

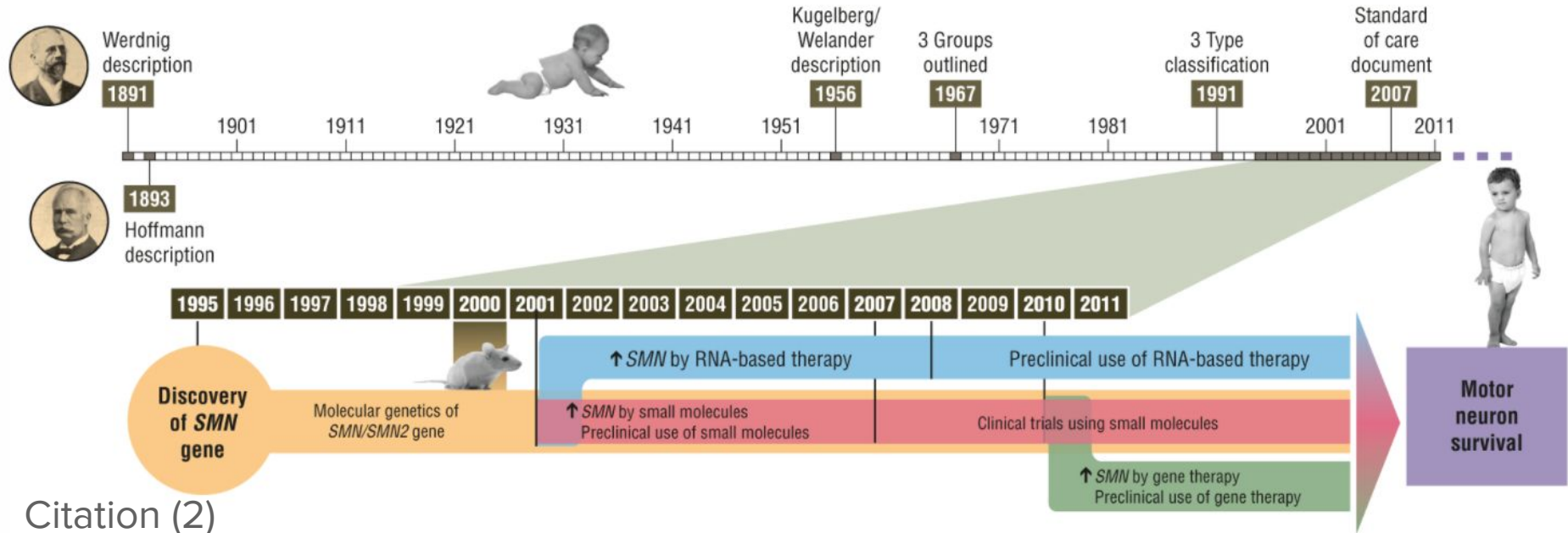
# RNA design rules from a massive open laboratory

---

CS 371 - Thursday January 18th  
Noor Siddiqui

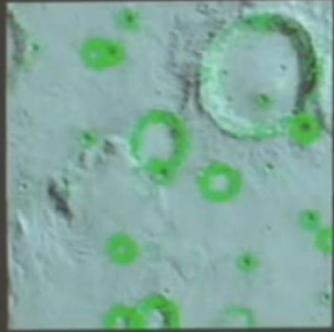
# Background

- RNA structure plays a key role in disease... RNA medicine
- Spinal Muscular Atrophy, many forms of Parkinsons, HIV/retroviruses can be traced to defects in the the way that RNA's fold up

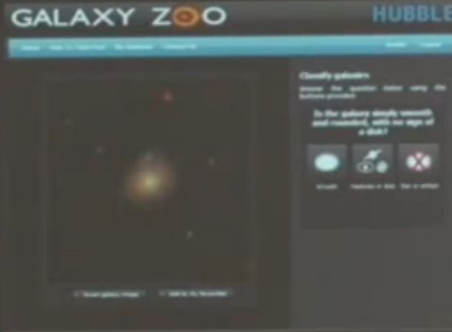


Citation (2)

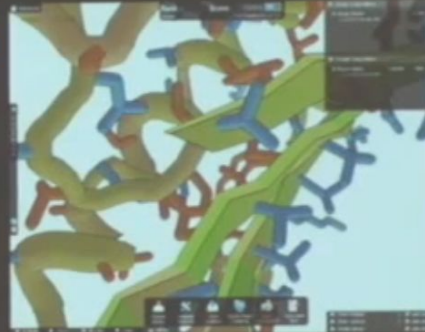
# Quick History of Citizen Science



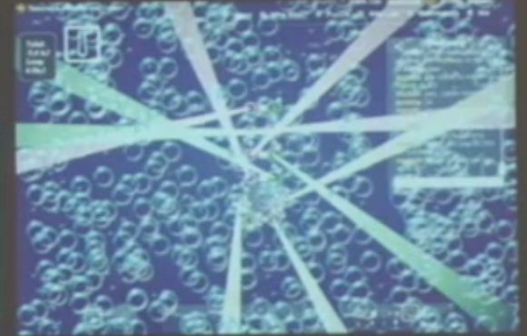
**NASA ClickWorkers**  
2000  
Image Analysis



**Zooniverse Projects**  
2007  
Image and  
Data Analysis



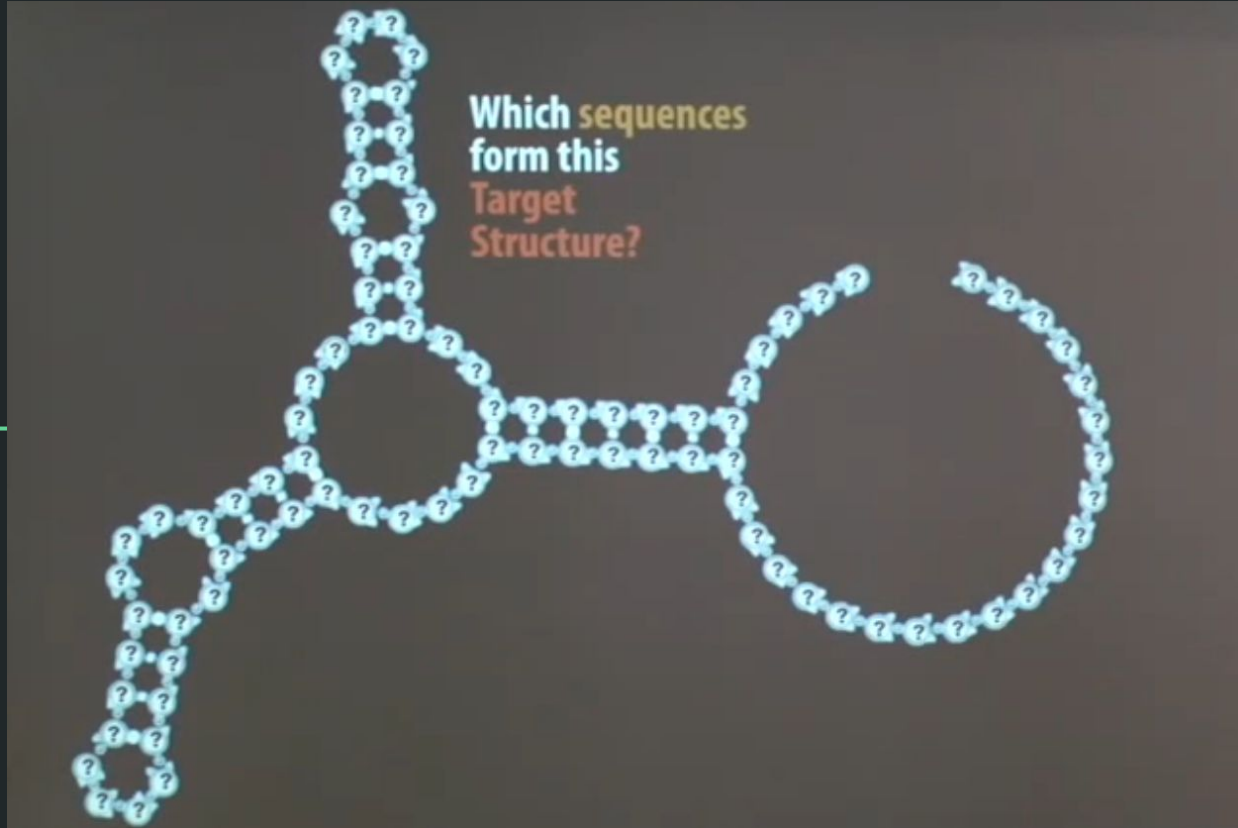
**Foldit**  
2008  
Computational  
Bioscience



**EteRNA**  
2011  
Experimental  
Bioscience

Citation (3)

# Goal - design sequence to fold into target



Citation (3)

DEMO

---

“Designed  
by humans,  
scored by  
nature”

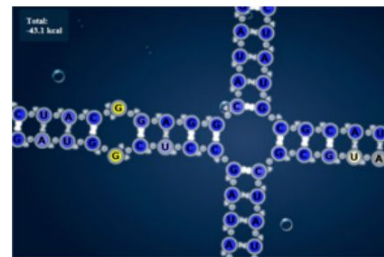
### A Design interface



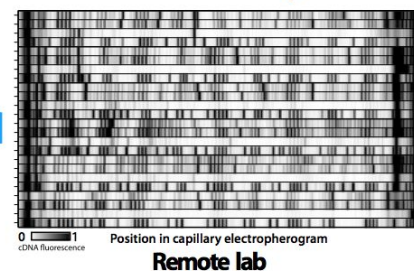
### B Voting interface

	5
Ding's Branches 1	GGAAA GGU
Branches	GGAAA GCU
Squid	GGAAACUC
Mat - Branches V1	GGAAA GGC
Y oh Y	GGAAA GCA
Deck the Halls	GGAAA GCG
JP-11-4-18	GGAAACCA
Chaen 202-	GGAA GCCU
Toad Hall	GGAAA GGU
Starry's Branches II	GGAAA GGC
Mat - Branches V1.1 (Submit...	GGAAA GGC

### D Results viewer



### C Synthesis results



### E Design rule collection

 aldo's Loops & Stacks	0.406	40
	Ordering	length
 deivad's deivad's strategy	0.382	60
	Ordering	length

# Main Results

37,000+ members in the EteRNA online team, generated 100k+ data points

Developed human authored RNA design rules

Developed an algorithm: EteRNAbot using submissions from users and experimental validation scores

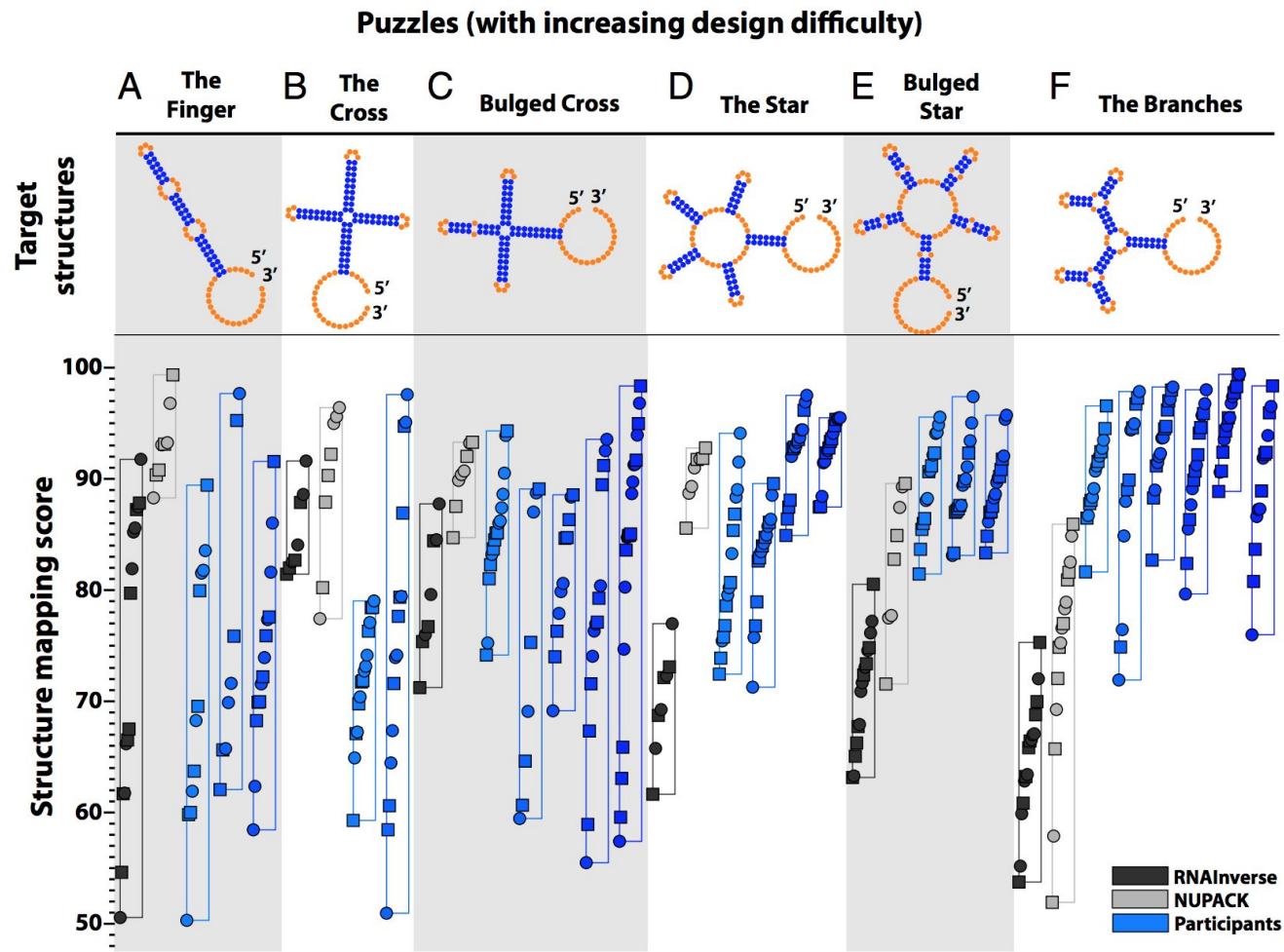
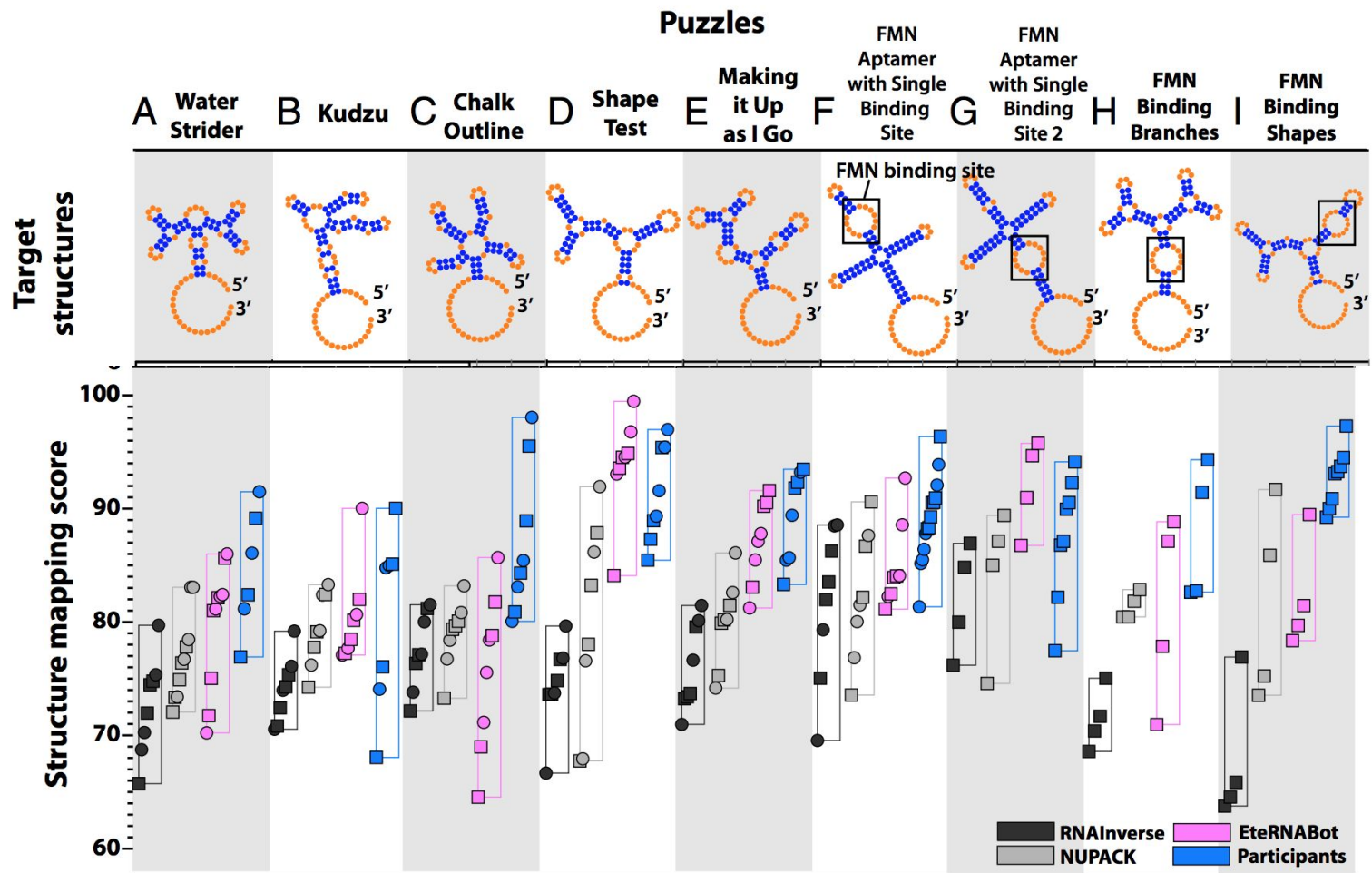


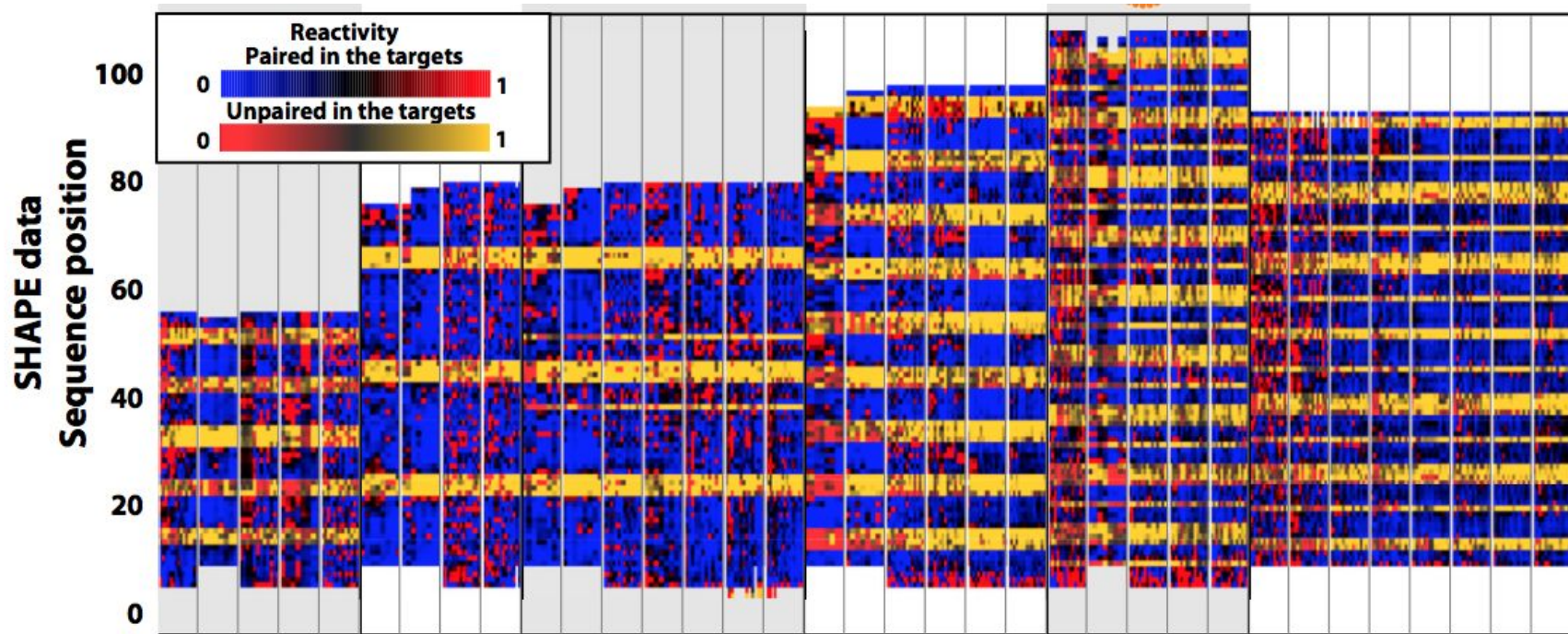
Fig. 2. Phase I puzzles and results in order of puzzle posting date.





**Fig. 4.** Phase II puzzles and results, including the EteRNABot algorithm, Citation (1)

# Scoring structures



Citation (1)

# Strengths

- Hugely popular

**NOVA LABS**

SUPPORT PROVIDED BY ARGOSY FOUNDATION

THE LABS VIDEOS ABOUT OPPORTUNITIES EDUCATORS LOG-IN

YOU ARE HERE → NOVA RNA VIRTUALLAB → ETERNA PUZZLES + RNA SCIENCE VIDEOS → CARRY-OVER 5000 PTS → eterna RESEARCH COMMUNITY @ CMU

## RNA VirtuaLab

Nature's best kept secret is a wonder molecule called RNA. It is central to the origin of life, evolution, and the cellular machinery that keeps us alive. In this Lab you'll play the role of a molecular engineer by solving RNA folding puzzles. Then take your skills to Eterna, where you can design RNAs that could be at the heart of future life-saving therapies.

**PLAY GAME**

GAME ENGINE BY THE ETERNA PROJECT  
CARNEGIE MELLON UNIVERSITY

VIDEO INTRO ABOUT THIS LAB EDUCATOR GUIDE

## RNA Game Lets Players Help Find a Biological Prize



Jim Wilson/The New York Times

Dr. Rhiju Das, left, and Daniel Cantu, a researcher, checked an RNA design created online.

11

[Carnegie Mellon University](#) and [Stanford University](#) are attempting to harness the wisdom of crowds with the creation of an online video game that challenges players to design new ways to fold RNA molecules.

TWITTER

LINKEDIN

PRINT

REPRINTS

# Strengths

- Hugely popular - but small number of experts
- Successful crowdsourced hypothesis generation and validation
- EteRNAbot outperformed the best software structure prediction packages to date (NUPACK and RNAInverse)
- Novel angle: experimental validation of designs

# Weaknesses

Complicated to orchestrate

- governance, what to incentivize, manage limited experimental resources
  - ~84k registered players, ~6k lab players, 1k designs/round, 32 syntheses/round
- Many participants complained the “voting” was a popularity contest
- Could have gone deeper into debugging software edge cases experimentally
- Training tool for scientists?


# Extension of work since publication

Collected design summaries submitted by users into academic papers

Collaborating on new tasks



OpenCRISPR - FMN

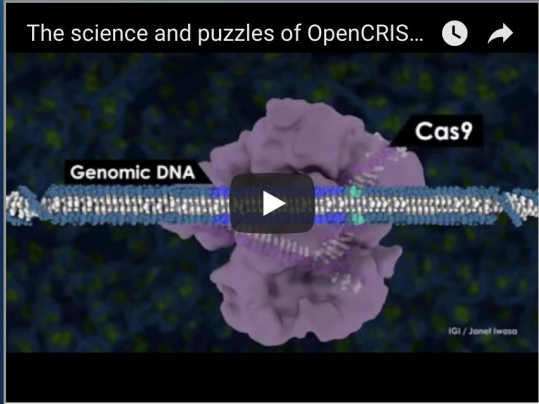


**Help us make gene editing technologies smarter and safer.**

The CRISPR technology is an RNA-based method that can target essentially any DNA gene in a living organism for genetic changes. Since its first demonstration five years ago, CRISPR has been revolutionizing biology and promises to change how we tackle numerous human diseases from malaria to cancer.

Stanford's Center for Personal Dynamic Regulomes and Berkeley's Innovative Genomics Institute are now challenging Eterna players to solve a remaining hurdle in making this technology safe for use.

The science and puzzles of OpenCRISPR...



IGI / Jamel Hessa

Thanks!

---

# Works Cited

1. RNA design rules from a massive open laboratory
2. Kolb SJ, Kissel JT. Spinal Muscular Atrophy A Timely Review. Arch Neurol. 2011;68(8):979–984. doi:10.1001/archneurol.2011.74
3. EteRNA: A Videogame and a Massive Open Lab - Presentation by Rhiju Das (YouTube Video <https://www.youtube.com/watch?v=JMAv0NS6gnl&t=2s>)



# Determining crystal structures through crowdsourcing and coursework

Scott Horowitz, Brian Koepnick, Raoul Martin, Agnes Tymieniecki, Amanda A. Winburn, Seth Cooper, Jeff Flatten, David S. Rogawski, Nicole M. Koropatkin, Tsinatkeab T. Hailu, Neha Jain, Philipp Koldewey, Logan S. Ahlstrom, Matthew R. Chapman, Andrew P. Sikkema, Meredith A. Skiba, Finn P. Maloney, Felix R.M. Beinlich, Foldit Playersw, University of Michigan studentsw, Zoran Popovic, David Baker, Firas Khatib & James C.A. Bardwell

# New feature in Foldit: Electron Density Puzzle

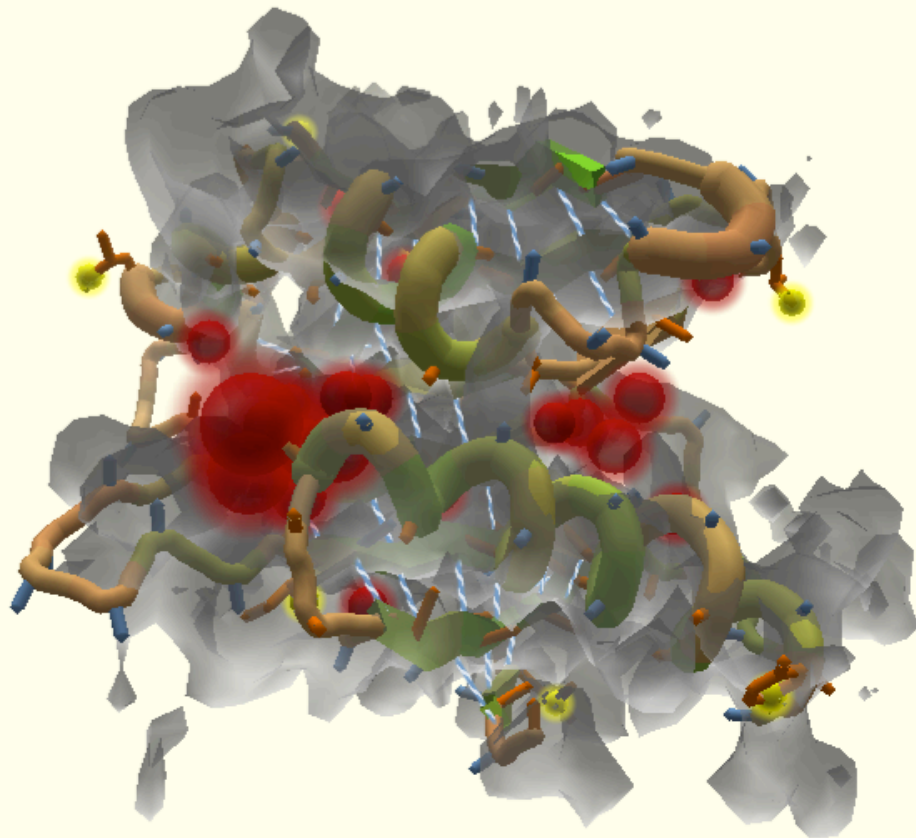
Rank: 74      Score: 11745

Soloist      Beginner Puzzle (<1...s Electron Density)

[Wiki Help for this Puzzle](#)

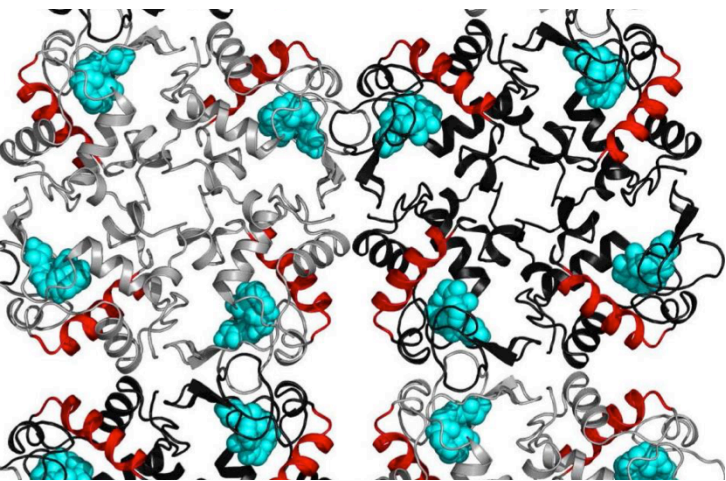
Expires 1/18/2018 15:00 PST(3 days, 19 hours)

▶ [Conditions satisfied: 1 of 1, No bonuses.](#)

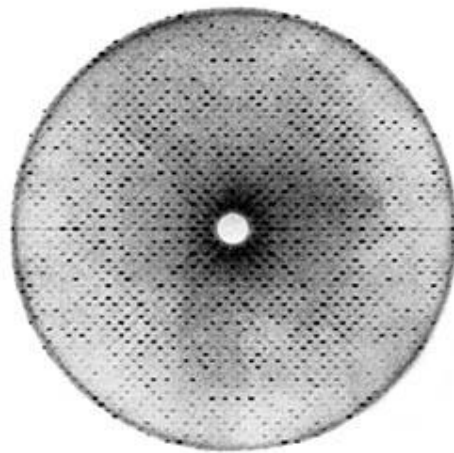


1. What is this puzzle
2. Who is best at solving it
3. What does the result mean

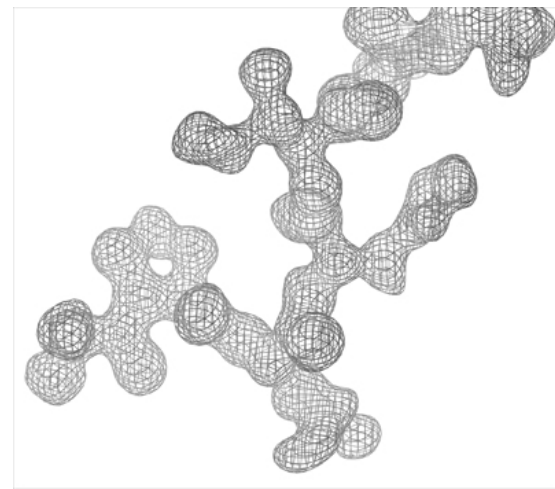
1. What is this puzzle
2. Who is best at solving it
3. What does the result mean



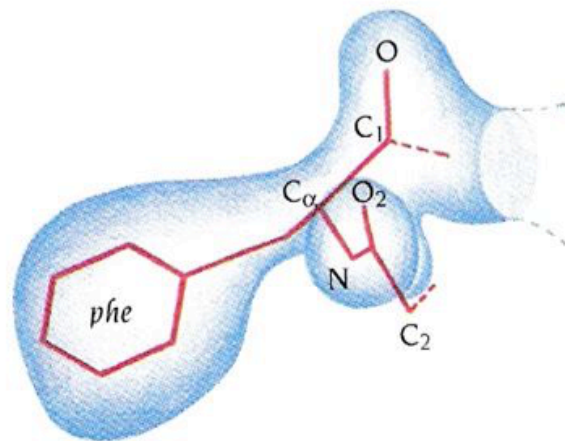
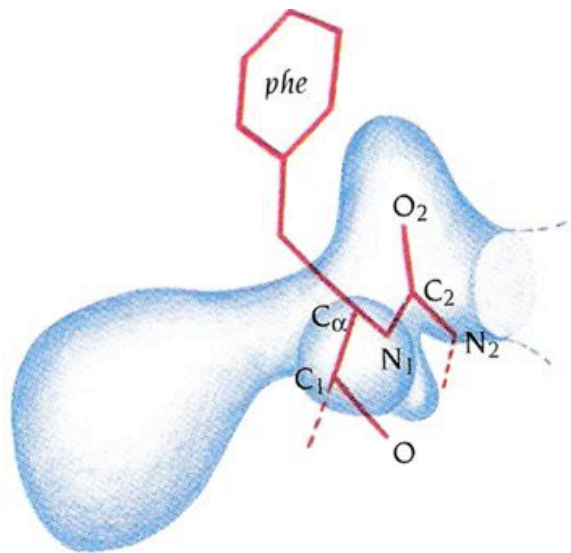
protein crystals



diffraction pattern



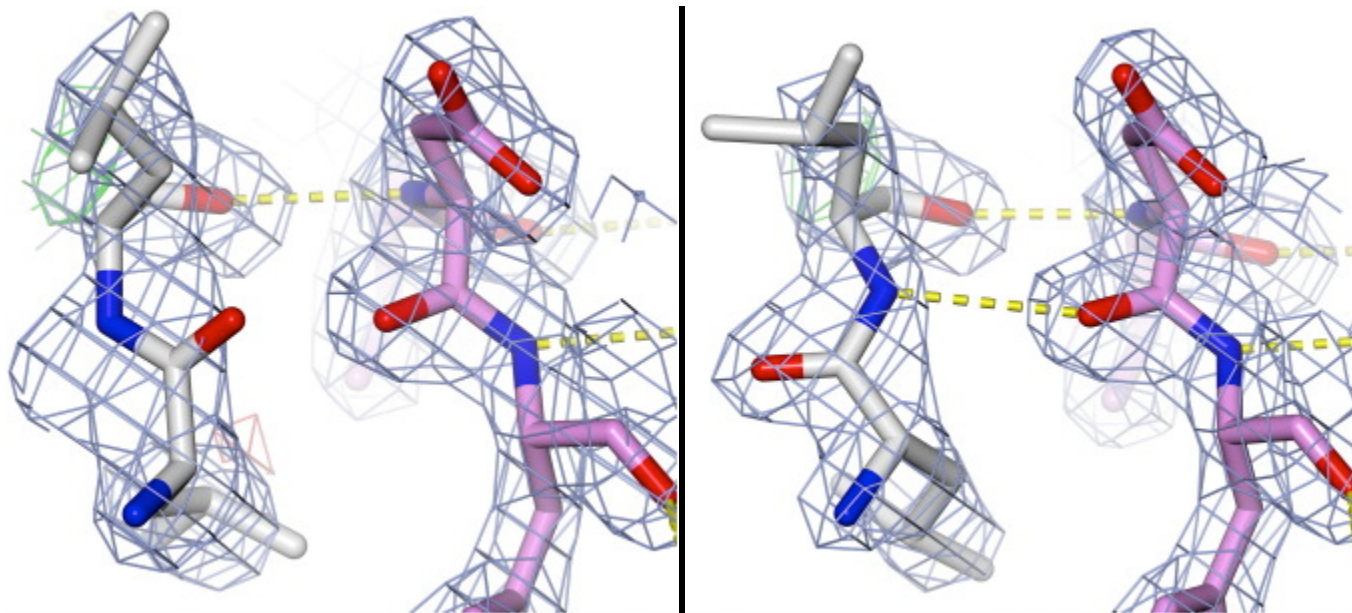
electron density map



“**85**% of deposited protein crystal structures  
contain discernable errors”



Which one is better?



1. What is this puzzle
- 2. Who is best at solving it**
3. What does the result mean

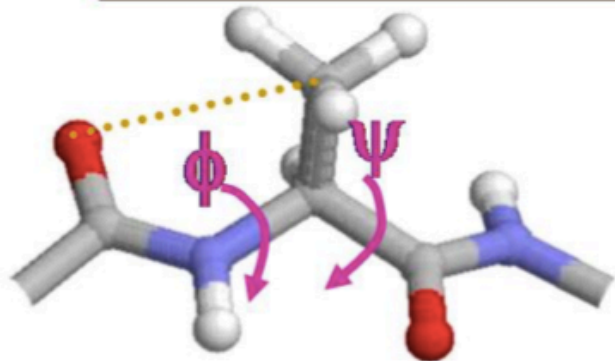
- **Goal:** Build the best structure for YPL067C (yeast protein)
- **Provided:** protein sequence, a secondary structure prediction, electron density map
- **Competitors:**
  - Foldit players
  - Crystallographers
  - Undergrads
  - Phenix autosolve, MR-Rosetta

- **Goal:** Build the best structure for YPL067C (yeast protein)
- **Provided:** protein sequence, a secondary structure prediction, electron density map
- **Competitors:**
  - **Foldit players**
  - **Crystallographers**
  - Undergrads
  - Phenix autosolve, MR-Rosetta

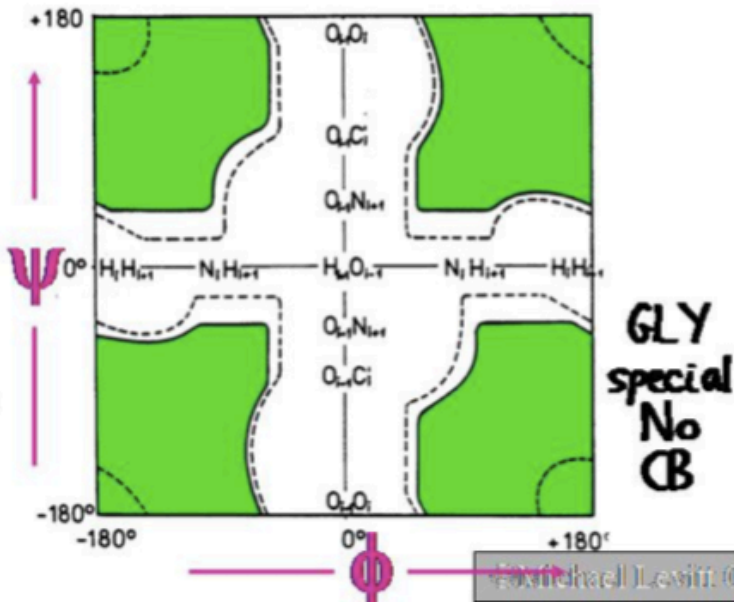
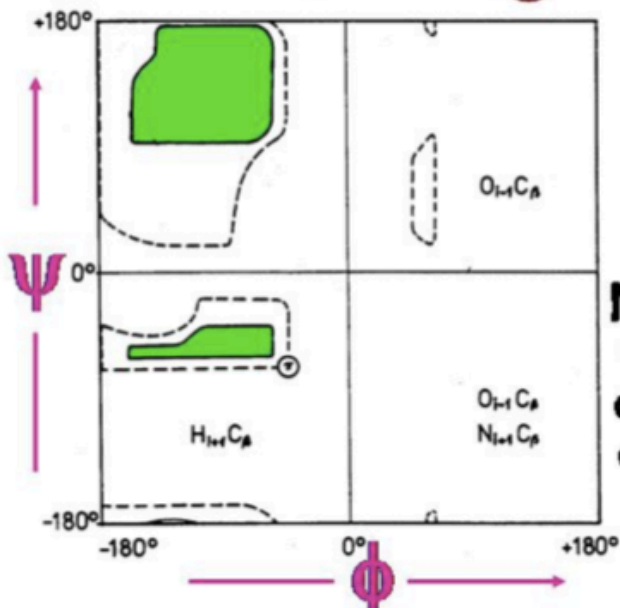
# Metric

- Root mean square deviation of bonds and angles
- Number and severity of steric clashes
- $R_{\text{free}}$
- Ramachandran outliers

# RAMACHANDRAN DIAGRAM



- Map the clashes. (Done in 1963).
- Green is allowed.



# Metric

- Root mean square deviation of bonds and angles
- Number and severity of steric clashes
- $R_{\text{free}}$
- Ramachandran outliers

Trained crystallographers Phenix autosolve MCDB 411 students FoldIt players (Pruned) MR-Rosetta

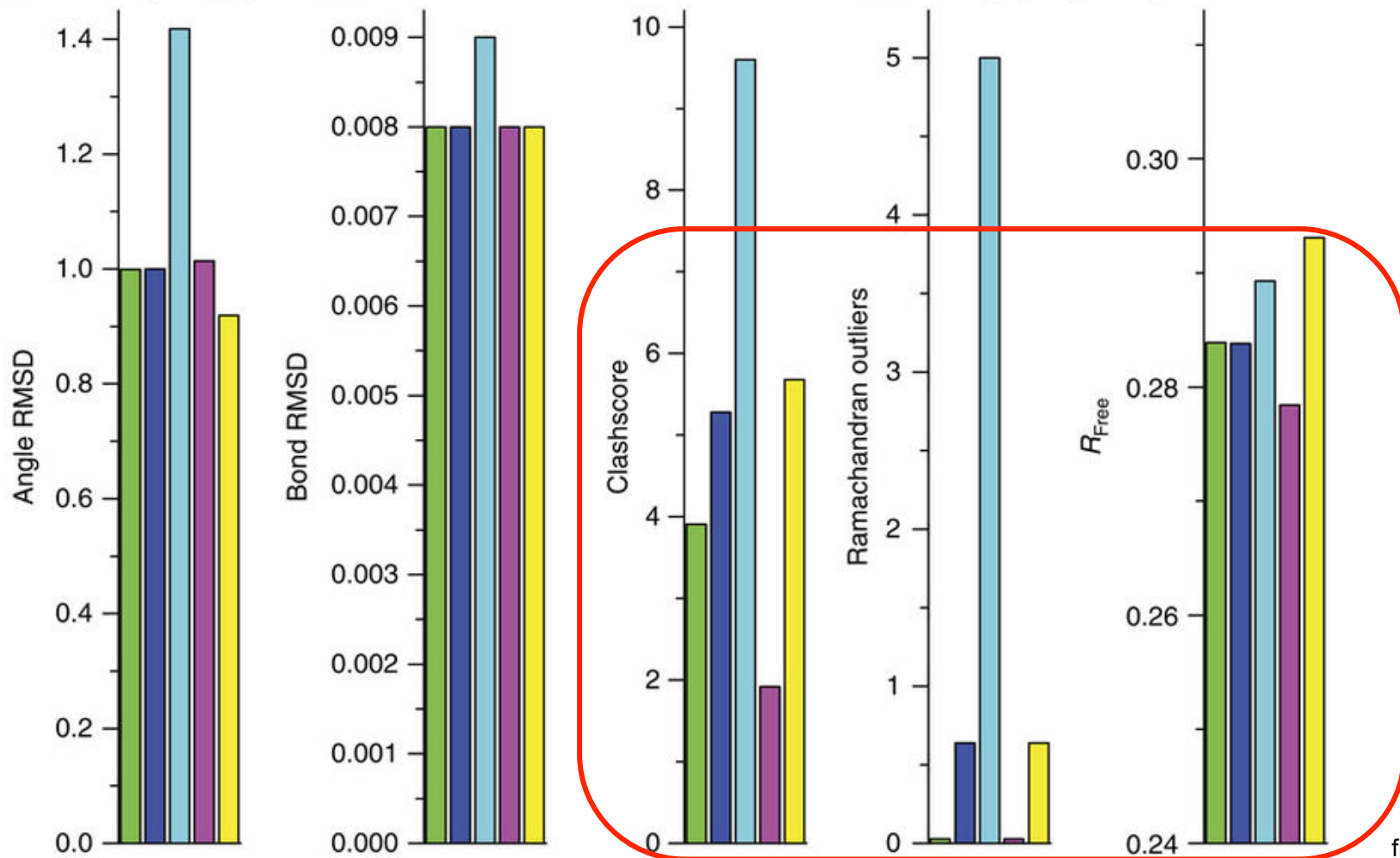


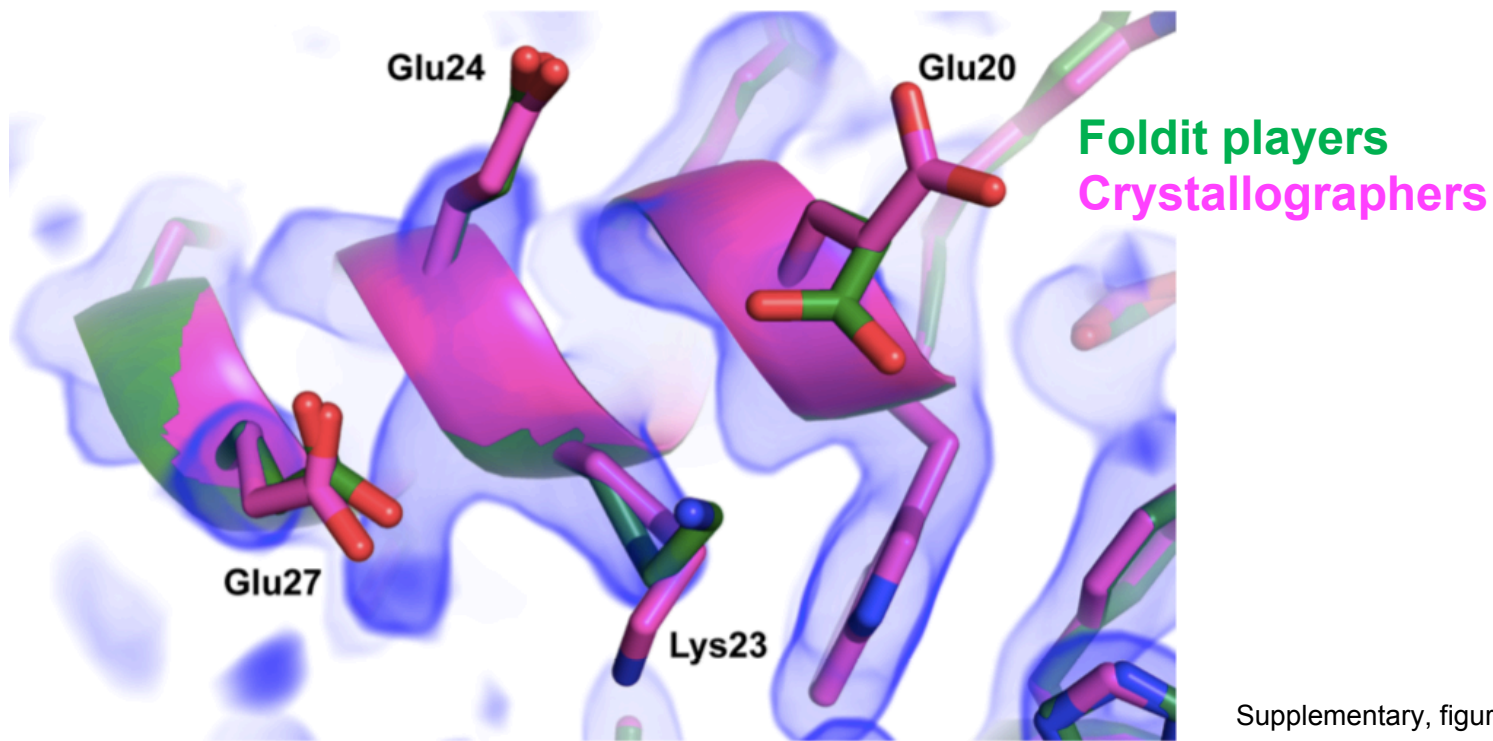
figure 2



1. What is this puzzle
2. Who is best at solving it
- 3. What does the result mean**

# Differences on the building result

- Better side-chain conformations

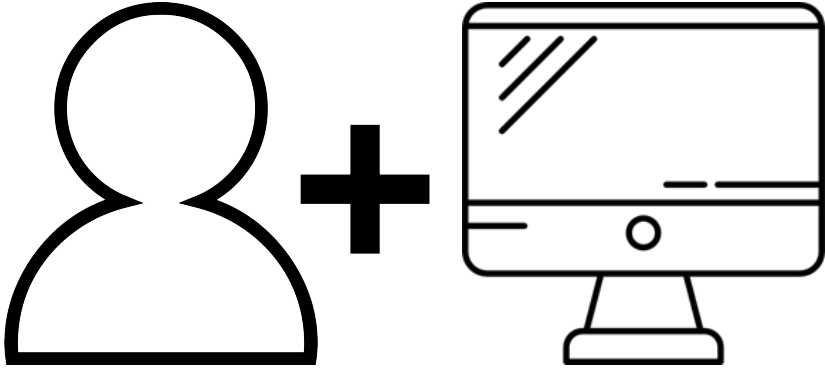


## Differences on the building process

- Foldit players used various types of tools, including modifications on secondary structure
- Foldit players sampled conformational space widely

# Implication

Human intervention



Collaboration

More extensive algorithm

# Critique

Trained crystallographers Phenix autosolve MCDB 411 students FoldIt players (Pruned) MR-Rosetta

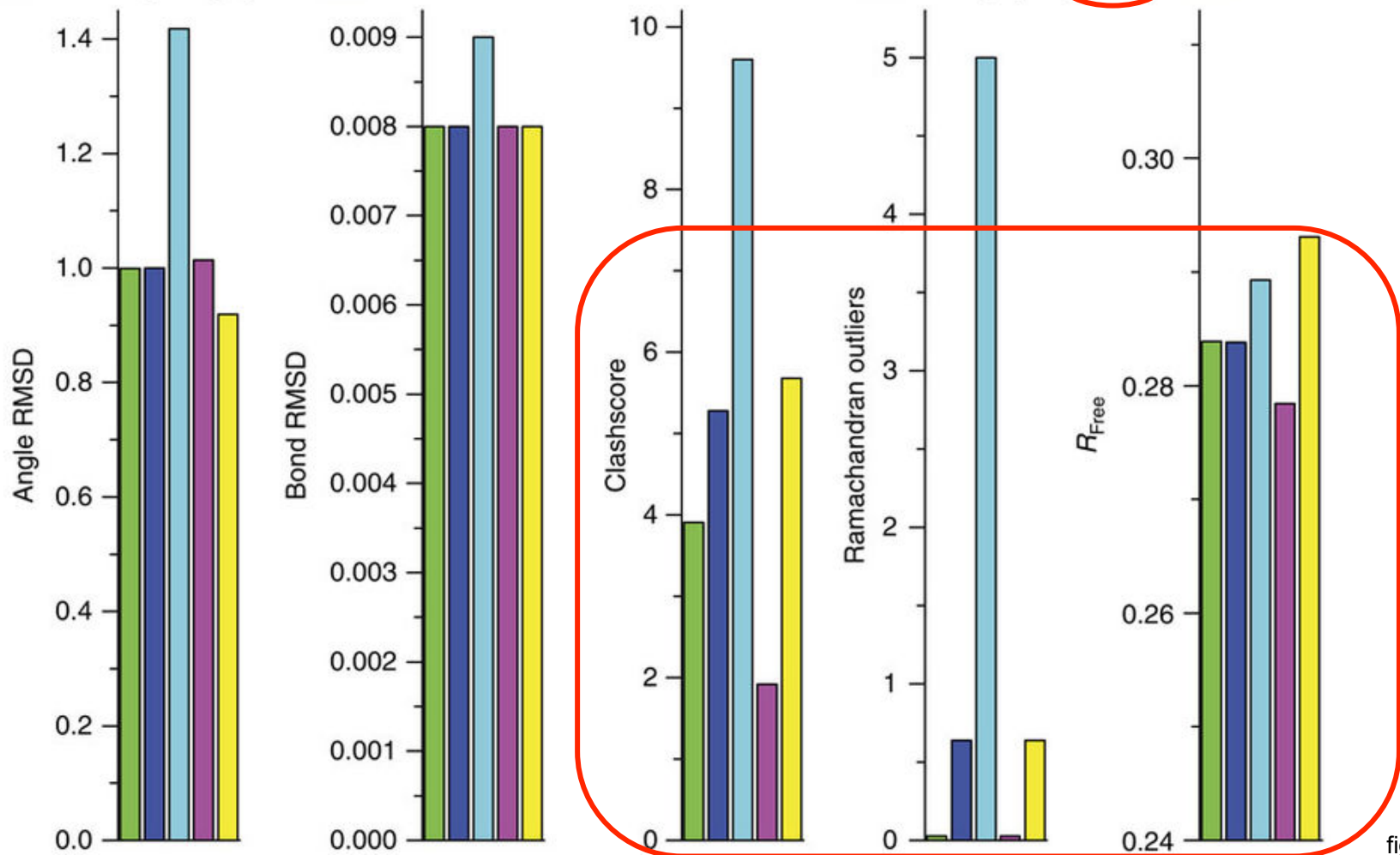
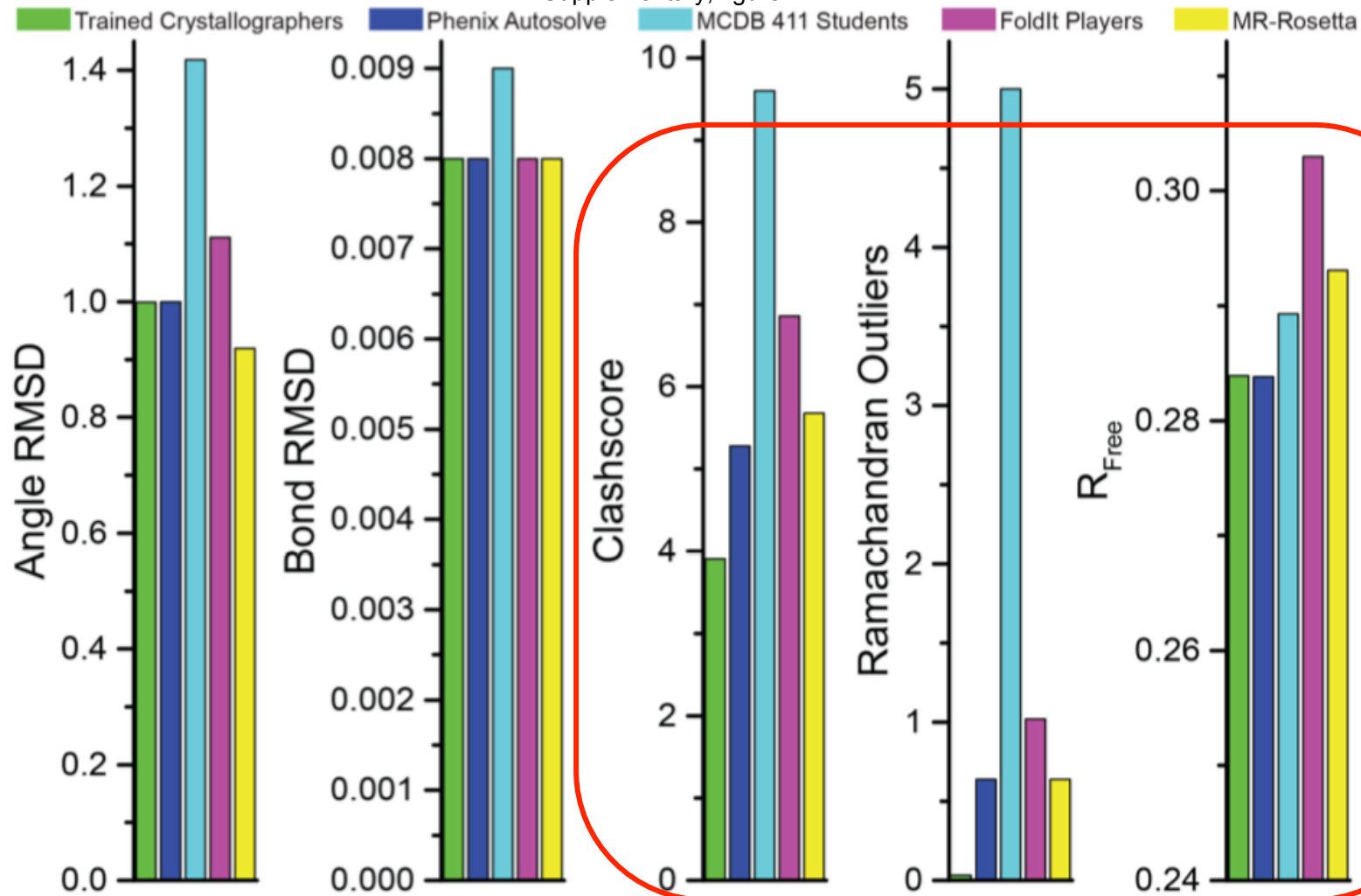


figure 2

Supplementary, figure 4



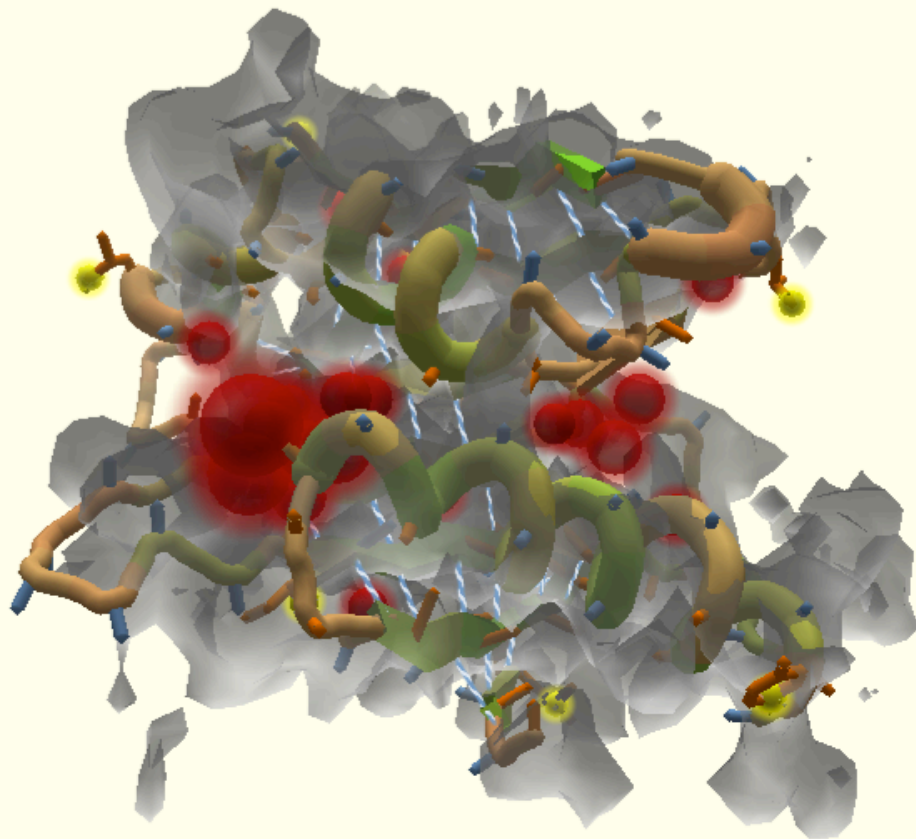
Rank: 74      Score: 11745

Soloist      Beginner Puzzle (<1...s Electron Density)

[Wiki Help for this Puzzle](#)

Expires 1/18/2018 15:00 PST(3 days, 19 hours)

▶ Conditions satisfied: 1 of 1, No bonuses.





- **Goal:** Build the best structure for YPL067C (yeast protein)
- **Provided:** protein sequence, a secondary structure prediction, electron density map
- **Competitors:**
  - **Foldit players**
  - **Crystallographers**
  - Undergrads
  - Phenix autosolve, MR-Rosetta

- **Goal:** Build the best structure for YPL067C (yeast protein)
- **Provided:** protein sequence, a secondary structure prediction, “**experimentally phased, density refined**” map
- **Competitors:**
  - **Foldit players**
  - **Crystallographers**
  - Undergrads
  - Phenix autosolve, MR-Rosetta

# Critique

- Foldit players' models pruned based on crystallographers' models
  - "Regions where the density was very poor and decisions ... about whether to keep trying to build or not proved to be the **hardest** part"
- Phase information not easily available