

蛋白质结构预测的边角故事

周耀旗

系统与物理生物学所

深圳湾实验室

2022.7

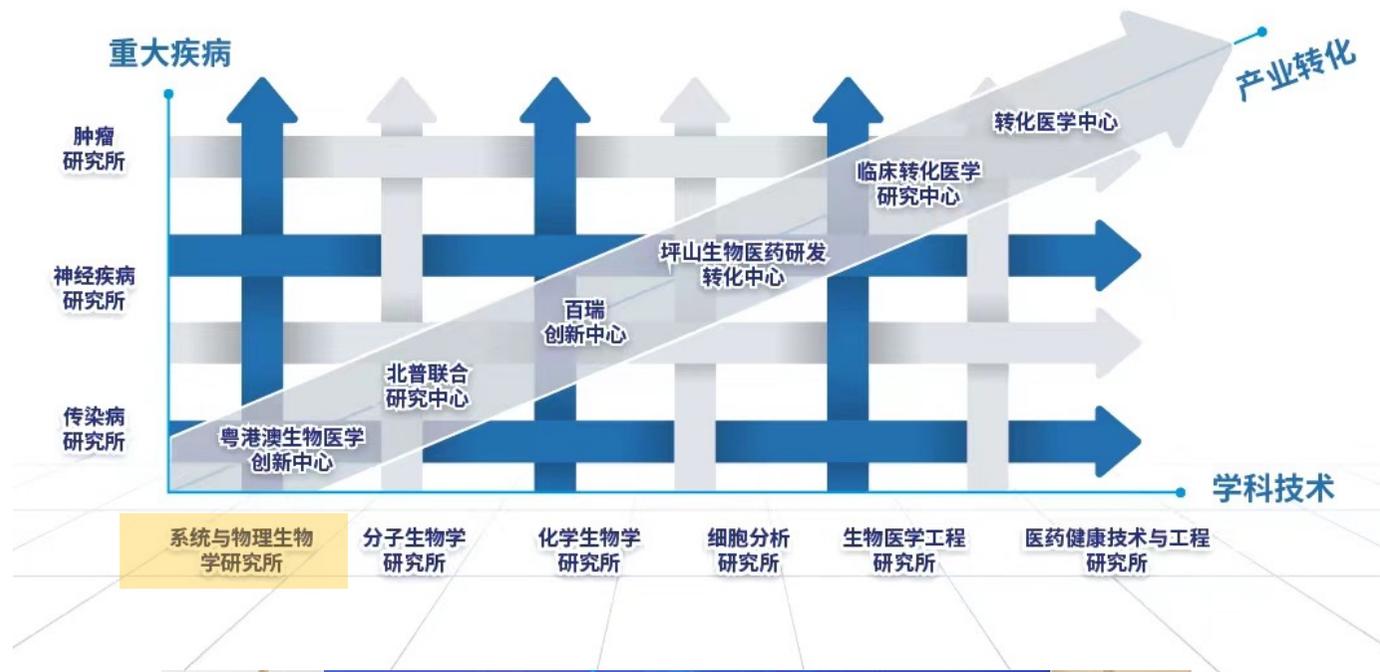


深圳湾实验室
Shenzhen Bay Laboratory



深圳湾实验室

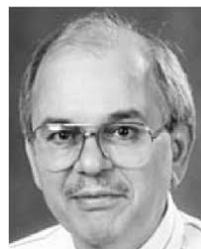
Shenzhen Bay Laboratory



感谢

我在Buffalo, Indiana, Griffith Universities的学生和博士后

Hongyi Zhou, Song Liu, Chi Zhang, Ofer Dor, Bin Xue, Eshel Faraggi, Yuedong Yang, Shesheng Zhang, Tuo Zhang, Huiying Zhao, Zhixiu Li, Thomas Litfin, Peter Brown, Jian Zhan



Prof Kuldip Paliwal
& his team @Griffith Univ

招聘：
研究助理
博士后
助理研究员
副研究员

方向：
计算/AI
软件/硬件
分子生物
结构生物
细胞生物

课题组网站



科学网博客

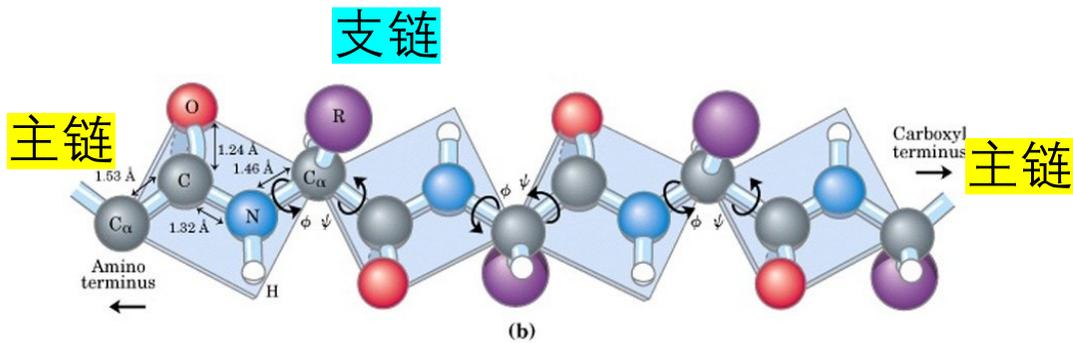


杨跃东和他的团队
@中山大学

生物分子蛋白质

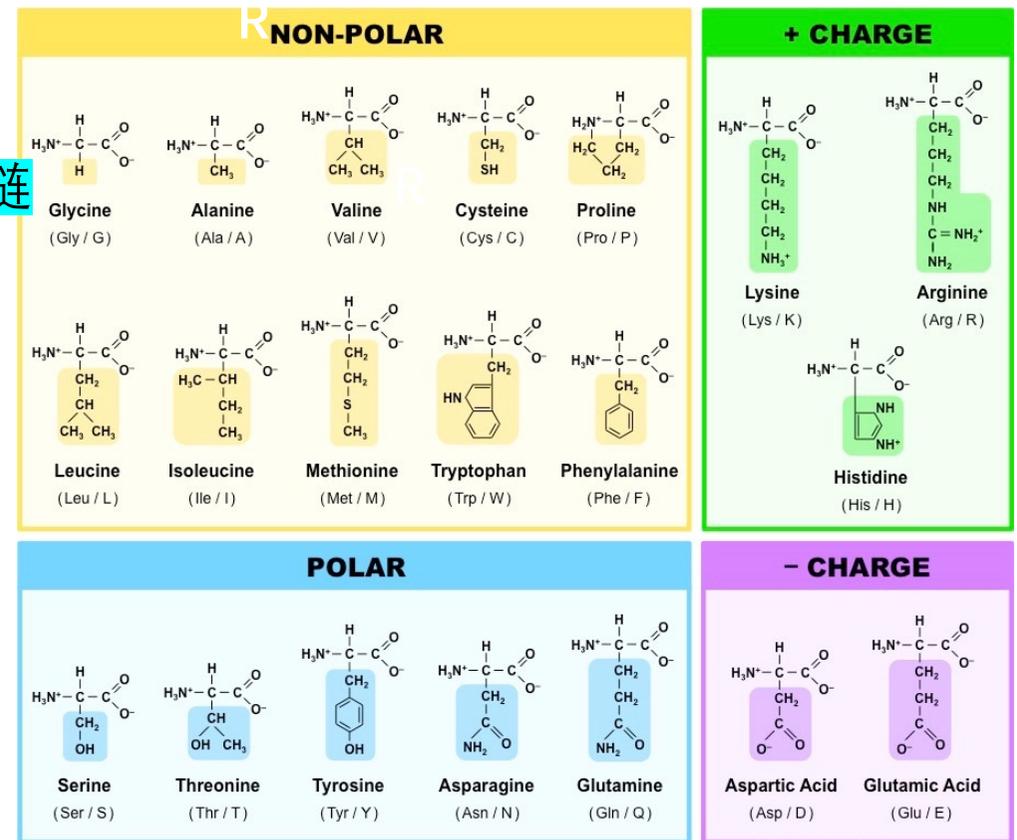
蛋白质：20个氨基酸、不同组合链接而成的高分子链

ASNDCEQGHILKMFPSTWYV



<https://www.quora.com/>

支链



<https://ib.bioninja.com.au>

万能的蛋白质

<http://geneontology.org/>

>10,000分子功能

抗体

酶

信号传递

结构支撑

运输

储存

马达

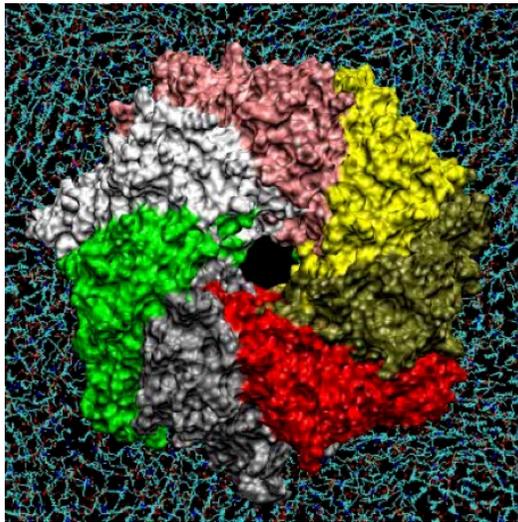
生长

分化...

The screenshot shows the Gene Ontology Resource website. The main heading is "THE GENE ONTOLOGY RESOURCE". Below it, there is a search bar with the text "Search GO term or Gene Product in AmiGO ...". To the right, there is a "GO Enrichment Analysis" section with a search box for "Your gene IDs here...", a dropdown menu for "biological process", and a dropdown menu for "Homo sapiens". There are also "Examples" and "Launch" buttons. The website footer includes the text "Current release 2022-06-15: 43,613 GO terms | 7,511,461 annotations | 1,473,261 gene products | 5,183 species (see statistics)".

无所不能，无所不在

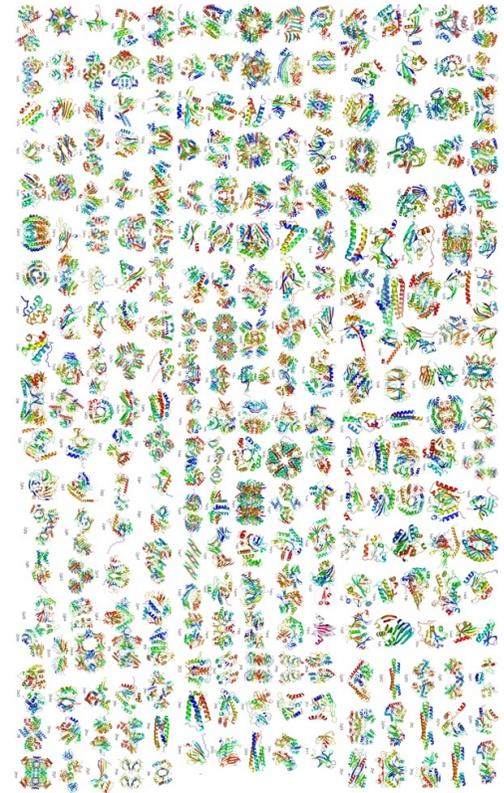
万能的蛋白质靠千变的结构



细菌毒素

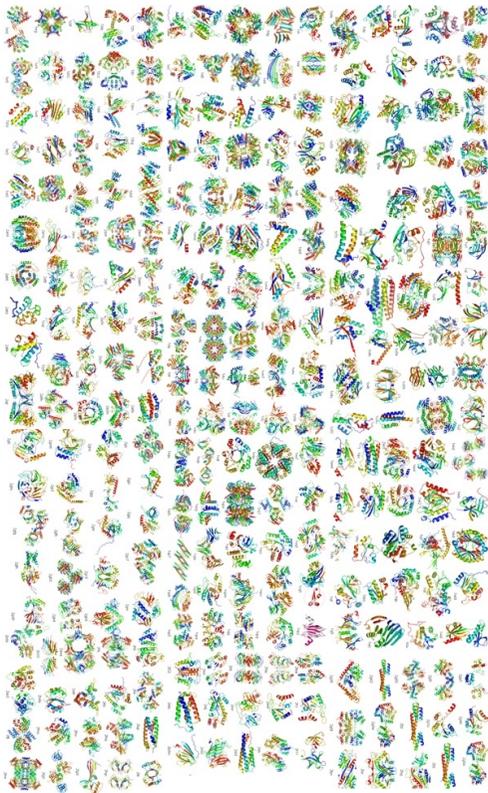


分子马达Kinesin



Protein Data Bank (19万结构)

高精度原子结构靠昂贵的实验



结构解析方法

NMR
X-ray Crystallography
Cryo-Electron Microscopy
X-ray Free Electron Laser

>\$10,000 per protein

Protein Data Bank (50年 : 19万结构)

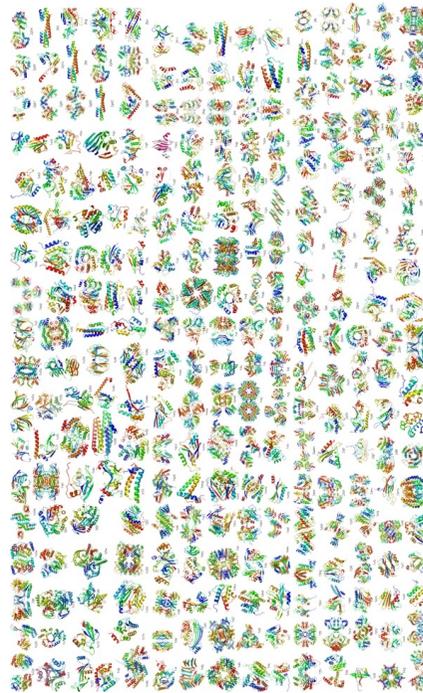
高精度原子结构靠昂贵的实验

10^8 (uniprot.org)

10^5 (PDB)

结构解析方法

....GADTIFG
KIIRKEIPAKII
FEDDRCLAF
HD....



NMR
X-ray Crystallography
Cryo-Electron
Microscopy
X-ray Free Electron
Laser

\$100 基因组

>\$10,000 per protein

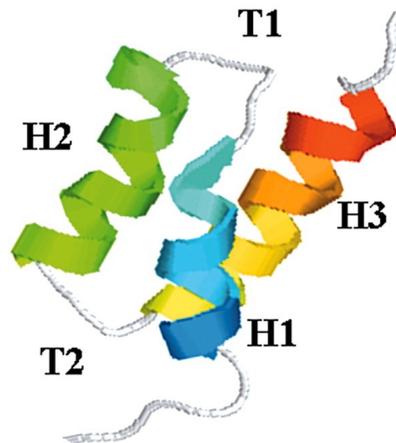
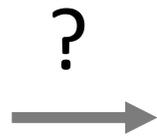
→ Impractical by experimental techniques

蛋白质结构计算预测

50年从序列到结构的奋斗

从头预测

....GADTIFG
KIIRKEIPAKII
FEDDRCLAF
HD....



- Large Conformational Space
- Poor Classical Energy Function (Need Quantum Accuracy)
- **Statistical Energy Function**

H. Zhou and **Y. Zhou**, "Distance-scaled, finite ideal-gas reference state improves structure-derived potentials of mean force for structure selection and stability prediction, *Protein Science*, 11 , 2714-2726 (2002).

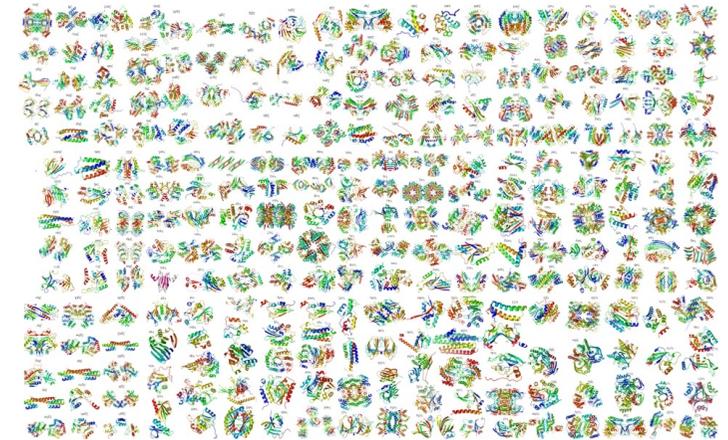
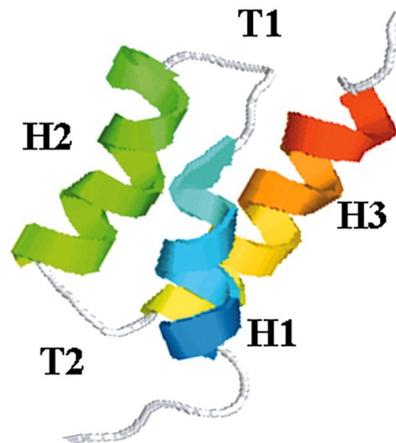
蛋白质结构计算预测

50年从序列到结构的奋斗

基于模板

....GADTIFG
KIIRKEIPAKII
FEDDRCLAF
HD....

?



模板：RAPTOR, SPARKS,...

H. Zhou and Y. Zhou, "Single-body residue-level knowledge-based energy score combined with sequence-profile and secondary structure information for fold recognition" , *Proteins*, 55, 1005-1013 (2004).

蛋白质结构计算预测

50年从序列到结构的奋斗

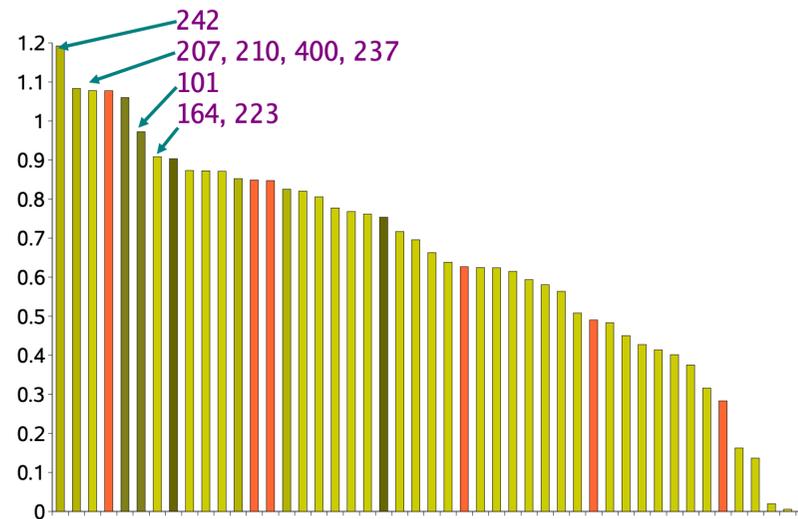
基于模板

CASP 6, 2004

#1: Comparative Modeling by SPARKS

(Group 242)

Results for Servers



Alfonso Valencia CNB-CSIC
Gaeta, 2004

H. Zhou and Y. Zhou, "SPARKS 2 and SP3 servers in CASP 6.", Proteins (Supplement CASP issue), **Suppl7** 152-156 (2005).

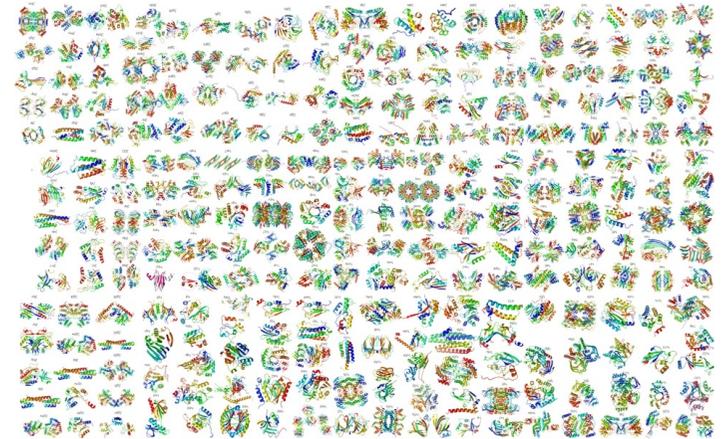
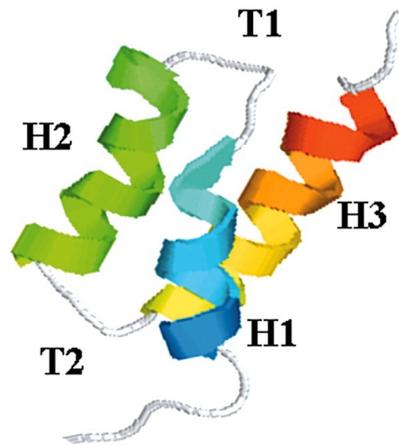
蛋白质结构计算预测

50年从序列到结构的奋斗

基于模板 / 结构碎片

....GADTIFG
KIIRKEIPAKII
FEDDRCLAF
HD....

?



模板/结构碎片 : ROSETTA, TASSER, ...

蛋白质结构的表述

序列

一级结构

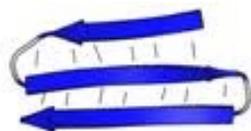
....GADTIFG
KIIRKEIPAKII
FEDDRCLAF
HD....

主链的形状

二级结构



螺旋

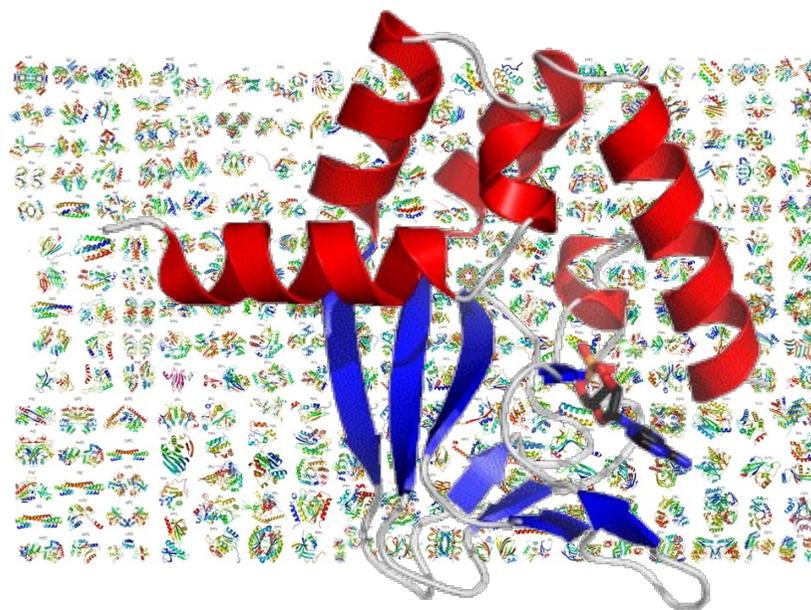


片条状

无规线圈状

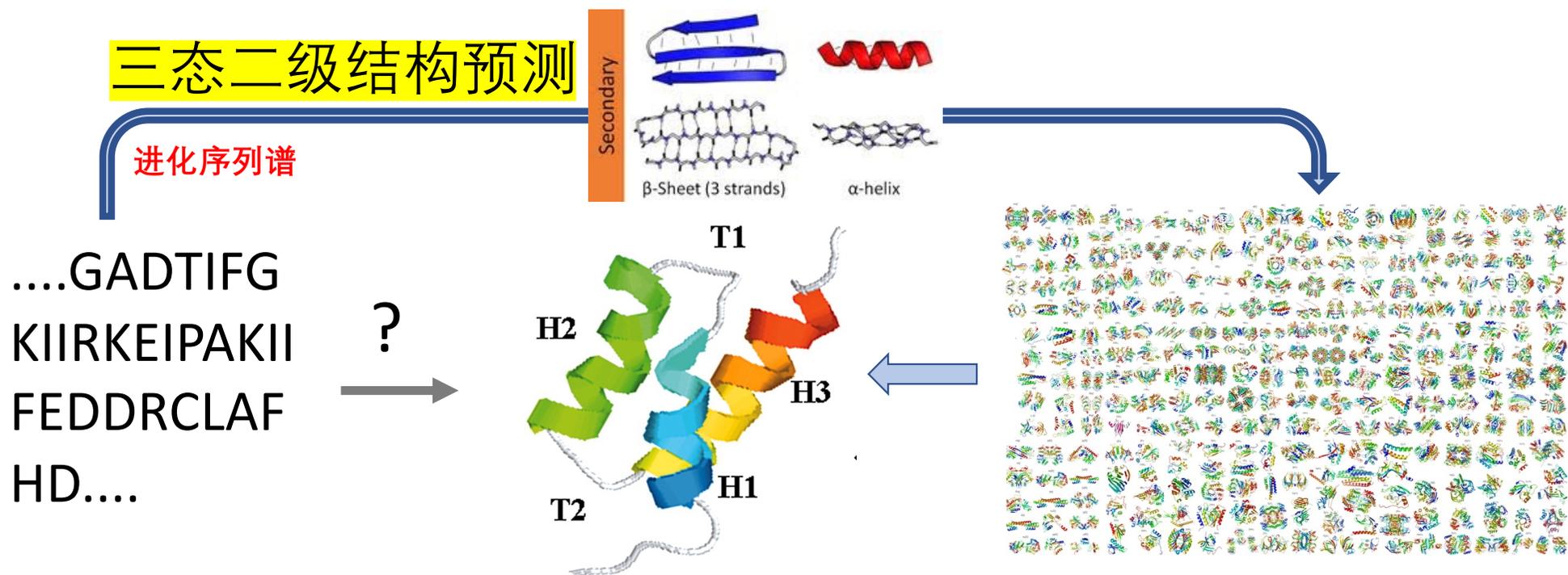
支链的填充

三级结构



四级结构：蛋白复合物

蛋白质结构计算预测

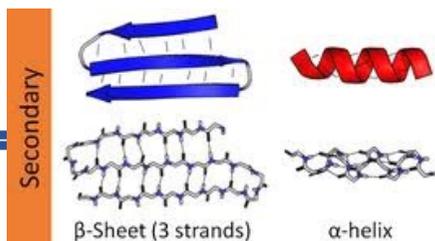


SPINE, 2007– 80% **SPIDER 2**, 2015 – 82%, **SPIDER 3**, 2017 – 84%
(Regular NN) (Deep Learning NN) (LSTM-BRNN)

蛋白质结构计算预测

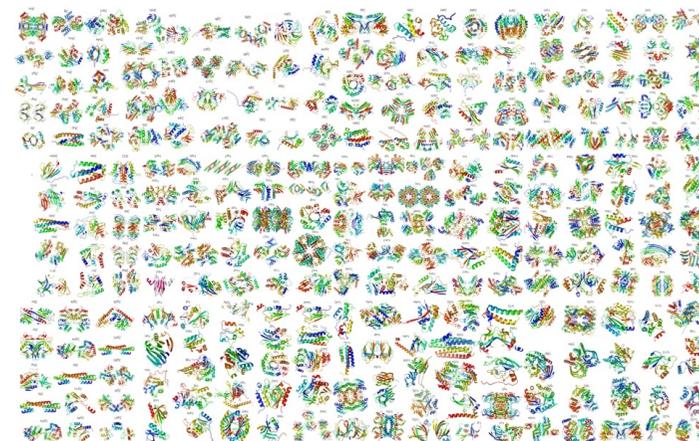
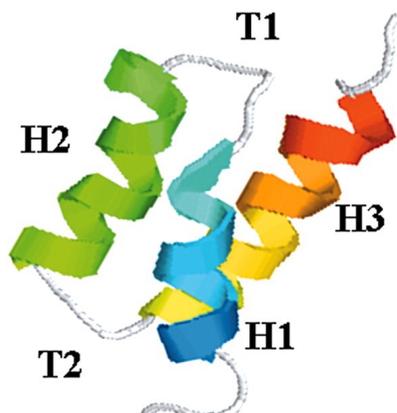
三态二级结构预测

进化序列谱

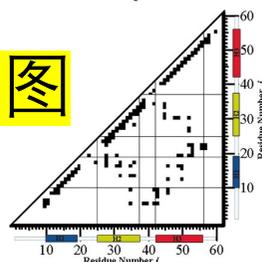


....GADTIFG
KIIRKEIPAKII
FEDDRCLAF
HD....

?



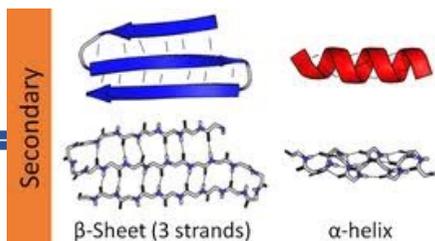
二态氨基酸接触图



蛋白质结构计算预测

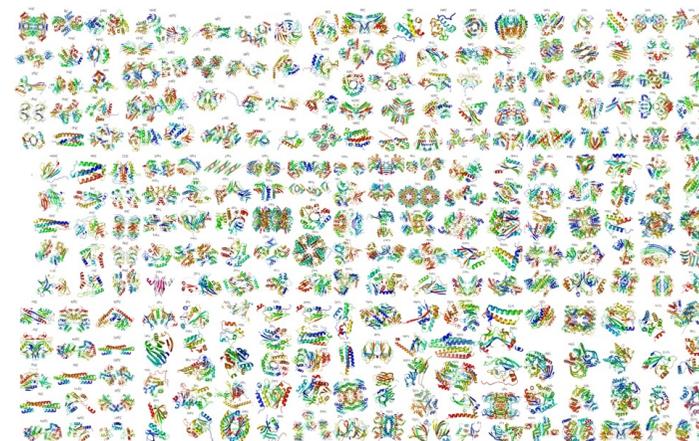
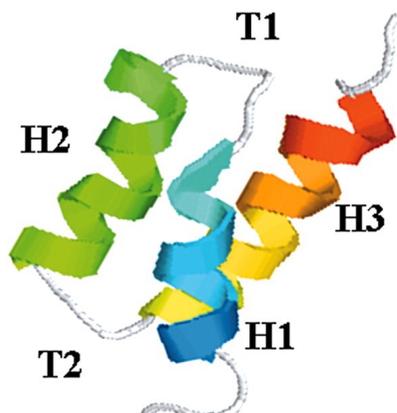
三态二级结构预测

进化序列谱



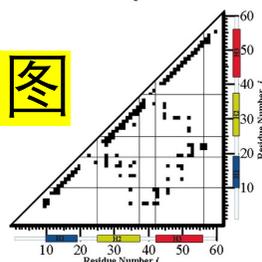
....GADTIFG
KIIRKEIPAKII
FEDDRCLAF
HD....

?



二态氨基酸接触图

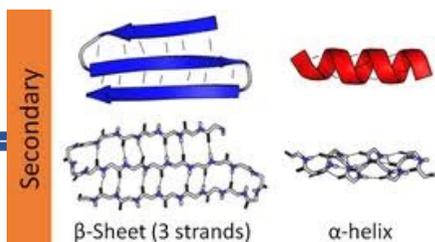
共进化直接相关图谱



蛋白质结构计算预测

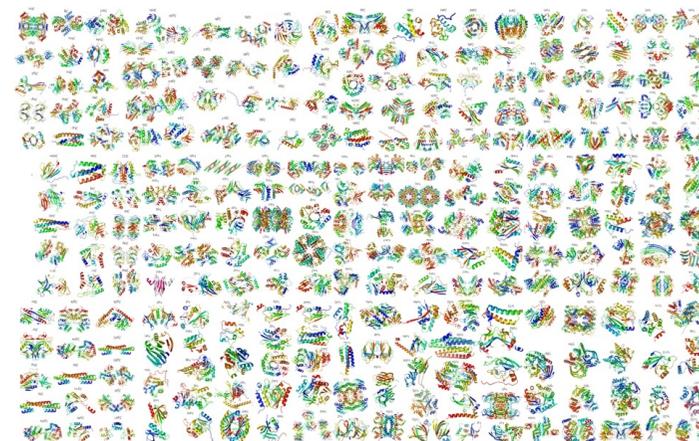
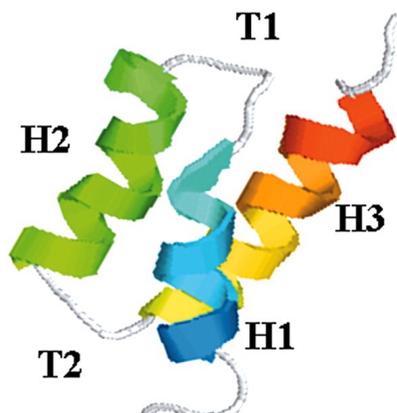
三态二级结构预测

进化序列谱



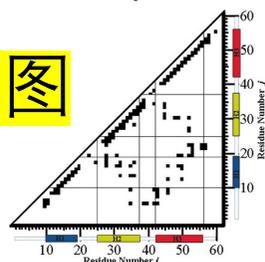
....GADTIFG
KIIRKEIPAKII
FEDDRCLAF
HD....

?



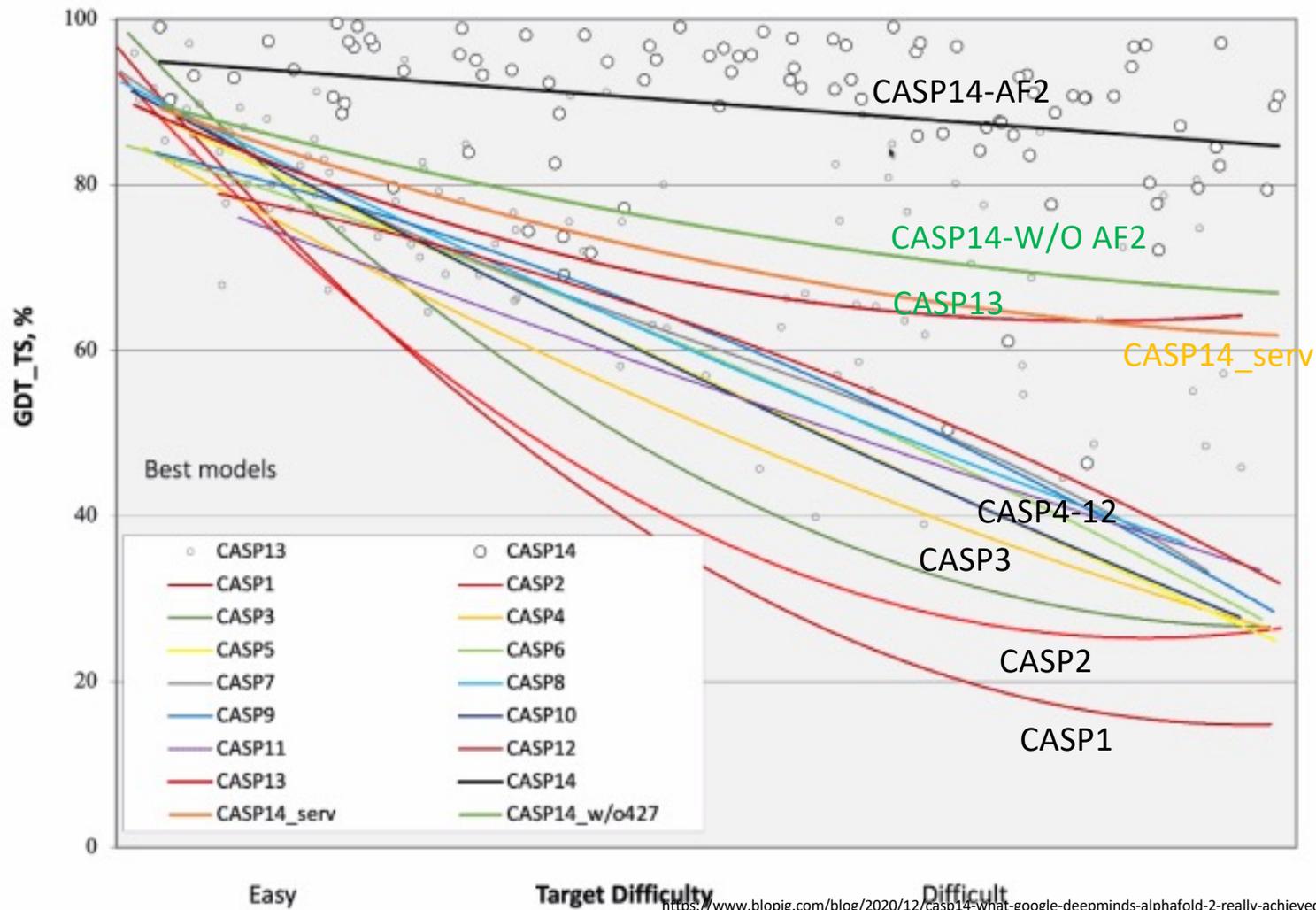
二态氨基酸接触图

共进化直接相关图谱



Pre-2020主流方法

2020年底AlphaFold革命

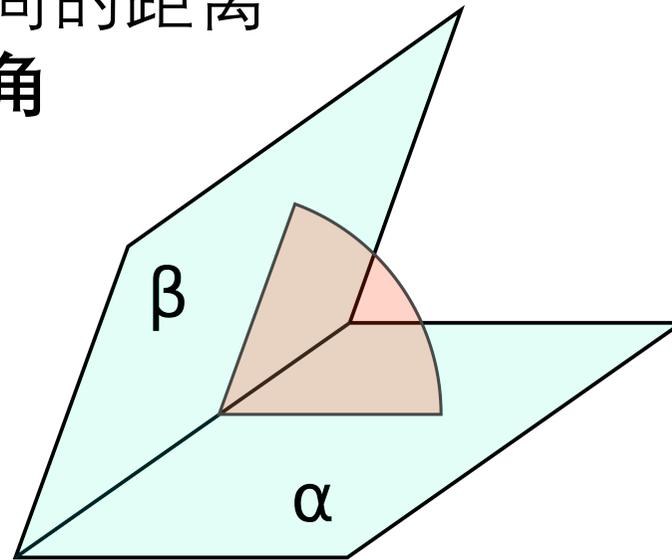


到底发生了什么？

蛋白质结构预测的边角故事

边：原子间的距离

角：二面角



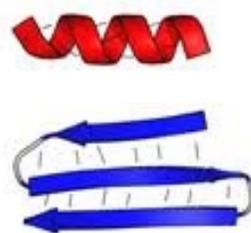
主链结构的预测

主链结构预测 = 二级结构预测

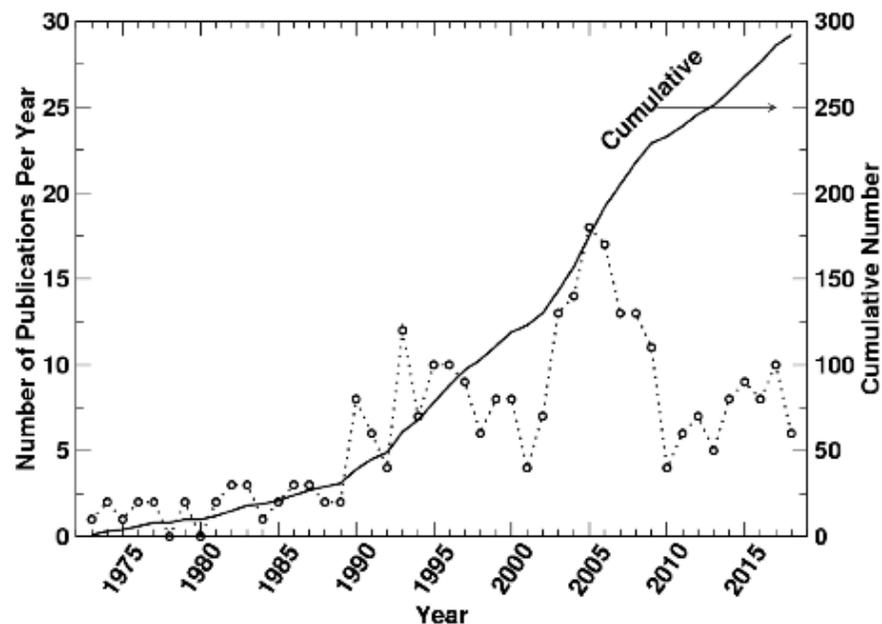
>300种方法，70年的历史

....GADTIFG
KIIRKEIPAKII
FEDDRCLAF
HD....

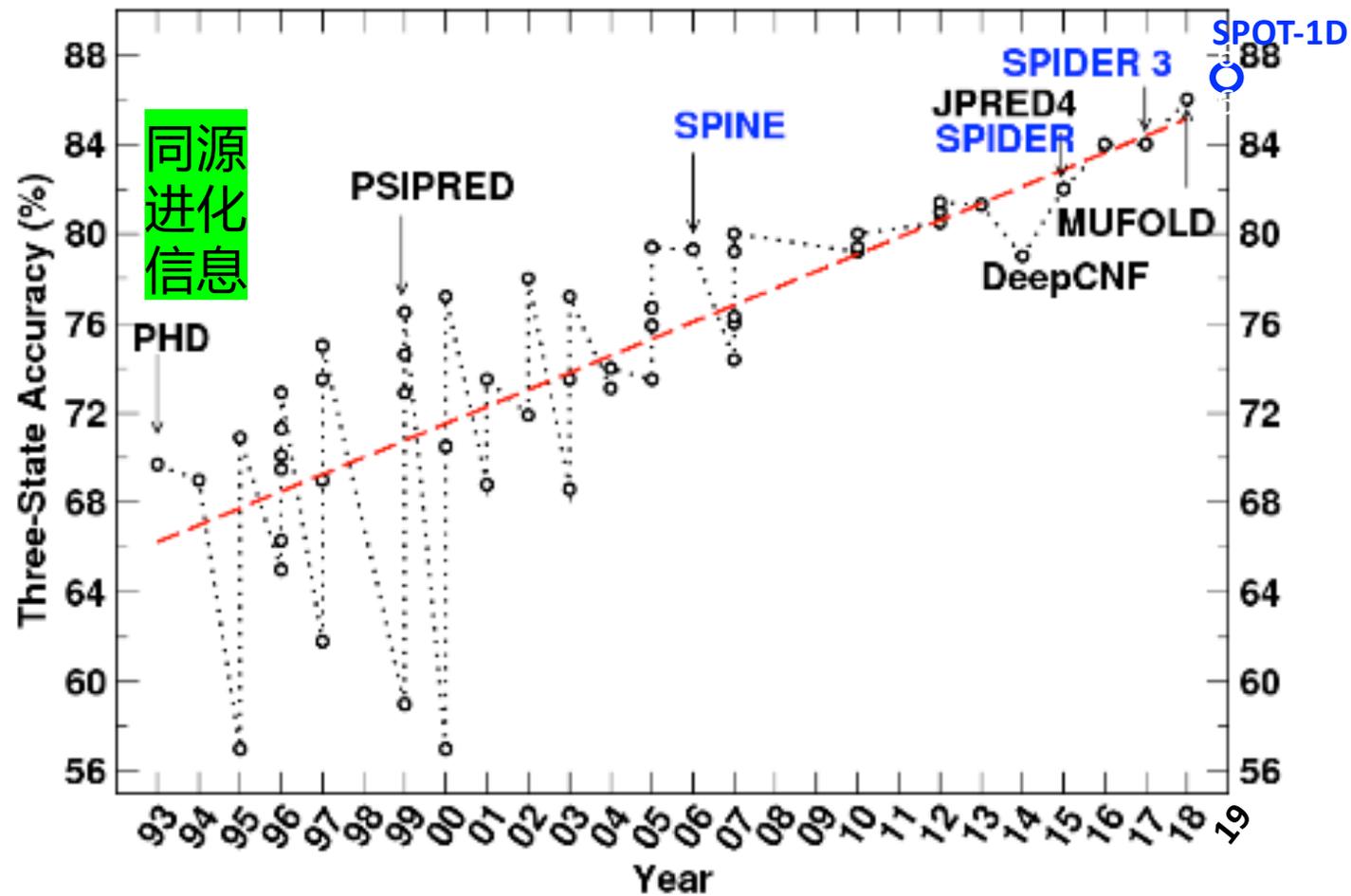
三态预测



Helix, Sheet, Coil



蛋白质三态二级结构预测



蛋白质三态二级结构预测

问题：

- 粗粒化
- 人为的分类，分界线不清不楚
- 不存在理想的螺旋、片条结构
- 对无规线圈的结构一无所知



蛋白质DSSP八态二级结构预测

问题：

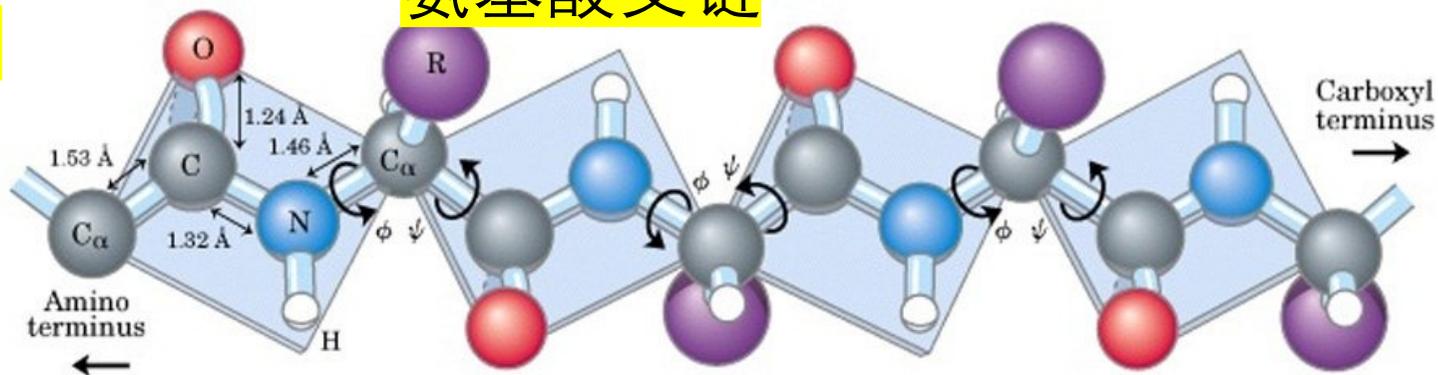
- 还是粗粒化
- 几乎没有理想八态结构
- 人为的分类

蛋白质主链的结构由二面角决定

主链二面角： ϕ , ψ , ω 决定整个主链的走向, $\omega = 180^\circ$

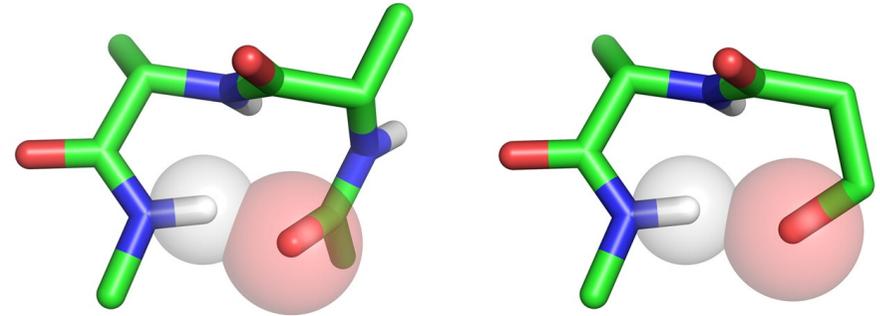
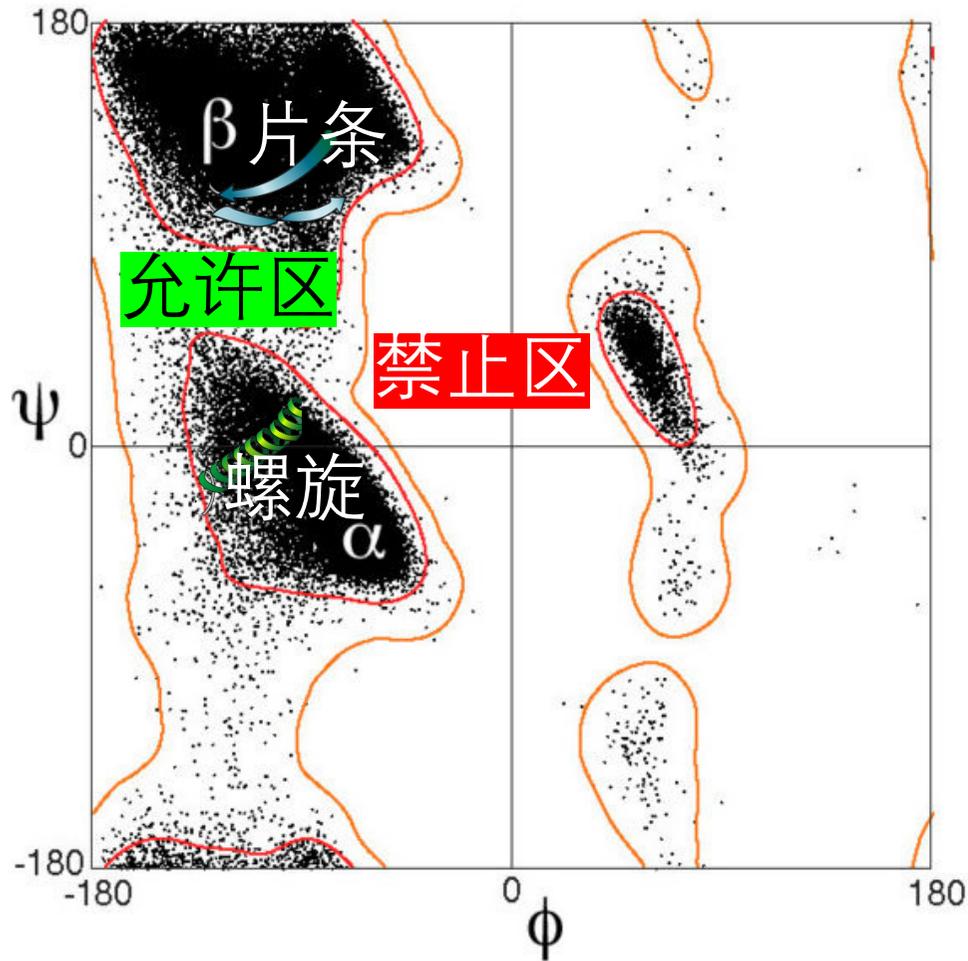
主链

氨基酸支链

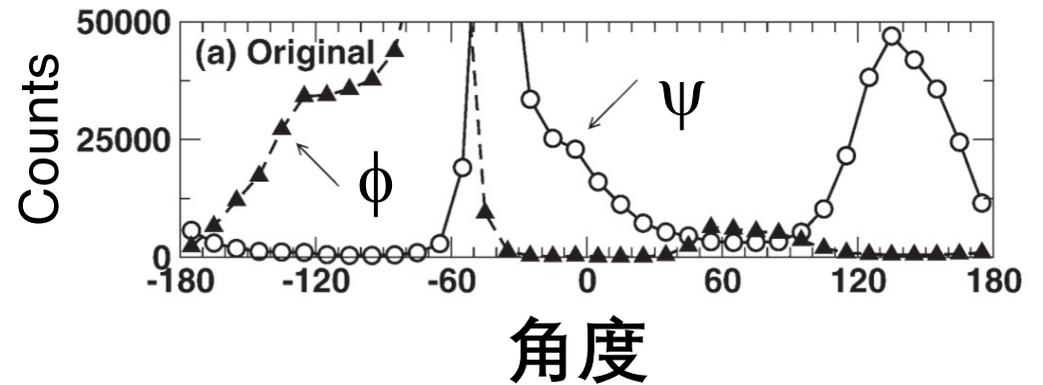


- ϕ , the rotational angle about the C_α -N bond
- ψ , the rotational angle about the C- C_α bond
- ω , the rotational angle about the N-C bond

主链二面角 ϕ 和 ψ 的分布



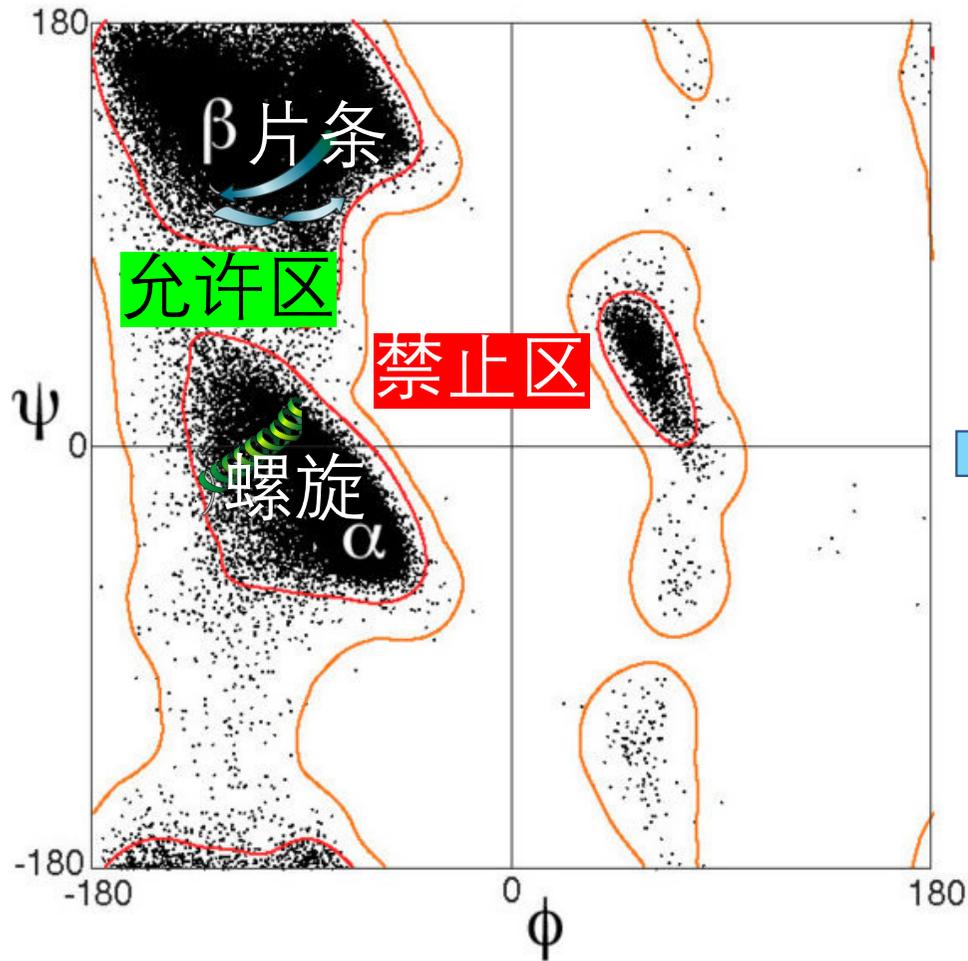
PNAS, 108, 109-113 (2010)



角度的周期性： $-180^\circ = 180^\circ$

主链二面角的分区预测

Sternberg et al. J. Mol. Biol. 1997



	A=a	B=b+p	G=g+l+G+U+N		E=e+M+E				
π	b	b	b	p	o	M	e	e	e
b	b	b	b	p	o	M	M	e	E
b	b	b	b	p	*	l	g	s	E
ψ	a	a	a	T	*	l	g	g	N
N	a	a	a	a	*	U	g	g	N
N	a	a	a	a	*	U	g	g	N
l	a	a	a	a	*	G	G	G	l
E	F	F	F	F	O	e	e	e	E
$-\pi$	b	b	b	p	O	e	e	e	e
	$-\pi$			ϕ					π

主链二面角的分区预测

优点：

- 通过角度可以直接建立主链结构
- 一些角度区域可以不在预测区域范围

问题：

- 人为的分类 Q4=74% (Bystroff et al., 2000)
- 分区越多，预测越不准确 Q4=77% (Kuang et al., 2004)
- Q3=79% (Zimmermann, 2008)

和3态二级结构预测比：没有优势

是不是可以进行真实主链二面角预测？

优点：

- 不需要人为的划分区域
- 通过角度可以直接建立高精度主链结构

创新点：

- 把分类问题（classification）变成回归问题（regression）

真实主链二面角预测



PROTEINS: Structure, Function, and Bioinformatics 68:76-81 (2007)

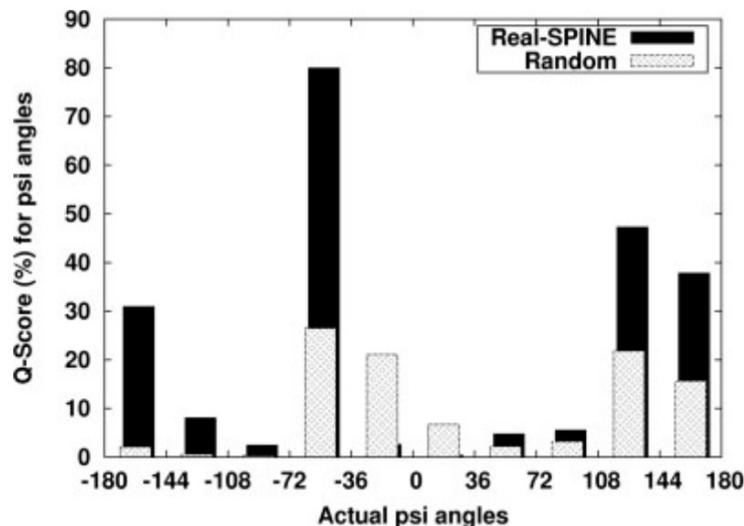
Proteins 68:76-81 (2007).

Real-SPINE: An Integrated System of Neural Networks for Real-Value Prediction of Protein Structural Properties

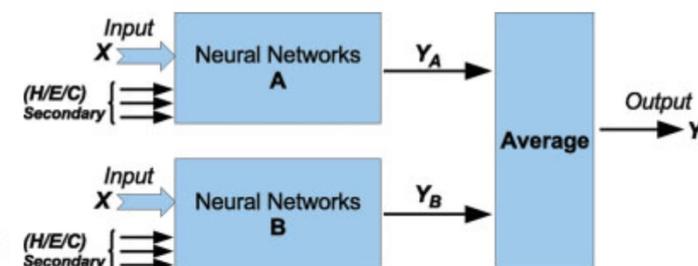
Ofer Dor¹ and Yaoqi Zhou^{1,2*}

¹Department of Physiology and Biophysics, Howard Hughes Medical Institute Center for Single Molecule Biophysics, State University of New York at Buffalo, Buffalo, New York 14214

²Indiana University School of Informatics, Indiana University-Purdue University and Center for Computational Biology and Bioinformatics, Indiana University School of Medicine, Indianapolis 46202



预测 ψ



One hidden layer, 200 hidden units

平均预测 ψ 角度误差 = 54°



真实主链二面角预测

Proteins 72:427-433 (2008).

Real-value prediction of backbone torsion angles

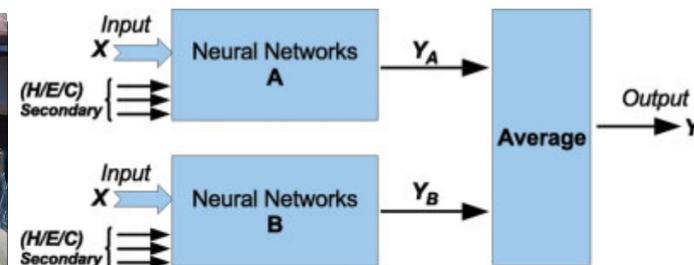
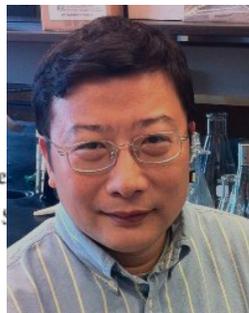
Bin Xue,^{1,2} Ofer Dor,³ Eshel Faraggi,^{1,2} and Yaoqi Zhou^{1,2*}

¹ Indiana University School of Informatics, Indiana University-Purdue University

² Center for Computational Biology and Bioinformatics, Indiana University

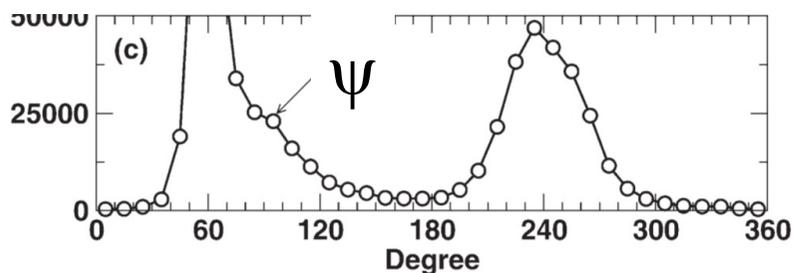
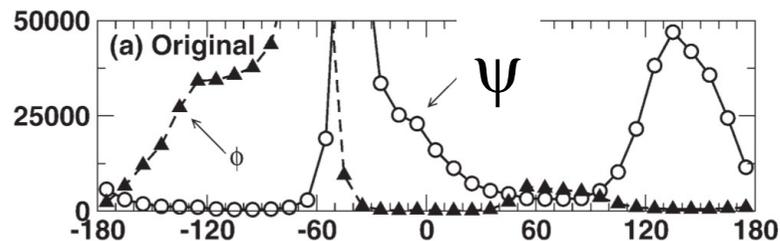
³ Faculty of Engineering, Tel Aviv University, Tel Aviv 69978, Israel

ψ/ϕ 同时预测

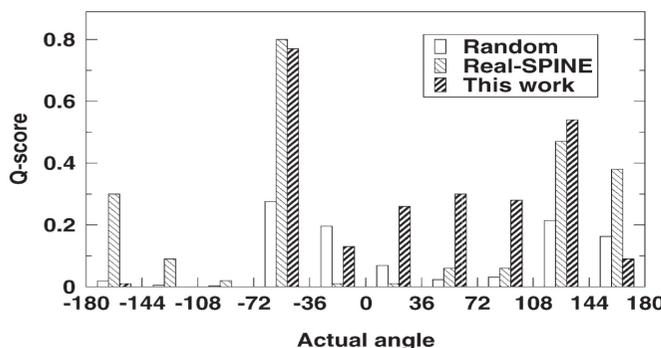


54° Real-SPINE
2007 (1层)

ψ 误差 38° Real-SPINE2
2008 (1层)



同样的神经网络，同样的输入特征，简单的角度位移 (+180+100)



真实主链二面角预测

Proteins 72:427-433 (2008).

Real-value prediction of backbone torsion angles

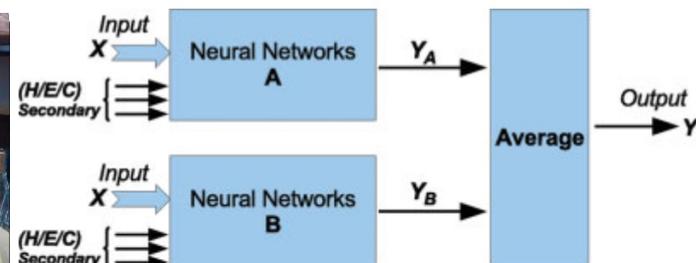
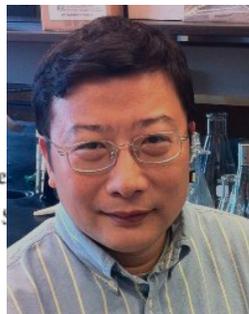
Bin Xue,^{1,2} Ofer Dor,³ Eshel Faraggi,^{1,2} and Yaoqi Zhou^{1,2*}

¹ Indiana University School of Informatics, Indiana University-Purdue University

² Center for Computational Biology and Bioinformatics, Indiana University

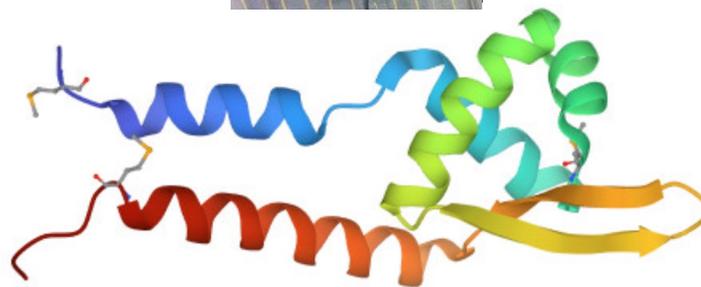
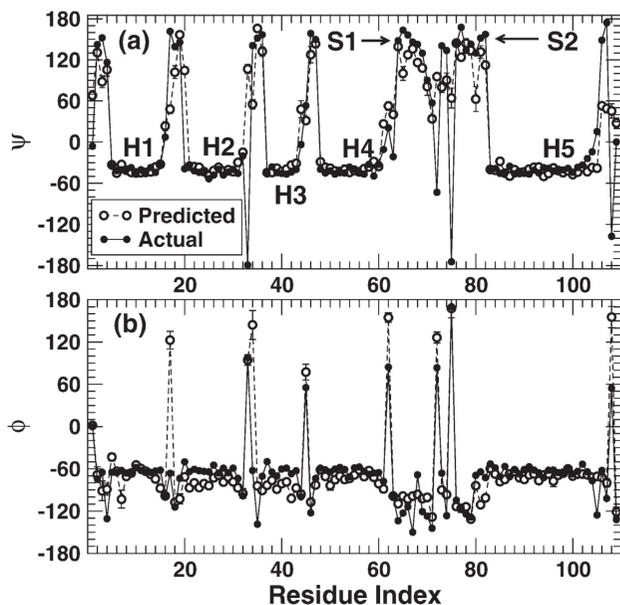
³ Faculty of Engineering, Tel Aviv University, Tel Aviv 69978, Israel

ψ/φ同时预测



54° Real-SPINE
2007 (1层)

ψ 误差 38° Real-SPINE2
2008 (1层)



误差从54° 下降到38°
(下降30%)。



深圳湾实验室
Shenzhen Bay Laboratory

真实主链二面角预测

Proteins 74:847-856 (2009).

Improving the prediction accuracy of residue solvent accessibility and real-value backbone torsion angles of proteins by guided-learning through a two-layer neural network

Eshel Faraggi,^{1,2} Bin Xue,^{1,2} and Yaoqi Zhou^{1,2*}

¹Indiana University School of Informatics, Indiana University-Purdue University, Indianapolis, IN 46202

²Center for Computational Biology and Bioinformatics, Indiana University School of Medicine, 719 Indiana Ave., Walker Plaza Building Suite 319, Indianapolis, IN 46202, USA



Guided learning

→ 远处氨基酸, 更小贡献

2 hidden layers

54° Real-SPINE
2007 (1层)



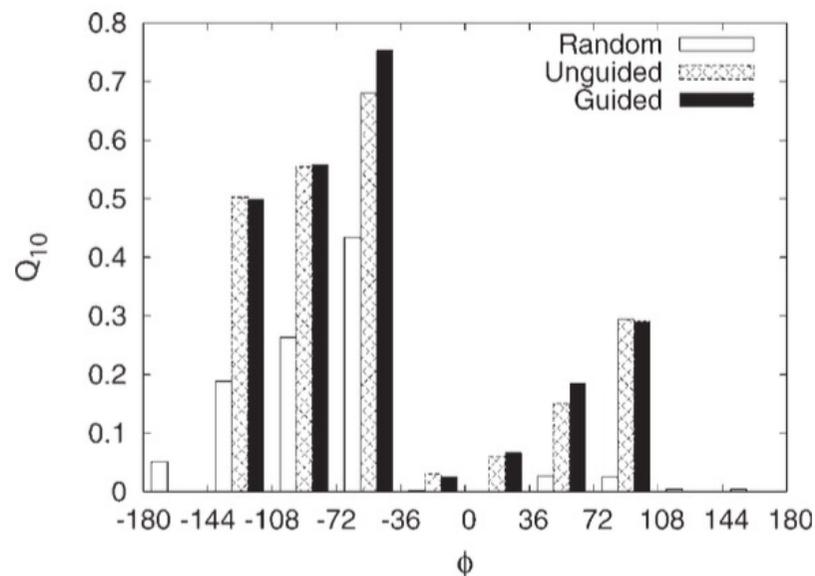
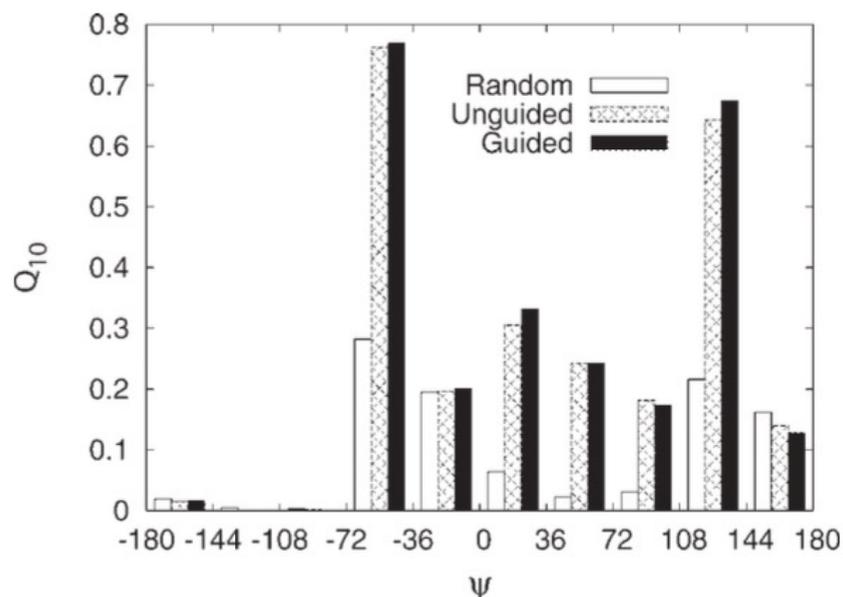
ψ 误差 38° Real-SPINE2
2008 (1层)



37°



36° Real-SPINE3
2009 (2层)



真实主链二面角预测

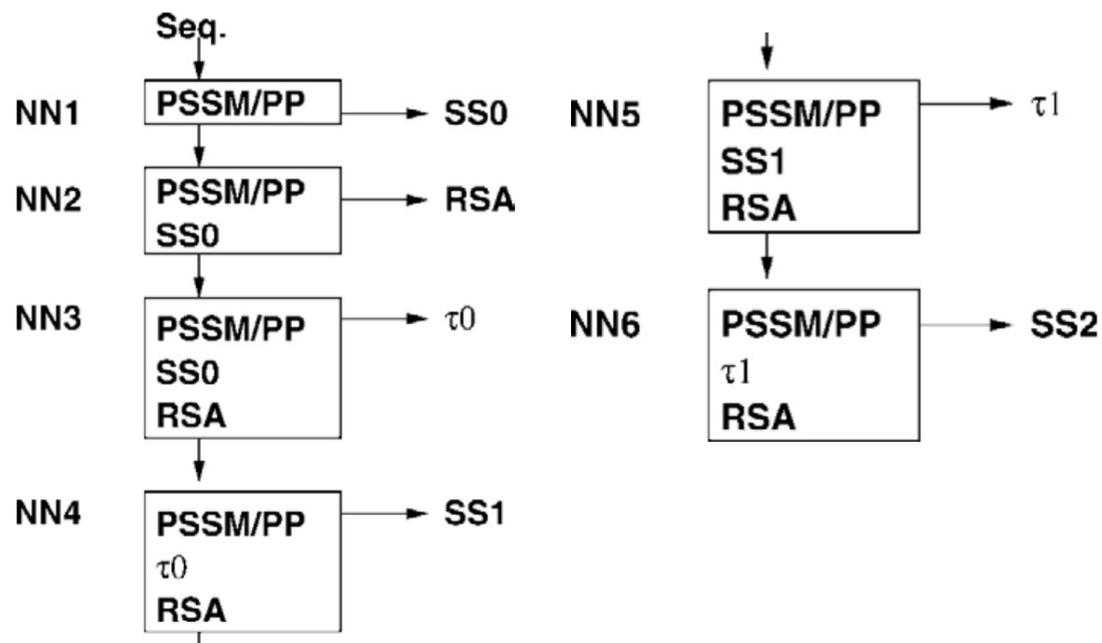
J. Compu Chem 33:259-267 (2012).

ORIGINAL ARTICLES

DOI 10.1002/jcc.21968

SPINE X: Improving Protein Secondary Structure Prediction by Multistep Learning Coupled with Prediction of Solvent Accessible Surface Area and Backbone Torsion Angles

Eshel Faraggi,^[a,b] Tuo Zhang,^[a,b] Yuedong Yang,^[a,b] Lukasz Kurgan,^[b,c] and Yaoqi Zhou*^[a,b]



迭代改进

54° Real-SPINE
2007 (1层)

↓

ψ 误差 38° Real-SPINE2
2008 (1层)

↓

37°

↓

36° Real-SPINE3
2009 (2层)

↓

35° SPINE X
2012 (2层)

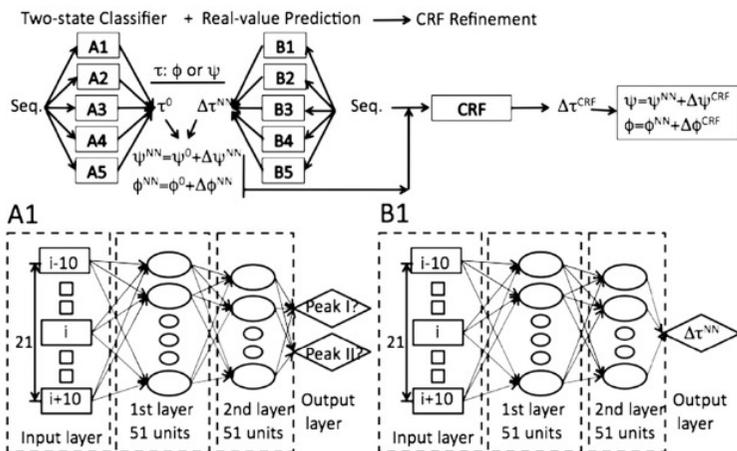
真实主链二面角预测

Structure
Article

Structure 17:1515-1527 (2009).

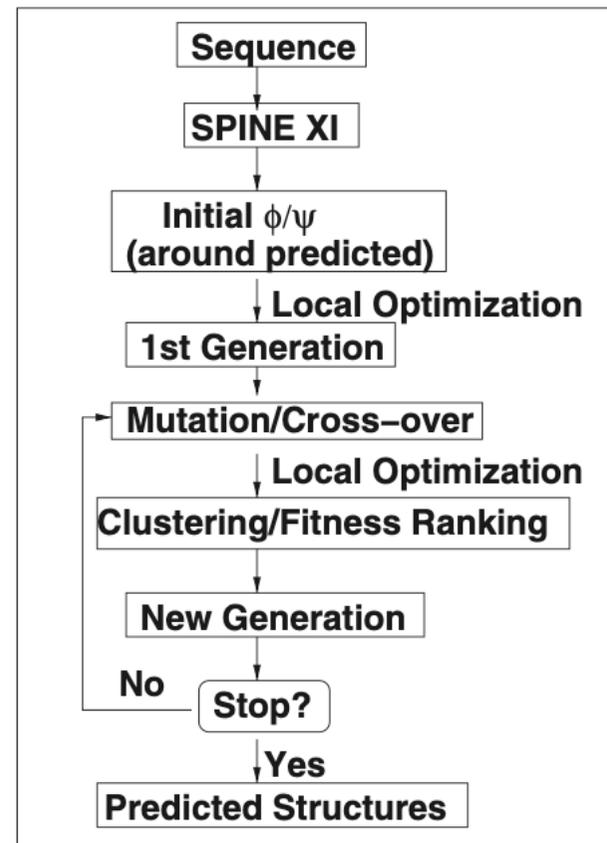
Predicting Continuous Local Structure and the Effect of Its Substitution for Secondary Structure in Fragment-Free Protein Structure Prediction

Eshel Faraggi,^{1,3} Yuedong Yang,^{1,3} Shesheng Zhang,¹ and Yaoqi Zhou^{1,2,*}



- Predict Angle Peak first and then the fluctuation around the peak.
- Using angles as restraints for structure prediction

54° Real-SPINE 2007 (1层)
 ↓
 ψ 误差 38° Real-SPINE2 2008 (1层)
 ↓
 37°
 ↓
 36° Real-SPINE3 2009 (2层)
 ↓
 35° SPINE X 2012 (2层)
 ↓
 SPINE XI 2009 (2层)



真实主链二面角预测

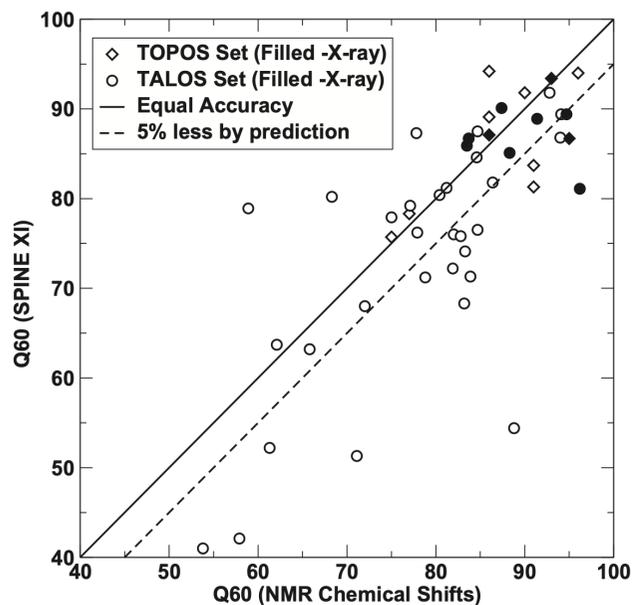
Structure
Article

Structure 17:1515-1527 (2009).

Predicting Continuous Local Structure and the Effect of Its Substitution for Secondary Structure in Fragment-Free Protein Structure Prediction

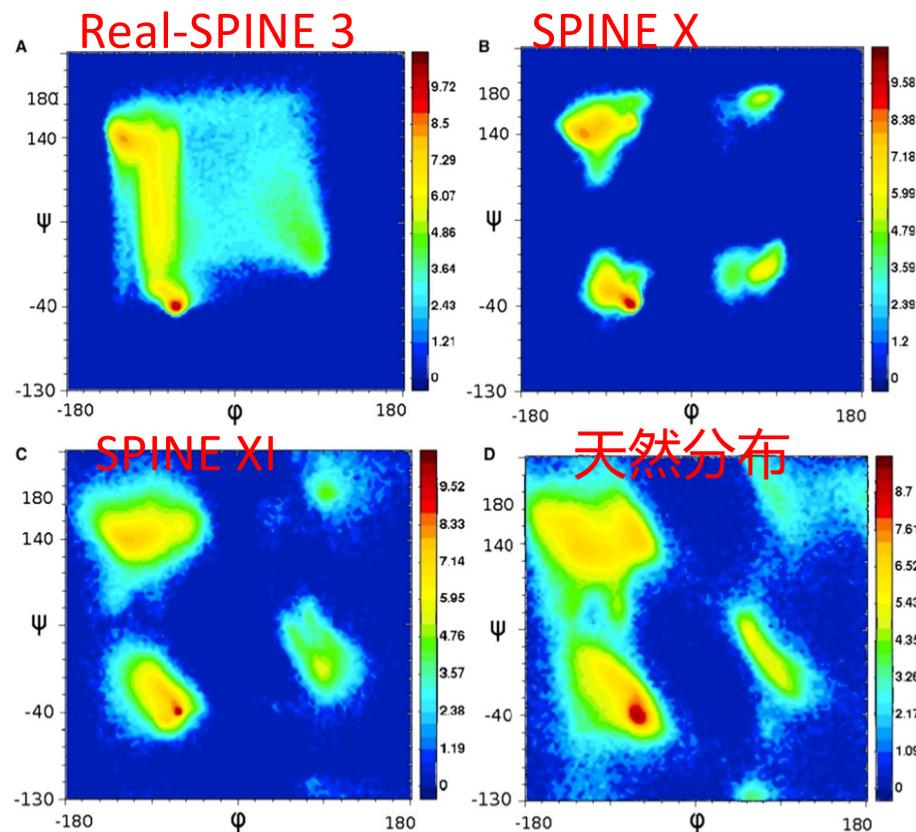
Eshel Faraggi,^{1,3} Yuedong Yang,^{1,3} Shesheng Zhang,¹ and Yaoqi Zhou^{1,2,*}

和NMR化学位移推导的结果比：几乎一样



Cell
PRESS

预测的角度分布与真实的比：几乎一样



54° Real-SPINE
2007 (1层)



ψ 误差 38° Real-SPINE2
2008 (1层)



37°



36° Real-SPINE3
2009 (1层)



35° SPINE X
2012 (2层)



33° SPINE XI
2009 (2层)



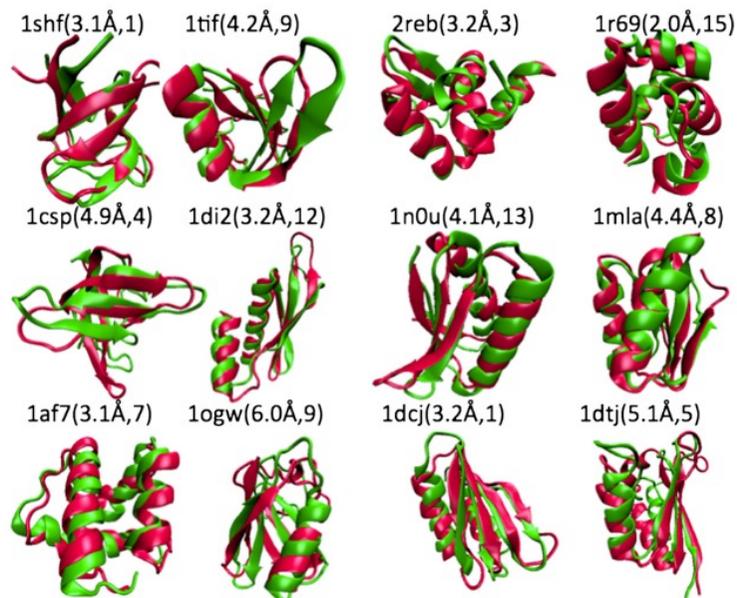
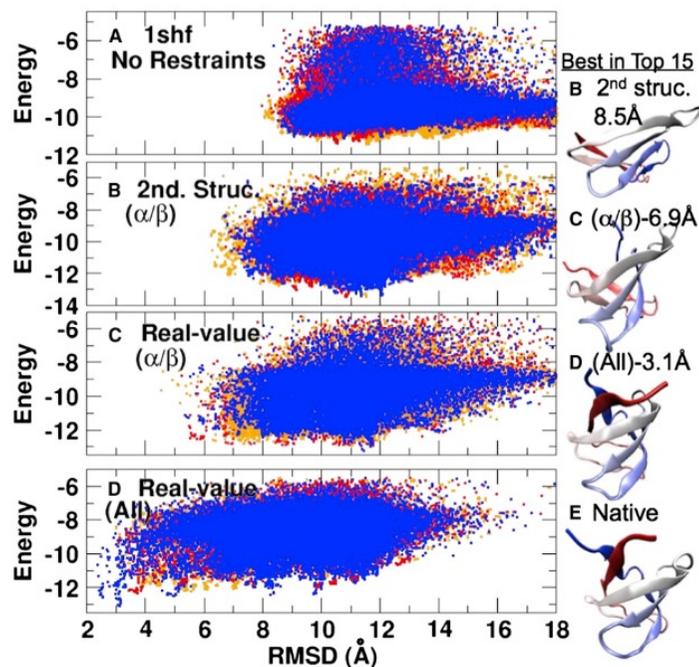
真实主链二面角预测

Structure
Article

Structure 17:1515-1527 (2009).

Predicting Continuous Local Structure and the Effect of Its Substitution for Secondary Structure in Fragment-Free Protein Structure Prediction

Eshel Faraggi,^{1,3} Yuedong Yang,^{1,3} Shesheng Zhang,¹ and Yaoqi Zhou^{1,2,*}



54° Real-SPINE 2007 (1层)
 ↓
 ψ 误差 38° Real-SPINE2 2008 (1层)
 ↓
 37°
 ↓
 36° Real-SPINE3 2009 (1层)
 ↓
 35° SPINE X 2012 (2层)
 ↓
 33° SPINE XI 2009 (2层)

真实主链二面角预测

SCIENTIFIC REPORTS

Sci Rep 5: 11476 (2015).

OPEN

Improving prediction of secondary structure, local backbone angles, and solvent accessible surface area of proteins by iterative deep learning

Received: 04 March 2015
Accepted: 19 May 2015
Published: 22 June 2015

Rhys Heffernan¹, Kuldip Paliwal¹, James Lyons¹, Abdollah Dehzangi^{1,2}, Alok Sharma^{2,3}, Jihua Wang⁴, Abdul Sattar^{2,5}, Yuedong Yang⁶ & Yaoqi Zhou^{6*}

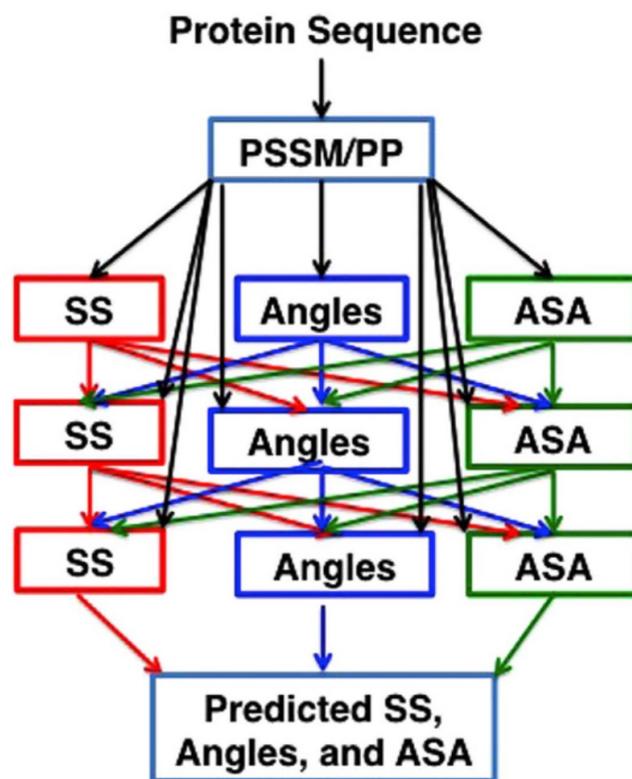
- 深度学习 (3个隐藏层)
- 预测 Sin(Angle), Cos(Angle)
- ARCTAN



Iteration 1

Iteration 2

Iteration 3



	54°	Real-SPINE 2007 (1层)
ψ 误差	↓	
	38°	Real-SPINE2 2008 (1层)
	↓	
	37°	
	↓	
	36°	Real-SPINE3 2009 (2层)
	↓	
	35°	SPINE X 2012 (2层)
	↓	
	33°	SPINE XI 2009 (2层)
	↓	
	30°	SPIDER 2 2015 (3层)

真实主链二面角预测

SCIENTIFIC REPORTS

Sci Rep 5: 11476 (2015).

OPEN

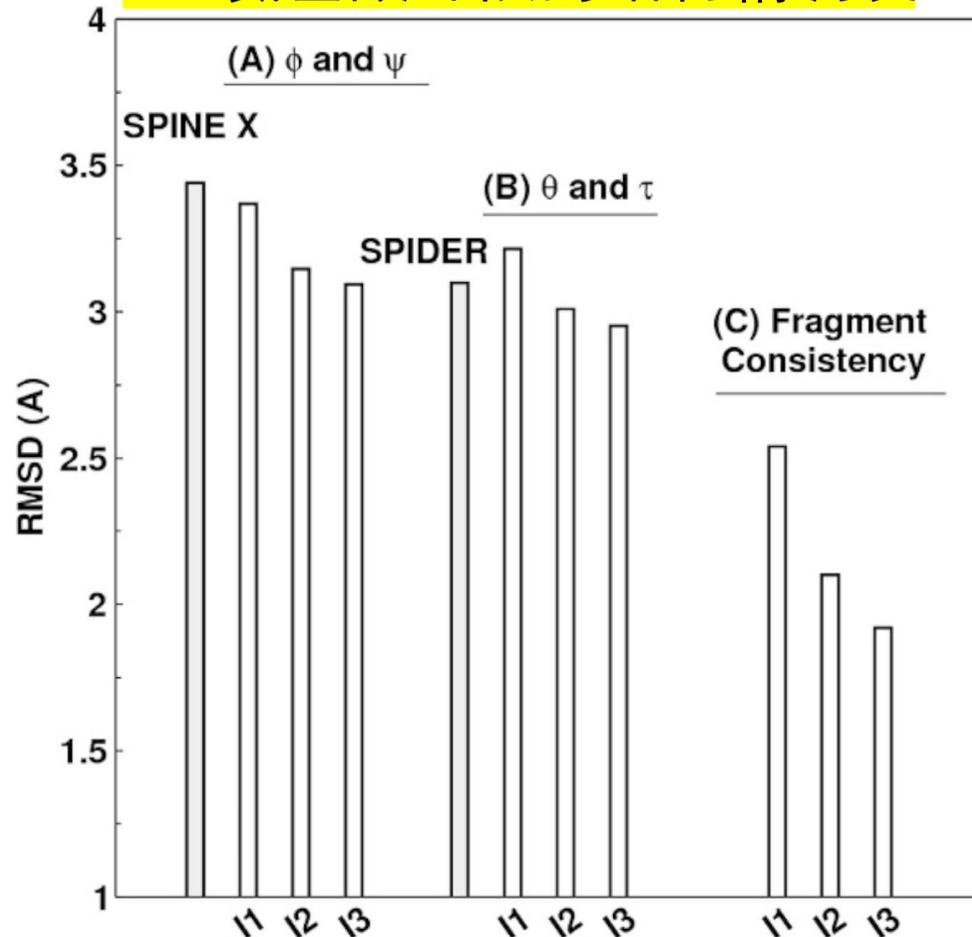
Improving prediction of secondary structure, local backbone angles, and solvent accessible surface area of proteins by iterative deep learning

Received: 04 March 2015
Accepted: 19 May 2015
Published: 22 June 2015

Rhys Heffernan¹, Kuldip Paliwal², James Lyons¹, Abdollah Dehzangi^{1,2}, Alok Sharma^{2,3}, Jihua Wang⁴, Abdul Sattar^{2,5}, Yuedong Yang⁶ & Yaoqi Zhou^{6*}

- 深度学习 (3个隐藏层)
- 预测 Sin(Angle), Cos(Angle)
→ ARCTAN

15-氨基酸片段的结构精确度



54° Real-SPINE 2007 (1层)
↓
ψ 误差 38° Real-SPINE2 2008 (1层)
↓
37°
↓
36° Real-SPINE3 2009 (2层)
↓
35° SPINE X 2012 (2层)
↓
33° SPINE XI 2009 (2层)
↓
30° SPIDER 2 2015 (3层)

真实主链二面角预测

SCIENTIFIC REPORTS

Sci Rep 5: 11476 (2015).

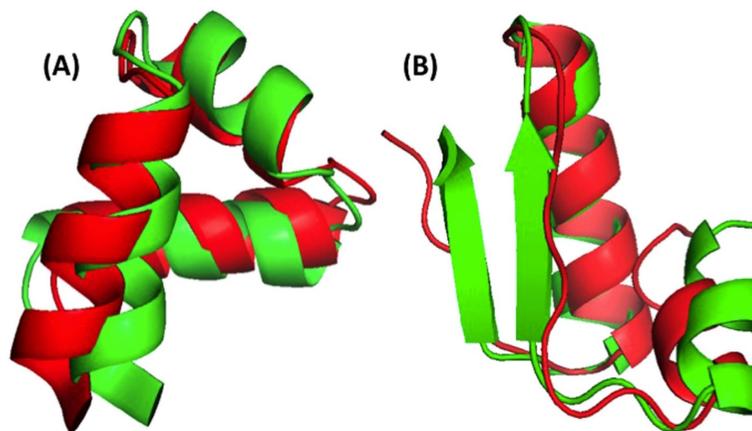
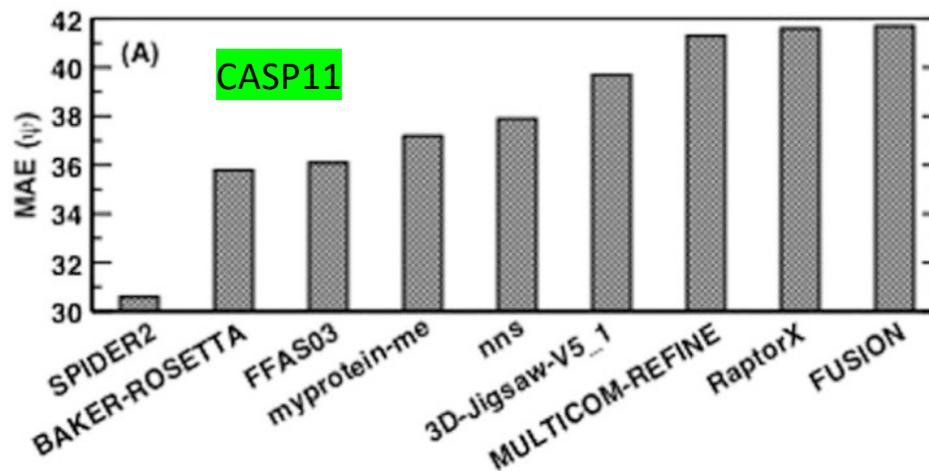
OPEN

Improving prediction of secondary structure, local backbone angles, and solvent accessible surface area of proteins by iterative deep learning

Received: 04 March 2015
Accepted: 19 May 2015
Published: 22 June 2015

Rhys Heffernan¹, Kuldip Paliwal², James Lyons¹, Abdollah Dehzangi^{1,2}, Alok Sharma^{2,3}, Jihua Wang⁴, Abdul Sattar^{2,5}, Yuedong Yang⁶ & Yaoqi Zhou^{6*}

- 深度学习 (3个隐藏层)
- 预测 Sin(Angle), Cos(Angle)
→ ARCTAN



54° Real-SPINE 2007 (1层)
↓
ψ 误差 38° Real-SPINE2 2008 (1层)
↓
37°
↓
36° Real-SPINE3 2009 (2层)
↓
35° SPINE X 2012 (2层)
↓
33° SPINE XI 2009 (2层)
↓
30° SPIDER 2 2015 (3层)

真实主链二面角预测

Bioinformatics, 33(18), 2017, 2842–2849
 doi: 10.1093/bioinformatics/btx218
 Advance Access Publication Date: 18 April 2017
 Original Paper

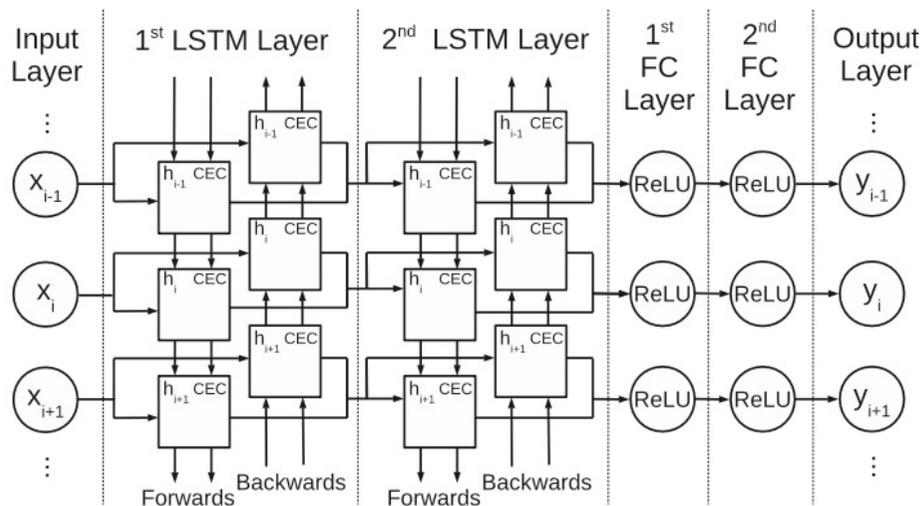


Bioinformatics 33: 2842-2849 (2017)

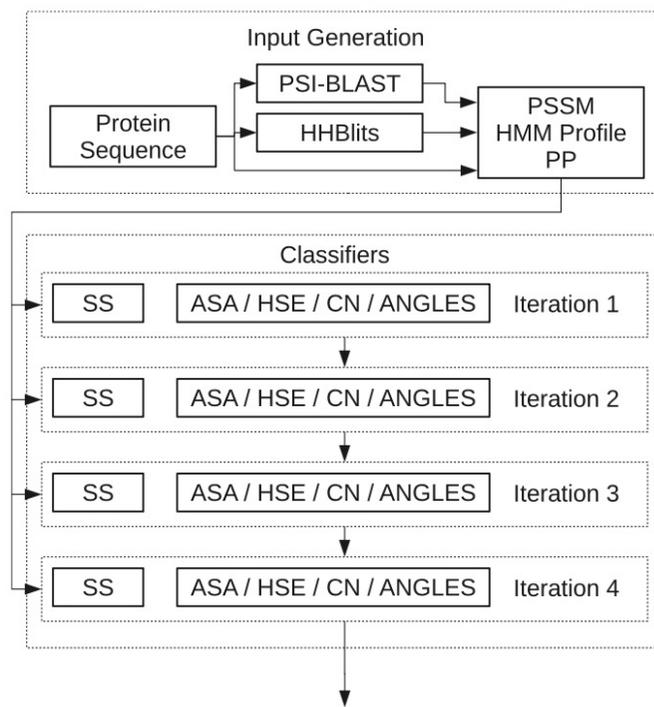
Structural bioinformatics

Capturing non-local interactions by long short-term memory bidirectional recurrent neural networks for improving prediction of protein secondary structure, backbone angles, contact numbers and solvent accessibility

Rhys Heffernan¹, Yuedong Yang^{2,*}, Kuldip Paliwal¹ and Yaoqi Zhou^{2,*}



LSTM Bidirectional neural network



- 54° Real-SPINE 2007 (1层)
- ↓
- ψ 误差 38° Real-SPINE2 2008 (1层)
- ↓
- 37°
- ↓
- 36° Real-SPINE3 2009 (2层)
- ↓
- 35° SPINE X 2012 (2层)
- ↓
- 33° SPINE XI 2009 (2层)
- ↓
- 30° SPIDER 2 2015 (3层)
- ↓
- 27° SPIDER 3 2017 (4层)

真实主链二面角预测

Bioinformatics, 33(18), 2017, 2842–2849
 doi: 10.1093/bioinformatics/btx218
 Advance Access Publication Date: 18 April 2017
 Original Paper

Bioinformatics 33: 2842-2849 (2017)

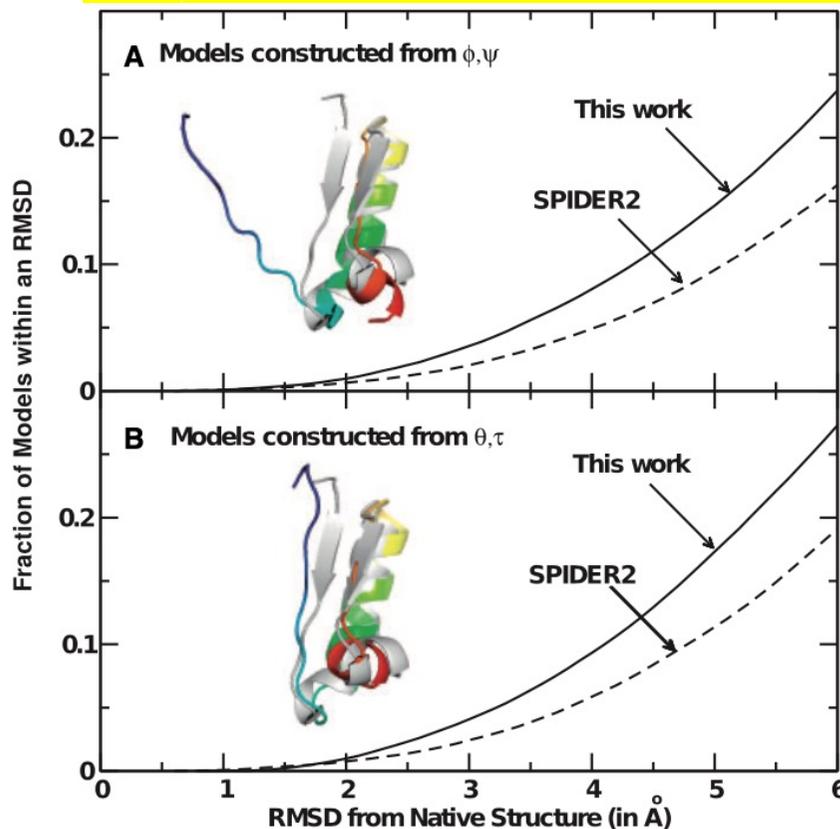
Structural bioinformatics

Capturing non-local interactions by long short-term memory bidirectional recurrent neural networks for improving prediction of protein secondary structure, backbone angles, contact numbers and solvent accessibility

Rhys Heffernan¹, Yuedong Yang^{2,*}, Kuldip Paliwal¹ and Yaoqi Zhou^{2,*}



40-氨基酸片段的结构精确度



- 54° Real-SPINE 2007 (1层)
- ↓
- ψ 误差 38° Real-SPINE2 2008 (1层)
- ↓
- 37°
- ↓
- 36° Real-SPINE3 2009 (2层)
- ↓
- 35° SPINE X 2012 (2层)
- ↓
- 33° SPINE XI 2009 (2层)
- ↓
- 30° SPIDER 2 2015 (3层)
- ↓
- 27° SPIDER 3 2017 (4层)

真实主链二面角预测

Bioinformatics 35: 2403-2410 (2019)

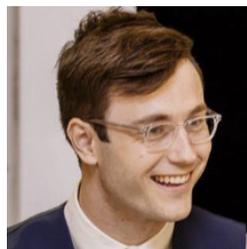
Bioinformatics, 35(14), 2019, 2403–2410
doi:10.1093/bioinformatics/bty1006
Advance Access Publication Date: 7 December 2018
Original Paper



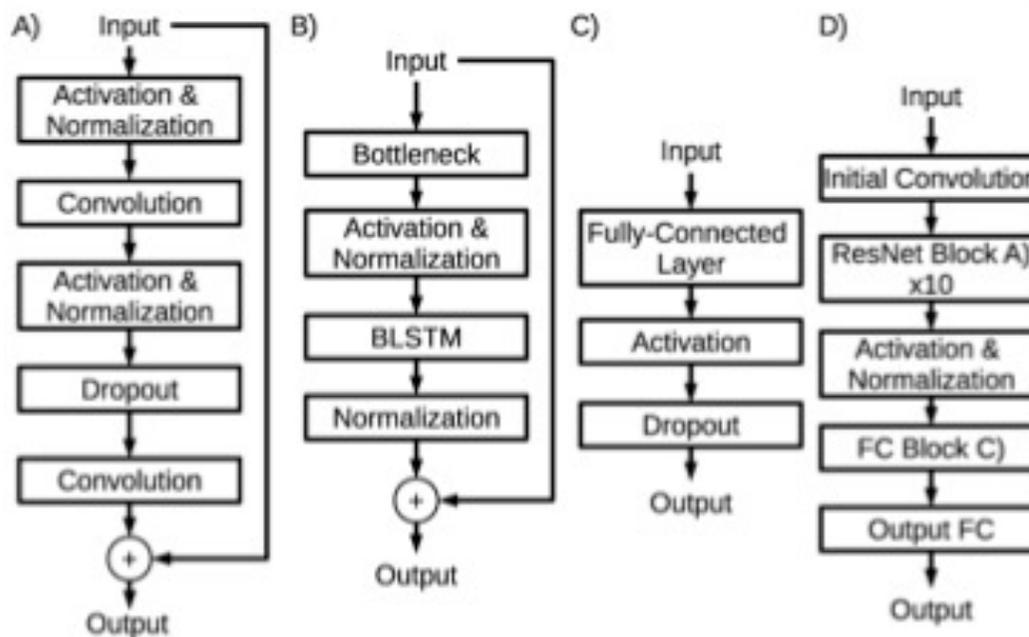
Structural bioinformatics

Improving prediction of protein secondary structure, backbone angles, solvent accessibility and contact numbers by using predicted contact maps and an ensemble of recurrent and residual convolutional neural networks

Jack Hanson^{1,*}, Kuldip Paliwal¹, Thomas Litfin², Yuedong Yang³ and Yaoqi Zhou^{2,4,*}



LSTM + ResNet Ensemble



Method	Year	Layers	ψ 误差
Real-SPINE	2007	(1层)	54°
Real-SPINE2	2008	(1层)	38°
Real-SPINE3	2009	(2层)	36°
SPINE X	2012	(2层)	35°
SPINE XI	2009	(2层)	33°
SPIDER 2	2015	(3层)	30°
SPIDER 3	2017	(4层)	27°
SPOT-1D	2019	(>10)	23°

真实主链二面角预测

Bioinformatics 35: 2403-2410 (2019)

Bioinformatics, 35(14), 2019, 2403-2410
doi:10.1093/bioinformatics/bty1006
Advance Access Publication Date: 7 December 2018
Original Paper

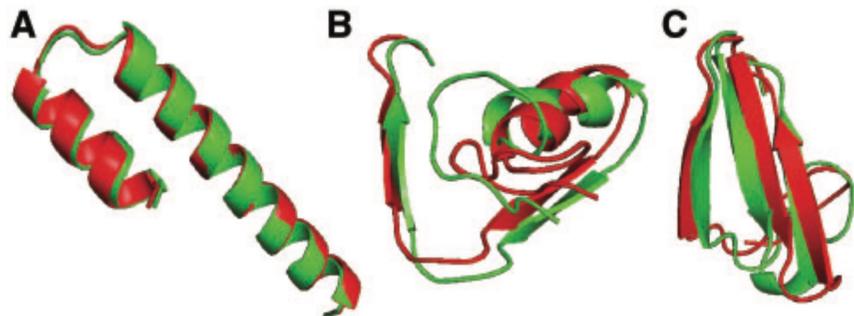


Structural bioinformatics

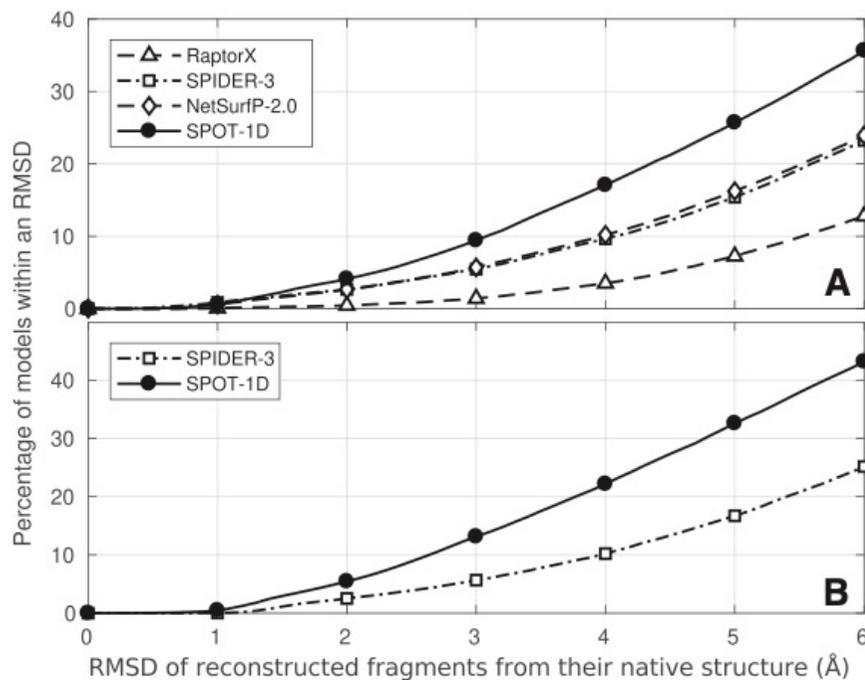
Improving prediction of protein secondary structure, backbone angles, solvent accessibility and contact numbers by using predicted contact maps and an ensemble of recurrent and residual convolutional neural networks

Jack Hanson^{1,*}, Kuldip Paliwal¹, Thomas Litfin², Yuedong Yang³
and Yaoqi Zhou^{2,4,*}

40-氨基酸片段的结构精确度

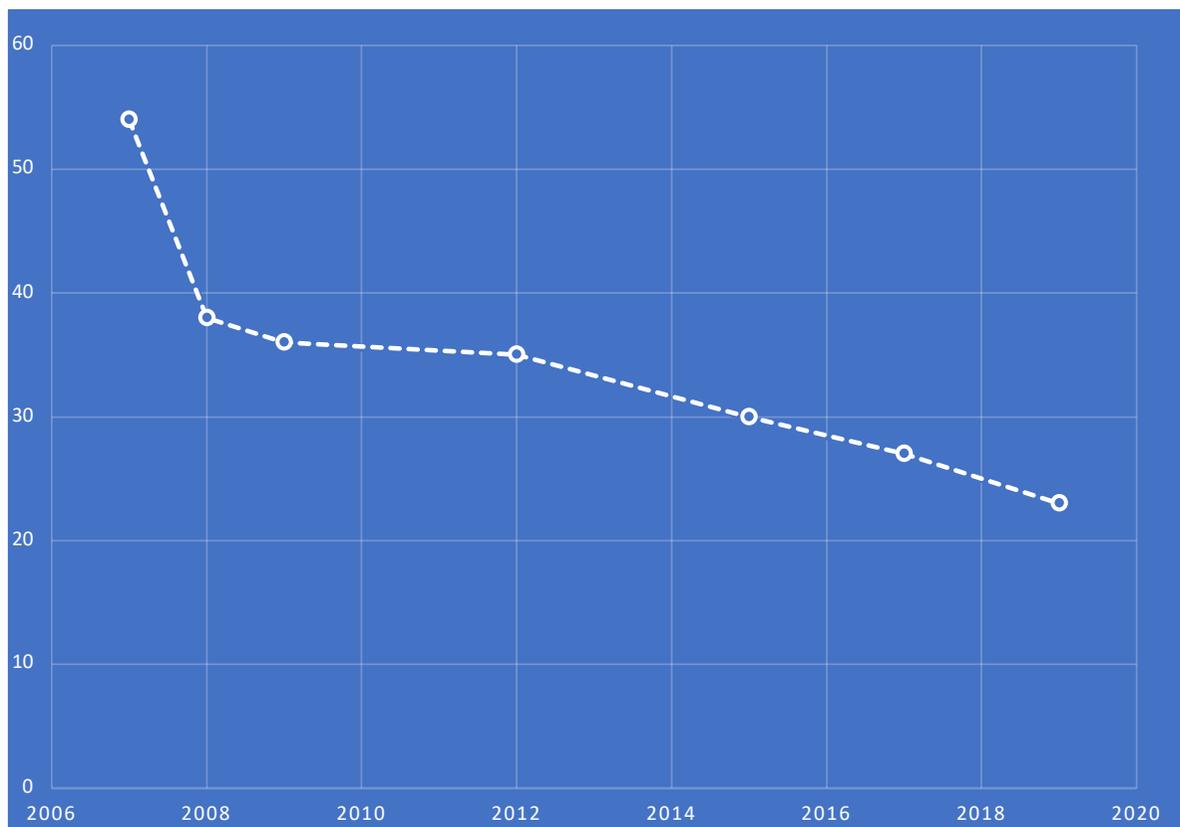


40-氨基酸片段的结构精确度



- 54° Real-SPINE 2007 (1层)
- ↓
- ψ 误差 38° Real-SPINE2 2008 (1层)
- ↓
- 37°
- ↓
- 36° Real-SPINE3 2009 (2层)
- ↓
- 35° SPINE X 2012 (2层)
- ↓
- 33° SPINE XI 2009 (2层)
- ↓
- 30° SPIDER 2 2015 (3层)
- ↓
- 27° SPIDER 3 2017 (4层)
- ↓
- 23° SPOT-1D 2019 (>10)

真实主链二面角预测



12年：从几乎无用变成了可以直接用来构建可靠的主链结构

- 54° Real-SPINE 2007 (1层)
- ↓
- ψ 误差 38° Real-SPINE2 2008 (1层)
- ↓
- 37°
- ↓
- 36° Real-SPINE3 2009 (2层)
- ↓
- 35° SPINE X 2012 (2层)
- ↓
- 33° SPINE XI 2009 (2层)
- ↓
- 30° SPIDER 2 2015 (3层)
- ↓
- 27° SPIDER 3 2017 (4层)
- ↓
- 23° SPOT-1D 2019 (>10)

蛋白质结构无结构碎片预测

Theor Chem Acc (2011) 128:3–16
DOI 10.1007/s00214-010-0799-2

FEATURE ARTICLE

Trends in template/fragment-free protein structure prediction

Yaoqi Zhou · Yong Duan · Yuedong Yang ·
Eshel Faraggi · Hongxing Lei

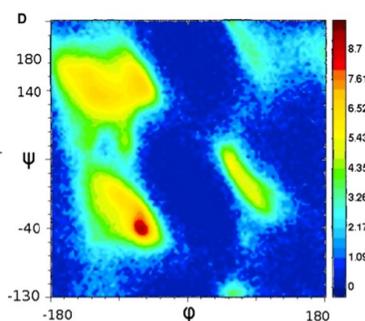
连续变化

主链二面角预测

蛋白质序列

....GADTIFG
KIIRKEIPAKII
FEDDRCLAF
HD....

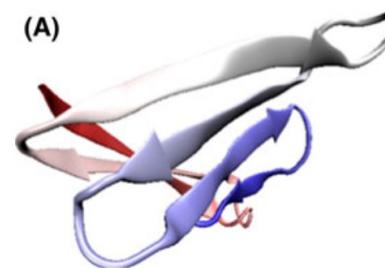
同源
进化
信息



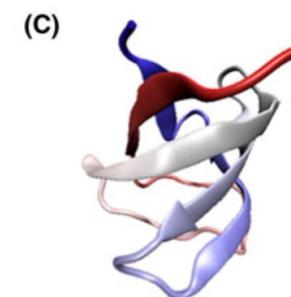
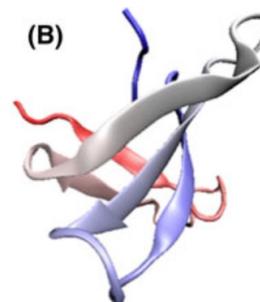
Reconstruct
From Angles
& Refinement

Angle Restraint + DFIRE energy (Top 15)

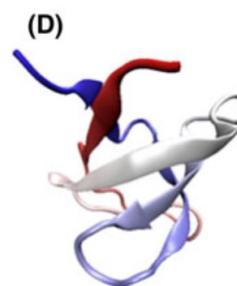
预测的二级结构



预测的片条
区域二面角



预测的所有角

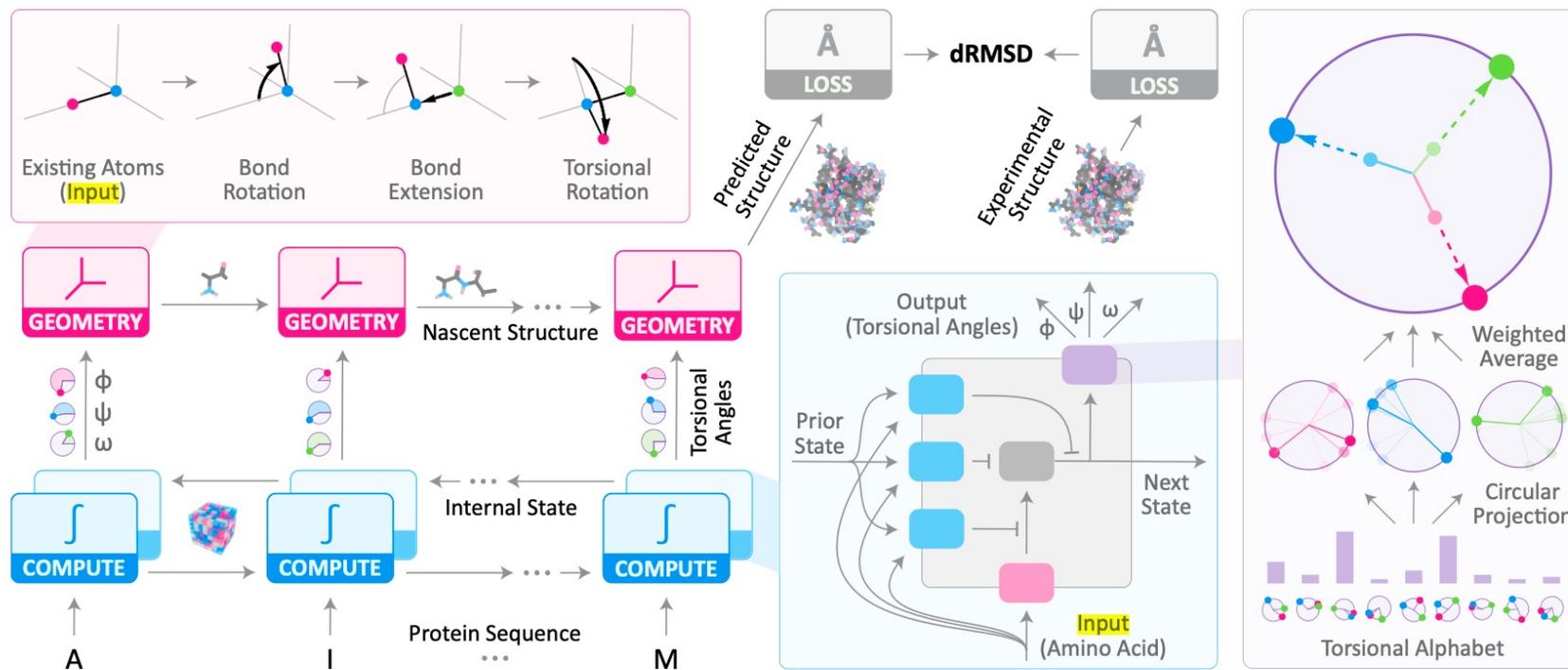


真实结构

蛋白质结构无结构碎片预测

端到端预测

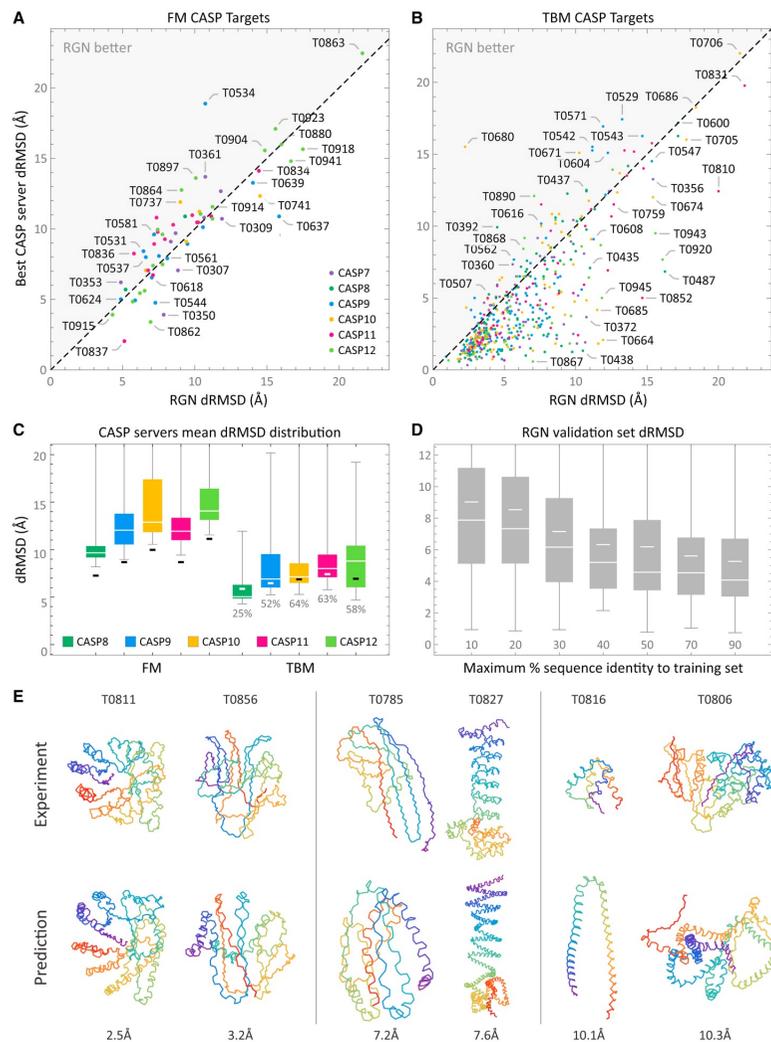
$$A = \text{atan2}(\sin A; \cos A)$$



Mohammed AlQuraishi, End-to-End Differentiable Learning of Protein Structure, Cell Systems, (2019)

蛋白质结构无结构碎片预测

端到端预测



蛋白质结构无结构碎片预测



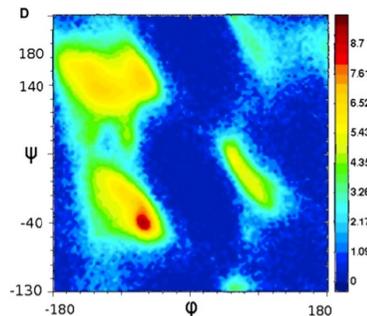
SPOT-Fold: Fragment-Free Protein Structure Prediction Guided by Predicted Backbone Structure and Contact Map

Yufeng Cai,^{†[a]} Xiongjun Li,^{†[a]} Zhe Sun,^[a] Yutong Lu,^[a] Huiying Zhao,^[b] Jack Hanson,^[c]
Kuldip Paliwal,^[c] Thomas Litfin,^[d] Yaoqi Zhou,^{*,[d]} and Yuedong Yang^{*,[a]}

....GADTIFG
KIIRKEIPAKII
FEDDRCLAF
HD....
蛋白质序列

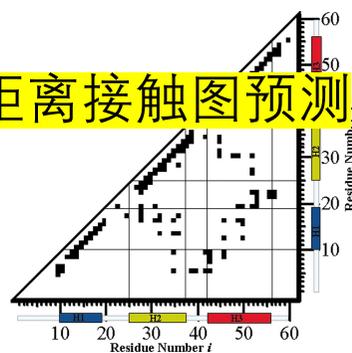
同源
进化
信息

共进
化信
息

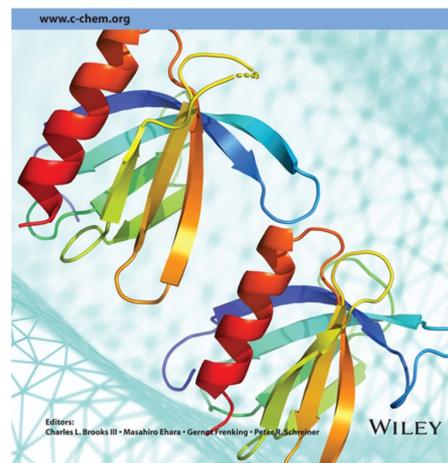


连续角度预测

距离接触图预测



Restrains



三级结构



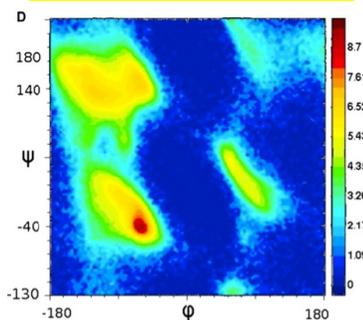
蛋白质结构无结构碎片预测

蛋白质序列
GADTIFG
 KIIRKEIPAKII
 FEDDRCLAF
 HD....

同源
进化
信息

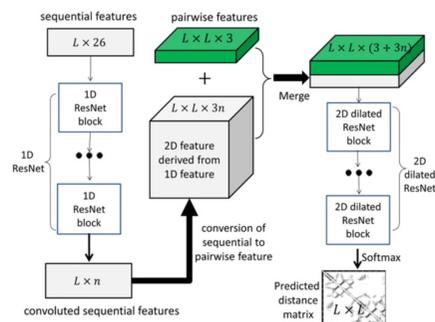
共进
化
信息

连续角度预测

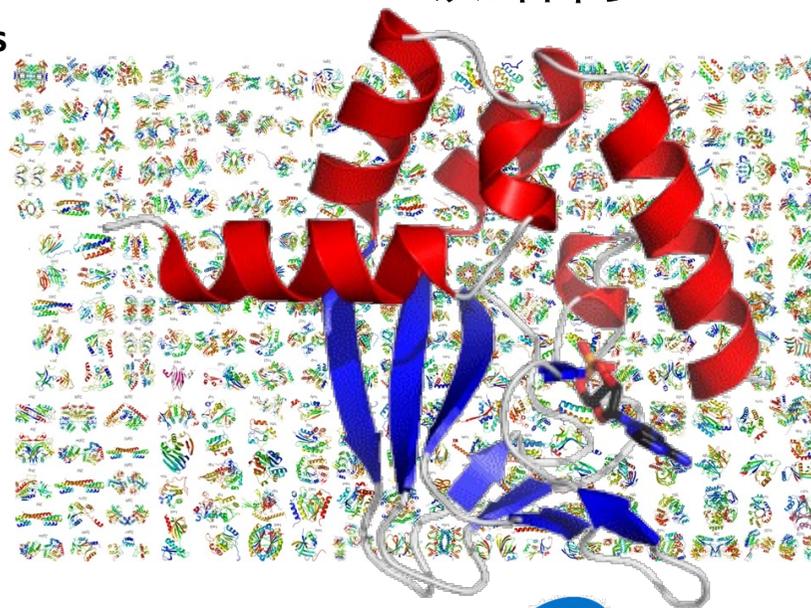


Restrains

连续距离预测



三级结构

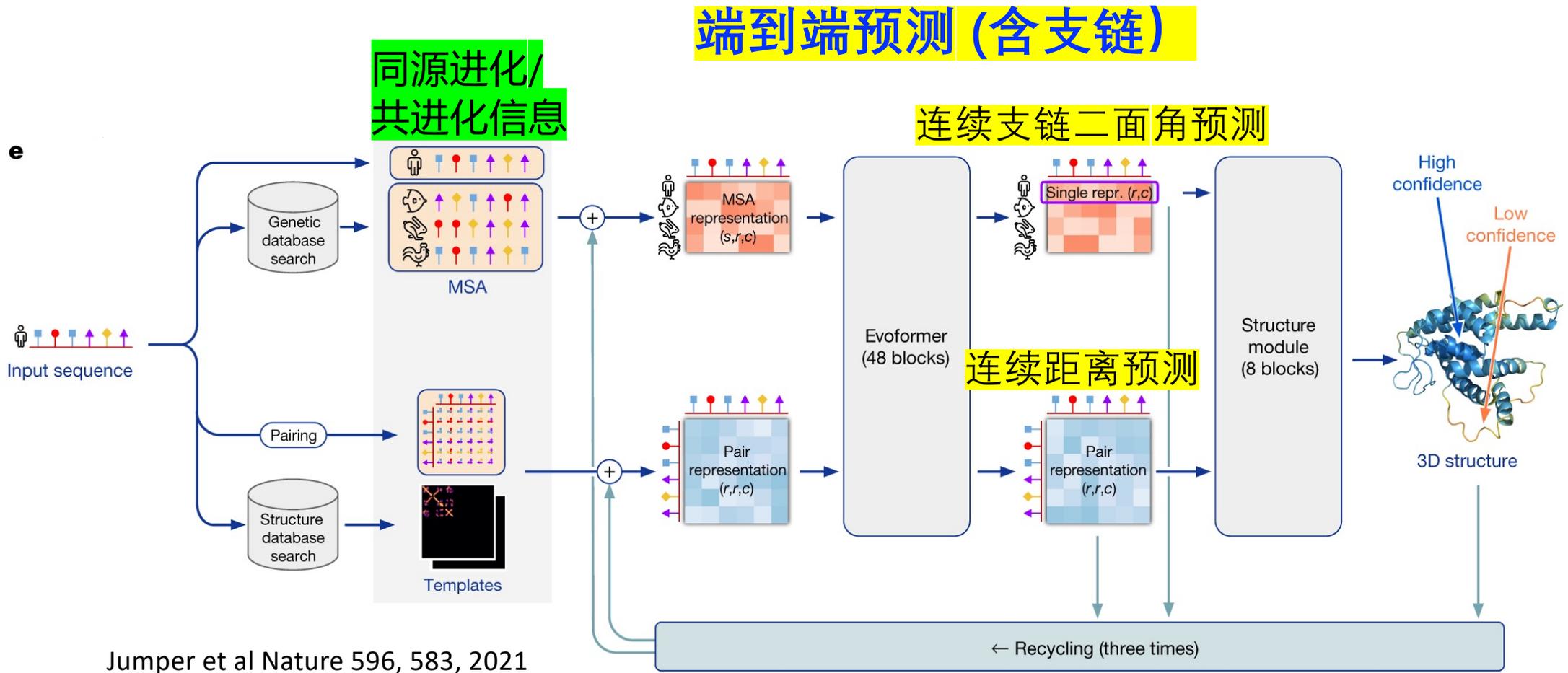


Xu, PNAS, 116, 16856, 2019



深圳湾实验室
 Shenzhen Bay Laboratory

AlphaFold2高精度蛋白质结构预测

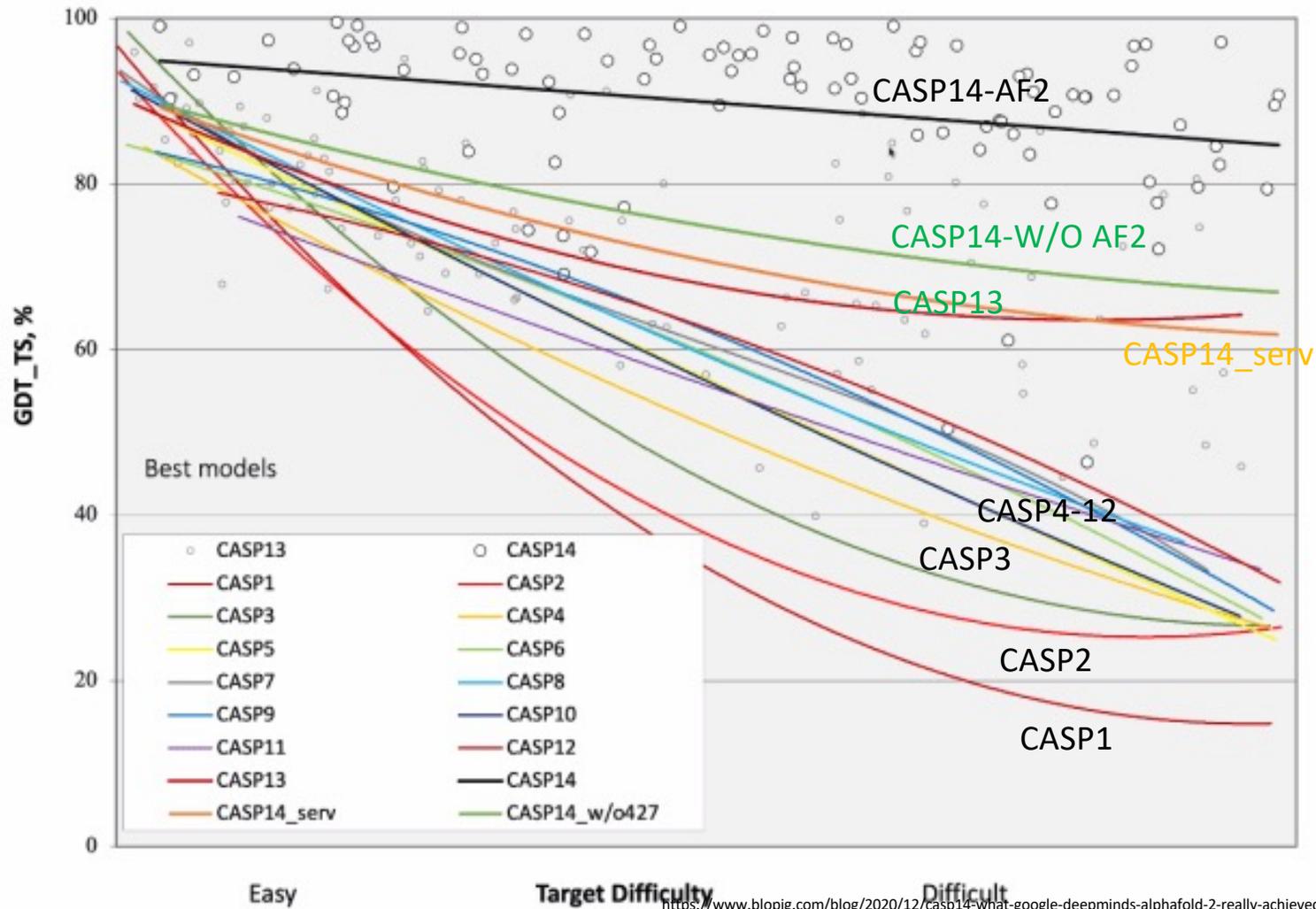


Jumper et al Nature 596, 583, 2021

AlphaFold2高精度蛋白质结构预测

- 连续二面角的准确预测：预测 $\text{Sin}(A)$ 和 $\text{Cos}(A)$ ，ARCTAN
- 连续距离的准确预测
- 端到端结构预测
- 抓住了蛋白质结构的稳定性是靠疏水支链的紧密堆积这个主要矛盾，先支链位置，再主链链接
- 三角形中，两边之和长于第三边
- 结构：平移/旋转不改变结构

2020年底AlphaFold革命



总结

蛋白质结构预测，在精确度上的革命，是在点点滴滴的积累以及一环扣一环的进步之后才能一跃而成的。其中，从分类到连续真实二面角的预测是重要的一环！

后AlphaFold时代展望

1. 蛋白质功能预测：

蛋白-蛋白，多肽，RNA，DNA，Glycan，ligand

2. 蛋白质设计

3. RNA结构预测和设计

感谢

我在Buffalo, Indiana, Griffith Universities的学生和博士后

Hongyi Zhou, Song Liu, Chi Zhang, Ofer Dor, Bin Xue, Eshel Faraggi, Yuedong Yang, Shesheng Zhang, Tuo Zhang, Huiying Zhao, Zhixiu Li, Thomas Litfin, Peter Brown, Jian Zhan



Prof Kuldip Paliwal
& his team @Griffith Univ

招聘：
研究助理
博士后
助理研究员
副研究员

方向：
计算/AI
软件/硬件
分子生物
结构生物
细胞生物

课题组网站



科学网博客



杨跃东和他的团队
@中山大学