Finding Cinderella's slipper—proteins that fit

Marc Ostermeier and Stephen J. Benkovic

When Cinderella disappeared in great haste at the last stroke of midnight, her prince was left with a daunting task. He didn't know anything about her. How could he find her? Luckily, she left behind a glass slipper. Find the foot that fits the slipper and he would find his true love. Exhaustively, the prince searched door to door and eventually found Cinderella. Now imagine that instead of one slipper, he had as many different slippers as there were feet in his kingdom, but only one slipper was the correct one. How does the prince simultaneously find the slipper and Cinderella? In this issue, Joelle Pelletier et al.[1] successfully address the protein–protein interaction embodiment of this significantly more complex problem using an enzymatic two-hybrid-like system.

Since its development 10 years ago[2], researchers have used the two-hybrid system to study and identify protein–protein interactions in a genetic system. Its success and widespread use stems from its ability to detect most protein–protein interactions independently of their function, to search a library for proteins that interact with a known protein, and to exploit protein–protein interactions that are critical to most biological functions. The field has grown rapidly, with many clever derivations of the original concept allowing the study of protein–protein interactions, DNAprotein interactions, RNA–protein interactions, small molecule–protein interactions, ligand–receptor interactions, and trimeric interactions. Screens have even been devised to identify mutations, peptides, or small molecules that dissociate protein interactions[3].

The majority of two-hybrid-like systems described to date involve interactions between the desired proteins mediating the transcriptional control of a gene that offers convenient screening or selection. As such, the systems usually require host-specific processes or enzymes and cannot be used when the interacting partners bind with components of the transcription machinery or transcription repressors. The authors have already described an enzymatic two-hybrid system that subverts these limitations, which they term a protein-fragment complementation assay (PCA)[4].

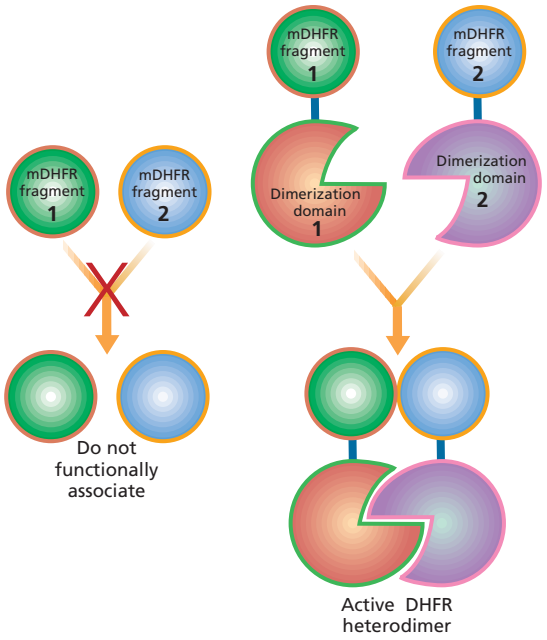

Figure 1. Two designed fragments of murine dihydrofolate reductase (mDHFR) do not functionally associate unless fused to dimerization domains. Hence, restoration of mDHFR activity can be used to demonstrate or select for association between any domains fused to the mDHFR fragments. In this manner, libraries of complementary heterodimeric coiled-coil forming sequences were fused to both mDHFR fragments and panned agasint each other to select for "the best" coiled-coil heterodimer.

The PCA consists of two designed fragments of murine dihydrofolate reductase (mDHFR), each of which is fused to an interacting domain (see Figure). The two mDHFR domains without the interacting domains do not associate efficiently enough to produce DHFR activity. Association between the two interacting domains drives the association of the mDHFR fragments and is essential for complementation of *Escherichia-coli* grown in the presence of the anti-folate drug trimethoprim. By introducing fragment interface mutations that partially disrupt reassembly of the fragments, the stringency of the assay can be controlled. The mDHFR PCA has also been shown to be useful in an in vivo fluorescence assay, both for quantitative analysis of ligand-induced protein dimerization in mammalian cells[5] and to demonstrate ligand-induced conformation change in receptors on the cell surface[6].

One potential caveat of the mDHFR PCA that has not been clearly addressed is the seeming requirement for the C termini of the interacting proteins to be proximal. It is quite possible that proteins that interact with their C termini being distal would not reconstitute an active mDHFR, even with the inclusion of a long linker between the mDHFR domain and the interacting protein. However, alternative constructs of mDHFR (or another enzyme) suitable for "antiparallel" interacting proteins can certainly be envisioned.

As in conventional two-hybrid systems, the screen involves fusing the protein of interest to one mDHFR fragment and a library to the second mDHFR fragment, and then selecting for proteins that interact with the protein of interest (i.e., finding the foot that fits the glass slipper). In this issue, however, the authors take mDHFR PCA to "the second dimension," to study how interacting partners are affected by variations in either partner.

In this system, both fragments of mDHFR are fused to libraries (i.e., many slippers and many feet). The authors panned two designed libraries of complementary heterodimeric coiled-coil forming sequences against each other. By use of selections of increasing stringency and finally a competitive selection, the sequences of $2x10^6$ combinations of 10^5 variants essentially converged to one pair. A competitive selection scheme allows discrimination be-tween subtle differences in suitability so the best performing coiled-coil pair could be selected. Interestingly, by following the rate of sequence convergence at several positions throughout the competitive selection, they found that the rates of selection were position dependent.

Library-versus-library two-hybrid systems had been implemented previously to identify 25 interactions among 55 proteins of phage T7[7] and to evaluate the importance of ionic interactions at specific positions in GCN4 leucine zipper mutants[8]. However, this is the first description of large-scale library-versus-library selection of protein interactions for the optimization of pro-

Marc Ostermeier (mao10@psu.edu) is a postdoctoral fellow and Stephen Benkovic (sjb1@psu.edu) is a professor in the chemistry department at Pennsylvania State University, University Park, PA 16802.

tein–protein interactions. Since the final heterodimerizing peptides were selected in vivo, they are optimized not only for heterodimerization, but also for expression and proteolytic stability. They should thus be particularly useful in protein heterodimerizing strategies—as in, for example, the design of mini-antibodies. In addition, the system should be very useful in the elucidation of the network of interactions among an organism's proteins—an advance that is critical for a better understanding of cellular functions and is analogous for protein interactions to sequencing an entire genome. Thus, as in the story of Cinderella, the mDHFR PCA story is well on its way to a happy ending.

1. Pelletier, J.N. et al. *Nat. Biotechnol* **17**: 683–690 (1999).
2. Fields, S. & Song, O.–K. *Nature* **340**, 245–246 (1989).
3. Vidal, M. & Legrain, P. *Nucleic Acids Res.* **27**, 919–929 (1999).
4. Pelletier, J.N. et al. *Proc. Nat. Acad. Sci. USA* **95**, 12141–12146 (1998).
5. Remy, I. & Michnick, S.W. *Proc. Nat. Acad. Sci. USA* **96**, 5394–5399 (1999).
6. Remy, I. et al. *Science* **283**, 990–993 (1999).
7 Bartel, P. et al. *Nat. Genet.* **12**, 72–77 (1996).
8. Zeng, X. et al. *Protein Sci.* **6**, 2218–2226 (1997).

On the wagon—DNA polymerase joins "H-bonds anonymous"

Myron F. Goodman

The concept that hydrogen bonding between a DNA template base and incoming dNTP substrate base is required to achieve high-fidelity DNA synthesis emanates from the Watson–Crick model of the DNA double-helix and subsequent discovery of *Escherichia coli* DNA polymerase I in the 1950s. In the ensuing four decades, we have seen few challenges to the primacy of hydrogen bonding in the replication hierarchy. However, recent studies by Eric Kool and colleagues[1,2] reveal that polymerases can indeed be "persuaded" to abandon their penchant for using hydrogen-bonding partners when synthesizing DNA.

In their latest paper, appearing in *Nature*[1], it is shown that a pyrene nucleotide can be incorporated efficiently opposite a template site lacking a base. This study can be viewed as "complementary" (pun intended) to earlier studies showing that DNA polymerase incorporates difluorotoluene opposite a template adenine and vice versa, even though this isosteric analog of thymine is unable to hydrogen-bond with adenine[2].

DNA polymerases normally synthesize DNA with exceptional accuracy. Base substitution errors typically occur with very low frequencies, ranging from 10^{-4} to 10^{-7} depending on the identity of the polymerase, on the base mispair, and on the DNA sequence context. Polymerase-associated 3'-exonuclease proofreading can excise from 95 to 99.5% of nucleotide mismatches at the growing replication fork, thereby reducing error rates to around 10^{-6}–10^{-9}. A 3'-exonuclease-deficient modified form of *E. coli* pol I (Klenow fragment exo$^-$) was used by Kool and colleagues to measure just the insertion specificity of pyrene and difluorotoluene substrate bases in the absence of proofreading.

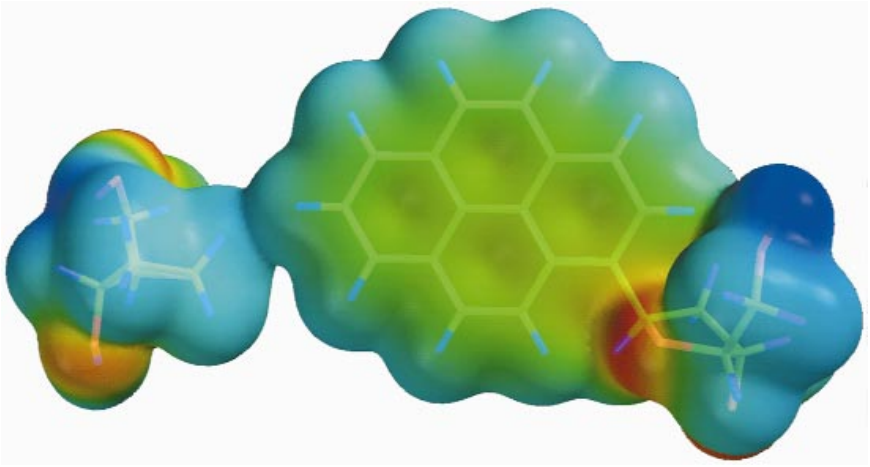

Figure 1. Spacefilling model of a nonpolar base-pair analog consisting of pyrene opposite an abasic site in B-form geometry (Courtesy E.T. Kool).

Just how do polymerases distinguish between a right and a wrong nucleotide for insertion into DNA, and what is the contribution of hydrogen bonding to pol insertion fidelity? Before Kool's studies, Harrison Echols, John Petruska, and I had expressed serious doubts about the importance of hydrogen bond contributions to fidelity because free-energy differences between base pairs and mispairs in DNA in aqueous solution appeared far too small to account for pol insertion fidelity[3]. These small free-energy differences can be attributed to a phenomenon called "enthalpy–entropy compensation," in which enthalpic differences between right and wrong base pairs are counteracted by strongly correlated entropic differences, thereby minimizing △△G values (△△G = △△H–T △△S). Instead, we proposed a geometric selection principle in which insertion was strongly favored for substrates with bond angles and distances that conformed most closely to Watson–Crick base-pairing geometry. By restricting the movement of substrates in the active cleft, polymerases might suppress entropic differences between right and wrong bound dNTPs, perhaps by primarily using the enthalpic component as a source of free energy, which is sizable enough to account for pol insertion fidelity[4].

In accordance with this principle of geometric selection, Matray and Kool[1] suggested that canonical purine and pyrimidine shapes might not be required for insertion by polymerase if both partners were accommodated in the double helix without steric strain. So they designed a non–hydrogen-bonded pair having one partner as small as possible—an abasic moiety, and the other large enough to take up the rest of the available space in the helix. A space-filling model depicting pyrene aligned opposite an abasic moiety, shown in Fig. 1, is close in surface area to a T•A base pair[1].

In their experiments, a comparison was made between incorporation efficiencies of pyrene opposite the abasic site with

Myron F. Goodman is professor in the department of biological sciences, University of Southern California, Los Angelas, CA (mgoodman@mizar.usc.edu).