Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Chemical Biology

Computational protein design of ligand binding and catalysis

Kaspar Feldmeier and Birte Höcker

The vision of custom-made proteins by computation appears closer than ever. Computational methods have advanced rapidly in recent years and proteins have been designed to catalyze new reactions. A number of second-generation enzyme designs analyzed possible bottlenecks and started tackling emergent problems. Detailed experimental analysis combined with structure determination and molecular dynamics simulations as well as design optimization with directed evolution techniques have led to important insights. While ligand recognition seems to be particularly problematic, new approaches focus on this design aspect and promising improvements have been made.

Addresses
Max Planck Institute for Developmental Biology, Spemannstrasse 35, 72076 Tübingen, Germany

Corresponding author: Höcker, Birte (birte.hoecker@tuebingen.mpg.de)

Current Opinion in Chemical Biology 2013, **17**:929–933

This review comes from a themed issue **Synthetic biomolecules**

Edited by **Shang-Cheng Hung** and **Derek N Woolfson**

For a complete overview see the Issue and the Editorial

Available online 26th October 2013

1367-5931/$ – see front matter,

http://dx.doi.org/10.1016/j.cbpa.2013.10.002

Introduction

Custom-made protein design has been a long-standing goal for biochemists. Since molecular biology techniques became available to tinker with protein sequences, it enabled exciting insights into catalytic mechanisms and simultaneously prompted the dream to rationally build new reactivities and specificities into enzymes [1] or even design entire proteins from first principles [2].

Immense advances have been made on both ends in the past years, mainly driven by the increasing use of powerful algorithms as well as the vastly growing knowledge of protein and enzyme structures. Just last year Koga *et al.* showed that *de novo* design of small, idealized single domain proteins based on general design rules can be performed with high success rates [3••]. It has yet to be combined with the design of functional sites that generally are in contrast to idealized structures and whose destabilizing effects will have to be compensated [4]. However, rational enzyme design has been successful in the repurposing of existing enzymes and the use of natural proteins as scaffolds for the construction of new functions. In this essay we focus on recent progress in the computer-based design of active site and ligand binding pockets. Recent advances and persistent challenges of current available methods are discussed.

A blueprint for design and early progress

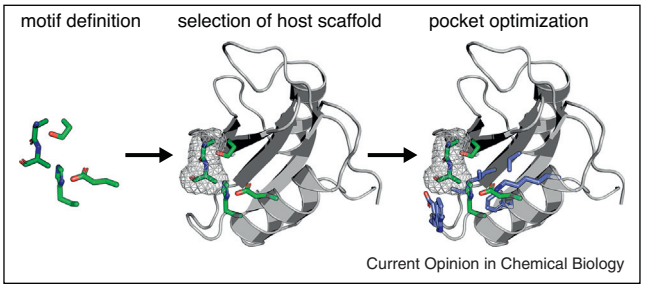
The general approach still used today is well described by Hellinga and Richards already in their 1991 paper on one of the first molecular modeling programs that aim to introduce new binding sites into proteins: 'In order to construct a new enzymatic function it is necessary to model simultaneously the geometry of functional groups and the shape of a complementary surface around a ligand' [5]. Particularly attractive is the wide applicability of this approach to virtually any kind of reaction as well as protein fold. In practice, design procedures usually divide this problem into two stages: the first step models the selected geometrically defined functional motif into a protein or set of proteins, while the second step deals with the optimization of the complementary surface in the intended binding pocket (Figure 1).

Early examples of computational designs of catalytic reactions based on this principle include a non-heme iron superoxide dismutase reaction, which was constructed simply via the first step by placing a metal-binding site that allows the metal to react with the superoxide anion [6], and a basic hydrolase activity that was built in a two-step approach by first placing a histidine nucleophile and then optimizing the active site through 2–3 additional mutations [7]. Both designs were realized in the protein thioredoxin.

Scaling up to tackle more complex reactions

A critical requirement for the success of this rational approach is the initial definition of the catalytic motif based on the chosen mechanism for the target reaction. In an idealized enzymatic motif, also called a theozyme, the functional residues have to be placed optimally for transition-state stabilization, which can be calculated by quantum mechanics (QM) [8]. To increase the possibilities for finding a good insertion site, the search is not limited to a single protein anymore. Search programs have been developed to identify the best match for a given motif in a large set of protein scaffolds [9,10]. Once a scaffold is selected and a motif docked into it, the surrounding residues have to be adjusted to interact optimally with the ligand and support the functional role of the catalytic residues. Programs developed to improve the packing of a protein fold such as the Rosetta suite [11] and ORBIT [12] can be used to optimize packing around a theozyme. Designs can then be further improved using directed evolution techniques to enhance the newly established function. This combined approach has been

Figure 1



Schematic overview of the general design procedure. First a catalytic geometry is defined (shown in green) and protein databases are searched for the optimal insertion site in a protein scaffold. In a second step the pocket around the insertion site is optimized (shown in blue) to accommodate the inserted motif and to promote interaction with the ligand.

used successfully first for the design of a retro-aldol activity [13], a proton transfer from carbon known as Kemp elimination [14], and a Diels–Alder reaction [15] (Table 1).

Recent achievements

With the basic design strategy established, the usability of the methods was explored for a number of additional enzymatic reactions. One example is the design of a cysteine–histidine dyad that catalyzes the cleavage of an ester [16]. The reaction has to overcome a significant energy barrier and also requires a tightly defined catalytic geometry without much tolerance for variation. The first motif used for the scaffold search comprised the dyad as well as backbone nitrogen atoms and an asparagine or glutamine that would serve as an oxyanion hole; but in a data set of 214 structures no insertion site could be found. However, a second, smaller motif consisting of only the dyad and one backbone amide led to the discovery of possible insertion sites. The binding pocket was optimized with Rosetta [11], 55 designs based on four different theozymes were tested, 32 expressed solubly, and four showed measurable activity. The designs provided an initial rate acceleration of up to 4000-fold, but they did not efficiently catalyze multiple turnover due to slow hydrolysis of the acyl-enzyme intermediate. Further analysis of crystal structures and molecular dynamics (MD) simulations indicated that this might be due to suboptimal positioning of the dyad. Another possibility is that the theozyme that was the basis for all four active designs does not capture all steps of the reaction. The analysis gives some clear directions to improve the strategy for future designs.

Another challenge was met by the redesign of a metalloenzyme to catalyze an organophosphate hydrolysis [17]. Using a natural zinc-binding site as a template, insertion sites were identified with RosettaMatch [9]. Twelve of these sites were redesigned considering shape complementarity and hydrogen bond interactions. Of the twelve designs, six were solubly expressed and one, a redesigned adenosine deaminase, showed activity in the phosphatase assay (k_{cat}/K_M of 4 M^{-1} s^{-1}). This basal activity enabled further improvements via directed evolution leading to a variant with a k_{cat}/K_M of 10^4 M^{-1} s^{-1}. The computational success was evaluated by screening of saturation mutagenesis libraries, which identified additional residues in the second shell. This suggests to include residues of a wider berth around the catalytic motif in the calculations. Structural analysis further revealed conformational differences between model and design, displaying the well-known need for backbone flexibility in the design methods.

The most recently reported activity design was on the Morita–Baylis–Hillman reaction, a multistep reaction that leads to carbon–carbon bond formation. In the first design step a particularly large set of scaffolds was searched [18]. Because of the complexity of the catalytic geometry motif, no insertion site was found in the initial search of a database containing 244 structures. Using a modified, faster RosettaMatch algorithm the search was expanded to the complete Protein Data Bank. Of the 48 designs

Table 1

Designed enzymes and their rate accelerations. Many designs were optimized subsequently by directed evolution. Here we list the best rates achieved after the computation before additional rounds of optimization

Reaction	Rate acceleration	k_{cat}/K_M [M^{-1} s^{-1}]	Remarks
Retro-aldol (2008)	$k_{cat}/k_{uncat} = 2.3 \times 10^4$	0.74 M^{-1} s^{-1}	[13]
Retro-aldol (2012)	$k_{cat}/k_{uncat} = 2.4 \times 10^4$	0.27 M^{-1} s^{-1}	78% success rate on a broad range of scaffolds [19•]
Kemp elimination (2008)	$k_{cat}/k_{uncat} = 2.5 \times 10^5$	163 M^{-1} s^{-1}	[14]
Kemp elimination (2012)	$k_{cat}/k_{uncat} = 5.9 \times 10^5$	425 M^{-1} s^{-1}	Based on structure analysis and MD simulations [21••]
Diels–Alder (2010)	Effective molarity: 89 M	4.7 M^{-1} s^{-1}	[15]
Diels–Alder (2011)		87.3 M^{-1} s^{-1}	Improved original design by crowdsourcing (foldit) [22•]
Ester hydrolysis (2001)	$k_{cat}/k_{uncat} = 180$	2.7 M^{-1} s^{-1}	Histidine nucleophile [7]
Ester hydrolysis (2012)	$k_{cat}/k_{uncat} = 4000$ in burst phase	405 $M^{-1}s^{-1}$	Cys His dyad & oxyanion hole [16]
Organophosphate hydrolysis (2012)	$k_{cat}/k_{uncat} = 4000$	4 M^{-1} s^{-1}	Uses a natural zinc-binding site [17]
Morita–Baylis–Hillman (2013)	$v/v_0 = 27$		Optimized algorithm allowed full PDB search [18]