარგებლო ორგანიზმისათვის (შეუნიშნოთ, რომ ერთი PAM ერთეულით დივერგირებული ასოთა რიგი არ გულისხმობს აუცილებლად ცვლილებებს პოზიციათა ერთ პროცენტში, რადგანაც ერთ პოზიციამში შეიძლება დაფიქსირდეს ერთზე მეტი მუტაცია).

PAM ერთეულის ცნების შემოღება ორ ძირითად პრობლემასთანაა დაკავშირებული:

1. პრაქტიკულად, დღეისათვის ცნობილი ყველა თანამიმდევრობა თანამედროვე ორგანიზმებიდანაა მოპოვებული. ჩვენ არ შეგვიძლია ვთქვათ, რომ ერთი თანამიმდევრობა მეორიდანაა წარმოებული. წინაპარ ცილათა თანამიმდევრობების არცოდნის პრობლემის გადალახვა

შესაძლებელია, თუ დავუშვებთ, რომ ამინმჟავათა მუტაციის პროცესი შექცევადია და ტოლფასია ორივე მიმართულებით. ეს დაშვება საშუალებას გვაძლევს ორი S_i და S_j თანამიმდევრობისთვის, რომელთა წინაპარს წარმოადგენს, ვთქვათ S_j დავეწროთ შემდეგი თანაფარდობა

$$d(S_i, S_j) = d(S_i, S_{ij}) + d(S_{ij}, S_j) \quad (12.2)$$

$d(i, j)$ წარმოადგენს i and j ამინმჟავურ თანამიმდევრობებს შორის PAMდისტანციას

2. მეორე, უფრო ძნელად გადასალახი პრობლემა უკავშირდება ურთიერთსწორების დამაჯერებლობას. ამ პრობლემის გადალახვა ხდება გამოტოვებითა და ჩამატებითა იგნორირებით. ამავე დროს გამოტოვებებსა და ჩამატებებს ევოლუციის პროცესში მუდმივად აქვს ადგილი. შესაბამისად, არ არსებობს მყარი გარანტია იმისა, რომ ურთიერთსწორების შედეგად ამინმჟავათა პოზიციები სწორ შესაბამისობაში მოყვანილი. ანუ არ არსებობს გარანტია, რომ ისტორიული გამოტოვებები სწორადაა დაფიქსირებული. განსაკუთრებით მცირდება გარანტია, თუ თანამიმდევრობები მნიშვნელოვნადაა დივერგირებული.

მიღებულ მუტაციათა ალბათობებზე დაფუძნებული მატრიცები

PAM მატრიცა წარმოადგენს ამინმჟავათა ჩანაცვლების მატრიცას და ახასიათებს ევოლუციის პროცესში ამინმჟავურ თანამიმდევრობაში მოსალოდნელ ცვლილებებს. თითოეული PAM მატრიცა ისეა წარმოებული, რომ შეძლოს PAM-ის შესაბამისი ერთეულით დაშორებულ თამიმდევრობათა შედარება. ასე, PAM120-ს შეუძლია 120PAM-ით დაშორებული თანამიმდევრობების შედარება. მატრიცის ელემენტი, რომლითაც ამინმჟავათა წყვილს ვახასიათებთ, წარმოადგენს 120PAM დისტანციაზე ამინმჟავათა ურთიერთჩანაცვლების შესაძლებლობის ალბათობას (ალბათობის ლოგარითმს). ამრიგად PAM m მატრიცის (i, j) ელემენტი გვიჩვენებს სიხშირეს, რომლითაც m PAM ერთეულ დისტანციაზე ერთი თანამიმდევრობის მეორეში დივერგირების პროცესში A_i ამინმჟავა შეიძლება შეიცვალოს A_j -თი. ამ სიხშირეთა შეფასება ხდება ამინმჟავური ჩანაცვლებების შესახებ მონაცემთა სტატისტიკური ანალიზის საფუძველზე.

როგორც ზემოთ აღვნიშნეთ, PAM m მატრიცის ფორმირების მიზნით ამინმჟავათა ჩანაცვლებების შესახებ მონაცემების შეგროვება შედარებით ძნელია, თუ თანამიმდევრობებს შორის დივერგირების პროცესი გრძელია. მაგრამ დიდი მსგავსების მქონე თანამიმდევრობათა პოზიციებს შორის სწორი შესაბამისობის

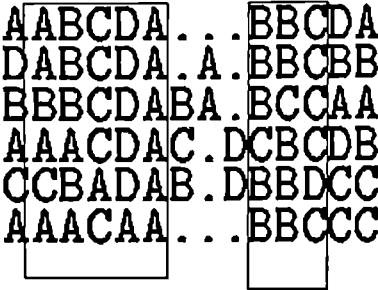
კი, A_j ამინმჟავის შეხვედრის სიხშირეა. ასეთი მიდგომა გულისხმობს, რომ დროის განმავლობაში არ ხდება უკანასკნელ სიხშირეთა სიდიდეების ცვლილება, და IPAM ანუ ერთეულ დისტანციაზე მიმდინარე ცვლილებების გამომწვევი მუტაციური პროცესი იდენტურია უფრო დიდ დისტანციაზე მიმდინარე პროცესისა (პროცესი და მექანიზმები უცვლელია). საბოლოო მატრიცის, ანუ ურთიერთსწორებულ წყვილებზე დასარიცხი ქულების მომცველი მატრიცის მისაღებად ზემოთ მოცემული გამოსახულებიდან იღებენ ლოგარითმს იმ მიზნით, რომ საქმე გვექონდეს აჯამვასთან და არა გამრავლებასთან. ანუ

$$PAM^n(i,j) = 10 \log_{10} ((M_{ij}^n / p)). \tag{12.4}$$

PAM მატრიცა, როგორც წესი, ისეა ორგანიზებული რომ მოხდეს მსგავს ამინმჟავათა დაჯგუფება. მაგალითის სახით 12.1 სურათზე მოყვანილია PAM 250 მატრიცა. სურვილის შემთხვევაში PAM მატრიცის ფორმირებისას გამოყენებული მიდგომის უფრო დანვრილებით აღწერას შეგიძლიათ გაეცნოთ ქვემოთ.

BLOSUM - ბლოკებით განსაზღვრული ჩანაცვლებების მატრიცა

BLOSUM მატრიცა ასევე წარმოადგენს ამინმჟავათა შორის შესაძლო ჩანაცვლებების დამახასიათებელ მატრიცას, რომელიც პირველად წარმოადგინა Henikoff და Henikoff-მა 1992 წელს. ამ მატრიცის მისაღებად გამოიყენება ამინმჟავური თანამიმდევრობების მხოლოდ ის ბლოკები (ცხადია მიღებული ურთიერთსწორების შედეგად), რომელთა შორის განსხვავება მცირეა, თუმცა სრულ თანამიმდევრობათა შორის განსხვავება შეიძლება შედარებით დიდი იყოს.



სურათი 12.2 რამდენიმე თანამიმდევრობის ერთდროული ურთიერთსწორება. კონსერვატული ბლოკები გამოყოფილია.

როგორც BLOSUM-*x*. ასე მაგალითად, თუ მატრიცის აგების პროცესში გამოყენებულ თანამიმდევრობათა შორის იდენტურობა არ აღემატება 50%-ს, მატრიცას უწოდებენ BLOSUM-50-ს .

მეორე ეტაპი გულისხმობს მრავალწევრიანი ერთდროული ურთიერთსწორების შედეგად, შესაბამისობაში მოყვანილ ამინმჟავათა სვეტში ამინმჟავათა წყვილების რიცხვის შეფასებას. ასე მაგალითად, AABACA შემადგენლობის (12.2 სურათზე მეორე სვეტი) სვეტში გვაქვს 6 AA, 4 AB, 4 AC და 1 BC წყვილი. ამის შემდეგ ფასდება სვეტში მოცემული (A_i, A_j) ამინმჟავური წყვილის გამოჩენის q_{ij} ალბათობა და ასევე კონკრეტულ პოზიციასში A_i ამინოჟავის გამოჩენის p_i ალბათობა.

მესამე ეტაპზე ითვლება უკანასკნელ ალბათობათა შეფარდების ლოგარითმი შემდეგი ფორმულის შესაბამისად $s_{ij} = \log_2(q_{ij} / p_i)$. საბოლოოდ გამოიანგარიშება $2s_{ij}$ სიდიდის დამრგვალებული მნიშვნელობა და სწორედ იგი თამაშობს BLOSUM-*x* მატრიცის M_{ij} ელემენტის როლს. PAM-თან შედარებით BLOSUM-ის ფორმირება თანამიმდევრობათა უფრო დიდი რაოდენობის ანალიზის საფუძველზე ხდება. ამასთან, თანამიმდევრობები ხასიათდება მსგავსების სხვადასხვა დონით, და აქედან გამომდინარე, მატრიცათა განსხვავებულ ტიპებთან გვაქვს საქმე.

12.3 სურათზე მაგალითის სახით მოცემულია BLOSUM45 მატრიცა.

ნუკლეოტიდურ თანამიმდევრობათა შედარებისას გამოყენებული ქულათა დარიცხვის სისტემები

ისევე როგორც ცილების შემთხვევაში, ნუკლეინის მჟავათა შედარების პროცესშიც არსებობს დასარიცხ ქულათა სხვადასხვა სისტემები. ცხადია, ამ შემთხვევაში არჩევანი უფრო მწირია. როგორც წესი გამოიყენება PAM 47 მატრიცა, რომელიც თანაბარ ქულას არიცხავს როგორც ტრანზიციას (transitions - პურინების პურინებით ან პირიმიდინების პირიმიდინებით შეცვლას), ისე ტრანსვერსიას (transversions - პირიმიდინების პურინებით ან პირიქით შეცვლას). ანუ ქულათა მხოლოდ ორი მნიშვნელობა გამოიყენება, 5 იდენტურობის წასახალისებლად და - 4 არაიდენტურობის დასაჯარიმებლად. მაგრამ არსებობს ქულათა დარიცხვის სხვა სისტემაც, რომელიც არ გულისხმობს ტრანზიციასა და ტრანსვერსიაზე ერთნაირი ქულების დარიცხვას. ეს უკანასკნელი სისტემა იყენებს PAM 50 მატრიცას, რომლის ტრანზიციასზე დასარიცხი ქულის მნიშვნელობა სამჯერ

აღმატება ტრანსვერსიაზე დასარიცხ ქულის მნიშვნელობას. 12.4 სურათზე მოყვანლია შესაბამისი მატრიცა.

A	1.36			
G	-0.37	1.36		
C	-1.60	-1.60	1.36	
T	-1.60	-1.60	-0.37	1.36
	A	G	C	T

სურათი 12.4. ნუკლეინის მჟავათათვის გათვალისწინებული ქულათა დარიცხვის PAM 50 მატრიცა, ტრანზიციასა და ტრანსვერსიას შორის თანაფარდობა შეესაბამება 3:1

გამოტოვებებისა და ჩამატებების დამაჯარიმებელ ქულათა დარიცხვის სისტემა

ურთიერთსწორების პროცესში წარმოშობილ გამოტოვებათა დაჯარიმება, როგორც წესი, წარმოებს შემდეგი სახის ორწევრა, წრფივი ფუნქციით

$$s = o + kg \quad (12.5)$$

სადაც გამოტოვებების გახსნაზე დასარიცხი (o) საჯარიმო ქულის სიდიდე განსხვავდება k რაოდენობა მიმდევრობით განლაგებულ შემდგომ გამოტოვებებზე დასარიცხ (g) ჯარიმის სიდიდისაგან. დაჯარიმების ასეთ სისტემას გააჩნია გარკვეული ბიოლოგიური საფუძველი და შესაბამისად უფრო ეფექტურია.

უფრო ვრცლად PAM მატრიცის მიღებასა და თავისებურებებზე

მულტიპლ ჩანაცვლებებზე ზემოქმედების უნარის მქონე ფაქტორთა სიმრავლის გამო, ამინმჟავურ ნაშთთა მსგავსების მახასიათებელი სიდიდეები უპრიანია ვაწარმოთ უშუალოდ რეალურ ჩანაცვლებათა სიხშირების შეფასების გზით. სწორედ ამ გზით არის შემუშავებული ქულათა დარიცხავი მატრიცების მნიშვნელოვანი ოჯახი, ე. წ. PAM

მატრიცები. აბრევიატურა PAM იზიფრება როგორც *Point Accepted Mutations (მიღებული წერტილოვანი მუტაციები)*, ან როგორც *Percent (probability) of Accepted Mutations (მიღებულ მუტაციათა წილები(ალბათობები))*. ბაზური PAM1 მატრიცა შეესაბამება ევოლუციის იმ პერიოდს, რომლის განმავლობაშიც 100 ამინმჟავაზე მოდის საშუალოდ 1 მიღებული მუტაცია. ამ პარაგრაფში მოკლედ ავლწერთ PAM-ის არსს და მისი გამოთვის ხერხს.

უპირველეს ყოვლისა, უნდა აირჩეს ის ევოლუციური დისტანცია, რომელშიც მოხდება თანამიმდევრობათა შედარება. მატრიცა დამოკიდებულია ამ დისტანციაზე. ასე მაგალითად, PAM250 მატრიცა გამოდგება იმ თანამიმდევრობათა შესადარებლად, რომლებიც ევოლუციურ გზაზე 250 ერთეულით არის დამორებული. უპრინაია პირველ რიგში მოხდეს PAM1-ის შესაბამისი მატრიცის ფორმირება და შემდეგ მასზე დაყრდნობით, მიღებულ იქნეს სხვა დისტანციის მატრიცები. ასევე უნდა აღინიშნოს, რომ მუტაციები განიხილება მხოლოდ ამინოჟავური თანამიმდევრობის და არ დნმ-ის დონეზე.

თითოეული ევოლუციური დისტანციისთვის ფორმირდება გადასვლათა ალბათობების მატრიცა (*probability transition matrix*) M და ქულათა ჯამების (*scores matrix*) მატრიცა S . PAM1 მატრიცის შესაქმნელად საჭიროა:

- მიღებულ მუტაციათა ჩამონათვალი
- თითოეული a ამინმჟავის შეხვედრის p ალბათობა

მიღებულ მუტაციებში იგულისხმება ის მუტაციები, რომლებიც იქნა დაფიქსირებული, და რომლებმაც ორგანიზმში არ გამოიწვია რაიმე დრამატული ცვლილება, ანუ გარემოსთან შეუუბნის თვალსაზრისით ორგანიზმმა შეძლო ბუნებრივი შერჩევის გავლა. მიღებულ მუტაციათა გამოვლენისა და მათ შესახებ მონაცემების შეგროვების ერთ-ერთ გზას წარმოადგენს სხვადასხვა სახეობიდან გამოყოფილი, ჰომოლოგიურ ცილათა თანამიმდევრობის ურთიერთსწორება. მაგალითად, ადამიანისა და ორანგუტანგის ჰემოგლობინის ალფა ჯაჭვების თანამიმდევრობებისა. თითოეული ის პოზიცია, სადაც ადგილი აქვს თანამიმდევრობათა შორის განსხვავებას შეიძლება ჩათვალოს მიღებულ (რეალიზებულ) მუტაციად. ეს მიღებული მუტაცია განიხილება როგორც არამიმართული მოვლენის შედეგი; ანუ, ამინმჟავათა მოცემული, ურთიერთსწორებული (a,b) წყვილისათვის ჩვენ არ ვიცით დანამდვილებით რომელი მათგანის ჩანაცვლება ხდება რომლით. ანუ ჩვენ არ გავაჩნია ინფორმაცია იმის შესახებ, თუ რომელია წინაპარი. უფრო მეტიც, შესაძლებელია რომ წინაპარს წარმოადგენდა სულაც მესამე თანამიმდევრობა, რომელსაც ამ პოზიციაში გააჩნდა სულაც სხვა, მესამე ამინმჟავა. თუმცა ამ უკანასკნელი მოვლენის შესაძლებლობის ალბათობა მინიმუმამდე დაიყვანება, თუ განვიხილავთ ძალზე ახლო ნათესაობაში მყოფ თანამიმდევრობებს. ბაზისური PAM1 მატრიცის ფორმირებისას ძალზედ მნიშვნელოვანია უშუალო $a \rightarrow b$ და არა შუალედური $a \rightarrow c \rightarrow b$ მუტაციების განხილვა.

ამინმჟავის შეხვედრის ალბათობის შეფასება ხდება მარტივად. გამოითვლება ამინმჟავის შეხვედრის სიხშირე, რომელიც აკმაყოფილებს პირობას

$$\sum_a P_a = 1 \quad (12.6)$$

მიღებულ მუტაციათა ჩამონათვალის ანალიზის საფუძველზე ფასდება f_a მუტაციათა რაოდენობა. რადგანაც საქმე გვაქვს მხოლოდ არამიმართულ მუტაციებთან $f_a = f_{ab}$. დაგვიკრდება ასევე შემდეგი ჯამები

$$f_a = \sum_{b \neq a} f_{ab} \quad (12.7)$$

რომელიც წარმოადგენს a -ს მომცველ მუტაციათა სრულ რიცხვს, და

$$f = \sum_a f_a \quad (12.8)$$

მუტაციებში ჩართულ ამინმჟავათა სრულ რიცხვს. f სიდიდე ორჯერ აღემატება მუტაციათა სრულ რიცხვს.

IPAM ინტერვალისათვის გადასვლათა ალბათობების M მატრიცის შესაქმნელად საკმარისია f_a სიხშირისა და p_a ალბათობის ცოდნა. M მატრიცა წარმოადგენს 20×20 მატრიცას, რომლის M_{ab} ელემენტი გვიჩვენებს a ამინმჟავის b ამინმჟავით შეცვლის ალბათობას. შევნიშნოთ, რომ როცა a უდრის b -ს, ელემენტი გვიჩვენებს მოცემული ევოლუციურ ინტერვალში მოცემული ამინმჟავის არშეცვლის ალბათობას. M_{aa} ელემენტების შეფასება წარმოებს ამინმჟავათა ფარდობით მუტაბელურობაზე დაყრდნობით, რომელიც გამოითვლება ასე:

$$m_a = f_a / (100 f p_a) \quad (12.9)$$

ამინმჟავის მუტაბელურობა წარმოადგენს იმის ზომას, თუ რამდენად ხშირად ხდება მისი შეცვლა. იგი წარმოადგენს მოცემულ ევოლუციურ პერიოდში მოცემული ამინმჟავის შეცვლის ალბათობას. შესაბამისად ალბათობა იმისა, რომ a -ს შეცვლა არ მოხდეს უდრის

$$M_{aa} = 1 - m_a \quad (12.10)$$

მეორეს მხრივ a -ს b -თი შეცვლის ალბათობა შეიძლება წარმოდგენილ იქნეს, როგორც ნამრავლი a -ს b -თი შეცვლის პირობითი ალბათობისა და a -ს შეცვლის ალბათობისა. პირობითი ალბათობა ფასდება, როგორც $a \leftrightarrow b$ მუტაციების რიცხვის ფარდობა a -ს მომცველ მუტაციათა სრულ რიცხვთან. აქედან გამომდინარე:

$$\begin{aligned} M_{ab} &= Pr(a \rightarrow b) = \\ &= Pr(a \rightarrow b / a \text{ ცვლებადია}) Pr(a\text{-ს შეცვლისა}) \quad (12.11) \\ &= (f_{ab} / f_a) m_a \end{aligned}$$

შევნიშნოთ, რომ ამ ალბათობის შეფასებისას გამოიყენება ცილათა ევოლუციის გამარტივებული მოდელი. ასე მაგალითად, დაიშვება, რომ ამინმჟავათა მუტაცია არ არის დამოკიდებული პრეისტორიაზე, რაც ალბათ არ წარმოადგენს ჭეშმარიტებას გენეტიკური კოდის ბუნებიდან გამომდინარე. ასევე მხედველობაში არ მიიღება სხვა ამინმჟავების შესაძლო გავლენა მოცემული ამინმჟავას მუტაციის პროცესზე. პრეისტორიაზე დამოუკიდებლობას მივყავართ მარკოვის ევოლუციურ მოდელამდე, რომელიც უზრუნველყოფილია საკმაოდ კარგი მათემატიკური აპარატით.

M ხასიათდება შემდეგი თვისებებით, რაც ადვილი შესამოწმებელია:

$$\sum_b M_{ab} = 1 \quad (12.12)$$

$$\sum_a p_a M_{aa} = 0.99 \quad (12.13)$$

(12.12) განტოლება აჩვენებს, რომ თუ ავჯამავთ a -ს არშეცვლის ალბათობას და მისი ყველა სხვა შესაძლო ამინმყავათი შეცვლის ალბათობებს მივიღებთ 1-ს, რის გამოც ეს სიდიდეები მოიხსენიება როგორც "ალბათობები". კვლავ შევნიშნოთ, რომ ეს ალბათობები შეესაბამება ევოლუციური ცვლილებების ერთეულს. ბუნებრივია ვიფიქროთ ევოლუციური ერთეულისა და დროითი ერთეულის შესაბამისობაზე. მაგრამ რადგანაც ზოგადად მიღებულია, რომ ცვლილებათა რეალიზების სიჩქარე ცვლადი სიდიდეა, ამიტომ დროისა და ევოლუციური ინტერვალების სიდიდე არაა პირდაპირპროპორციული.

მოცემულ მოდელში ევოლუციის ერთეული შეესაბამება საშუალოდ ყოველ 100 ამინმყავაზე 1 ცვლილებას. ასეთი ერთეული მოიხსენიება, როგორც 1-PAM ევოლუციური დისტანცია. როგორც (12.11) გამოსახულებიდან ჩანს უკანასკნელი დაშვების ასახვის მიზნით ხდება გადასვლათა ალბათობების მატრიცის ნორმირება. თუ შემოვიღებთ სხვა ერთეულს, ვთქვათ, თუ 100-ის მაგივრად გამოვიყენებთ 50-ს, მივიღებთ ზუსტად იგივე თვისებების მატრიცას, მხოლოდ იმ განსხვავებით, რომ (12.11)-ში უნდა აისახოს ის ფაქტი, რომ ერთეული გულისხმობს საშუალოდ 1 ცვლილებას ყოველ 50 ამინმყავაზე და არშეცვლის ალბათობა უკვე გახდება 0.98.

თუ ჩამოვადობდა ბაზური M მატრიცა, შესაძლებელია გადასვლათა ალბათობების წარმოება უფრო დიდი ევოლუციური მანძილებისათვის. ასე მაგალითად, როგორ შეიძლება შევასდეს ევოლუციის 2-PAM ერთეულზე a -ს b -თი ცვლილების ალბათობა? შეიძლება ვივარაუდოთ, რომ პირველი ერთეულის პერიოდში a იცვლება ნებისმიერი c ამინმყავათი M_{ac} ალბათობით (შემცვლელი შეიძლება თვითონ a -ც იყოს). მეორე პერიოდში უკვე მოხდება c -ს b -თი შეცვლა M_{cb} ალბათობით. თუ ყოველივე ამას შევავაძებთ, აღმოჩნდება, რომ საბოლოო ალბათობა არ აღემატება M_{ab} -ს კვადრატს, ანუ M -ის ელემენტის კვადრატს. თუ განვაზოგადებთ, შეიძლება ითქვას, ევოლუციის k ერთეულის ტოლი პერიოდის შესაბამის გადასვლათა ალბათობების მატრიცა გამოითვლება როგორც M^k .

ჩვენ უკვე შეგვიძლია დასარიცხ ქულათა მატრიცის ჩამოვადობება. მატრიცის ელემენტები განისაზღვრება, როგორც ორი ალბათობის შეფარდება, კერძოდ, წყვილთა ურთიერთმუტირებისა და მათი შემთხვევით ურთიერთშეცვლის ალბათობებისა. ასეთი შეფარდება მოიხსენიება, როგორც შემთხვევითთან უპირატესობის მაჩვენებელ შეფარდებად.

ორი a და b ამინმყავისათვის შევავასოთ უკანასკნელი შეფარდება. დავუშვათ, რომ მოცემულმა ურთიერთსწორებამ მოახდინა a -ს და b -ს დაწყვილება. დავიწყოთ a -ს განხილვით, მაშინ ალბათობა იმისა, რომ სხვა თანამიმდევრობის იგივე პოზიციამში მუტირების შედეგად გამოჩნდება b ტოლია M_{ab} -ის. მეორეს მხრივ, ალბათობა იმისა, რომ იგივე პოზიციამში b შემთხვევით აღმოჩნდება, უდრის p_b -ს. აქედან გამომდინარე, შეფარდებას შემდეგი სახე ექნება

$$M_{ab} / p_b \quad (12.14)$$

იგივე მსჯელობით მიიღება შეფარდება როცა $a=b$. რეალური დასარიცხი ქულა წარმოადგენს უკანასკნელი შეფარდების ლოგარითმს გამრავლებულს 10-ზე, რათა რამდენიმე წყვილის ურთიერთსწორებისას ცალკეულ ქულათა აჯამვის შედეგი შესაბამებოდეს შეფარდებათა ნამრავლს. გამოთვლების დაჩქარების მიზნით პრაქტიკაში გამოიყენება რეალური სიდიდეების უახლოეს მთელამდე დამრგვალებული მნიშვნელობები. სწორედ ნამდვილ სიდიდესა და მთელამდე დამრგვალებით მიღებულ მიახლოებას შორის ცდომილების შემცირების მიზნით ხდება ლოგარითმის 10-ზე გამრავლება.

წინამდებარე განხილვა ეხებოდა I-PAM -ს, თუმცა ნებისმიერი ევოლუციური დისტანციისათვისაც გამოსადეგია იგივე სქემის გამოყენება. ასე მაგალითად, დასარიცხ ქულათა მატრიცა k-PAM დისტანციისათვის განისაზღვრება ასე :

$$score(a,b) = 10 \log_{10} (M_{ab} / p_b) \quad (12.15)$$

ასეთი მატრიცა სიმეტრიულია. დასარიცხ ქულათა მატრიცა სიმეტრიული უნდა იყოს, რათა უზრუნველგვეყოს ისეთ დასარიცხ ქულათა სისტემით, რომელიც არ იქნება დამოკიდებული იმაზე, თუ რომელ თანამიმდევრობას ჩავთვლით წინაპრად. სიმეტრიულობის ჩვენება $k=1$ შემთხვევისათვის ადვილია

$$M_{ab} / p_b = (f_a m_a) / (f_a p_b) = f_b / (100 f_p p_b) \quad (12.16)$$

დიდი k-ების შემთხვევაშიც შესაძლებელია სიმეტრიულობის ჩვენება.

მიუხედავად იმისა, რომ დასარიცხ ქულათა მატრიცა გენერირებულია მიღებული მუტაციებისა და ამინმჟავათა შეხვედრის სიხშირების მონაცემებზე დაყრდნობით, მასში ვლინდება ამინმჟავათა მთელი რიგი მნიშვნელოვანი ქიმიური და ფიზიკური თვისებები, როგორცაა წყლისადმი მიდრეკილება და გვერდითი ჯაჭვის ზომა. როგორც წესი, მსგავსი თვისებების მქონე ამინმჟავათა ურთიერთსწორებულ წყვილზე დასარიცხი ქულა მაღალია, მაშინ როცა არამსგავსი თვისებების მქონე ამინმჟავათათვის ანალოგიური ქულა მცირეა. ეს არაა გასაკვირი, რადგან გავრცელებულ მუტაციებში უფრო ხშირად ადგილი აქვს მსგავსი ამინმჟავების ურთიერთჩანაცვლებას, ვიდრე არამსგავსისას. ევოლუციის პროცესში ცილის საჭირო თვისებების შენარჩუნების მოთხოვნა უზრუნველყოფილია სწორედ მსგავს ამინმჟავათა უპირატესი ურთიერთჩანაცვლებებით.

13. მსგავსება და დისტანცია. ფილოგენეტიკური ხე

მსგავსებისა და დისტანციის ცნებები.

მსგავსება და დისტანცია წარმოადგენს ორ მახასიათებელს, რომელთა საშუალებითაც შეიძლება თანამიმდევრობების (ასოთა რიგების) შედარება. მსგავსების შეფასებისას ხდება ორ რიგს შორის საუკეთესო ურთიერთსწორების მოძიება, და ასეთი ურთიერთსწორების შესაბამისი ქულათა ჯამი გვაძლევს რიგების მსგავსების ზომას. დისტანციის განსაზღვრისას კი, ხდება რედაქტირების ელემენტარულ ოპერაციათა რიცხვის შეფასება და ერთი ასოთა რიგის მეორეში ტრანსფორმაციისათვის საჭირო ოპერაციათა ყველაზე მინიმალური ფასის მქონე კომბინაციის მოძიება. ასე რომ, დისტანცია წარმოადგენს ერთი რიგის მეორისაგან განსხვავების შემფასებელ სიდიდეს. ორივე შემთხვევაში საქმე გვაქვს რიცხვით სიდიდესთან, რომელიც ზომავს იმ ხარისხს, რა ხარისხითაც თანამიმდევრობები მსგავსია ან განსხვავებულია.

აქამდე ჩვენ ვიხილავდით მსგავსებას. ამ პარაგრაფში შევეხებით დისტანციის ცნებას და მის დამოკიდებულებას მსგავსებასთან. მრავალ შემთხვევაში ეს ორი საზომი დაკავშირებულია ერთმანეთთან საკმაოდ მარტივი ფორმულით, და აქედან გამომდინარე ადვილადაა შესაძლებელი ერთი საზომიდან მეორის მიღება. აქ ხაზი უნდა გაესვას იმას, რომ დისტანცია განხილული იქნება მხოლოდ გლობალური შედარების ჭრილში. დისტანცია გამოუსადეგია ლოკალური შედარების პროცესში.

დავიწყოთ მსგავსებისა და დისტანციის ზუსტი განმარტებით. დავუშვათ, რომ s და t წარმოადგენენ Σ ანბანით წარმოებულ ასოთა თანამიმდევრობებს. მართალია, როგორც წესი, ეს ანბანი ან ამინმუაჟურ ან დნმ-ს ანბანს წარმოადგენს, მიღებული შედეგები სამართლიანია ნებისმიერი სხვა ანბანის შემთხვევაში (ანუ ზოგადია). გავიხსენოთ, რომ ვეძებთ რიცხვს, რომელიც ზომავს თუ რამდენადაა ასოთა რიგები მსგავსი ან განსხვავებული. მსგავსებას კვლავ აღვნიშნავთ $sim(s, t)$ -ით, ხოლო დისტანციისათვის გამოვიყენებთ $dist(s, t)$.

მსგავსება

მსგავსების შეფასება, როგორც წესი, ეფუძნება ურთიერთსწორებას. მოდით ახლა გავიხსენოთ და დავაზუსტოთ ურთიერთსწორების განმარტება. s და t

თანამომდევრობათა შორის ურთიერთსწორებას უწოდებენ თანამომდევრობათა ისეთ (s', t') წყვილს, რომელიც მიღებულია შესაბამისად s -სა და t -საგან მათში გამოტოვებათა ჩამატების გზით. $\alpha(s', t')$ ურთიერთსწორება აკმაყოფილებს შემდეგ მოთხოვნებს:

1. $|s'| = |t'|$
2. s' -გან ყველა გამოტოვების ამოგდების შემთხვევაში მიიღება s .
3. t' -გან ყველა გამოტოვების ამოგდების შემთხვევაში მიიღება t .
4. არ არსებობს ისეთი i , რომ $s'[i]$ -ც და $t'[i]$ -ც ერთადროულად იყოს გამოტოვება.

ურთიერთსწორება ახდენს ერთსა და იმავე i პოზიციაში განთავსებული $s'[i]$ და $t'[i]$ სიმბოლოების დაკავშირებას. ამბობენ, რომ α -ს საშუალებით $s'[i]$ და $t'[i]$ სიმბოლოები ურთიერთსწორდნენ.

მსგავსება შეესაბამება შესაძლო ურთიერთსწორებათა შორის მაქსიმალურ ქულათა ჯამთან ვარიანტს. იგულისხმება, რომ საქმე გვაქვს ადიტიურ ქულათა დარიცხვის სისტემასთან. დავაზუსტოთ, თუ რას წარმოადგენს ქულათა დარიცხვის სისტემა. ქულათა დარიცხვის სისტემას ქმნის (p, g) წყვილი, რომელიც წარმოადგენს ურთიერთსწორებულ ასოთა წყვილზე ქულის დამრიცხავ $p: \Sigma \times \Sigma \rightarrow R$ ფუნქციას და გამოტოვებაზე დასარიცხ g ჯარიმას. როგორც წესი $g < 0$, მაგრამ ეს აუცილებელი არაა. ასეთი ქულათა დარიცხვის სისტემით შესაძლებელია თითოეული ურთიერთსწორების რაღაც რიცხვით, *ქულათა ჯამით*, დახასიათება. ყოველთვის, როცა α -ს მიხედვით a შესაბამისობაში მოდის b -თან ხდება $p(a, b)$ სიდიდის დარიცხვა, ხოლო როცა წყვილდება a და გამოტოვება, მაშინ ვარიცხავთ g -ს. ამგვარი პროცედურების შედეგად მიღებულ ჯამურ სიდიდეს მოიხსენიებენ როგორც α ურთიერთსწორების ქულათა ჯამს, და აღნიშნავენ *score* (α). ქულათა დარიცხვის ასეთი სისტემის შესაბამისი, ორ s და t თანამომდევრობას შორის *მსგავსება* ფასდება ასე

$$\text{sim}(s, t) = \max_{\alpha \in A(s, t)} \text{score}(\alpha) \tag{13.1}$$

სადაც $A(s, t)$ შეესაბამება s -სა და t -ს შორის ყველა შესაძლო ურთიერთსწორებათა სიმრავლეს. ქულათა დარიცხვის სისტემა ადიტიურია. ანუ თუ ნებისმიერ ურთიერთსწორებას დავანაწევრებთ ორ თანამომდევრულ ბლოკად, მაშინ სრული ურთიერთსწორების შესაბამისი ქულა ტოლი იქნება ცალკეული ბლოკების შესაბამის ქულათა ჯამისა.

დისტანცია

E სიმრავლეზე განსაზღვრული დისტანცია წარმოადგენს ისეთ $d: E \times E \rightarrow R$ ფუნქციას, რომ

1. $d(x,x) = 0$ ყველა $x \in E$ და $d(x,y) > 0$ თუ $x \neq y$
2. $d(x,y) = d(y,x)$ ყველა $x,y \in E$ (დისიმეტრიულია)
3. $d(x,y) \leq d(x,z) + d(y,z)$ ყველა x,y და $z \in E$

მესამე პირობა ცნობილია სამკუთხა უტოლობის სახელწოდებით. იგი მეტად სასარგებლოა მრავალი თვალსაზრისით და მას ეყარება მრავალი ალგორითმი. დისტანცია შესაძლებელია შეფასდეს Σ ანბანის საფუძველზე ჩამოყალიბებულ ასოთა ნებისმიერი ორი რიგისათვის, თუ შევფასებთ ოპერაციათა იმ რიცხვს, რომელიც საჭიროა ერთი რიგის მეორეში სრული ტრანსფორმირებისათვის. საჭირო ოპერაციების წარმატებული შერჩევის საფუძველზე შესაძლებელია ნებისმიერი რიგის სხვა რიგში ტრანსფორმირება. თუ მოვახდენთ თითოეული საჭირო ოპერაციის შეფასებას და ამის საფუძველზე ვიპოვოთ ტრანსფორმაციისათვის საჭირო მინიმალური ფასს, შევძლებთ მოცემულ ორ რიგს შორის დისტანციის განსაზღვრას. საჭირო ოპერაციები შეიძლება იყოს:

1. a ასოს b ასოთი ჩანაცვლება.
2. ნებისმიერი ასოს ჩამატება ან ამოგდება.

ასეთი ოპერაციების შესაფასებლად გამოიყენება (c,h) შემფასებელი სისტემა, სადა c წარმოადგენს $c: \Sigma \times \Sigma \rightarrow R$ ფუნქციას, ხოლო h კი ნამდვილ რიცხვს. a -ს b -თი ჩანაცვლება ფასდება $c(a,b)$ ფასით, ასოს ჩამატება ან ამოგდება კი h -ით. ოპერაციათა σ სერიის სრული ფასი აღინიშნება $cost(\sigma)$ -ით და წარმოადგენს ცალკეულ ოპერაციათა ფასების ჯამს. ამრიგად, შეფასების მოცემული სისტემით შეფასებული დისტანცია ორ s და t თანამიმდევრობას შორის შეიძლება შემდეგნაირად ჩაიწეროს

$$dist(s,t) = \min_{\sigma \in S(s,t)} cost(\sigma) \quad (13.2)$$

სადაც $S(s,t)$ წარმოადგენს s -ის t -ში ტრანსფორმაციის ოპერაციათა ყველა შესაძლო სერიების სიმრავლეს.

ზემოთ ჩამოყალიბებული პირობების დასაკმაყოფილებლად (c,h) -ს უწესდება შეზღუდვების მთელი რიგი. უპირველეს ყოვლისა საქმე უნდა ვიქონიოთ მხოლოდ არაუარყოფით ფასებთან, რადგანაც წინააღმდეგ შემთხვევაში აზრს კარგავს მინიმუმის მოძიება. ასევე საჭიროა, რომ შემფასებელი ფუნქცია c იყოს სიმეტრიული, წინააღმდეგ შემთხვევაში დისტანცია ვერ იქნება სიმეტრიული. ასეთი მინიმალური შეზღუდვები უკვე იძლევა იმის გარანტიას, რომ $dist(s,t)$ იქნება სიმეტრიული და დააკმაყოფილებს სამკუთხა უტოლობას. ამასთან დისტანციის განსაზღვრისას საჭიროა გავითვალისწინოთ, რომ $a \neq b$ -სთვის $c(a,b) > 0$ და ასევე $h > 0$. ეს მოთხოვნები თავიდან აგვაცილებენ ისეთ შემთხვევას, როცა s

და t არ წარმოადგენს იდენტურ თანამიმდევრობებს, მაგრამ $dist(s, t) = 0$. ჩვენი მოთხოვნაა, რომ დისტანცია მხოლოდ მაშინ გაუტოლდეს ნულს, როცა თანამიმდევრობები იდენტურია. ასევე საჭიროა, რომ c აკმაყოფილებდეს სამკუთხა უტოლობას, ანუ

$$c(x, y) \leq c(x, z) + c(y, z) \quad (13.3)$$

მაგალითად: თუ გამოვიყენებთ

0 როცა $x=y$

$$c(x, y) = \begin{cases} 0 & \text{როცა } x=y \\ 1 & \text{როცა } x \neq y \end{cases}$$

და $h=1$, მივიღებთ ორ თანამიმდევრობას შორის ე.წ. *სარედაქციო დისტანციას*, რომელიც ასევე ცნობილია ლევინშტეინის დისტანციის სახელწოდებით

მსგავსებასა და დისტანციას შორის ურთიერთდამოკიდებულება

ოპერაციათა შეფასების (c, h) სისტემისა და M კონსტანტის კონკრეტული მნიშვნელობის შემთხვევაში შესაძლებელია ურთიერთსწორების ქულათა დარიცხვის (p, g) სისტემის შემდგენიარად წარმოდგენა:

$$p(a, b) = M - c(a, b) \quad (13.4)$$

$$g = -h + M/2 \quad (13.5)$$

ამ პირობებში, ორ s და t თანამიმდევრობათა შედარებისას წარმოებულ მსგავსებასა და დისტანციას შორის კავშირი მართებულია წარმოვადგინოთ შემდეგი თანაფარდობის სახით

$$sim(s, t) + dist(s, t) = (M/2)(m+n) \quad (13.6)$$

სადაც $m=|s|$ და $n=|t|$.

ჩვენ აქ არ შევუდგებით ამ ტოლობის დამტკიცებას. დაინტერესებულ პირებს ვურჩევთ გაეცნონ ქვეპარაგრაფს „მსგავსებასა და დისტანციას შორის დამოკიდებულება“. დამოკიდებულების სამართლიანობის დამტკიცება“ (მოცემულია პარაგრაფის ბოლოს წვრილი შრიფტით).

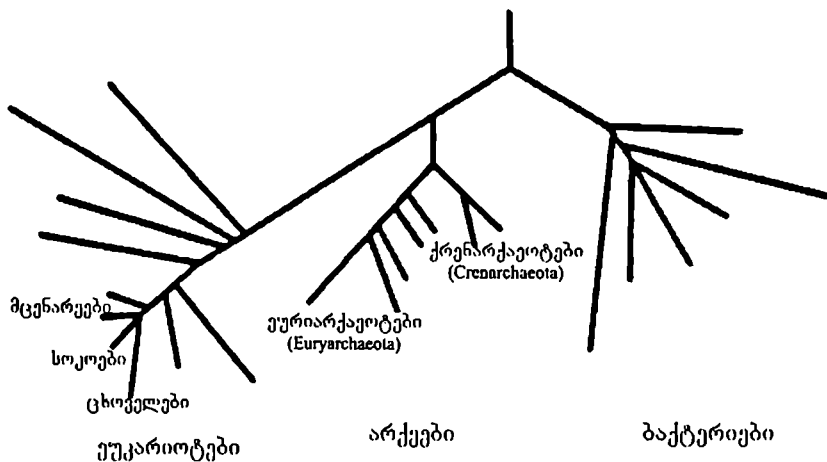
ფილოგენეტიკური ხე

ფილოგენეზი ბერძნულად ნიშნავს სახეობათა წარმოქმნას. ფილოგენეზის საფუძველს წარმოადგენს ევოლუციური პროცესი. ევოლუციურ ურთიერთდამოკიდებულებათა განმსაზღვრელი ძირითადი პრინციპი მდგომარეობს შემდეგში: მსგავსების საფუძველს წარმოადგენს საერთო წინაპრის ყოლა. ამ პრინციპის გადმოტანა შესაძლებელია ბიომაკრომოლეკულების (ცილისა თუ დნმ-ს) დონეზეც. თუმცა არსებობს გამონაკლისებიც, რომლებიც არ ექვემდებარება ამ პრინციპს. მაგალითად გამოდგება კონვერგენტული ევოლუციის პროცესი. მიუხედავად გამონაკლისებისა, უკანასკნელი პრინციპის მნიშვნელობა, როგორც თანამედროვე დაკვირვებათა ინტერპრეტირების, ისე ცოცხალი სამყაროს ისტორიული პროცესის გაანალიზებისთვის, დიდია. ფილოგენეტიკის მიზანია სახეობებს, პოპულაციებს, ინდივიდებს, გენებს, ზოგადად ტაქსონომიურ ჯგუფებს შორის ევოლუციურ ურთიერთმიმართებათა გარკვევა. ურთიერთმიმართება ვლინდება ნათესაობა-მსგავსების ანუ გენეოლოგიის სახით და შეიძლება წარმოდგენილ იქნეს შთამომავლობების საერთო წინაპართან და ერთმანეთთან ურთიერთკავშირის აღმწერი სქემით. სქემა, როგორც წესი, წარმოადგენს განშტოებად დიაგრამას - კლადოგრამას, რომელიც, მსგავსად საგვარეულო ხისა, აღწერს ნათესაურ კავშირებს. არ არის აუცილებელი კლადოგრამის შტოთა სიგრძეები ორგანიზმთა ან თანამიმდევრობათა შორის ევოლუციური განვითარების გზის მონაკვეთის სიგრძის პროპორციული იყოს. ფილოგენეტიკური ხე წარმოადგენს კლადოგრამის სპეციალურ სახეს, სადაც შტოების სიგრძეები პროცესის ხანგრძლივობისა. ამრიგად შეიძლება ითქვას, რომ ფილოგენეტიკური ხე აღწერს ევოლუციურ ურთიერთმიმართებას ტაქსონომიურ ჯგუფებს შორის. ტაქსონომია გულისხმობს ორგანიზმთა კლასიფიცირების სისტემას, რომელიც მსგავსების შეფასების საფუძველზე ახდენს მათ კატეგორიების მიხედვით დაჯგუფებას. ცოცხალი სამყაროს ძირეული ტაქსონომიური სურათის ამსახველი ფილოგენეტიკური ხე, რომელიც რიობოსომულ რნმ-თა თანამიმდევრობების ანალიზის (ურთიერთშედარების) საფუძველზეა აგებული, წარმოდგენილია სურათ 13.1-ზე.

ბიოლოგიურ კლასიფიკაციასა და ფილოგენეზთან დაკავშირებული ცნებები

ტერმინი ჰომოლოგიურობა იხმარება საერთო წინაპრის მქონე შთამომავლებს შორის ურთიერთმიმართების დასახასიათებლად.

ტერმინი მსგავსება იხმარება მსგავსებისა თუ განსხვავების დონის დახასიათების პროცესში და ახასიათებს მსგავსებას დამოუკიდებლად იმისა, თუ რა არის ამ მსგავსების წარმოშობის მიზეზი (არსებობს თუ არა საერთო წინაპარი).



სურათი 13.1 ცოცხალი სამყაროს ბირეული ტაქსონომიის აღმწერი ფილოგენეტიკური ხის გამარტივებული ფორმა (სრული ფორმა იხილეთ Woese C R, Kandler O, Wheelis M L-ის 1990 -ის ნაშრომში). ხე აგებულია რიბოსომულ რნმ თანამიმდევრობათა შედარების შედეგად მიღებულ მონაცემებზე დაყრდნობით.

მსგავსება ეყრდნობა დღეისთვის არსებულ დაკვირვების შედეგად მიღებულ მონაცემებს და არა ჰიპოტეზებს. ხოლო, ჰომოლოგიურობის შეფასებისას კი, აუცილებელია მოვიძიოთ ისტორიული წინაპირობები, რომლებიც, როგორც წესი, არაა დაკვირვების ობიექტები.

ცნება კლასტერიზაცია გულისხმობს მსგავსებულ ობიექტთა თავმოყრას, ანუ იმ ობიექტთა კლასის გამოყოფას, რომლებიც უფრო მსგავსია ერთმანეთის, ვიდრე სხვა კლასში გაერთიანებული ობიექტების. მსგავსების დონის განსაზღვრა შეიძლება მივიჩნიოთ ობიექტურ პროცესად, მაშინ, როცა კლასტერიზაცია ცალსახად სუბიექტურია.

ცნება იერარქიული კლასტერიზაცია გულისხმობს კლასტეთა კლასტერების, მათი კლასტერების და ა.შ. ფორმირებას.

ფილოგენეტიკური წარმოდგენა გულისხმობს ბიოლოგიური ურთიერთკავშირების აღწერას, რომელიც, როგორც წესი, გამოიხატება ხის სახით.

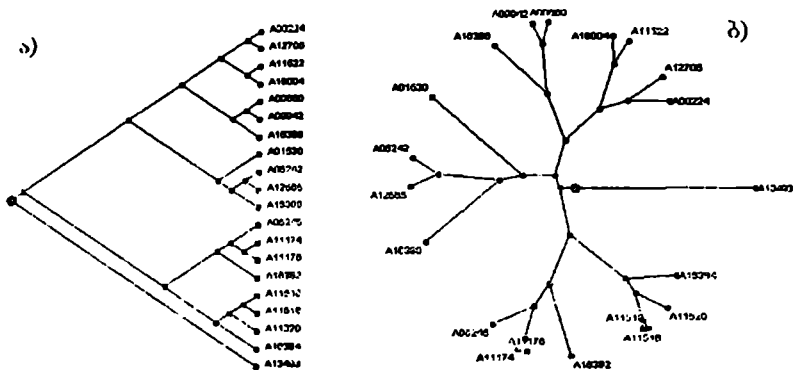
ობიექტებს შორის ფილოგენეტიკური წარმოდგენის ფორმირება გულისხმობს ჰომოლოგიურობის შეფასებებს და დამოკიდებულია კლასტეროზაციის სახეზე. ფილოგენეტიკური წარმოდგენის ტოპოლოგიას განსაზღვრავს მსაგავსების შესაფასებლად შერჩეული მახასიათებლები და ევოლუციური პროცესის ასაღწერად გამოყენებული მოდელი. ფილოგენეტიკურ წარმოდგენათა სურათები, შეფასებული სხვადასხვა მახასიათებლების საშუალებით, უნდა იყოს თანხვედრაში და ავსებდეს ერთმანეთს. თუ სხვადასხვა მახასიათებლებს მივყავართ განსხვავებულ ფილოგენეტიკურ ურთიერთკავშირამდე, მათი ქმედითუნარიანობა საეჭვოა. ამასთან, მსგავსებათა აღმწერი ერთი და იგივე მონაცემი შესაძლებელია წარმოვადგინოთ სხვადასხვა ტოპოლოგიის მქონე ხით. ხეს, რომლის მიხედვითაც ყველა შთამომავალს ერთადერთი საერთო წინაპარი ყავს, უწოდებენ დაფესვილ ხეს. მაგალითად გამოდგება სურათი 13.2ა. ალტერნატივას წარმოადგენს ხე, რომელიც ურთიერთკავშირების განსაზღვრისას არ გულისხმობს ისტორიულ ჭრილში განლაგებას. ამ უკანასკნელს უწოდებენ უფესვო ხეს (იხ. სურათი 13.2ბ). ხეზე ურთიერთკავშირები შეიძლება გამოხატული იყოს მხოლოდ დამაკავშირებელი მონაკვეთებითა და მათი განლაგებით (ტოპოლოგიით), ხოლო დამაკავშირებელ მონაკვეთთა სიგრძე არ იყოს ინფორმაციის მატარებელი, ისე, როგორც ეს ზემოთ მოყვანილ მაგალითებშია მოცემული. მაგრამ შესაძლებელია მეტი ინფორმაციის მომცველი ხის კონსტრუირებაც, სადაც ტაქსონებს (ხის წვეროებს) შორის მანძილი აღწერილი იქნება რაოდენობრივად. ვთქვათ, დამაკავშირებელი მონაკვეთები-სათვის (წიბოებისათვის) მითითებული რიცხვები შესაბამისობაშია საერთო წინაპრისგან დივერგირების დროსთან (ევოლუციურ მანძილთან).

ფილოგენეტიკური ხე

როგორც ზემოთ აღინიშნა ფილოგენეტიკური კავშირების წარმოდგენა ხდება ე.წ. ხის სახით. ხეს განიხილავენ, როგორც მათემატიკური გრაფის ერთ-ერთ ნაირსახეობას.

გრაფი განიმარტება, როგორც სტრუქტურა, რომელიც შედგება ერთმანეთთან წიბოებით (წრფის მონაკვეთებით) დაკავშირებული წვეროებისაგან (წარმოსახვითი წერტილებისაგან). ერთი წვეროდან მოერემდე ბილიკი წარმოადგენს ამ წვეროებს შორის განლაგებულ თამანიდევრულ წიბოთა ერთობას.

დაკავშირებული გრაფი წარმოადგენს გრაფს, რომლის ორ ნებისმიერ წვეროს შორის არსებობს ერთი ბილიკი მანინგ.



სურათი 13.2. ა) დაფესვილი და ბ) უფესვო ფილოგენეტიკური ხეები. ხეები აღწერენ *e.coli* -ს 20 შტამისაგან შემდგარი ერთი და იგივე კომპლექტის ტაქსონომიურ სურათს. ხეები აგებულია ბაქტერიულ დნმ თანამიმდევრობათა შედარების მონაცემებზე დაყრდნობით.

ასეთ განმარტებებზე დაყრდნობით ხე შეიძლება განვსაზღვროთ შემდეგნაირად: ხე წარმოადგენს დაკავშირებულ გრაფს, რომლის ორ ნებისმიერ წვეროს შორის არსებობს მხოლოდ ერთი ბილიკი. ხე შეიძლება ისე იქნეს აგებული, რომ რომელიღაც წვერო გამოვიდეს ფესვის როლში, თუმცა ეს არაა აუცილებელი. აგებენ როგორც დაფესვილ ისე უფესვო ხეებს. უფესვო ხე გვიჩვენებს ურთიერთკავშირთა ტოპოლოგიას, მაგრამ არ ასახავს ისტორიულ (წინაპართან მიმართებაში) დამოკიდებულებებს.

დაფესვილი ხე, რომლის თითოეულ წვეროს გააჩნია ორი შთამომავალი, მოიხსენიება ბინარულ ხედ.

გრაფთა სხვა სახეს წარმოადგენს ვექტორული გრაფი, რომლის შემთხვევაშიც თითოეულ წიბოს ახასიათებს მიმართულება და გრაფის გასწვრივ მხოლოდ ამ მიმართულებით შეიძლება მოძრაობა. დაფესვილი ფილოგენეტიკური ხე შეიძლება განხილულ იქნეს, როგორც ვექტორული გრაფი, რადგანაც სწორედ წინაპარ-შთამომავლობითი ურთიერთკავშირები განაპირობებს თითოეული წიბოს მიმართულებას.

წიბოს სიგრძე გამოიხატება თითოეული წიბოსათვის მიწერილი რიცხვით და განსაზღვრავს მოცემული წიბოთი დაკავშირებულ წვეროებს შორის მანძილს.

ბილიკის სიგრძე წარმოადგენს ბილიკის შემქმნელ წიბოთა სიგრძეების ჯამს.

ფილოგენეტიკური ხის შემთხვევაში წიბოს სიგრძე გვაძლევს ინფორმაციას როგორც ორ სახეობას შორის განსხვავების, ისე მათი დამოუკიდებელი ევოლუციონირების დაწყების მომენტიდან განვლილი დროის შესახებ. აქვე აუცილებლად უნდა აღინიშნოს, რომ დაშვება იმის თაობაზე, რომ სახეობათა შორის განსხვავებანი განმსაზღვრელია დივერგირების დროისა, სწორი იქნება იმ შემთხვევაში, თუ დივერგირების სიჩქარე ევოლუციური პროცესის სხვადასხვა ეტაპზე, ანუ ხის სხვადასხვა შტოს შემთხვევაში, ერთნაირია. გამომდინარე იქიდან, რომ რიგ მონაცემებზე დაყრდნობით იკვეთება დივერგირების სიჩქარის ცვლილების ფაქტი, დივერგირების დროის განმსაზღვრელად უშუალოდ განსხვავებათა მახასიათებლების გამოყენება არაა გამართლებული.

ფილოგენეტიკური ხის ფორმირების მეთოდები

თუ მივმართავთ უხეშ დაყოფას, ფილოგენეტიკური ხის ფორმირების მიდგომები შეიძლება დაყოფილ იქნეს ორ ტიპად. პირველი ტიპის მიხედვით ურთიერთკავშირების განსაზღვრისას არაა გათვალისწინებული რაიმე ისტორიული ინფორმაცია. ხის ფორმირებისას ფასდება მხოლოდ სახეობათშორისი დისტანციები და ამ დისტანციებზე დაყრდნობით ხორციელდება იერარქიული კლასტერიზაციის პროცედურა. ასეთი ტიპის მიდგომებს უწოდებენ ფენეტიკურს (*phenetic* - მსგავსებაზე დაფუძნებული კლასიფიკაცია). მათ ალტერნატივად შეიძლება წარმოვადგინოთ მეორე ტიპი, რომელსაც უწოდებენ კლადისტიკურ (*cladistic*-ქრონოლოგიაზე, ევოლუციურ ისტორიაზე დაფუძნებულ) მიდგომებს. კლადისტიკური მიდგომები, ეფუძნება რა ევოლუციის რაღაც შესაძლო გზას, აკეთებს გარკვეულ დაშვებას თითოეული წევრის შესაძლო წინაპრის თაობაზე და არჩევს ევოლუციურ ცვლილებათა გარკვეული მოდელის შესაბამის ოპტიმალურ ხეს. ამრიგად, ფენეტიკური მიდგომები ეფუძნება მსგავსებას, კლადისტიკური, კი გენეოლოგიას.

კლასტერიზაციაზე დაფუძნებული მეთოდები.

ფენეტიკური, ანუ კლასტერიზაციაზე დაფუძნებული, ფილოგენეტიკური ურთიერთკავშირების ამსახველი მიდგომები არაისტორიული ბუნებისაა. იერარქიული კლასტერიზაციით ხის ფორმირება ევოლუციური ურთიერთკავშირების შესახებ ცოდნის არ არსებობის პირობებშიცაა შესაძლებელი.

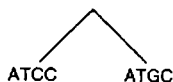
უმარტივესი კლასტერიზების პროცესი მიმდინარეობს შემდეგნაირად: ვთქვათ, გვინდა მოვახდინოთ სახეობათა გარკვეული სიმრავლის (კომპლექტის) კლასტერიზაცია, პირველ რიგში კომპლექტის თითოეული წევლისათვის უნდა განისაზღვროს მსგავსების ან განსხვავების დონე. მსგავსების ან განსხვავების შესაფასებლად შეიძლება გამოყენებულ იქნეს სრულიად განსხვავებული მახასიათებლები, იმისდა მიხედვით, თუ რა ობიექტების კლასტერიზაციას ვახდენთ. ვთქვათ, მიტოქონდრიული დნმ-ების კლასტერიზაციის პროცესში მოხდება დნმ-ს თანამიმდევრობათა ურთიერთსწორება და მსგავსება-განსხვავება შეფასდება ურთიერთსწორებაში მოსული ფუძეების მსგავსება-განსხვავების საფუძველზე. თუ ხის ფორმირებას არამსგავსებებზე დაფუძნებით გადავწყვეტთ, პირველ რიგში მოგვიწევს ორი ყველაზე მსგავსი სახეობის (წვერის) შერჩევა და მათი საერთო წინაპრის შესაბამისი წვეროს ფორმირება. ამის შემდგომ ორი უკანსკნელი სახეობა უნდა ჩავანაცვლოთ ერთი ჰიბრიდული სახეობით, რომელიც ორი მანამდე ცალკე არსებული სახეობის შერწყმის შედეგია. მანძილები, რომელიც მოცემული იყო ორი ცალკეული სახეობებისათვის, უნდა ჩავანაცვლოთ მანძილებით, რომლებიც გამოითვლება ერთ ახლადფორმირებულ ჰიბრიდულ სახეობამდე და რომელიც წარმოადგენს ცალკეულ სახეობებამდე მანძილების საშუალოს. ანუ ვიხილავთ წვეილურ არამსგავსებებს არა მარტო ინდივიდუალურ სახეობებს შორის, არამედ ჰიბრიდულ სახეობასა და ინდივიდუალურ სახეობებს შორისაც. ანალოგიური პროცესი გაგრძელდება მანამ, სანამ არ მივიღებთ კლასტერთა სრულ ჩამონათვალს. თვალსაჩინოების მიზნით კლასტერიზაციის პროცესს დავაკვირდეთ სახეობათა კონკრეტული კომპლექტის კლასტერიზაციის მაგალითზე.

მაგალითი 1

ოთხი სხვადასხვა სახეობის როლში განვიხილოთ ოთხი ჰომოლოგიური ნუკლეოტიდური თანამიმდევრობა ATCC, ATGC, TTCG, და TCGG და კლასტერიზაციის პროცედურაზე დაფუძნებით ავაგოთ ფილოგენეტიკური ხე. თანამიმდევრობათა შორის განსხვავებათა რიცხვი გამოვიყენოთ მათ შორის არამსგავსების მახასიათებელ სიდიდედ და ჩამოვაყალიბოთ დისტანციათა მატრიცა

	ATCC	ATGC	TTCG	TCGG
ATCC	0	1	2	4
ATGC		0	3	3
TTCG			0	2
TCGG				0

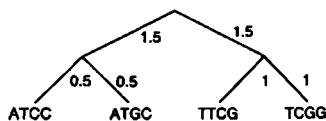
რადგანაც მატრიცა სიმეტრიულია, ვავსებთ მხოლოდ მის ზედა მარჯვენა სამკუთხედს. ყველაზე ახლო მანძილით 1-ით ხასიათდება ATCC და ATCG. მათ საფუძველზე ფორმირდება პირველი კლასტერი {ATCC, ATCG} და შესაბამისად, ხის ფრაგმენტი



დისტანციათა მატრიცაში განხორციელდება ცვლილება და იგი მიიღებს შემდეგ სახეს:

	{ATCC, ATGC}	TTCG	TCGG
{ATCC, ATGC}	0	$\frac{1}{2}(2+3) = 2.5$	$\frac{1}{2}(4+3) = 3.5$
TTCG		0	2
TCGG			0

შემდეგ კლასტერად წარმოგვიდგება {TTCG, TCGG}, დისტანციით 2. და ბოლოს, თუ გავაერთიანებთ {ATCC, ATGC} და {TTCG, TCGG} კლასტერებს მივიღებთ ხის საბოლოო სახეს:



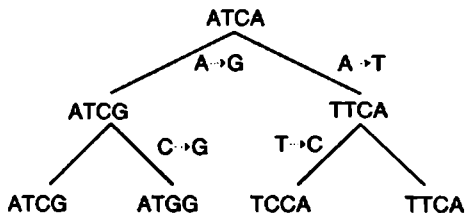
შტოთა სიგრძეების მინიჭებისას გამოიყენებოდა შემდეგი წესი: x და y წვეროების დამაკავშირებელი წიბოების სიგრძე ტოლია x და y შორის დისტანციის ნახევრისა.

ხის აგების ასეთ მიდგომას უწოდებენ UPGMA (Unweighted Pair Group Method with Arithmetic mean) მეთოდს.

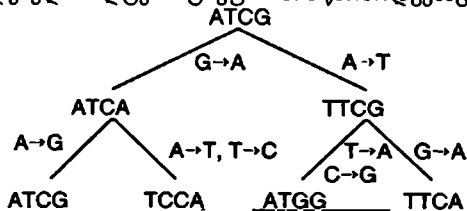
კლადისტიკური მეთოდები

კლადისტიკური მეთოდით განიხილება და ირჩევა გენეოლოგიათა (ანუ წარმოშობის ფორმებისა) და შესაბამის ხეთა ის ვარიანტები, რომლებიც მისაღებია ტაქსონთა მოცემული კომპლექტისათვის. შესაბამისად, მეთოდის მიზანს წარმოადგენს ევოლუციური პროცესის გარკვეულ მოდელზე დაფუძნებით სწორი ხის შერჩევა. კლადისტიკურ მეთოდებს შორის ყველაზე პოპულარულია მაქსიმალურ ეკონომიურობასა და მაქსიმალურ ალბათურობაზე დაფუძნებული მიდგომები. ორივე ეს მიდგომა ეფუძნება თანამიმდევრობათა შესახებ მონაცემებს, კერძოდ კი, ერთდროულ მრავალწევრა ურთიერთსწორებას.

მაქსიმალურ ეკონომიურობაზე დაფუძნებული W. Fitch-ის მიერ შემუშავებული მეთოდის მიხედვით წარმოებს მინიმალური რაოდენობა მუტაციების დამშვები ხის მოძიება. ასე მაგალითად, თუ გვაქვს ჰომოლოგიური თანამიმდევრობები ATCG, ATGG, TCCA და TTCA, მათთვის შესაძლებელია ისეთი ხის (იხ. სურათი)



აგება, რომელიც უშვებს ოთხი მუტაციის არსებობას. ზემოთ მოყვანილ თანამიმდევრობათათვის შესაძლებელია ალტერნატიული ხის წარმოდგენაც (იხ. სურათი)



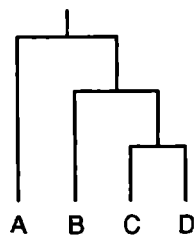
უკანასკნელი ხის შემთხვევაში უკვე საქმე გვაქვს შვიდ მუტაციასთან. ამასთან, იგი გულისხმობს, რომ G→A მუტაციას მოეთხე პოზიციასში ადგილი აქვს ორ დამოუკიდებელ შემთხვევაში. მაქსიმალურ ეკონომიურობაზე დაფუძნებული მიდგომის შესაბამისად წინა (პირველი) ხე წარმოადგენს ოპტიმალურს, რადგანაც სხვა, უფრო ნაკლებ რაოდენობა მუტაციაზე დაფუძნებული ხე არ მოიძებნება. შესაძლებელია შემთხვევა, როცა რამდენიმე ხე ხასიათდება მუტაციათა

ერთნაირი რაოდენობით. ასეთ შემთხვევაში მაქსიმალურ ეკონომიურობაზე დაფუძნებული მეთოდით ვერ ვიღებთ ცალსახა ამონახსნს.

მაქსიმალურ ალბათურობაზე დაფუძნებული მეთოდი მუტაციათა თითოეულ ტიპს ახასიათებს კონკრეტული რიცხვითი მახასიათებლით, რომელიც უკავშირდება ამ მუტაციის მოხდენის ალბათობას და არ შემოიფარგლება მათი რიცხვის უბრალო შეფასებით. აქ, ისევე როგორც მაქსიმალურ ეკონომიურობაზე დაფუძნებული მიდგომის შემთხვევაში ხდება ხის თითოეული წვეროსათვის წინაპრის შერჩევა, ამასთან ნაგულისხმევი მუტაციებიდან გამომდინარე შტოს სიგრძეს მიეწერება მოვლენის მოხდენის ალბათობის შესაბამისი რიცხვი. ხის თითოეული შესაძლო ტოპოლოგიისათვის ხდება ჩანაცვლებათა სიბშირეების სიდიდეთა შერჩევა საკვლევი თანამიმდევრობების მიღების ალბათობების მაქსიმალური მნიშვნელობების მიღწევის მიზნით. ყველაზე დიდი ალბათობის შესაბამისი ხე წარმოადგენს ამონახსნს, ოპტიმალურ ხეს.

ევოლუციის სიჩქარის ცვლილებასთან დაკავშირებული პრობლემები

ვთქვათ A, B, C, და D სახეობათა კომპლექტს აღწერს შემდეგი ფილოგენეტიკური ხე:



ხე აგებულია არამსგავსებათა შემდეგ მატრიცაზე დაფუძნებით:

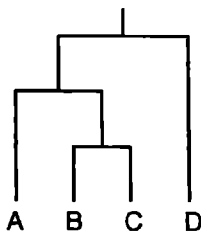
	A	B	C	D
A	0	3	3	3
B		0	2	2
C			0	1
D				0

ახლა კი დავუშვათ, რომ D-ს შესაბამისი მახასიათებლის ცვლილების სიჩქარე სხვა სახეობათა ანალოგიურ მახასიათებელთა ცვლილებებთან შედარებით, ძალზედ დიდია, მიუხედავად იმისა, რომ ფილოგენეზის სურათი ვიცი, რომ

იგივეა. ასეთ პირობებში ცხადია არამსგავსებათა (დისტანციათა) მატრიცა იცვლება და ვთქვათ, აქვს შემდეგი სახე

	A	B	C	D
A	0	3	3	20
B		0	2	20
C			0	20
D				0

რაც (შეცვლილი მატრიცა) თავის მხრივ, გამოიწვევს არასწორი ფილოგენეტიკური ხის ფორმირებას (იხ. სურათი)



თუ ევოლუციურ ცვლილებათა სიჩქარეები ხის სხვადასხვა შტოსათვის განსხვავდება, მაშინ ყველა ზემოთ წარმოდგენილ მეთოდში შეიძლება ადგილი ქონდეს ასეთ შეცდომას. იმის შესამოწმებლად, იცვლება თუ არა ევოლუციურ ცვლილებათა სიჩქარე საკვლევე კომპლექტში გაერთიანებულ სახეობათათვის, მოიშველიებენ ე.წ. „გარენათესავ“ (outgroup) სახეობას. გარენათესაობა გულისხმობს განსაზღვრულ სახეობებთან უფრო შორეულ ნათესაობას, ვიდრე განსაზღვრული კომპლექტის შიგნით არსებული ნათესაობაა. ასე მაგალითად, თუ შესწავლის საგანს წარმოადგენს პრიმატთა სახეობები, გარენათესავად შეიძლება განვიხილოთ რომელიმე არაპრიმატი ძუძუმწოვარი, ვთქვათ ძროხა. თუ პრიმატთა შიგნით ევოლუციურ ცვლილებათა სიჩქარე მუდმივია, ჩვენ უნდა ველოდოთ ერთი და იგივე რიგის (ზომის) არამსგავსებებს ძროხასა და პრიმატთა ყველა განხილულ სახეობას შორის. ზოგჯერ პრიმატთა შიგნით ადგილი აქვს ევოლუციურ ცვლილებათა სიჩქარის არამუდმივობას და მოცემული მახასიათებლით ფილოგენეტიკური ხის აგება გაუმართლებელია.

მსგავსებასა და დისტანციას შორის დამოკიდებულება. დამოკიდებულების სამართლიანობის დამტკიცება.

რედაქტირების ოპერაციათა ფასების (c, h) სისტემისა და M კონსტანტის კონკრეტული მნიშვნელობის შემთხვევაში შესაძლებელია ურთიერთსწორების ქულათა დარიცხვის (p, g) სისტემის შემდეგნაირად წარმოდგენა:

$$p(a, b) = M - c(a, b) \quad (13.7)$$

$$g = -h + M/2 \quad (13.8)$$

ვაჩვენოთ ასეთი დამოკიდებულების სამართლიანობა. ვთავაზოთ ორ s და t თანამიმდევრობათა ურთიერთსწორება წარმოიდგინება α -თი. შესაძლებელია ოპერაციათა ისეთი σ სერიის მოძიება, რომ სამართლიანი იყოს შემდეგი:

$$score(\alpha) + cost(\sigma) = (M/2)(m+n) \quad (13.9)$$

სადაც $m=|s|$ და $n=|t|$. σ -ს შეფასების მიზნით α ურთიერთსწორება დავანაწევროთ სვეტებად. თითოეულ სვეტს შევუსაბამოთ მისთვის ზუნებრივად დასაშვები ოპერაცია. ასე მაგალითად, ასოთა თანხვედრას შევუსაბამოთ ჩანაცვლების ოპერაცია, გამოტოვებებსა და/ან ჩამატებებს კი, დელეტიის ოპერაციები. ოპერაციათა გამოყენების რიგი ნებისმიერია, გამომდინარე იმ დამებიდან, რომ ისინი შეესაბამება უბნებს, რომლებსაც არა აქვთ გავლენა ერთმანეთზე.

ამრიგად σ ვუწოდებთ თანამიმდევრულ (მარცხნიდან მარჯვნივ) ოპერაციათა სერიას. განვსაზღვროთ α -ს ქულათა j -ში და σ -ს ფასი. დავუშვათ, რომ α -ში გვაქვს შესაბამისობაში მოყვანილ ასოთა ზუსტად l რაოდენობა. ვთქვათ შესაბამისობაში მოყვანილი i -ური წყვილი (სვეტი) მოიცავს a -ს s -დან და b -ს t -დან. გარდა ამისა დავუშვათ, რომ α -ში გვხვდება ზუსტად r გამოტოვება. α - ს ქულათა j -ში მისთვის ჩვენ შეგვიძლია დავწეროთ:

$$score(\alpha) = \sum_{i=1}^l p(a_i, b_i) + rg \quad (13.10)$$

მეორეს მხრივ σ -ს ფასი

$$cost(\sigma) = \sum_{i=1}^l c(a_i, b_i) + rh \quad (13.11)$$

ამ ტოლობათა წევრობრივი შეკრებითა და (13.7) და (13.8) თანაფარდობების გათვალისწინებით მივიღებთ

$$score(\alpha) + cost(\sigma) = lM + rM/2 \quad (13.12)$$

შეგნიშნოთ, რომ l და r სიდიდეები არ არის დამოუკიდებელი. ისინი უნდა აკმაყოფილებდეს განტოლებას, რომელშიც ასევე ფიგურირებს თანამიმდევრობების ასოთა რაოდენობები. მართლაც, თითოეული პოზიციური თანხვედრა გულისხმობს ორ ასოს,

მაშინ, როცა გამოტოვების შემთხვევაში საქმე გვაქვს მხოლოდ ერთ ასოსთან. ამიტომ ასოთა სრული რიცხვი $m+n$ უტოლდება

$$m + n = 2l + r \quad (13.13)$$

ყოველივე ამის გათვალისწინებით (13.12) -ის გადაწერა შეიძლება ასე

$$score(\alpha) + cost(\sigma) = (M/2) (m+n)$$

რაც ზუსტად შეესაბამება (13.9) განტოლებას.

ეს უკანასკნელი სამართლიანია ნებისმიერი ურთიერთსწორებისათვის. კერძოდ, იმ შემთხვევისათვისაც, როცა α ოპტიმალური ურთიერთსწორებაა. ასე რომ

$$sim(s,t) + cost(\sigma) = (M/2) (m+n)$$

სადაც σ α -სგან წარმოებულ ოპერაციათა სერიას წარმოადგენს (მაგრამ იგი შეიძლება არ შეესაბამებოდეს დისტანციას). და რადგანაც დისტანცია, s და t -ს შორის ტრანსფორმირების მინიმალურ ფასიან ოპერაციათა სერია უნდა იყოს, ჩვენ ვწერთ

$$sim(s,t) + dist(s,t) \leq (M/2) (m+n) \quad (13.14)$$

ახლა მოდით უკანასკნელი შედეგი ჩამოვაცალიოთ თეორემის სახით, რადგან შემდგომ სწორედ ამ თეორემას მივმართოთ ხოლმე.

თეორემა 13.1 მოცეული (c,h) -ისა და გარკვეული M -ით განსაზღვრულ ქულათა დარიცხვის შესაბამისი სისტემის პირობებში, ნებისმიერი s და t წყვილისათვის შეიძლება დავეწროთ

$$dist(s,t) + sim(s,t) \leq (M/2) (|s|+|t|)$$

ანუ 13.1 თეორემა განსაზღვრავს $dist(s,t)$ -ს ზედა საზღვარს. ვაჩვენოთ ახლა, რომ იგივე გამოსახულება შეიძლება ასევე შეესაბამებოდეს ქვედა საზღვარს და აქედან გამომდინარე დავასკვნით, რომ დისტანცია შესაძლებელია პირდაპირ იქნეს მიღებული მსგავსებიდან.

თეორემა 13.2. 13.2 თეორემით განსაზღვრულ პირობებში ასევე სამართლიანია შემდეგი უტოლობა

$$sim(s,t) + dist(s,t) \geq (M/2) (|s|+|t|)$$

დამტკიცება. მივმართოთ იმ პროცედურების ანალოგიურ პროცედურებს, რომლებმაც მიგვიყვანა 13.1 თეორემამდე. უპირველეს ყოვლისა აღვნიშნოთ, რომ დასაშვებ ოპერაციათა მოცეული σ სერიისათვის ჩვენ შეგვიძლია მოვძებნოთ ისეთი α ურთიერთსწორება, რომ

$$score(\alpha) + cost(\sigma) \geq (M/2) (m+n) \quad (13.15)$$

შევნიშნოთ, რომ წინა თეორემით განსაზღვრულ შემთხვევისაგან განსხვავებით ზოგადად აქ არ შეიძლება ვიხმართ ტოლობა. უკანასკნელი შემთხვევა მცირედ განსხვავებულია წინასაგან. ურთიერთსწორება გვიჩვენებს მხოლოდ გარკვეულ ჯამურ ეფექტს და მას შეიძლება შევესაბამოთ მრავალი განსხვავებული ოპერაციათა სერია. ასე მაგალითად, თუ

σ -ს რომელიმე ეტაპზე ხდება a -ს ჩამატება და შემდეგ იქვე მისი უკუგდება, ურთიერთსწორებისათვის ჯამური ეფექტი წულია. ამასთანავე ტრანსფორმაციის ოპერაციათა მოცემული სერიის შეფასებისას ვახდენთ ორივე ოპერაციაზე ერთი და იმავე სიდიდის დარიცხვას. ანუ დარიცხვა ხდება ოპერაციათა მთელი სერიისათვის, იმისდა მიუხედავად, ეფექტურია თუ არა ოპერაცია, და აქედან გამომდინარე ასეთ შემთხვევაში გვექნება უტოლობა და არა ტოლობა.

ახლა დავამტკიცოთ (13.15) თანაფარდობის სამართლიანობა ინდუქციის მეთოდით. თუ $|\sigma|=0$, ანუ არავითარ ოპერაციას არა აქვს ადგილი, შეგვიძლია დავწეროთ, რომ $s=t$. (s,s) ურთიერთსწორების შემთხვევაში ქულათა ჯამი განისაზღვრება როგორც $Mm = Mn = M(m+n)/2$ და ე.ი. ამ შემთხვევაში დასამტკიცებელი თანაფარდობა სამართლიანია.

თუ σ წარმოადგენს ერთ ან მეტ ოპერაციათა ერთობას, გამოვყოთ ბოლო ოპერაცია და დავწეროთ $\sigma = \sigma' u$. u ბოლო ოპერაციაა.

$$\begin{array}{ccc} \sigma' & u & \\ s \rightarrow t' & \rightarrow & t \end{array}$$

სადაც t' მიიღება σ -გან ბოლოს გარდა ყველა ოპერაციის განხორციელების ხარჯზე და t მიიღება t' -გან u ოპერაციით.

ჩვენი ინდუქციური ჰიპოტეზა მდგომარეობს შემდეგში, σ -სა და t' შორის გვაქვს ისეთი α' ურთიერთსწორება, რომ

$$score(\alpha') + cost(\sigma') \geq (M/2)(m+n)$$

სადაც $\sigma' = t'/t$. რადგანაც $cost(\sigma) = cost(\sigma') + cost(u)$, ერთადერთი რაც გვჭირდება არის s -სა და t -ს შორის ისეთი α ურთიერთსწორების მოძიება (თუ ეს შესაძლებელია), რომელიც დააკმაყოფილებს შემდეგ უტოლობას

$$score(\alpha) \geq score(\alpha') + (M/2)(n-n') - cost(u) \quad (13.16)$$

ამ ორი უტოლობის წევრობრივი შეკრება მიგვიყვანს სასურველ შედეგამდე. განხილულ უნდა იქნეს სამი შემთხვევა დასაშვებ u ოპერაციათა ტიპების შესაბამისად, რომლებიც, როგორც ვიცით, არის: ჩანაცვლება, ჩამატება და ამოგდება.

შემთხვევა 1: u წარმოადგენს ჩანაცვლებას.

ამ შემთხვევაში $cost(u) = c(a,b)$, $n=n'$ და t განსხვავდება t' -გან მხოლოდ ერთ პოზიციაში, როგორც ეს ნაჩვენებია ქვემოთ

$$\begin{array}{ccc} s & & \dots x \dots \\ t' & & \dots a \dots \\ t & & \dots b \dots \end{array}$$

α -ს მისაღებად საკმარისია α' -ში a ჩანაცვლდეს b -თი, რაც ზემოთ მოყვანილი სურათის შემთხვევაში გულისხმობს შუა სტრიქონის იგნორირებას. თუ x s რიგის სტანდარტული ასოა, მაშინ

$$\begin{aligned}
 \text{score}(\alpha) &= \text{score}(\alpha') - p(a,x) + p(b,x) \\
 &= \text{score}(\alpha') - M + c(a,x) + M - c(b,x) \\
 &= \text{score}(\alpha') + c(a,x) - c(b,x) \\
 &\geq \text{score}(\alpha') - c(a,b)
 \end{aligned}$$

უტოლობაზე გადასვლის უფლება მოგვცა c -სთვის სამკუთხა უტოლობის არსებობის ფაქტმა.

თუ x წარმოადგენს გამოტოვებას, α და α' ქულათა ჯამები ტოლია და შესაბამისად

$$\text{score}(\alpha) = \text{score}(\alpha') \geq \text{score}(\alpha') - c(a,b),$$

რადგანაც $c(a,b) \geq 0$ ჰიპოტეზის თანახმად. და ამით სასურველი ნაჩვენებია იმ შემთხვევისთვის როცა u ჩანაცვლება.

შემთხვევა 2: u წარმოადგენს ჩამატებას.

ამ შემთხვევაში $\text{cost}(u) = h$ და $n' = n - 1$. α' ურთიერთსწორება გარკვეულ პოზიციებში იძლევა ასოს ჩამატების საშუალებას, როგორც ეს ნაჩვენებია ქვემოთ:

$$\begin{array}{cccc|cccc}
 s & & \dots & & s & & \dots & & | & - & | & \dots \\
 t' & & \dots & & t & & \dots & & | & a & | & \dots
 \end{array}$$

s -ში ჩნდება დამატებითი გამოტოვება. აქედან გამომდინარე α -სა და α' -ს შორის ქულათა სხვაობა განპირობებულია α -ში გაჩენილი დამატებითი გამოტოვებით:

$$\begin{aligned}
 \text{score}(\alpha) &= \text{score}(\alpha') + g \\
 &= \text{score}(\alpha') - h + M/2
 \end{aligned}$$

და ამრიგად (12.16) თანაფარდობის სისწორე იმ შემთხვევაში, როცა u წარმოადგენს ჩამატებას ნაჩვენებია.

შემთხვევა 3. u წარმოადგენს ამოკლებას.

ამჯერად კვლავ $\text{cost}(u) = h$, მაგრამ $n' = n + 1$. ჩვენ საქმე გვაქვს შემდეგ სიტუაციასთან:

$$\begin{array}{cccc|cccc}
 s & & \dots & x & \dots & & & & & & & \\
 t' & & \dots & a & \dots & & & & & & & \\
 t & & \dots & - & \dots & & & & & & &
 \end{array}$$

განვასხვავებთ ორ შემთხვევას, იმის მიხედვით x წარმოადგენს გამოტოვებას თუ არა.

თუ x ასოა, მაშინ

$$\begin{aligned}
 \text{score}(\alpha) &= \text{score}(\alpha') - p(a,x) + g \\
 &= \text{score}(\alpha') - M + c(a,x) - h + M/2 \\
 &= \text{score}(\alpha') - h - M/2 + c(a,x) \\
 &\geq \text{score}(\alpha') - h - M/2
 \end{aligned}$$

რაც მოითხოვებოდა, გამომდინარე იქიდან რომ $c(a,x) \geq 0$.

თუ x გამოტოვება, შევიძლია დავწეროთ

$$\begin{aligned} \text{score}(\alpha) &= \text{score}(\alpha') - g \\ &= \text{score}(\alpha') - h + M/2 \\ &\geq \text{score}(\alpha) - h - M/2, \end{aligned}$$

რადგანაც $h > 0$.

ამით ინდუქციური ანალიზი დამთავრებულია და შეიძლება ითქვას, რომ (13.15) თანაფარდობის სამართლიანობა ნაჩვენებია. კერძოდ ეს უტოლობა სამართლიანია σ -ს მინიმალური მნიშვნელობისთვისაც

$$\text{score}(\alpha) + \text{dist}(s, t) \geq (M/2)(m+n)$$

თუ გავითვალისწინებთ, რომ $\text{sim}(s, t)$ წარმოადგენს s -სა და t -ს შორის ურთიერთსწორების შესაძლო მაქსიმალურ ქულათა ჯამს, მითუმეტეს გვექნება

$$\text{sim}(s, t) + \text{dist}(s, t) \geq (M/2)(m+n)$$

ამრიგად 13.2 თეორემა დამტკიცებულია.

13.1 და 13.2 თეორემებით დეკლარირებული უტოლობების ერთდროული შესრულება კი შესაძლებელია მხოლოდ, თუ

$$\text{sim}(s, t) + \text{dist}(s, t) = (M/2)(m+n) \quad (13.17)$$

ამრიგად დისტანციის დათვლა დაიყვანება მსგავსების შეფასებაზე. ერთადერთი რაც ჩვენ ამისათვის დაგვჭირდება, ესაა შესაბამისი M -ის შერჩევა, (13.7) და (13.8) თანაფარდობათა დახმარებით ქულათა დამრიცხავი პარამეტრების p და q -ს განსაზღვრა და გლობალური შედარების ერთ-ერთი ალგორითმის გამოყენება. ზემოთ მოყვანილი ფორმულის საშუალებით კი, შესაძლებელია მიღებული მსგავსების დისტანციაში გადაყვანა.

ასე მაგალითად, თუ განვიხილავთ სარედაქციო დისტანციის შესაფასებლად მიღებულ ქულათა დარიცხვის სისტემას (იხ. ზევი), M -სთვის შევიძლია გამოვიყენოთ 0, ხოლო მსგავსების ალგორითმის რეალიზების პროცესში თანხვედრას დავარიცხოთ 0, ართანხვედრას = -1 და გამოტოვებასაც = -1. ან ჩვენ შევიძლია ვიზმაროთ $M=2$ და თანხვედრა გაუტოლდება 2-ს, ართანხვედრა 1-ს და გამოტოვება კი 0-ს. ორივე ქულათა სისტემა მიგვიყვანს ერთსა და იმავე ოპტიმალურ ურთიერთსწორებამდე, თუმცა, განსხვავებული ქულათა ჯამით. თუმცა (13.17) ფორმულის გამოყენებით მივიღებთ ერთსა და იმავე დისტანციას.

14. მონაცემთა ბაზების სწრაფი ანალიზი

ნუკლეინის მჟავათა და ცილთა თანამიმდევრობების დადგენის პროცედურის სისწრაფისა და სანდოობის ზრდასთან ერთად იქმნება სულ უფრო მრავალფეროვანი ცენტრალიზებული მონაცემთა ბაზები, რომლებშიც ხდება მსოფლიოს სხვადასხვა სამეცნიერო ლაბორატორიებში გაშიფრულ თანამიმდევრობათა შენახვა. მონაცემთა ბაზები უზრუნველყოფილია მათში დაგროვილი ინფორმაციის ანალიზისათვის განკუთვნილი პროგრამათა პაკეტებით. მრავალფეროვანია იმ ბიოლოგიურ ამოცანათა სპექტრი, რომელთა გადაწყვეტაც უკანასკნელი პაკეტებით ხორციელდება. მაგალითისათვის შეიძლება მოვიყვანოთ ასეთი ტიპური ამოცანა: გვაქვს საკვლევი თანამიმდევრობა, რომელიც ლოკალური მსგავსების გამოვლენის მიზნით უნდა შედარდეს ბაზაში განთავსებულ ყველა თანამიმდევრობას. ასეთი კვლევის პროცესი გულისხმობს ათეულ ათასობით თანამიმდევრობათა შედარების ქვეპროცესს.

დინამიკურ პროგრამირებაზე დაფუძნებული, ორ თანამიმდევრობას შორის მსგავსების და ოპტიმალური ურთიერთსწორების განმსაზღვრელი მეთოდის მონაცემთა ბაზების ანალიზისათვის გამოყენება მოითხოვს დიდ სამანქანო დროს. შესაბამისად გაჩნდა აუცილებლობა ახალი და სწრაფი მეთოდების შექმნისა. ისინი როგორც წესი, ემყარებიან ევრისტიკულ მიდგომებს. მიუხედავად ამისა, მათზე დაფუძნებული პროგრამული პაკეტები მეტად ეფექტური და პოპულარულია.

ამ პარაგრაფში შევჩერდებით მონაცემთა ბაზების კვლევის ორ ყველაზე პოპულარულ მიდგომაზე. ორივე მიდგომა საწყის ეტაპზე ევრისტიკულია. ამასთან, ევრისტიკული მეთოდებით მიღებული მონაცემების საბოლოო ურთიერთსწორებამდე მისაყვანად ორივე იყენებს დინამიკური პროგრამირების მოდიფიცირებულ, ორიგინალურ ვარიანტს.

BLAST

BLAST წარმოადგენს პროგრამათა პაკეტს, რომლის სამუშაო (შემავალი) მონაცემების როლში ფიგურირებს ორი თანამიმდევრობა: საკვლევი (query ან target) თანამიმდევრობა და თანამიმდევრობათა ბაზა. BLAST თანამიმდევრობათა ბაზაში ნახულობს თანამიმდევრობათა ისეთ ქვერიგებს (ფრაგმენტებს), რომლებიც გარკვეული დონით მსგავსია საკვლევის კონკრეტული ქვერიგისა. როგორც წესი, საკვლევი თანამიმდევრობა გაცილებით პატარაა, ვიდრე მონაცემთა ბაზა; ასე მაგალითად, საკვლევი თანამიმდევრობის სიგრძე შეიძლება აღწევდეს ათას ნუკლეოტიდს, მაშინ როცა მონაცემთა ბაზა შეიძლება მოიცავდეს რამდენიმე მილიარდ ნუკლეოტიდს.

საკვლევსა და მონაცემთა ბაზაში გაერთიანებულ თანამიმდევრობებს შორის მაღალქულიანი ურთიერთსწორების ძიების პროცესში BLAST საწყის ეტაპზე

ემყარება ევრისტიკულ ალგორითმს, რომელიც ახდენს სმიტ-ვატერმანის ალგორითმის (ლოკალური ურთიერთსწორება) აპროქსიმირებას. სმიტ-ვატერმანის ბაზური ალგორითმის საშუალებით ძალიან ნელა ხდება ისეთი ამოცანის გადაწყვეტა, როგორცაა, მაგალითად სრული GenBank-ის კვლევა მსგავს თანამიმდევრობათა მოძიების მიზნით. ამიტომ BLAST ალგორითმი იყენებს ევრისტიკულ მიდგომას, რომელიც სმიტ-ვატერმანის ბაზურ ალგორითმზე ნაკლებად ზუსტია, მაგრამ ამასთანავე 50-ჯერ სწრაფია. სწორედ სისწრაფე განაპირობებს BLAST-ის დიდ პოპულარობას.

BLAST პროცედურაში შესაძლებელია გამოიყოს სამი ძირითადი საფეხური (ეტაპი):

- პირველ ეტაპზე საკვლევი და მონაცემთა ბაზის თანამიმდევრობების შედარების პროცესში ხდება ფიქსირებულ, W სიგრძის ასოთა ქვერიგებს შორის თანხვედრების გამოვლენა. ასე მაგალითად, თუ გვაქვს თანამიმდევრობები AGTTAC და ACTTAG და საძიებელი სიტყვის (ჩარჩოს) სიგრძე $W = 3$, BLAST პოულობს თანხვედრაში მყოფ ქვერიგებს TTA-ს. ქვერიგთა ასეთ წყვილებს მოიხსენიებენ პირველწყარობად. სტანდარტულ რეჟიმში, ნუკლეოტიდურ პირველწყაროთა ძიების პროცესში გამოიყენება $W = 11$.
- მეორე ეტაპი გულისხმობს თანხვედრილ ფრაგმენტთა სიგრძის ორივე მიმართულებით გაზრდას. ჩამატებათა და გამოტოვებათა გამოყენებლობის პირობებში პირველწყაროს ორივე მიმართულებით დაგრძელების მიზანია გაზარდოს ურთიერთსწორების ქულათა ჯამი. ქულათა ჯამის შეფასებისას გამოიყენება პოპულარული ჩანაცვლებათა მატრიცები. ზემოთ მოყვანილი მაგალითის შემთხვევაში, ანუ TTA საერთო სიტყვაზე დაფუძნებულ, AGTTAC-სა და ACTTAG-ს შორის გამოტოვებების გარეშე ურთიერთსწორებას ექნება შემდეგი სახე:

```

..AGTTAC..
   |||
   |||
..ACTTAG..

```

მონაცემთა ბაზის ის თანამიმდევრობები, რომელთათვისაც ასეთი ურთიერთსწორების საფუძველზე მოძიებულ იქნა მაღალქულიანი ფრაგმენტები ექვემდებარება დამუშავების მესამე ეტაპს.

- მესამე ეტაპზე საკვლევ თანამიმდევრობასა და პირველ ორ ეტაპზე შერჩეულ ბაზის თანამიმდევრობას შორის ხდება უკვე ჩამატებების პირობებში ურთიერთსწორება და ამისათვის გამოიყენება სმიტ-ვატერმანის ბაზური ალგორითმის მოდიფიცირებული ვარიანტი. და ბოლოს, მომხმარებელს მიეწოდება სტატისტიკურად დამაჯერებელი ურთიერთსწორების ვარიანტი.

დამაჯერებლობის შეფასება გულისხმობს მოცემული ურთიერთსწორების შესაბამისი ქულათა ჯამის შედარებას შემთხვევით თანამიმდევრობათა ურთიერთსწორებისას მოსალოდნელ ქულათა ჯამებთან. ამისათვის საჭიროა შემთხვევით თანამიმდევრობათა ურთიერთსწორების შესაძლო ქულათა ჯამების განაწილების ფუნქციის ცოდნა. შესაბამისად, შემთხვევით თანამიმდევრობათა შედარებისათვის ჩამოყალიბებულ იქნა გამოტოვებების გარეშე ლოკალური ურთიერთსწორებების ქულათა ჯამების განაწილების აღმწერი ანალიზური, ასიმპტოტური სტატისტიკური თეორია (Altschul SF და თანმშ. 1990). გასაგები რომ გახდეს BLAST-ის გამოსავალი პარამეტრების არსი, წარმოგიდგენთ თეორიის ზოგად პრინციპებს.

თეორიის თანახმად, ორი m და n სიგრძის შემთხვევითი თანამიმდევრობებისათვის, λ -ზე მეტი ან ტოლი ქულათა ჯამით დახასიათებულ ლოკალურ ურთიერთსწორებათა რიცხვის განაწილება აპროქსიმირდება პუასონის განაწილებით, და შესაბამისად, ამ რიცხვის მოსალოდნელი მნიშვნელობა გამოითვლება ასე:

$$E = Kmn e^{-\lambda S}, \tag{14.1}$$

სადაც K და λ გამოთვლადი სიდიდეებია და დამოკიდებულია ქულათა დარიცხვის სისტემასა და ამინმუჯვურ კომპოზიციაზე. პუასონის განაწილება გულისხმობს, რომ მაქიმალური ქულათა ჯამი შეესაბამება სიდიდეთა განაწილების ექსტრემალურ არეს და λ -ზე მეტი ან ტოლი ქულათა ჯამის მქონე ლოკალურ ურთიერთსწორების შეხვედრის ალბათობისათვის გვაძლევს შემდეგ დამოკიდებულებას:

$$P = 1 - e^{-E} = 1 - \exp(-Kmn e^{-\lambda S}). \tag{14.2}$$

ეს უკანასკნელი შეიძლება გადავწეროთ ასე

$$E = -\ln(1 - P). \tag{14.3}$$

გამოტოვების გარეშე განხორციელებული ლოკალური ურთიერთსწორების შემთხვევაში λ განისაზღვრება მხოლოდ ქულათა დარიცხვის ისეთი სისტემისთვის, როცა ქულათა სიდიდეების საშუალო მნიშვნელობა უარყოფითია. ასეთ პირობებში იგი წარმოადგენს ქვემოთ მოყვანილი განტოლების დადებით ამონახსნს

$$\sum_{i=1}^n p_i p'_i e^{a_i x} = 1, \tag{14.4}$$

სადგ s_j შესაბამება i -ს j -თან შესაბამისობაში მოყვანის შემთხვევაში დასარიცხ ქულას, p_i და p_j -ს, კი პირველ თანამიმდევრობაში i -ს, ხოლო მეორე თანამიმდევრობაში j -ს გამოჩენის ალბათობებია (შეხვედრის სიხშირეები), შესაბამისად. ემპირიული გამოთვლები აჩვენებს, რომ გამოტოვებათა დამშვები ოპტიმალური ლოკალური ურთიერთსწორებები ექვემდებარება იგივე განაწილების კანონს, თუმცა განაწილების პარამეტრთა λ და K -ს ანალიზურად გამოთვლა შეუძლებელია და შეფასება ხდება შემთხვევით რიცხვთა გენერატორების საშუალებით.

ურთიერთსწორებების ქულათა ჯამების შეფასებები გამოხატული E ან P სიდიდებით, დამოკიდებულია თანამიმდევრობათა სიგრძეებზე. E ან P სიდიდეები შესაძლებელია მოცემული იყოს ან წყვილური ურთიერთსწორების, ან მონაცემთა ბაზის კვლევის (სკანირების) კონტექსტში. მონაცემთა ბაზის სკანირების შემთხვევაში (14.1) და (14.2) ფორმულებში ცალკეული თანამიმდევრობის სიგრძე n ჩაინაცვლება აგრეგირებული სიგრძით N -ით, რომელიც წარმოადგენს მონაცემთა ბაზაში გაერთიანებულ ამინმჟავათა სრულ რიცხვს. BLAST, როგორც წესი, გვაძლევს მონაცემთა ბაზის შესაბამის E სიდიდეს.

არსებობს BLAST-ის განსხვავებული ვარიანტები. მაგალითად უფრო ნაკლებად მგრძნობიარე, მაგრამ გაცილებით სწრაფი BLAT (Blast Like Alignment Tool), რომელიც შექმნილია ნუკლოტიდური თანამიმდევრობის გენომთან შედარების ამოცანისათვის. არსებობს ამ უკანასკნელის ვერსიაც - BLASTZ, რომელსაც იყენებენ განსაკუთრებით დიდი გენომების ან ქრომოსომების შედარების პროცესში. ასევე აღსანიშნავია სტრუქტურულ ბიოლოგებში მეტად პოპულარი PSI-BLAST, რომელიც ამინმჟავური თანამიმდევრობების ურთიერთსწორების შემთხვევაში გამოიყენება და მისი დანიშნულება შორეული პომოლოგების წარმოჩენაა.

FASTA

FAST წარმოადგენს თანამიმდევრობათა შესახებ მონაცემთა ბაზების კვლევისათვის განკუთვნილ პროგრამათა სხვა ოჯახს. ამ ოჯახის პირველი პროგრამათა პაკეტის, FASTP -ს მიზანს წარმოადგენდა მონაცემთა ბაზაში საკვლევი თანამიმდევრობის მსგავსი თანამიმდევრობების მოძიება. FASTP-ის მიერ ხდება მონაცემთა ბაზაში განთავსებული ასოთა თითოეული რიგის შედარება საკვლევ თანამიმდევრობასთან და გამოვლინდება ის თანამიმდევრობები, რომლებიც ჩაითვლება ყველაზე მეტად მსგავსად საკვლევთან. გარდა ამისა, ალგორითმი უზრუნველგვყოფს შესაბამისი ურთიერთსწორებებით და სხვა თანმზღები ინფორმაციით. ცხადია FASTP-ის სიჩქარეს განსაზღვრავს ის, თუ რამდენად სწრაფადაა შესაძლებელი ორი თანამიმდევრობის შედარება. შესაბამისად, ყურადღება გავამახვილოთ იმაზე, თუ როგორ ხორციელდება ეს ძირითადი ეტაპი. ვთქვათ s და t შესადარებელი თანამიმდევრობებია და დავუშვათ, რომ

ისინი ცილის თანმიმდევრობებს შეესაბამება. მათი სიგრძეებია $m = |s|$ და $n = |t|$, შესაბამისად.

- შედარების პირველი ეტაპი მდგომარეობს თანამიმდევრობათათვის საერთო k -სიგრძის სიტყვების (k -tuples) გამოვლენაში, სადაც $k = 1$ -ს ან 2 -ს. პროგრამაში k -ს მნიშვნელობა განისაზღვრება $ktup$ პარამეტრით. ასევე დგინდება ერთნაირ k -tuples-თა წანაცვლებები (*offset*). ასე მაგალითად, თუ $s[i]$ და $t[j]$ ერთნაირ k -tuples-ების საწყისი პოზიციებია, მაშინ წანაცვლებები განისაზღვრება, როგორც $i-j$.

შესაბამისად, პირველ ეტაპზე ფორმირდება შემდეგი სტრუქტურირებული მონაცემები: დაკვირვებათა ცხრილი და ინდექსირებულ წანაცვლებათა ვექტორი (იხ. სურათი 14.1). ამისათვის ჯერ ხდება s -ის სკანირება და კონკრეტულ k -tuples პოზიციათა ჩამონათვალთაგან ცხრილის ფორმირება და შემდეგ t -ს სკანირება. ამის მერე დარდება ეს ცხრილები და თითოეული ერთნაირი k -tuples წყვილისთვის გამოითვლება წანაცვლება. ბოლოს ფორმირდება ინდექსირებულ წანაცვლებათა რაოდენობის განმსაზღვრელი ვექტორი. 14.1 სურათზე წარმოდგენილია $s = \text{HARFYAAQJVL}$ და $t = \text{VDMAAQIA}$ ამინმჟავურ თანმიმდევრობათა შედარების მიზნით ფორმირებული წანაცვლებათა ვექტორი და დაკვირვებათა ცხრილი. შევნიშნოთ, რომ $+2$ წანაცვლება გვხვდება ყველაზე დიდი რაოდენობით, რაც იმაზე მიუთითებს, რომ ყველაზე მეტი რაოდენობის თანხვედრა შეესაბამება ასეთ წანაცვლებას. ეს მეთოდი ცნობილია *დიაგონალური* მეთოდის სახელწოდებით, რადგანაც უმაღლესი წანაცვლება გამოიყენება დინამიკური პროგრამირების ალგორითმის ხმარების ადგილის განსაზღვრაში. ალგორითმი მხოლოდ მოცემული წანაცვლების შესაბამისი დიაგონალის მიმდებარე არეს ითვალისწინებს.

- მეორე ეტაპზე FASTP ახდენს საერთო k -tuples-თა უფრო დეტალურ ანალიზს და აერთიანებს ორ ან მეტ ასეთ k -tuples, თუ ისინი ერთმანეთთან ახლოს, ერთსა და იმავე დიაგონალზეა განლაგებული. ალგორითმი და კრიტერიუმი ევრისტიკულია. გაერთიანებული k -tuple-ები ქმნიან ე.წ. უბნებს. ეს უბნები შეიძლება შევუსაბამოთ BLAST -სათვის განსაზღვრულ პირველწყაროებს (სეგმენტთა წყვილებს), ანუ გამოტოვებების გარეშე განხორციელებულ ლოკალურ ურთიერთსწორებებს. უბნები ხასიათდება გარკვეული ქულათა ჯამებით, რომლებიც განისაზღვრება თანხვედრათა და არათანხვედრათა რიცხვით. მნიშვნელოვანია გვახსოვდეს, რომ უბნები არ მოიცავს გამოტოვებებს.

```

      1 2 3 4 5 6 7 8 9 10 11
s = H A R F Y A A Q I V L
A 2,6,7      1 2 3 4 5 6 7 8
F 4          t = V D M A A Q I A
H 1
I 9          | | | | | | | |
L 11        წანაცვლება +9      -2 -3 +2 +2 -6
Q 8          +2 +1              -2
R 3          +3 +2              -1
V 10
Y 5
სხვა -

```

წანაცვლებათა კოეფიციენტი

-7	-6	-5	<u>-4</u>	-3	-2	-1	0	+1	+2	+3	+4	+5	+6	+7	+8	+9	+10
----	----	----	-----------	----	----	----	---	----	----	----	----	----	----	----	----	----	-----

	1			1	2	1		1	4	1						1	
--	---	--	--	---	---	---	--	---	---	---	--	--	--	--	--	---	--

სურათი 14.1 FASTP -ის დაკვირვებათა ცხრილისა და წანაცვლებათა ვექტორის მაგალითი $ktup = 1$ მნიშვნელობის შემთხვევაში.

- მესამე ეტაპი გულისხმობს ზემოთაღწერილი გზით მიღებული გარკვეული რაოდენობა (განისაზღვრება შესაბამისი პარამეტრით. ვთქვათ ხუთი) საუკეთესო უბანში ქულათა ჯამის ხელახალ შეფასებას უკვე ჩანაცვლებათა მატრიცების საშუალებით (როგორც წესი, BLOSUM62 ან PAM250). ახლადშეფასებულ ქულათა ჯამებს შორის საუკეთესო წარმოადგენს s და t -ს შორის მასგავსების პირველ საზომს და მოიხსენიება როგორც *საწყისი ქულათა ჯამი*. საკვლევი თანამიმდევრობის მიმართ საწყისი ქულათა ჯამის გამოთვლა წარმოებს მონაცემთა ბაზის თითოეული თანამიმდევრობისათვის. საწყისი ქულათა ჯამი გამოიყენება აგრეთვე მონაცემთა ბაზაში გაერთიანებულ თანამიმდევრობათა რიგის მიხედვით დასალაგებლად (რანჟირებისათვის).
- და ბოლოს, ზღვრულ ქულათა ჯამზე მეტი ქულათა ჯამით შეფასებულ თანამიმდევრობათა წყვილისათვის, დინამიკური პროგრამირების ბაზური ალგორითმის გამოყენებით ხდება ოპტიმალურ ქულათა ჯამიანი ურთიერთსწორების მოძიება. ამასთან ალგორითმი შემოიფარგლება საწყისი ურთიერთსწორების (საწყისი ქულათა ჯამის) შესაბამისი დიაგონალის მიმდებარე ზოლით. გამოიყენება სმიტ-ვატერმანის ალგორითმი.

$ktup$ პარამეტრის სიდიდე განსაზღვრავს ალგორითმის შერჩევითობას და მგრძობელობას. *მგრძობელობა* ახასიათებს მეთოდის უნარს წარმოაჩინოს

შორეულად ნათესავური თანამიმდევრობები. შერჩევითობა კი მცდარი პოზიტივის უკუგდების უნარს (მცდარ პოზიტივი (false positives) გულისხმობს არანათესაურ თანამიმდევრობათა მსგავსებად გამოვლენას). FASTP-სა და ამ ოჯახის სხვა პროგრამების ხმარების პროცესში ითვლება, რომ *ktup*-ის დაბალი მნიშვნელობა ზრდის მგრძობელობას, ხოლო მაღალი *ktup* უზრუნველგვცოფს შერჩევითობის მაღალ დონით.

FASTP აგრეთვე აფასებს მსგავსების სტატისტიკურ დამაჯერებლობას. ამ მიზნით ხდება, ვთქვათ, *t* თანამიმდევრობის ამინმჟავათა ისეთი შემთხვევითი გადაჯგუფებები, რომ შენარჩუნებულ იქნეს ამინმჟავური შემადგენლობა, მაგრამ ირღვეოდეს მათი რიგითობა (ბუნებრივთან შედარებით) და ამის შემდეგ დინამიკური პროგრამირების ალგორითმის საშუალებით ხდება *t*-ს ამ გადაჯგუფების შედეგად მიღებულ ვერსიათა (ვარიანტთა) *s*-ის ორიგინალურ ვერსიასთან შედარება. ასეთი მრავალჯერადი პროცედურების საფუძველზე წარმოებს მსგავსების დამახასითებელ ქულათა ჯამების საშუალო მნიშვნელობისა და ამ საშუალოდან სტანდარტული (საშუალო კვადრატული) გადახრის შეფასება. ამ მონაცემების საფუძველზე კი, ქვემოთ მოცემული ფორმულის გამოყენებით გამოითვლება *z* - სიდიდე:

$$z = \frac{\text{ქულათა ჯამს} - \text{საშუალო ქულათა ჯამი}}{\text{სტანდარტული გადახრა}} \tag{14.5}$$

FASTP პროგრამის დახვეწის შედეგად მიღებულ იქნა FASTA. მისი ერთ-ერთი განმასხვავებელი თვისება მდგომარეობს იმაში, რომ მას შეუძლია მუშაობა როგორც ცილების, ისე დნმ-ს თანამიმდევრობებთან. ამ თვალსაზრისით FASTA შეიძლება განხილულ იქნეს, როგორც FASTP-სა და FASTN-ის კომბინაცია. ეს უკანასკნელი შეიქმნა უშუალოდ დნმ-ს თანამიმდევრობებთან სამუშაოდ. ამ ოჯახის კიდევ ერთი წარმომადგენელია AFASTA, რომლის დანიშნულებაც საკვლევი ცილის თანამიმდევრობის შედარება დნმ-ს თანამიმდევრობათა შესახებ მონაცემთა ბაზასთან. ამ შემთხვევაში პროგრამა თვითონ ახდენს ბანკიდან აღებული თანამიმდევრობის ტრანსლაციას ამინმჟავურ თანამიმდევრობაში. სხვა განსხვავებები დაკავშირებულია საწყისი ქულათა ჯამის შეფასების პროცედურასთან. მას შემდეგ, რაც მოხდება საუკეთესო უბნების შერჩევა, FASTA-ში იყენებენ მათი გაერთიანების განსხვავებულ პროცედურას, რაც გულისხმობს ერთსა და იმავე დიაგონალზე არა მყოფი (თუმცა ახლოს მყოფი) უბნების გაერთიანების მცდელობას. ამით ნათესავური თანამიმდევრობების შემთხვევაში საწყისი ქულათა ჯამი მნიშვნელოვნად უმჯობესდება და უახლოვდება გაუმჯობესებულ ქულათა ჯამს (FASTA -ს ენაზე “ოპტიმალურ ქულათა ჯამს”). უფრო მეტიც, FASTA-ში FASTP-საგან განსხვავებით (სადაც ირჩევა 5 საუკეთესო უბანი) ირჩევა უკვე 10 საუკეთესო უბანი. FASTA-ს შემთხვევაში ასევე გაუმჯობესებულია სტატისტიკური დამაჯერებლობის შეფასება. ამ თვალსაზრისით ყველაზე

ხაზგასასმელი სიახლე მდგომარეობს იმაში, რომ ადგილი აქვს თანამიმდევრობათა მხოლოდ ლოკალურ გადაჯგუფებას. გადაჯგუფებებს ექვემდებარება 10 და 20 ამინომჟვანი ბლოკები, რის შედეგადაც მიიღება გადაჯგუფებული, შემთხვევითი თანამიმდევრობები, მაგრამ ორიგინალურთან მსგავსი ლოკალური ამინმჟვური კომპოზიციებით. შედეგად, თუ მაღალი ქულათა ჯამი განპირობებულია ლოკალური კომპოზიციით, მაშინ ეს ფაქტი ქულათა ჯამი იქონიებს ლოკალურად გადაჯგუფებულ თანამიმდევრობათა საშუალო ქულათა ჯამზეც. სხვა უპირატესობებია: ქულათა დასარიცხ მატრიცათა არჩევის უფრო მოქნილი მექანიზმი და თითოეული გადასაჯგუფებელი თანამიმდევრობისათვის მეტი რაოდენობა გადაჯგუფებული ვარიანტების შეფასების უნარი.

BLAST-სა და FASTA-ს თავისებურებანი

ბოლოს ყურადღება გავამახვილოთ BLAST-სა და FASTA-ს თავისებურებებზე და შევხვით მათ შორის განსხვავებებსაც.

BLAST-სთვის დამახასიათებელია :

- ურთიერთსწორების ლოკალურობა. BLAST წარმოაჩენს რეგიონალურ მსგავსებას, მაგრამ არ გვაძლევს საკვლევ თანამიმდევრობასა და მონაცემთა ბაზის თანამიმდევრობას შორის საუკეთესო ურთიერთსწორებას.
- ურთიერთსწორება გამოტოვებების გარეშე. BLAST-ით წარმოებული ურთიერთსწორება მის პირველ ორ ეტაპზე არ გულისხმობს გამოტოვებებს. სწორედ ესაა საფუძველი BLAST-ის სიჩქარისა და სტატისტიკური შეფასების მოდელისა. თუმცა ამით მცირდება მისი მგრძნობელობა.
- მკაცრი სტატისტიკური თეორია. BLAST ეფუძნება Samuel Karlin-ისა და Steven Altschul-ის მიერ შემუშავებულ მკაცრ სტატისტიკურ თეორიას.
- სისწრაფე. BLAST ძალიან სწრაფია. თქვენ შეგიძლიათ გაუშვათ პროგრამა ლოკალურად ან გააგზავნოთ საკვლევ თანამიმდევრობა და მიიღოთ კვლევის შედეგი ელ.-ფოსტით.
- ევრისტიკულობა. BLAST ვერ იძლევა იმის გარანტიას, რომ მოძებნილ იქნება საკვლევ თანამიმდევრობასა და მონაცემთა ბაზას შორის საუკეთესო ურთიერთსწორება. მართალია BLAST-ის სტრატეგიაა, რომ მოძებნოს ყველა შესაბამისობა, მაგრამ ხშირად სიჩქარის გაზრდის მოთხოვნას ეწირება მგრძნობელობა და შესაძლებელია, რომ რომელიმე ბიოლოგიურად მნიშვნელოვანი შესაბამისობა ვერ იქნეს აღმოჩენილი BLAST-ით.
- ჩანაცვლებათა მატრიცები გამოიყენება ურთიერთსწორების პროცესის ყველა ფაზაში. ჩანაცვლებათა მატრიცები გამოიყენება ურთიერთსწორების პროცესის როგორც სკანირების, ისე წაგრძელების ფაზაში. ეს არის სწორედ მისი ძირითადი და მნიშვნელოვანი განსხვავება

FASTA-საგან, რომელიც მატრიცებს იყენებს პროცედურის მხოლოდ მესამე ეტაპზე. საერთოდ ჩანაცვლებათა მატრიცების გამოყენება მნიშვნელოვნად ზრდის მგრძობელობას.

FASTA -სთვის დამახასიათებელია :

- ურთიერთსწორების ლოკალურობა, ისევე როგორც BLAST-ითვის
- ურთიერთსწორება შეიძლება მოიცავდეს გამოტოვებებს.
- სისწრაფე. საკმაოდ სწრაფია, მაგრამ არა ისეთი როგორც BLAST.
- ევრისტკულობა. ისევე როგორც BLAST-ის შემთხვევაში FASTA ვერ იძლევა იმის გარანტიას, რომ მოძებნილ იქნება საკვლევ თანამიმდევრობასა და მონაცემთა ბაზას შორის საუკეთესო ურთიერთსწორება.
- ჩანაცვლებათა მატრიცები გამოიყენება ურთიერთსწორების პროცესის მხოლოდ მესამე ფაზაში

FASTA ფორმატი

მონაცემთა ბაზების მასკანირებელი პროგრამათა პაკეტები, მათ შორის BLAST და FASTA, ხშირ შემთხვევაში ძირითადი შემავალი მონაცემის, საკვლევ ამინმჟავური თანამიმდევრობის წარდგენას მოითხოვს FASTA ფორმატში. ეს ფორმატი გულისხმობს პირველი აღმწერი სტრიქონის არსებობას, რომელსაც მოყვება უშუალოდ თანამიმდევრობის მომცველი სტრიქონები. აღმწერი სტრიქონის პირველ პოზიციას აუცილებლად უნდა იყოს მეტობის („>“) ნიშანი. სტრიქონთა სიგრძე არ უნდა აღემატებოდეს 79-ს. FASTA ფორმატის მაგალითებია:

> თანამიმდევრობის სახელი

```
ctgcgagNcgcgcatgatagMMM-
NNNnnnnncggcggagcatgtagctgtagctgtctcggagcactUUUURRrrrrrrr
cggccgagatcaggcagatcagcgcaggagcagcgcgacgagcacagcatgtagctgatgcatgctaVvvvcgtaggcagc
cggcggagagacgatggagctgc
```

და
>გი|44010|emb|X64011.1| Listeria ivanovii sod gene for superoxide dismutase
CGTTATTTAAGGTGTTACATAGTTCTATGGAAATAGGGTCTATACCTTTCGCCTT
CAATGTAATTTCTTTTACATAAATAATAACAATCCGAGGAGGAATTTTTAATGA
CTTACGAATTACAAAAATTACCTTATACTTATGATGCTTTGGAGCCGAATTTTAT
AAAGAAACAATGGAAATTCACTATACAAAGCACCAACAATTTTATGTAACAAAAC
TAAATGAAGCAGTCTCAGGACACGCAGAACTTGCAAGTAAACCTGGGGAAGAATT
AGTTGCTAATCTAGATAGCGTTTCTGAAGAAATTCGTGGCGCAGTACGTAACCAC

GGTGGTGGACATGCTAACCATACTTTATTCTGGTCTAGTCTTAGCCCAAATGGTGG
TGGTGCTCCAACCTGCTAACTTAAAAGCAGCAATCGAAAGCGAATTCGGCACATTT
GATGAATTCAAAGAAAAATTCAATGCGGCAGCTGCGGCTCGTTTTGGTTCAGGAT
GGGCATGGCTAGTAGTGAACAATGGTAAACTAGAAAATTGTTTCCACTGCTAACCA
AGATTCTCCACTTAGCGAAGGTAAAACCTCCAGTTCCTTGGCTTAGATGTTTTGGGAAC
ATGCTTATTATCTTAAATTCAAAAACCGTCGTCCTGAATACATTGACACATTTTGG
AATGTAATTAACCTGGGATGAACGAAATAAACGCTTTGACGCAGCAAAATAATTAT
CGAAAGGCTCACTTAGGTGGGTCTTTTTATTCTA

წასაკითხად რეკომენდირებული ლიტერატურა

Needelman SB, Wunsch CD. A general method applicable to the search for similarities in the amino acid sequences of two proteins. *J. Mol. Biol.* 1970;48:443-453.

Smith TF, Waterman MS. The identification of common molecular subsequences. *J. Mol. Biol.* 1981;147:195-197.

Dayhoff O, Schwartz RM, Orcutt BC. A model of evolutionary change in proteins, In: *Atlas of protein Sequence and Structure*. Ed. (Dayhoff M.O., Washington, DC, National Biomedical Research Foundation); 1978. p. 345-352.

Henikoff S, Henikoff JG. Amino Acid Substitution Matrices from Protein Blocks. *Proc. Natl. Acad. Sci.* 1992; 89:10915-10919.

Johnson, MS, Overington JP. A structural basis for sequence comparisons. An evaluation of scoring methodologies. *J. Mol. Biol.* 1993; 233:716-738.

Li, W-H. (1997) *Molecular Evolution* (Sunderland, MA, USA: Sinauer.)

Penny, D., Hendy, M.D., Zimmer, E.A., and Hamby, R.K. (1990) 'Trees from sequences: Panacea or Pandora's box?', *Australian Systematic Botany* 3, 21-38.

Woese C R, Kandler O, Wheelis M L (1990) „Towards a natural system of organisms: Proposal for the domains Archaea, Bacteria, and Eucarya“ *Proc Natl Acad Sci USA* 87:4576-4579,

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990). 'Basic local alignment search tool'. *J Mol Biol* 215 (3): 403-410.

Lipman DJ, Pearson WR (1985) „Rapid and sensitive protein similarity searches“ *Science*. 227 (4693):1435-41

თავი 5

სტრუქტურათა შედარება და კლასიფიკაცია

15. ცილებისა და რნმ-ების სტრუქტურული ურთიერთსწორება

სტრუქტურული ურთიერთსწორება წარმოადგენს ურთიერთსწორების ერთ-ერთ ფორმას, როცა ურთიერთსწორება ეფუძნება სივრცული სტრუქტურების, ცილათა ფორმების შედარებას. ორი ან მეტი პოლიმერის ფორმათა და სამგანზომილებიან კომპორმაციათა შესახებ ინფორმაციის გამოყენების საშუალებით ურთიერთსწორების მიზანია შეფასდეს სტრუქტურათა ეკვივალენტობა. სტრუქტურული ურთიერთსწორება, როგორც წესი, გამოიყენება ცილათა მოლეკულების შედარებისათვის, მაგრამ მისი გამოყენება ასევე შეიძლება დიდი ზომის რნმ-თა სტრუქტურების მსგავსების შესაფასებლად. მოლეკულათა სტრუქტურების შედარების ამოცანები შეიძლება დავყოთ სტრუქტურათა სუპერპოზიციონებისა (სუპერშეთავსების) და სტრუქტურული ურთიერთსწორების ამოცანებად.

სტრუქტურათა სუპერპოზიციონება მარტივი პროცედურაა და გულისხმობს ეკვივალენტურ პოზიციებში მყოფ ნაშთთა სტრუქტურული ურთიერთმიმართების შესახებ ინფორმაციის მოპოვებას. გამოიყენება მაშინ, როცა უკვე არსებობს თანამიმდევრობათა ურთიერთსწორება, და საჭიროა შესაბამისი სტრუქტურული მსგავსების შეფასება, ან კიდევ როცა შესადაარებელ სტრუქტურათა თანამიმდევრობები ეკვივალენტურია.

სტრუქტურული ურთიერთსწორება არ მოითხოვს *a priori* ეკვივალენტურ პოზიციათა ცოდნას. როცა თანამიმდევრობათა მსგავსების დაბალი დონის გამო ცილებს შორის ევოლუციური სიახლოვის გამოვლენა სტანდარტული, თანამიმდევრობათა შედარების (ურთიერთსწორების) მეთოდებით ვერ ხერხდება,

ასპარეზზე გამოდის სტრუქტურული ურთიერთსწორება, რომელიც აფასებს ცილათა შორის სიახლოვეს. ამრიგად, სტრუქტურული ურთიერთსწორება გამოიყენება ისეთ ცილათა შორის ევოლუციური ნათესაობის გამოსავლენად, რომელთა თანამიმდევრობებს შორის რაღაც საერთო წარმოჩენა ძალზედ გაძნელებულია. აქ ხაზი უნდა გაესვას იმ გარემოებას, რომ ჰომოლოგიურობის (საერთო ევოლუციური წინაპრის არსებობის) შეფასების ჭრილში სტრუქტურული ურთიერთსწორების შედეგების ინტერპრეტირებისას მოგვეთხოვება სიფრთხილე, რადგან სრულიად არამონათესავე (არასაერთო წინაპრის მქონე) თანამიმდევრობათა კონვერგენტული ევოლუციონირების შედეგად შეიძლება მივიღოთ მსგავსი სივრცული, მესამეული სტრუქტურები.

სტრუქტურული ურთიერთსწორების საფუძველზე, როგორც ზემოთ აღინიშნა, შესაძლებელია შედარდეს ორი ან რამდენიმე ცილა. ურთიერთსწორება ეფუძნება საკვლევ ცილათა მესამეული სტრუქტურების შესახებ ინფორმაციას. შესაბამისად, მეთოდი შესაძლებელია გამოყენებულ იქნეს მხოლოდ იმ თანამიმდევრობათათვის (ცილათათვის), რომელთა მესამეული სტრუქტურაც ცნობილია. მესამეულ სტრუქტურათა დადგენა ჩვეულებრივ ხდება რენტგენული კრისტალოგრაფიის, ან ბირთვული მაგნიტური რეზონანსის მეთოდების გამოყენებით. ასევე შესაძლებელია განხორციელდეს კომპიუტერული წინასწარმეტყველების მეთოდებით მიღებულ სტრუქტურათა ურთიერთსწორება. როგორც წესი, სტრუქტურული ურთიერთსწორების მიდგომებს იყენებენ კომპიუტერული მეთოდებით შემოთავაზებულ სტრუქტურათა მოდელების სისწორის შეფასების პროცესში. ამ დროს კომპიუტერით პროგნოზირებული სტრუქტურა დარდება ცნობილ, ექსპერიმენტულად დადგენილ მესამეულ სტრუქტურას. სტრუქტურული ურთიერთსწორება ასევე მეტად სასარგებლოა სტრუქტურული გენომიკისა და პროტეომიკის ფარგლებში განხორციელებული კვლევების პროცესშიც, რადგანაც იგი თანამიმდევრობებზე დაფუძნებული შედარების ბიოინფორმატიკული მეთოდებით მოპოვებული მონაცემების სიზუსტის შეფასების შესაძლებლობას იძლევა.

სტრუქტურული ურთიერთსწორებას ახასიათებენ სივრცულად ურთიერთშესაბამისობაში მოყვანილ ატომთა კოორდინატების კომპლექტის საფუძველზე შეფასებულ სტრუქტურათშორისი განსხვავებების მახასიათებელი საშუალო კვადრატული გადახრებით (RMSD). ურთიერთსწორების მახასიათებელი RMSD გვიჩვენებს სტრუქტურათა დივერგირების ხარისხს. სტრუქტურული ურთიერთსწორება რთულდება, თუ ერთ-ერთი ან რამდენიმე შესადაარებელი სტრუქტურა მოიცავს ერთზე მეტ დომენს, რადგან დომენტშორისი ორიენტაციის უმნიშვნელო ცვლილებამაც კი, შეიძლება გამოიწვიოს RMSD-ის შეფასებებში მნიშვნელოვანი არტიფაქტები.

სტრუქტურული ურთიერთსწორების შედეგად წარმოებული მონაცემები

წარმატებული სტრუქტურული ურთიერთსწორება, როგორც მინიმუმ, იძლევა შესაბამისობაში მოყვანილ ატომთა სივრცულ კოორდინატებს (შეწინაშთ, რომ შესაძლებელია ურთიერთსწორების ერთ-ერთ სტრუქტურასთან შედარების მიზნით ჩატარება. ამ დროს უკანასკნელის სივრცული კოორდინატები უცვლელი დარჩება). შესადარებელთან შესაბამისობაში მოყვანილი სტრუქტურა გამოიყენება RMSD-ის ან სხვა უფრო რთულ მახასიათებელთა შესაფასებლად. სტრუქტურული ურთიერთსწორება გულისხმობს თანამიმდევრობათა წყვილის სტრუქტურული შესაბამისი ურთიერთსწორებით წარმოდგენას. უკანასკნელი უკვე შესაძლებლობას მოგვცემს შეფასდეს სტრუქტურული ურთიერთსწორების საფუძველზე შესაბამისობაში მოყვანილ, იდენტურ ამინმჟავურ ნაშთთა მომცველ, წყვილთა რიცხვი და ამ სიდიდის საფუძველზე მოხდეს სტრუქტურულად ურთიერთსწორებულ ცილათა ნათესაურობის დონის (ჰომოლოგიურობის) დახასიათება.

სტრუქტურათა შედარების პროცესში გამოყენებული მონაცემები.

ცილები პოლოპეპტიდური ჯაჭვებია, საერთო ბუნების ძირითადი ჯაჭვითა და გვერდით ჯაჭვთა (ამინმჟავურ ნაშთთა) განსხვავებული თანამიმდევრობებით. შესაბამისად, ცილათა სტრუქტურების შედარება და RMSD-ის შეფასება შესაძლებელია დაეყრდნოს ცილის ატომთა სრულიად განსხვავებულ ქვეკომპლექტებს. როცა სტრუქტურულ ურთიერთსწორებაში მოსაყვან ცილათა თანამიმდევრობები მნიშვნელოვნად განსხვავდება ერთმანეთისაგან (როგორც წესი, ეს ასეა) გვერდით ჯაჭვთა ატომებზე დაყრდნობა აზრს კარგავს, რადგან შესაბამისობაში მოყვანილ გვერდით ნაშთთა ატომური შემადგენლობები შეიძლება მნიშვნელოვნად განსხვავდება ერთმანეთისაგან. სწორედ ამის გამო სტრუქტურული ურთიერთსწორების მეთოდების (პროგრამული პაკეტების) უმრავლესობა სტანდარტულ რეჟიმში (როცა მითითება არ ხდება) გულისხმობს მხოლოდ ძირითადი ჯაჭვის ატომთა კომპლექტებზე დაყრდნობას. უფრო მეტიც, პეპტიდური ბმის ჯგუფის ატომთა ერთ სიბრტყეში განლაგების დაშვების პირობებში შესადარებელ ატომთა ჩამონათვალი გამართლებულია შემოიფარგლოს მხოლოდ α -ნახშირბადის ატომებით. დამაჯერებლობის გაზრდის მიზნით, სტრუქტურული მსგავსების შეფასებისას RMSD-სთან ერთად ხშირად გამოიყენება სხვა კრიტერიუმებიც, რომლებიც შეიძლება ემყარებოდეს ინფორმაციას მეორეული სტრუქტურის ელემენტების, ნატივური კონტაქტების, წყალბადურ ბმათა ქსელების შესახებ და ა.შ.

სტრუქტურათა სუპერპოზიცია.

ხშირად ამოცანა არ მოითხოვს სტრუქტურათა ურთიერთსწორებას. მაგალითად, გვაქვს უკვე რაღაც ურთიერთსწორება ისეთი ამინმჟავური თანამიმდევრობებისა, რომელთა სივრცული სტრუქტურები ცნობილია. ასეთ შემთხვევაში შესაძლებელია შეწყვილებულ, შესაბამისობაში მოყვანილ ამინოჟავების კოორდინატებზე დაფუძნებით RMSD-ის შეფასება. ანუ, ცნობილია რა ცილათა სივრცული სტრუქტურები, თანამიმდევრობათა უკვე არსებული ურთიერთსწორების მონაცემთა საფუძველზე შესაძლებელია სტრუქტურათა სუპერპოზიცირება (სუპერმეთავსება). ამრიგად, სტრუქტურული სუპერპოზიცირება შესაძლებელია გამოყენებულ იქნეს ან ერთი და იგივე ცილის სხვადასხვა შესაძლო კონფორმაციათა შედარების ამოცანის გადაწყვეტის პროცესში (ამ შემთხვევაში ურთიერთსწორება არაა საჭირო, რადგან კონფორმაციებს ერთი და იგივე თანამიმდევრობა შეესაბამება), ან, როგორც ზემოთ აღვნიშნეთ, იმ თანამიმდევრობათა ურთიერთსწორების ხარისხის შეფასების პროცესში, რომელთა სივრცული სტრუქტურებიც ცნობილია. სუპერპოზიცირების მეთოდი გულისხმობს ერთი სტრუქტურის მეორის მიმართ მობრუნებისა და გადატანის ოპერაციების ოპტიმალური ვარიანტების შერჩევას, სტრუქტურათშორისი შეთავსების საუკეთესო ვარიანტის მოძიების მიზნით. მოძიების პროცესი ტრადიციულად ეფუძნება სტრუქტურათშორისი საშუალო კვადრატული გადახრების (მანძილების) ჯამის მინიმალური მნიშვნელობის პოვნის პროცედურას.

სტრუქტურათა ურთიერთსწორება

სტრუქტურათა ურთიერთსწორებისას სასურველია მათი წარმოდგენა შედარებისათვის მოსახერხებელ სივრცეში. ამ მიზნით სტრუქტურებს წარმოადგენენ კოორდინატთაგან დამოუკიდებელი ფორმით. ეს, როგორც წესი, მიიღწევა ისეთი მატრიცების საშუალებით, რომლებიც ეფუძნება არა ფიქსირებულ, კოორდინატთა სისტემის მიმართ განსაზღვრულ აბსოლუტურ მანძილებს, არამედ ფარდობით საზომებს (მეტრიკებს). ერთ-ერთი ასეთი წარმოდგენა გულისხმობს ცილის სტრუქტურის დისტანციური მატრიცით აღწერას. უკანასკნელი წარმოადგენს ორგანზომილებიან მატრიცას, რომლის ელემენტები ცალკეული სტრუქტურის გარკვეულ ატომთა (მაგალითად α -ნახშირბადის) წყვილებს შორის მანძილებია. შედარებით უხეში მეტრიკის (მიახლოების) პირობებში, მაგალითად, მეორეული სტრუქტურული ელემენტების მიახლოებაში ჩამოყალიბებული დისტანციური მატრიცის საფუძველზე წარმოებულ შედარებას საკმაოდ რეალურ ურთიერთსწორებამდე მივყავართ. უნდა აღინიშ-

ნოს, რომ უკანასკნელ შემთხვევაში მთელი რიგი მანძილების შესახებ ინფორმაციის დანაკლისი კომპენსირდება შემთხვევითი შემფოთებების (ხმაურების) შემცირებით. სტრუქტურათა წარმოდგენის ფორმის სწორ შერჩევას გადამწყვეტი მნიშვნელობა ენიჭება ეფექტური ურთიერთსწორების ალგორითმის შემუშავების პროცესში.

სტრუქტურული ურთიერთსწორების მეთოდები

სტრუქტურული ურთიერთსწორების მეთოდები გამოიყენება როგორც ცალკეული სტრუქტურების, ისე სტრუქტურათა კომპლექტების ურთიერთშესადარებლად, და შესაბამისად, ყველა ყველასთან შედარების შედეგების მომცველ მონაცემთა ბაზების შესაქმნელად. ეს მონაცემთა ბაზები გვაძლევს ცილათა სტრუქტურების შესახებ მონაცემთა ბაზაში (PDB) შემავალ სტრუქტურათა ურთიერთშესაბამისების აღმწერ სურათს და დახვევის ტიპების მიხედვით ცილათა კლასიფიკაციის საშუალებას იძლევა.

DALI

სტრუქტურათა ურთიერთსწორების ერთ-ერთ ფართოდ გავრცელებულ მეთოდს წარმოადგენს DALI (distance alignment matrix), რომელშიც შესადარებელი სტრუქტურები წარმოდგენილია დისტანციათა მატრიცების სახით. მეორეული სტრუქტურები შედგება ჯაჭვის გასწვრივ მეზობლად განლაგებული ნაშთებისაგან. შესაბამისად, მეორეული სტრუქტურის თავისებურებანი აისახება მატრიცის მთავარი დიაგონალის სიახლოვეს. სხვა დიაგონალები გვიჩვენებს ჯაჭვის გასწვრივ არამეზობელ ნაშთთა შორის კონტაქტების სურათს. ამასთან, თუ ეს დიაგონალი პარალელურია მთავარი დიაგონალის, შესაძლებელია ვილაპარაკოთ შესაბამის ფრაგმენტთა პარალელურობაზე, და თუ დიაგონალი მთავარი დიაგონალის პერპენდიკულარულია, უკვე საქმე გვაქვს ანტიპარალელურობასთან.

როცა ორი შესადარებელი ცილის დისტანციათა მატრიცები ერთსა და იგივე პოზიციებში ამქლავნებს იდენტურ ან მსგავს თვისებებს, შეიძლება ითქვას, რომ მათი დახვევის ტიპები მსგავსია. DALI-ის მიერ წარმოებული ურთიერთსწორება ეფუძნება ცილების დისტანციათა მატრიცების შედარების პროცედურებს. პროცესი იწყება ურთიერთგადაფარვად, 6×6 ზომის ქვემატრიცათა შედარებით. შემდგომ სტანდარტული, ქულათა მაქსიმიზაციის ალგორითმით ხდება წყვილურ შესაბამისობაში მოყვანილ ქვემატრიცათა საერთო ურთიერთსწორებად რეასამბლირება. DALI-ის ორიგინალური ვერსია ქულათა ჯამის მაქსიმიზის

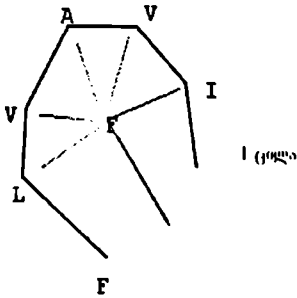
მოძიებისას იყენებს მონტე-კარლოს მეთოდს. ქულათა ჯამი წარმოადგენს შესაბამისობაში მოყვანილ ატომთა შორის მანძილების ფუნქციას. ხაზი უნდა გაესვას იმ გარემოებას, რომ განსაკუთრებით დამორებულ ატომთა წყვილების წვლილის ხელოვნური შემცირებით წარმოებს სხვადასხვა მიზეზებით გამოწვეულ ხმაურთა (შემთხვევით შემფოთებათა) ეფექტების ნიველირება. ხმაურთა მიზეზებად შეიძლება მოვიყვანოთ: მოსახვევთა მობილურობა, სპირალთა ტორსიონული თავისუფლებების გარკვეული ხარისხი და სტრუქტურაში მინორული (უმნიშვნელო) ვარიაციების შესაძლებლობა.

DALI გამოყენებულ იქნა FSSP (Fold classification based on Structure-Structure alignment of Proteins, or Families of Structurally Similar Proteins) მონაცემთა ბაზის კონსტრუირების პროცესში. FSSP მოიცავს ცნობილ სტრუქტურათა ერთმანეთთან წყვილურად შედარების, ურთიერთსწორების შედეგებს. ამ შედეგების საფუძველზე შესაძლებელი ხდება სტრუქტურულად მსგავს ცილათა ყველა ჯგუფის გამოვლენა და შესაბამისად დახვევის ტიპებად კლასიფიკაცირება.

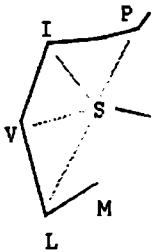
SSAP

SSAP (Sequential Structure Alignment Program) მეთოდი იყენებს რა სტრუქტურულ სივრცეში განსაზღვრულ ატომ-ატომურ ვექტორების, სტრუქტურული ურთიერთსწორების პროცესს წარმართავს დინამიკურ პროგრამირების ალგორითმზე დაყრდნობით. SSAP სტრუქტურული ურთიერთსწორების პროცესში საზოგადოდ გამოყენებულ α -ნახშირბადების ნაცვლად ეყრდნობა β -ნახშირბადებს და მათი საშუალებით განსაზღვრავს ვექტორებს. SSAP პირველ ეტაპზე ახდენს ცალკეული ცილის თითოეული ნაშთიდან ჯაჭვის გასწვრივ არამომიჯნავე, მაგრამ სივრცულად ახლო მყოფ ნაშთებამდე დამორებების ვექტორთა კონსტრუირებას (იხ. სურათი 15.1). შესადარებელი ცილებიდან აღებული ნაშთების თითოეული წყვილისათვის (წყვილს ქმნიან ნაშთები სხვადასხვა ცილებიდან) ნაშთთა შორის დამორების ვექტორების შესახებ მონაცემების საფუძველზე იგება ორგანოზომილებიანი მატრიცა, რომლის ელემენტები წარმოადგენს სხვადასხვა ნაშთთა შესაბამის ვექტორებს შორის სხვაობებს. იწარმოება ასეთ მატრიცათა გარკვეული სიმრავლე. შემდგომ ეტაპზე გამოიყენება რა დინამიკური პროგრამირების მეთოდი, მატრიცათა უკანასკნელი სიმრავლის თითოეულ წევრზე დაყრდნობით წარმოებს ოპტიმალური ლოკალური ურთიერთსწორების მოძიება. შემდგომში ხდება ასეთნაირად წარმოებული ლოკალურ ურთიერთსწორებათა გაერთიანება ე.წ. ჯამურ მატრიცად, რომლიდანაც ასევე დინამიკური

პროგრამირების მეთოდზე დაყრდნობით ვიღებთ საბოლოო სტრუქტურულ ურთიერთსწორებას



1 (კლა)



2 (კლა)

სურათი 15.1. გვიჩვენებს SSAP-ში გამოყენებულ ატომ-ატომური ვექტორების შინაარსს. მეთოდის შესაბამისად ადგილი აქვს ვექტორულ სხვაობათა კონსტრუირებას. მაგალითად, პირველი ცილის (FA) ვექტორსა და მეორე ცილის (SI) ვექტორს შორის. ასეთი გზით შესაძლებელია ცილათა შორის სხვაობების მატრიცათა გარკვეული სიმრავლის ფორმირება. სწორედ ასეთნაირად ფორმირებული მატრიცებზე დაყრდნობით დინამიკური პროგრამირების ალგორითმის გამოყენებით ხდება ლოკალურად ოპტიმალური ურთიერთსწორების მოძიება. შემდგომ ლოკალურად ოპტიმალური ურთიერთსწორებათა მონაცემებზე დაყრდნობით ფორმირდება ჯამური ორგანოზომილებიანი მატრიცა, რომლის საფუძველზეც, ასევე დინამიკური პროგრამირების გამოყენებით, მოძიებულ იქნება საბოლოო ურთიერთსწორება.

SSAP, როგორც წესი, გამოიყენება წყვილური ურთიერთსწორებისათვის. მაგრამ იგი შეიძლება მოდიფიცირდეს, რათა მოვახდინოთ ერთდროული მრავალწევრა ურთიერთსწორება. SSAP წარმოადგენს იმ სქემის მნიშვნელოვან რგოლს, რომელიც გამოყენებულ იქნა დახვევათა ტიპების იერარქიული კლასიფიცირების ცნობილი სისტემის CATH-ის (Class, Architecture, Topology, Homology) ფორმირებისას. ანუ სწორედ მასზე დაყრდნობითაა შექმნილი უკანასკნელი კლასიფიკაცია.

რნმ-თა სტრუქტურული ურთიერთსწორების თავისებურებანი.

შეიძლება ითქვას, რომ ტრადიციულად სტრუქტურული ურთიერთსწორება გამოიყენება ცილათა სტრუქტურების შესადარებლად. თუმცა რნმ მოლეკულებსაც შეიძლება გააჩნდეთ გამოკვეთილი შესამეული სტრუქტურა, რომელიც ძირითადად შედგება კომპლემენტარულ ფუძეებს შორის წყალბადურ ბმათა ბადისა და კონტაქტში მყოფ ფუძეთა სტეკინგისა (ჰიდროფობული ეფექტი, ვან-დერ-ვაალსის ძალები, ელექტროსტატიკა). საერთოდ, როგორც ცილების, ისე რნმ-ს სტრუქტურა მათ თანამიმდევრობასთან შედარებით გაცილებით კონსერვატიულია. ამას ემატება რნმ-ანბანის შეზღუდულობა, რისი გამოხატულებაცაა კონფორმაციულ შესაძლებლობათა სიმცირე, რაც აძნელებს გენომურ თანამიმდევრობათა მონაცემებიდან არამაკოდირებელ რნმ მოლეკულათა ფუნქციურად მსგავსი ოჯახის გამოყოფას. ერთი სიტყვით, მხოლოდ თანამიმდევრობათა შესახებ ინფორმაციაზე დაფუძნებით ფუნქციონალური მსგავსება-განსხვავებათა გამოვლენა რნმ-ების შემთხვევაში ძალზედ სარისკო საქმეა.

თუმცა ახლახან შემოთავაზებულ იქნა 7მცირე თანამიმდევრული იდენტურობის მქონე რნმ-თა სტრუქტურული ურთიერთსწორების მეთოდი. ამ მეთოდის საფუძველზე შექმნილია პროგრამათა პაკეტი FOLDALIGN. უნდა აღინიშნოს, რომ ამ მეთოდს ცოტა აქვს საერთო ცილათა სტრუქტურული ურთიერთსწორებისას გამოყენებულ მეთოდებთან, რადგან იგი თვითონ აწარმოებს სტრუქტურის წინასწარმეტყველებას და არ იყენებს ექსპერიმენტულად დადგენილ სივრცულ სტრუქტურათა შესახებ ინფორმაციას. (უფრო დაწვრილებით რნმ-ს სტრუქტურის წინასწარმეტყველების შესახებ ინფორმაციას შეგიძლიათ გაეცნოთ მეექვსე თავში, მე-17 პარაგრაფი). ცილის სტრუქტურის წინასწარმეტყველების მეთოდთა ეფექტურობა დღესდღეობით არდამაკმაყოფილებელია, მაშინ, როცა ფსევდომარყულების (pseudoknots) არმქონე რნმ-ს სტრუქტურის წინასწარმეტყველება თავისუფალ ენერგიაზე დაფუძნებულ ქულათა დარიცხვის მეთოდის ან სტატისტიკურ მიდგომის საშუალებით, შესაძლებელია. აქ შევნიშნავთ, რომ თავისუფალი ენერგია კომპლემენტარულ ფუძე წყვილებისა და სტეკინგის მახასიათებელი პარამეტრების ფუნქციაა.

16. ცილათა სტრუქტურული კლასიფიკაცია

შეიძლება ჩაითვალოს, რომ ცილის სტრუქტურა განსაზღვრავს მის ფუნქციასა და ევოლუციის თავისებურებებს. შესაბამისად, ფუნქციისა და ევოლუციური ისტორიის შესახებ ცოდნის დაგროვების პროცესში მოითხოვება ცილის სტრუქტურისა და ამ სტრუქტურის სხვა ცილის სტრუქტურებთან მიმართების შესახებ ცოდნის მოპოვება. ამისათვის სასურველია გვექონდეს ცოდნა როგორც ცილების დახვევათა რაღაც ზოგადი ტიპების, ასევე ცალკეულ ცილათა ნატივი სტრუქტურების შესახებ.

დღეისათვის შეიძლება ითქვას, რომ არ არსებობს ცილა, რომლისთვისაც არ შეიძლებოდეს მასთან საერთო ევოლუციური წინაპრის მქონე, ან სტრუქტურულად მსგავსი სხვა ცილის მოძიება. ინფორმაციას ევოლუციური კავშირების შესახებ დიდი მნიშვნელობა ენიჭება თანამედროვე ბიოლოგიის განვითარებისთვის, კერძოდ კი ცილათა სტრუქტურის წარმოქმნისა და ევოლუციის საკითხების გარკვევისათვის, უცნობ გენთა ფუნქციის დადგენისათვის და სხვა. დღეს რენტგენოსტრუქტურული ანალიზისა და ბირთვული მაგნიტური რესონანსის საშუალებით დადგენილ ცილათა სივრცული სტრუქტურების რიცხვი სწრაფად იზრდება. დიდია კომპიუტერული მიდგომების საფუძველზე სწორად პროგნოზირებულ სივრცულ სტრუქტურათა რაოდენობაც. ამასთან აღმოჩნდა, რომ მსგავსი ტოპოლოგია შესაძლებელია ახასიათებდეს არა მხოლოდ მსგავსი ფუნქციისა და მსგავსი თანამიმდევრობების მქონე ცილებს, არამედ საკმაოდ განსხვავებული ფუნქციის და არამსგავსი თანამიმდევრობების მქონე ცილებსაც. შესაბამისად, ცილათა სტრუქტურების კლასიფიკაცია მართალია დამოკიდებულია თანამიმდევრობათა კლასიფიკაციასთან, მაგრამ მაინც დამოუკიდებელ ამოცანას წარმოადგენს. თანამიმდევრობათა კლასიფიკაცია, როგორც წესი, ეფუძნება თანამიმდევრობათა წყვილური შედარების, ურითიერთსწორების შედეგებს. ბუნებრივია, სტრუქტურული კლასიფიკაცია, ასევე უნდა დაეფუძნოს სტრუქტურული ურითიერთსწორების შედეგებს. სტრუქტურული ურითიერთსწორების მეთოდებს ჩვენ გავეცანით წინა პარაგრაფში. ახლა მოკლედ მიმოვიხილოთ ამ მეთოდთა გამოყენებით განხორციელებული ცილათა სტრუქტურების კლასიფიცირების შედეგები.

SCOP (Structural Classification of Proteins) კლასიფიკაცია

ცილათა სტრუქტურული კლასიფიკატორი (SCOP) წარმოადგენს ცნობილი სტრუქტურის მქონე ცილათა შორის სტრუქტურულ და ევოლუციურ ურთიერთ-მიმართებათა აღმწერ მონაცემთა ბაზას. იგი მოწყობილია იერარქიულად და ქმნის რამდენიმე სხვადასხვა დონის გაერთიანებებს. კლასიფიკაციის პროცესში ფუნდამენტალურ ერთეულს წარმოადგენს დომენი. ცილების დომენთა სხვადასხვა დონის დაჯგუფებების ფორმირებისას გამოიყენებენ თანამიმდევრობათა, სტრუქტურათა და ფუნქციათა მსგავსებების შესახებ ინფორმაციას.

SCOP იერარქია

ქვევიდან ზევით SCOP -ის იერარქიულ დონეებია:

- სახეობები (Species) შეესაბამება ცალკეულ ცილათა თანამიმდევრობებს, ან მათ ბუნებრივ თუ ხელოვნურად წარმოებულ ვარიანტებს;
- ცილები (Protein) აერთიანებს სხვადასხვა სახეობის, თითქმის იდენტური ფუნქციის მქონე მსგავს თანამიმდევრობებს;
- ოჯახი (Family) აერთიანებს მსგავსი თანამიმდევრობის (იდენტურობა $\geq 30\%$), ან არამსგავსი თანამიმდევრობის, მაგრამ მსგავსი ფუნქციის და სტრუქტურის მქონე ცილებს;
- ზეოჯახი (Superfamilie) აერთიანებს იმ ოჯახებს, რომელთათვისაც გარკვეული სტრუქტურული და ფუნქციური მსგავსების საფუძველზე შესაძლებელია საერთო ევოლუციური წინაპრის არსებობა.

უფრო ზედა იერარქიულ დონეებზე კლასიფიკაცია წმინდად სტრუქტურული ხდება. შესაბამისად ჩნდება შემდეგი დონეები:

- დახვევა (Folds) წარმოადგენს სტრუქტურულად მსგავს, მაგრამ სხვადასხვა ბიოლოგიური თვისებების მქონე სუპეროჯახთა გაერთიანებას;
- კლასი წარმოადგენს, მხოლოდ მეორეული სტრუქტურული ელემენტების ტიპებსა და მათ ორგანიზებაზე დაფუძნებულ დახვევათა გაერთიანებას.

მონაცემთა ბაზის 2007 წლის (1.73 ვერსია) გამოშვების 7 ძირითად კლასში გაერთიანებულია 92 927 დომენი, რომლებიც ჯგუფდება 3464 ოჯახად, 1777 ზეოჯახად და 1086 დახვევად. 7 ძირითადი და 4 დამატებითი სტრუქტურული კლასის დასახელება და მოკლე აღწერა ასეთია:

1. α -სპირალური (α -helical) დომენები
2. β -სტრუქტურული (β -sheet) დომენები
3. α/β დომენები, რომლებიც შედგება " β - α - β " სტრუქტურული მოტივისაგან. β - სტრუქტურა ძირითადად პარალელურია
4. ($\alpha+\beta$) დომენი, ფორმირებული დამოუკიდებელი α -სპირალებისა და ძირითადად ანტიპარალელური β -სტრუქტურისაგან
5. მულტიდომენური ცილები
6. მემბრანული და უჯრედული ზედაპირის ცილები და პეპტიდები (აქ არ შედის იმუნური სისტემის ცილები)
7. „მცირე“ ("small") ცილები
8. სპირალ-სპირალური (coiled-coil) ცილები
9. დაბალი გარჩევითობით დადგენილი ცილათა სტრუქტურები
10. პეპტიდები და ფრაგმენტები
11. დაპროექტებული (designed, არაბუნებრივი) ცილები

SCOP-ის დღევანდელ ვარიანტში დომენების მოძიებისას გამოიყენება 34 495 PDB ფაილი. SCOP-ში დომენთა კლასიფიკაცია ემყარება ამინმჟავურ თანამიმდევრობათა და სამგანზომილებიან სტრუქტურათა მსგავსებას და წარმოებს ძირითადად ხელით, ექსპერტების მიერ. კლასიფიკაციის პირველი ვარიანტი შემოთავაზებულ იქნა 1995 წელს და მას შემდეგ ყოველწლიურად ხდება მისი განახლება. ამ ხნის განმავლობაში ცილების ცნობილ სივრცულ სტრუქტურათა რიცხვი გაიზარდა 20-ჯერ, დახვევათა, ზეოჯახთა და ოჯახთა რიცხვი კი, შესაბამისად 4-ჯერ, 5-ჯერ და 7-ჯერ.

ახლად დადგენილ სტრუქტურათა SCOP-ში წარმატებულად და დროულად განთავსებისათვის საჭიროა კლასიფიცირების პროცესში გარკვეული მოქნილობის გამოჩენა. ასე მაგალითად, შესაძლებელია ზეოჯახთა მნიშვნელოვანი გაფართოება მათში ახალ-ახალი ოჯახებისა და ცილების ჩართვის ხარჯზე. განსაკუთრებით პოპულარული ზეოჯახების ოჯახებად დაყოფის პროცესში დასაშვებია კლასიკური სქემიდან გადახრა: ითვლება, რომ ოჯახი შედგება მსგავსი სტრუქტურის მქონე ისეთ ცილათაგან, რომელთა შორის შეიძლება ან შეიმჩნეოდეს, ან არა მნიშვნელოვანი თანამიმდევრული მსგავსება. ოჯახის შიგნით ცილებს შორის ნათესაობა უფრო მკვეთრად არის გამოხატული, ვიდრე სხვადასხვა, სტრუქტურულად დივერგირებულ ოჯახებში გაერთიანებულ ცილებს შორის. აქვე უნდა აღინიშნოს, რომ ახალ სტრუქტურათა დიდი წილი სტრუქტურული გენომიკის ეგიდით განხორციელებული პროექტების განხორციელების შედეგია. როგორც წესი, ასეთ სტრუქტურათა სრული ფუნქციური დახასიათება არ არსებობს, რაც ბუნებრივია აძნელებს მათ კლასიფიკაციას ცილების და ზეოჯახების დონეებზე. შესაბამისად, ასეთ ცილათა პირველადი კლასიფიკაცია პირობითია. შემდგომი მონაცემები ან დადასტურებს ან უარყოფს ამ კლასიფიკაციას.

როგორც აღვნიშნეთ SCOP წარმოადგენს ექსპერტთა მიერ ძირითადად ხელით წარმოებულ კლასიფიკაციას. არსებობს ნახევრად ავტომატური კლასიფიკაციები, რომელთა შორის გამოვყოფთ CATH კლასიფიკაციას.

CATH კლასიფიკაცია

CATH ასევე წარმოადგენს PDB-ში გაერთიანებული ცილოვანი სტრუქტურების დომენთა იერარქიულ კლასიფიკაციას. გამოიყენება მხოლოდ 4.0 Å და უკეთესი გარჩევითობით გაშიფრული კრისტალური სტრუქტურები და ასევე NMR-ის საფუძველზე დადგენილი სტრუქტურები. კლასიფიციებისას არ ხდება არაცილოვანი მოლეკულებისა და მოდელების განხილვა. კლასიფიციებისას გამოიყენება როგორც ავტომატური, ისე ხელით მიდგომები. იერარქია მოიცავს ოთხ ძირითად დონეს. ესენია: კლასი (Class), არქიტექტურა (Architecture), ტოპოლოგია (დახვევათა ოჯახი) (Topology (fold family)) და ჰომოლოგიურთა ზეოჯახი (Homologous superfamily). თითოეული დონის აღწერა მოცემულია ქვევით.

დომენთა საზღვრების დადგენა

კლასიფიციებისას განიხილება ცილათა ცალკეული დომენები და შესაბამისად დგება დომენთა საზღვრების დადგენის პრობლემა. მრავალდომენიანი ცილის შემადგენელ დომენებად დაყოფისას გამოიყენება როგორც ავტომატური, ისე ხელით მიდგომები. თუ დომენებად დასაყოფი ცილა ავლენს მნიშვნელოვან თანამიმდევრობით და სტრუქტურულ მსგავსებას (თუ სტრუქტურული ურთიერთსწორების შედეგად ურთიერთშესაბამისობაში მოყვანილ ამინმჟავურ წყვილებს შორის 80% იდენტური წყვილებია, SSAP ქულათა ჯამი ≥ 80 -ზე) სხვა ცილასთან, რომლისთვისაც დაყოფა უკვე არსებობს, მაშინ დაყოფა ხდება არსებულ მონაცემებთან სრული შესაბამისობით, ავტომატურად. სხვა შემთხვევებში დომენთა საზღვრების დადგენა ცალკეული ცილისათვის წარმოებს ინდივიდუალურად, ხელით, რაც გულისხმობს: სტრუქტურაზე და თანამიმდევრობაზე დაფუძნებული მეთოდების გამოყენებას, ლიტერატურული მონაცემების მოძიებას და ამ ინფორმაციის ზედმიწევნით ანალიზს.

CATH იერარქია და კლასიფიცირების პროცედურები

ავტომატური პროცედურები

თუ დომენი ავლენს მნიშვნელოვან თანამიმდევრობით და სტრუქტურულ მსგავსებას (იდენტურობა $\geq 35\%$ -ზე, SSAP ქულათა ჯამი ≥ 80 -ზე) CATH-ში უკვე კლასიფიცირებულ სხვა დომენთან, მისთვის ავტომატურად გამოყენებული იქნება არსებულის (მსგავსი დომენის) კლასიფიკატორი. სხვა შემთხვევაში კლასიფიცირება წარმოებს ექსპერტების მიერ ხელით და ემყარება შესაძარბეელი ალგორითმების (მაგალითად SSAP, HMM (Hidden Markov Model) და სხვა) გამოყენების შედეგად მოპოვებულ ინფორმაციასა და შესაბამის ლიტერატურულ მონაცემებს.

იერარქიულ დონეთა შესაბამისად კლასიფიცირების პროცედურები

კლასი (Class), C-დონე

კლასს განსაზღვრავს მეორეულ სტრუქტურულ ელემენტთა კომპოზიცია და მათი ურთიერთგანლაგება სტრუქტურაში. განასხვავებენ სამ ძირითად კლასს: ძირითადად α ; ძირითადად β ; და $\alpha - \beta$ კლასები. ეს უკანასკნელი თავის მხრივ იყოფა α/β და $\alpha+\beta$ ქვეკლასებად. გამოყოფენ ასევე მეოთხე კლასს, რომელიც მოიცავს მეორეული სტრუქტურის არ მქონე, ან მცირე წილის მქონე ცილათა დომენებს.

არქიტექტურა (Architecture), A-დონე

A-დონე მეორეულ სტრუქტურათა ორიენტაციის საფუძველზე აღწერს დომენის სტრუქტურის ზოგად ფორმას, თუმცა იგნორირებულია მათ შორის კავშირების ხასიათი. არქიტექტურის განსაზღვრა ხდება ხელით, მეორეულ სტრუქტურათა ურთიერთგანლაგების მარტივი აღწერის საშუალებით. არქიტექტურული ტიპების მაგალითებია: კასრი (barrel) ან სამ-შრიანი სენდვიჩი (3-layer sandwich). გამოიყენება საზოგადოდ მიღებული, ცნობილი არქიტექტურული ტიპებიც, ვთქვათ β -პროპელერი (beta-propeller), ან ოთხი α -სპირალისაგან შემდგარი შეკვრა (alpha four helix bundle).

ტოპოლოგია (Topology (Fold family)), T-დონე

დახვევის ტიპებად სტრუქტურათა დაჯგუფება ეფუძნება როგორც სტრუქტურის საერთო ფორმას, ასევე მეორეულ სტრუქტურათა შორის კავშირების შესახებ ინფორმაციას. ასეთნაირი დაჯგუფებისათვის გამოიყენება სტრუქტურული ურთიერთსწორების მეთოდები SSAP (Taylor & Orengo, 1989) და CATHEDRAL

(Harrison *et al.* 2003). სტრუქტურები, რომელთა SSAP ქულათა ჯამი 70 ან მეტია და იდენტურობა > 60%, ერთიანდება ერთსა და იმავე T-დონეში ანუ დახვევის ტიპში.

ზოგიერთი დახვევა ფართოდაა გავრცელებული, სხვები ნაკლებად. ფართოდ გავრცელებულთა შორისაა β -ორშრიანი სენდვიჩი და α - β -სამშრიანი სენდვიჩი.

ჰომოლოგიურთა ზოგჯახი (Homologous Superfamily), H-დონე

ამ დონეზე გაერთიანება გულისხმობს ცილათა დომენტათვის საერთო წინაპრის არსებობას და შესაბამისად, მათი ჰომოლოგიურად განხილვის შესაძლებლობას. მსგავსების გამოვლენა ხდება როგორც თანამიმდევრობათა მაღალი იდენტურობის დაფიქსირებით, ისე სტრუქტურული მსგავსების შეფასებით (SSAP-ზე დაფუძნებით). სტრუქტურათა ერთსა და იმავე ჰომოლოგიურ ზოგჯახში გაერთიანება ხდება მაშინ, თუ ისინი აკმაყოფილებენ შემდეგ მოთხოვნებს:

- თანამიმდევრობათაშორისი იდენტურობა $\geq 35\%$, გადაფარვა $\geq 60\%$, ანუ დიდი სტრუქტურის ასეთი ნაწილია ეკვივალენტური პატარისა.
- SSAP ქულათა ჯამი ≥ 80.0 , თანამიმდევრობათა შორისი იდენტურობა $\geq 20\%$, გადაფარვა $\geq 60\%$.
- SSAP ქულათა ჯამი ≥ 70.0 , გადაფარვა $\geq 60\%$ და არსებობს ლიტერატურული ან Pfam ცილათა ბაზის ინფორმაცია, რომ დომენები მსგავსი ფუნქციის მატარებლებია.
- SAM (Hughey & Krogh, 1996), HMMER (<http://hmmer.wustl.edu>) და PRC (<http://supfam.org/PRC>) მეთოდებით ჩატარებული თანამიმდევრობათა HMM- ანალიზისა და HMM-HMM შედარების საფუძველზე ვლინდება მნიშვნელოვანი მსგავსება.

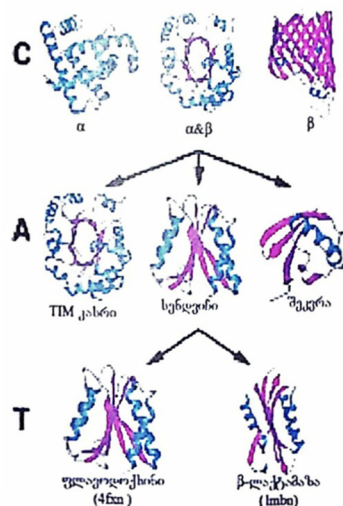
თანამიმდევრობათა ოჯახების დონეები: (S,O,L,I, D)

შემდგომში ხდება H-დონეზე დაჯგუფებულთა შემდგომი სუბკლასტერიზაცია თანამიმდევრობათა გჯგუფებად (ოჯახებად). თანამიმდევრობათა მსგავსების მიხედვით განასხვავებენ შემდეგი დონის ოჯახებს:

დონე	სახელი ინდენტურობიდან გამომდინარე	გადაფარვა
S	35%	80%
O	60%	80%
L	95%	80%
I	100%	80%

და ბოლოს D-დონე, რომელიც წარმოადგენს აღმრიცხველს და გამოიყენება თითოთეული დომენისათვის თავისი უნიკალური CATHSOLID კლასიფიკაციის მისანიჭებლად. თანამიმდევრობათშორისი იდენტურობისა და გადაფარვის შეფასებები ეფუძნება ნიდელმან-ვონჩის ალგორითმს, სადაც გამოტოვებაზე ჯარიმა შეადგენს -3-ს. თანამიმდევრობათშორისი იდენტურობის პროცენტი გამოითვლება როგორც (100 x იდენტურ ნაშთთა რიცხვი / მოკლე თანამიმდევრობის სიგრძე) და გადაფარვის პროცენტი კი, როგორც (100 x ურთიერთსწორებულ ნაშთთა რიცხვი / გრძელი თანამიმდევრობის სიგრძე).

16.1 სურათზე წარმოდგენილია იერარქიული დონეები და ამ დონეთა შიგნით დაყოფის მაგალითები



სურათი 16.1 CATH-ის იერარქიულ დონეთა (C-კლასი, A -არქიტექტურა, T -ტოპოლოგია) შესაბამისი კლასიფიცირების მაგალითი.

FSSP -კლასიფიკაცია (Families of structurally similar proteins)

FSSP კლასიფიკაცია, რომელიც საფუძვლად უდევს FSSP მონაცემთა ბაზას მთლიანად ავტომატური პროცედურების ნაყოფია. ეს პროცედურები, მხოლოდ DALI-ის სტრუქტურული ურთიერთსწორების ალგორითმზე დაყრდნობით,

ავტომატურად გამოყოფს სტრუქტურულად შესაბამისობაში მყოფ ცილათა ჯგუფებს. ჯგუფებში გაერთიანება ხდება DALI-ის Z-ქულათა ჯამის საფუძველზე (უფრო დაწვრილებითი ინფორმაციისათვის იხილეთ წინა პარაგრაფი „სტრუქტურული ურთიერთსწორება“).

წასაკითხად რეკომენდირებული ლიტერატურა

Theobald DL, Wuttke DS (2006). “THESEUS: Maximum likelihood superpositioning and analysis of macromolecular structures.” *Bioinformatics* 22(17):2171-2172.

Holm L, Sander C (1996) “Mapping the protein universe.” *Science* 273(5275):595-603

Mount DM. (2004). *Bioinformatics: Sequence and Genome Analysis* 2nd ed. Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY.

Havgaard JH, Lyngso RB, Stormo GD, Gorodkin J. (2005). “Pairwise local structural alignment of RNA sequences with sequence similarity less than 40%.” *Bioinformatics* 21(9):1815-24.

Bray JE, Todd AE, Pearl FM, Thornton JM, Orengo CA. (2000) “The CATH Dictionary of Homologous Superfamilies (DHS): a consensus approach for identifying distant structural homologues.” *Protein Eng.* 13(3):153-65.

Harrison A, Pearl F, Sillitoe I, Slidel T, Mott R, et al. (2003) “Recognizing the fold of a protein structure.” *Bioinformatics* 19: 1748–1759.

Holm L, Sander C (1993) “Protein structure comparison by alignment of distance matrices.” *J Mol Biol* 233: 123–138.

Taylor WR, Orengo CA (1989) “Protein structure alignment.” *J Mol Biol* 208: 1–22.

Hughey, R. & Krogh, A. (1996). “Hidden Markov models for sequence analysis: Extension and analysis of the basic method.” *CABIOS*, 12 (2), 95-107.

თავი 6

სივრცული სტრუქტურისა და ფუნქციის წინასწარმეტყველება

17. მეორეული სტრუქტურის წინასწარმეტყველება

მეორეული სტრუქტურის წინასწარმეტყველების მეთოდები მხოლოდ პირველადი სტრუქტურის შესახება ინფორმაციაზე დაყრდნობით ახდენენ ცილების ან რნმ-ს ლოკალური მეორეული სტრუქტურის წინასწარმეტყველებას. ცილების შემთხვევაში წინასწარმეტყველება გულისხმობს ამინმჟავური თანამიმდევრობის ცალკეული ფრაგმენტის მიერ ალფა-სპირალური, ბეტა-სტრუქტურული (ხშირად მოიხსენიება როგორც გოფირებულ-ფურცლოვანი სტრუქტურა, რომლის ცალკეული ზოლის, ჯაჭვის სეგმენტის კონფორმაციას უწოდებენ „გაჭიმულს“), მოსახვევის ტიპის კონფორმაციების მიღების პოტენციალის შეფასებას. შეფასების ხარისხი მოწმდება წინასწარმეტყველების მონაცემების კრისტალოგრაფულ მონაცემებთან შედარების საფუძველზე. შედარება ხორციელდება DSSP ალგორითმის გამოყენებით. რნმ-ს თანამიმდევრობის შემთხვევაში წინასწარმეტყველება გულისხმობს წყალბადურ ბმათა ოპტიმალური ქსელის მოძიებას. შემუშავებულია სპეციალზირებული ალგორითმები, გათვლილი მეორეული სტრუქტურის ისეთი სპეციფიკური ელემენტების საწინასწარმეტყველოდ, როგორებიცაა ტრანსმემბრანული სპირალი ან სპირალური-სპირალი ცილებში და კანონიკური მიკრო-რნმ-ული სტრუქტურა რნმ-ში.

ცილებში მეორეული სტრუქტურების წინასწარმეტყველების თანამედროვე მეთოდების სიზუსტე აღწევს 80% .

რნმ-ს მეორეული სტრუქტურის წინასწარმეტყველების მეთოდები ძირითადად ეფუძნება კომპლემენტარულ ფუძეთა შეწყვილებისა და სტეკინგის შესახებ ინფორმაციას. რნმ-ს მოლეკულათა უმრავლესობას გააჩნია გარკვეული სივრცული სტრუქტურა. მაგალითებად გამოდგება ტრანსპორტული და მიკრო რნმ-ები. უნდა აღინიშნოს, რომ რნმ-ს მეორეული სტრუქტურის წინასწარმეტყველებისთვის შექმნილი მეთოდების უმრავლესობა ეფუძნება დინამიკური პროგ-

რამირების ბაზურ ალგორითმს და შესაბამისად არ გააჩნიათ, ვთქვათ, ფსევდო-ნოტების წარმატებულად წინასწარმეტყველების უნარი.

ცილის მეორეული სტრუქტურის ტიპის განსაზღვრა სივრცული სტრუქტურის შესახებ ინფორმაციაზე დაყრდნობით

DSSP (Define Secondary Structure of Proteins) ალგორითმი გულისხმობს სტანდარტულ მიდგომას, რომლის საფუძველზეც წარმოებს ცილის თითოეული ამინომჟავისათვის მეორეული სტრუქტურის ტიპის დადგენა ცილის სივრცული მოწყობის ანუ ატომთა სივრცული კოორდინატების შესახებ ინფორმაციაზე დაყრდნობით.

DSSP პროცედურა იწყება ცილის სივრცულ სტრუქტურაში წყალბადური ბმების ადგილმდებარეობის დადგენით. წყალბადური ბმის წარმოჩენა ეფუძნება მხოლოდ ელექტორსტატიკურ ურთიერთქმედებათა ენერგიების შეფასებებს. ენერგიის გამოსათვლელად გამოიყენება დამოკიდებულება

$$E = q_1 q_2 (1/ r_{ON} + 1/ r_{CH} - 1/ r_{OH} - 1/ r_{CN}) \quad (17.1)$$

სადაც $q_1 = 0.42$ და $q_2 = 0.20$. ასეთ მნიშვნელობათა საფუძველს ქმნის კარბონილის ჟანგბადისა და ამიდური წყალბადის პარციალურ მუხტათათვის განსაზღვრული მნიშვნელობები, რომლებიც უტოლდება $-0.42 e$ -ს და $+0.2 e$ -ს, შესაბამისად და კარბონილის ნახშირბადისა და ამიდური აზოტის პარციალური მუხტები, რომლებსაც ნიშნით საწინააღმდეგო, მაგრამ სიდიდით იგივე მნიშვნელობები აქვთ. თუ E ენერგია ნაკლებია -0.5 -ზე, DSSP ალგორითმი წყალბადური ბმის არსებობას ადასტურებს.

DSSP ალგორითმი განსაზღვრავს მეორეული სტრუქტურის რვა ტიპს და თითოეულის აღსანიშნავად იყენებს ცალკე სიმბოლოს. ასე მაგალითად 3_{10} (3_{10} helix), α (alpha helix) და π (pi helix) სპირალების საიდენტიფიკაციოდ იყენებენ შესაბამისად G, H და I სიმბოლოებს და მათი არსებობა გულისხმობს ძირითადი ჯაჭვის გასწვრივ შესაბამისად სამ, ოთხ და ხუთი პეპტიდური ბმით დამორეზულდონორებს შორის განხორციელებულ წყალბადურ ბმათა თანამიმდევრულად განმეორებადი ქსელის არსებობას. β (გოფირებულ ფურცლოვანი) სტრუქტურის იდენტიფიცირებისას DSSP ასხვავებს წყალბადურ ბმაში მონაწილე ორი ტიპის წყვილებს, ესენია: ანტიპარალელური და პარალელური ხიდაკები. თუ გამოკვეთილია მხოლოდ ერთი ხიდაკის არსებობა გამოიყენება აღნიშვნა B (a beta bridge), ხოლო როცა წყალბადურ ბმათა უფრო მასშტაბური ქსელი იდენტიფიცირდება, იხმარება აღნიშვნა E (extended), რაც მიანიშნებს ბეტა გოფირებულ-ფურცლოვანი (beta sheet) სტრუქტურის არსებობაზე. ბეტა ბურცოების (beta bulges) მომცველი ფურცლები ასევე E -თი აღინიშნებიან. მეორეულ სტრუქტურათა

დანარჩენ ტიპებია: T მოსახვევები (ალფა სპირალის მსგავსი ერთი წყალბადური ბმით), მკვეთრი S მოსახვევები (მოსახვევი მკვეთრია თუ C₁ C₂ და C₂ C₁ ვექტორებს შორის კუთხე 70°-ზე ნაკლებია). ზემოთ მოხსენიებულთაგან განსხვავებული შემთხვევების აღსანიშნავად გამოიყენება ცარიელი სიმბოლო. მეორეული სტრუქტურების კლასიფიცირება შესაძლებელია უფრო მსხვილი გაერთიანებების სახითაც. არსებობს სამი მსხვილი გაერთიანება : სპირალები (სადაც ერთიანდება G, H და I), ფურცლოვანი სტრუქტურის ცალკეული სეგმენტები (სადაც ერთიანდება E და B), და ყულფები (სადაც ერთიანდება ყველა დანარჩენი).

16.1 ცხრილში მოცემულია ცილის სივრცულ სტრუქტურებში დამზერილი მეორეული სტრუქტურის ტიპები. ასევე სხვადასხვა ამინმჟავების დახასიათება მათი მეორეული სტრუქტურის ცალკეული ტიპებისადმი მიდრეკილების თვალსაზრისით.

ცხრილი 17.1

მეორეული სტრუქტურის ტიპები	
სპირალები:	α-სპირალი 3 ₁₀ სპირალი π- სპირალი β- სპირალი პოლიპროლინის სპირალი კლაგენის სპირალი
გაჭიმული სტრუქტ.	β-ზოლი მოსახვევი ბეტა სამაგრი ბეტა ბურცოზი α-ზოლი
სუპერმეორული:	სპირალური-სპირალი სპირალი-მოსახვევი-სპირალი EF ხელი
ამინმჟავების მეორეულ სტრუქტურებისადმი მიდრეკილება	
სპირალისადმი:	მეთიონინი ალანინი ლეიცილი გლუტამინის მჟავა გლუტამინი ლიზინი
გაჭიმული სტრუქტ.:	ტრეონინი იზოლეიცილი ვალინი ფენილალანინი ტიროზინი ტრიფტოფანი
არარეგულარული:	გლიცილი სერინი პროლინი ასპარაგინი ასპარაგინის მჟავა
არ არსებობს მიდრეკილება:	ცისტეინი ჰისტიდინი არგინინი

ცილის მეორეული სტრუქტურის წინასწარმეტყველება ამინმჟავური თანამიმდევრობის შესახებ ინფორმაციაზე დაყრდნობით

ცილის მეორეული სტრუქტურის წინასწარმეტყველების პირველი მეთოდები გაჩნდა წინა საუკუნის 60-ანი წლების ბოლოსა და 70-ანის დასაწყისში. ისინი ძირითადად აწარმოებდნენ ალფა სპირალების წინასწარმეტყველებას. შედარებით უფრო სრულყოფილი წინასწარმეტყველების მეთოდები, რომლებიც ახდენდნენ ბეტა გოფირებული ფურცლების წინასწარმეტყველებასაც გამოჩნდა 70-იან წლებში. ისინი იყენებდნენ ცნობილ სივრცულ სტრუქტურათა შესახებ მონაცემთა სტატისტიკური ანალიზის შედეგებს. ეს მეთოდები ეფუძნებოდა რა ერთი კონკრეტული თანამიმდევრობის შესახებ ინფორმაციას, ამჟღავნებდა დაახლოებით 60-65%-იან სიზუსტეს და ახასიათებდათ ბეტა ფურცლების გადამეტებული წინასწარმეტყველება. შემდგომში ამინმჟავების მეორეული სტრუქტურის ელემენტებისადმი მიდრეკილების შეფასება დაიწყო ჰომოლოგიურ თანამიმდევრობათა მრავალწევრა ურთიერთსწორების პროცესში ფორმირებულ ურთიერთსწორების სვეტთა ანალიზის საფუძველზე. შესაბამისად ცილათა ცნობილი სტრუქტურების შესახებ მონაცემთა ბაზის ზრდისა და თანამედროვე ალგორითმების (მაგალითად ნეირონული ქსელების - neural nets) შემუშავებასთან ერთად სიზუსტემ გლობალური ცილებისათვის მიაღწია 80%-ს. მეორეული სტრუქტურების წინასწარმეტყველების მეთოდების სიზუსტის თეორიულად შეფასებული ზღვარი შეადგენს 90%-ს. იმ ობიექტურ მიზეზთა შორის, რომლებიც პრინციპულად ზღვრავენ წინასწარმეტყველების სიზუსტეს შეიძლება დავსახვლოთ ის, რომ ა) კრისტალიზაციის პროცესში შეიძლება ადგილი ჰქონდეს ნატივური სტრუქტურის ცვლილებას. ცვლილებას როგორც წესი ადგილი აქვს მეორეული სტრუქტურების ბოლოებზე; ბ) წინასწარმეტყველების პროცესში არ გაითვალისწინება შესამეული (შორი) ურთიერთქმედებები. ხშირ შემთხვევაში შორეული ურთიერთქმედებები გადამწყვეტ როლს თამაშობენ მეორეული სტრუქტურის საბოლოო ტიპის ჩამოყალიბებაში.

წინასწარმეტყველების პირველი, უმარტივესი მეთოდები

ჩოუ-ფასმანის მეთოდი (Chou-Fasman)

ჩოუ-ფასმანის მეთოდი იყო პირველი, რომელიც შეიქმნა მეორეული სტრუქტურის წინასწარმეტყველებისათვის. ის ძირითადად ეყრდნობა ალბათობებს, რომლებიც ისაზღვრებოდა მეორეულ სტრუქტურათა თითოეულ ტიპში ცალკეულ ამინმჟავათა შეხვედრის ფარდობითი სიხშირეების სახით. ჩოუ-ფასმანის მიერ პირველად, 70-ანი წლებში გამოყენებული წინასწარმეტყველების პარამეტრები

განისაზღვრებოდა სტრუქტურათა მცირე სიმრავლეზე დაფუძნებით და ამიტომ ნაკლებად ეფექტური იყო. თუმც, შემდგომში მოხდა ამ პარამეტრთა ხელახალი შეფასება სტრუქტურათა უფრო დიდი სიმრავლის გამოყენებით. მეორეული სტრუქტურის წინასწარმეტყველების პროცესში ჩოუ-ფასმანის მეთოდის სიზუსტე დაახლოებით 50-60%-ა.

გორ-ის მეთოდი (GOR)

გორ-ის მეთოდი შექმნილია გარნიერის, ოსგუტორფისა და რობსონის მიერ და მისი სახელი სწორედ ამ მეცნიერების გვარების აბრევიატურას წარმოადგენს. ეს არის ინფორმაციულ თეორიაზე დაფუძნებული მეთოდი. გორ-ის მეთოდი ჯაჭვის კონკრეტულ პოზიციამი სტრუქტურის წინასწარმეტყველებისას ითვალისწინებდა არა მხოლოდ ამ პოზიციამი მყოფი ამინომჟავის გარკვეული მეორეული სტრუქტურისადმი მიდრეკილებას (ამ სტრუქტურის მიღების ალბათობას), არამედ ასევე პირობითი ალბათობას იმისა, რომ მის მეზობელ ამინომჟავასაც ექნება იგივე სტრუქტურა. ეს მეთოდი ჩოუ-ფასმანზე უფრო მგრძობიარე და ზუსტია. იგი უმრავლეს ამინომჟავათა მეორეული სტრუქტურის ტიპისადმი მიდრეკილებას საკმაოდ წარმატებულად წარმოაჩენს. თუმცა არსებობს გამოწვევებიც, მაგალითად პროლინი და გლიცინი. გორ-ის მეთოდის სიზუსტე დაახლოებით 65%-ა და ბევრად ეფექტურია α -სპირალის წინასწარმეტყველებისას. β -ფურცლების შემთხვევაში ეს მეთოდი ხშირად უძლურია. მას ასევე უჭირს მარყუქებისა და არარეგულარული უბნების წინასწარმეტყველება.

წინასწარმეტყველების თანამედროვე მეთოდები

მეორეული სტრუქტურის წინასწარმეტყველების ამოცანის გადაჭრის საქმეში რევოლუციური გარდევვა მოხდა მას შემდეგ, რაც ხელმისაწვდომი გახდა კომპლუგური თანამიმდევრობათა დიდი ოჯახების შესახებ ინფორმაცია. ცილათა ოჯახების შესახებ ინფორმაციის გამოყენების პირობებში ბევრად უფრო იზრდება მეორეული სტრუქტურის ელემენტის წინასწარმეტყველების სიზუსტე. მანქანური თვითსწავლების მიდგომებზე დაფუძნებული მეთოდიკების დანერგვამ კი (მაგალითად ნეირონულ ქსელზე დაფუძნებულმა მეთოდმა), წინასწარმეტყველების სიზუსტემ 70%-ზე მაღლა ასწია.

ჰომოლოგიურ ოჯახში შემავალ ცილათა თანამიმდევრობების მრავალწევრა ერთდროულ ურთიერთსწორებაზე დაფუძნებული მეთოდები

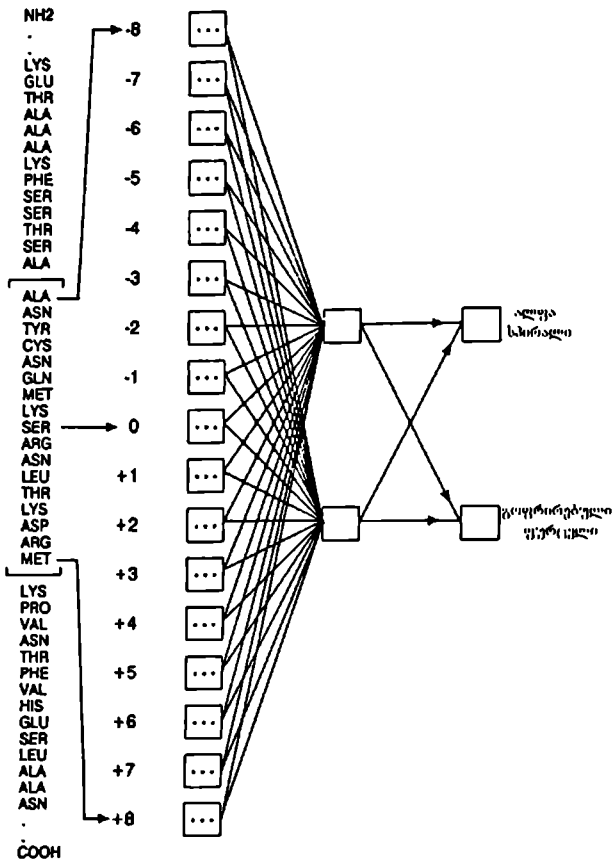
ცილათა თანამიმდევრობების მრავალწევრა ურთიერთსწორების საშუალებით მეორეული სტრუქტურის წინასწარმეტყველების ავტომატური მეთოდები გულისხმობს ჰომოლოგიურ ცილათა ოჯახების შესახებ ინფორმაციის გამოყენებას. ასეთ მეთოდთა სპექტრი საკმაოდ მრავალფეროვანია. ერთი პარაგრაფის ფარგლებში შეუძლებელია მათი სრულფასოვანი მიმოხილვა. ამიტომ, მხოლოდ შესაბამის სერვერებსა და ლიტერატურის ჩამონათვალს შემოგთავაზებთ:

- Zvebil, M.J.J.M., Barton, G.J., Taylor, W.R. & Sternberg, M.J.E. (1987). Prediction of Protein Secondary Structure and Active Sites Using the Alignment of Homologous Sequences *Journal of Molecular Biology*, 195, 957-961. (ZPRED)
- Rost, B. & Sander, C. (1993), Prediction of protein secondary structure at better than 70 % Accuracy, *Journal of Molecular Biology*, 232, 584-599. (PHD)
- Salamov A.A. & Solovyev V.V. (1995), Prediction of protein secondary structure by combining nearest-neighbor algorithms and multiply sequence alignments. *Journal of Molecular Biology*, 247, 1 (NNSSP)
- Geourjon, C. & Deleage, G. (1994), SOPM : a self optimised prediction method for protein secondary structure prediction. *Protein Engineering*, 7, 157-16. (SOPMA)
- Solovyev V.V. & Salamov A.A. (1994) Predicting alpha-helix and beta-strand segments of globular proteins. (1994) *Computer Applications in the Biosciences*, 10, 661-669. (SSP)
- Wako, H. & Blundell, T. L. (1994), Use of amino-acid environment-dependent substitution tables and conformational propensities in structure prediction from aligned sequences of homologous proteins. 2. Secondary Structures, *Journal of Molecular Biology*, 238, 693-708.
- Mehta, P., Heringa, J. & Argos, P. (1995), A simple and fast approach to prediction of protein secondary structure from multiple aligned sequences with accuracy above 70 %. *Protein Science*, 4, 2517-2525. (SSPRED)
- King, R.D. & Sternberg, M.J.E. (1996) Identification and application of the concepts important for accurate and reliable protein secondary structure prediction. *Protein Sci*, 5, 2298-2310. (DSC).

მანქანურ თვითსწავლებაზე დაფუძნებული ნეირონული ქსელის მეთოდი

მანქანური თვითსწავლების მეთოდთა შორის განვიხილოთ ყველაზე ფართოდ გავრცელებული, ნეირონულ ქსელზე (neural network - NN) დაფუძნებული მეთოდი. NN მეთოდები ემყარება რა ცნობილი სტრუქტურის მქონე ცილათა სტრუქტურის კომპლექტს, წარმოაჩენს მოცემულ მეორეულ სტრუქტურასთან შესაბამისობაში მყოფ ამინმჟავურ თანამიმდევრობათა მოტივებს. NN მეთოდებით წინასწარმეტყველების სიზუსტე 70%-ზე მეტია, თუმცა მათი საშუალებით მ-სტრუქტურების პროგნოზირება კვლავ გაძნელებულია. ქვემოთ შევეცდებით წარმოდგინა შეგიქმნათ ნეირონულ ქსელზე დაფუძნებული მიდგომით მეორეული სტრუქტურის წინასწარმეტყველების ზოგად თავისებურებებზე.

წინასწარმეტყველება ნეირონული ქსელებით. ცილის მეორეული სტრუქტურის წინასწარმეტყველების მიზნით გამოყენებული ნეირონული ქსელის მეთოდზე წარმოდგენის შესაქმნელად განვიხილოთ 17.1 სურათზე მოყვანილი მაგალითი. საზოგადოდ NN მეთოდით განხორციელებული პროცედურები (გამოთვლები) გულისხმობს შემავალი ფენის (input layer), ფარული ფენებისა (hidden layer) და გამოსავალი ფენის (output layer) არსებობას. 17.1 სურათზე განხილული მაგალითის შემთხვევაში წარმოდგენილია ერთი ფარული ფენა. ზოგადად, ფარული ფენა შეიძლება არ გვექნოდეს ან მათი რაოდენობა შეიძლება ერთზე მეტი იყოს. როგორც წესი, მეორეული სტრუქტურის წინასწარმეტყველების ამოცანის შემთხვევაში შემავალი ფენა აკოდირებს ამინმჟავური თანამიმდევრობის გასწვრივ მოძრავ ფანჯარას, რომლის ცენტრალური ამინმჟავა წარმოადგენს სწორედ წინასწარმეტყველების ობიექტს. ფანჯრის სიგრძე ჩვეულებრივ (ისევე როგორც სურათზე) უტოლდება 17 (შეიძლება გამოყენებულ იქნეს სხვა სიგრძის ფანჯარაც), ანუ წინასწარმეტყველების ობიექტის (ამინმჟავას) ორივე მხარეს განიხილება რვა-რვა ამინმჟავა და გამოიყენება მათ შესახებ ინფორმაცია. ქსელის შესავალზე ინფორმაციის წარსადგენად გამოიყენება ორობითი სისტემით გამოხატული კოდირება. თითოეული პოზიციის დაკოდირების პროცესში კოდის შერჩევა ხდება შემავალი სიდიდეების 21 კომპლექტიდან. ამათგან 20 კომპლექტი აკოდირებს ოც ამინმჟავას, ხოლო ერთი კომპლექტი კი გამოიყენება იმ შემთხვევაში, როცა ფანჯრის მიერ ხდება ამინო ან კარბოქსილის ზოლოს გადაფარვა. ამრიგად შემავალი ფენა შედგება ზემოთ მოხსენიებულ 21-გან შერჩეულ 17 კომპლექტისაგან. და ამ 17 კომპლექტში გაერთიანებულ შემავალ მონაცემებში ერთიანთა რაოდენობა უტოლდება 17-ს, დანარჩენი ნულებია. 17.1 სურათზე წარმოდგენილ შემთხვევაში გამოიყენება ერთი ფარული ფენა, რომელიც შედგება ორი ერთეულისაგან (ერთეულთა რიცხვი შეიძლება იყოს სხვა). გამოსავალი ფენაც მოიცავს ორ ერთეულს. გამოსავალი მონაცემების ფორმირებისას გამოიყენება მეორეული სტრუქტურის შემდეგი კოდირება: (1,0) = სპირალს, (0,1) = გოფირებულ ფურცელს და (0,0) = გორგალს (არარეგულარულ სტრუქტურას). როგორც წესი, გამოსავალ მონაცემთა მნიშვნელობები განათავსებულია 0.0–1.0 ინტერვალში

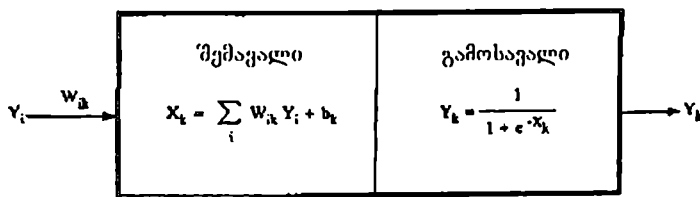


ამონაბეჭედი შემაჯავლი უჯრა ფარული უჯრა გამოსავალი უჯრა
 იანაბინდებობა (17x21 ერთეული) (2 ერთეული) (2 ერთეული)

სურათი 17.1. ნეირონული ქსელის ტოპოლოგია. შემოსავალზე ნაჩვენებია 17 ბლოკი. თითოეული ბლოკი, რომელიც მოიცავს 21 მონაცემს (ნულს ან ერთს) შეესაბამება ფანჯრის მოცემულ პოზიციაში განთავსებული ამინმეცვის კოდს. ამრიგად ამინმეცვური თანამიმდევრობის მოცემული ფანჯრისათვის ქსელში შეშავალი 357 მონაცემიდან 17 ერთიანია, დანარჩენი კი ნული. ფარულ ან გამოსავალ ფენაში განთავსებული ბლოკი შეესაბამება ცალკეულ ერთეულს. წინასწარმეტყველების ობიექტს წარმოადგენს ფანჯრის ცენტრალური ნაშთი.

და მათი წინასწარმეტყველებად გარდასახვის პროცესი ეფუძნება ზღვრულ სიდიდე $t-x$. სპირალად ჩაითვლება ნებისმიერი ოთხი (Kabsch და Sander-ის 236

კლასიფიკაციის მიხედვით) თანამიმდევრულ ამინმჟავათაგან შემდგარი ფრაგმენტი, რომელთა სპირალობის დამახასიათებელი გამოსავალ მონაცემთა სიდიდეები აღემატება ბეტა ფურცლის გამოსავალს და ამასთანავე მეტია ϵ -ზე. ბეტა ფურცლად ჩათვლება ნებისმიერი ორი (ბეტა ფურცლისათვის მინიმუმი Kabsch და Sander-ის კლასიფიკაციის მიხედვით) თანამიმდევრულ ამინმჟავათაგან შემდგარი ფრაგმენტი, რომელთა დამახასიათებელ გამოსავალ მონაცემთა სიდიდეები აღემატება სპირალობის გამოსავალს და ამასთანავე მეტია ϵ -ზე. ამინმჟავებს, რომლებიც არ მიიჩნევიან არც სპირალურად და არც ბეტა-ფურცელთა წარმომადგენლად, ენიჭებათ არარეგულარი სტრუქტურის (გორგლის) სტატუსი. ზღვრული პარამეტრის სიდიდე შეირჩევა ცილათა საწერტნელი კომპლექტისათვის წინასწარმეტყველების სიზუსტის ოპტიმიზირების პროცესში და შეესაბამება წინასწარმეტყველების მაქსიმალურ სიზუსტეს. როგორც წერტნის, ისე წინასწარმეტყველების პროცესში ქსელის გასწვრივ მონაცემთა მოძრაობის გზა განისაზღვრება ქსელის ტოპოლოგიით, ხოლო ამ მოძრაობის პროცესში მონაცემთა გარდაქმნის წესები მოცემულია 17.2 სურათზე. წესებში იგულისხმება კოდირებისა და დეკოდირების წესები. წინასწარმეტყველების სქემა სწორედ ამ უკანასკნელი წესების, ასევე წერტნის შედეგად დადგენილ წონათა (weights) და მიდრეკილებათა (biases) მნიშვნელობების (მაგალითისათვის იხილეთ ცხრილი 17.2) საშუალებით აღიწერება.



k-ური ერთეული

სურათი 17.2. ქსელში მიმდინარე გამოთვლები. ქსელის გასწვრივ ფარული და გამოსავალი ფენების გავლის პროცესში თითოეული ერთეულის (ფენის) შესავალზე და გამოსავალზე წარმოებს გამოთვლები, რომელთა სახეე წარმოდგენილია სურათზე. როგორც სურათიდან ჩანს წარმოებს წინა ფენის ყველა ერთეულიდან მიღებული გამოსავლების Y_i -ების, დამაკავშირებელი წონების W_{ki} გათვალისწინებით, აჯამვა უკანასკნელი k -ამის კონკრეტული ერთეულის მიმართ დარეგულირება ხდება b_k მიდრეკილებით. ამის შემდეგ წარმოებს k -ური ერთეულისათვის გამოსავლის გენერირება სურათზე მითითებული ფორმულის მიხედვით. ყოველივე ამის შემდეგ მზადაა მონაცემები ქსელის შემდეგ ფენაზე გადასასვლელად. ერთეულიდან გამოსავალის სიდიდე დევს (0,1) ინტერვალში. დამაკავშირებელი წონები შეიძლება იყოს როგორც დადებითი, ისე უარყოფითი მნიშვნელობის.

ცხრილი 17.2. 48 ცილისაგან შემდგარ კომპლექტზე, 0 ფარული ფენის პირობებში, 17.2 სურათზე მოცემული წესით წერთის შედეგად დახვეწილ წონათა μ -ს და მიდრეკილებათა b -ს სიდიდეები. ცხრილი აღებულია Holley-ისა და Carplus-ის 1989 ნაშრომიდან

Amino acid	Weight at window position																	
	-8	-7	-6	-5	-4	-3	-2	-1	0	1	2	3	4	5	6	7	8	
	Part A. Input layer to helix output																	
-	0.10	0.20	-0.33	0.02	-0.27	-1.42	-1.56	-1.22	0.04	-1.39	-0.81	-0.45	-0.80	-0.39	0.10	-0.19	0.82	
ALA	0.07	0.09	0.10	0.16	0.20	0.46	0.58	0.56	0.65	0.63	0.43	0.42	0.41	0.28	0.12	0.16	0.14	
ARG	0.11	-0.05	-0.06	-0.04	-0.10	0.81	0.45	0.43	0.45	0.45	0.46	0.52	0.17	0.03	-0.17	0.05	-0.21	
ASN	-0.11	-0.15	-0.01	-0.05	-0.17	-0.14	-0.30	-0.25	-0.35	-0.23	-0.29	-0.28	-0.17	-0.13	-0.18	-0.06	-0.09	
ASP	-0.04	-0.05	-0.05	-0.07	0.04	0.06	-0.04	-0.06	-0.14	-0.18	-0.50	-0.45	-0.24	0.02	-0.02	-0.03	0.08	
CYS	0.08	-0.15	-0.01	-0.19	-0.20	-0.04	-0.13	-0.34	-0.24	-0.02	-0.43	-0.60	-0.33	-0.43	-0.38	-0.48	-0.64	
GLN	0.21	-0.03	-0.09	-0.28	-0.25	-0.18	-0.34	0.08	0.29	0.49	0.37	0.47	0.23	0.04	-0.27	0.38	-0.39	
GLU	0.45	0.81	0.48	0.51	0.53	0.62	0.59	0.68	0.76	0.69	0.25	-0.15	0.02	-0.24	-0.33	-0.29	-0.26	
GLY	0.15	0.12	0.07	-0.08	-0.19	-0.26	-0.68	-1.00	-1.26	-0.80	-0.60	-0.62	-0.51	-0.58	-0.49	-0.46	-0.50	
His	0.05	0.06	0.11	0.16	0.19	0.12	-0.03	0.22	0.25	0.34	0.43	0.61	0.59	0.58	0.51	0.62	0.63	
ILE	-0.07	-0.13	0.20	0.02	0.11	0.25	0.18	0.24	0.26	0.74	0.10	0.70	0.04	0.72	0.79	0.33	0.19	
LEU	-0.23	-0.18	-0.04	0.13	-0.02	0.38	0.29	0.36	0.43	0.45	0.41	0.64	0.36	0.30	0.24	0.23	0.19	
LYS	-0.72	-0.24	-0.05	-0.07	-0.14	-0.00	0.07	0.10	0.89	0.27	0.38	0.42	0.45	0.35	0.24	0.30	0.41	
Met	-0.17	-0.21	0.04	-0.16	0.30	-0.05	0.33	0.34	0.27	0.58	0.64	0.66	0.65	0.40	0.21	0.43	0.23	
PHE	-0.76	-0.17	-0.13	-0.05	-0.02	0.05	0.33	0.40	0.32	0.35	0.30	0.37	0.07	0.19	0.26	0.03	0.17	
PRO	-0.53	-0.58	-0.57	-0.61	-0.58	-0.77	-0.66	-0.86	-1.15	-2.07	-1.77	1.11	-1.02	0.99	0.95	-0.61	-0.62	
SER	-0.20	-0.14	0.01	0.02	-0.11	-0.05	-0.22	-0.29	-0.63	-0.24	-0.38	-0.37	-0.24	-0.23	-0.34	-0.40	-0.37	
THR	-0.20	-0.41	-0.30	-0.38	-0.36	-0.36	-0.28	-0.20	-0.37	-0.33	-0.42	-0.25	-0.19	-0.21	0.75	0.21	0.20	
TRP	0.37	0.58	0.28	0.08	0.00	0.22	0.13	0.05	-0.20	-0.39	-0.55	-0.51	-0.14	-0.02	0.12	0.34	0.34	
TYR	-0.00	0.05	0.04	-0.03	-0.03	0.07	0.07	0.22	0.21	-0.03	0.60	0.12	0.24	0.13	0.14	0.06	-0.45	
VAL	0.08	-0.04	0.01	-0.06	-0.10	-0.01	-0.08	0.09	0.16	0.19	0.20	0.06	0.04	0.12	0.18	0.13	0.14	
	Part B. Input layer to sheet output																	
ALA	-0.07	0.30	0.98	0.06	-0.22	-1.41	-1.41	-1.50	0.10	-1.39	-0.42	-0.18	0.32	0.29	0.37	0.16	-0.78	
ARG	-0.45	-0.30	0.63	0.02	-0.01	-0.05	-0.23	-0.27	-0.28	-0.28	-0.19	0.50	-0.69	-0.60	-0.08	0.36	0.51	
ASN	0.20	0.38	0.38	0.24	0.26	0.10	-0.57	-1.09	-0.74	-0.03	-0.08	0.07	0.75	0.10	0.37	0.56	0.21	
ASP	-0.12	-0.13	-0.07	-0.01	0.04	-0.19	-0.66	-1.21	-1.31	-0.67	-0.48	-0.25	-0.01	-0.04	-0.03	-0.14	-0.26	
CYS	-0.78	0.22	0.00	-0.32	-0.58	-0.81	-0.40	0.07	0.20	0.37	0.09	-0.29	-0.25	-0.06	0.21	0.37	-0.37	
GLN	0.38	0.36	0.26	0.23	0.08	0.02	-0.02	0.11	0.05	-0.23	-0.18	-0.07	0.14	0.11	-0.24	0.08	0.26	
GLU	-0.36	-0.48	-0.21	-0.17	-0.14	-0.13	-0.55	-0.53	-0.61	-0.36	0.31	-0.34	-0.33	-0.14	0.09	0.34	-0.25	
GLY	0.15	0.15	0.24	0.47	0.45	0.35	0.24	-0.92	-0.24	0.23	0.42	0.57	0.56	0.47	0.12	0.22	0.22	
His	0.03	0.07	-0.33	0.00	0.10	0.23	-0.23	-0.57	-0.34	-0.32	-0.43	0.25	0.17	0.11	-0.31	0.11	0.11	
ILE	-0.43	-0.20	-0.50	-0.16	0.00	0.19	0.52	0.99	1.02	0.86	0.21	-0.77	-0.43	-0.66	-0.51	-0.83	-0.23	
LEU	-0.31	-0.35	-0.24	-0.34	-0.43	-0.29	0.27	0.50	0.62	0.50	0.21	-0.08	-0.33	-0.23	-0.20	-0.19	-0.13	
LYS	-0.34	-0.39	0.13	0.13	0.08	0.17	-0.21	-0.32	-0.56	-0.72	-0.66	-0.42	-0.44	-0.66	-0.68	-0.39	-0.34	
Met	0.01	-0.39	-0.60	-0.70	-0.73	0.07	0.26	0.49	0.67	0.70	0.69	-1.53	-1.13	-0.35	-0.50	-0.21	-1.18	
PHE	-0.73	-0.28	-0.30	-0.33	-0.21	-0.23	0.35	0.69	0.82	0.52	0.00	0.15	0.31	-0.01	-0.39	-0.61	-0.04	
PRO	0.35	0.39	0.33	0.61	0.24	0.12	-0.17	-0.81	-2.08	-2.30	-0.59	0.12	0.20	0.24	-0.04	-0.22	-0.21	
SER	0.37	0.23	0.18	0.19	0.26	0.09	-0.17	-0.26	-0.29	-0.04	0.25	0.38	0.42	0.48	0.51	0.34	0.27	
THR	-0.08	0.07	0.31	0.41	0.29	0.23	0.37	0.60	0.31	0.13	0.76	0.39	0.08	0.22	0.34	0.37	0.28	
TRP	-0.07	-0.52	-0.94	-1.12	-0.86	0.10	0.01	0.36	0.49	0.69	0.48	0.04	0.12	-0.22	-0.36	-0.03	0.19	
TYR	0.05	0.12	-0.30	-0.31	-0.36	-0.10	0.55	0.60	0.60	0.31	0.19	0.12	-0.17	-0.09	0.27	0.27	0.16	
VAL	-0.47	-0.33	-0.24	-0.08	-0.10	0.07	0.57	0.94	1.19	1.21	0.61	0.18	0.15	-0.06	-0.20	-0.14	-0.13	

Bias in the output layer are 0.87 for helix and -1.67 for sheet.

წერთის პროცესი ამომწურავადაა აღწერილი Rumelhart et al.-ის ნაშრომში. მოკლედ კი იგი შეიძლება ასე წარმოვადგინოთ. თავიდან ქსელურ წონებს ენიჭებათ (-0.1 ; 0.1) ინტერვალში განსაზღვრული შემთხვევით სიდიდეთა მნიშვნელობები. წერთის თითოეული ციკლის განმავლობაში ფანჯრის თითო ამინმავით წანაცვლებების ხარჯზე ხდება საწერთნელ კომპლექტში შემავალი ყველა ცილის თანამიმდევრობის გავლა. მოცემული ციკლის გავლის პროცესში დახვეწილი წონების გამოყენება ხდება შემდეგ ციკლში. წონათა დახვეწის პროცესი, რომელიც გამოსავალზე რეალურ მნიშვნელობებთან მინიმუმალური განსხვავების მიღწევას ეფუძნება, იყენებს ქვემოთ მოცემულ, განსხვავების დამა-

ხასიათებელი ფუნქციის გრადიენტის საწინააღმდეგოდ მოძრაობის, ანუ მინიმიზირების პროცედურას. განსხვავების აღმწერ გამოსახულებას, რომელიც წონათა და მიდრეკილებათა ფუნქციაა, შემდეგი სახე აქვს:

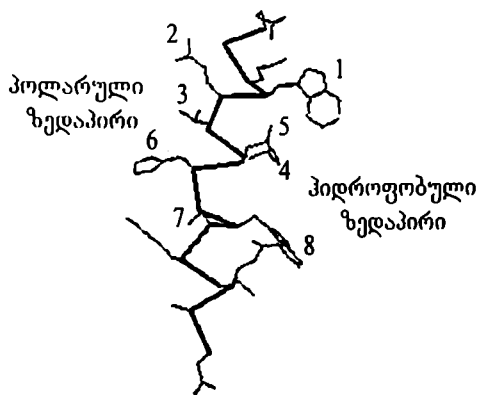
$$E = \sum_c \sum_j (O_{j,c} - D_{j,c})^2,$$

სადაც $O_{j,c}$ წარმოადგენს c -ური ციკლის შედეგად მიღებულ გამოსავალს j -ური ერთეულისათვის $D_{j,c}$ კი სასურველი გამოსავალია. ტრენირების დასამთავრებლად საჭიროა, რომ E -მ შემცირება დაიწყო ასიმპტოტური კანონით (ანუ მომდევნო ციკლებს შორის ფარდობითი ცვლილება იყოს ძალიან მცირე, ვთქვათ არ აღემატებოდეს $2 \cdot 10^{-4}$).

თვალთ ანალიზი

შეიძლება ითქვას, რომ თანამიმდევრობაში ამინმჟავურ ნაშთთა განაწილების ხასიათი მიანიშნებს კონკრეტული მეორეული სტრუქტურის ტიპზე. ასე მაგალითად, α -სპირალის პერიოდულობა 3.6-ია, რაც ნიშნავს, რომ $i, i+3, i+4$ და $i+7$

კემოგლობინის პირველი სპირალი



სურათი 17.3. კემოგლობინის პირველი სპირალის ჰიდროფობული და პოლარული ზედაპირების სქემატური წარმოდგენა.

პოზიციებში განლაგებული ნაშთები სპირალის ზედაპირის ერთ გარკვეულ მხარეს განლაგდება. ასეთი მხარე შეიძლება იყოს მიმართული ცილის შიგთავსისკენ და შეიძლება გამხსნელისაკენ. მართლაც, ცილებში უმრავლესობა α -სპირალებისა ამფიფატურია, რაც ნიშნავს იმას, რომ სპირალის ერთი მხარე მიმართულია ჰიდროფობული შიგთავსისკენ (ბირთვისაკენ) და მეორე გამხსნელისაკენ. ანუ ერთი მხარე ჰიდროფობულია და მეორე პოლარული. ამრიგად, ჰიდროფობულ ნაშთთა განაწილების კონკრეტული სურათი, როცა $i, i+3, i+4, i+7$ პოზიციებში ძირითადად ჰიდროფობული ნაშთებია განთავსებული, მიანიშნებს

α -სპირალის არსებობის ფაქტზე. ამფიფატური სპირალის კლასიკურ ნიმუშად

გამოდგება ცილა ჰემოგლობინის პირველი სპირალი, რომელიც 17.3 სურათზეა წარმოდგენილი.

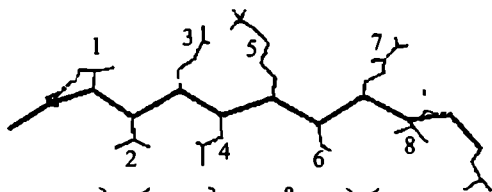
β -გოფირებულ ფურცლოვანი სტრუქტურის ცალკეულ გოფირებულ ზონართა (სეგმენტთა) გომეტრია გულისხმობს, რომ ზონრის მეზობელი ამინმჟავური ნაშთების გვერდითი ჯაჭვები საპირისპიროდაა მიმართული. ხშირად β -სეგმენტის ერთი მხარე მოთავსებულია ცილის შიგნით და სასურველია ჰიდროფობული ნაშთები განლაგდნენ $i, i+2, i+4, i+8$ და ა.შ. პოზიციებში, ხოლო პოლარული ნაშთები $i+1, i+3, i+5$ და ა.შ. პოზიციებში. 17.4 სურათზე წარმოდგენილი CD8-ის β -გოფირებული ფურცლის მეორე სეგმენტი სწორედ ამის კლასიკური ნიმუშია.

β -სეგმენტები, რომლებიც მთლიანად შიგნითაა მოთავსებული (რასაც ხშირად აქვს ადგილი იმ ცილების შემთხვევაში, რომლებიც შეიცავს როგორც α -სპირალს, ასევე β -სეგმენტებს) ჩვეულებრივ შეიცავს ჰიდროფობული ნაშთების

CD8 ცილის გოფირებული ფურცლის მეორე სეგმენტი

სერიას. ასეთი განაწილება იხსნება იმით, რომ გოფირებული ზედაპირის ორივე მხარე მოთავსებულია ცილის შიგნით. 17.5 სურათზე მოცემული ცილა CheY ამის კარგი მაგალითია.

პორალური ზედაპირი



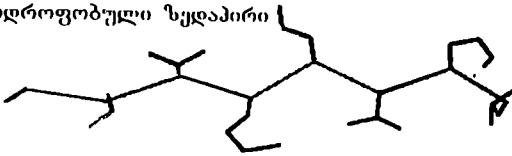
ჰიდროფობული ზედაპირი

მეორეული სტრუქტურის წინასწარმეტყველების მეთოდთა უმრავლესობის პრინციპი მდგომარეობს იმაში, რომ მოიძებნოს ამინმჟავათა ნაშთების კონსერვაციის ისეთი ნიმუშები, რომლებიც ხელს შეუწყობს ზევით მოცემული მეორეული სტრუქტუ-

სურათი 17.4. CD8-ის β -გოფირებული ფურცლის მეორე სეგმენტის ჰიდროფობული და პოლარული ზედაპირების სქემატური წარმოდგენა.

ქემოტაქსისის ცილა CheY
ამინომჟაეური ნაშთები 81-88

ჰიდროფობული ზედაპირი



ჰიდროფობული ზედაპირი

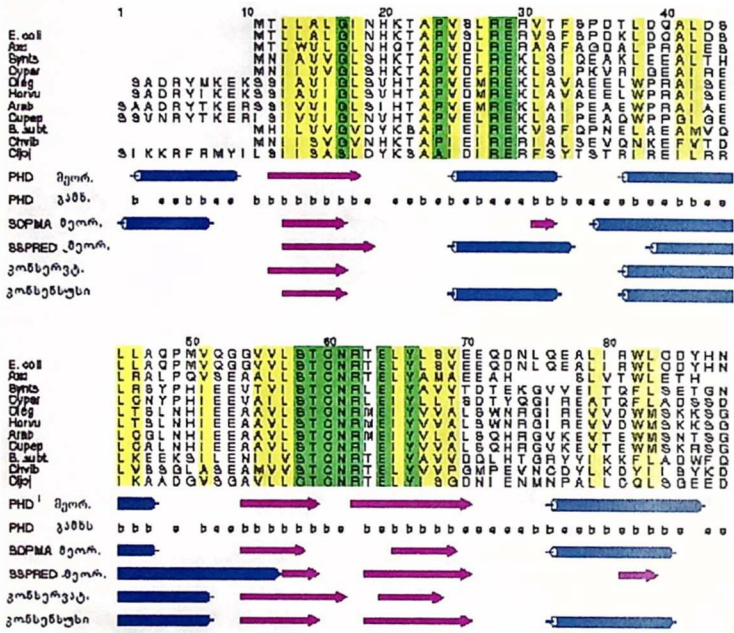
სურათი 17.5. ქემოტაქსისის ცილა CheY-ის გოგორიებული ფურცლის ცილის ინტერიერში განთავსებული სეგმენტის ჰიდროფობული ზედაპირების სქემატური წარმოდგენა

რის ელემენტების სწორ წარმოჩენას. მთელი რიგი მაგალითების საფუძველზე ნაჩვენებია, რომ ეს სტრატეგია ხშირ შემთხვევებში იძლევა თითქმის იდეალური წინასწარმეტყველების შესაძლებლობას.

ცილის მეორეული სტრუქტურის წინასწარმეტყველების სტრატეგია

პრაქტიკაში რეკომენდირებულია გამოყენებულ იქნეს რაც შეიძლება მეტი წინასწარმეტყველების მეთოდი და ცილათა ჰომოლოგიური ოჯახის არსებობის პირობებში, მრავალწევრა ურთიერთსწორებაც. წინასწარმეტყველებათა შორის კონსერვაციის ძიებისას მოშველიებულ უნდა იქნეს მკვლევარის ინტუიცია. ანუ მიზანშეწონილია მოვახდინოთ ყველა ჩვენს მიერ განხორციელებული წინასწარმეტყველების (მათ შორის ამინმჟაეთა ნაშთების განაწილების კონსერვაციაზე დაფუძნებულის) შეჯერება თანამიმდევრობათა მრავალწევრა ურთიერთსწორებასთან, და ამის საფუძველზე შევეცადოთ მივიღოთ სტრუქტურის წინასწარმეტყველების შეთანხმებული სურათი. ქვემოთ, 17.6 სურათზე მაგალითის სახით მოყვანილია ცილათა ოჯახის წევრთა თანამიმდევრობების ერთდროული ურთიერთსწორებისა და სხვადასხვა მეთოდებით მეორეული სტრუქტურის წინასწარმეტყველების მონაცემები. კერძოდ, შეჯერებულია გლუტამილ t-რნმ რედუქტაზას ოჯახში შემავალი 12 ცილის თანამიმდევრობებზე დაყრდნობით მეორეული სტრუქტურის წინასწარმეტყველებისა და ამ თანამიმდევრობათა მრავალწევრა ურთიერთსწორების შედეგები. მეორეული სტრუქტურის წინასწარმეტყველება ხორციელდებოდა სამი სხვადასხვა ავტომატური მეთოდით (PHD, SOPMA და SSPRED). ყვითლადაა აღნიშნული ურთიერთსწორების ის სეგმენტები, რომლებიც აჩვენებენ ჰიდროფობული გვერდითი ნაშთების განაწილების კონსერვაციას.

არპიდროფობული ნაშთების კონსერვაციის მიმანიშნებელი (ზოირად მიმანიშ-
ნებელი აქტიურ უბანზე) სეგმენტები კი, გამოყოფილია მწვანედ.



სურათი 17.6 მეორეული სტრუქტურის წინასწარმეტყველება t-რნმ რედუქტაზას
ოჯახში შემავალ ცილათათვის. წინასწარმეტყველება დაფუძნებულია თანამიმ-
დევრობათა მრავალწევრა ურთიერთსწორებისა და წინასწარმეტყველების განსხვავ-
ებული მეთოდების ერთდროულ გამოყენებაზე.

სურათზე ასევე მოცემულია გამხსნელთან ამინმჟავური ნაშთის მისაწვდომობის
PHD (PHD Acc. Pred.) მეთოდით წინასწარმეტყველების შედეგები (b აღნიშნავს
ცილის შიგნით და e ცილის ზედაპირზე განლაგებულ ამინმჟავებს).
წინასწარმეტყველებისა და ურთიერთსწორების შედეგების ანალიზისა პროცესში
შეიძლება მაგალითად გამოვეყნოთ 38-45 პოზიციები, რომლებიც მიანიშნებენ
ჰიდროფობული ამინმჟავების კლასიკური ამფიფატიური სპირალის შესაბამისად
განაწილების კონსერვაციაზე. ანუ i, i+3, i+4 და i+7 პოზიციები ძირითადად
დაკავებულია ჰიდროფობული ამინმჟავებით, ხოლო მათ შორის ძირითადად
პოლარული ამინმჟავებია. 13-16 პოზიციები უკვე მიანიშნებს ჰიდროფობული

ამინმეავების წ-სტრუქტურისათვის დამახასიათებელი განაწილების კონსერვაციაზე.

სხვადასხვა მეთოდებით მეორეული სტრუქტურის წინასწარმეტყველების შედეგების შედარება აჩვენებს, რომ უმრავლესობა უბნებისათვის ავტომატური და ხელით შესრულებული მეთოდებით მიღებული შედეგები ერთმანეთთან შესაბამისობაშია. მეორეული სტრუქტურის წინასწარმეტყველების სხვადასხვა მეთოდით მიღებული შედეგების შეჯერების საფუძველზე შესაძლებელია აიგოს მეორეული სტრუქტურის ჯაჭვის გასწვრივ განაწილების საერთო სურათი. 17.6 სურათზე შემაჯამებელი განაწილება წარმოდგენილია ბოლო სტრიქონში.

რნმ-ს სტრუქტურის წინასწარმეტყველება

რნმ-ს მეორეული სტრუქტურის წინასწარმეტყველება ეფუძნება დინამიკური პროგრამირების ალგორითმს. მიდგომა გულისხმობს რნმ-ს ნუკლეოტიდურ ჯაჭვში შინაგან, ფრაგმენტშორის, მაქსიმალურ ქულათა ჯამიან ურთიერთსწორების მოძიებას. ქულათა ჯამი პროპორციულია ურთიერთსწორების შედეგად შესაბამისობაში მოყვანილ კომპლემენტარულ ფუძეთა წყვილების რაოდენობისა. ურთიერთსწორების პროცესში სრულდება პირობა - თითოეული ფუძე მხოლოდ ერთი, კონკრეტული კომპლემენტარული წყვილის მონაწილე შეიძლება იყოს. ასეთნაირად წარმოჩენილ, თანამიმდევრულად განლაგებულ კომპლემენტარულ წყვილთა ერთობა ქმნის რნმ-ს მეორეული სტრუქტურის ისეთ ტიპს, როგორცაა ორმაგ-სპირალური ფრაგმენტი; ორმაგ სპირალურ ფრაგმენტის შემქნელ უბნებს შორის განთავსებული, ჯაჭვის სხვა ფრაგმენტებთან კომპლემენტარულად არადაწყვილებადი რნმ-ს უბანი მოიხსენიება ღეროვან ყულფად; ასეთ ყულფთა ერთობამ შეიძლება წარმოქმნას ტრანსპორტული რნმ-ს მოლეკულებისათვის დამახასიათებელი "ნეკერჩხლის ფოთლის" მაგვარი სტრუქტურა (იხ. სურათი 4.10).

დინამიკური პროგრამირების ალგორითმით ურთიერთსწორებისას, ქულათა დარიცხვის სისტემა ეფუძნება ფუძე-ფუძე ურთიერთქმედების თავისუფალი

ენერჯის შეფასებას. შესაბამისად ხდება სხვადასხვა კომპლემენტარულ წყვილთა წარმოქმნაზე სხვადასხვა ქულის დარიცხვა. ალგორითმის რიგ გართულებულ ვარიანტებში ხერხდება ფუძეთა სტეკინგის ეფექტების გათვალისწინებაც. აქვე უნდა აღინიშნოს, რომ დინამიკურ პროგრამირებაზე დაფუძნებული მეთოდით გამწვანებულია ფსევდონოტების (ფსევდომარყუყუების) წარმატებული წინასწარმეტყველება. თუმცა ალგორითმის გართულებით, რომელიც დაკავშირებულია სამანქანო რესურსის დიდ დანახარჯებთან, პრინციპში შესაძლებელია ამ მიმართულებითაც გარკვეული პროგრესის მიღწევა.

არსებობს რნმ-ს მეორეული სტრუქტურის წინასწარმეტყველების სხვა, დინამიკურ პროგრამირებაზე არ დაფუძნებული მიდგომა. იგი ცნობილია „თანამიმდევრობათა კოვარიაციული მეთოდით ანალიზის“ სახელით. კოვარიაციული მიდგომა იყენებს არსებულ მონაცემთა ბაზებს, რომლებიც მოიცავენ ჰომოლოგიურ რნმ-თა დიდი ოჯახებს. ჰომოლოგებში იგულისხმება განსხვავებული თანამიმდევრობების, მაგრამ ნათესაური, საერთო წინაპრის მქონე რნმ-ები. მიდგომა გულისხმობს ევოლუციური პროცესის ანალიზს, რაც ითვალისწინებს კონკრეტულ პოზიციებში ფუძეთა კოვარიაციური ანალიზის საფუძველზე კანონზომიერების გამოვლენას და მათზე დაყრდნობას. ასე მაგალითად, ჰომოლოგიურ რნმ-თა ევოლუციური განვითარების პროცესში ჯაკვის გასწვრივ საკმაოდ დაშორებულ პოზიციებში ერთმანეთთან კომპლემენტარულად დაწყვილებულ ფუძეთა შენარჩუნების (კონსერვაციის) მოვლენა მიუთითებს სტრუქტურისათვის ამ პოზიციებს შორის არსებული წყალბადური ბმების (ფუძეთა დაწყვილების) მნიშვნელობაზე.

18. ცილის სივრცული მოწყობის წინასწარმეტყველება

წინასწარმეტყველების აუცილებლობის მიზეზები

გენომთა სრული ნუკლეოტიდური თანამიმდევრობების (ტექსტების) დასადგენად უკანასკნელი ათწლეულების განმავლობაში დიდი სამუშაო ჩატარდა. გენომთა ტექსტების გაშიფვრის პროცესში მათემატიკურ და კომპიუტერულ მეცნიერებათა მიღწევების გამოყენების ზრდის ტენდენცია მკაფიოდ იკვეთება. შედეგებიც სახეზეა, მიღებულია ასობით ორგანიზმის, მათ შორის ადამიანის გენომის სრული ტექსტები. გენომთა ტექსტების დადგენის პროცესი გრძელდება და ამასთან, ინფორმაციული ნაკადის სიჩქარე სულ უფრო იზრდება. თუმცა უნდა აღინიშნოს, რომ მხოლოდ ტექსტების ცოდნა, ანუ გენთა და შესაბამის ცილათა თანამიმდევრობების ცოდნა ცილათა ფუნქციის ცოდნის გარეშე ვერ წყვეტს თანამედროვე ბიოლოგიისა და მედიცინის უმნიშვნელოვანეს ამოცანას, ცოცხალი ორგანიზმის სასიცოცხლის სისტემათა წარმართველი მოლეკულური მექანიზმების გაგებისა და საჭიროების შემთხვევაში ამ მექანიზმებზე ზემოქმედების საშუალებების გენერირების ამოცანას. თუ გავითვალისწინებთ, რომ ცილის ფუნქციის რეალიზების საფუძველს წარმოადგენს მისი სივრცული სტრუქტურა, ბუნებრივია დნმ-ს ნუკლეოტიდური და ცილათა ამინმჟავური თანამიმდევრობების (ტექსტების) შესახებ მონაცემთა მოპოვების და დამუშავების პარალელურად გაჩნდა საჭიროება ცილათა სივრცულ მოწყობათა შესახებ ინფორმაციის მოპოვებისა, შენახვისა და ანალიზისა. როგორც პარაგრაფ 9-ში ითქვა, სივრცული სტრუქტურების შესახებ მონაცემთა ბაზა (PDB) წარმოადგენს ორთოგონალურ კოორდინატთა სისტემაში განსაზღვრულ ცილის ატომთა კოორდინატების შესახებ ინფორმაციის საცავს. ინფორმაცია ძირითადად ექსპერიმენტული მეთოდების (ცილათა კრისტალოგრაფული და ხსნარში ზირთვული მაგნიტური რეზონანსის (ბმრ)) გამოყენებით მოიპოვება. აღსანიშნავია ისიც, რომ ამ ბაზაში გაერთიანებულ ცილათა რიცხვი გაცილებით მცირეა, ვიდრე ცილათა რიცხვი, რომელთათვის მხოლოდ ტექსტებია ცნობილი და განთავსებული შესაბამის ბაზებში. დნმ-ტექსტთა ბაზები ყველაზე მოცულობითია. შესაბამისად ნათელია, რომ გაშიფრულ გენომთა მთელი რიგი ORF-სთვის (Open Reading Frame) უცნობია შესაბამისი პროდუქტის (ცილის) ბიოლოგიური ფუნქცია, სტრუქტურის თავისებურებანი და ამრიგად ფუნქციონირების მოლეკულური მექანიზმები. ყოველივე ზემოთ თქმულიდან გამომდინარე ნათელია, რომ დნმ-ტექსტების გაშიფვრის (რომელიც დიდ დანახარჯებთან არის დაკავშირებული) შედეგად დაგროვილი ინფორმაციის ეფექტურად გამოყენების შესაძლებლობა დიდად არის დამოკიდებული იმ მეთოდების არსებობასა და განვითარების

დონეზე, რომლებიც თითოეულ გენურ ტექსტს შესაბამისი ცილის სტრუქტურითა და ფუნქციით უზრუნველყოფენ. სწორედ ეს გახდა მიზეზი სტრუქტურული გენომიკის განვითარების წახალისებისა. რაც გამოიხატა იმ ექსპერიმენტული მეთოდების (რენტგენოსტრუქტურული ანალიზისა და ორგანოზომილებიანი ბმრის) დახვეწისა და ეფექტურობის გაზრდის მცდელობაში, რომლებიც ცილის სტრუქტურის დადგენის მძლავრ საშუალებებს წარმოადგენენ. ამ მცდელობებმა მართალია თავისი შედეგი გამოიღო, მაგრამ შეიძლება ითქვას, რომ ეს ექსპერიმენტული მეთოდები დღესაც საკმაოდ ძვირადღირებული და დროში გაწელილი საშუალებებია. ამას გარდა ამ მეთოდებს გარკვეული პრინციპული შეზღუდვებიც ახასიათებთ. მაგალითად, თუ ვერ ხერხდება ცილის კრისტალიზება, ვერ გამოვიყენებთ რენტგენოსტრუქტურულ ანალიზს. ბმრ-ს კი, დღეისათვის მხოლოდ მცირე ზომის (50-100 ამინომჟავას შემცველი) ცილების სტრუქტურის დადგენის შესაძლებლობა გააჩნია. სწორედ ექსპერიმენტთან დაკავშირებული პრობლემები გახდა საფუძველი შექმნილიყო სტრუქტურული გენომიკის პრობლემების გადამწყვეტი ახალი, სწრაფი და იაფი, კომპიუტერ მეცნიერებათა მიღწევებზე დაფუძნებული მიდგომები. მიდგომების საშუალებით შესაძლებელია გენის ნუკლეოტიდური ან შესაბამისი ცილის ამინომჟავური თანამიმდევრობის შესახებ ინფორმაციაზე დაყრდნობით ცილის სტრუქტურის ტიპისა და ფუნქციის გამოცნობა, ან ნატივურთან მიახლოებული სივრცული მოდელის აგება.

ამრიგად, ცილის მოდელირების ძირითად მიზანს წარმოადგენს თანამიმდევრობაზე დაყრდნობით ისეთი სტრუქტურის წინასწარმეტყველება, რომელიც სიზუსტით შესადარი იქნება ექსპერიმენტულ მონაცემებთან. ასეთი მიზნის მიღწევის შემთხვევაში თამამად შეგვეძლება გამოვიყენოთ სულ უფრო მეტი *in silico* მოდელები თანამედროვე ბიოლოგიისა თუ მედიცინის მთელ რიგ სფეროებში. მაგალითად: სტრუქტურულ მონაცემებზე დაყრდნობით ბიოლოგიურად აქტიურ ნაერთთა (წამალთა) პროექტირებისას; ცილის ფუნქციის კვლევისას; ანტიგენურ დეტერმინანტთა გამოვლენის პროცესში; სხვა მოლეკულებთან ურთიერთქმედების შესაძლებლობის დადგენის პროცესში; გაზრდილი სტაბილურობის ან ახალი ფუნქციის მატარებელ ცილათა პროექტირების პროცესში და სხვა მრავალ შემთხვევაში. არის შემთხვევები, როცა ექსპერიმენტი უძლურია დაადგინოს ცილის სტრუქტურა და ასეთ შემთხვევებში კომპიუტერული მოდელირება წარმოადგენს ერთადერთ საშუალებას. მაგალითად, როგორც ზემოთ აღვნიშნეთ, ვერ ხერხდება მთელი რიგი დიდი ზომის ცილების სტრუქტურათა ბირთვული მაგნიტური რეზონანსის მეთოდით დადგენა. ზოგიერთი ცილისთვის კი პრობლემას კრისტალიზაცია ქმნის.

ცილის სივრცული სტრუქტურის წინასწარმეტყველების მეთოდები

ფიზიკურ კონონზომიერებებზე დაფუძნებული მოდგომები. ab initio წინასწარმეტყველება.

მოდით მიმოვიხილოთ ისტორიულ ჭრილში თანამედროვე ბიოლოგიის უმნიშვნელოვანესი ამოცანის, კერძოდ კი სტრუქტურული გენომიკის ცენტრალური პრობლემის კომპიუტერული მიდგომებით გადაწყვეტის პროცესი. შეიძლება ითქვას, რომ კომპიუტერული მოლეკულური ბიოლოგიის პირველი მეთოდები შექმნილი იყო სტრუქტურული ბიოლოგიის ამოცანების, მათ შორის სტრუქტურის წინასწარმეტყველების პრობლემის გადაწყვეტის მიზნით. პირველი მეთოდები მხოლოდ ფიზიკურ-მათემატიკურ ცოდნას ეფუძნებოდა და შესაბამის კონონზომიერებებს ითვალისწინებდა. ამიტომ მათ ფიზიკურ მიდგომებად მოვიხსენიებთ და აღვნიშნავთ, რომ ისინი ემყარებოდნენ შემდეგ პოსტულატს: პოლიპეპტიდური ჯაჭვის ნატივური სტრუქტურა შეესაბამება ცილა-გამხსნელი-საგან შემდგარი ჩაკეტილი სისტემის წინასწორულ მდგომარეობას. ამ ფიზიკური მოდელის შესაბამისი მათემატიკური ამოცანა გულისხმობდა ცილისა და გამხსნელის ატომებისაგან შემდგარი სისტემის კონფორმაციული მდგომარეობის დახასიათებას ატომ-ატომურ ურთიერთქმედებათა აღმწერი გარკვეული პოტენციური ფუნქციით და მის მრავალგანზომილებიანი ზედაპირის კვლევას ლოკალური მინიმუმების დაფიქსირებისა და გლობალურის პოვნის მიზნით. ნატივური კონფორმაცია, როგორც უკვე აღინიშნა შეესაბამება სისტემის წინასწორულ მდგომარეობას და ე.ი. მისი აღმწერი ფუნქციის გლობალურ მინიმუმს. ამრიგად გლობალური მინიმუმის პოვნა ნიშნავს ნატივური ცილის სივრცული მოწყობის დადგენის ამოცანის გადაწყვეტას. ამრიგად ცილის ab initio მოდელირების (მხოლოდ ამომომავური თანამიმდევრობის შესახებ ინფორმაციაზე დაყრდნობით) ამოცანის (რომელიც ასევე მოიხსენიება, როგორც “დახვევის პირდაპირი” ამოცანა) გადაწყვეტისათვის შემუშავებული მიდგომების წარმატება დიდწილად დამოკიდებულია მრავალი ცვლადის ფუნქციის გლობალური მინიმუმის მოძიების ამოცანის წარმატებულ გადაწყვეტაზე. შესაბამისად სტრუქტურის მოდელირების მიზნით შექმნილი კომპიუტერულ მეთოდთა სიმრავლე კონფორმაციული ენერჯის მინიმუმის მოძიების საშუალებებიდან გამომდინარე, შეიძლება დაყოფილ იქნეს სამ ძირითად ჯგუფად:

- კლასიკური კონფორმაციული ანალიზი (მოლეკულური მექანიკის მეთოდი)
- მოლეკულური დინამიკის მეთოდები
- მონტე-კარლოს მეთოდები

მოკლედ მიმოვიხილოთ თითოეული.

მოლვეკულური მექანიკის მეთოდი (მმ) გლობალური მინიმუმის მოძიებისას გულისხმობს სუფთა მათემატიკურ მიდგომას ანუ პოტენციური ფუნქციის მრავალგანზომილებიანი ზედაპირის კვლევას მრავალი ცვლადის ფუნქციის მინიმიზირებისათვის განკუთვნილი რიცხვითი მეთოდების გამოყენებით. ასეთ რიცხვით მეთოდთა მაგალითად შეიძლება მოვიყვანოთ შეუღლებულ გრადიენტთა მეთოდი. შეუღლებულ გრადიენტთა მეთოდის გამოყენების შემთხვევაში საჭირო ხდება ფუნქციის პირველი და მეორე რიგის ყველა კერძო წარმოებულის შეფასება, რადგან იტერაციის ნებისმიერ ეტაპზე თითოეული ცვლადისათვის ნაზრდის მინიჭება და შესაბამისად ფუნქციის ზედაპირის მოცემული წერტილიდან მინიმუმისაკენ მიმართულების განსაზღვრა სწორედ მინუს ნიშნით აღებული გრადიენტით (კერძო წარმოებულებით) წარმოებს. ალგორითმი ყოველი იტერაციული საფეხურის ბოლოს გულისხმობს ენერჯის მნიშვნელობის შემცირებას წინა საფეხურთან შედარებით და ასეთი გზით ლოკალური მინიმუმის მიღწევას. იმისათვის, რომ მოხდეს მთელი კონფორმაციული სივრცის დახასიათება და შესაბამისად, გლობალური მინიმუმის მოძიება, აუცილებელია მინიმიზაციის პროცესი განხორციელდეს დიდი რაოდენობა საწყისი წერტილებიდან. ამომწურავი ანალიზისათვის აუცილებელი საწყისი წერტილების რაოდენობა კი, უშუალო კავშირშია საკვლევი პოლიპეპტიდის (როგორც ძირითადი, ისე გვერდითი ჯაჭვის) სიგრძესთან, ანუ კონფორმაციული თავისუფლების ხარისხებთან. ამ უკანასკნელის ზრდასთან ერთად, სრული ანალიზისათვის საჭირო საწყისი წერტილების რაოდენობა დრამატულად იზრდება. საერთოდ, ამ მეთოდით სრული ანალიზი შესაძლებელია მხოლოდ რამდენიმე ამინმჟავას მომცველი პეპტიდისათვის. ასე მაგალითად, ჩატარდა სხვადასხვა ამინმჟავის, მათ შორის ალანინისა და გლიცინის მომცველი დიპეპტიდების კონფორმაციული ენერჯის აღმწერი პოტენციური ფუნქციის კვლევა. შედეგად ალანინისათვის კონფორმაციულ სივრცეში გამოიყო ოცამდე განსხვავებული ლოკალური მინიმუმის მომცველი უბანი. გლიცინისათვის ცოტა მეტი (იხილეთ ალანინისა და გლიცინის რამაჩანდრანის რუქები 5 პარაგრაფში). მინიმუმები, რომლებიც გამოვლინდა ალანინისათვის, გამოყენებულ იქნა სხვა, C^β ატომის შემცველი ამინმჟავების ანალიზის პროცესში ძირითადი ჯაჭვის საწყის კონფორმაციათა განსასაზღვრავად. ამრიგად, შეიძლება ითქვას, რომ თითოეული ამინმჟავას ძირითადი ჯაჭვის კონფორმაციული სივრცე საშუალოდ ხასიათდება არა ნაკლებ ოცი მეტნაკლებად სტაბილური კონფორმაციით. გვერდითი ჯაჭვების დამატებითი თავისუფლების ხარისხების გამო, ალანინზე გრძელი გვერდითი ჯაჭვის მქონე ამინმჟავების სრული კონფორმაციული შესაძლებლობანი უკვე შესაძლებელია განისაზღვროს ოცზე მეტი შესაძლო კონფორმაციით. ამიტომ n რაოდენობა ამინმჟავასაგან შემდგარი პეპტიდის სრული ანალიზის პროცესში განსახილველ საწყის წერტილთა რაოდენობა ვერ იქნება 20^n -ზე ნაკლები. აქედან გამომდინარე, პოლიპეპტიდის თავისუფლების ხარისხის შემზღუდველი დამატებითი პირობების არარსებობის შემთხვევაში, კომპიუტერულ საშუალებათა შესაძლებლობებიდან გამომდინარე,

საშუალოზე მეტი სიგრძის ($m > 10$) პეპტიდებისათვის (არაფერს ვამბობთ ცილებზე) სრული კონფორმაციული ანალიზი პრაქტიკულად შეუძლებელი ხდებოდა. უფრო მეტიც, უკვე პენტაპეპტიდების შემთხვევაში სრულ კონფორმაციულ ანალიზს ატარებდნენ ე.წ. "ბლოკური" მიდგომით, როცა პეპტიდს ანაწევრებდნენ გადაფარვად დიპეპტიდებად და შემდეგ ტრიპეპტიდებად და ცალცალკე ჩატარებული ანალიზის შედეგების ურთიერთშეჯერების საფუძველზე არჩევდნენ სრული პეპტიდის ანალიზისათვის აუცილებელ საწყის წერტილთა ერთობლიობას. მოკლედ, მოცემული მეთოდის შეზღუდული შესაძლებლობები გახდა მიზეზი ახალი მიდგომების ძიებისა.

მოლეკულური დინამიკის მეთოდი (მდმ) ასევე გულისხმობს კონფორმაციული ენერჯის მრავალგანზომილებიანი ზედაპირის აღმწერი ფუნქციის კვლევას გლობალური მინიმუმის შემცველი უბნის გამოვლენის მიზნით. მაგრამ მმ მეთოდისაგან განსხვავებით გლობალური მინიმუმისაკენ მოძრაობა მოდელირდება სითბურ პროცესთა სიმულირებაზე დაყრდნობით. სხვანაირად რომ ვთქვათ, მდმ პოლიპეპტიდური ჯაჭვის სივრცულ სტრუქტურად დახვევის სიმულირებას ეწევა კომპიუტერის დახმარებით. კომპიუტერზე სიმულირებისას პოლიპეპტიდის ნატივურ სტრუქტურას განსაზღვრავენ იგივე ფაქტორები, რომლებზეც ზემოთ იყო საუბარი, მაგრამ ამასთან ითვლება, რომ პოლიპეპტიდური ჯაჭვის დახვევის პროცესს სითბური პროცესები წარმართავენ. ამრიგად მდმ-ის შემთხვევაში მოცემული საწყისი კონფორმაციიდან წონასწორული მდგომარეობისაკენ (ან გლობალური მინიმუმისაკენ) მოძრაობის სიმულირებისას ითვლება, რომ მოძრაობას სითბური პროცესები წარმართავენ. მდმ-სგან განსხვავებით, მმ მეთოდის გამოყენებისას მინიმუმისაკენ მოძრაობა წარმართება წმიდნა მათემატიკურ, აბსტრაქტულ მიდგომაზე დაფუძნებით და აქედან გამომდინარე, უფრო დიდია პოტენციური ფუნქციის მრავალგანზომილებიანი ზედაპირის სრული სკანირებისათვის აუცილებელი საწყისი კონფორმაციების რაოდენობა. სითბური პროცესით განპირობებული ცილის მოლეკულის სტრუქტურის ცვლილების დინამიკის აღსაწერად შემუშავებული ალგორითმი მარტივია და გულისხმობს განსახილველი მოლეკულის თითოეული ატომისათვის კლასიკური მოძრაობის განტოლების ამოხსნას

$$dU/dR = m d^2R/dt^2 \quad (18.1)$$

აქ U კონფორმაციული ენერჯიაა, R - გადაადგილება. ალგორითმი მოკლედ ასე შეიძლება აღიწეროს. გარკვეულ სითბურ ველში, ანუ T -ს გარკვეული მნიშვნელობისათვის საკვლევი მოლეკულის (რომელსაც გააჩნია გარკვეული საწყისი კონფორმაცია) ყოველი ატომი ხასიათდება კოორდინატისა და სიჩქარის გარკვეული მნიშვნელობებით. საწყის მომენტში გარკვეული მნიშვნელობა გააჩნია კონფორმაციულ ენერჯიასაც. ამ საწყისი პირობების გათვალისწინებით, დროის გარკვეული ინტერვალისათვის (როგორც წესი პიკოსეკუნდის რიგის) იხსნება ზემოთხსენებული (18.1) განტოლება, რის შედეგადაც მოცემული ატომი ხვდება

ახალ სივრცულ მდგომარეობაში და ასეთი მდგომარეობის მიღების ან მიუღებლობის შესაძლებლობა განპირობებულია ენერჯის მნიშვნელობის შემცირების მოთხოვნით. საკვლევი მოლეკულის თითოეული ატომისათვის ზემოთ აღწერილი პროცესის მრავალჯერად გამოყენებას მივყავართ რაღაც სტაციონალურ მდგომარეობამდე, ანუ მრავალგანზომილებიანი კონფორმაციული სივრცის ისეთ უბნამდე, საიდანაც თავის დაღწევა მოცემულ ტემპერატურაზე (ანუ სითბური პროცესებით გამოწვეული შემდგომი ისეთი გადაადგილება, რომ თავი დაავაღწიოთ პოტენციური ფუნქციის მრავალგანზომილებიანი ზედაპირის მოცემულ უბანს) შეუძლებელი ხდება. თუ ასეთი უბანი შესაბამება მხოლოდ ლოკალურ მინიმუმს, მაშინ გლობალურისაკენ შემდგომი მოძრაობა და ე.ი. იმ ენერგეტიკული ბარიერის დაძლევა (აქვე აღვნიშნავთ, რომ ბარიერის ასეთი დაძლევა მშეთოდის შემთხვევაში, როცა გამოიყენება ფუნქციის მინიმუმაციის რიცხვითი მეთოდები, შეუძლებელია), რომელიც შემოსაზღვრავს მოცემულ ლოკალურ მინიმუმს, შესაძლებელია მხოლოდ გარემოს ტემპერატურის ხელოვნური გაზრდით. ერთი სიტყვით, მდმ დროის ინტერვალისა და გარემოს ტემპერატურის მნიშვნელობის რეგულირების ხარჯზე პროცესის გარკვეული საწყისი კონფორმაციიდან დაწყების შემთხვევაში იძლევა პოლიპეპტიდური ჯაჭვის კონფორმაციული სივრცის სრული სკანირების საშუალებას. ამასთან უნდა აღვნიშნოს, რომ ამ ალგორითმის რეალიზებისათვის საჭიროა მძლავრი კომპიუტერული სისტემები, რადგან იგი მოითხოვს მეხსიერების დიდ მოცულობასა და ცენტრალური პროცესორის დიდ სიჩქარეს. კომპიუტერული ტექნიკის განვითარების დღევანდელი დონე საშუალებას იძლევა ამ მეთოდით სრულად იქნეს გაანალიზირებული საშუალო სიდიდის სისტემის (საშუალოდ 150-200 ატომის მომცველი) კონფორმაციული შესაძლებლობები. მაგრამ კიდევ ერთხელ აღვნიშნავთ, რომ მდმ მოითხოვს მძლავრი ეგზის დიდ სამანქანო დროს, რაც ამოცანას მეტად ძვირადღირებულს ხდის და ამიტომ ამ მეთოდით კონფორმაციული ამოცანის გადაწყვეტა არაა ყველასათვის მისაწვდომი. ქვემოთ შევეცდებით ავღწეროთ მდმ-ზე უფრო იაფი მეთოდი და ამასთან, მმ-ზე მეტი შესაძლებლობების მქონე (რეალურ სამანქანო დროში 80-100 ატომიანი სისტემის კონფორმაციული ანალიზის შესაძლებლობა).

მონტე-კარლოს მეთოდი გულისხმობს შემთხვევით რიცხვთა გენერატორისა და სტატისტიკური მეთოდების გამოყენებას. ასეთი მიდგომა საშუალებას იძლევა დახასიათდეს საკვლევი პეპტიდის თითოეული ამინმჟავას კონფორმაციული სივრცის თითოეული უბანი მასში შემავალი კონფორმაციების რეალიზაციის ალბათობით. ამისათვის აუცილებელია სტატისტიკური წონების განმსაზღვრელი შემდეგი ინტეგრალების შეფასება:

$$Z(\beta) = \int \dots \int \exp[-\beta U\{\phi, \psi, \chi\}] d\{\phi, \psi, \chi\} \quad (18.2)$$

$$P_i(\phi_i, \psi_i) = 1/Z \int \dots \int \exp[-\beta U\{\phi', \psi', \chi\}] d\{\phi', \psi', \chi\} \quad (18.3)$$

სადაც ϕ, ψ ძირითადი ჯაჭვის, χ კი გვერდითი ჯაჭვის კონფორმაციის განმსაზღვრელი ორწახნაგა კუთხეებია. U კონფორმაციული, პოტენციური ენერგიაა. რომელიც ბმათა სიგრძეებისა და ვალენტური კუთხეების მუდმივობის პირობებში მხოლოდ ორწახნაგა კუთხეების ფუნქციაა.

პოტენციური ფუნქცია

ცილის სტრუქტურის მოდელირებისას გამოიყენება პოტენციური ფუნქციების მრავალი სახეობა. ჩვენი მიზანია წარმოვადგინოთ მათი ზოგადი სახე. ზოგადი სახით U ატომ-ატომური ურთიერთქმედების პოტენციური ფუნქცია (ააპფ) წარმოადგენს მთელი რიგი წევრების, ცალკეული ფუნქციების ჯამს. ასე მაგალითად

$$U = U_{\text{ბმ}} + U_{\text{სკ}} + U_{\text{სგ}} + U_{\text{ტრ}} + U_{\text{ჯლ}} + U_{\text{ფ}} + U_{\text{სოლო}} \quad (18.4)$$

ფუნქცია წარმოადგენს ჯამს, სადაც თითოეულ წევრს, ფუნქციას აქვს თავისი დატვირთვა. ქვევით შევეცდებით წევრების დახასიათებას.

ვალენტური კუთხეებისა ($U_{\text{სკ}}$) და ბმათა სიგრძეების ($U_{\text{ბმ}}$) დეფორმაციის ენერგიების გამოსათვლელად გამოიყენებიან შემდეგი ტიპის გამოსახულებები:

$$U_{\text{ბმ}} = K_i (l_i - l_i^0)^2 \quad (18.5)$$

$$U_{\text{სკ}} = C_i (\alpha_i - \alpha_i^0)^2 \quad (18.6)$$

სადაც K_i და C_i დრეკაობის კოეფიციენტებია, l_i და α_i ბმის სიგრძეთა და ვალენტურ კუთხეთა რეალური მნიშვნელობები, ხოლო l_i^0 და α_i^0 მათი იდეალური მნიშვნელობები.

არავალენტურ ურთიერთქმედებათა ენერგიას $U_{\text{სგ}}$ შემდეგი სახე გააჩნია:

$$U_{\text{სგ}} = -A' r^{-6} + B' r^{-12} \quad (18.7)$$

A' პარამეტრები მიიღება კვანტურ-ქიმიური გათვლების საფუძველზე, ხოლო B' კი დაბალმოლეკულურ, მოდელოვან ნაერთთა კრისტალების რენტგენოგრაფიული მონაცემების ანალიზის საფუძველზე.

ტორსიონული წევრის $U_{\text{ტრ}}$ ჩართვა გამოწვეულია იმით, რომ არავალენტური ურთიერთქმედება არასწორად აღწერს ქიმიური ბმის ირგვლივ ბრუნვით გამოწვეულ ენერგიის ცვლილების ხასიოს. ასე მაგალითად, ეთანის ($\text{H}_3\text{C} - \text{CH}_3$) მოლეკულაში C-C ბმის ირგვლივ ბრუნვისას C-H ბმების ურთიერთადაფარვით გამოწვეული ბარიერის სიდიდე 12 კჯ/მოლეკულა, მაშინ როცა არსებული

პოტენციალები არ ასახავდა ასეთი ბარიერის არსებობის ფაქტს. ამ მდგომარეობის გამოსწორების, ანუ გამოთვლების ექსპერიმენტულ მონაცემებთან შესაბამისობაში მოყვანის მიზნით პოტენციალური ფუნქციის გამოსახულებაში დამატებულია ტრასიონული წევრი. ძირითადად გამოიყენება ასეთი სახის ფუნქცია:

$$U(\phi) = (U_0/2) (1 + \cos n\phi) \quad (18.8)$$

მაგალითად C-C ბმისათვის $n=3$ და $U_0=12$ კჯ/მოლი.

U_{Σ} -ის გამოთვლისას ხმარობენ ან მონოპოლ-მონოპოლურ, ან დიპოლ-დიპოლურ მიახლოებას. კონფორმაციული გათვლებისას უფრო პოპულარულია მონოპოლ-მონოპოლური მიახლოება, რომელიც ემყარება ცალკეულ ატომებზე პარციალური მუხტის სიდიდის განსაზღვრას. ამ მუხტის სიდიდეებს, ჩვეულებრივ აფასებენ კვანტური ქიმიის მეთოდებით გათვლების საფუძველზე. თუ პარციალური მუხტის სიდიდე გამოსახულია მუხტის ელექტრონულ ერთეულებში (ელექტრონის მუხტის ერთეულებში), მანძილი ნანომეტრებში, ხოლო ენერგია კჯ/მოლი, მაშინ

$$U_{ij}^{\Sigma} = 131 (q_i q_j) / (\epsilon r_{ij}) \quad (18.9)$$

დიპოლ-დიპოლურ მიახლოებას არ გააჩნია უპირატესობა მონოპოლურთან შედარებით. ამასთან უნდა აღინიშნოს, რომ დიპოლ-დიპოლური ურთიერთქმედებათა შეფასებისას ვალენტური ბმების დიპოლური მომენტები მიიღება არა კვანტურ-ქიმიური გათვლებით, არამედ მცირე, მოდელებად შერჩეულ მოლეკულათა ექსპერიმენტულად მიღებული დიპოლური მომენტების მდგენელებად დაშლით. დიპოლ-დიპოლური ურთიერთქმედების ენერგიის შეფასებისას გამოიყენება ფორმულა:

$$U_{ij}^{\Sigma} = \mu_i \mu_j [\cos \chi_{ij} - 3 \cos(\alpha_{1ij}) \cos(\alpha_{2ij})] / (\epsilon r_{ij}^3) \quad (18.10)$$

სადაც μ_i და μ_j ბმათა დიპოლური მომენტებია. χ_{ij} , α_{1ij} და α_{2ij} კუთხეებია შესაბამისად μ_i და μ_j -ს, μ_i -სა და r_{ij} -ს, და μ_j -სა და r_{ij} -ს შორის.

წყალბადური ბმის ენერგიის შეფასებისას გამოიყენება უამრავი სხვადასხვა სახის პოტენციალი, მათ შორის გამოვიყოფთ ერთ-ერთ ყველაზე პოპულარულს

$$U_{\text{ბ}} = -C r^{-6} + D r^{-10} \quad (18.11)$$

r - ატომებს შორის მანძილია, C და D - ემპირიული პარამეტრებია.

ბიოლოგიურ კონონზომიერებებზე დაფუძნებული მიდგომები

ბიომაკრომოლეკულების სამგანზომილებიანი სტრუქტურის წინასწარმეტყველების ფიზიკურ საფუძველს დამყარებული მიდგომები უკვე რამდენიმე ათეული წელია მეტნაკლები წარმატებით აკეთებენ თავის საქმეს. თუმცა შეიძლება ითქვას, რომ მათ სათანადოდ ვერ უპასუხეს იმ გამოწვევას, რომელიც თანამედროვე ბიოლოგიისა და მედიცინის მიერ სტრუქტურული გენომიკის ძირითადი პრობლემის გადაჭრის მოთხოვნის სახით იქნა წარმოდგენილი. შესაბამისად გაჩნდა ახალი ტალღა კომპიუტერული მიდგომებისა, რომლებიც ცილის დახვევის ტიპის წინასწარმეტყველებისა და სივრცული მოდელის აგების ამოცანის გადაწყვეტას უკვე არა ფიზიკურ (ან არა მხოლოდ ფიზიკურ) არამედ ბიოლოგიურ საფუძველზე დაყრდნობით ახორციელებენ.

სანამ შევხებოდეთ ახალ მიდგომათა ბიოლოგიურ საფუძველებს, კიდევ ერთხელ ჩამოვაცალიბოთ სტრუქტურული გენომიკის ამოცანა: არსებობს გენის ტექსტი და მოსაპოვებელია ინფორმაცია შესაბამისი ცილის სივრცული მოწყობისა და ფუნქციის შესახებ.

გამართლებულია მოსაზრებები: ა) ბიოლოგიური ფუნქციის თვალსაზრისით მსგავსი ცილების სივრცული სტრუქტურები მსგავსია; ბ) არსებობს მართალია არა ცალსახა, მაგრამ დამოკიდებულება ცილის სივრცულ მოწყობის ტიპსა და მის ტექსტს შორის (შებრუნებული დამოკიდებულება ცალსახაა); უკანასკნელი მოსაზრებიდან გამომდინარე, მსგავსი ფუნქციის მქონე ცილების ტექსტებში შეიძლება ვივარაუდოთ საერთო ტექსტური მოტივების არსებობა, ანუ ტექსტებს შორის რაღაც დონის მსგავსება. ყოველივე ეს იძლევა შესაძლებლობას ვამტკიცოთ, რომ თუ გენური, ან ცილოვანი ტექტები მსგავსია, შეგვიძლია თამამად ვილაპარაკოთ ცილათა ფუნქციურ მსგავსებაზე. ამრიგად, უნდა იქნას სტრუქტურის მქონე საკვლევი ტექსტის ცნობილი სტრუქტურისა და ფუნქციის ტექსტთან მსგავსებამ შეიძლება უზრუნველგვეყოს საკვლევი ტექსტის ფუნქციის შესახებ ინფორმაციით. ამდენად, დღესდღეობით მეტად აქტუალურია ტექსტთა (ასოთა რიგთა) შედარების მეთოდები, რომლებიც გავრცელებულია სხვადასხვა კომერციული პაკეტების სახით. მიუხედავად იმისა, რომ ეს პაკეტები სრულიად განსხვავებული ბიოლოგიური პრობლემების გადასაწყვეტადაა შექმნილი, ისინი წარმოადგენენ ერთი და იმავე დინამიკური პროგრამირების ბაზური ალგორითმის წარმოებულს. ბაზური ალგორითმი შეიქმნა სპეციალურად ბიოლოგიური ტექსტების შედარების მიზნით, მაგრამ შეიძლება გამოყენებულ იქნეს ნებისმიერი ტექსტისათვის (იხ ლექცია 10). ბიოლოგიურ საფუძველზე დამყარებულ წინასწარმეტყველების მეთოდების მიმოხილვა დაიწყოთ ჰომოლოგიის საფუძველზე მოდელირების მეთოდით.

ჰომოლოგიის საფუძველზე მოდელირება

იმ ძირითად მიდგომებს შორის, რომლებიც გამოიყენება ცილის სამგანზომილებიანი სტრუქტურის წინასწარმეტყველების პროცესში, ჰომოლოგიურობის საფუძველზე მოდელირება ყველაზე მარტივია და ეფუძნება შემდეგ ორ კანონზომიერებას:

1. ცილის სივრცული სტრუქტურა ცალსახად განისაზღვრება მისი ამინმჟავური თანამიმდევრობით. თანამიმდევრობის ცოდნა, თეორიულად მაინც, საკმარისია ცილის სივრცული სტრუქტურის განსასაზღვრავად.

2. ევოლუციური ზეწოლის პირობებში სტრუქტურა უფრო სტაბილურია და უფრო ძნელად იცვლება, ვიდრე თანამიმდევრობა. ასე რომ, მსგავსებულ თანამიმდევრობათა სტრუქტურები თითქმის იდენტურია და უფრო მეტიც, შესაძლებელია შორეულ ნათესაობაში მყოფ თანამიმდევრობებსაც ახასიათებდეს მსგავსი სივრცული მოწყობები. ასეთ ურთიერთმიმართებათა არსებობა პირველად წარმოჩინდა 1986 წელს Chothia და Lesk -ის მიერ, ხოლო შემდგომ, 1991 წელს რაოდენობრივადაც დახასიათდა Sander და Schneider-ის მიერ. 1999 წელს კი PDB ბანკის მნიშვნელოვნად გაზრდილი მონაცემების ანალიზის საფუძველზე დადგინდა ურთიერთმიმართების განმსაზღვრელი ისეთი ზღვრული მახასიათებლები, რომელთა საშუალებითაც ხდება იმ ზონის გამოვლენა, სადაც თამამად შეიძლება თანამიმდევრობათა შორის მსგავსების შეფასების საფუძველზე ნათესაობაზე საუბარი. ამ ზონის განსაზღვრა შეიძლება ურთიერთსწორებულ ამინმჟავათა რიგების იდენტურ ამინმჟავათა პროცენტზე დამოკიდებულების სურათით, რომელიც მიიღება მსგავსი სტრუქტურების მქონე ცილათა მონაცემების ანალიზის საფუძველზე. ურთიერთსწორებულ ამინმჟავათა რიგებისა და იდენტურ ამინმჟავათა პროცენტის ზღვრული მიახლოებითი მნიშვნელობები, რომლებიც განსაზღვრავენ ნათესაობის დამაჯერებლად დადგენის არეს, შემდეგია: ურთიერთსწორებულ ამინმჟავათა რიგებისათვის 50 და იდენტურ ამინმჟავათა პროცენტისათვის 35. ანუ, ვთქვათ, თუ A თანამიმდევრობის (რომლის სიგრძეა 100 და სტრუქტურა უცნობია) B-სთან (სიგრძეა 200 და სტრუქტურა ცნობილია) ურთიერთსწორებისას ურთიერთსწორებულ ფრაგმენტთა საერთო სიგრძე შეადგენს 100-ს (>50) და ამ დროს ურთიერთსწორებულ წყვილთა 40% (>35%) იდენტურია, დიდი დამაჯერებლობით შეიძლება ითქვას, რომ A ცილა (შესაბამისად მისი სტრუქტურა) ნათესაურია B ცილის შესაბამისი ფრაგმენტის, ანუ მათ გააჩნიათ საერთო წინაპარი. A-ს უწოდებენ საკვლევ თანამიმდევრობას, B-ს კი, მოძიებულ ნიმუშს, რომელზეც დაყრდნობითაც უნდა წარიმართოს მოდელირება. რეალურად ჰომოლოგიურობის საფუძველზე მოდელირება მრავალეტაპიანი პროცესია და შესაძლებელია გამოიყოს შეიდი ძირითადი ეტაპი:

1. ნიმუშის მოძიების მიზნით თანამიმდევრობათა ურთიერთსწორება;
2. ურთიერთსწორების კორექტირება;
3. ძირითადი ჯაჭვის სტრუქტურის გენერირება;
4. ყულფების მოდელირება;
5. გვერდით ჯაჭვთა მოდელირება;

6. ერთიანი მოდელის ოპტიმიზაცია;
7. მოდელის დადასტურება (რატიფიკაცია);

ყველა ამ ეტაპზე მოდელირების შემსრულებელს უხდება სუბიექტური არჩევანის გაკეთება. საჭიროა ამ არჩევანის კანონიკურ ჩარჩოებში მოქცევა და პროცესის ავტომატიზირება. მოდელირების ხარისხისა და ავტომატიზირების დღევანდელმა დონემ შესაძლებელი გახადა გენომებში გენებით კოდირებულ ამინმჟავათა საშუალოდ 25%-ის სტრუქტურის მოდელირება, რითაც მნიშვნელოვანი წვლილი შეიტანა სტრუქტურული გენომიკის ამოცანის გადაჭრის საქმეში. გამოდის, რომ გენურ თანამიმდევრობათა დაახლოებით 75%-სათვის ჰომოლოგიურობის საფუძველზე მოდელირება არ მოხერხდა, რადგან თანამიმდევრობათა შედარების საფუძველზე არ მოიძებნა შესაბამისი ჰომოლოგიური ნიმუში. ამ 75%-ისათვის საჭიროა სხვა მიდგომების მოშველიება, მაგრამ სანამ მათზე ვისაუბრებდეთ, განვიხილოთ თანამიმდევრობათა მსგავსებასა და სტრუქტურათა მსგავსებას შორის ურთიერთმიმართება და ჩამოვყალიბოთ ჰომოლოგიურობის უფრო რეალური და ზუსტი განსაზღვრება. სწორი იქნება თუ ჰომოლოგიურობას ვიზმართ მხოლოდ ევოლუციური ნათესაობის, ანუ საერთო წინაპრის არსებობის აღსანიშნავად. ზემოთ ვისაუბრეთ თანამიმდევრობათა შორის მსგავსებების განსაზღვრელ სივრცეში ისეთი ზონის არსებობაზე, სადაც თამამად შეიძლება თანამიმდევრობათა შორის ნათესაობაზე, ანუ ჰომოლოგიურობაზე საუბარი. ამ ზონის გარეთ კი, ჰომოლოგიურობაზე დაბეჯითებით ვერაფერს ვიტყვით. ანუ, თუ ურთიერთსწორებისას იდენტური ამინმჟავები 30%-ზე ნაკლებია, ცალსახად ვერ ვიტყვით აქვს თუ არა ადგილი თანამიმდევრობათა შორის ჰომოლოგიურობას. ამ დროს ხშირია შემთხვევები, როცა შესაბამის ცილებს აღმოაჩნდებათ მსგავსი სტრუქტურები და ფუნქციები, ანუ არსებობს დიდი ალბათობა საერთო წინაპრის არსებობისა და ე.ი ჰომოლოგიურობისა. თანამიმდევრობათაშორისი მსგავსების ზონა, რომელიც ურთიერთსწორებისას იდენტურ ამინმჟავათა 20-30 %-ით განისაზღვრება, წარმოადგენს ე.წ „ბუნდოვან ზონას“. ამ ზონაში მხოლოდ თანამიმდევრობათა შედარების, მათ შორის მსგავსების შეფასების საფუძველზე ჰომოლოგიურობის შესახებ დასკვნის გაკეთება შეუძლებელია. გამომდინარე იქიდან, რომ ევოლუციის პროცესში სტრუქტურა ცილის ყველაზე კონსერვატიული მახასიათებელია, გაჩნდა მოსაზრება ცილის სტრუქტურის წინასწარმეტყველების პროცესში გამოყენებულიყო სტრუქტურული ინფორმაცია და შეიქმნა მიდგომები (ფიზიკურ-ბიოლოგიურ საფუძველზე დაყრდნობით), სადაც წინასწარმეტყველება ხდება უცნობ თანამიმდევრობასა და ცნობილ სტრუქტურას შორის შესაბამისობის შეფასების საფუძველზე. ანუ ფასდება ალბათობა საკვლევი (უცნობი სტრუქტურის მქონე) თანამიმდევრობის ჰიპოტეტური სტრუქტურის მსგავსებისა მოცემულ ცნობილ სტრუქტურასთან. ასეთი მიდგომით შესაძლებელი ხდება გენთა იმ წილის შემცირება, რომელთათვისაც შეუძლებელია ჰომოლოგიურობის საფუძველზე მოდელირება. თუმცა, სტრუქტურული მსგავსების არსებობაც არ იძლევა ჰომოლოგიურობის გარანტიას. სტრუქტურული მსგავსება შესაძლებელია კონვერგენტული ევოლუციის შედეგი იყოს, ანუ არ იყოს განპირო-

ბებული საერთო წინაპრით. შესაბამისად არსებობს მიდგომები (ძირითადად ბიოლოგიურ ასპექტებზე დაფუძნებული), სადაც საკვლევი თანმიმდევრობისათვის ჰომოლოგიურთა გამოვლენის პროცესი ეფუძნება დივერგენტული ევოლუციის სურათის ანალიზს. როგორც ფიზიკურ, ისე ბიოლოგიურ ასპექტებზე დაფუძნებულ ჰომოლოგიურ თანმიმდევრობათა გამოსავლენ მეთოდებს ეწოდებათ „დახვევის ტიპის გამოცნობი მეთოდები“.

დახვევის ტიპის გამოცნობა

მიუხედავად იმისა, რომ ცნობილია დახვევის წარმართველ ძალთა ხასიათი, ჯერ კიდევ ვერ ხერხდება მხოლოდ ძირითად პრინციპებზე დაყრდნობით, ანუ ფუნდამენტალური ფიზიკური განტოლებების ამოხსნების საფუძველზე ცილის სტრუქტურის წინასწრამეტყველება. ამავე დროს, გაცილებით ადვილი გადასაწყვეტია არაპირდაპირი ამოცანა, რომლის საშუალებითაც შეიძლება გაირკვეს თუ ცნობილ დახვევის ტიპებს შორის რომელია ყველაზე უფრო კარგ შესაბამისობაში საკვლევი ცილის თანამიმდევრობასთან, რომლის დახვევის ტიპი ჯერ კიდევ დასადგენია. დახვევის შეზღუდვებული ამოცანის (ხდება არა სტრუქტურის შერჩევა თანამიმდევრობისათვის, როგორც ეს ab initio ამოცანების შემთხვევაშია, არამედ თანამიმდევრობის შერჩევა არსებული სტრუქტურისათვის) რამდენიმე ვარიანტი არსებობს, მათ შორის გამოყოფთ: მორგების ამოცანას (თანამიმდევრობასა და სტრუქტურას შორის შესაბამისობის შეფასების მიზნით ხდება თანამიმდევრობის მორგება (threading) სტრუქტურაზე) და დახვევის ტიპის გამოცნობის ზოგად ამოცანას.

საერთოდ დახვევის ტიპის გამოცნობის ამოცანის გადაწყვეტა წარმოადგენს აუცილებელ წინაპირობას დახვევის პროცესის განმსაზღვრელ კანონზომიერებათა დადგენისათვის. ხოლო, რაც შეეხება ცილის სივრცული სტრუქტურის მოდელირებას, აქ დახვევის ტიპის გამოცნობის ამოცანის ნაწილობრივი გადაწყვეტაც კი, მნიშვნელოვნად ზრდის მოდელირების შესაძლებლობებს. დახვევის ტიპის გამოცნობის მეთოდებზე მოთხოვნილება დაკავშირებულია იმ ფაქტთან, რომ თითქოსდა არამონათესავე (თანამიმდევრობათა შედარებაზე დაფუძნებით გაკეთებული დასკვნა) ცილებს შეიძლება დახვევის ერთნაირი ტიპები გააჩნდეთ. ითვლებოდა, რომ ახლად აღმოჩენილ ცილათა თითქმის ნახევარს არ გააჩნდა ნათესავი ცნობილ სტრუქტურათა შორის (ასევე თანამიმდევრობათა ურთიერთსწორებაზე დაფუძნებით გაკეთებული დასკვნა), მაგრამ აღმოჩნდა, რომ მათ უმრავლესობის სტრუქტურა შეესაბამება კარგად ცნობილ დახვევის ტიპს. დღეისათვის შემუშავებულ დახვევის ტიპის გამოცნობის მეთოდებს შორის შეიძლება განვასხვავოთ თანამიმდევრობათა ერთდროულ მრავალწევრა ურთიერთსწორებაზე დაფუძნებული მიდგომები (სუფთად ბიოლოგიური მიდგომები), რომლის მაგალითად განვიხილავთ PSI-BLAST-ს, და ენერგიაზე (პოტენ-

ციალზე) დაფუძნებული მიდგომები (ფიზიკო-ბიოლოგიური მიდგომები), რომლის მაგალითადაც წარმოვადგენთ წინასწარმეტყველებას მორგებით (Threading).

PSI-BLAST

PSI-BLAST, როგორც სახელიდან ჩანს, წარმოადგენს BLAST ალგორითმზე დაფუძნებულ მიდგომას (იხ. თანამიმდევრობათა შესახებ მონაცემთა ბაზების სწრაფი ანალიზის მეთოდები). მიდგომა იტერაციულია და ურთიერთსწორების პროცესში ეფუძნება პოზიცია-სპეციფიკურ ქულათა დარიცხვის PSSM მატრიცებს (Position-Specific Score Matrices), სხვაგვარად მათ პროფილებსაც უწოდებენ.

ამრიგად, PSI-BLAST საწყისი იტერაციის გარდა, ანუ მეორე და შემდგომ ეტაპებზე ეფუძნება პოზიცია-სპეციფიკურ მატრიცას (PSSM). ურთიერთსწორება წარმოებს ჩვეულებრივად, ისე, როგორც ამას ადგილი აქვს BLAST ალგორითმით ნებისმიერი ორი თანამიმდევრობის ურთიერთსწორებისას. განსხვავება მხოლოდ ურთიერთშესაბამისობაში მოყვანილ პოზიციებზე ანუ ამინმჟავათა აღმნიშვნელ ასოთა წყვილზე დასარიცხ ქულათა მნიშვნელობებშია. ჩვეულებრივი ურთიერთსწორებისას დასარიცხ ქულათა მნიშვნელობები მიიღება ამინმჟავათა ჩანაცვლებების ტრადიციული მატრიცებიდან (მაგალითად PAM, BLOSUM ან სხვა), PSI-BLAST-ის შემთხვევაში კი PSSM -დან. საკვლევი ცილისათვის, რომლის ამინმჟავური თანამიმდევრობის სიგრძეა L , ურთიერთსწორების პროცესში გამოყენებული 20×20 განზომილების ჩანაცვლებათა მატრიცის ნაცვლად გამოიყენება $L \times 20$ განზომილების PSSM. ისევე როგორც თანამიმდევრობათა წყვილის ურთიერთსწორების პროცესში, აქაც შესაძლებელია PSSM -სა და ბანკის თანამიმდევრობას შორის გლობალური ურთიერთსწორების პოვნა, სრულ PSSM-სა და ბანკის თანამიმდევრობის სეგმენტს შორის ურთიერთსწორების მოძიება, ან PSSM-სა და ბანკის თანამიმდევრობას შორის საუკეთესო ლოკალური ურთიერთსწორების მოძიება.

PSSM -ის საწარმოებლად საჭიროა გვექონდეს მრავალწევრა ერთდროული ურთიერთსწორება (multiple alignment). გამომდინარე იქიდან, რომ პირველ ეტაპზე, ანუ ნულოვანი იტერაციის დონეზე ჯერ კიდევ არაა ფორმირებული მრავალწევრა ურთიერთსწორება და შესაბამისად, არაა წარმოებული საწყისი PSSM, საკვლევი თანამიმდევრობით მონაცემთა ბაზის სკანირებისას გამოიყენება ჩანაცვლებათა რომელიმე ტრადიციული მატრიცა (PAM, BLOSUM ან სხვა). სკანირების შედეგად მიღებულ მონაცემთა, ანუ BLAST პროცედურით მიღებული შედეგების ანალიზის საფუძველზე ირჩევა მონაცემთა ბაზის იმ თანამიმდევრობათა სეგმენტები, რომელთათვისაც საკვლევ თანამიმდევრობასთან მსგავსების E კრიტერიუმი რაღაც ზღვრული მნიშვნელობის ტოლია ან მასზე ნაკლებია (ანუ ამ ზღვრული სიდიდით ან მეტით არის მსგავსი). მითითების გარეშე (ავტომატურად) ამ

ზღვრული მნიშვნელობის სიდიდედ პროგრამა იყენებს 0.01. ასეთნაირად შერჩეული სეგმენტებით საკვლევ თანამიმდევრობაზე, როგორც საფუძველზე (ნიმუშზე) იგება მრავალწევრა ერთდროული ურთიერთსწორება *M*. საჭიროა აღინიშნოს, რომ თუ ურთიერთსწორების ორ სტრიქონს, ანუ ორ თანამიმდევრობას შორის იდენტურობა აღემატება 98%-ს, ამ ორიდან რჩება ერთ-ერთი სტრიქონი. ასეთნაირად ფორმირებული მრავალწევრა ერთდროული ურთიერთსწორება წარმოადგენს სწორედ PSSM-ის წარმოების საფუძველს. ანუ ურთიერთსწორების *i*-ურ სვეტში *k*-ური (ოციდან რომელიმე) ამინმჟავას გამოჩენის სიხშირე განსაზღვრავს სწორედ მოცემულ *i*-ურ პოზიციამში კონკრეტული *k*-ური ამინმჟავის გამოჩენის ალბათობას, ანუ $L \times 20$ განზომილებიანი პოზიგია-სპეციფიკური მატრიცის კონკრეტულ (*i,j*) ელემენტს ($i=1,L$ და $j=1,20, L$ საკვლევი ცილის თანამიმდევრობის სიგრძეა). ასეთნაირად წარმოებული PSSM უკვე გამოიყენება მონაცემთა ბაზის ახალი სკანირების ჩასატარებლად, ანუ პირველი იტერაციული პროცედურის განსახორციელებლად, რომელიც მიგვიყვანს ახალ მრავალწევრა (თუ მოხდა ახალ წევრთა ჩართვა) ურთიერთსწორებამდე და შესაბამისად ახალ, უფრო დახვეწილ PSSM -ის ფორმირებამდე. ახალი PSSM -ით ხდება შემდგომი, მეორე იტერაცია ანუ მიმდინარეობს ანალოგიური პროცედურა და ასეთი განმეორებადი პროცესი გაგრძელდება მანამ, სანამ PSSM -ის შემდგომი დახვეწა მოცემული საკვლევი თანამიმდევრობისა და მოცემული მონაცემთა ბაზის პირობებში აღარ იქნება შესაძლებელი.

ასეთია მოკლედ PSI-BLAST ალგორითმის არსი. იგი იყენებს მოცემული ცილის ევოლუციური განვითარების პროცესში კონკრეტულ პოზიციებში დაგროვილ ცვლილებათა შესახებ ინფორმაციას და გარკვეულწილად მონაცემებს, რომლებიც ახასიათებს მოცემული ცილის თანამიმდევრობაში თითოეული პოზიციის მნიშვნელობას ცილის კონკრეტული სტრუქტურის და მოცემული ფუნქციის რეალიზებისათვის. როცა ადგილი აქვს შორეულ ჰომოლოგიურობას (თანმომდევრობათა მსგავსება არ ვლინდება, თუმცა სტრუქტურული სახეზეა) ასეთი მიდგომა გაცილებით წარმატებულია დახვეწის ტიპის ამოცნობის ამოცანის გადაწყვეტისას, ვიდრე ტრადიციული ჩანაცვლებათა მატრიცებითა და ტრადიციული FASTA და BLAST ალგორითმებით განხორციელებული პოცედურები.

მორგებით წინასწარმეტყველება (Threading)

როგორც უკვე აღვნიშნეთ, აღიარებული თერმოდინამიკული ჰიპოტეზის თანახმად, ცილის ნატივური სტრუქტურა შეესაბამება ცილა-გამსხნელისაგან შემდგარი სისტემის თავისუფალი ენერჯიის მინიმუმს. ამიტომ ბუნებრივია, რომ კონფორმაციულ სივრცეში ნატივური სტრუქტურის მოძიების ამოცანა დაყვანილ

იქნა კორექტული ენერგეტიკული ფუნქციის შერჩევის ამოცანასა და ამ ფუნქციის მინიმუმის მოძიებისათვის კომპიუტერული ფიზიკის ეფექტური მეთოდების შემუშავების პრობლემის გადაწყვეტაზე. მიუხედავად იმისა, რომ ამ მიმართულებით წარმატებები სახეზეა, მხოლოდ თანამიმდევრობაზე დაყრდნობით უნობი სტრუქტურის წინასწარმეტყველების ამოცანა ჯერ კიდევ შორსაა საბოლოო გადაწყვეტამდე. შეიძლება აღვნიშნოთ ორი ძირითადი პრობლემა, რომელიც აფერხებს *ab initio* წინასწარმეტყველებას: ესენია ადეკვატური პოტენციური ფუნქციის არ ქონა და უზარმაზარი კონფორმაციული სივრცე. დახვევის შეზღუდული ამოცანაც ენერგიაზე დაფუძნებული მეთოდია. მაგრამ დახვევის ტიპის წინასწარმეტყველების შემთხვევაში, იმის გამო, რომ მოხსნილია კონფორმაციული სივრცის უსასრულობის პრობლემა (სივრცე სასრულია), ამოცანის გადაწყვეტა პირდაპირ ამოცანასთან შედარებით გაადვილებულია.

დახვევის ტიპის გამოცნობის ენერგიაზე დაფუძნებული მიდგომა შესაძლებელია შედარდეს მინიმუზაციის პროცესს, თუ მას წარმოვადგენთ როგორც დისკრეტულ წერტილთა სიმრავლეში მინიმუმის მოძიების ამოცანად. ენერგიათა დისკრეტული მნიშვნელობები, რომელთა შორისაც ხდება მინიმალურის შერჩევა, მიიღება ცილათა ცნობილი სტრუქტურების შესახებ ინფორმაციის (ატომური კოორდინატების) გამოყენების საფუძველზე. დახვევის ტიპის გამოცნობის ფიზიკურ-ბიოლოგიურ კანონზომიერებებზე დაფუძნებულ მიდგომაში გამოიყენება საკვლევი ცილის (რომლის სტრუქტურაც უცნობია) თანამიმდევრობა და ცნობილ ცილათა სტრუქტურები (ატომური კოორდინატები). საკვლევ თანამიმდევრობის ათავსებენ ცნობილი სტრუქტურების ძალოვან ველებში და აფასებენ საკვლევი თანამიმდევრობის ცნობილ სტრუქტურებთან შესაბამისობის დამახასიათებელ ენერგიათა სიდიდეებს (ამ დროს გამოიყენება ცნობილი სტრუქტურის ატომთა მხოლოდ კოორდინატები, არა თანამიმდევრობა). გამოდინარე იქიდან, რომ ადგილი აქვს საკვლევი თანამიმდევრობის ცნობილ სტრუქტურასთან შესაბამისობის შეფასებას, ანუ სტრუქტურასთან მორგების შესაძლებლობის შეფასებას, მეთოდს უწოდებენ მორგების მეთოდს. იგი მიეკუთვნება პოტენციალებზე (ენერგიაზე) დაფუძნებულ მიდგომებს. რადგანაც, მორგების მეთოდის გამოყენების შემთხვევაში აუცილებელია ცილათა სტრუქტურების შესახებ დიდი მოცულობის მონაცემთა ბაზების სკანირება და მინიმალური ენერგიის შესაბამისი სტრუქტურის მოძიება, საჭიროა, რომ ალგორითმი იყოს ძალიან სწრაფი. დღეს ასეთი ალგორითმების დიდი არჩევანი არსებობს. იქმნება ახალი ალგორითმებიც. ძალზე მრავალფეროვანია გამოყენებულ პოტენციალეთა სიმრავლეც. თუმცა შესაძლებელია იმ საერთო მახასიათებელთა გამოყოფა, რომლებსაც ეს პოტენციალები ეყრდნობა. მახასიათებელთა მაგალითებად, რომლებიც ცილის სივრცული მოწყობის ზოგად კანონზომიერებებს აღწერენ, შეგვიძლია მოვიყვანოთ: ამინმჟავიდან ცილის გლობულის გომეტრიულ ცენტრამდე მანძილი, გვერდით ჯაჭვთა ორიენტაცია, ამინმჟავათშორისი კონტაქტები, გამხსნელისათვის მისაწვდომი ზედაპირის ფართობი, მეორეული სტრუქტურა და სხვა.

დახვევის ტიპის ამოცნობის ამოცანის გადაწყვეტისათვის მორგების მიდგომა სხვა მეთოდებზე უფრო წარმატებულია. წარმატებულობა განსაკუთრებით თვალში-საცემია მაშინ, როცა შორეულ ჰომოლოგიასთან გვაქვს საქმე, ანუ როცა თანამიმ-დევერობათა მსგავსება არ შეინიშნება, თუმცა სტრუქტურული სახეზეა.

19. მრავალწევრა ერთდროული ურთიერთსწორება

ცილის ამინმჟავური თანამიმდევრობის დადგენასთან ერთად დღის წესრიგში დგება ფუნქციის გარკვევის მეტად მნიშვნელოვანი ამოცანის გადაწყვეტის საკითხი. პირველი მგდელობა ამ ამოცანის გადაწყვეტის გზაზე, როგორც წესი, გულისხმობს საკვლევი ცილის თანამიმდევრობის შედარებას მონაცემთა ბაზებში დეპონირებულ ნუკლეოტიდურ თუ ამინმჟავურ თანამიმდევრობებთან. ამ შედარების შედეგად შესაძლებელია გამოვლინდეს ერთი ან მეტი ისეთი თანამიმდევრობა, რომელიც ავლენს მსგავსებას საკვლევ თანამიმდევრობასთან, ან მის ნაწილთან. იმ შემთხვევაში, როცა მსგავსებულად წარმოჩენილ თანამიმდევრობათა რიცხვი ერთზე მეტია, სასურველია განხორციელდეს ანალიზის შემდგომი ეტაპი, თანამიმდევრობათა მრავალწევრა ერთდროული ურთიერთსწორება. მრავალწევრა ურთიერთსწორება იძლევა იმ ბაზისურ საწყის მონაცემს, რომელზე დაფუძნებითაც შესაძლებელია სტრუქტურული ბიოლოგიის თვალსაზრისით ისეთი მნიშვნელოვანი წინასწარმეტყველებების განხორციელება, როგორებიცაა: ცილის მეორეული სტრუქტურის, ამინმჟავური ნაშთის გამხსნელისადმი მისაწვდომობის, ფუნქციის წინასწარმეტყველება და ცილის სპეციფიკურობის განმსაზღვრელ ამინმჟავათა იდენტიფიცირება. ანუ წარმატებული მრავალწევრა ურთიერთსწორების ანალიზის საფუძველზე შესაძლებელია მოვიპოვოთ ინფორმაცია იმ ამინმჟავათა შესახებ, რომლებიც მნიშვნელოვან როლს თამაშობენ ცილის ფუნქციონირების პროცესში, ან რომლებიც განაპირობებენ ცილის მეორეული და მესამეული სტრუქტურის სტაბილიზაციას.

მოცემულ პარაგრაფში შევეცდებით ზოგადი წარმოდგენა შევქმნათ მრავალწევრა ერთდროული ურთიერთსწორების (მეუ) განმარტებული მოდგომებზე. მოკლედ მიმოვიხილათ დღეისათვის ფართოდ გამოყენებად პროგრამათა პაკეტებს.

რას წარმოადგენს მრავალწევრა ერთდროული ურთიერთსწორება ?

ვიცით, რომ ცილის ამინმჟავურ თანამიმდევრობას წარმოადგენენ ასოთა გარკვეული რიგის სახით. გამომდინარე ბუნებრივ ამინმჟავათა რიცხვიდან, გამოიყენება 20 სხვადასხვა ასო. გავიხსენოთ, რომ თანამიმდევრობათა ოპტიმალური ურთიერთსწორება ნიშნავს ამ ასოთა რიგების ისეთ ურთიერთგანლაგებას, რომ

მივიღოთ იდენტურ ან შესაბამის ასოთა ერთმანეთის ქვეშ, ანუ ერთსა და იმავე სვეტში განლაგების მაქსიმალური რიცხვი (მაქსიმალური ქულათა ჯამი). ორი თანამიმდევრობის ოპტიმალური ურთიერთსწორების მისაღწევად დასაშვებია ასოთა გამოტოვება-ჩამატება ერთერთ, ან ორივე თანამიმდევრობაში. ორი ამინმჟავური ნაშთის ურთიერთსწორება გვაძლევს საშუალებას ვივარაუდოთ, რომ სხვადასხვა ცილის შესაბამის პოზიციებში განლაგებულ ამინმჟავებს მსგავსი როლი აკისრიათ. ასე მაგალითად, თუ ცნობილი ფერმენტის აქტიური ცენტრის მოწყობის შესახებ არსებობს სრული ინფორმაცია და ამ ცენტრის მონაწილე ამინმჟავური ნაშთები ურთიერთსწრდნენ სხვა ცილის თანამიმდევრობის ანალოგიურ ამინმჟავებთან, შეგვიძლია ვივარაუდოთ, რომ უკანასკნელ ცილას ახასიათებს ცნობილი ფერმენტის აქტივობის ანალოგიური კატალიზური აქტივობა. ასეთი ჰიპოტეზის სისწორე დამოკიდებულია თანამიმდევრობათა გლობალური მსგავსების თაობაზე შეფასებების სისწორეზე, რომელსაც თავის მხრივ განაპირობებს განხორციელებული ურთიერთსწორების დამაჯერებლობა. ორი თანამიმდევრობის ურთიერთსწორების უამრავი შესაძლო ვარიანტი არსებობს. ძირითადი ამოცანა მდგომარეობს ამ ვარიანტებს შორის ისეთის შერჩევაში, რომელიც სწორად ასახავს ორ ცილას შორის არსებულ ქიმიურ თუ ბიოლოგიურ მსგავსებებს. მრავალწევრა ურთიერთსწორება გულისხმობს ერთოდროულ ურთიერთსწორებას, განხორციელებულს ორზე მეტი (როგორც წესი, ჰომოლოგიური) თანამიმდევრობის მომცველი კომპლექტისათვის. იმ შემთხვევაშიც კი, როცა ჩვენ ინტერესს წარმოადგენს ასეთ კომპლექტში შემავალ მხოლოდ ორ თანამიმდევრობას შორის მსგავსების შეფასება, მიზანშეწონილია მოვახდინოთ კომპლექტში შემავალი ყველა თანამიმდევრობის მეუ. დამატებითი თანამიმდევრობების ჩართვა მრავალწევრა ურთიერთსწორების პროცესში, როგორც წესი, აუშკობებს ჩვენთვის საინტერესო თანამიმდევრობათა წყვილის ურთიერთსწორების ხარისხს.

იერარქიული მიდგომა

დღეისათვის ყველაზე სრულყოფილი მრავალწევრა ერთოდროული ურთიერთსწორების მეთოდები ეფუძნება იერარქიულ მიდგომას. მიდგომა გულისხმობს პროცედურას, რომლის საწყისი ეტაპი ითვალისწინებს ფილოგენეტიკური ხის (იხ. პარაგრაფი 13) აგებას. ურთიერთსწორების ფორმირების შემდგომ ეტაპზე ხდება სწორედ საწყის ეტაპზე აგებულ ფილოგენეტიკურ ხის გამოყენება. 19.1 სურათზე სქემატურად ასახულია მეუ-ს განმახორციელებელი პროცედურის ეტა-

პები. როგორც სურათიდან ჩანს, პროცედურის პირველი ეტაპი გულისხმობს მეუ-ში მონაწილე ყველა თანამიმდევრობათა დაწყვილებას და ურთიერთშედარებას ორი თანამიმდევრობის შედარებისათვის განკუთვნილი მეთოდის გამოყენებით. ამ შედარებათა შედეგია წყვილებში მსგავსებათა მახასიათებელი ქულათა ჯამები. ამ ჯამებზე დაფუძნებით ხდება სწორედ თანამიმდევრობათა კლასტერიზაცია და შესაბამისი ფილოგენეტიკური ხის აგება. ხის მეზობელ შტოებზე განლაგებულია უფრო მსგავსი თანამიმდევრობები, მაშინ, როცა ნაკლები მსგავსების შესაბამისი შტოები შედარებით დაშორებულია. მეუ-ს აგების პროცესი იწყება თანამიმდევრობათა იმ წყვილის ურთიერთსწორებაზე დაფუძნებით, რომელთა შორის მსგავსებაც ყველაზე დიდია. შემდეგ ეტაპზე ერთვება შემდგომი, უფრო ნაკლებად მსგავს წყვილთა ურთიერთსწორება, შემდეგზე კიდევ უფრო ნაკლებად მსგავსის და ა.შ. ურთიერთსწორების წყვილი შეიძლება შექმნას არა მხოლოდ ცალკეულმა თანამიმდევრობებმა, არამედ ადრეულ ეტაპზე ფორმირებულ ურთიერთსწორებებმაც. ურთიერთსწორებების შედარებისას ურთიერთსწორებაში არსებული გამოტოვებები ნარჩუნდება. შეიძლება გაჩნდნეს ახლებიც.

ასეთ მიდგომაზე დაფუძნებული მეუ-ს უამრავი სხვადასხვა ვარიანტი არსებობს. მიდგომის სპეციფიკიდან გამომდინარე ცხადია, რომ საწყის ეტაპზე განხორციელებულ ურთიერთსწორებაში დაშვებულმა შეცდომამ შეიძლება შემდგომი შეცდომების დაგროვება გამოიწვიოს. ამიტომ ზოგი მეუ მეთოდი საწყის ეტაპზე განხორციელებული ურთიერთსწორების კვლავ განმეორება-გადამოწმების შესაძლებლობას უშვებს. მეუ-ს სხვა მეთოდი სხვადასხვა ეტაპზე გამოიყენებს მსგავსებათა სხვადასხვა მატრიცას.

ზოგადად, ცილის თანამიმდევრობებში გაჩენილ გამოტოვებათა განაწილება არაა შემთხვევითი. გამომდინარე იქიდან, რომ არსებობს სტაბილური ცილის შენარჩუნების მოთხოვნა, შეიძლება ვივარაუდოთ მნიშვნელოვანი ევოლუციური ზეწოლის არსებობა იმ გამოტოვებების საწინააღმდეგოდ, რომლებიც შეიძლება გაჩნდეს მეორეული სტრუქტურის ელემენტებში. სწორედ ამის შედეგია ალბათ ის, რომ ჩამატება - გამოტოვებებს ძირითადად ვხვდებით ცილის ყულფების ნაწილში. ასეთი ტენდენციები შეიძლება გამოყენებულ იქნეს მეუ-ს დასახვეწად. კერძოდ, თუ მეუ-ში მონაწილე ერთი ან რამდენიმე ცილის მეორეული სტრუქტურის ელემენტების შესახებ ინფორმაცია ცნობილია, გამართლებულია მეორეულ სტრუქტურაში გაჩენილი გამოტოვებები ყულფების ნაწილში არსებულ გამოტოვებებზე მეტად დაჯარბდეს.

CLUSTAL W და სხვა იერარქიული ურთიერთსწორების განმახორციელებელი პროგრამათა პაკეტები

CLUSTAL W (Thompson et al., 1994) წარმოადგენს მრავალწევრა ერთობლივი ურთიერთსწორების იერარქიული ვარიანტისა და ადვილად მოსახმარი ინტერფეისის ერთობას. იგი მუშაობს უმრავლესობა კომპიუტერული სისტემების პირობებში. CLUSTAL W უზრუნველგვყოფს სხვადასხვა ჩანაცვლებათა მატრიცების ფართო არჩევანით და ურთიერთსწორების დახვეწის მოთხოვნიდან გამომდინარე, იძლევა ხელახალი ურთიერთსწორების განხორციელების საშუალებას. იგი ამასთანავე ითვალისწინებს მეორეულ სტრუქტურათა ელემენტების განაწილების სურათს და გამოტოვებათა დაჯარიმებას ახდენს ამ სურათთან შესაბამისად.

CLUSTAL W წარმოადგენს ყველაზე ფართოდ გავრცელებულ და მრავალი თვალსაზრისით ყველაზე ადეკვატურ მეუ-ის განმახორციელებელ პროგრამათა პაკეტს. თუმცა არსებობენ სხვა პაკეტებიც, რომლებიც გამოირჩევიან გარკვეულწილად განსხვავებული შესაძლებლობებით. მაგალითის სახით შეგვიძლია მოვიყვანოთ AMPS (Barton, 1990), რომელიც თანამიმდევრობათა წყვილური შედარების პროცესში ეფუძნება Z-ქულათა ჯამებს და ამ უკანასკნელთა გამოთვლისას იყენებს თანამიმდევრობათა შემთხვევითი გადაჯგუფების პროცედურას. გარდა ამისა, ურთიერთსწორების გენერირება AMPS-ს შეუძლია ფილოგენეტიკური ხის აგების გარეშე. ეს უკანასკნელი თვისება განსაკუთრებით მომგებიანია დიდი რაოდენობა თანამიმდევრობებათა ურთიერთსწორების შემთხვევაში, რადგანაც თავიდან გვაცილებს მრავალ წყვილურ ურთიერთსწორებათა განხორციელების აუცილებლობას და ამცირებს მეუ-ზე დასახარჯ სამანქანო დროს. ამასთან AMPS უზრუნველყოფილია ფილოგენეტიკური ხის ვიზუალიზაციისათვის საჭირო პროგრამათა პაკეტით, რომელიც გვეხმარება ურთიერთსწორებაში ჩასართველ თანამიმდევრობათა შერჩევაში. პროგრამათა პაკეტის ნაკლს წარმოადგეს რთული „მენიუთი“ აღჭურვილი ინტერფეისი.

PSI-BLAST-ით განხორციელებული მრავალწევრა ერთდროული ურთიერთსწორება

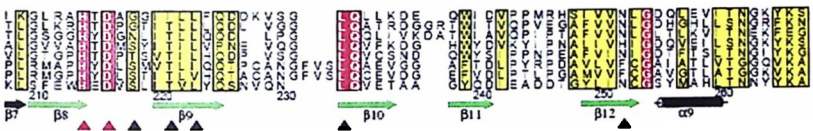
საკვლევი თანამიმდევრობის მსგავსი თანამიმდევრობების წარმოჩენის მიზნით თანამიმდევრობათა შესახებ მონაცემთა ბაზების ანალიზი შესაძლებელია განხორ-

ციელდეს უფრო სრულყოფილად, თუ დავეყრდნობით მრავალწევრა ერთდროულ ურთიერთსწორებას და არ შემოვიფარგლებით მხოლოდ ორი თანამიმდევრობის შედარების შედეგებით. PSI-BLAST (Altschul et al., 1997) პროგრამათა პაკეტის დახმარებით შესაძლებელი გახდა განხორციელებულიყო მრავალწევრა ერთდროულ ურთიერთსწორებისას მიღებულ პროფილებზე დაფუძნებული ბაზის ანალიზი. ამრიგად PSI-BLAST-ით მონაცემთა ბაზის ანალიზის პროცესში გენერირდება მრავალწევრა ურთიერთსწორება. ოღონდ უნდა აღინიშნოს, რომ ამ უკანასკნელ ურთიერთსწორებას ცოტა აქვს საერთო CLUSTAL W, AMPS, ან სხვა ტრადიციული მეთოდებით განხორციელებულ მრავალწევრა ურთიერთსწორებასთან. ტრადიციულ ურთიერთსწორებისას საბოლოო ურთიერთსწორების სიგრძე, როგორც წესი, აღემატება თითოეული ცალკეული თანამიმდევრობის სიგრძეს, რადგანაც ურთიერთსწორება გამოტოვებულია ოპტიმიზირებული. ამის საპირისპიროდ PSI-BLAST-ით განხორციელებული მრავალწევრა ურთიერთსწორება ყოველთვის საკვლევი თანამიმდევრობის სიგრძის ტოლია. თუ საკვლევი თანამიმდევრობის (ან მისი პროფილის) მონაცემთა ბაზის თანამიმდევრობასთან ურთიერთსწორებისას საჭირო ხდება გარკვეული ჩამატების განხორციელება, მაშინ ჩამატების გამომწვევი ფრაგმენტი უკუიგდება მონაცემთა ბაზის თანამიმდევრობიდან. საბოლოო ურთიერთსწორება ამგვარად გამოაშკარავებს მხოლოდ იმ ამინმჟავებს, რომლებიც საკვლევის მოცემული პოზიციას შესაბამებიან. აქედან გამომდინარე PSI-BLAST-ით განხორციელებული მრავალწევრა ურთიერთსწორება და მასთან ასოცირებული სიხშირეთა ცხრილები და პროფილები წარმოადგენს მეტად ეფექტურ საწყის მონაცემებს იმ მეთოდთათვის, რომელთა მიზანია ცილის ფუნქციისა და მეორეული სტრუქტურის ელემენტების წინასწარმეტყველება (Jones, 1999; Cuff and Barton, 2000).

მრავალწევრა ერთდროული ურთიერთსწორების ინტერპრეტირების საშუალებები (პროგრამათა პაკეტები).

მრავალწევრა ერთდროული ურთიერთსწორება შეიძლება ახდენდეს 500 ან მეტი ამინმჟავისაგან შემდგარ ასეულობით თანამიმდევრობის ერთდროულ შედარებას. მონაცემთა ასეთი მოცულობის პირობებში საკვანძო კანონზომიერებათა გამოვლენის მიზნით ანალიზის ჩატარება გამძლეებულია და ამიტომ, მეტად მნიშვნელოვანია ურთიერთსწორების ისეთი სახით წარმოდგენა, რომ შესაძლებელი იყოს მისი ინტერპრეტირება. ასეთი წარმოდგენის ფორმირების ერთადერთ საშუალებას ადრე ურთიერთსწორების მრავალფურცლიანი ამონაბეჭდის გაკეთება, ამ ფურცლების შესაბამისი ფორმით შეწებება, შეწებების შედეგად მიღებული პლაკატის ანალიზი და ანალიზის შედეგად წარმოჩენილ კანონზომი-

ერებათა სხვადასხვა ფერებით გამოყოფა წარმოადგენდა. ასეთი მიდგომა შეიძლება დღესაც სასარგებლო იყოს, მაგრამ ცხადია, რომ იგი მოუხერხებელია. ურთიერთსწორების ვიზუალიზაციისათვის მოხერხებული საშუალებების შექმნა საკმაოდ მნიშვნელოვან ამოცანას წარმოადგენს. ასეთი ამოცანის გადაწყვეტა ემსახურება როგორც ანალიზის, ისე პუბლიკაციის ხარისხის გაუმჯობესებას. ვიზუალიზაციის საშუალებებია ერთ-ერთ მაგალითად გამოდგება პროგრამათა პაკეტი ALSCRIPT (Barton, 1993). ALSCRIPT იყენებს მრავალწევრა ერთდროული ურთიერთსწორების მონაცემებს, როგორც შემავალ მონაცემებს და იძლევა PostScript ფორმატში ჩაწერილ ფაილს, რომლის ანალიზის მიზნით მოხმარება გაადვილებულია. 19.2 სურათზე მოცემულია ALSCRIPT-ის გამოსავლის ფრაგმენტი. როგორც სურათიდან ჩანს, თუ ურთიერთსწორების სვეტი იდენტური ამინმჟავებისაგან შედგება, იგი გამოიყოფა წითელი ფერით, ხოლო მსგავსი ფიზიკურ-ქიმიური თვისებების მქონე ამინმჟავებისაგან შემდგარი სვეტი კი ყვითლით. დანომვრა შეესაბამება ქვედა თანამიმდევრობას. მწვანე ისრები მიუთითებს ბეტა ფურცლოვან ფრაგმენტზე, შავი ცილინდრები კი -ალფა სპირალებზე.

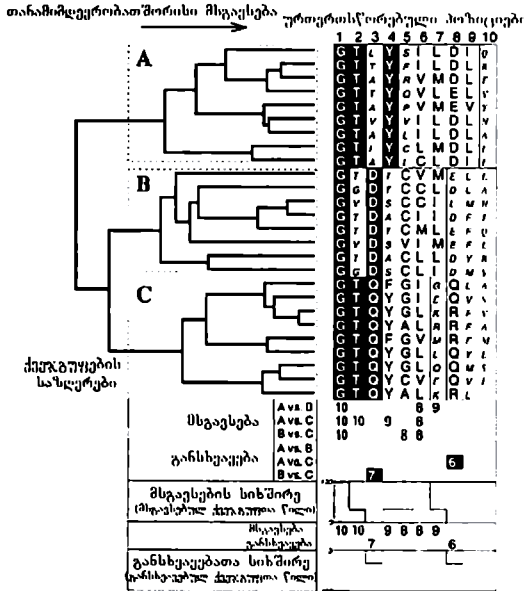


სურათი 19.2. ALSCRIPT (Barton, 1993) პროგრამათა პაკეტის გამოსავალი მონაცემები.

სუბურთიერთსწორება - AMAS პროგრამათა პაკეტი

ALSCRIPT საშუალებას იძლევა ცილოვანი ოჯახის შიგნით დაფიქსირდეს კონსერვატული პოზიციები და გარკვეული შეფერილობით გამოიყოს ისინი. შესაბამისად, ALSCRIPT უფრო სადემონსტრაციო საშუალებაა, ვიდრე ანალიზის. ამის საპირისპიროდ არსებობს პროგრამათა პაკეტი AMAS (Analysis of Multiply Aligned Sequences; Livingstone and Barton, 1993), რომელიც სწავლობს მრავალწევრა ურთი-

ერთსწორებაში მონაწილე თანამიმდევრობათა შორის ურთიერთდამოკიდებულებებს და იძლევა ფუნქციისათვის პოტენციურად მნიშვნელოვან ამინმეავეებს. ეს ხორციელდება იმის ხარჯზე, რომ AMAS ავლენს და აანალიზებს მსგავსებებსა და განსხვავებებს მრავალწევრა ურთიერთსწორებაში მონაწილე თანამიმდევრობათა ქვეჯგუფებს შორის. ასე მაგალითად, ვთქვათ, გასაანალიზებელია თანამიმდევრობათა ოჯახი, რომელშიც ადგილი აქვს გარკვეულ ვარიაციებს, მაგრამ გვხვდება პოზიციები, რომლებიც მთელი ოჯახის მიმართ ავლენენ კონსერვატულობას.



სურათი 19.3. AMAS პროგრამათა პაკეტის გამოსავალი მონაცემების მაგალითი. სხვადასხვა ქვეჯგუფების შიგნით ნაპოვნი კონსერვატულობის საილუსტრაციოდ გამოიყენება სხვადასხვა შეფერილობა. კონსერვატულობის სიდიდე (Livingstone and Barton, 1993; Zvelebil et al., 1987) იცვლება 0-დან 10-მდე და ახასიათებს თითოეული ურთიერთსწორების სვეტის შიგნით ფიზიკურქიმიურ თვისებათა მსგავსების ხარისხს. ურთიერთსწორების ქვემოთ განთავსებულ სტრიქონში სახელწოდებით „მსგავსება“, მოცემულია კონსერვატულობის სიდიდე

ქვეჯგუფთა დაწველებისა და კომბინირების შედეგად წარმოქმნილ ჯგუფთა გაერთიანებისათვის, თუ კონსერვატულობის სიდიდე არაა ნაკლები რაღაც ზღვრულ მნიშვნელობაზე. ასე მაგალითად, მე-7 პოზიციაში A და B ქვეჯგუფები ერთიანდებიან კონსერვატულობის სიდიდით 9. შემდგომი სტრიქონი სახელწოდებით „განსხვავება“ (“Different Pairs”) გვიჩვენებს წვეილებს იმ ქვეჯგუფებისა, რომელთა კომბინაციამ გამოიწვია კონსერვატულობის სიდიდის ზღვრულ მნიშვნელობაზე დაბლა დაწევა. ასე მაგალითად, მე-3 პოზიციაში B და C ქვეჯგუფების შიგნით ადგილი აქვს სრულ იდენტურობას. მაგრამ როცა ხდება მათი კომბინირება კონსერვატულობა ეცემა 7-მდე, ანუ ზღვრულ მნიშვნელობაზე, 8-ზე დაბლა. მსგავსებებისა და განსხვავებების შესახებ ჯამური მონაცემები მოცემულია სიხშირეთა პისტოგრამის სახით.

შეიძლება ვივარაუდოთ, რომ ეს უკანასკნელი მნიშვნელოვანია როგორც სტრუქტურის სტაბილურობისათვის, ისე საერთო ფუნქციის რეალიზებისათვის. იმ პოზიციებისათვის კი, რომლებიც ავლენენ კონსერვატულობას მხოლოდ თანამიმდევრობათა გარკვეული ქვეჯგუფის შიგნით, მაგრამ იცვლებიან ოჯახის დანარჩენი წევრების შემთხვევაში, შეიძლება ვივარაუდოთ, რომ ისინი მნიშვნელოვანია მოცემულ ქვეჯგუფში შემავალი ცილების რაღაც სპეციფიკური ფუნქციის განხორციელებისათვის. AMAS ახდენს უკანასკნელთა იდენტიფიცირებას. 19.3 სურათზე მოცემულია AMAS-ით განხორციელებული ანალიზის საფუძველზე მოპოვებული ტიპური ინფორმაციული სურათი.

წასაკითხად რეკომენდირებული ლიტერატურა

- Kabsch W, Sander C (1983). "Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features". *Biopolymers* 22 (12): 2577-637.
- Garnier, J., Osguthorpe, D. J. & Robson, B. (1978). *Journal of Molecular Biology*, 120, 97-120.
- D. G. Kneller, F. E. Cohen & R. Langridge (1990) Improvements in Protein Secondary Structure Prediction by an Enhanced Neural Network, *Journal of Molecular Biology*, 214, 171-182.
- Rumelhart, D., Hinton, G and Williams, R. (1986) in *Parallel Distributed Processing*, eds. Rumelhart, D.E. & McClelland, J.E. (MIT Press, Cambridge, MA), Vol. 1, pp. 318-362
- L. Howard Holley and Martin Karplus (1989) „ Protein secondary structure prediction with a neural network“ Proc. Nati. Acad. Sci. USA Vol. 86, pp. 152-156
- Holm L, Ouzounis C, Sander C, Tuparev G, Vriend G. A (1992) „database of protein structure families with common folding motifs.“ *Protein Sci.*, 1:1691-1698.
- Hobohm U, Sander C. (1995) „A sequence property approach to searching protein databases.“ *J. Mol. Biol.*, 251:390-399.
- Bowie JU, Luethy R, Eisenberg D. (1991) „A method to identify protein sequences that fold into a known three-dimensional structure.“ *Science*. 253:164-169.

- Fischer D, Elofsson A, Rice DW, LeGrand S, Eisenberg D. (1996) „Assessing the Performance of Fold Recognition Methods By Means of a Comprehensive Benchmark.“ Pac. Symp. Biocomput. 300-318.
- Jones TD, Taylor WR, Thornton JM. (1992) „A new approach to protein fold recognition.“ *Nature* 358:86-89.
- Zhang Y, Arakaki AK, Skolnick J. (2005) „TASSER: an automated method for the prediction of protein tertiary structures in CASP.“ *Proteins* 61 Suppl 7:91-8.
- Eddy, S. (1996) 'Hidden Markov models', *Current Opinion in Structural Biology* 6, 361-5. [Readable introduction to an important mathematical technique providing powerful tools for detection of distantly-related sequences, and protein fold recognition.]
- Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994). Clustal W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positionspecific gap penalties and weight matrix choice. *Nucl. Acids Res.* 22, 4673-4680.
- Barton, G. J. (1990). Protein multiple sequence alignment and flexible pattern matching. *Methods Enz.* 183, 403-428.
- Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997). Gapped blast and psi-blast: a new generation of protein database search programs. *Nucl. Acids Res.* 25, 3389-3402.
- Jones, D. T. (1999). Protein secondary structure prediction based on position-specific scoring matrices. *J. Mol. Biol.* 17, 195-202.
- Cuff, J. A., and Barton, G. J. (1999). Evaluation and improvement of multiple sequence methods for protein secondary structure prediction. *Proteins* 34, 508-519.
- Barton, G. J. (1993). ALSRIPT: A tool to format multiple sequence alignments. *Prot. Eng.* 6, 37-40.
- Livingstone, C. D., and Barton, G. J. (1993). Protein sequence alignments: A strategy for the hierarchical analysis of residue conservation. *Comp. App. Biosci.* 9, 745-756.
- Zvelebil, M. J. J. M., Barton, G. J., Taylor, W. R., and Sternberg, M. J. E. (1987). Prediction of protein secondary structure and active sites using the alignment of homologous sequences. *J. Mol. Biol.* 195, 957-961.

მოლეკულათმორისი ურთიერთქმედებების წინასწარმეტყველება

20. ცილა - ცილა ურთიერთქმედების წინასწარმეტყველება

უჯრედის მოლეკულურ ბიოლოგიაში მნიშვნელოვანი ცოდნის დაგროვების კვალობაზე ნათელი ხდება, რომ საჭიროა გენების და შესაბამისი ცილოვანი მოლეკულების არა მხოლოდ დამოუკიდებელ ერთეულებად განხილვა, არამედ მათი გარკვეული ქსელებისა თუ ბიოლოგიური ციკლების შემადგენელ ნაწილებად მოაზრება. უჯრედში ბიომაკრომოლეკულათა ფუნქციური გამოვლინებების თავისებურებათა სრულფასოვანი ახსნა შეუძლებელია მხოლოდ საკუთარ თვისებათა კრილში განხილვის პირობებში, აუცილებელია უჯრედული კონტექსტების გათვალისწინება.

გენეტიკური ინფორმაცია შესაძლებელია განხილულ იქნეს როგორც ცოცხალი უჯრედის მოწყობის და გამართულად ფუნქციონირებისათვის აუცილებელი საშუალებების და ამ საშუალებათაგან შემდგარი ბიოლოგიური სისტემების შესახებ ინფორმაცია. დღეს, თანამედროვე ბიოლოგიის უმნიშვნელოვანესი ამოცანის, გენეტიკური ინფორმაციის სრულად წარმოჩენის ამოცანის გადაწყვეტის პროცესი თვისობრივად ახალ, სისტემურ მიდგომებს საჭიროებს. შესაბამისად, გაჩნდა აუცილებლობა, ცოცხალი უჯრედის ფუნქციონირებისათვის მნიშვნელოვან საშუალებათა ფუნქცია მოვიაზროთ გარკვეულ ბიოლოგიურ სისტემათა (უჯრედულ სტრუქტურათა, მოლეკულურ ქსელთა თუ მეტაბოლურ ციკლთა) ფუნქციონირების კრილში. ცილა - ცილა ურთიერთქმედებათა შესწავლა წარმოადგენს სწორედ უპირველეს და მნიშვნელოვან ეტაპს ცოცხალი უჯრედის ცხოველყოფილობის წარმმართველ უჯრედული სტრუქტურების, მეტაბოლური ციკლებისა და შესაბამისი ბიომაკრომოლეკულური ქსელების მოწყობისა და ფუნქციონირების მოლეკულურ მექანიზმებში გარკვევის გზაზე.

როგორც აღინიშნა, ცილათმორისი ურთიერთქმედებები ბევრი ბიოლოგიური ფუნქციის რეალიზების საფუძველს ქმნის. ასე მაგალითად, უჯრედგარე სივრ-

ციდან უჯრედშიგნით სიგნალის გადაცემის პროცესი უფუძნება სასიგნალო მოლეკულებს შორის ცილა-ცილა ურთიერთქმედებებს. თავის მხრივ, სიგნალის გადაცემის (ტრანსდუქციის) პროცესი ფუნდამენტურ როლს თამაშობს უამრავ, როგორც ნორმალურ, ისე ანომალურ (პათოლოგიურ) ბიოლოგიურ პროცესში.

ცილები შესაძლებელია ურთიერთქმედებდნენ დიდი ხნის განმავლობაში და ქმნიდნენ სტაბილურ ცილოვან კომპლექსს. ცილები შეიძლება წარმოადგენდნენ სხვა ცილების გადამტანებს (მაგალითად ციტოპლაზმიდან ბირთვში ან პირიქით, როგორც ეს ხდება ბირთვულ ფორებში მოქმედ იმპორტინების შემთხვევაში). და ბოლოს ცილები შეიძლება მყისიერად, მცირე დროის განმავლობაში ურთიერთ-ქმედებდნენ სხვა ცილებთან, ამ უკანასკნელთა მოდიფიცირების მიზნით (მაგალითად პროტეინკინაზა, რომელსაც გადააქვს ფოსფატის ჯგუფი სამიზნე ცილაზე). მოდიფიკაციებმა შესაძლებელია გავლენა იქონიოს ცილა-ცილა ურთიერთ-ქმედებაზე. ასე მაგალითად, SH2 დომენის მომცველი ცილების სხვა ცილებთან დაკავშირების უნარი ვლინდება მხოლოდ მაშინ, როცა ფოსფორილირებულია ტიროზინის ამინმჟავური ნაშთი. ცილა-ცილა ურთიერთქმედებებს ცენტრალური როლი უნიჭებათ ცოცხალ უჯრედში მიმდინარე პრაქტიკულად ყველა ბიოლოგიურ პროცესში და შესაბამისად ამ ურთიერთქმედებათა შესახებ ინფორმაცია უდავოდ უზრუნველგვყოფს პათოლოგიათა მიზეზებში გარკვევისათვის აუცილებელი ცოდნით და ქმნის წინაპირობას ახალი თერაპიული საშუალებების შექმნისათვის.

საზოგადოდ, ცილა-ცილა ურთიერთქმედებათა შესწავლასთან დაკავშირებულ ამოცანათა სიმრავლეში შეიძლება გამოვყოთ ორი განსხვავებული ბუნების ამოცანა: ურთიერთქმედებაში უშუალოდ ჩართულ ამინმჟავათა ან ცილის ფრაგმენტთა წარმოჩენის (წინასწარმეტყველების) ამოცანა და პოტენციურ პარტნიორთა (მოცემულ ცილასთან ურთიერთქმედებაში შესვლის შესაძლებლობის მქონეთა) წარმოჩენის (წინასწარმეტყველების) ამოცანა. ამ განსხვავებულ ამოცანათა გადაწყვეტისათვის გამოსადეგია როგორც ბიოფიზიკური და ბიოქიმიური მეთოდები, რომელთაც საშუალებითაც შესაძლებელია უშუალოდ დაფიქსირდეს კომპლექსის წარმოშობა (მაგალითებად გამოდგება ქრომატოგრაფული მეთოდი, კო-იმუნოპრეციპიტაციის მეთოდი, ქიმიური გზით (კოვალენტური ბმით) კომპლექსის დაფიქსირების მეთოდი და სხვა), ასევე არაპირდაპირი გენეტიკური მეთოდები (მაგალითად გენთა სუპრესიის, მუტაგენეზის და სხვა). დღეს, თანამედროვე გენომური და პოსტგენომური ტექნოლოგიების პირობებში მნიშვნელოვნად შეიცვალა შესაძლებლობები და ცილა-ცილა ურთიერთქმედებათა შესახებ გაცილებით მოკლე დროში გაცილებით მეტი მონაცემის მიღება ხდება. პროგრესის საფუძველი გახდა ისეთი ექსპერიმენტული მიდგომების ავტომატიზაცია, როგორცაა მაგალითად, საფუარში ორნაწილა ჰიბრიდულობაზე დაფუძნებული მიდგომა და მასსპექტრომეტრული მოდგომა. ამის პარალელურად შემუშავებულ იქნა მთელი რიგი ბიოინფორმატიკული მეთოდებისა, რომლებიც ეფუძნებიან თანამიმდევრობათა და გენომთა

შესახებ მონაცემთა ბაზებში თავმოყრილ დიდძალ ინფორმაციას. ქვევით შევყვებით წარმოდგენა შევქმნათ ამ ბიოინფორმატიკულ მიდგომებზე. მაგრამ უშუალოდ მათზე გადასვლამდე შევხებით შესაბამისი მიდგომების წარმოშობის ბიოლოგიურ საფუძვლებს .

სტრუქტურითა და ფუნქციით განპირობებული თავისებურებანი ასახულნი ევოლუციურ სურათში.

ბიომპრომოლეკულების, მათ შორის ცილათა თანამიმდევრობების ევლუციურ ცვლილებათა სურათი, როგორც წესი აღიწერება მრავალწევრა ერთდროული ურთიერთსწორების (მეუ) საშუალებით. ურთიერთსწორებულ პოზიციებში (სვეტებში) მუტაციების (ცვლილებების) ანალიზის საფუძველზე შესაძლებელია ცილის ფუნქციური და სტრუქტურული თავისებურებების თაობაზე მეტად მნიშვნელოვანი ინფორმაციის მოპოვება.

კონსერვაცია

ერთ-ერთი მნიშვნელოვანი ინფორმაცია, რომლის მოპოვებაც შესაძლებელია მრავალწევრა ერთდროული ურთიერთსწორების (მეუ) საფუძველზე ეს არის ინფორმაცია კონსერვაციული პოზიციების შესახებ (იხ. სურათი 20.1).



სურათი 20.1. მრავალწევრა ურთიერთსწორების შედეგების ანალიზის პროცესში გამოკვეთილი სტრუქტურასა და ფუნქციასთან დაკავშირებული პოზიციათა მაგალითები: კონსერვაციული პოზიციები, ხის სტოთა განმსაზღვრელი პოზიციები და კორელირებალად მუტირებადი პოზიციათა წყვილები.

უცვლელი შემადგენლობის მქონე პოზიციების არსებობის ახსნა მარტივია - ამ პოზიციებში განლაგებული ამინმჟავები მეტად მნიშვნელოვანია სტრუქტურისა და ფუნქციისათვის და ამის გამო დაუშვებელია მათი შეცვლა ევოლუციის პროცესში. კონსერვატული პოზიციები, როგორც წესი, გვხვდებიან ან ცილის სტრუქტურული ბირთვის ნაწილში (ე.ი. სტრუქტურულად მნიშვნელოვანია), ან ცილის აქტიური ცენტრის ნაწილში (ე.ი. ფუნქციურად მნიშვნელოვანია). თუმცა არსებობს მონაცემები, რომლებიც მიუთითებს გარკვეულ დამოკიდებულებაზე ცილის სხვა ცილებთან დამაკავშირებელ უბნებს, კონსერვატულ პოზიციებსა და ამ პოზიციებში განლაგებულ ამინმჟავებს შორის.

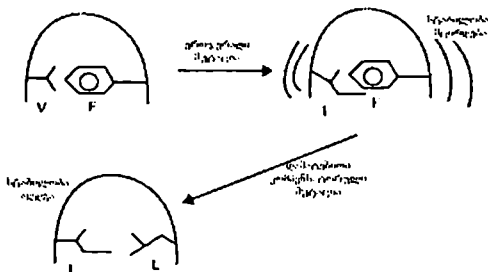
ოჯახზე დამოკიდებული კონსერვაცია (ევოლუციური ხის შტოთა განმსაზღვრელები)

გამოარჩევენ კონსერვაციის უფრო ნატიფ სახესაც, ე.წ. ოჯახზე დამოკიდებულ კონსერვაციას. ასეთი ტიპის კონსერვატულ პოზიციებში განლაგებულ ამინმჟავებს სხვანაირად უწოდებენ ევოლუციური ხის შტოთა განმსაზღვრელებს. ოჯახზე დამოკიდებულების შემთხვევა გულისხმობს პოზიციების კონსერვაციას მხოლოდ იმ ქვეოჯახთა შიგნით, რომლებიც წარმოადგენენ ფილოგენეტიკური ხის ცალსახად გამოკვეთილ შტოებს (ტოტებს) და ერთმანეთისაგან განსხვავდებიან მოცემულ პოზიციასში, მხოლოდ მოცემული ქვეოჯახისათვის დამახასიათებელი კონკრეტული ამინმჟავათი (იხ. სურათი 20.1). ამრიგად, ასეთი პოზიციები მოიცავს ოჯახის ფილოგენეტიკური ხის სტრუქტურის განსაზღვრისათვის საკვანძო ინფორმაციას და ალბათ განაპირობებს ქვეოჯახის ისეთ თავისებურებებს, როგორცაა დაკავშირების უნარი სხვა ცილასა თუ სუბსტრატთან.

ურთიერთქმედებაში მონაწილე (ანუ ურთიერთქმედებაზე პასუხისმგებელ) ამინმჟავათა წინასწარმეტყველების მიზნით შექმნილ პირველ მიდგომათა საერთო მიზანს წარმოადგენდა მრავალწევრა ერთდროული ურთიერთსწორებების (მეუ) საფუძველზე იმ პოზიციათა მოძიება, რომლებიც დამახასიათებელია ცილათა დიდი ოჯახის შემადგენლობაში მყოფ თანამიმდევრობათა ცალკეული ჯგუფებისათვის. მიუხედავად იმისა, რომ ეს პირველი მიდგომები შეიქმნა 90-ანი წლების დასაწყისში, მხოლოდ ახლა როგორც თანამიმდევრობათა ისე სტრუქტურების შესახებ მონაცემთა ბაზების მნიშვნელოვანი ზრდის პირობებში შესაძლებელი გახდა იმის ზუსტი შეფასება, თუ რამდენად ცალსახაა დამოკიდებულება ევოლუციური ხის შტოთა განმსაზღვრელ ამინმჟავებსა და ურთიერთქმედებაზე პასუხისმგებელ ფუნქციურ უბნებს შორის.

კოევილუცია (კორელირებადი მუტაციები)

კიდევ ერთი თანამიმდევრობაზე-დაფუძნებული მიდგომა, შექმნილი ცილის სტრუქტურისა თუ ცილური კომპლექსების შესაძლებლობის წინასწარმეტყველების მიზნით, რომელიც იყენებს ასევე მრავალწევრა ერთდროულ ურთიერთსწორებას (მეუ), გულისხმობს კორელირებადი მუტაციების დეტექტირებას და მათ გამოყენებას ან ერთი ცილის შიგნით განთავსებულ, ან სხვადასხვა ცილაში



სურათი 20.2. დემონსტრირება ჰიპოტეზისა, რომელიც კორელირებად მუტაციას უკავშირებს შესაბამის პოზიციებში მყოფ ამინმჟავათა წყვილის სივრცეში დაახლოებას, და ამით ხსნის კონტაქტში მყოფ ამ წყვილის დიდ როლს ცილის სტაბილური სტრუქტურის, თუ ცილა-ცილა სტაბილური კომპლექსის ჩამოყალიბების საქმეში.

მყოფ ამინმჟავათა ურთიერთქმედების ფაქტის დასაფიქსირებლად. კორელირებადი მუტაცია გულისხმობს ცალსახად გამოხატული კოვარიაციული ბუნების პოზიციათა წყვილის არსებობას (იხ. სურათი 20.1 და 20.2). ევოლუციური მოდელი, რომელიც ამართლებს კორელირებად მუტაციებში მონაწილე ამინმჟავათათვის სივრცული მეზობლობის დაშვებას, ეფუძნება სინქრონულ ცვლილებათა მოდელს. უკანასკნელი მოდელი გულისხმობს, რომ დამზერილ კორელირებად მუტაციათა ნაწილი მაინც შეესაბამება კომპენსატორულ მუტაციებს, ანუ შემთხვევებს, როცა სტაბილური სტრუქტურის შენარჩუნების მოთხოვნიდან გამომდინარე რაღაც პოზიციამი განხორციელებული მუტაცია კომპენსირებულია სინქრონულად განხორციელებული მუტაციით სივრცულად მომიჯნავე პოზიციამი. ამრიგად, ადგილი აქვს ცილის (ან ცილოვანი კომპლექსის) სტაბილობისათვის აუცილებელ კონტაქტთა შენარჩუნების მოვლენას. ასეთ ბიოლოგიურ მიდგომებზე დაყრდნობით 1994 წელს Gobel U, Sander C, Shneider R და Valencia A -ს მიერ შემოთავაზებულ იქნა მეთოდი, რომელიც ახდენდა ამინმჟავათა სივრცული სიახლოვის წინასწარმეტყველებას (მართალია დაბალი ეფექტურობით). შემდგომში კორელირებად მუტაციებთან ერთად სხვა მახასიათებლების (მაგალითად კონსერვატულობა ან ჰიდროფობულობა) გამო-

ყენებით, შესაძლებელი გახდა ამინმეჯვურ კონტაქტთა წინასწარმეტყველების გაუმჯობესება. მიუხედავად დაბალი სიზუსტისა, კონტაქტთა წინასწარმეტყველება რიგ შემთხვევებში მეტად სასარგებლოა, ვთქვათ მოდელთა ფილტრიებისა თუ ab initio სიმულირების პროცესში.

ურთიერთქმედ უბანთა წინასწარმეტყველება

სტრუქტურაზე დაფუძნებული მეთოდი (ფიზიკური დოკინგი)

ცნობილი სტრუქტურების მქონე ცილათა კომპლექსისათვის ურთიერთქმედების ფიზიკური სურათის დადგენის ამოცანა წარმოდგენს დოკინგის ზოგადი ამოცანის შემადგენელ ნაწილს. მიუხედავად იმისა, რომ დოკინგის მეთოდის (განსაკუთრებით კი ცილისა და დაბალმოლეკულური ნაერთების დოკინგის) დახვეწა-განვითარების მიზნით დიდი ძალისხმევაა გაწეული და ამ მეთოდის დახმარებით უამრავი ბიოლოგიურად აქტიური ნაერთია (წამალია) დაპროექტებული, ცილა-ცილა ურთიერთქმედების განსაზღვრის ამოცანა ჯერ კიდევ შორსაა საბოლოო გადაწყვეტისაგან. პრობლემა ძირითადად უკავშირდება იმას, რომ ცილის ზედაპირის ურთიერთქმედებაში შესვლის უნარის მქონე ნაწილების განსხვავება ცილის დანარჩენი ზედაპირისაგან ძნელია.

მიუხედავად სიმწიფისა, მკვლევართა გარკვეულმა ჯგუფებმა მაინც შეძლო ფიზიკური დოკინგის განმარტოვებელი, გარკვეული დონით წარმატებული პროგრამათა პაკეტების (მეთოდების) შექმნა. ამ მეთოდთა ნაწილი დოკინგის პროცესში ცილებს განიხილავს მყიფე (ფიქსირებული) სტრუქტურის მქონე ობიექტების სახით და მათი ურთიერთ დაშორებისა და ორიენტაციის ცვლილების პირობებში ახდენს ცილის ზედაპირზე განლაგებულ, ურთიერთქმედ ანუ ურთიერთშეთავსებად ფრაგმენტთა მოძიებას. არის მეთოდები, რომლებშიც გათვალისწინებულია ზედაპირზე განლაგებულ ამინმეჯვართა გვერდით (და ძირითად) ჯაჭვთა მოქნილობა და ასეთ პირობებში ხორციელდება დოკინგი.

დოკინგის პროცესში, ცილის ზედაპირთა ურთიერთშესაბამისობის შესაფასებლად გამოიყენება სხვადასხვა პოტენციალი. პოტენციური ფუნქციები შეიძლება დაიყოს ორ ჯგუფად; ერთს ქმნიან კლასიკური ტიპის ნახევრად-ემპირიული, ატომ-ატომურ ურთიერთქმედებათა აღმწერი ფუნქციები (იხ. პარაგრაფი 18) , რომელთა ფორმირებისას ნაწილი პარამეტრებისა მიიღება კვანტურ-ქიმიური გათვლების საფუძველზე , ხოლო მეორე ნაწილი კი დაბალმო-

ლეკულურ, მოდელურ ნაერთთა კრისტალების რენტგენოგრაფული მონაცემების ანალიზის საფუძველზე; მეორე ჯგუფი აერთიანებს ცოდნაზე დაფუძნებულ პოტენციურ ფუნქციებს, რომლებიც იწარმოება ცნობილი ცილური კომპლექსების სტრუქტურულ მონაცემებზე დაყრდნობით. სტრუქტურული მონაცემების ანალიზის პროცესში აქცენტი კეთდება ცილის ისეთ მახასიათებლებზე, როგორც ბიგაა ჰიდროფობულობა, ელექტროსტატიკური მახასიათებლები და სხვა.

უფრო დაწვრილებით ინფორმაციას დოკინგ მეთოდების თაობაზე შეგიძლიათ გაეცნოთ პარაგრაფ 21-ში და Halperin I, Ma B, Wolfson H, Nussinov R -ის 2002 წლის მიმოხილვას.

თანამიმდევრობაზე დაფუძნებული მიდგომები

ევოლუციური ხის შტოთა განმსაზღვრელი ამინმჟავები. ურთიერთქმედების ზედაპირსა და ევოლუციური ხის შტოთა განმსაზღვრელ ამინმჟავათა შორის ურთიერთდამოკიდებულის არსებობის შესაფასებლად ჩატარებულ იქნა ცილათა მრავალი ოჯახის ამომწურავი ანალიზი. განსაკუთრებით მნიშვნელოვანი იყო პირდაპირი ექსპერიმენტებით იმის ჩვენება, რომ ცილათა სხვადასხვა ქვეოჯახების წარმომადგენელთა შორის ხის შტოთა განმსაზღვრელ ამინმჟავათა ურთიერთგაცვლას მივყავართ ცილათა შორის ურთიერთქმედების თვალსაზრისით სპეციფიკურობის შესაბამის ცვლილებამდე. ასეთ ექსპერიმენტულ მონაცემთა მაგალითებად შეიძლება მოვიყვანოთ 1999 წელს Azuma და თანამშ. -ის მიერ მცირე GTP აზთა ras ოჯახის წარმომადგენელთა კვლევის შედეგები.

კორელირებადი მუტაციები. ითვლება, რომ ცილათშორის კორელირებადი მუტაციები მიანიშნებს იმ ამინმჟავებსა თუ უბნებზე, რომლებიც მონაწილეობენ ცილა-ცილა ურთიერთქმედებაში. ამინმჟავათა სივრცეში დაახლოების (კონტაქტის) წინასწარმეტყველების საქმეში კორელირებად მუტაციებზე დაფუძნებული მიდგომა არ გამოირჩევა დიდი ეფექტურობით. მისი გამოყენებით გამართლებულია ვიწინასწარმეტყველოთ მხოლოდ ამინმჟავების ურთიერთქმედ ცილათა ინტერფეისურ ზედაპირებში განლაგება.

კორელირებად მუტაციებზე დაფუძნებული მიდგომის მიზიდიველობა მდგომარეობს იმაში, რომ წინასწარმეტყველების პროცესი არ მოითხოვს სტრუქტურული ინფორმაციის ქონას და ამით ჩნდება შესაძლებლობა წინასწარმეტყველება ვაწარმოოთ იმ შემთხვევაშიც, როცა ვფლობთ მხოლოდ თანამიმდევრობის შესახებ ინფორმაციას. კონფორმაციულ ინფორმაციაზე დამოუკიდებლობა იმითა-ცაა მიზიდიველი, რომ არაა საჭირო ცილის სტრუქტურის ისეთი თავისებურების გათვალისწინება, როგორცაა ვთქვით თავისუფალ და დაკავშირებულ მდგომა-

რეობებს შორის კონფორმაციული განსხვავებები. ასეთ განსხვავებათა გათვალისწინების აუცილებლობა დიდად ართულებს ფიზიკური დოკინგის ამოცანას.

ჰიბრიდული მიდგომა

მრავალი მეთოდი ურთიერთქმედების ზედაპირის წინასწარმეტყველების პროცესში ერთდროულად იყენებს როგორც თანამიმდევრობაზე დაფუძნებულ, ისე სტრუქტურაზე დაფუძნებულ მიდგომებს. ვთქვათ გამოიყენება კორელირებადი მუტაციების შესახებ ინფორმაცია სტრუქტურულ ინფორმაციასთან კომბინაციაში. ასეთი მეთოდები გამოიყენება მაშინ, როცა ურთიერთქმედებაში შემავალი ცალკეული ცილების სტრუქტურები ცნობილია. ამ დროს კორელირებადი მუტაციის წარმოჩენის საფუძველზე ხდება იმ ამინმჟავურ წყვილთა წინასწარმეტყველება, რომელთა შორისაც შესაძლებელია ვიწინასწარმეტყველოთ ურთიერთქმედების უზრუნველყოფი კონტაქტის არსებობა. თავის მხრივ ასეთნაირად მიღებული ინფორმაცია გამოიყენება საკვლევ ცილათა დოკინგ ანალიზის საფუძველზე წარმოჩენილ შესაძლო კომპლექსთა კომპლექტების ფილტრირებისათვის.

ურთიერთქმედ პარტნიორთა წინასწარმეტყველება.

ექსპერიმენტული მიდგომები

ორნაწილა ჰიბრიდულობაზე დაფუძნებულ ახალ სტრატეგიათა შემუშავებამ საფუძველი დაუდო ურთიერთქმედ პარტნიორთა წარმომჩენ ექსპერიმენტულ მეთოდთა მნიშვნელოვან გუმჯობესებას. საფუარსა და სხვა მისი რიგის ორგანიზმებში მოხდა სრული პროტეომის სკანირება და შედეგად შეიქმნა შესაძლებლობა ყველა ურთიერთქმედ ცილათა წყვილის გამოვლენისა. „ინტერაქტომის“ სრულფასოვანი წარმოდგენის მიზნით შემუშავებულ იქნა სხვა ექსპერიმენტული მიდგომებიც. მაგალითად, დაფუძნებული მასსპექტრომეტრულ ანალიზზე. მართალია, მთელ რიგ, დღეს გამოყენებად ექსპერიმენტულ მეთოდთა სიზუსტე შესაძლებელია კითხვის ქვეშ დავაყენოთ, მაგრამ მაინც

თამამად შეიძლება ვთქვათ, რომ უახლოეს მომავალში იარსებებს ზღვა ინფორმაცია ცილა-ცილა ურთიერთქმედებათა შესახებ.

ურთიერთქმედ ცილათა გაერთიანებები და მონაცემთა ბაზები.

ცილა-ცილა ურთიერთქმედებათა შესახებ ზღვა ინფორმაციის დაგროვება განაპირობებს ურთიერთქმედებათა შესახებ ახალი მონაცემთა ბაზების შექმნას. მიუხედავად ამისა, ურთიერთქმედ ცილათა დიდი, სტანდარტიზებული გაერთიანების (კოლექციის) ჩამოყალიბების პრობლემა ჯერ კიდევ არსებობს. არსებულ ბაზათა მაგალითებად შეიძლება მოვიყვანოთ:

- SPIN-PP, რომელიც აერთიანებს ცნობილი სტრუქტურების მქონე ცილათა კომპლექსებს და მოიცავს მათი ინტერფეისების ფიზიკურ-ქიმიური თვისებების შესახებ ინფორმაციას (<http://trantor.bioc.columbia.edu/cgi-bin/SPIN/>)
- MIPS, რომელიც მოიცავს საფუარის ცილათა შორის არსებულ ურთიერთქმედებათა დიდი წილის შესახებ ინფორმაციას (<http://www.mips.biochem.mpg.de/proj/yeast/tables/interaction/>)
- ProNet მოიცავს ინფორმაციას ადამიანის ცილათშორისი ურთიერთქმედებების შესახებ (<http://pronet.doublet.wisc.edu/>).
- DIP აერთიანებს ურთიერთქმედ ცილათა სხვადასხვა ტიპებს, შესაბამისი ანოტაციითა და ლიტერატურის მითითებით (<http://www.ampere.due-mbi.ucla.edu:8801/dip.html>)
- BIND აერთიანებს ურთიერთქმედებებს, რომლებშიც არა მხოლოდ ცილები არიან ჩართულნი, არამედ ასევე ნუკლეინის მჟავებები და სხვადასხვა დაბალმოლეკულური ნაერთები. აქვე შესაძლებელია ურთიერთქმედებათა დამახასიათებელი ისეთი ინფორმაციის მოძიებაც, როგორცაა ვთქვათ, ურთიერთქმედების თანმხლები კონფორმაციული ცვლილებები, ბიოლოგიური ციკლები რომელთა პროცესშიც მოცემულ ურთიერთქმედებას აქვს ადგილი და სხვა. (<http://www.bind.ca>)

გენომურ ინფორმაციაზე დაფუძნებული კუმპიუტერული მეთოდები

ურთიერთქმედ ცილათა წყვილების დეტექტირების ამოცანის გადასაწყვეტი საკმაოდ ეფექტურ ექსპერიმენტულ მეთოდთა განვითარება-დახვეწის პარალელურად შემუშავებულ იქნა ბიოინფორმატიკური მიდგომებიც. თუმცა ბიოინფორმატიკულ მიდგომათა პირველი გენერაცია სპეციალიზირდებოდა ფუნქციურად დაკავშირებული (ერთი და იგივე ფუნქციის განხორციელებაში მონაწილე) ცილების დეტექტირების ამოცანის გადაწყვეტაზე და არ აკეთებდა აქცენტს კონკრეტული ფიზიკური ურთიერთქმედების დეტექტირებაზე. ასე მაგალითად, წყდებოდა მოცემული სასიგნალო გზის რაღაც ნაწილის განხორციელებაში მონაწილე ცილათა კომპლექტის გამოვლენის ამოცანა და ამავე დროს არ წყდებოდა სიგნალის გადაცემის პროცესში ცილათშორისი ურთიერთქმედებების სურათის გარკვევის ამოცანა. ქვემოთ ზოგადად ავლწერთ სამ ძირითად მიდგომას, რომლებიც წარმოადგენს გენომურ ინფორმაციაზე დაფუძნებით ცილების ფუნქციურ კავშირში ყოფნის განმსაზღვრელ მეთოდთა საფუძველს.

ფილოგენეტიკური პროფილები. ეს მიდგომა გულისხმობს ისეთ გენთა დეტექტირებას, რომლებსაც გააჩნიათ მსგავსი სახეობათშორისი განაწილება. ანუ გენთა არსებობა/არარსებობის თვალსაზრისით ქვევის სურათი სახეობიდან სახეობამდე არ იცვლება. ამ მიდგომის საფუძველად არსებული ჰიპოტეზა გულისხმობს, რომ ასეთი გენები (და შესაბამისი ცილები) გამომდინარე იქიდან, რომ მათ არ შეუძლიათ დამოუკიდებლად არსებობა და ე.ი ფუნქციონირება, იმყოფება ფუნქციურ კავშირში. მიუხედავად მთელი რიგი რთული შემთხვევებისა, სადაც ეს ჰიპოტეზა არ მუშაობს, მოცემული იდეა ფუნქციურ კავშირში მყოფ ცილათა დეტექტირების ამოცანის გადაწყვეტის თვალსაზრისით აღმოჩნდა საკმაოდ ქმედითი.

გენური მეზობლობის კონსერვაცია. ასეთი მიდგომა ეფუძნება იმ მოვლენას, რომ შორეულ ნათესაობაში მყოფ სახეობათა გენომებში ვლინდება გენომის გასწვრივ გარკვეულ გენთა სიახლოვის უცვლელობა (კონსერვაცია). ჯერ კიდევ 1997 წელს Tamames J. et al. -ის მიერ იქნა ნაჩვენები, რომ კონკრეტული ბიოლოგიური ფუნქციის რეალიზებაში მონაწილე ცილათა შესაბამისი გენები გენომში განლაგებულია მეზობლად.

გენთა შერწყმა. მესამე მიდგომა ეფუძნება გარკვეულ გენომებში განცალკავებით არსებულ გენთა სხვა გენომებში შერწყმის მოვლენებს. ასეთ შემთხვევებში, ანუ როცა ორი ცილა გარკვეულ ორგანიზმებში კოდირებულია განცალკავებული გენებით, და სხვა ორგანიზმებში კი ერთი (შერწყმის შედეგად მიღებული) გენით, ლოგიკურია ვივარაუდოთ, რომ ეს ორი ცილა, ერთ შეთხვევაში წარმოდგენილი

როგორც ორი დამოუკიდებელი ერთეული და მეორე შემთხვევაში როგორც ერთი და იგივე ცილის ორი დომენი, ფუნქციურ კავშირში უნდა იმყოფებოდეს. კონკრეტულ მეტაბოლურ პროცესებში მონაწილე ცილების მაგალითზე ნაჩვენები იქნა, რომ ასეთ მოვლენებს ხშირად აქვს ადგილი.

თანამიმდევრობათა შესახებ ინფორმაციაზე დაფუძნებული კუმპიუტერული მეთოდები

ზემოთ აღწერილი მიდგომები არ ითვალისწინებდა თანამიმდევრობათა შესახებ არსებულ ინფორმაციას და ცილათა შორის ურთიერთქმედებათა არსებობის წინასწარმეტყველებისას ეფუძნებოდა გენომის შემადგენლობასთან, მის სტრუქტურულ მოწყობასთან დაკავშირებულ თავისებურებებს (მაგალითად, ინფორმაციას იმის თაობაზე: არის თუ არ არის ესა თუ ის გენი მოცემულ გენომში, რა პოზიცია უჭირავს მოცემულ გენს მოცემულ გენომში და ა.შ.). უახლოეს წარსულში შემოთავაზებულ იქნა მეთოდები, რომლებიც ასევე ეფუძნება ევოლუციის პროცესთან დაკავშირებულ ინფორმაციას, ოღონდ ამჯერად ინფორმაციის მოპოვება ხდება თანამიმდევრობათა ანალიზის საფუძველზე (ანუ გამოიყენება ინფორმაცია თანამიმდევრობათა ევოლუციის თავისებურებების თაობაზე). ქვემოთ განვიხილავთ ორ ასეთ მიდგომას.

კორელირებადი მუტაციები. პირველი მიდგომა ეფუძნება კორელირებად მუტაციათა იდეას, რომლის შესახებაც ზევით გვქონდა საუბარი. როგორც ზემოთ იყო ნათქვამი, ურთიერთქმედების ზედაპირების წინასწარმეტყველების ამოცანის ამოხსნა დაიყვანებოდა ორი ცილის ამინმჟავებს შორის ცილათშორის ურთიერთქმედებაში პოტენციურად შემსვლელი წყვილების წარმოჩენაზე. პარტნიორის მოძიების ამოცანის გადაწყვეტის შემთხვევაში მრავალწევრა ერთდროული ურთიერთსწორებისას (მეუ) შემოიფარგლებიან მხოლოდ ერთი და იგივე სახეობების თანამიმდევრობების გამოყენებით და ვლინდება სამი ტიპის კორელირებად მუტაციებში მონაწილე ამინმჟავური წყვილები. ესენია: ერთი ცილის შიგნით დამზერილი კორელირებად მუტაციებში მონაწილე ამინმჟავური წყვილები, მეორე ცილის შიგნით დამზერილი კორელირებად მუტაციებში მონაწილე წყვილები და ცილათშორისი კორელირებად მუტაციებში მონაწილე წყვილები. ამ სამი ტიპის წყვილთა შორისი კორელაციის დამახასიათებელ სიდიდეთა განაწილებებზე დაფუძნებით ხდება ორი ცილის ურთიერთქმედების ინდექსის შეფასება. გამოვლინდა ცალსახა კავშირი ინდექსის მაღალ მნიშვნელობასა და რეალური ურთიერთქმედების არსებობას შორის.

ფილოგენეტიკურ ხეთა მსგავსება. მეორე მიდგომა ეფუძნება ურთიერთქმედ ან ფუნქციურ კავშირში მყოფ ცილათა ფილოგენეტიკური ხეების მსგავსებას. ასეთი მსგავსება პირველად დაკვირვებულ იქნა ინსულინისა და ინსულინის რეცეპტორის შემთხვევაში. პაზოსმა და ვალენსიამ 2001 წელს ურთიერთქმედ ცილათა წყვილების დიდ სიმრავლეზე ჩატარებული ანალიზის საფუძველზე, ე.ი. დიდი სტატისტიკური დამაჯერებლობით აჩვენეს, რომ არსებობს მკაფიოდ გამონათული დამოკიდებულება ფილოგენეტიკურ ხეთა მსგავსებასა და ცილა-ცილა ურთიერთქმედების არსებობას შორის. მათვე წარმოადგინეს ასეთი შედეგის ამხსნელი შემდეგი ჰიპოტეზა: ურთიერთქმედებაში მყოფი ცილები კოვეოლუციის სუბიექტებია.

ფილოგენეტიკური ხეების მსგავსებაზე დაფუძნებული მეთოდის რეალიზებისას ქმედებათა პირველი ეტაპი ემთხვევა კორელირებად მუტაციებზე დაფუძნებული მეთოდის რეალიზების პირველ ეტაპს, ანუ აქაც ცილათა ურთიერთსწორება შემოიფარგლება ერთი და იგივე სახეობების თანამიმდევრობების განხილვით. ურთიერთსწორების მონაცემების საფუძველზე თითოეული ცილისათვის (წყვილიდან) იგება მატრიცა, რომლის ელემენტებიც გვიჩვენებს მოცემული ცილის თანამიმდევრობის სხვა ორგანიზმებიდან აღებულ თანამიმდევრობებთან მსგავსებას (დისტანციას). ამგვარად მიღებული მატრიცები მოიცავენ ცილათა ფილოგენეტიკური ხეების ასაგებ მონაცემებს. ამიტომ, ფილოგენეტიკურ ხეთა მსგავსების შესაფასებლად საკმარისია მოცემულ მატრიცათა შორის მსაგვსების შეფასება. რაც ხორციელდება მათემატიკური სტატისტიკის ცნობილ, კორელაციური ანალიზის მეთოდებზე დაფუძნებით.

პერსპექტივები

უნდა ვივარაუდოთ, რომ მომავალში უჯრედში ცილა-ცილა ურთიერთქმედებათა სრული სურათის წარმოდგენის მიზნით მოხდება მძლავრ ექსპერიმენტულ მეთოდთა (საფუარში ორნაწილა ჰიბრიდულობაზე დაფუძნებული მიდგომა და მასსპექტროსკოპული მიდგომა) ძალისხმევით, თანამედროვე ბიოინფორმატიკულ მიდგომათა ძალისხმევით, სტრუქტურული გენომიკის პროექტთა ფარგლებში გაიფრულ სტრუქტურათა დოკინგის ძალისხმევით, დნმ- მიკრო-მატრიცებზე განხორციელებული ფართომასშტაბიანი ექსპერიმენტების ძალისხმევით მიღებულ შედეგთა ურთიერთშეჯერება. ამ ზღვა მონაცემთა სამართავად და შესანახად, ასევე შეიქმნება ახალი მონაცემთა ბაზები. ასეთი მძლავრი ინფორმაციული ნაკადების პირობებში მარტივი მოდელური სისტემებისათვის (მაგალითად საფუარები და ბაქტერიები) ჩვენ მალე შევძლებთ ცოცხალი უჯრედის ცხოველმყოფელობის წარმმართველ ცილურ ურთიერთქმედებათა ურთულესი ქსელის სტრუქტურაზე ზედაპირული წარმოდგენის შექმნას.

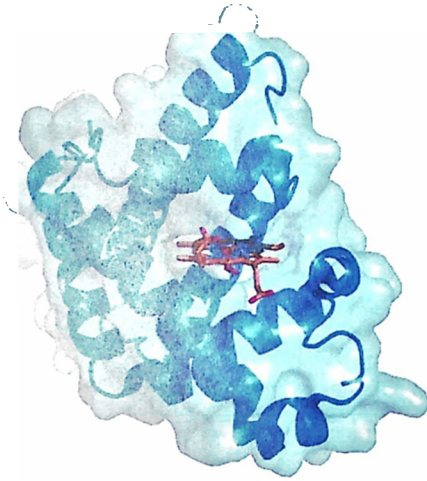
დღეს უკვე ვართ იმის მოწმე, თუ ურთიერთქმედებათა შესახებ ინფორმაციის დაგროვება როგორ ასტიმულირებს ფუნქციური და სტრუქტურული კავშირების შესახებ ახალი ინფორმაციის მოდინებას. ეს კი წარმოადგენს არა მხოლოდ უჯრედის ფუნქციონირების მექანიზმებში უფრო ღრმად ჩაწვდომის გარანტიას, არამედ ამ ფუნქციის მართვის საშუალებების, ახალი, ბიოლოგიურად აქტიურ ნაერთთა შექმნის (პროექტირების) გარანტიასაც.

21. დოკინგი. ლიგანდთა (წამალთა) პროექტირება.

სტრუქტურულ ბიოინფორმატიკას თავისი წვლილი შეაქვს ახალი ქიმიური ნაერთების აღმოჩენაში, პროექტირებასა და ოპტიმიზაციაში. ბიოინფორმატიკის დახმარებით წარმოებულ ქიმიურ ნაერთთა სპექტრი მრავალფეროვანია და მოიცავს წამლებს, ბიომასალებს, კატალიზატორებს, ზოგადად ახალ მაკრომლეკულებს. მოლეკულათა პროექტირება მნიშვნელოვან როლს თამაშობს ისეთი დარგების განვითარებაში როგორცაა ორგანული ქიმია, ფიზიკური ქიმია, ქიმიური ინჟინერია, ქიმიური ფიზიკა, ბიოინჟინერია, მოლეკულური ბიოლოგია. ახალი თვისებების მატარებელ ბიომასალათა პროექტირებისათვის გამოყენებული მეთოდები ასევე მრავალფეროვანია. ჩვენ შევეცდებით ჯერ წარმოვადგინოთ ის ძირითადი პრინციპები, რომლებსაც ეფუძნება ზოგადად პროექტირების პროცესი და შემდგომ განვიხილავთ სივრცულ სტრუქტურაზე დაფუძნებული წამლის პროექტირებას, რომელიც წარმოადგენს პროექტირების იმ სფეროს, სადაც დაგროვილია უდიდესი გამოცდილება.

რა არის ის არსებითი მოთხოვნა, რამაც განაპირობა ახალ ლიგანდთა პროექტირების აუცილებლობა? სიტყვა „ლიგანდი“ (ligand) წარმოებულია სიტყვიდან (ligator), რაც ნიშნავს „თასმას“, კავშირს. ლიგანდი, როგორც წესი, იხმარება იმ მოლეკულის აღსანიშნავად, რომელიც სხვა მოლეკულასთან მხოლოდ არაკოვალენტურად ურთიერთქმედებს (იხ. სურათი 21.1). მეორე მოლეკულა, რომელსაც მოიხსენიებენ როგორც „სამიზნეს“ ან „რეცეპტორს“, ჩვეულებრივ უფრო დიდი ზომისაა, ვიდრე ლიგანდი. ლიგანდ-რეცეპტორულ კომპლექსთა ფიზიკურ-ქიმიურ-ბიოლოგიურ თავისებურებათა მრავალფეროვნებას განსაზღვრავს ლიგანდთა მრავალფეროვნება. ლიგანდსა და რეცეპტორს შორის არსებულ ურთიერთქმედებათა ბუნება განისაზღვრება ბალანსით იმ ძალებს შორის, რომლებიც მოქმედებს ლიგანდსა და რეცეპტორს შორის ერთის მხრივ და ცალკეულ მოლეკულებსა და გამხსნელის მოლეკულებს შორის მეორეს მხრივ. ეს ძალები ატომის შემადგენელ ნაწილაკთა შორის ურთიერთქმედებების შედეგია და ფუნდამენტურ მიახლოებაში მათი შესწავლა უნდა წარმოებდეს კვანტური მექანიკის მიდგომებით. თუმცა კვანტური თეორიის გამოყენება ისეთი ბიოლოგიური სისტემების შესწავლის პროცესში, რომლებიც შედგება რამდენიმე ამინმჟავსაგანან კი შეზღუდულია (არარეალური კომპიუტერული რესურსების საჭიროებიდან გამომდინარე). ამიტომ მოლეკულურ ურთიერთქმედებათა აღსაწერად გამოყენებულ კომპიუტერულ მეთოდთა უმრავლესობა გულისხმობს ნახევრადემპირიული ან სულაც ემპირიული პოტენციალებით სარგებლობას და შესაბამისად, არც თუ ისე ზოგადია. თუმცა ექსპერიმენტის საშუალებით შესაძლებელია თეორიის მკაცრი შეფასება.

ლიგანდთა პროექტირების პროცესში ძირითადად ხდება ისეთ თერმოდინამიკურ მახასიათებლებზე დაყრდნობა, როგორიცაა: დაკავშირების თავისუფალი ენერგია, პოლარულ და არაპოლარულ გარემოში ხსნადობა, და ა.შ. კინეტიკური მახასიათებლები, გამომდინარე იქიდან, რომ მათი როგორც ექსპერიმენტული, ისე



სურათი 21.1. მიოგლობინის მოლეკულა (ცისფრად) და მისი ლიგანდი, ჰემი (ნარინჯისფრად)

თეორიული შეფასება გაძნელებულია, ნაკლებადაა გათვალისწინებული. თუმცა კინეტიკა მნიშვნელოვან როლს თამაშობს ისეთ ფარმაკოლოგიურად საინტერესო პროცესებში, როგორიცაა ფერმენტული კატალიზი, სიგნალთა გადაცემა და მოლეკულური სისტემების რეორგანიზაცია.

შეჯერდეთ დაკავშირების თავისუფალ ენერგიაზე და მის შემადგენელ დაკავშირების ენტალპიასა და ენტროპიაზე. დაკავშირების ენერგია განიმარტება ასე:

$$\Delta G_{\text{bound}} = \Delta G_{\text{complex}} - (\Delta G_{\text{ligand}} + \Delta G_{\text{receptor}}) \quad (21.1)$$

ასეთნაირად განსაზღვრული ΔG_{bound} წარმოადგენს ტემპერატურის, იონური ძალის, pH-ის, გამხსნელის ტიპის და მონაწილე ქიმიური ნერთების კონცენტრაციების ფუნქციას. Δ ნიშანი მიუთითებს იმაზე, რომ ექსპერიმენტულად შეუძლებელია თავისუფალი ენერგიის აბსოლუტური მნიშვნელობის შეფასება. ამიტომ რეალურად, გარკვეულ პირობებში მყოფი სისტემის თავისუფალი ენერგია გამოისახება რაღაც ათვლის მდგომარეობის მიმართ, ანუ როგორც სხვაობა ორ მდგომარეობას შორის. ზოგადად ΔG_{bound} შეფასება ხდება კომპლექსის წონასწორობის მახასიათებელ მუდმივაზე დაფუძნებით.

$$\Delta G_{bound} = -RT \ln K_{eq} = RT \ln K_d \quad (21.2)$$

სადაც R გაზური მუდმივია, ხოლო T კი აბსოლუტური ტემპერატურა (კელვინებში). ნათელია, რომ ლიგანდთა პროექტირებისას დიდ ინტერესს იწვევს K_d შერჩევა, რადგან სწორედ ის ახასიათებს ლიგანდის მეტად მნიშვნელოვან თვისებას - ლიგანდის რეცეპტორისადმი სწრაფვის ხარისხს. ფარმაკოლოგიურ ლიტერატურაში ხშირად იხმარება K_d დისოციაციის კონსტანტა, რომელიც K_d შებრუნებულ სიდიდეს წარმოადგენს. ასევე იხმარება IC_{50} , ანუ 50%-ანი აქტივობის შესაბამისი ლიგანდის კონცენტრაცია.

ამ პარაგრაფში მოკლედ მიმოვიხილავთ წამალთა პროექტირებისათვის გამოყენებულ კომპიუტერულ მიდგომებს. საზოგადოდ ეს მიდგომები გულისხმობს ისეთ მოლეკულათა გამოვლენა/პროექტირებას, რომლებსაც შეუძლიათ სწრაფვის დიდი ხარისხითა და სპეციფიკურობით დაუკავშირდნენ ცნობილი სივრცული სტრუქტურის მქონე ბიოლოგიურ სამიზნეებს. იმისათვის, რომ სწრაფვის ხარისხზე დაფუძნებით შერჩეული მოლეკულები საბოლოოდ ჩამოყალიბდნენ პოტენციურ წამლებად, შეიძლება საჭირო იყოს მათი შემდგომი მოდიფიკაცია, რათა შეიცვალოს მათი ხსნადობის ხარისხი, მოლეკულური წონა, მეტაბოლური მახასიათებლები და ა.შ. რაც შეეხება სტრუქტურაზე დაფუძნებულ მიდგომებს, მათდამი ინტერესი უპირველეს ყოვლისა განპირობებულია იმით, რომ სივრცულ სტრუქტურათა ცოდნა (როგორც რეცეპტორის, ისე ლიგანდის) მნიშვნელოვნად აჩქარებს პროექტირების პროცედურას და ზრდის პროექტირების ხარისხს. თუმცა პროექტირების განხორციელება სივრცული სტრუქტურების შესახებ ინფორმაციის არარსებობის პირობებშიცაა შესაძლებელი.

ლიგანდთა კომპიუტერული პროექტირების მეთოდები.

კომპიუტერული პროექტირებისას უპირველეს ამოცანას წარმოადგენს საკვანძო (ცალსაბა აქტივობის მაჩვენებელ) ნაერთთა იდენტიფიცირება. უახლოეს წარსულში, ბაზარზე არსებულ წამალთა უმრავლესობა სწორედ ასეთ საკვანძო ნაერთთა საფუძველზე იქმნებოდა. საკვანძო ნაერთების წარმოჩენა კი ან არსებულ ბუნებრივ ნაერთთა ბიოქიმიური მეთოდებით სკრინინგის, ან არსებულ სუბსტრატთა თუ ლიგანდთა ანალოგების გამოცდის შედეგად წარმოებდა. ანუ ახალ წამალთა შექმნის პროცესში სამიზნის სივრცული მოწყობის შესახებ ინფორმაცია, გამომდინარე მისი მოპოვების შეუძლებლობიდან, არ გამოიყენებოდა. ბოლო წლებში, სივრცული სტრუქტურის დამდგენ მეთოდთა დახვეწასთან ერთად შეგვიძლია ვივარაუდოთ, რომ უახლოეს მომავალში ინფორმაცია უმრავლეს სამიზნეთა სივრცული მოწყობის შესახებ ხელმისაწვდომი გახდება. ასე რომ,

წამალთა პროექტირების პროცესში მეთოდები, რომლებიც ეფუძნება სივრცული მოწყობის შესახებ ინფორმაციას მეტ აქტუალობას შეიძენს.

დღეს კომპიუტერული მეთოდები წამალთა პროექტირებასთან დაკავშირებული თითქმის ყველა შესაძლო ამოცანის გადაწყვეტის ძირითად საშუალებას წარმოადგენს. ამოცანები, რომელთა გადაწყვეტაც კომპიუტერულ მიდგომებს ეფუძნება შემდეგია: კომბინატორულ ბიბლიოთეკათა სკრინინგი, სამიზნე მოლეკულათა სივრცული სტრუქტურის დადგენა, ახალ საკვანძო ნაერთთა პროექტირება ან არსებულთა გაუმჯობესება.

ლიგანდთა (პოტენციურ წამალთა) კომპიუტერული პროექტირების მეთოდები შესაძლებელია წარმოვადგინოთ ორ სტრატეგიულად განსხვავებულ ტიპად: ლიგანდზე დაფუძნებული პროექტირება (ანუ ანალოგიაზე) და სამიზნეზე დაფუძნებული პროექტირება (ანუ სტრუქტურაზე). კონრეტული ამოცანის სპეციფიკიდან გამომდინარე, ამ ორ სტრატეგიას იყენებენ როგორც ცალ-ცალკე, ისე ერთად.

კომპიუტერზე დაფუძნებული ლიგანდთა (წამალთა) პროექტირების პროცესში გამოიყენება საკმაოდ ფართო სპექტრი კომპიუტერული მიდგომებისა. ერთ-ერთი მიდგომის მიხედვით გაითვალისწინება კლასიკური მექანიკის კანონზომიერებები ფასდება მოლეკულის გეომეტრიის, ენერჯისა და დინამიკური მახასიათებლები. ტიპიურ ურთიერთქმედებათა აღწერა ხდება ატომ-ატომური ურთიერთქმედებების მიახლოებაში (იხ. 18 პარაგრაფში აღწერილი პოტენციური ფუნქცია). ერთი-ერთი ამგვარი, მოლეკულური მექანიკის სახელით ცნობილი, მიდგომის ძირითად ამოცანას მოლეკულური კომპლექსის ენერჯის ოპტიმიზაცია წარმოადგენს. ოპტიმიზაცია გულისხმობს შესაბამისი პოტენციური ფუნქციის ჰიპერზედაპირის ანალიზს ენერჯის მინიმუმის პოვნის მიზნით. სხვა მიდგომის, კერძოდ კი მოლეკულური დინამიკის შემთხვევაში, ოპტიმიზაციის პროცესი ითვალისწინებს სისტემის კონფორმაციულ სივრცეში მოძრაობის სითბური პროცესით სიმულირებას. კონფორმაციულ სივრცეში მოძრაობის სიმულირება გულისხმობს დროის მცირე თანამიმდევრული მონაკვეთებისათვის მექანიკის განტოლებათა რიცხვითი მეთოდით ამოხსნას, რათა მიღებულ იქნეს ენერჯის ოპტიმიზისაკენ მიმავალი ტრაექტორია, ანუ გლობალური ენერგეტიკული მინიმუმისაკენ მიმავალ გზაზე განლაგებულ თანამიმდევრულ კონფორმაციათა სიმრავლე („ფოტო კადრები“). ამ დროს საკვლევ ბიომაკრომოლეკულებთან ერთად სისტემის შემადგენელ ნაწილებად მოიაზრება გამხსნელის მოლეკულებიც. არსებობს აგრეთვე სტოქასტიკური, მონტე-კარლოს მიდგომა, როცა სისტემის მდგომარეობათა სივრცე ალბათობათა განაწილების ენაზე აღიწერება, ხოლო მდგომარეობათა განაწილება ითვლება, რომ შესაბამებ ბოლცმანის განაწილებას.

ანალოგიაზე დაფუძნებული პროექტირება.

ანალოგიაზე დაფუძნებით პროექტირებისას ფარმაკოფორული ხატი და აქტივობის-სტრუქტურაზე დამოკიდებულების აღმწერი რიცხვითი მახასიათებლის (QSAR-ს – quantitative structure-activity relationship) გამოყენებით, რეცეპტორის სივრცული სტრუქტურის არ არსებობის პირობებში, ვლინდება საკვანძო ნაერთი ან ხდება არსებული საკვანძო ნაერთის მოდიფიცირებული ვარიანტის შემოთავაზება.

ფარმაკოფორულ ხატში იგულისხმება ლიგანდის ძირითად (როგორც წესი, ფიზიკურ-ქიმიურ) თავისებურებათა შესახებ ჰიპოტეტური, გამარტივებული გეომეტრიული სურათი (ხატი). თავისებურებათა ჩამონათვალი, როგორც წესი, მოიცავს ინფორმაციას: წყალბადურ ბმათა დონორებისა და აქცეპტორების, დამუხტული ჯგუფების და ჰიდროფობული სტრუქტურების შესახებ. ჰიპოტეტური ხატის გამოყენებით შესაძლებელია ქიმიურ ნაერთთა ბაზების სკანირება და საკვანძოდ ვარგის ნაერთთა წარმოჩენა. საკვანძოდ შერჩეულ ნაერთთა გეომეტრიული ურთიერთსწორების სურათის მისაღებად აუცილებელია ნაერთების დაკავშირებულ მდგომარეობათა შესაბამის კონფორმაციათა ცოდნა. ამრიგად, სწორი ფარმაკოფორული ჯგუფების კონსტრუირების, მათი გეომეტრიული მოწყობის წარმოჩენის მიზნით, საწყის ეტაპზე გამართლებულია ვაწარმოოთ მხოლოდ შედარებით ხისტი (ნაკლები კონფორმციული თავისუფლების ხარისხების მქონე) მოლეკულების ურთიერთსწორება.

იმის გამო, რომ აქტივობის საწინასწარმეტყველოდ უბრალო ფარმაკოფორული თითის ანაბეჭდის არსებობა არაა საკმარისი, ასევე უნდა მოხდეს სასურველი ფარმაკოფორული თვისებების მატარებელი, მაგრამ არააქტიური ნაერთების განხილვაც. ასე მაგალითად, აქტიურ და არააქტიურ ნაერთთა მოცულობების შედარების საფუძველზე მოხდება რეცეპტორულ უბანთან (რომლის სტრუქტურაც უცნობია) შეთავსებადი მოცულობის შერჩევა და ამ მახასიათებლის, ფარმაკოფორული მოდელის შემდგომი დახვეწისა და დამატებით საკვანძო ნაერთთა წარმოჩენის მიზნით, ბაზის ხელახალი სკანირებისთვის გამოყენება.

QSAR ანალიზის მთავრ მიზანს ქიმიურ შემადგენლობაზე დაფუძნებით ნაერთის აქტივობის წინასწარმეტყველება წარმოადგენს. აქ ძირითადი დაშვება მდგომარეობს იმაში, რომ ბიოლოგიური აქტივობა ნაერთის შემადგენლობაში მყოფი ფრაგმენტების აქტივობების მარტივი ჯამია, ანუ კორელაციაშია ცალკეულ ფრაგმენტთა არსებობასთან. ასეთ დაშვებას სხვანაირად მოიხსენიებენ ხაზოვანი თავისუფალი ენერჯის პრინციპად. ინფორმაცია ურთიერთქმედების ძალების შესახებ მოიგება სტერიულ, ელექტროსტატიკურ და ჰიდროფობულ მახასიათებელთა შეფასებებზე დაყრდნობით. QSAR ზე დაფუძნებული წინასწარმეტყველება

შემოიფარგლება იმ ახალ ნაერთთა აქტივობების წინასწარმეტყველებით, რომლებიც მსგავსია საცდელ კომპლექტში გაერთიანებულ ნაერთებთან. უნდა აღინიშნოს, რომ არსებობს შემთხვევითი მსგავსების არსებობით გამოწვეული რისკიც.

სამიზნის სტრუქტურაზე დაფუძნებული პროექტირება.

წამალთა საიტ-დამოკიდებული პროექტირების საწყისი ეტაპი ყოველთვის გულისხმობს სამიზნე მოლეკულის სამგანზომილებიანი სტრუქტურის დადგენას ან რენტგენული კრისტალოგრაფიის და ბირთვული მაგნიტური რეზონანსის, ან კომპიუტერული (ჰომოლოგიურობის საფუძველზე მოდელირება, ab initio მოდელირება, იხ. პარაგრაფი 18) წინასწარმეტყველების მეთოდებზე დაყრდნობით.

რეცეპტორის ზედაპირის შებრუნებული გამოსახულება მიგვითითებს ლიგანდის დასაკავშირებლად გამოსადეგ ადგილზე. ასეთ ადგილად შეიძლება შეირჩეს რამდენიმე. რეალური დაკავშირების ადგილის გამოსვლენად შესაძლებელია გამოყენებულ იქნეს მონაცემები მონათესავე ლიგანდ-ცილოვანი კომპლექსის შესახებ (ცხადია, თუ ასეთი არსებობს და მისი სტრუქტურა ცნობილია). როგორც წესი, დაკავშირების ადგილი შეესაბამება ყველაზე დიდ ჩაღრმავებას ცილის ზედაპირზე.

ყველაზე ძნელად გადასაწყვეტ ამოცანას კონფორმაციული ანალიზი წარმოადგენს. თუ ლიგანდი არა მარტივი ნაერთი, სითბური მოძრაობის გამო მის სტრუქტურას რაღაც კონკრეტული გეომეტრიული ფორმით ვერ ავლწერთ. დაკავშირების ენერგიას განსაზღვრავს: თავისუფალ მდგომარეობაში ლიგანდის მინიმალური ენერგია და რეცეპტორთან დაკავშირებული მდგომარეობის ენერგია, უფრო ზუსტად კი მათი სხვაობა. დასაკავშირებლად გამოსადეგ კონფორმაციათა რიცხვი მნიშვნელოვნად ნაკლებია დასაკავშირებლად გამოუდგე კონფორმაციათა რიცხვზე, რაც მიუთითებს დაკავშირებისას ენტროპიის შემცირებაზე. გამომდინარე იქიდან, რომ სხვადასხვა კონფორმაციაში ლიგანდს გააჩნია სხვადასხვა გეომეტრია და ფიზიკური მახასიათებლები, საჭიროა ლიგანდის კონფორმაციული ანალიზის განხორციელება. კონფორმაციული ანალიზისას მიმართავენ სხვადასხვა სტრატეგიას.

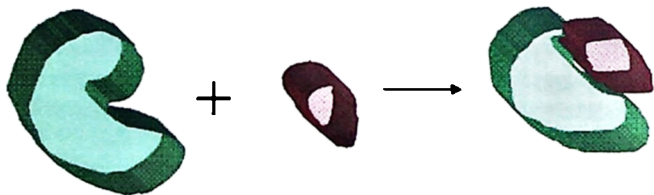
ლიგანდთა საიტ-დამოკიდებულად პროექტირების მეთოდები შესაძლებელია დაიყოს ორ ტიპად. ესენია : დოკინგი და de novo შენება (პროექტირება).

დოკინგი გულისხმობს აქტიურ ცენტრთან შესაბამისობაში მყოფი ნაერთის წარმოჩენას მონაცემთა ბაზებში ძიების საფუძველზე.

de novo შენება კი - კონკრეტული რეცეპტორის აქტიური ცენტრის მოწყობის თავისებურებების შესაბამისად შერჩეული ატომების ან მოლეკულათა ცალკეული ფრაგმენტების ერთ ნაერთად გაერთიანებით ახალი ლიგანდის გენერირებას.

დოკინგი.

მოლეკულათა დოკინგის ძირითად ამოცანას წარმოადგენს ცნობილი სივრცული სტრუქტურის მქონე სამიზნესთან დაკავშირების უნარის მქონე ლიგანდის შერჩევა (იხ. სურათი 21.2). შერჩევის პროცესი გულისხმობს დაკავშირების ფორმების ანალიზს. დაკავშირების ფორმებში, რომელსაც ხშირად მოიხსენიებენ, როგორც დაკავშირების მოდებს ან პოზიციებს, გულისხმობენ როგორც ლიგანდის პოზიციონირებას (განლაგებას) რეცეპტორის მიმართ, ისე ლიგანდისა და რეცეპტორის კონფორმაციულ მდგომარეობებს. ერთის მხრივ კონფიგურაციულ და კონფორმაციულ სივრცეთა კვლევა და მეორეს მხრივ თითოეული გეომეტრიის ენერგეტიკული შეფასება, დოკინგის მეტად მნიშვნელოვან ორ შემადგენელ, განსხვავებულ ამოცანას წარმოადგენს. დოკინგ-მეთოდის ეფექტურობა ფასდება მისი უნარით სწრაფად და აკურატულად განახორციელოს დიდი რაოდენობა დაბალმოლეკულურ ნაერთთა დოკინგი რეცეპტორის აქტიურ ცენტრში და მოახდინოს მათი რანჟირება რეცეპტორთან ურთიერთქმედების სიძლიერის (სწრაფვის) მიხედვით. ლიგანდ-რეცეპტორის ურთიერთქმედების უნარის სწორი შეფასება, ცხადია, ცალსახადაა დამოკიდებული ურთიერთქმედების თავისუფალი ენერჯის კორექტულ შეფასებაზე.



სურათი 21.2. დოკინგი გულისხმობს ლიგანდის (ყავისფერი), რეცეპტორის (მწვანე) მიერთების უზნათან შორჯებას.

ამრიგად დოკინგის პროცედურა ზოგადად გულისხმობს სამი დამოუკიდებელი ამოცანის გადაწყვეტას. ესენია: 1) დაკავშირების უბნის (საიტის) დახასიათება; 2) ლიგანდის პოზიციონირება (ორიენტირება) დაკავშირების ცენტრში (საიტში); 3) ლიგანდ-რეცეპტორული კომპლექსის თითოეული კონფიგურაციისათვის ურთიერთქმედების ძალის შეფასება (ქულის მინიჭება). ქულათა დარიცხვის პროცესში

გამოყენებად ფუნქციათა სპექტრი ძალიან ფართოა. გამოყენებად ფუნქციათა ერთ-ერთ მაგალითად გამოდგება 18 პარაგრაფში აღწერილი პოტენციალი. აქ შესამე ამოცანაზე დეტალურად არ შევჩერდებით და ქვემოთ ძირითადად ვისაუბრებთ პირველ ორ ამოცანაზე.

იმისათვის, რომ შესაძლებელი იქნეს დიდი მოცულობის მონაცემთა ბაზების სკანირება, სასურველია, რომ დოკინგის მეთოდი იძლეოდეს ავტომატური ძიების საშუალებას. ავტომატური ძიების მეთოდები შეიძლება დავეყთ: გეომეტრიაზე დაფუძნებულ ძიების მეთოდებად და ენერგიაზე დაფუძნებულ ძიების მეთოდებად. გეომეტრიული შეთავსებადობის შეფასებებზე დაფუძნებული მეთოდები იყენებს ევრისტიკულ მიდგომებს, რათა კომბინატორიკაზე დაფუძნებული ძიებით მოხდეს ლიგანდსა და რეცეპტორის მახასიათებლებს შორის შეთავსებადი ვარიანტების შერჩევა. ენერგეტიკაზე დაფუძნებული ძიებისას კი, ლიგანდის საბოლოო განლაგების შერჩევისათვის გამოიყენება ლიგანდ-რეცეპტორის ურთიერთქმედების ენერჯის მინიმიზირების პროცედურა. როგორც წესი, მინიმიზირება ხორციელდება ან მონტე-კარლოს, ან მოლეკულური დინამიკის, ან გენეტიკურ ალგორითმზე (იგი, როგორც მონტე-კარლო სტოქასტიკურია) დაყრდნობით.

დაკავშირების უბნის (საიტის) დახასიათება. დაკავშირების უბნების დახასიათებლად შექმნილ პროგრამათა პაკეტების მაგალითად შეიძლება მოვიყვანოთ DOCK (Erwing et al. 2001) რომელიც დაკავშირების უბნის შებრუნებულ გამოსახულებას წარმოადგენს გადაფარვად სფეროთა ფორმით. დაკავშირების უბანში (საიტში) ლიგანდის ხელსაყრელი ორიენტაცია განისაზღვრება მისი ამ სფეროთა ცენტრებთან ურთიერთშეთავსების პროცედურით. შემდგომი, ლიგანდ-რეცეპტორული ურთიერთქმედების ენერჯის მინიმიზირება ემსახურება სტერიული კომპლემენტარობის გაუმჯობესებას.

რეცეპტორის დაკავშირების უბნის აღწერა შესაძლებელია ქიმიურ მახასიათებელთა ენაზე, ისე როგორც ეს ხდება CLIX (Lawrence MC and Davis PC 1992) ალგორითმის შემთხვევაში. ამ დროს აღწერა, ანუ შეთავსების წერტილების დაფიქსირება, გულისხმობს აქტიური ცენტრის საკვანძო ფუნქციონალური ჯგუფებისა (უბნებისა) და მათი გეომეტრიის აღწერას. დოკინგის პროცესში, როცა ლიგანდის ფუნქციონალური ჯგუფი კონტაქტში მოდის აქტიური ცენტრის შესაბამის საკვანძო ფუნქციონალურ უბანთან, ადგილი აქვს დაკავშირების მახასიათებელ ქულათა ჯამის ზრდას.

ცილა-ცილა დოკინგის შემთხვევაში, როგორც წესი, ხდება ცილათა ზედაპირების ანალიზი და მათზე ურთიერთქმედების უბნების მოძიება. ამოცანის გამარტივების მიზნით, პირველ ეტაპზე, ზედაპირის წარმოდგენა ხდება დაბალ-გარჩევით მიახლოებაში. ატომურ მოწყობასთან დაკავშირებული დეტალების დაზუსტებაზე ზრუნვა კი შემდგომ ეტაპებზე წარმოებს (Vakser IA 1995). ისევე

როგორც ცილა-ლიგანდის შემთხვევაში, აქაც შესაძლებელია ზედაპირის მოწყობის ისეთ ფიზიკურ-ქიმიურ მახასიათებელთა ენაზე აღწერა, როგორცაა ჰიდროფობულობა ან ელექტროსტატიკური მახასიათებლები.

ორიენტაციის შერჩევა. თუ ლიგანდი ან მისი რაიმე ფრაგმენტი შეგვიძლია განვიხილოთ როგორც ხისტი, უცვლელი გეომეტრიის მქონე მოლეკულა, მისი რეცეპტორის დაკავშირების უბანში ორიენტაციის განსაზღვრის პროცედურა გულისხმობს ორიენტაციის განმსაზღვრელი ექვსგანზომილებიანი სივრცის გარკვეული ბიჯით კვლევას. განხილულ ორიენტაციას შორის ხელსაყრელთა შერჩევა წარმოებს ურთიერთქმედების აღმწერ პოტენციური ფუნქციის მნიშვნელობათა შეფასებების მეშვეობით.

ორიენტაციის შერჩევის თანამედროვე მიდგომების ეფექტურობის გაზრდის მზნით გამოიყენება დამატებითი ინფორმაცია. ასე მაგალითად, დაბალმოლეკულურ ნაერთთა დოკინგის განმახორციელებელ მეთოდთა უმრავლესობა ეფუძნება მახასიათებელთა შეთავსებაზე დაფუძნებულ მიდგომას. ამ მიდგომის თანახმად წარმოებს ლიგანდის მახასიათებელ-დისკრიპტორების (წერტილების, რომლებსაც მიეწერებათ რაღაც, ლიგანდისათვის დამახასიათებელი, მისი უნიკალობის განმსაზღვრელი თვისება) რეცეპტორის დაკავშირების უბნის მახასიათებელ-დისკრიპტორებთან სივრცეში შეთავსების მცდელობა. შეთავსების თაობაზე მოთხოვნა აადვილებს ლიგანდის რეცეპტორის დაკავშირების ცენტრში სწორ (მისაღებ) ორიენტაციას. მისაღებ კონფიგურაციათა შემდგომი დახვეწის მიზნით უკვე გამოიყენება ქულათა დამრიცხავი (ურთიერთქმედების ენერჯის აღმწერი) ფუნქციის ოპტიმუმის მოძიების პროცედურა.

de novo პროექტირება.

de novo პროექტირების (შენების) ძირითად მიზანს წარმოადგენს ადრე არ არსებულ ნაერთთა (მოლეკულათა) პროექტირება-კონსტრუირება. არსებობს de novo პროექტირების მეთოდთა სამი კლასი. ესენია: ფრაგმენტთა პოზიციონირების მეთოდი, ფრაგმენტთა დაკავშირების მეთოდი და თანამიმდევრული შენების მეთოდი.

ფრაგმენტთა პოზიციონირება.

ეს მეთოდი არ ითვალისწინებს ლიგანდის მოლეკულის, როგორც მთლიანის სწორ განლაგების გამოვლენას, არამედ ავლენს რეცეპტორის დაკავშირების უბანში ლიგანდის მხოლოდ ცალკეულ ატომთა ან ფრაგმენტთა სწორ განლაგებას. ასეთი მიდგომის შემთხვევაში ძირითადი დაშვება შემდეგია: მცირე რაოდენობა სწორად განლაგებული ფრაგმენტებისა უზრუნველყოფს დაკავშირებისათვის საკმარისი ურთიერთქმედებებით და მოლეკულის დანარჩენი ნაწილი მხოლოდ დამხმარე, ჩონჩხის როლს თამაშობს. ჩონჩხი უზრუნველყოფს დასაკავშირებელ ფრაგმენტთა ურთიერთ და რეცეპტორის დაკავშირების (საკვანძო) უბნების მიმართ სწორ ორიენტაციას. ფრაგმენტები შეირჩევა ფუნქციონალური ჯგუფებისა და სტრუქტურული სუბერთეულების შესახებ მონაცემთა სხვადასხვა ბაზებიდან. შერჩევის პირობა შეიძლება იყოს მაგალითად, ფრაგმენტთა უნარი მონაწილეობა მიიღოს მოლეკულათშორის ურთიერთქმედებებში (მაგალითად წყალბადური ბმებში, ჰიდროფობულ ურთიერთქმედებებში და სხვა). ფრაგმენტთა განთავსების პროცედურა გულისხმობს ან მოლეკულური მექანიკის ძალოვან ველში განსაზღვრულ, ან სტრუქტურულ მონაცემთა ბაზების სტატისტიკური ანალიზის საფუძველზე შექმნილ (ე.წ. „გოდნაზე -დაფუძნებული“) პოტენციალების გამოყენებას. ფრაგმენტთა პოზიციონირების მეთოდით შერჩეული ფრაგმენტები შესაძლებელია გამოყენებულ იქნეს როგორც საწყისი წერტილები ფრაგმენტთა დაკავშირებისა, თუ თანამიმდევრული ზრდის მიდგომების რეალიზების პროცესში.

ფრაგმენტთა ერთიან ნაერთად გაერთიანება (დაკავშირება).

შეთავსებად ფრაგმენტთა ერთიან ნაერთად გამაერთიანებელი მეთოდის მიზანია ფრაგმენტები ერთიან, დაბალმოლეკულური ნაერთის სახით წარმოადგინოს და ამასთან შესაძლებელი იყოს შესაბამისი ფუნქციონალური ჯგუფების სწორად განლაგება რეცეპტორის დაკავშირების უბანში. დაკავშირების უბნის მახასიათებელ (საკვანძო) წერტილებისათვის ლიგანდის ფრაგმენტების შერჩევის თაობაზე ზემოთ იყო ლაპარაკი. შერჩეულ ფრაგმენტთა ერთ ნაერთად გაერთიანების პროცედურა კი გულისხმობს მონაცემთა ბაზებიდან შესაბამისი გამაერთიანებელი ნაერთის როგორც საფუძვლის მოძიებას. მოძიებული ნაერთი უნდა წარმოადგენდეს კოვალენტურ კავშირთა ისეთი ჩონჩხს, რომელიც უზრუნველგვყოფს გასაერთიანებელ ფრაგმენტთა საჭირო განლაგებისათვის აუცილებელი გეომეტრიით. საფუძველი შეიძლება იყოს როგორც ხისტი, ისე მოქნილი.

თანამიმდევრული შენება.

პოტენციური ლიგანდის ნაბიჯ-ნაბიჯი კონსტრუირებით ფორმირება წარმოადგენს კიდევ ერთ მიდგომას, რომლის მიზანია ან ახალი საკვანძო ნაერთების შექმნა ანაც ცნობილის ოპტიმიზირება. ამ მიდგომის თანახმად კონსტრუირების

პროცესი იწყება საფუძველი (ჩანასახოვანი) ფრაგმენტის ან ატომის განლაგებით დაკავშირების უბანში, რასაც მოყვება ამ უკანასკნელზე, როგორც საფუძველზე სხვა სტრუქტურული ელემენტების თანამიმდევრული მიერთების გზით ახალი ლიგანდის მშენებლობის პროცესი. მშენებლობის პროცესი კონტროლდება ქულათა დარიცხვის პროცედურით, რომელსაც ადგილი აქ ყოველ თანამიმდევრულ ეტაპზე მორიგი სტრუქტურული ელემენტის მიერთებისა. საბოლოო შედეგი ცხადია დამოკიდებულია საფუძველად შერჩეულ ფრაგმენტის ან ატომის სახეზე. ამ მიდგომის რეალიზებასთან დაკავშირებულ ერთ-ერთ პრობლემას შეიძლება წარმოადგენდეს მშენებლობის რომელიმე ეტაპზე განსახილველ ვარიანტთა მოზღვავეების შესაძლებლობა. პრობლემა ასევე შეიძლება წარმოიქმნეს შესაძლო ქიმიურ ნაერთთა სიმრავლესა და ქიმიური სინთეზის შეზღუდულ შესაძლებლობებს შორის წინააღმდეგობის გამოც.

შესაძლებელია გამოიყოს *de novo* მეთოდთათვის დამახასიათებელი საერთო პრობლემებიც. მაგალითად იმის გამო რომ გენერირებული ნაერთის საერთო ფორმას, როგორც წესი, განსაზღვრავს დაკავშირების სახე, არ არსებობს იმის გარანტია, რომ ლიგანდის გენერირებული კონფორმაცია არის ენერგეტიკულად ოპტიმალური. ასევე არ არსებობს იმის გარანტია, რომ წერტილოვან მუხტთა სიდიდეები, რომლებიც გამოიყენება ძალოვან ველში განსაზღვრული პოტენციალის შემთხვევაში, რეალურად არ იცვლება ლიგანდის კონსტრუირების პროცესში. და კიდევ, როგორც ზემოთ აღინიშნა, საჭიროა განისაზღვროს ქიმიური სინთეზის შესაძლებლობები.

წასაკითხად რეკომენდირებული ლიტერატურა

Del Sol A., Pazos F, Valencia A. (2002), "Automatic methods for the prediction of functionally important residues" , J. Mol. Biol.

Gobel U, Sander C, Shneider R, Valencia A , "Correlated mutations and residue contacts in proteins" v.18, pp309-17, 1994

Halperin I, Ma B, Wolfson H, Nussinov R, "Principles of docking : An overview of search algorithms and a guide to scoring functions", Proteins v. 47, pp. 409-43, 2002

Azuma et. all. " Model of the ran-RCC1 interaction using biochemical and docking experiments" J. Mol. Biol. v283, pp. 1119-30. 1999 .

Tsoka S. and Ouzounis CA. " Prediction of protein interactions: metabolic enzymes are frequently involved in gene fusion", Nat. Genet. v 26, pp 141-2, 2000

Pazos F. & Valencia A. "Similarity of phylogenetic trees as indicator of protein-protein interaction" , Protein Eng. , v14, pp.609-14, 2001.

Erwing et all. "DOCK 4.0 : Search strategies for automated molecular docking of flexible molecular databases " , J. Comput. – Aided Mol. Des. , 2001, v 15, pp. 411-28

Lawrence MC and Davis PC , "CLIX; A search algorithm for novel ligands capable of binding proteins of known three-dimensional structure", Proteins , 1992, v 12, pp. 31-41

Vakser IA " Protein docking for low-resolution structures " Protein, 1995, v 8. pp. 371-7

თავი 8

სტრუქტურულ ბიოლოგიური კვლევების დაგეგმარების სტრატეგია

22 . სტრუქტურული გენომიკა

გენომური თანამიმდევრობების შესახებ მონაცემების მოზღვავება "უდიდესი სამეცნიერო" ინიციატივების საწინდარი გახდა. მნიშვნელოვანი გარდაქმნების ზღვარზე აღმოჩნდა სტრუქტურული ბიოლოგიის შედარებით ახალგაზრდა დისციპლინაც. რენტგენო-კრისტალოგრაფიის, ხსნარის ბმრ სპექტროსკოპიის და კომპიუტერული ბიოლოგიის სპეციალისტები ცილების სტრუქტურათა სამყაროს სრულფასოვანი აღწერის მიზნით ჩაერთვნენ სტრუქტურათა დადგენის პროგრამებში. სტრუქტურული გენომიკის განვითარების დღევანდელი დონე იძლევა იმედს, რომ სტრუქტურების დიდი რაოდენობა (ათობით ათასი) ექსპერიმენტულად იქნება დადგენილი, ხოლო კიდევ უფრო მეტი (მილიონობით), მათთან ნათესაობაში მყოფი სტრუქტურები კი, მოდელური გათვლების საფუძველზე განისაზღვრება. ამასთან, განვითარდება და დაიხვეწება სტრუქტურული ბიოლოგიისათვის ისეთი მნიშვნელოვანი ტექნოლოგიები, როგორცაა ცილების ავტომატური ექსპრესია და გაწმენდა, რომოტული კრისტალიზაცია, ბმრ ან რენტგენოდიფრაქციული გაზომვები და მონაცემთა ანალიზი. სტრუქტურათა დამდგენ ეფექტურ მიდგომათა დახვეწით გამოწვეული სტრუქტურულ ინფორმაციათა მძლავრი ნაკადები უდავოდ დააჩქარებს ბიო-სამედიცინო მეცნიერებათა პროგრესს. უფრო მეტიც, მოსალოდნელია რევოლუციური ნახტომი, როგორც ცილის სტრუქტურის ევოლუციის, ისე სამგან-ზომილებიან სტრუქტურასა და ცილის ფუნქციას შორის ურთიერთ-მიმართების კანონზომიერებათა გაგების საქმეში.

სტრუქტურული გენომიკის ძირითადი ამოცანა ყველა ბუნებრივი ცილის შესახებ სტრუქტურული ინფორმაციის არსებობაა.

გენომში წარმოჩენილ გენთა პროდუქტების სრული კომპლექტისათვის სამგან-ზომილებიანი სტრუქტურების შესახებ ამომწურავი ინფორმაციის ფლობა ბუნებრივია, რომ მრავალ ახალ ინიციატივას გაუხსნის გზას. თამამად შეიძლება ითქვას, რომ უკვე დღეს, PDB-ში გაერთიანებული 50000-ზე მეტი ექსპერიმენტულად დადგენილი სტრუქტურის შესახებ ინფორმაციას უდიდესი სარგებელი მოაქვს თანამედროვე ბიოლოგიასა და ბიოქიმიასი მოღვაწე მკვლევარებისათვის. იგივე შეიძლება ითქვას კომპიუტერული კვლევების ისეთ მძლავრ მეთოდებზე, როგორცაა ცილის სტრუქტურის შედარებითი მოდელირება (ანუ ჰომოლოგიური საფუძველზე მოდელირება), დახვევის ტიპის გამოცნობა და *ab initio* წინასწარმეტყველება. მათ უდიდესი წვლილი შეაქვთ თანამედროვე ბიოლოგიის, ქიმიისა და მედიცინის განვითარებაში.

ცილების სტრუქტურათა სიმრავლე

სტრუქტურული გენომიკის ძირითადი ამოცანის გადაწყვეტის პროცესში სტრუქტურის დადგენის ექსპერიმენტული მიდგომების პარალელურად ფართოდ გამოიყენება ცილათა შედარების საფუძველზე მოდელირების (სტრუქტურული ბიოინფორმატიკის იხ. პარაგრაფი 18) მეთოდები. ასეთი რეალობა დიდწილად განპირობებულია იმ გარემოებით, რომ ცილათა სტრუქტურების სიმრავლე არაა ისე მრავალფეროვანი, როგორც გენთა სიმრავლეა. მიახლოებითი შეფასებების თანახმად ადამიანის გენომში 40000 გენია და დაახლოებით 400000 გენური პროდუქტია. დახვევის ტიპებთან დაკავშირებით ზუსტი რაოდენობრივი შეფასების გაკეთება გაძნელებულია. თუმცა, რეალური საფუძველი აქვს მოსაზრებას „კომპაქტურ გლობულარულ ცილათა სტრუქტურული დომენების, მათი დახვევის ტიპების სიმრავლე შედარებით მწირია“. დღევანდელი მონაცემებით ბუნებაში ვარაუდობენ მხოლოდ 1000-დან 5000-მდე სტაბილური დახვევის ტიპის არსებობას. დღეისათვის, სტრუქტურათა შესახებ მონაცემთა ბაზებში მხოლოდ 700 დახვევის ტიპის შესახებ მოიპოვება მონაცემები. ამ მონაცემთა ანალიზი აჩვენებს, რომ არა მხოლოდ ცილათა გარკვეულ ოჯახში გაერთიანებული თანამიმდევრობები, არამედ სხვადასხვა ოჯახის წევრებიც კი, დახვევის შედეგად იღებენ დახვევის ერთსა და იმავე ფორმას (ტიპს). ასე მაგალითად, ყველაზე პოპულარულ დახვევის ტიპს წარმოადგენს ტრიოზ ფოსფატ იზომერაზას ტიპის რვა-შრიანი *αβ* კასრი. ამასთან, რვა-შრიანი *αβ* კასრის ფორმის სტრუქტურა კიდევ 23 სხვადასხვა ოჯახში გაერთიანებულ ცილათა თანამიმდევრობების დახვევის ტიპია. აქ უპრიანია გავიხსენოთ, რომ ეუკარიოტულ გენთა უმრავლესობა აკოდირებს მრავალდომენიან ცილებს. დომენთა

საშუალო ზომაა 153 ± 87 ამინმჟავა. მრავალდომენიან ცილას ხშირად წარმოადგენენ როგორც ბურთულებიან მიმის, სადაც ბურთულები დომენებია. როგორც წესი, თითოეულ ბურთულას მიეწერება კონკრეტული ბიოქიმიური ამოცანის გადაწყვეტაზე პასუხისმგებლობა. დახვევათა ტიპების კლასიფიცირების პროცესში ჩვეულებრივ ეყრდნობიან სწორედ დომენებს (ხდება დომენთა და არა ცილათა სიმრავლის სტრუქტურული კლასიფიკაცია. იხ. პარაგრაფი 16).

ევოლუციური სურათი

გენურ თანამიმდევრობაში მომხდარი ცვლილებები, უპირველეს ყოვლისა გენური პროდუქტის ფუნქციის ანუ დომენის დონეზე ვლინდება, და ამრიგად სწორედ დომენი ექცევა ბუნებრივი შერჩევის ზეწოლის ქვეშ. ის ცვლილებები, რომლებიც დომენის სტრუქტურის დესტაბილიზირებას იწვევს ვერ უძლებს ზეწოლას. სტრუქტურისათვის უმტკივნეულო ცვლილებების უმრავლესობა, როგორც წესი, გვხვდება ცილის გლობულის ზედაპირსა ან დომენთა დამაკავშირებელ ლინკურულ უბნებში. გენთა დუბლირებამ და შემდგომმა მუტაციებმა შესაძლოა გამოიწვიოს ახალი ბიოქიმიური შესაძლებლობების გაჩენა. მუტაციები, რომლებიც პოზიტიურ ცვლილებებს უზრუნველყოფს, ვთქვათ ორგანიზმს შესძენენ უნარს გაუმკლავდეს მავნე ნივთიერებებს (მაგალითად გარემოში არსებულ ტოქსინებს), გაუძღვებს ევოლუციურ წნებს. ამით, ბუნებრივია, გაიზრდება ორგანიზმის გარემოსთან შეგუების უნარი. ასეთი წარმატებული მუტაციის მატარებელი ორგანიზმი, გამრავლება-განვითარების თვალსაზრისით შეიძენს უპირატესობას ნაკლებად წარმატებული მეზობლის მიმართ. შედეგად, ახლად წარმოქმნილი გენი და შესაბამისი ცილოვანი პროდუქტი შენარჩუნდება ევოლუციური წნეხის პირობებში.

ჰომოლოგიურობის საფუძველზე მოდელირება

კომპიუტერული ბიოლოგები, ითვალისწინებენ რა ცილათა ევოლუციური განვითარების პროცესის თავისებურებებსა და მისი მიმდინარეობის კონონზომიერებებს, ცდილობენ რენტგენოკრისტალოგრაფული ანალიზისა და ბმრ-ის საშუალებით მიღებული მონაცემების განვრცობა-გამრავლებას. ექსპერიმენტით დადგენილი თითოეული ცილის სტრუქტურა უამრავ სხვა ცილის თანამიმდევრობას (რომლის სტრუქტურაც დასადგენია) უზრუნველყოფს მეტად საჭირო სტრუქტურული ინფორმაციით. ამასთან, ჰომოლოგიურობის საფუძველზე მოდელირება შესაძლებელია მხოლოდ ისეთი ცილის თანამიმდევრობისათვის, რომლის ახლო ნათესავის (ჰომოლოგის) სტრუქტურული ნიმუში (ექსპერიმენტულად დადგენილი) ცნობილია. დღეისათვის ჰომოლოგიურობის საფუძველზე მოდელირების შესაძლებლობები წვდება *Saccharomeces cerevisiae*-ს გენომის ღია წასაკითხ ჩარჩოთა მხოლოდ 50%-ს. თუ გავითვალისწინებთ, რომ

უმრავლეს შემთხვევაში შესაძლებელია ცილის მხოლოდ გარკვეული ნაწილის მოდელირება, სურათი კიდევ უფრო მიმდებარე (Saccharomices cerevisiae-ის ცილათა სრული ამინმჟავური შემადგენლობის მხოლოდ 18%-ის მოდელირება შესაძლებელი). 29 გენომის ნუკლეოტიდური თანამიმდევრობის ანალიზმა (Liu & Rost, 2001) აჩვენა, რომ PDB-ში გენომურ ცილათა მხოლოდ 20-40%-ს მოეძებნება ჰომოლოგი. ამასთან ამინმჟავათა წილი, რომელთათვისაც შესაძლებელია ჰომოლოგიურობის საფუძველზე სტრუქტურის წინასწარმეტყველება 20-30%-ია.

სტრუქტურული ბიოლოგები ცილათა სტრუქტურული ტიპების სრული სიმრავლის მონაცემების მოპოვებით უდიდეს წვლილს შეიტანენ როგორც თანამედროვე ბიოლოგიის ფუნდამენტალური პრობლემების გადაწყვეტის, ისე ბიოსამედიცინო კვლევების წარმატებით განხორციელების საქმეში. თუ ცილის თითოეული ოჯახის შიგნით ერთი წარმომადგენლისათვის მაინც მოხერხდება სტრუქტურის ექსპერიმენტული გზით დადგენა, მაშინ 30%-ანი იდენტურობის ინტერვალში მოხვედრილი ყველა სხვა ცილისათვის იქმნება ჰომოლოგიურობის საფუძველზე მოდელირების შესაძლებლობა. ჩატარებულმა ანალიზმა (Viktup et al., 2001) აჩვენა, რომ ინფორმაცია მხოლოდ 16 000, ზედმიწევნით შერჩეული ცილოვანი სტრუქტურის შესახებ საკმარისია, რათა ჰომოლოგიურობის საფუძველზე მოდელირების საშუალებით განისაზღვროს ბუნებაში არსებულ ცილათა 90%-ის სტრუქტურა.

ჰომოლოგიურობის საფუძველზე მოდელირების მეთოდი შეიძლება პრიორიტეტულად ჩაითვალოს თანამიმდევრობათა შედარების საფუძველზე ურთიერთ-შესაბამისობის განმსაზღვრელ მეთოდებთან მიმართებაში, რადგანაც ის იძლევა ატომურ კოორდინატებს და შესაბამისად, რენტგენულ ან ბმრ სტრუქტურებთან უშუალო შედარების შესაძლებლობას. ცნობილ სტრუქტურებთან შედარების (blind test) საფუძველზე განსაზღვრული სიზუსტის მიხედვით მოდელები იყოფა სამ კალასად: მოდელები, რომლებიც იწარმოებს საკვლევთან 50%-ზე მეტი თანამიმდევრობათშორისი იდენტურობის მქონე ნიმუშის საფუძველზე. რენტგენოკრისტალოგრაფიულ სტრუქტურასთან შედარების თვალსაზრისით ისინი შესაბამეა 3 Å-ან გარჩევითობას, ხოლო ბმრ-ული სტრუქტურებთან შედარების თვალსაზრისით - საშუალო-გარჩევითობის ამონახსნს; საკვლევთან ნაკლებად მსგავსი ნიმუშების გამოყენებით (30-50% იდენტურობა) მიღებული მოდელები. მათი α -ნახშირბადის ატომების 85%-ზე მეტისათვის გადახრა სწორი პოზიციიდან არ აღემატება 3.5Å-ს; და მოდელები, როცა თანამიმდევრობათშორისი იდენტურობა < 30%-ზე. ისინი შეიძლება შეიცავდნენ მნიშვნელოვან უზუსტობებს. თუმცა ცილის დახვევის ტიპის იდენტიფიცირების მიზნით მაინც ახდენენ ასეთ მოდელირებას.

ცილოვან სტრუქტურათა შესახებ მონაცემთა ბაზები

PDB წარმოადგენს ექსპერიმენტულად მიღებული სტრუქტურების (რენტგენოკრისტალოგრაფული ანალიზი და ხსნარში ზმრ) ძირითად საცავს. მასში შესაძლებელია მოვიძიოთ 50 000-მდე პეპტიდის, ცილის თუ სხვადასხვა კომპლექსის სტრუქტურული მონაცემები (შემადგენელ ატომთა კოორდინატები). როგორც წესი, PDB-ში არ ინახება მოდელირებით წარმოებულ სტრუქტურათა შესახებ მონაცემები (მართალია ვხვდებით გამონაკლისებსაც).

შედარებითი მოდელირების შედეგად წარმოებულ სტრუქტურათა მონაცემებით სარგებლობის შესაძლებლობებს გვათავზობს ორი საჯარო მოხმარების სერვერი: MODBASE (<http://pipe.rockefeller.edu/modbase-cgi/index.cgi>) და SWISS-MODEL (http://www.expasy.org/swissmod/SM_3DCrunch_Search.html). ამასთან მკვლევარმა შეიძლება მიიღოს საკვლევი თანამიმდევრობის ერთი ან რამდენიმე შესაძლო სტრუქტურული მოდელი, თუ მოთხოვნას (თანამიმდევრობის სახით) წარადგენს შემდეგ სერვერებზე: MODWEB (<http://pipe.rockefeller.edu/mwtest-cgi/main.cgi>);

SWISS-MODEL (<http://www.expasy.ch/swissmod>);

FAMS (<http://physchem.pharma.kitasatou.ac.jp/FAMS/fams.html>);

3_DJIGSAW (<http://www.bmm.icnet.uk/servers/3djigsaw/>);

CPHmodels (<http://www.cbs.dtu.dk/services/CPHmodels/index.html>)

San Diego Supercomputing Center (<http://cl.sdsc.edu/hm.html>).

გარდა ამისა მკვლევარებს აქვთ შესაძლებლობა წარადგინონ MODWEB და SWISS-MODEL სერვერებზე ატომთა კოორდინატების ფაილი და მიიღონ ჰომოლოგიურობის საფუძველზე წინასწარმეტყველების მიდგომით წარმოებული მოდელი ყველა იმ თანამიმდევრობისათვის, რომელიც იმყოფება მოდელირების შესაძლებლობის ზღვრული რადიუსის ფარგლებში. უკანასკნელი შესაძლებლობის შეფასება განსაკუთრებით მნიშვნელოვანია იმ თვალსაზრისით, რომ გაირკვეს, თუ როგორია სტრუქტურული გენომიკის ცენტრის მიერ ექსპერიმენტულად დადგენილი თითოეული ახალი სტრუქტურის წვლილი სტრუქტურული გენომიკის ძირითადი ამოცანის გადაწყვეტის საქმეში.

სტრატეგია

სანამ ვილაპარაკებდეთ იმაზე, თუ როგორ იგეგმება სტრუქტურული ბიოლოგიის საპილოტე პროგრამები და როგორია ძირითადი ამოცანის - „უზრუნველყოთ ბუნებაში არსებული თითოეული ცილის თანამიმდევრობა სტრუქტურული ინფორმაციით“ - გადაწყვეტის პერსპექტივები, საჭიროა შევაფასოთ, თუ რამდენად ადეკვატურ შედეგებამდე მიგვიყვანს მასშტაბური სტრუქტურული გენომიკის პროექტების განხორციელება. ის რომ, არსებული ექსპერიმენტული მიდგომებით ამოცანის გადაწყვეტისთვის ძალისხმევის გაწევა აუცილებელია,

ეკვს არ იწვევს. განხილვის საგანია მხოლოდ ის, თუ როგორი უნდა იყოს ცალკეული ცილის სტრუქტურის დასადგენად საჭირო დრო და დანახარჯები. რენტგენოკრისტალოგრაფული და ხსნარში ბმრ ექსპერიმენტებით ყოველწლიურად ხორციელდება 2000-დან 3000-მდე ახალი ცილის სტრუქტურის ანოტაცია და PDB-ში განთავსება (დეპონირება). PDB-ში 1994 წელს წარდგენილ სტრუქტურათა ანალიზმა (Brenner, Chothia, & Hubbard, 1997) აჩვენა: სტრუქტურა-გამიფრულ ახალ ცილათა 70%-ისათვის PDB-ში მოიძებნა თითქმის ისეთივე თანამიმდევრობის ცილა, 21%-ს აღმოაჩნდა თანამიმდევრობის თვალსაზრისით ახლო ნათესავი, და მხოლოდ 9%-სათვის არ მოიძებნა PDB-ში ცხადად გამოხატული ჰომოლოგი. ახალ სტრუქტურათა დეპონირების სიჩქარის მაქსიმალური მნიშვნელობის (3000 სტრუქტურა წელიწადში) პირობებშიც კი, უახლოეს მომავალში სტრუქტურულ-ბიოლოგიური კვლევები ვერ შეძლებს მოახდინოს რაღაც მნიშვნელოვანი გარღვევა ჰომოლოგიურობის საფუძველზე მოდელირებით გამიფრულ სტრუქტურათა რიცხვის მნიშვნელოვანი ზრდის თვალსაზრისით. შესაბამისად, გამოიკვეთა კვლევის მიმართულება, რომელიც ითვალისწინებს ცილის ფუნქციაში გარკვევას, მისი არქიტექტურის არცოდნის პირობებშიც კი.

იმედია, ძალისხმევა მაღალი წარმადობის მეთოდების შემუშავების (აქცენტი კეთდება სტრუქტურის დამდგენი ექსპერიმენტის ჩატარებისა და მონაცემთა დამუშავების პროცესის ავტომატიზაციაზე) მიმართულებით თავის შედეგს გამოიღებს და შეამცირებს სტრუქტურის გამიფრის როგორც დროს, ისე დანახარჯებს. თუმცა დღეს, ცალკეული სტრუქტურის დადგენაზე გაწეული დანახარჯი (\$100000) აღარაა პრინციპულად მიუღებელი. თუ ზემოთ მოყვანილი შეფასებები რეალურია, მაშინ საჭირო ნიმუშთა მინიმალური რაოდენობისათვის(16000) სრული სტრუქტურული ინფორმაციის მიღებაზე გასაწევი დანახარჯები დაახლოებით 1,6 მილიარდს შეადგენს. გარკვეული ეკონომიის პირობებში მოსალოდნელია დანახარჯების მნიშვნელოვანი შემცირება.

სტრუქტურული გენომიკის საპილოტე პროგრამათა ამოცანები.

ჩრდილო და სამხრეთ ამერიკაში (Terwilliger, 2000), ევროპაში (Heinemann, 2000) და აზიაში (Yokoyama et. al., 2000) მიმდინარე საპილოტე კვლევების ფარგლებში ხორციელდება სტრუქტურის დამდგენი მაღალი წარმადობის ექსპერიმენტები. საპილოტე პროგრამის განხორციელების მეთოდებად, როგორც წესი, გამოიყენებენ ან რენტგენოკრისტალოგრაფიას, ან ხსნარში ბმრ სპექტროსკოპიას. ამასთან, საკვლევ ცილათა ჩამონათვალი შეირჩევა არქების, ელბაქტერიებისა და ეუკარიოტების გენომური თანამიმდევრობების ანალიზის საფუძველზე. წამყვან საპილოტე კვლევებს კოორდინირებულად აფინანსებს სხვადასხვა სახელმწიფო და საზოგადო ორგანიზაციები. მათ შორის ძირითად დონორებია: National

NIH-ის მიერ დაფინანსებული ჯგუფების პასუხისმგებლობით ხდება მაღალი წარმადობის ტექნოლოგიების განვითარება-დანერგვა. ამ ტექნოლოგიებში იგულისხმება უამრავი მიდგომა, გენის სეკვენირებით დაწყებული და სტრუქტურის დადგენით დამთავრებული. თუ გავანალიზებთ აშშ-ში მიმდინარე საპილოტე კვლევებს, აღმოჩნდება, რომ კვლევების ობიექტთა ნაწილს შეადგენს ჰომოლოგიურ თანამიმდევრობათა ოჯახების წარმომადგენლები. ამასთან, სტრუქტურის დასადგენად შერჩეულ სამიზნე ცილებს მოიძიებენ სხვადასხვა ბაქტერიულ და ეუკარიოტულ გენომებში. სხვა სტრატეგია მდგომარეობს იმაში, რომ სამიზნედ (საკვლევადა) შერჩევა ცალკეული მოდელოური ორგანიზმების, მათ შორის ადამიანის პათოგენთა გენური პროდუქტები. ოჯახებზე დაფუძნებული სტრატეგიის შემთხვევაში, ხშირად აქცენტი კეთდება ისეთ სამიზნეებზე, რომლებიც მონაწილეობენ მედიცინის თვალსაზრისით მნიშვნელოვან ბიოლოგიურ პროცესებში (მაგალითად სიგნალთა ტრანსდუქციაში, ანტიბიოტიკებისადმი რეზისტულობაში, კანცეროგენეზში).

NIGMS-ის მიერ დაფინანსებული ცენტრებში იქმნება ლაბორატორიული ინფორმაციული სისტემები და საჯარო ვებ-გვერდები. ასეთი მიდგომა გამჭვირვალეს ხდის როგორც სამიზნის შერჩევის პროცედურას, ისე სტრუქტურის დადგენის პროცესში მიღწეულ წარმატებებს. გარდა ამისა, აშშ მთავრობა აყენებს მოთხოვნას, რათა მოპოვებული სტრუქტურული ინფორმაცია განთავსებულ იქნეს PDB-ში. სტრუქტურული გენომიკის პროექტის რეალიზებისას PDB-ში ხდება ისეთი შუალედური შედეგების დაფიქსირებაც, რომელიც, მაგალითად, ეხება ცილის ექსპრესიას, გასუფთავებას და ბიოფიზიკურ დახასიათებას. NIGMS-ის მიერ წარმოებს პილოტური კვლევების პროცესში შერჩეული თუ ფორმირებული რეაგენტების, მაგალითად ექსპრესიის პროცესში გამოყენებულ პლაზმიდ/კლონების, გასუფთავებული ცილების ან სულაც კრისტალების არქივირება და თავისუფალი მოხმარებისათვის ხელსაყრელ მდგომარეობაში მოყვანა.

NIGMS-ის მიერ დაფინანსებულ საპილოტე პროგრამათა ტიპური მეთოდოლოგიური მიდგომის ზოგადი სქემა შეიძლება წარმოვადგინოთ ასე: 1) საკვლევი ობიექტის (სამიზნეს) შერჩევა; 2) გენომური ან კლონ-დნმ-ური მაკოდირებელი თანამიმდევრობის PCR-ით ამპლიფიკაცია; 3) მაკოდირებელი თანამიმდევრობის შესაბამის საექსპრესიო ვექტორში კლონირება; 4) მაკოდირებელი უბნის სწორად ამპლიფიცირებაში დასარწმუნებლად, კლონირებული გენის სეკვენირება; 5) ცილის ექსპრესია; 6) ექსპრესირებული ცილის იდენტიფიკაცია და მისი დახასიათება, რათა მომზადდეს პირობები მისი კრისტალიზაციისათვის; 7) საკმაო რაოდენობისა და სისუფთავის ცილის მიღება; 8) კრისტალიზაციის პირობების შერჩევა; 9) რენტგენული დიფრაქციის ექსპერიმენტის ფარგლებში

გაზომვები; 10) ექსპერიმენტის მონაცემებზე დაყრდნობით სტრუქტურის დადგენა და დახვეწა; 11) ახალი სტრუქტურული ნიშნით ჰომოლოგიურობის საფუძველზე მოდელირება; 12) ახლადდადგენილი სტრუქტურისა და მასზე დაფუძნებით წარმოებული მოდელების ანალიზის საფუძველზე ფუნქციის შესახებ ჰიპოტეზების შემოთავაზება.

გავცანით რა სტრუქტურული გენომიკის ტიპური პროექტის სქემას, შეგვიძლია შევნიშნოთ, რომ ჰომოლოგიურობის საფუძველზე მოდელირების მეთოდების, ფუნქციისა და ფუნქციის რეალიზების მექანიზმების ახსნის მიზნით სივრცული მოწყობის ანალიზისათვის განკუთვნილი მეთოდების, ანუ სტრუქტურული ბიოინფორმატიკის მეთოდების როლი სტრუქტურული გენომიკის ამოცანის წარმატებულად გადაწყვეტის საქმეში მნიშვნელოვანია.

შორეული პერსპექტივები, შესაძლო პრობლემები

მიუხედავად იმისა, რომ საპილოტე პროგრამების პირველი შედეგები მეტად იმედისმომცემია, უნდა ითქვას, რომ ცილოვან სტრუქტურათა სივრცეში მოიძებნებიან ისეთი უბნები, სადაც რენტგენოსტრუქტურული და ბმრ მეთოდებით სტრუქტურული გენომიკის ამოცანის გადაწყვეტა გაძნელებულია. მაგალითად, კვლავ პრობლემატურად რჩება მემბრანული ცილების კრისტალიზაციის საკითხი, თუმცა გარკვეული წარმატებები ცილის კრისტალიზაციის ტექნოლოგიების განვითარების საქმეში სახეუა. პრობლემებს ვაწყდებით ასევე ბმრ სპექტროსკოპიით ცილის სტრუქტურის დადგენის პროცესშიც, რომელთა მოგვარების საფუძველს იძლევა ახალი (TROSY) ტექნოლოგიების განვითარება. არიან ცილები, რომლებიც მხოლოდ გარკვეული კომპლექსების ფარგლებში წარმოადგენენ სტაბილურ მოლეკულებს და შესაბამისად, მათი სტრუქტურის დამოუკიდებლად შესწავლა შეუძლებელია. ასეთი რეალობა კითხვის ნიშნის ქვეშ აყენებს სტრუქტურული გენომიკის პრეტენზიას იყოს გენომური მასშტაბის პრობლემათა გადაწყვეტის ერთადერთი საშუალება. ზემოთ მითითებული მიზეზების გამო ცილების მთელი რიგი სტრუქტურული გენომიკის კვლევათა არეალს მიღმა რჩება.

ამასთან, სტრუქტურული გენომიკის პროექტები, როგორც წესი, ვერ უზრუნველგვყოფს ინფორმაციით ცილის ნატიფი სტრუქტურის შესახებ. ასე რომ, ბიოქიმიური მექანიზმების, ცილა-ლიგანდურ კომპლექსთა სტრუქტურების, მუტაციათა სტრუქტურაზე გავლენის, ცილა-ცილა ურთიერთქმედებების დეტალური შესწავლა ვერ მოხერხდება სტრუქტურული გენომიკის ფარგლებში. შესაბამისად, დღის წესრიგში დგება მთელ რიგ ცილათა ნატიფი სტრუქტურების დამოუკიდებლად, მაღალტექნოლოგიური მეთოდებით დადგენის აუცილებლობა.

მიუხედავად ზემოთ მითითებული პრობლემებისა, სტრუქტურული გენომიკის პროექტები იძლევა იმედს, რომ მალე უზრუნველყოფილი ვიქნებით ძალზე დიდი რაოდენობა ცილის დომენტა ექსპერიმენტური სტრუქტურებით. რაც თავის მხრივ საწინდარი გახდება მრავალი ახალი მოდელის გენერირებისა. დადგება დრო, როცა მოდელირება მოიცავს ამინმჟავური თანამიმდევრობების მთელ სივრცეს. სტრუქტურული გენომიკის პროექტების რეალიზების შედეგად დაგროვილი ცოდნა, რომელიც თანამიმდევრობათა სრულ სიმრავლეს უზრუნველყოფს სამგანზომილებიანი სტრუქტურებით, საშუალებას მისცემს თანამედროვე ბიოლოგებს ჩასწვდნენ ბიოქიმიურ და უჯრედულ ფუნქციათა სიღრმეებში. დაგროვილ სტრუქტურათა და ჰომოლოგიურობაზე დაფუძნებულ მოდელთა განუსაზღვრელი სიმდიდრე უდავოდ გადაიქცევა თანამედროვე ბიოლოგურ კვლევათა ფართო (თუ არა სრული) სპექტრისათვის ფასდაუდებელ იარაღად და შესაბამისად, სულ უფრო მეტ ბიოლოგს მოუწევს ამ სიმდიდრით სარგებლობა.

წასაკითხად რეკომენდირებული ლიტერატურა:

- Liu J, Rost B (2001) "Comparing function and structure between entire proteomes." *Protein Sci* , 10:1970-9.
- Vitkup D, Melamud E, Moults J, Sander C (2001) "Completeness in structural genomics." *Nat. Struct. Biol*, 8:559-66.
- Brenner SE, Chothia C, Hubbard T (1997) "Population statistics of protein structures: lessons from structural classifications." *Curr Opin Struct Biol*, 7:369-76.
- Heinemann U (2000) "Structural genomics in Europe: Slow start, strong finish?" *Nat Struct Biol* , 7:940-2.
- Terwilliger TC (2000) "Structural genomics in North America." *Nat Struct Biol*, 7:935-9.
- Yokoyama S, Hirota H, Kigawa T, Yabuki T, Shirouzu M, Terada T, Ito Y, Matsuo Y, Kuroda Y, Nishimura Y, Kyogoku Y, Miki K, Masui R, Kuramitsu S (2000) "Structural genomics projects in Japan." *Nat Struct Biol* ,7:943-5.

საგნობრივი სარჩევი

- ab initio წინასწარმეტყველება, 247
BLAST, 202
BLOSUM მატრიცები, 175
CATH, 224, 227
CLUSTALW, 265
DALI, 217
DDBJ ფორმატი, 93
de novo პროექტირება, 292
DSSP, 230
EMBL ფორმატი, 91
FASTA, 205
FSSP, 227
GenBank, 83
GenBank ფორმატი, 92
PAM მატრიცები, 173, 181
PDB ფორმატი, 107
PSI-BLAST, 257, 265
SCOP, 222
SSAP, 218
Swiss-Prot, 95
TrEMBL, 97
454 სეკვენირება, 27
ანოტირება, 22, 98
ასამბლირება, 20
გენომური პროექტები, 19
დახვევის მოდელები ცილის, 71, 72
დახვევის ტიპის გამოცნობა, 256
დაფანტვით სეკვენირება, 20
დისტანცია თანამიმდევრობებს შორის, 185
დნმ-მიკრომატრიცები, 16
დნმ-ს სტრუქტურა, 39
დოკინგი, 290
კლასტერიზაცია, 188
კოევილუცია, 275
ლოკალური ურთიერთსწორება, 166
მეოთხეული სტრუქტურა ცილის, 67
მეორეული სტრუქტურა ცილის, 60
მესამეული სტრუქტურა 65
მორგებით წინასწარმეტყველება, 258
მსგავსება თანამიმდევრობათა, 183
მრავალწევრა ურთიერთსწორება, 261
ნახევრადგლობალური ურთიერთსწორება, 159
ნეირონული ქსელის მეთოდი, 235
ოჯახ-დამოკიდებული კონსერვაცია, 274
პოლიპეპტიდური ჯაჭვი, 55
რნმ-ს სტრუქტურა, 43, 220
რნმ-ს სტრუქტურის წინასწარმეტყველება, 243
სტრუქტურათა სუპერპოზიციონირება, 216
სტრუქტურათა ურთიერთსწორება, 217
სტრუქტურული ბიოლოგია, 17
სტრუქტურული გენომიკა ამოცანა, 298
ურთიერთქმედ პარტნიორთა წინასწარმეტყველება, 278
ურთიერთქმედ უბანთა წინასწარმეტყველება, 276
ფარმაკოგენომიკა, 29
ფილოგენეტიკური წარმოდგენა, 188
ფილოგენეტიკური ხე, 189
შაპერონები, 77
ჩანაცვლების უმარტივესი მატრიცები, 171
ჩოუ-ფასმანის მეთოდი, 232
წამალთა პროექტირება, 286
ხის ფორმირების მეთოდები, 191
ჰიბრიდიზაციით სეკვენირება, 28
ჰომოლოგიურობა ცილების, 187
ჰომოლოგიურობის საფუძველზე მოდელირება, 254
ჯაჭვის ტერმინირებით სეკვენირება, 26