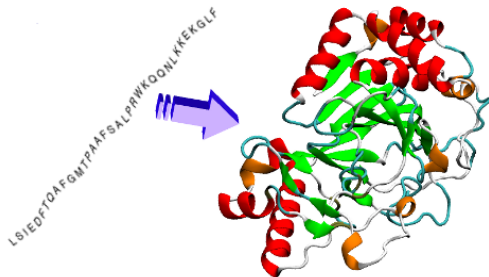


Rosetta & Foldit

9. říjen 2012

- 1 Protein folding
- 2 Rosetta
- 3 Rosetta@home
- 4 Foldit
- 5 Robetta
- 6 Shrnutí

Protein folding



- Proces skládání prostorového tvaru proteinu (terciární struktura)
- Struktura → funkce

Predikce vs. experimentální přístup

- Sekvenci aminokyselin proteinů lze snadno přečíst ze sekvence nukleotidů DNA
- Pouze zlomek proteinů se známou sekvencí má objasněnou strukturu
- Problém je přečíst informaci o prostorové struktuře proteinu ze sekvence aminokyselin
 - Ohromné množství konformací proteinu
 - Je obtížné vytvořit realistický model vodného prostředí

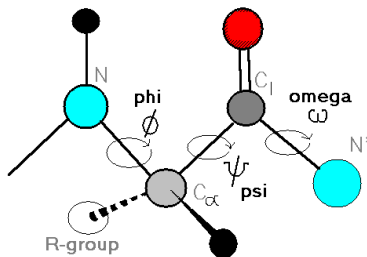
Rosetta

- Baker laboratory (University of Washington)
- Algoritmus pro předpověď struktury proteinu ze sekvence de novo
- Softwarový balík s mnoha funkčními moduly: Rosetta ab initio, RosettaNMR, RosettaDesign, RosettaDoc, Rosetta Fragment Selection, Rosetta Ligand, RosettaMembrane...

Rosetta

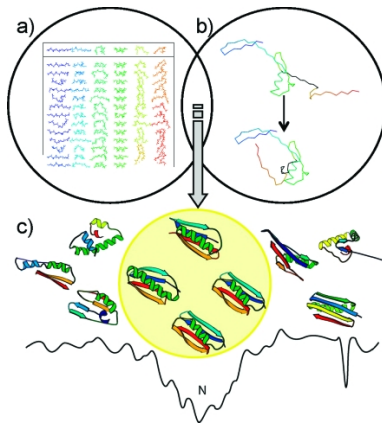
- Ab initio vs. de novo modelování
 - "Od počátku"
 - Sekvence aminokyselin obsahuje informaci o struktuře proteinu
 - Snaha najít stav s nejnižší energií
 - Potřeba zredukovat ohromné množství konformací
 - Potřeba ohodnotit jednotlivé stavy (funkce přidělující skóre na základě energie)

Odvození úhlů mezi aminokyselinami v polypeptidové kostře



- Rozdělení proteinu na segmenty
- Snaha o snížení počtu konformací pro jednotlivý segment: nemusí se vyzkoušet každý úhel mezi dvěma aminokyselinami, ale uvažují se pouze realistické úhly
- Využití experimentálně získaných struktur
- Knihovna proteinových fragmentů
- Fragments dlouhé 3 a 9 reziduí
- Fragments jsou mapovány v PDB pomocí PSIBLAST
- Kombinace fragmentů a vypočítání energie

Schéma de novo modelování Rosetta



- (a) Knihovny fragmentů zachycují lokální konformace
- (b) Kombinace různých fragmentů z různých knihoven, optimalizace vzdálených kontaktů
- (c) Snížení energie struktury na globální minimum

Hodnocení konformace proteinu

- Fyzikální základ
- Vlastnosti aminokyselin (hydrofobicita, velikost)
- Vzájemné interakce vedlejší řetězců
- Tvar kostry proteinu (vodíkové můstky, disulfidické můstky)
- Tvorba sekundárních struktur (alfa šroubovice, beta skládaný list)

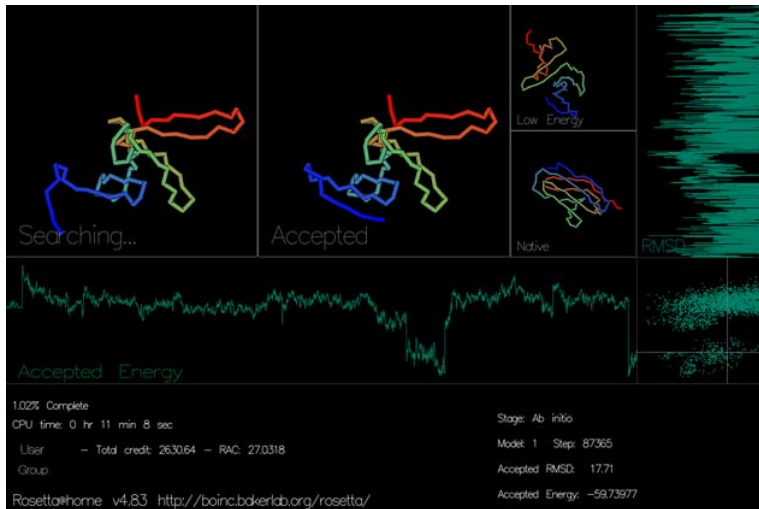


- BOINC (Berkeley Open Infrastructure for Network Computing)
- Open-source platforma vytvořená po úspěchu SETI@home
- Projekt založený na distribuovaných výpočtech
- Software využívá výpočetní kapacitu domácích počítačů v době, kdy jsou nečinné
- Stahuje problém z hlavního serveru a výsledky odesílá zpět



- Hledá strukturu proteinů s nejnižší energií
- Spuštěno v říjnu 2005

Grafické zobrazení výpočtu



Foldit

- Free online hra
- Vychází z projektu Rosetta
- Využívá Rosetta energy function
- Stejný cíl jako Rosetta@home řeší využitím veřejnosti
- Člověk řeší problém efektivněji než program
- Motivace uživatelů formou hry

Motivace uživatelů formou hry

The image shows a screenshot of the Foldit game interface. The central part of the screen displays a complex protein structure rendered in green and blue, with some parts highlighted in orange and blue. The structure is set against a light yellow background.

On the right side, there is a leaderboard panel with the following information:

Rank: 17 Score: 9092
 48: Pro Peptide

Group Competition

#	Group Name	Score
1	The Lone Folder	9388
2	Street Smarts	9367
3	Illinois	9303
4	Berkeley	9255

Player Competition

#	Player Name	Score
16	psen	9098
17	kathleen	9092
18	versat82	9091
19	darktorres	9081
20	ccarrico	9032
21	mbjorkegren	9048
22	sslickerson	9038

Chat

At the bottom left, there is a control panel with the following text and icons:

Shake sidechains to improve the protein.
 Hotkey: S

Shake Sidechains Wiggle Backbone Clear Locks and Bands Reset Puzzle Mouse Help

Actions History View File

At the bottom right, there is a "Pull Tool" button.

 Pull Mode

Rank: 686


Score:

3131 


Soloist

New Player Puzzle 2: Scorpion Toxin

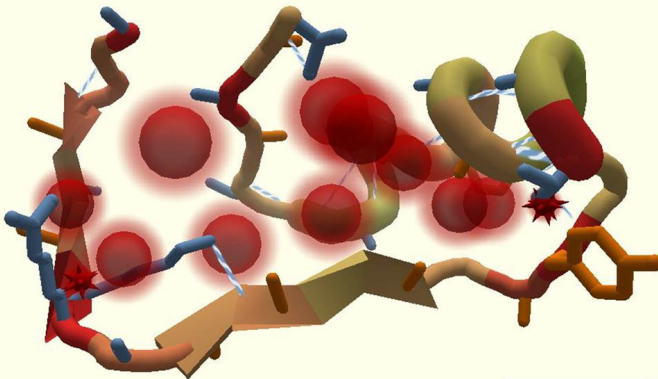
Expires 10/05/2011 16:59 PDT (1 days, 3 hours)

 No bonuses or conditions

 Group Competition

 Soloist Competition

Lookbook   



 Chat - Group

 auto show

 Chat - Puzzle








 auto show

 Chat - Global

 auto show

 Notifications

 auto show

 Actions  Undo  Social  Modes  Behavior  View  Menu

Člověk řeší problém efektivněji než program

▼ Dr. David Baker



Several **sheets** are lining up in this protein, but one is out of place. Pull the misaligned **sheet** back in to form **hydrogen bonds**! Don't forget you can control-click to lock, and use Shake and Wiggle.

Repeat Introduction
Clear Labels

Progress: of 10000
Level 4-2: A Sheet Out of Place
▶ Chat



Shake Sidechains
Wiggle Backbone
Clear Locks
Reset Puzzle

▲ Actions ▶ History ▶ File

Pull Tool

Algorithm discovery by protein folding game players

Firas Khatib^a, Seth Cooper^b, Michael D. Tyka^a, Kefan Xu^b, Ilya Makedon^b, Zoran Popović^b, David Baker^{a,c,1}, and Foldit Players

^aDepartment of Biochemistry, ^bDepartment of Computer Science and Engineering; and ^cHoward Hughes Medical Institute, University of Washington, Box 357370, Seattle, WA 98195

Contributed by David Baker, October 5, 2011 (sent for review June 29, 2011)

Foldit is a multiplayer online game in which players collaborate and compete to create accurate protein structure models. For specific hard problems, Foldit player solutions can in some cases outperform state-of-the-art computational methods. However, very little is known about how collaborative gameplay produces these results and whether Foldit player strategies can be formalized and structured so that they can be used by computers. To determine whether high performing player strategies could be collectively codified, we augmented the Foldit gameplay mechanics with tools for players to encode their folding strategies as “recipes” and to share their recipes with other players, who are able to further modify and redistribute them. Here we describe the rapid social evolution of player-developed folding algorithms that took place in the year following the introduction of these tools. Players developed over 5,400 different recipes, both by creating new algorithms and by modifying and recombining successful recipes developed by other players. The most successful recipes rapidly spread through the Foldit player population, and two of the recipes became particularly dominant. Examination of the algorithms encoded in these two recipes revealed a striking similarity to an unpublished algorithm developed by scientists over the same period. Benchmark calculations show that the new algorithm independently discovered by scientists and by Foldit players outperforms previously published methods. Thus, online scientific game frameworks have the potential not only to solve hard scientific problems, but also to discover and formalize effective new strategies and algorithms.

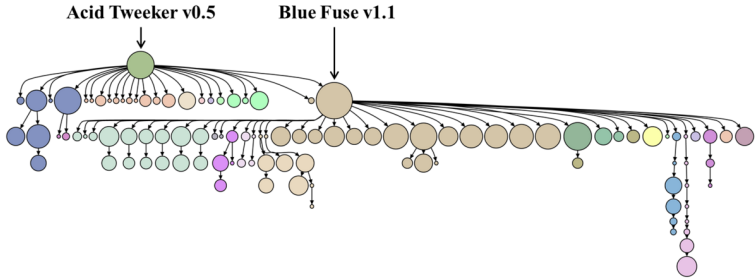
As the players themselves understand their strategies better than anyone, we decided to allow them to codify their algorithms directly, rather than attempting to automatically learn approximations. We augmented standard Foldit play with the ability to create, edit, share, and rate gameplay macros, referred to as “recipes” within the Foldit game (10). In the game each player has their own “cookbook” of such recipes, from which they can invoke a variety of interactive automated strategies. Players can share recipes they write with the rest of the Foldit community or they can choose to keep their creations to themselves.

In this paper we describe the quite unexpected evolution of recipes in the year after they were released, and the striking convergence of this very short evolution on an algorithm very similar to an unpublished algorithm recently developed independently by scientific experts that improves over previous methods.

Results

In the social development environment provided by Foldit, players evolved a wide variety of recipes to codify their diverse strategies to problem solving. During the three and a half month study period (see *Materials and Methods*), 721 Foldit players ran 5,488 unique recipes 158,682 times and 568 players wrote 5,202 recipes. We studied these algorithms and found that they fell into four main categories: (i) *perturb and minimize*, (ii) *aggressive rebuilding*, (iii) *local optimize*, and (iv) *set constraints*. The first category goes beyond the deterministic minimize function provided to Foldit players, which has the disadvantage of readily

Sociální evoluce foldit receptů



- Algorithm discovery by protein folding game players. Khatib, F., Cooper S., Tyka M. D., Xu K., Makedon I., Popovic Z., et al. Proc Natl Acad Sci USA (2011)

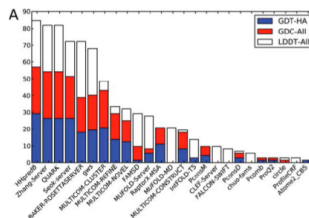
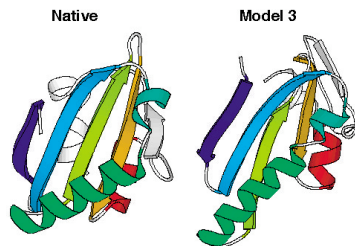


- Server pro predikci proteinových struktur
- Rozděluje proteiny na domény
- Využívá homologní a ab initio modelování (Rosetta)

Přínos

- Predikce struktury proteinů
- Návrh nových proteinů
- Návrh léčby chorob: rakovina, AIDS, Alzheimerova choroba, Creutzfeldt-Jakobova choroba
- Interakce proteinů s proteiny (dokování)
- Interakce proteinů s malými molekulami (návrh léčiv)
- Design nových proteinů (vakcíny, paliva)
- Testování nových metod ve strukturní bioinformatice

Závěr



- Přístup Rosetta byl inovativní díky využití fragmentových knihoven pro hledání vhodných konformací
- Se znalostí lokálních struktur lze snadněji získat globální strukturu
- Rosetta předpokládá, že informace o lokální struktuře lze získat ze strukturních databází
- Jeden z nejúspěšnějších programů na určení struktury proteinů: soutěže CASP, CAPRI