Структурная Биоинформатика

Лекция 12. Макромолекулярные взаимодействия. Возможность предсказания.

Головин А.В. [1]

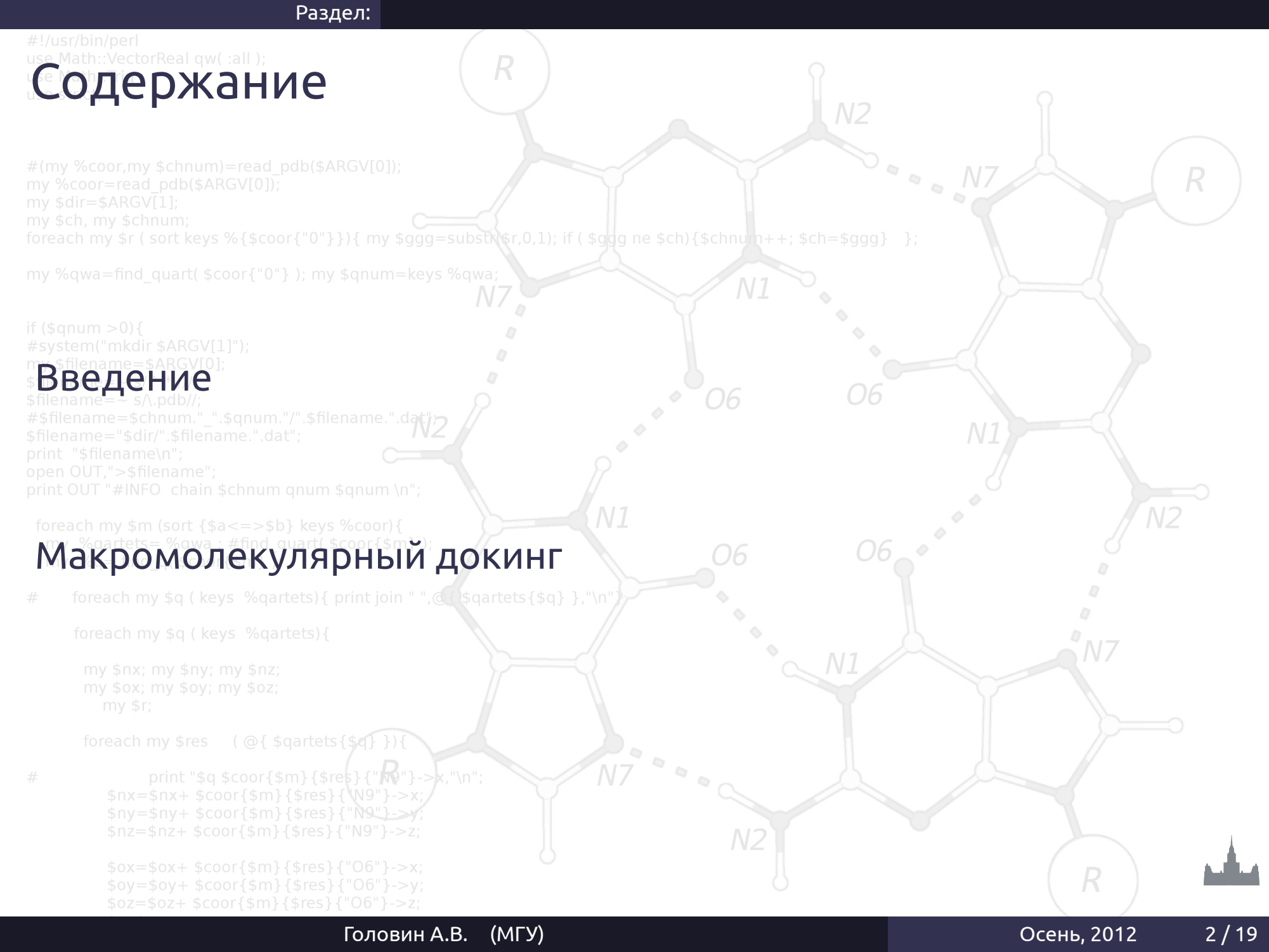
[1] МГУ им М.В. Ломоносова, Факультет Биоинженерии и Биоинформатики

Москва, 2012

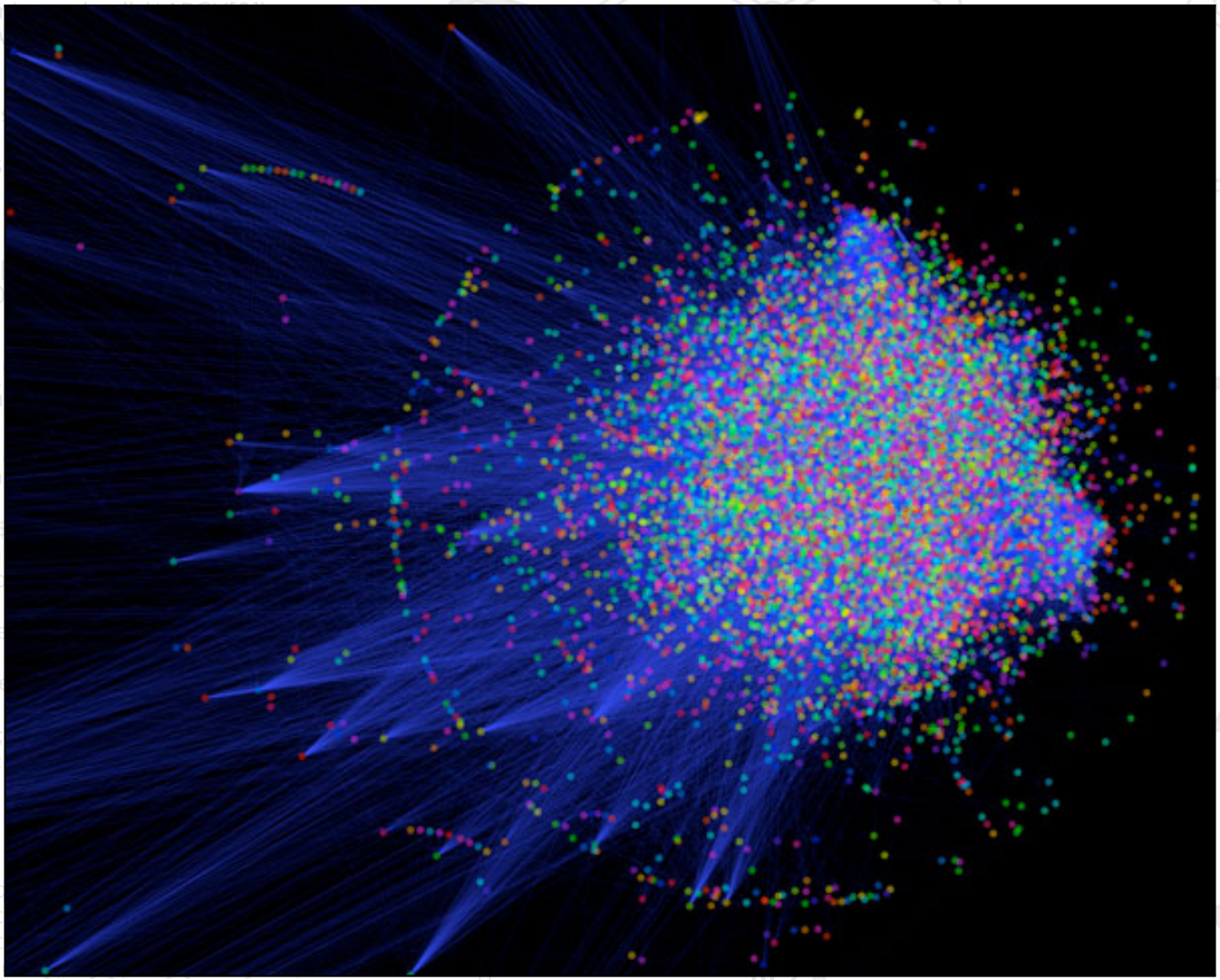
Содержание

Введение

Макромолекулярный докинг



Human "Interactome"



Способы предсказания белок-белковых взаимодействий

Взаимодействующие белки возможно ко-эволюционируют.

- Филогенетический профайлинг.
  Поиск пар белковых семейств среди широкого ряда видов.
  Появление и исчезновение пар семейств возможно указывает на взаимодействие.
- Предсказание на основе подобия филогенетических деревьев.
- Методы на основе классификации.
- Поиск гомологичных мест контакта.
- Ассоциативные методы. Это поиск характеристических последовательностей на основе профилей и мотивов.

Способы предсказания белок-белковых взаимодействий

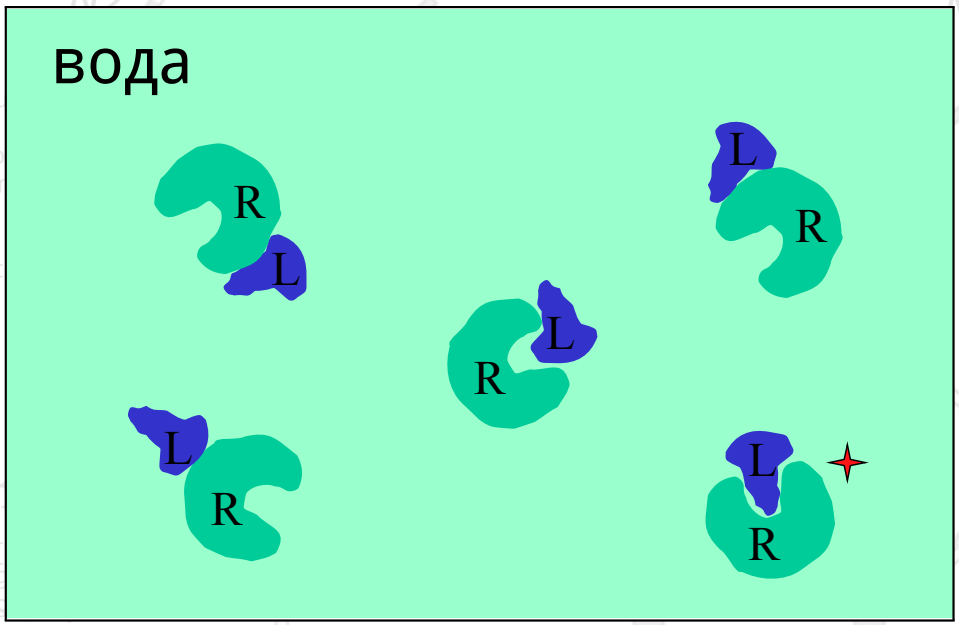
- Идентификация структурных паттернов на основе известных структурных данных.
  Построение библиотеки и сканирование по ней.
- Методы Байеса для анализа экспериментальных результатов с значимым уровнем шума.
- Методы исключения доменных пар.
- Моделирование структуры комплекса на основе известной структуры и оценка его качества.
- Макромолекулярный докинг.

Базы данных

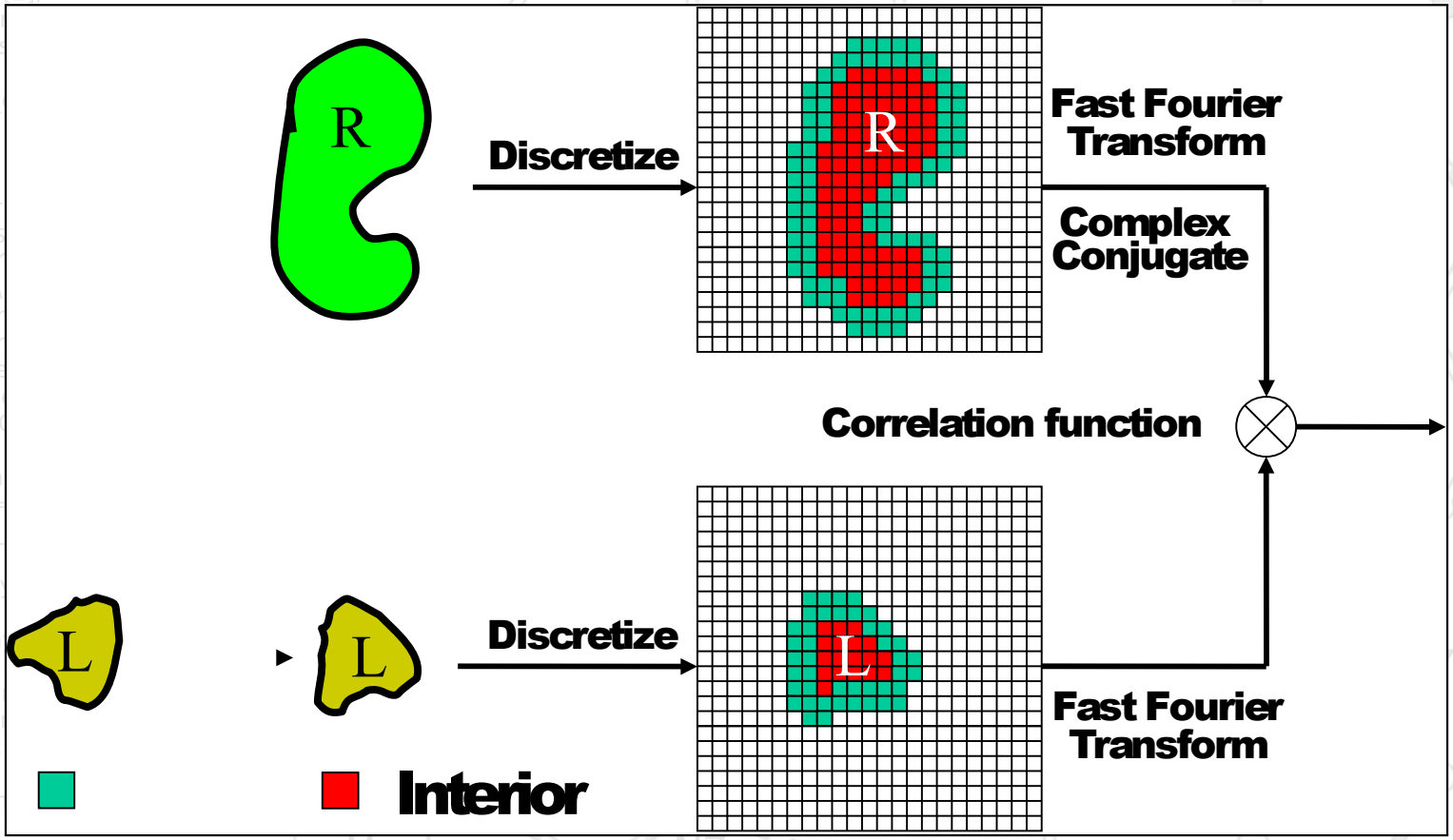
- **String** - база данных экспериментальных и предсказанных взаимодействий; отличная графика; http://string-db.org/
- **IntAct** - база данных на основе литературных данных или прямая информация от авторов. http://www.ebi.ac.uk/intact/
- **iHOP** - Информация слинкованая с другими белками. Построена на основе литературных данных. Представление в виде кусочков текста. http://www.ihop-net.org/
- **BioGRID** - Источники: литература и результаты high-throughput экспериментов; http://thebiogrid.org/
- **MIPS** Mammalian Protein-Protein Interaction Database, не работает :). http://mips.helmholtz-muenchen.de/proj/ppi

Поиск наименьшего ΔG

- Суть метода основывается на поиске соответствия поверхностей для достижения максимальной поверхности контакта.
- После нахождения возможных конфигураций происходит ранжирование.



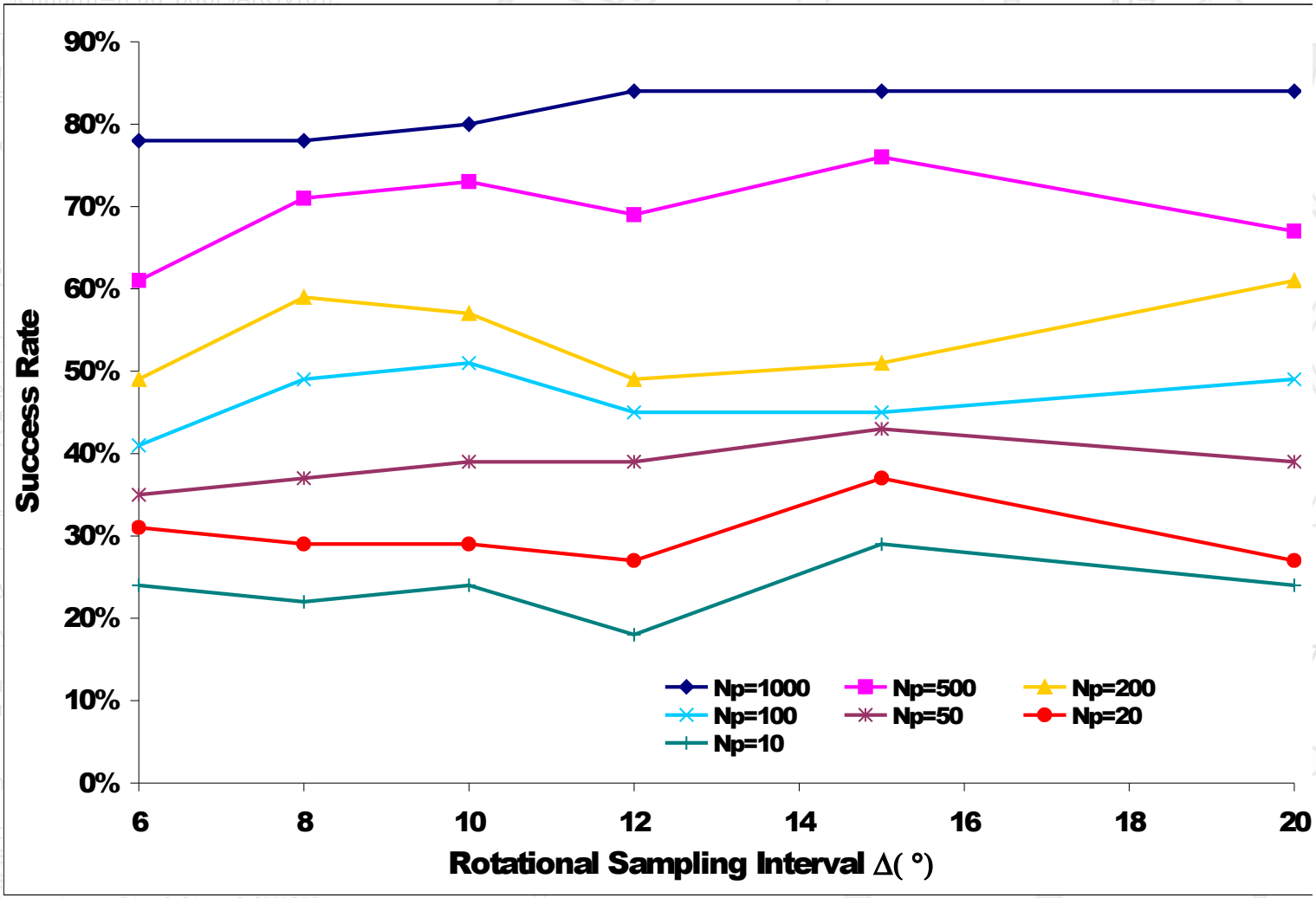
Белковый докинг с использованием FFT



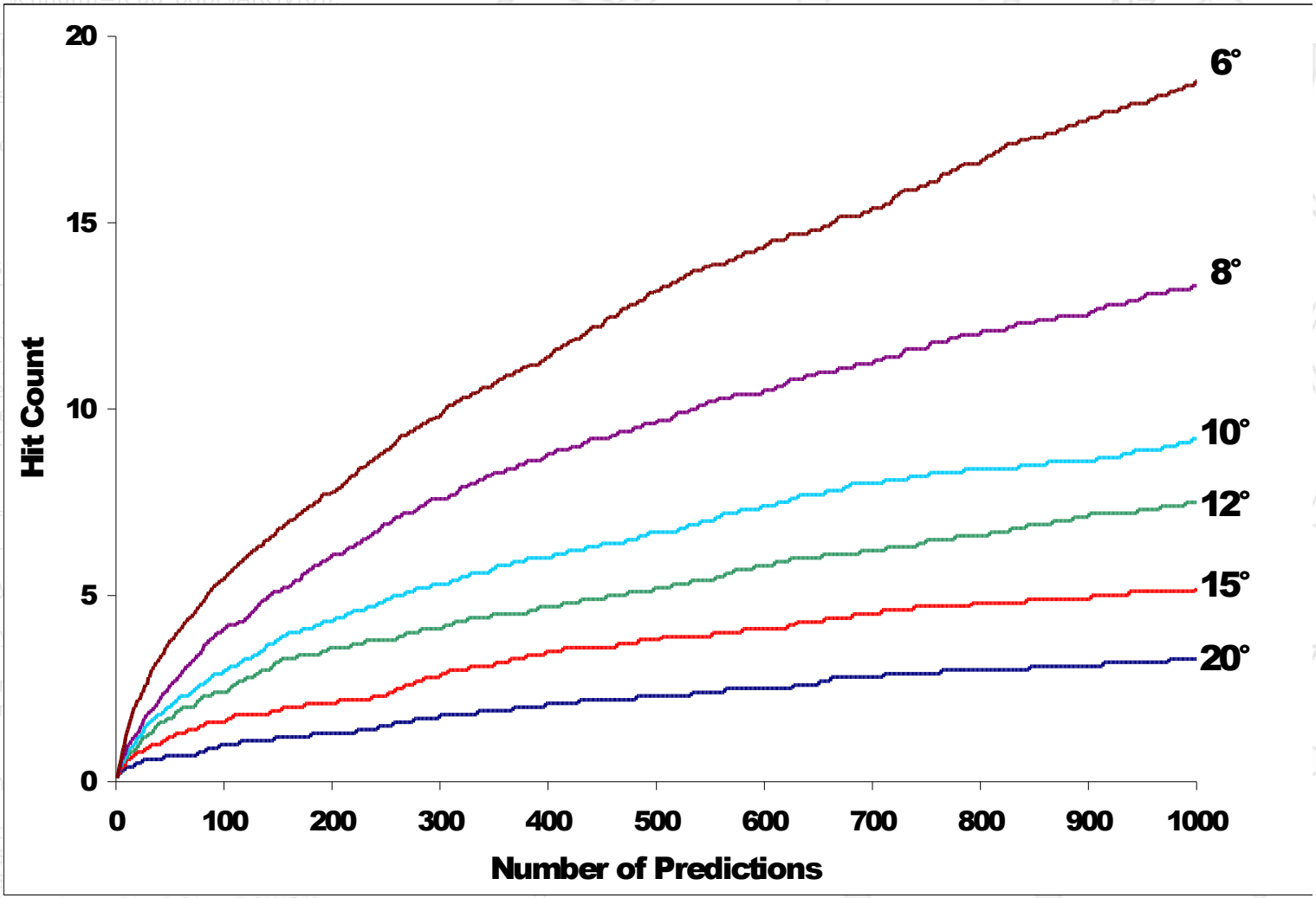
Оценка производительности

- Success Rate: для данного количества предсказаний (N_p), это процент структур для которых был найден как минимум один удачный результат
- Hit Count: среднее количество хитов при данном значении N_p.

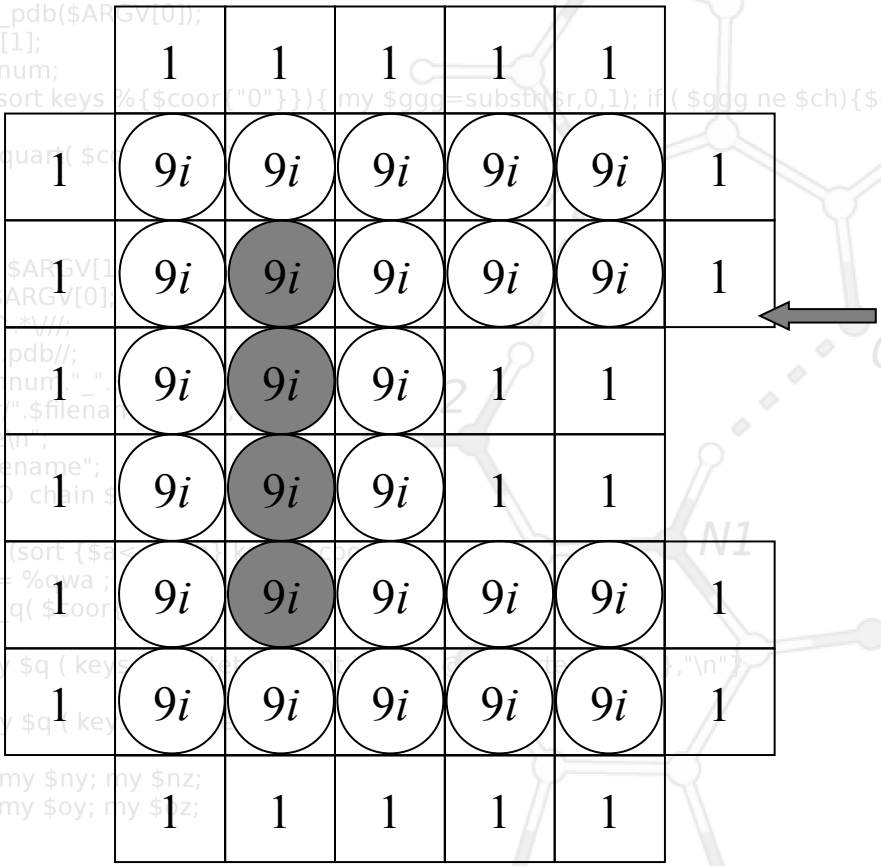
Зависимость Success Rate от шага вращения



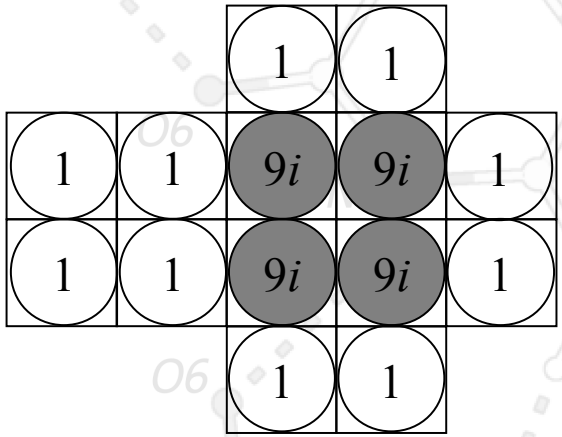
Зависимость Hint count от шага вращения



Решеточная комплементарность поверхности

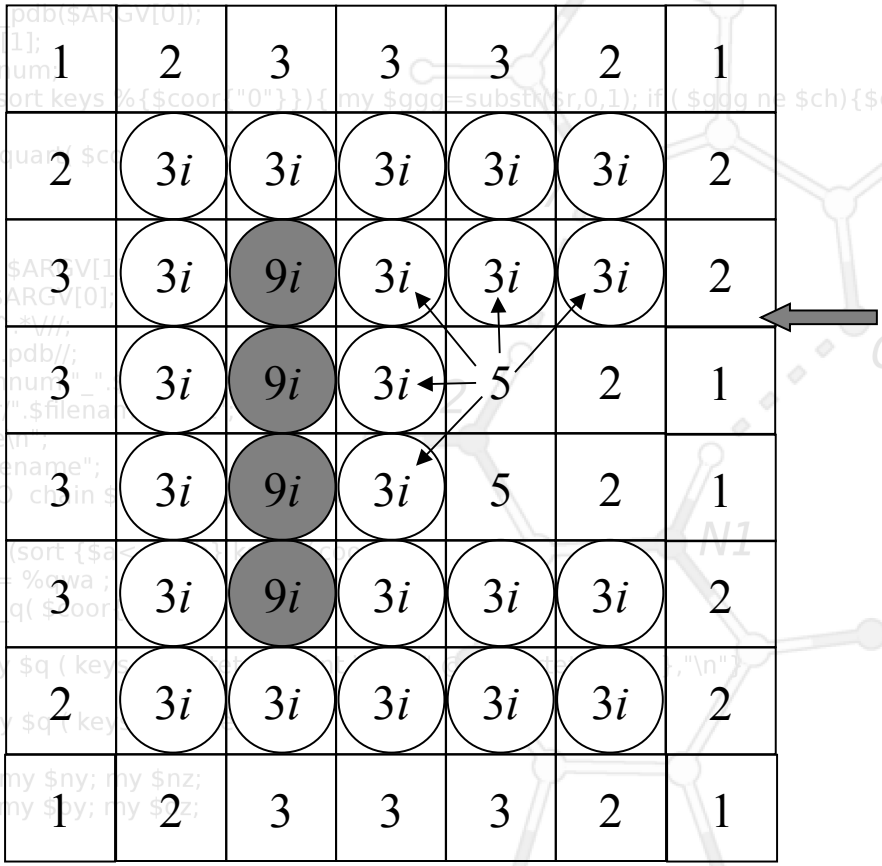


R_{GSC}

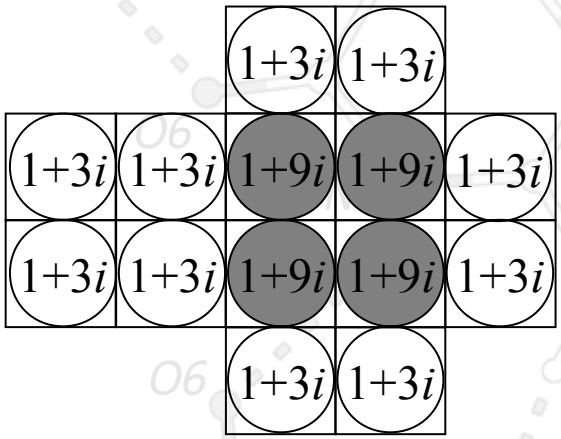


L_{GSC}

Парная комплементарность поверхности

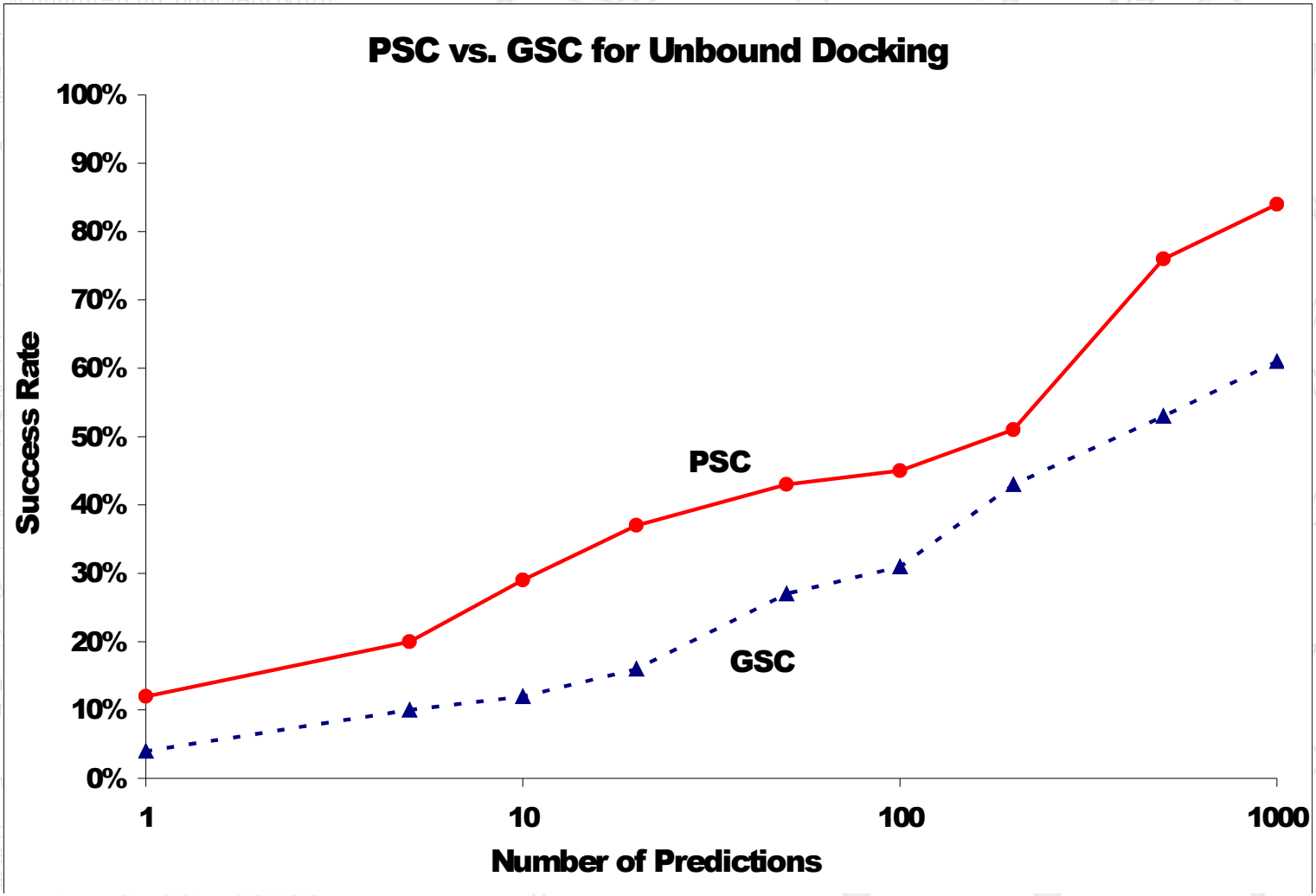


R_{PSC}



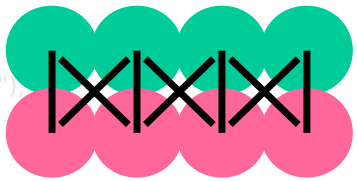
L_{PSC}

PSC vs. GSC и Success Rate

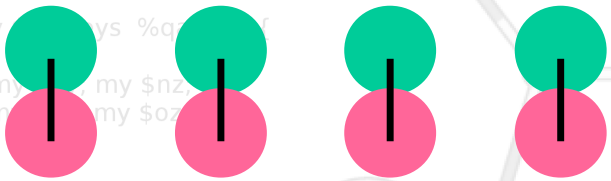


Почему так?

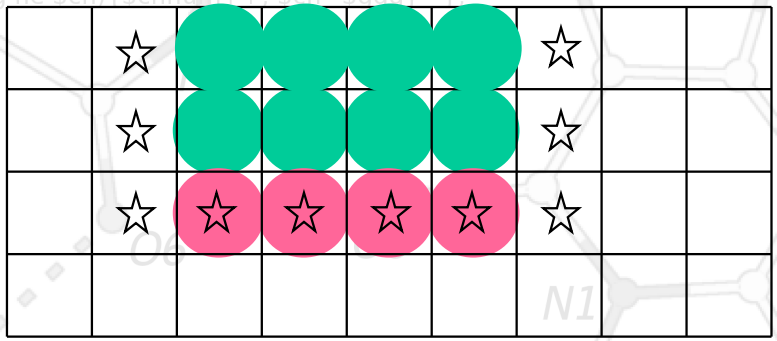
A



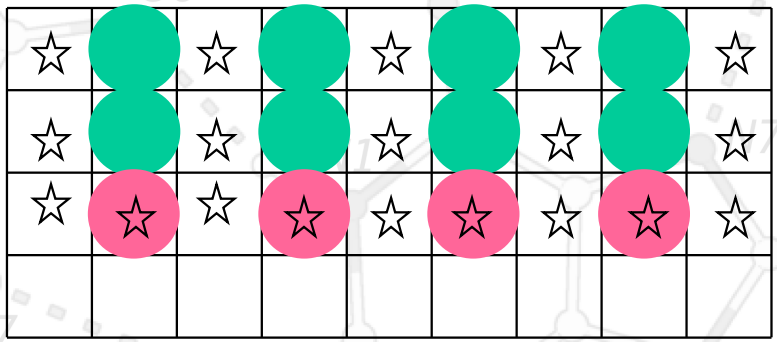
B



C



D



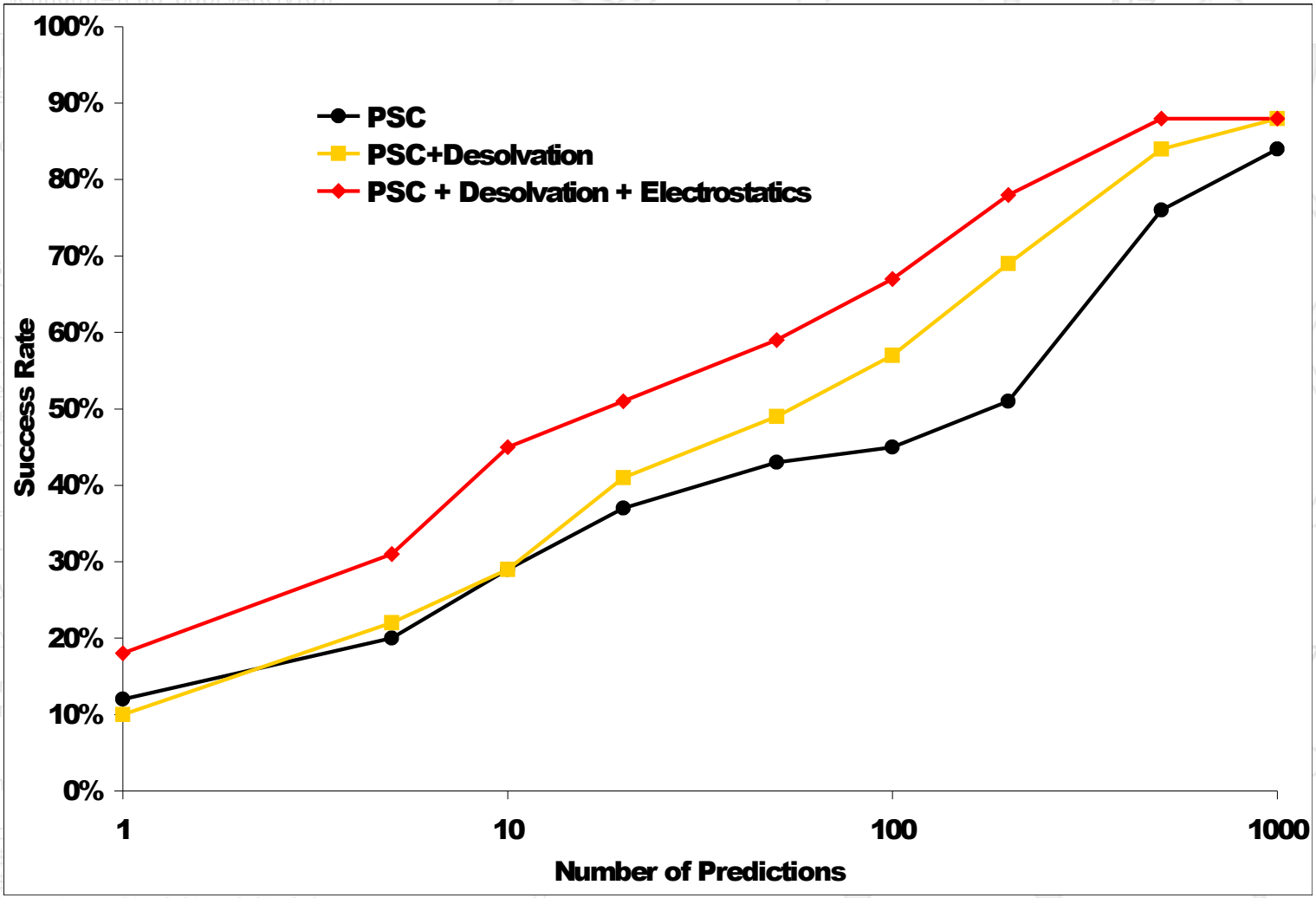
Функция оценки энергии связывания

$$\Delta G = \Delta E_{VdW} + \Delta E_{el.} + \Delta G_{desol} + \Delta G_{const}$$

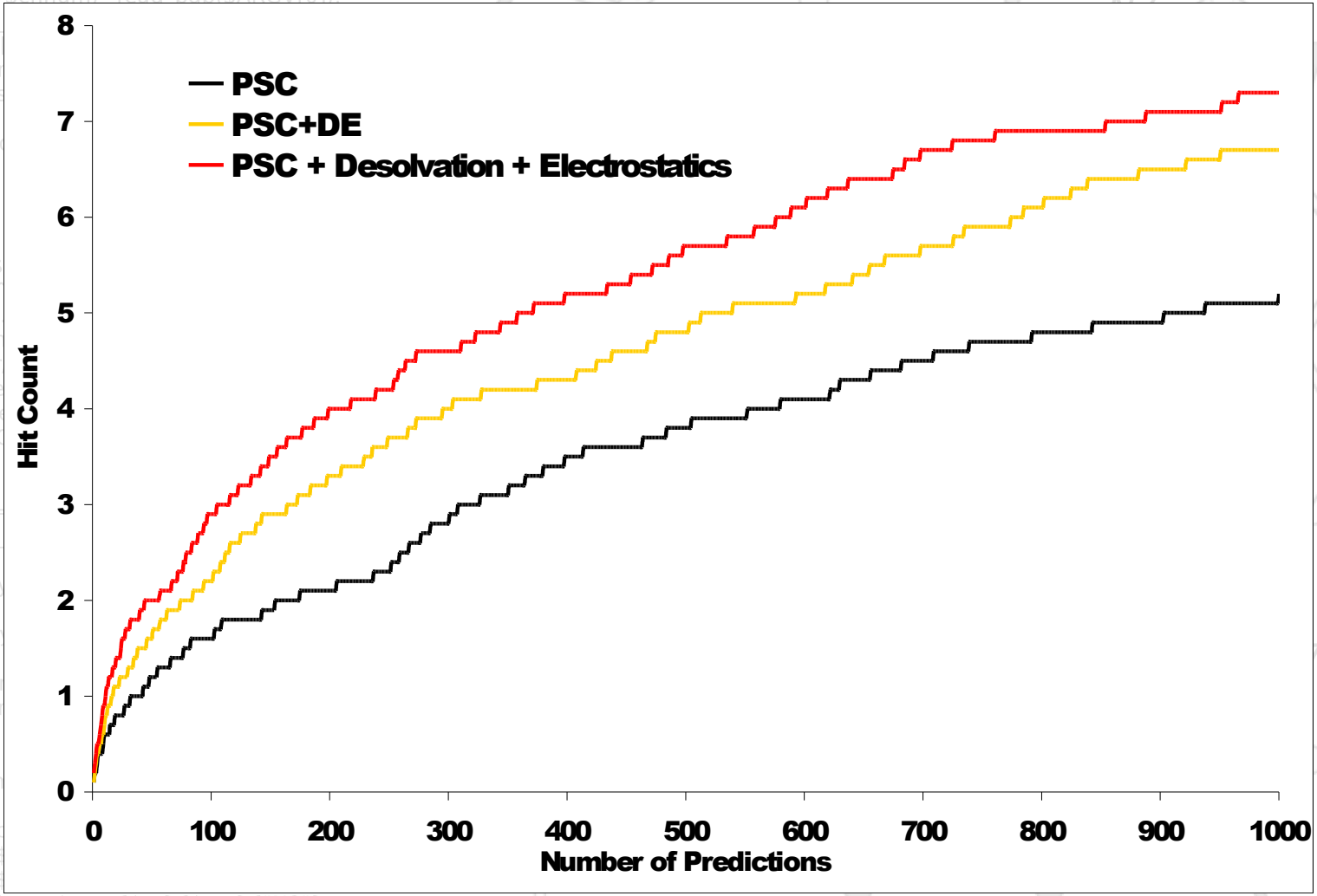
- ΔE_{VdW} - это комплементарность поверхности
- ΔG_{desol} - это гидрофобика
- $\Delta E_{el.}$ - это электростатика
- ΔG_{const} - это изменение вращательной и прочих энтропий.

$$\Delta G_{desol} = \sum_i \sum_j N_{ij} \Delta G_{ij}$$

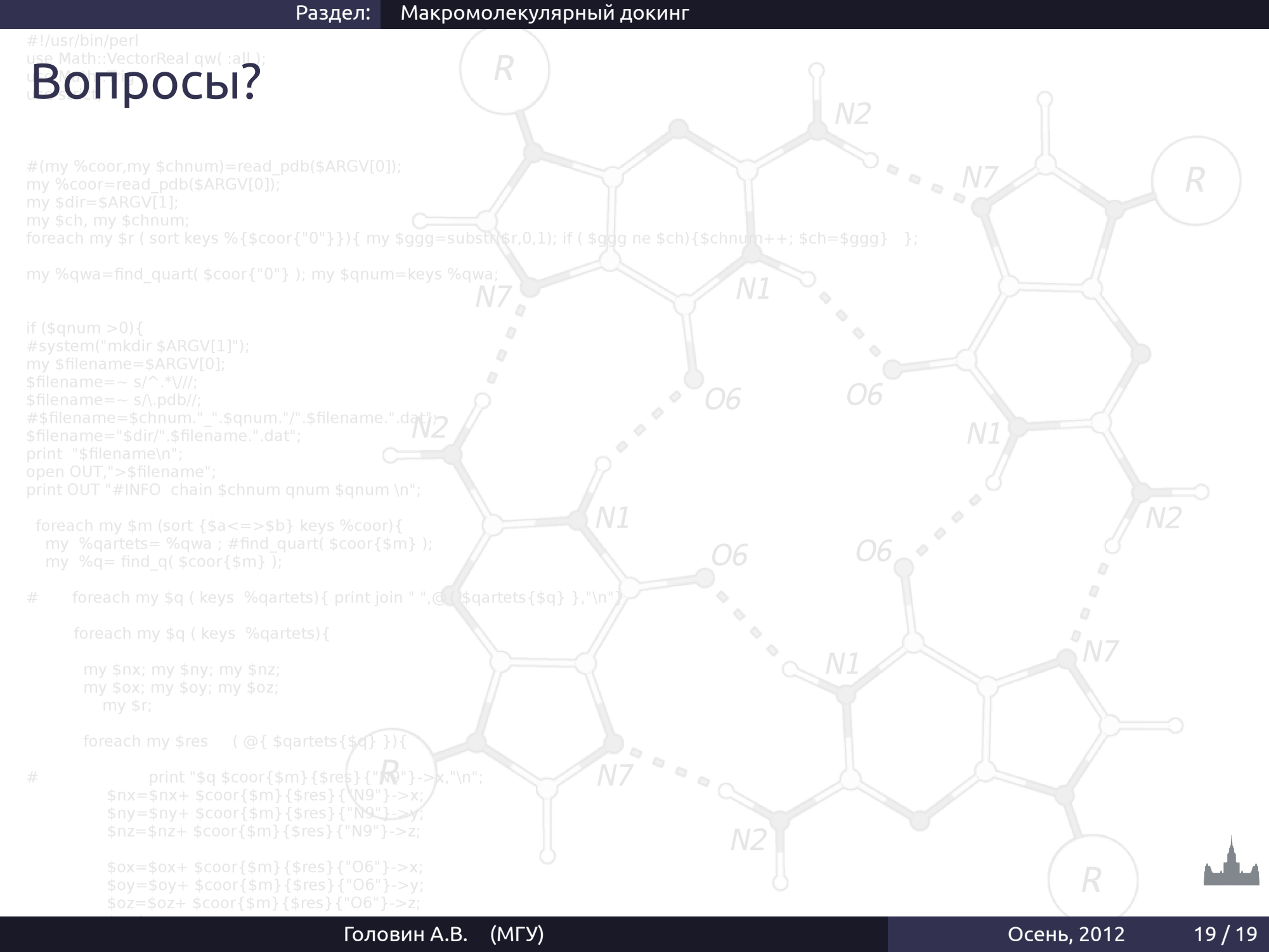
Влияние на Succes rate



Влияние на Hit Count



Вопросы?


#!/usr/bin/perl
use Math::VectorReal qw(:all);
#(my %coor,my $chnum)=read_pdb($ARGV[0]);
my %coor=read_pdb($ARGV[0]);
my $dir=$ARGV[1];
my $ch, my $chnum;
foreach my $r (sort keys %{$coor{"0"}}){ my $ggg=substr($r,0,1); if ($ggg ne $ch){$chnum++; $ch=$ggg} };
my %qwa=find_quart($coor{"0"}); my $qnum=keys %qwa;
if ($qnum >0){
#system("mkdir $ARGV[1]");
my $filename=$ARGV[0];
$filename=~ s/^.*\///;
$filename=~ s/\.pdb//;
#$filename=$chnum."_".$qnum."/".$filename.".dat";
$filename="$dir/".$filename.".dat";
print "$filename\n";
open OUT,">$filename";
print OUT "#INFO chain $chnum qnum $qnum \n";
foreach my $m (sort {$a<=>$b} keys %coor){
my %qartets= %qwa ; #find_quart($coor{$m});
my %q= find_q($coor{$m});
# foreach my $q (keys %qartets){ print join " ",@{ $qartets{$q} },"\n"}
foreach my $q (keys %qartets){
my $nx; my $ny; my $nz;
my $ox; my $oy; my $oz;
my $r;
foreach my $res (@{ $qartets{$q} }){
# print "$q $coor{$m}{$res}{"N9"}->x,"\n";
$nx=$nx+ $coor{$m}{$res}{"N9"}->x;
$ny=$ny+ $coor{$m}{$res}{"N9"}->y;
$nz=$nz+ $coor{$m}{$res}{"N9"}->z;
$ox=$ox+ $coor{$m}{$res}{"O6"}->x;
$oy=$oy+ $coor{$m}{$res}{"O6"}->y;
$oz=$oz+ $coor{$m}{$res}{"O6"}->z;
R
N2
N7
R
N7
N1
O6
O6
N2
N1
N1
N2
O6
O6
N1
N7
N7
N2
R