

What Is Bioinformatics?

报告人: Yu Lijia, Wang LingPing

生物信息小组:

Chen hua, Mao FengBiao, Wang LingPing, Wang Qi, Wang XiaoShan, Yu Lijia



Biowords

It looks like biologists are colonizing the dictionary with all these **bio**words: we have **bio-chemistry** (生物化学), **bio-metrics** (生物测定学), **bio-physics** (生物物理学), **bio-technology** (生物技术), **bio-hazards** (生物性危害), and even **bio-terrorism** (生物恐怖主义). Now what's up with the new entry in the **bio-sweepstakes**, **bio-informatics**?



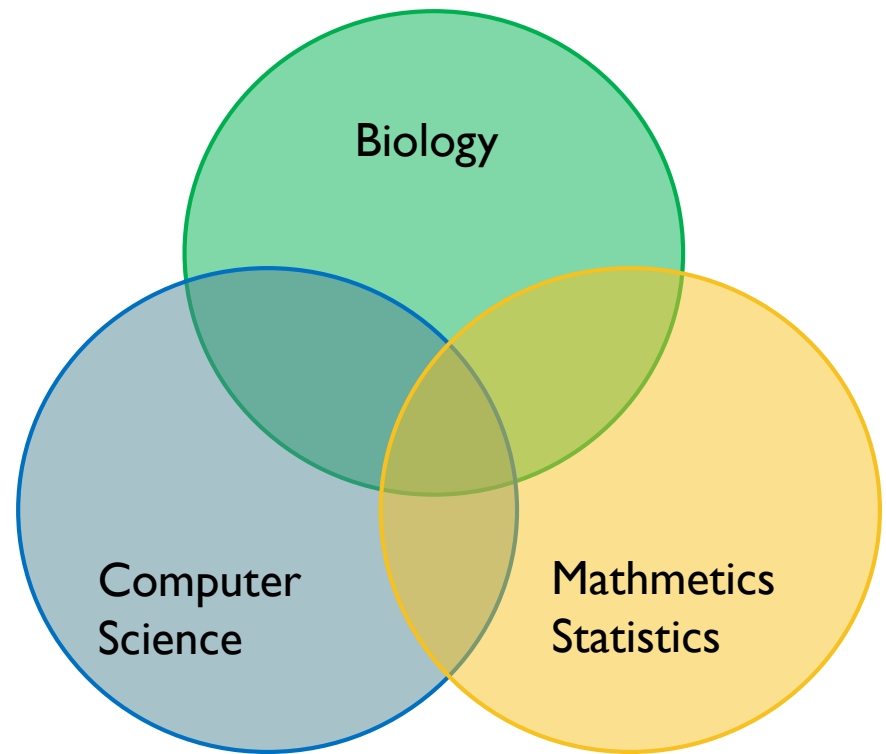
Bioinformatics?

▶ Definition

- ▶ Integration of computational and biological methods to promote biological discovery
- ▶ Combination of Biology, Mathmatics (Statistics), Computer Science

▶ Purpose

- ▶ Predict, Decipher, Visualize



Origins and history

- Symposium on information theory in biology, Gatlinburg, Tennessee, October 29-31, 1956

		A	T	T	C	G	T	A	C	T	T	A	G	T
	0	-1	-2	-3	-4	-5	-6	-7	-8	-9	-10	-11	-12	-13
C	-1	-1	-2	-3	-2	-4	-6	-7	-8	-9	-10	-11	-12	-13
T	-2	-2	0	0	-2	-3	-3	-5	-7	-7	-7	-9	-11	-11
T	-3	-3	0	1	-1	-3	-2	-4	-6	-6	-6	-8	-10	-10
A	-4	-2	-2	-1	0	-2	-4	-1	-3	-5	-7	-5	-7	-9
G	-5	-4	-3	-3	-2	1	-1	-3	-2	-4	-6	-8	-4	-6
C	-6	-6	-5	-4	-2	-1	0	-2	-2	-3	-5	-7	-6	-5
T	-7	-7	-5	-4	-4	-3	0	-1	-3	-1	-1	-3	-5	-5
A	-8	-6	-7	-6	-5	-5	-2	1	-1	-3	-2	0	-2	-4
A	-9	-6	-7	-8	-7	-6	-4	1	0	-2	-4	0	-1	-3
T	-10	-8	-5	-5	-7	-8	-4	-1	0	1	1	-1	-1	0
C	-11	-10	-7	-6	-4	-6	-6	-3	0	-1	0	0	-2	-2
A	-12	-10	-9	-8	-6	-5	-7	-3	-2	-1	-2	1	-1	-3
G	-13	-12	-11	-10	-8	-5	-6	-5	-4	-3	-2	-1	2	0

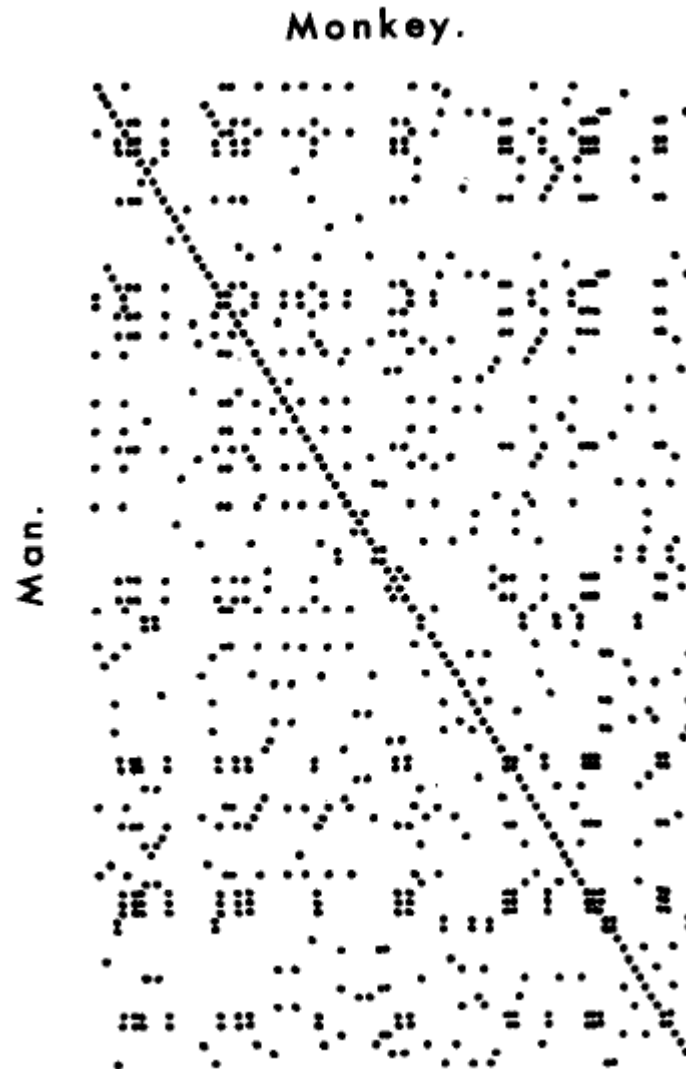
So the best alignment would be:

-- = gap
| = match

ATTCG--TACTTAGT
||| ||| |||
CTTAGCTAATCAG--

Origins and history

- ▶ Symposium on i
Tennessee, Octo



ogy, Gatlinburg,

Origins and history

- ▶ Symposium on information theory in biology, Gatlinburg, Tennessee, October 29-31, 1956



Origins and history

- ▶ Symposium on information theory in biology, Gatlinburg, Tennessee, October 29-31, 1956



Origins and history

- ▶ Symposium on information theory in biology, Gatlinburg, Tennessee, October 29-31, 1956



Origins and history

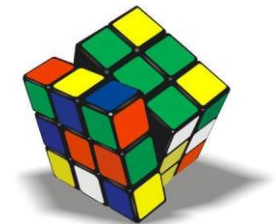
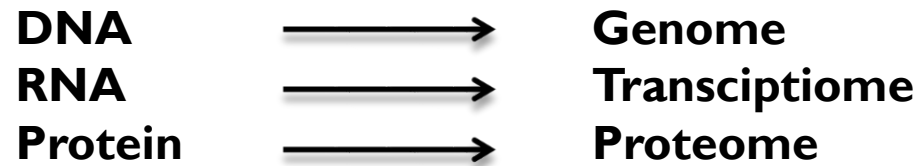
- ▶ Symposium on information theory in biology, Gatlinburg, Tennessee, October 29-31, 1956



Eric Steven Lander
lander@broad.mit.edu
Whitehead Institute

Units of information

DNA	Sequence	Pathways
RNA	Structure	Interactions
Protein	Evolution	Mutations



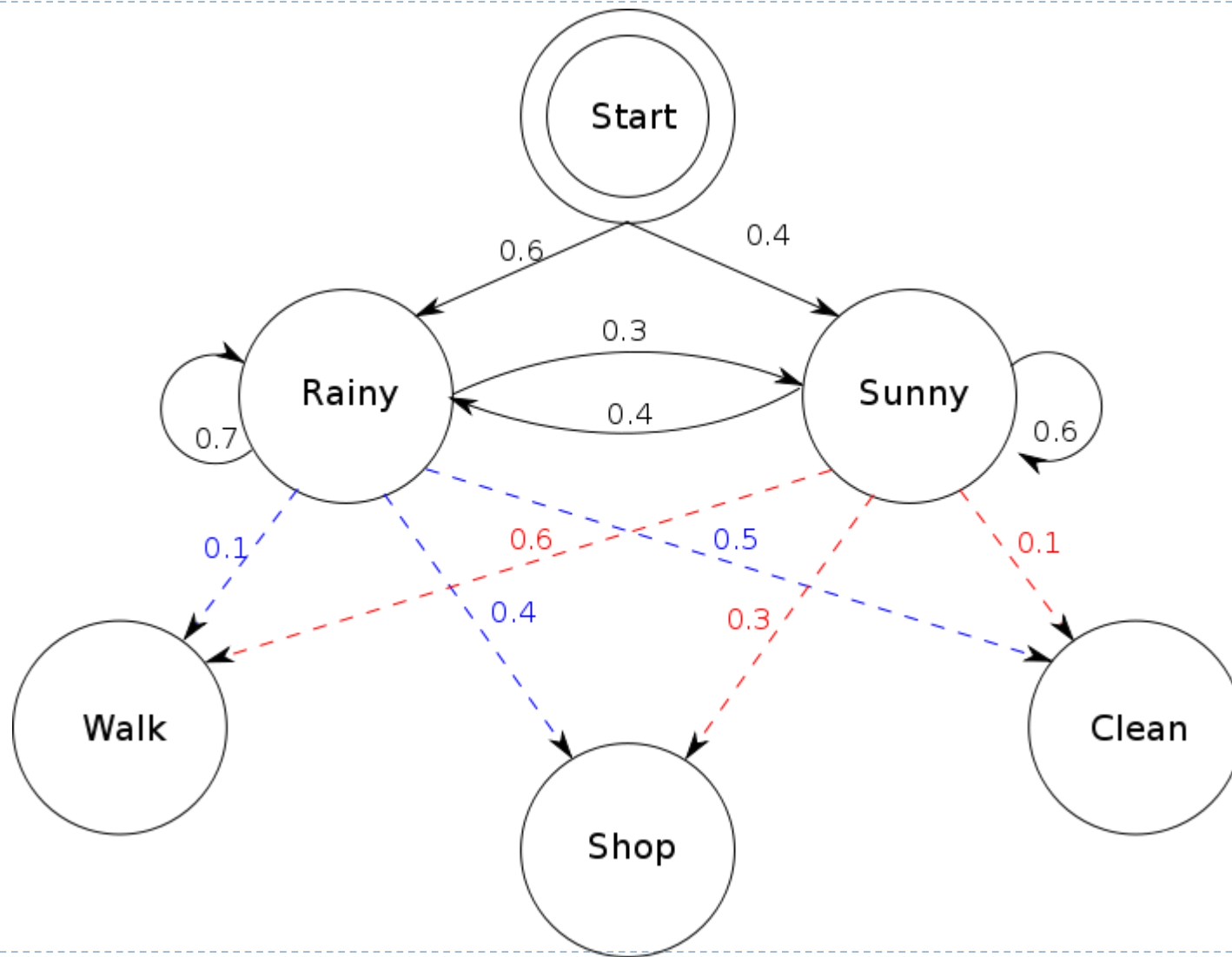
Work with DNA

- ▶ Simple sequence Analysis
 - ▶ database searching——BLAST
 - ▶ pairwise comparison——两序列比对
- ▶ Regulatory sequence——Sequence logo
- ▶ Gene finding——Hidden Markov model
- ▶ Comparative genomic(analyses between species and strains)

A	2	1	0	21	0	2	2	1	2
C	0	21	21	1	1	9	1	3	4
G	17	1	0	1	1	6	2	3	10
T	4	0	2	0	21	6	18	16	7



Markov Model and Hidden Markov model



Work with RNA and Protein

- ▶ Splice variants——GeneChip
- ▶ Tissue specific expression——GeneChip

Detection method?

——Shannon entropy

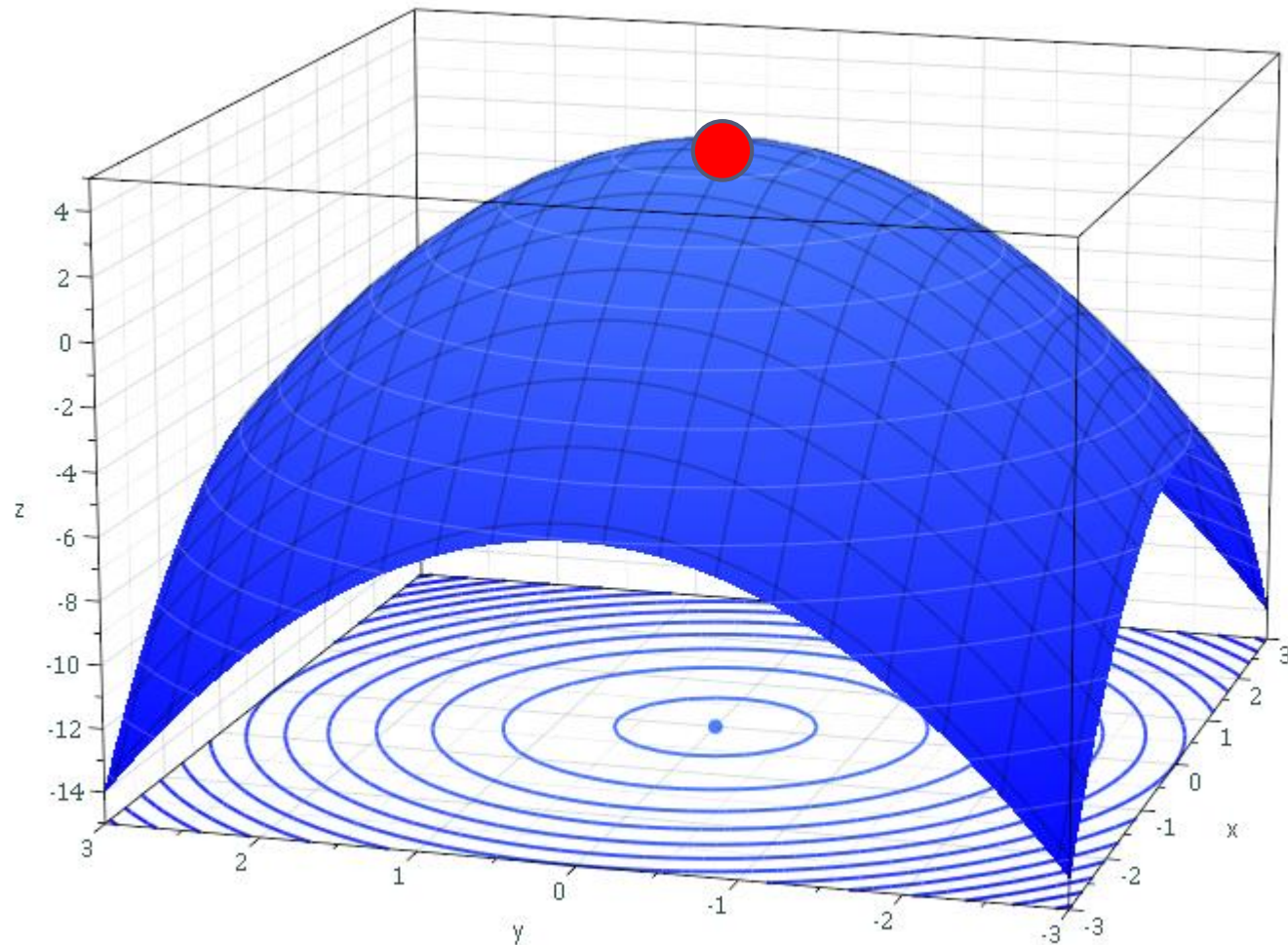
$$H(x)=E(I(x))$$

$$I(x)=-\log(P(x))$$

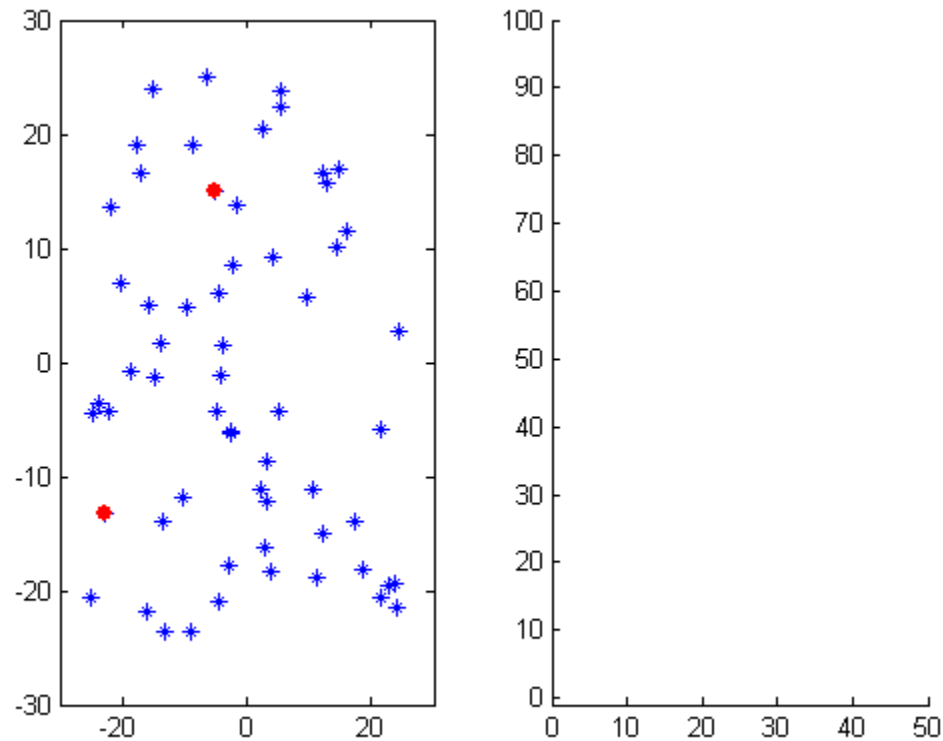
- ▶ 3D Structure
 - ▶ 科学发现游戏Foldit——Mathematical optimization



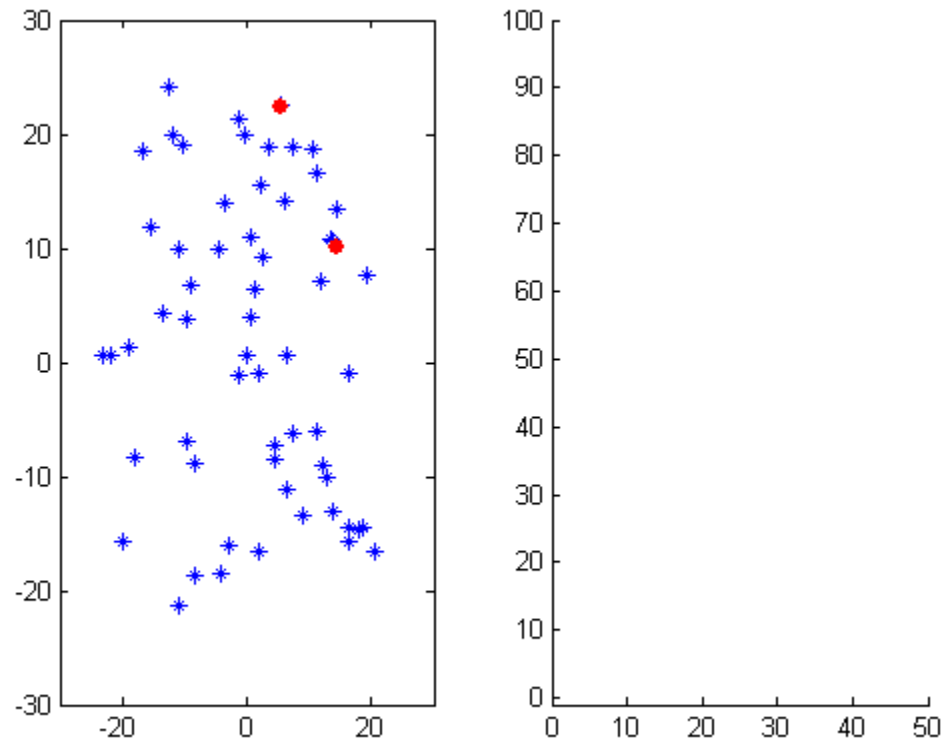
Mathematical optimization



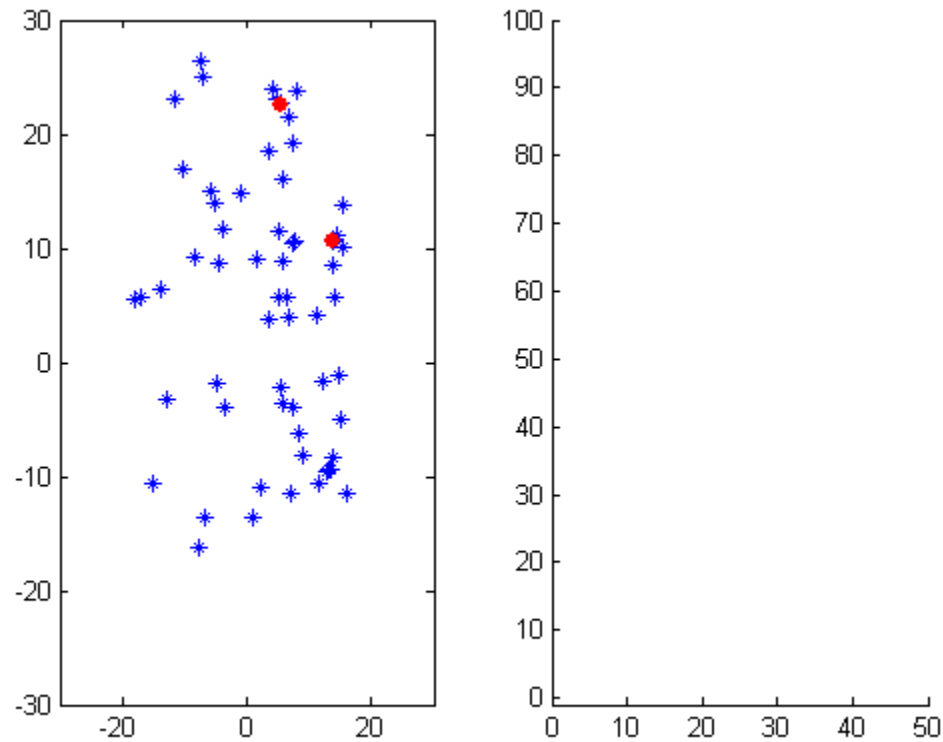
Mathematical optimization



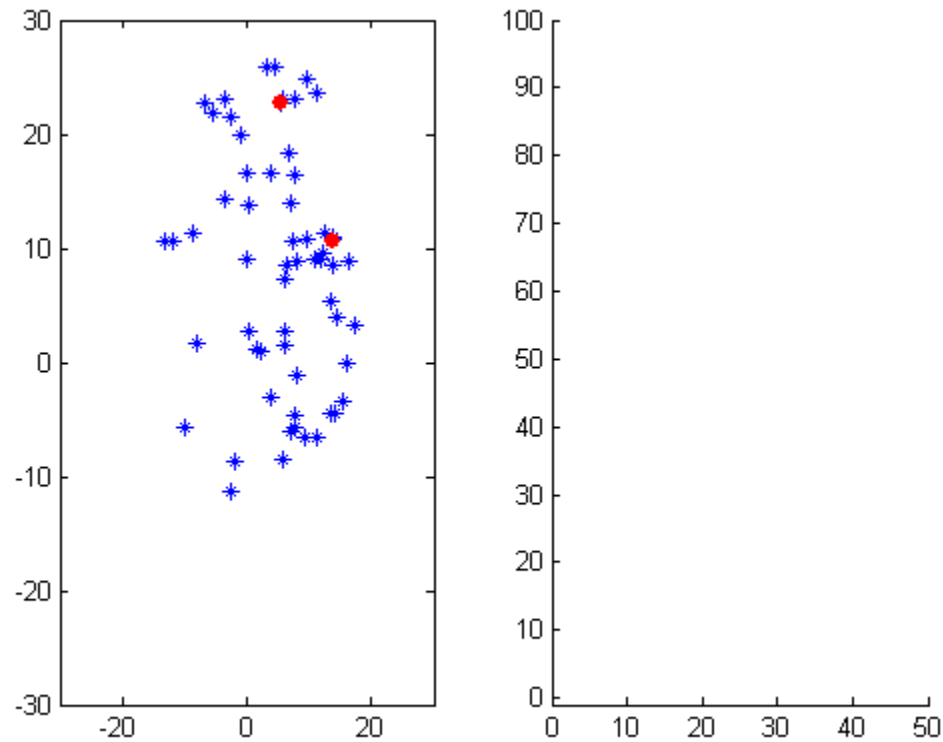
Mathematical optimization



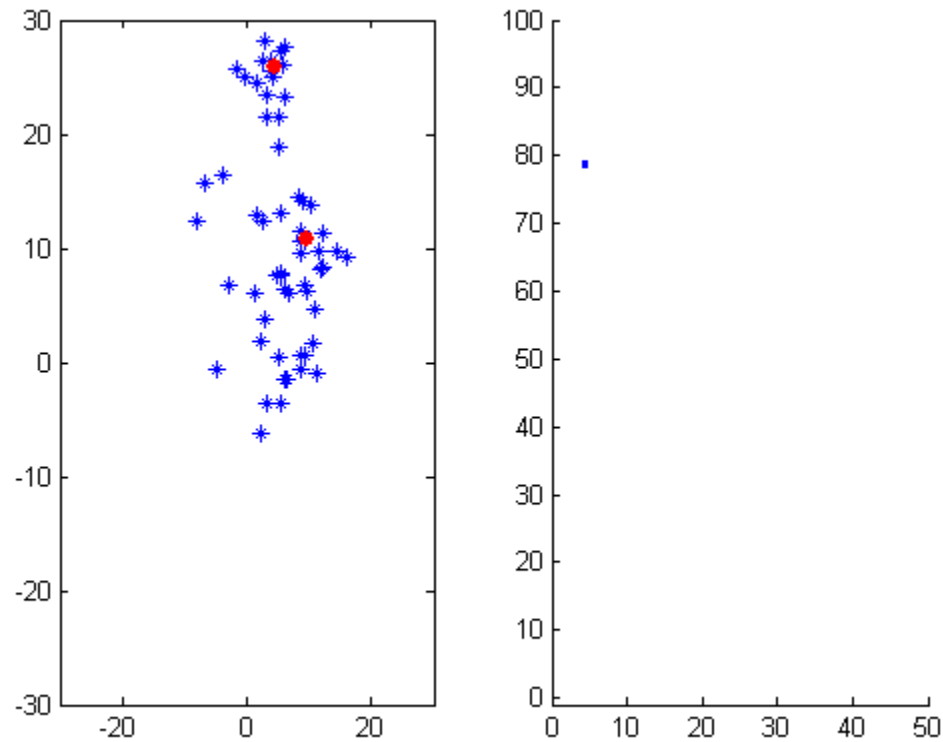
Mathematical optimization



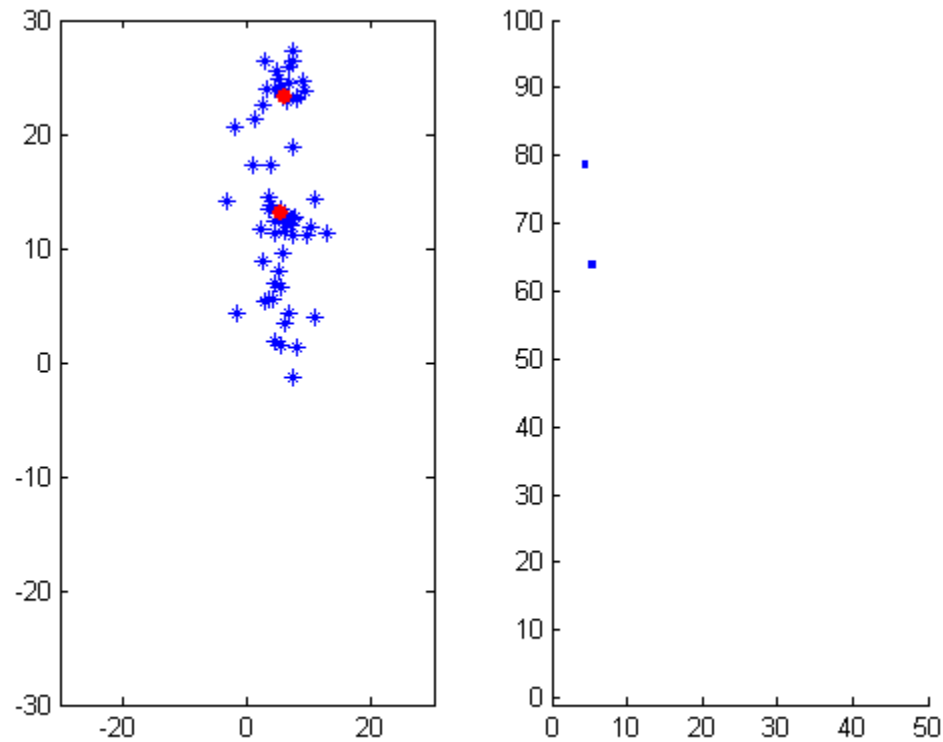
Mathematical optimization



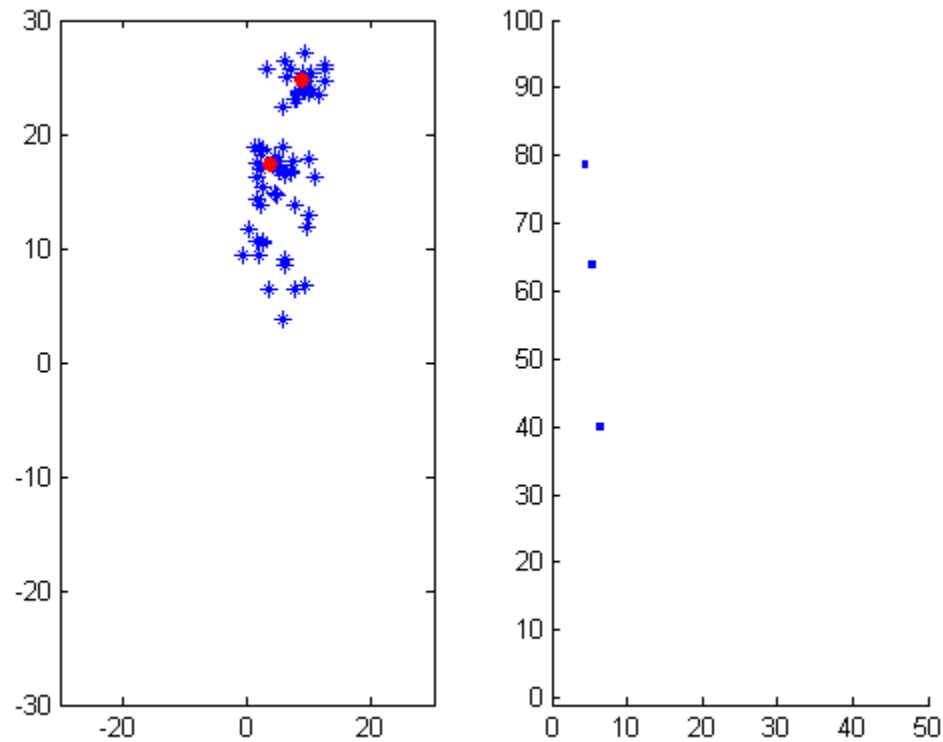
Mathematical optimization



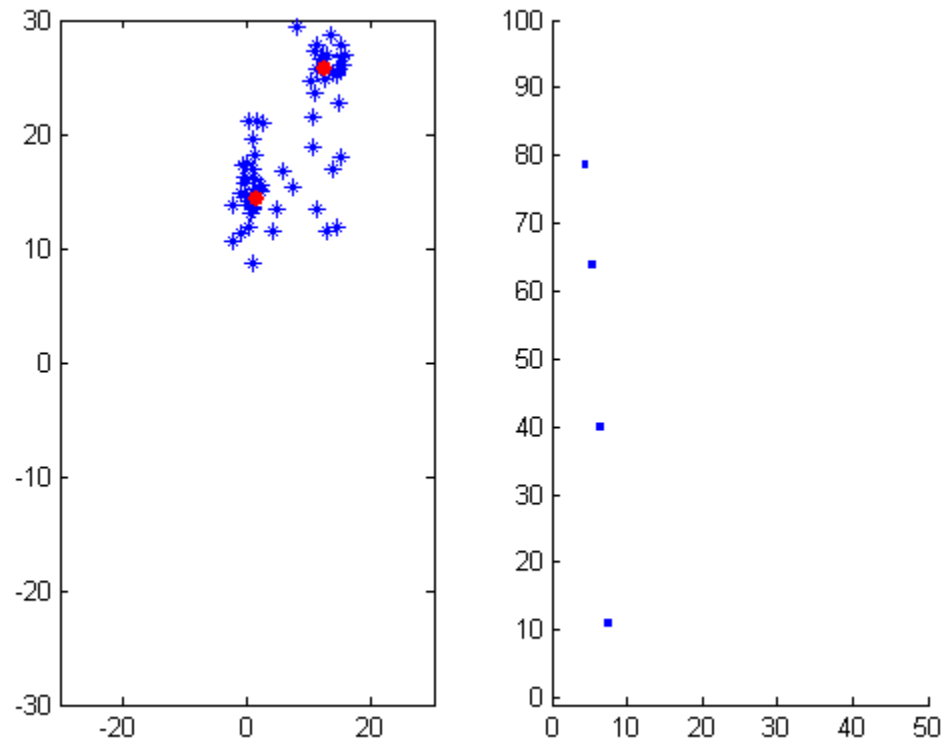
Mathematical optimization



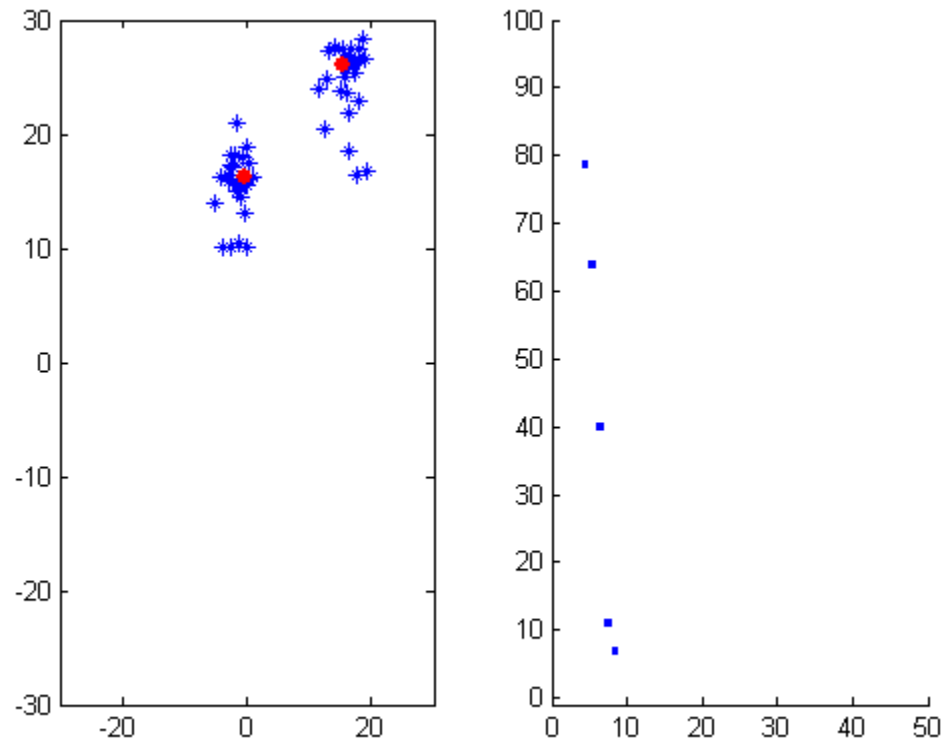
Mathematical optimization



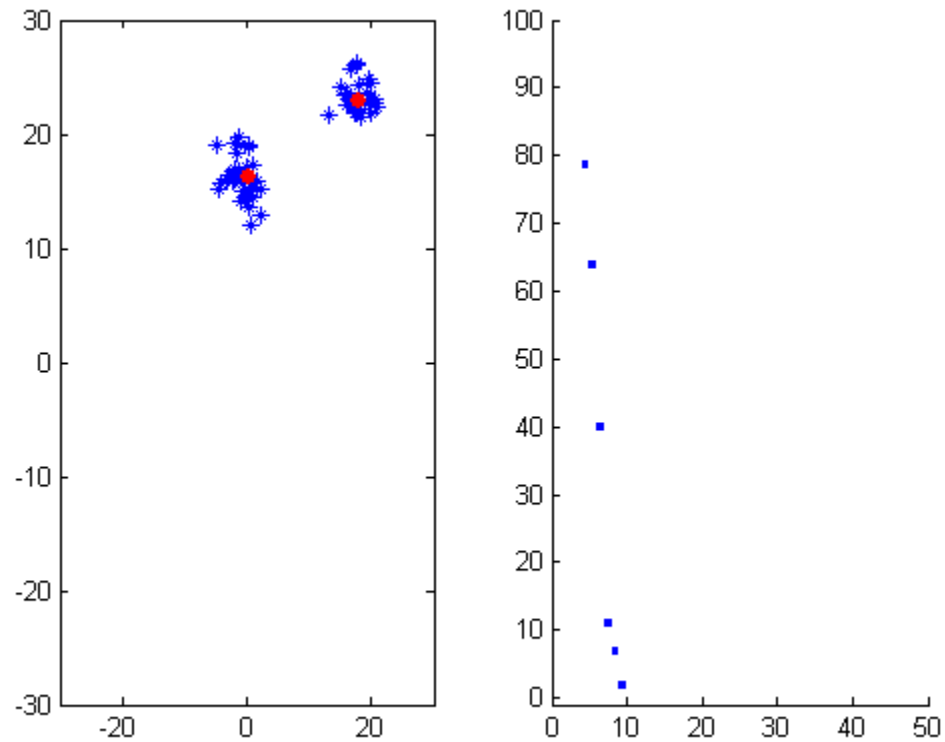
Mathematical optimization



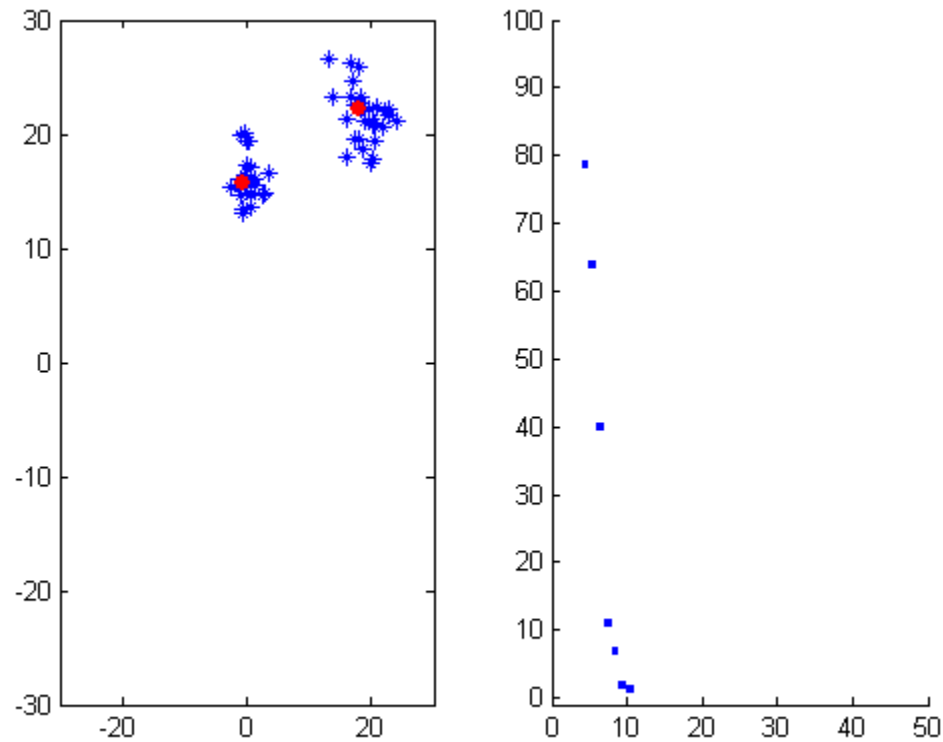
Mathematical optimization



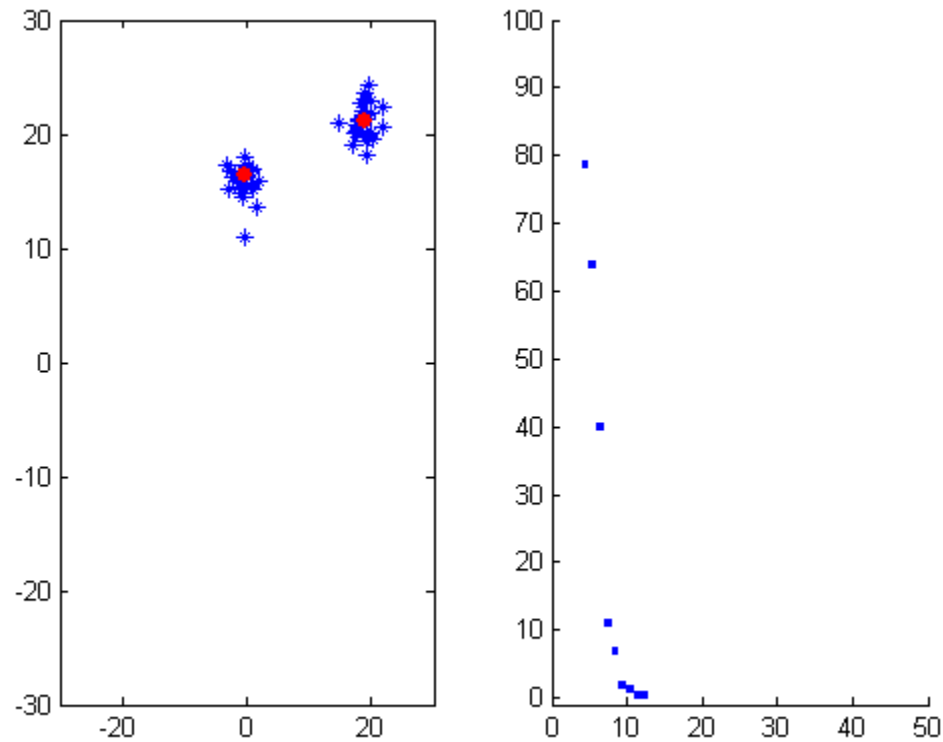
Mathematical optimization



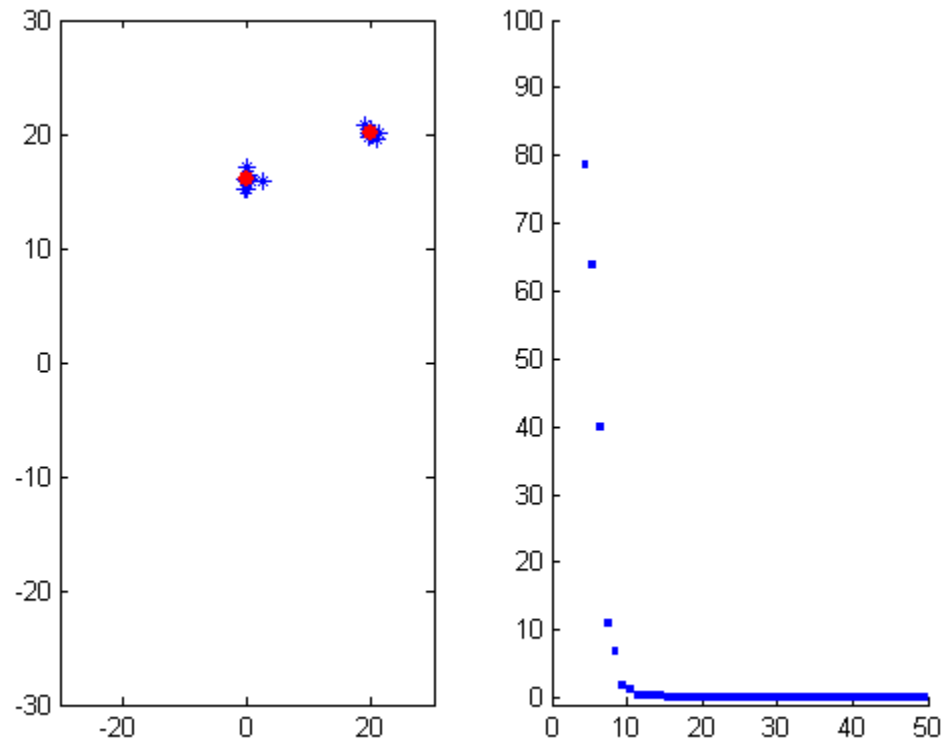
Mathematical optimization



Mathematical optimization



Mathematical optimization



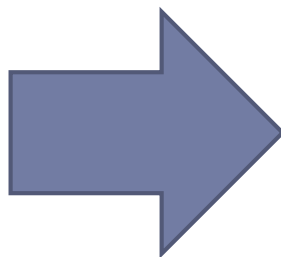
Foldit —— 在线蛋白质折叠游戏



Rosetta@home 是一项利用已联网的计算机来准确预测和设计蛋白质结构及聚合物的分布式计算项目。



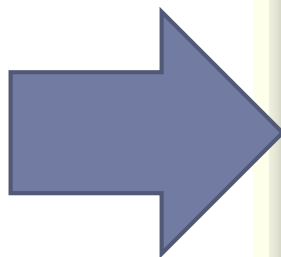
Foldit —— 在线蛋白质折叠游戏



遇到困境：计算机算法
总会陷入局部最优解



Foldit —— 在线蛋白质折叠游戏



在 2008 年 5 月 9 日，贝克实验室接受 Rosetta@home 用户关于交互式版本的建议，发布了 Foldit。



Foldit介绍

The screenshot displays the Foldit game interface. At the top left, the 'foldit' logo and 'Pull Mode' button are visible. The top center shows the player's 'Rank: 735' and 'Score: 8766'. Below the score, it indicates 'Soloist 478 (<15% New...berculosis Protein)' and 'Expires 12/08/2011 09:09:55 (5 days, 10 hours)'. A 'No bonuses or conditions' message is also present. The top right features a 'Group Competition' panel with a table of group names and scores, and a 'Soloist Competition' panel with a table of player names and scores. The central area shows a 3D protein structure with various colored segments (blue, green, orange, red, white). The bottom left contains a 'Cookbook' sidebar with icons for actions like 'Shake Sidechains', 'Mutate Sidechains', 'Wiggle All', 'Wiggle Backbone', 'Wiggle Sidechains', 'Help', 'Glossary', 'Freeze Protein', 'Remove Bands', 'Disable Bands', 'Reset Structures', 'Reset Puzzle', and 'Align Guide'. The bottom right has a 'Chat' panel with options for 'Chat - Puzzle', 'Chat - Global', and 'Notifications'. The bottom of the screen shows a Windows taskbar with the 'foldit' application running.

分数

排名

操作区

#	Group Name	Score
1	strong_base	9699
2	theroller	9693
3	senger	9666
4	mcammack	9659
5	j_johnso	9637
6	Bithalbier	9635
7	Cytochrome X	9619

#	Player Name	Current	Best
1	strong_base	-	9699
2	theroller	-	9693
3	senger	-	9666
4	mcammack	-	9659
5	j_johnso	-	9637
6	Bithalbier	-	9635
7	Cytochrome X	-	9619

296	Vicfung3	
297	loge	
298	pjd306	
299	abriggs	
300	Mong0	
301	fwjmath	9143
302	Aesir	
303	Dragon89	

连接，没啥用，而且不容易拉动

氨基酸残基，图中只标示出了根部的位置，橙色荧光的是疏水的，而蓝色的是亲水的。

这些边沿曲折的就是sheet，如果将不同的sheet适当拼起来就会形成蓝色的氢键

这些蓝白相间的就是氢键，因为它们键能很高，所以在大多数情况下越多氢键蛋白质就越稳定，分数也就越高

这个就是螺旋，上面有不少的氨基酸残基，摆放它们的时候最重要的一点就是要把疏水的基团藏起来，因为细胞内环境有很多水，如果疏水的基团露在外边的话不利于蛋白质的稳定性，也不利于得高分

自动调整氨基酸残基位置以降低能量

自动调整骨架以降低能量（与Rosetta算法相同）

www.equn.com/forum
by fwjmath

游戏玩家破解蛋白质谜题，艾滋病、癌症研究有望获重大突破

- ▶ 仅用了三周时间，游戏玩家就解决了一个困扰科学家好几年的难题。一群玩家通过玩游戏预测了逆转录病毒蛋白酶的结构，这种蛋白质在艾滋病毒生长过程中起到了至关重要的作用。该发现标志着人类有望在艾滋病毒（HIV）和艾滋病（AIDS）研究领域获得重大突破。这一成果刊登在Nature Structural & Molecular Biology杂志上。

So if you're looking for a game that can double as good volunteer work, go play. You just might help change the world.




生物信息学数据库

▶ 大量生物数据

-- 存储

-- 分析



-
- 基因组数据库
 - 人类、小鼠、果蝇、水稻
 - 核酸序列数据库
 - NCBI、EMBL、DDBJ
 - 蛋白质序列数据库
 - SWISS-PROT、PIR
 - 蛋白质结构数据库
 - PDB、SCOP、CATH
 - 二次数据库
 - 比较基因组学
 - 代谢途径和细胞调控
 - 农、林、医学
-
- 

[Genomes](#) - [Blat](#) - [Tables](#) - [Gene Sorter](#) - [PCR](#) - [VisiGene](#) - [Proteome](#) - [Session](#) - [FAQ](#) - [Help](#)

- 完成 (网页评级: 4 级)

EMBL



European Molecular Biology Laboratory

by category full

Grenoble | Hamburg | **Heidelberg** | EMBL-EBI Hinxton | Monterotondo

ABOUT US

RESEARCH

SERVICES

TRAINING

EMBL News



Grenoble, 13 October 2011

Intruder detected: raise the alarm!

How a molecular switch activates the anti-viral innate immune response > [more](#)

Heidelberg/Hinxton, 14 September 2011

Five countries and EMBL sign Memorandum of Understanding to make ELIXIR a reality

> [more](#)

> [More News](#)

Upcoming Events

Tuesday, 6 December 2011, 15:00, Small Operon

Bioorthogonal Staudinger reactions – from labeling to the functionalization of proteins

Christian Hackenberger, Freie Universität Berlin, Institut für Chemie und Biochemie

Monday 5 December 2011

EMBL Symposium The Use of Zinc Finger Nucleases for the Development of Next Generation Cell Lines and Animal Models

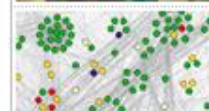
More events > [Seminars](#) > [Events](#)



Conferences and Courses



PhD Programme



Bioinformatics Services

EMBO Conference Series



EMBL Heidelberg, Germany | 12 - 17 Jun 2012

Click here
advent c

Click here
advent c

Quick Li

- > Job O
- > PhD P
- > Postc
- > Scien
- > Visito
- > Alum
- > Techn

Follow u

RSS

You

face

PIR

A UniProt CONSORTIUM MEMBER

Protein Information Resource

TLALPN----RKAVADH

LIGCLRNC SAVTAAAKQ

VTGFSN----AKTTAQH

Protein Se

About PIR

Databases

Search/Analysis

Download

Support

INTEGRATED PROTEIN INFORMATICS RESOURCE FOR
GENOMIC, PROTEOMIC AND SYSTEMS BIOLOGY RESEARCH

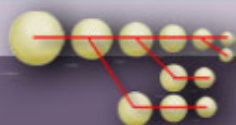
The Universal Protein Resource (UniProt) provides the centralized, authoritative resource for protein sequence and structure information.

UniProtKB | UniRef | UniParc

Current release: 2011_11

PRO

Protein Ontology



- Representation of protein objects with descriptions and relationships
- [Browse PRO](#)
- Annotate with [RACE-PRO](#)

Sample PRO report

iProClass

Integrated Protein Knowledgebase



- Value-added reports for [UniProtKB](#) and unique [UniParc](#) proteins
- Functional analysis and [protein ID mapping](#)

Sample protein report

iProLINK

Literature Information & Knowledge



- Source for text mining and ontology development
- [RLIMS-P](#) text mining tool, [BioThesaurus](#)
- [Bibliography mapping](#)

Sample Biblio. report

O OTHER RESOURCE

- [Representative Proteomes](#)
- [PIR Grid-Enablement](#): Data node on NCI's [caBIG](#)

P PEPTIDE SEARCH ?

DATABASE: UniProtKB

Use single letter amino acid code

T TEXT SEARCH ?

DATABASE: iProClass

Structural Classification of Proteins



Welcome to **SCOP: Structural Classification of Proteins**.
1.75 release (June 2009)

38221 PDB Entries. 1 Literature Reference. 110800 Domains. (excluding nucleic acids and theoretical models).

Folds, superfamilies, and families [statistics here](#).

[New folds](#) [superfamilies](#) [families](#).

[List of obsolete entries and their replacements](#).

Authors. Alexey G. Murzin, John-Marc Chandonia, Antonina Andreeva, Dave Howorth, Loredana Lo Conte, Bartlett G. Ailey, Steven E. Brenner, Tim J. P. Hubbard, and

Reference: Murzin A. G., Brenner S. E., Hubbard T., Chothia C. (1995). SCOP: a structural classification of proteins database for the investigation of sequences and structures.

Recent changes are described in: Lo Conte L., Brenner S. E., Hubbard T.J.P., Chothia C., Murzin A. (2002). SCOP database in 2002: refinements accommodate structural changes. [\[PDF\]](#).

Andreeva A., Howorth D., Brenner S.E., Hubbard T.J.P., Chothia C., Murzin A.G. (2004). SCOP database in 2004: refinements integrate structure and sequence family data.

Andreeva A., Howorth D., Chandonia J.-M., Brenner S.E., Hubbard T.J.P., Chothia C., Murzin A.G. (2007). Data growth and its impact on the SCOP database: new domain families. [DOI:10.1093/nar/gkm993](#) [\[PDF\]](#).

Postdoc Wanted

- Want to help us design and build the next generation of SCOP and ASTRAL?

[Get more details and apply here.](#)

Access methods

- Enter SCOP at the [top of the hierarchy](#)
- [Keyword search of SCOP entries](#)
- [SCOP parseable files](#)
- [All SCOP releases and reclassified entry history](#)

QUESTION?

