

# Computational Methods for Enzyme Design and Its Biological Significance

Prabodh Sarmah, Devajit Mahanta

**Abstract:** *Enzymes are large biological molecules responsible for the thousands of metabolic processes that sustain life. They are highly selective catalysts, greatly accelerating both the rate and specificity of metabolic reactions, from the digestion of food to the synthesis of DNA. Most enzymes are proteins, although some catalytic RNA molecules have been identified. Enzymes adopt a specific three-dimensional structure, and may employ organic (e.g. biotin) and inorganic (e.g. magnesium ion) cofactors to assist in catalysis. Multiple experimental approaches have been applied to generate nearly all possible mutations of target enzymes, allowing the identification of desirable variants with improved properties to meet the practical needs. Meanwhile, an increasing number of computational methods have been developed to assist in the modification of enzymes during the past few decades. With the development of bioinformatics algorithms, computational approaches are now able to provide more precise guidance for enzyme engineering and make it more efficient and less laborious. In this review, we summarize the recent advances of method development with significant biological outcomes to provide important insights into successful computational protein designs.*

**Keyword:** *Enzymes, computational approaches*

## I. INTRODUCTION

Enzymes are large biological molecules responsible for the thousands of metabolic processes that sustain life.[1][2] They are highly selective catalysts, greatly accelerating both the rate and specificity of metabolic reactions, from the digestion of food to the synthesis of DNA. Most enzymes are proteins, although some catalytic RNA molecules have been identified. Enzymes adopt a specific three-dimensional structure, and may employ organic (e.g. biotin) and inorganic (e.g. magnesium ion) cofactors to assist in catalysis.

In enzymatic reactions, the molecules at the beginning of the process, called substrates, are converted into different molecules, called products. Almost all chemical reactions in a biological cell need enzymes in order to occur at rates sufficient for life. Since enzymes are selective for their substrates and speed up only a few reactions from among many possibilities, the set of enzymes made in a cell determines which metabolic pathways occur in that cell.

Like all catalysts, enzymes work by lowering the activation energy ( $E_a$ ) for a reaction, thus dramatically increasing the rate of the reaction. As a result, products are formed faster and reactions reach their equilibrium state more rapidly. Most enzyme reaction rates are millions of times faster than those of comparable un-catalyzed reactions.

As with all catalysts, enzymes are not consumed by the reactions they catalyze, nor do they alter the equilibrium of these reactions. However, enzymes do differ from most other catalysts in that they are highly specific for their substrates. Enzymes are known to catalyze about 4,000 biochemical reactions.[3] A few RNA molecules called ribozymes also catalyze reactions, with an important example being some parts of the ribosome.[4][5] Synthetic molecules called artificial enzymes also display enzyme-like catalysis.[6]

Enzyme activity can be affected by other molecules. Inhibitors are molecules that decrease enzyme activity; activators are molecules that increase activity. Many drugs and poisons are enzyme inhibitors. Activity is also affected by temperature, pressure, chemical environment (e.g., pH), and the concentration of substrate. Some enzymes are used commercially, for example, in the synthesis of antibiotics. In addition, some household products use enzymes to speed up biochemical reactions (e.g., enzymes in biological washing powders break down protein or fat stains on clothes; enzymes in meat tenderizers break down proteins into smaller molecules, making the meat easier to chew. Rational design, the earliest approach applied to the modification of enzymes [3-5], requires the availability of detailed structural information and catalytic mechanism of the targets. Computational tools have been developed to deal with a large number of data produced in rational enzyme design. In the meanwhile, such development leads to the emergence of “de novo computational design” approach [6], which commonly refers to the generation of novel protein scaffolds or enzymatic activity. Limited but exciting goals have been achieved in this field [7-9], making de novo computational design a promising approach in enzyme engineering. As another common methodology, directed evolution, was only applied to improve desired properties of enzymes recently [10, 11], but it has quickly become a powerful and popular tool in enzyme engineering [12]. Nevertheless, the bottleneck of directed evolution lies in the development of an efficient high-throughput screening technology, despite that there are quite a few successful examples that used directed evolution to modify important commercial enzymes. Consequently, the combined approaches involving rational or de novo design with directed evolution may offer significant advantages over individual approaches [8].

In paper, we highlight the strengths of a number of effective computational methodologies/tools that can assist in the rational and de novo enzyme design (see Figure 2). Successful examples, especially those concerning improvement of enzymatic activity and stability, which are the most important properties from a practical perspective, are discussed in the following respective sections.

**Manuscript received on April 2014.**

**Dr. Prabodh Sarmah**, Associate Professor Department Of Botany Nalbari College, Assam India.

**Mr. Devajit Mahanta**, Department of Computer science MSJ College Of Management And Technology, Assam India.

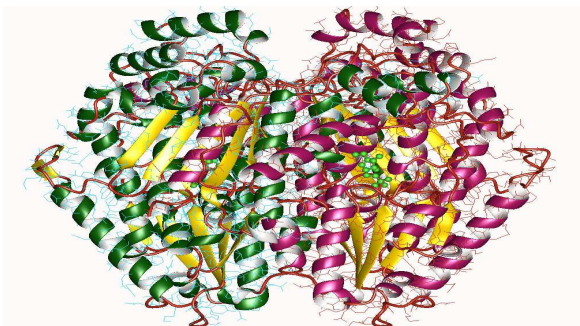


Figure 1: Natural Enzyme

## II. DESIGN AND TOOLS

The success of rational design depends on our in-depth knowledge about sequence and structure features of target proteins. A popular strategy to identify functionally related residues of unknown targets is the use of sequence features. Analysis of these features can provide enough information about evolutionary relationship, functional sites, correlated mutations and so on. The most useful tools for extracting sequence information are multiple sequence alignment (MSA) and coevolutionary analysis while the latter sometimes requires structural information. As a matter of fact, structure-based design is no doubt more efficient to locate key residues, because the execution of the protein function is directly linked with the maintenance of the 3D structure in functionally related regions. Structure-based rational design can benefit considerably from the rapidly growing number of solved protein structures, however, these account for only a small portion of naturally occurring proteins. To make a better use of structural information, 3D structure prediction or analysis tools are extremely important and greatly desired. Fortunately, a variety of computational methodologies/tools have been available to facilitate processing and data analysis, which have significantly contributed to the progress of rational enzyme design. Among them, several noteworthy tools are discussed below.

## III. CO EVOLUTIONARY ANALYSIS

Coevolution (also known as covariation, correlated mutation or co-substitution) refers to “reciprocal evolutionary change in evolutionarily interacting loci” [43], and occurs at many levels in biology. In this review, only the correlated mutations between amino acids within a protein are discussed. Coevolutionary analysis methods have a number of important applications in the prediction of protein structure identification of functional sites and candidate design sites. The identified coevolving residues have been experimentally validated in some studies implying the potential application of coevolutionary analysis in rational enzyme design.

In the past few decades, a number of coevolutionary analysis algorithms have been developed [56]. These methods share a common procedure of three steps: MSA construction, coevolutionary measure calculation and experimental validation. Most coevolutionary analyses start with the construction of an MSA of the query protein. Although certain automatic software can be applied (see Table 1), manual refinement, including filtering of sequences with large gaps, low homology or wrong annotation, is often required to ensure a high-quality MSA [57]. The second step is to calculate coevolutionary

measures, which can be done by using different correlated mutation algorithms, followed by statistical significance tests and analyses to extract significant convolution values, eliminate background noise [58] and evaluate the performance and robustness of the convolution measures [59]. Finally, “wet” experiments need to be performed to validate the obtained co evolutionary results. For experimental scientists, co evolutionary web servers seem to be more straightforward, attractive and practical. Up to now, several online tools have been made publicly available [56, 60]. However, how to choose an optimal scoring function of co evolutionary measures in the second step remains to be a critical factor that will determine the quality of co evolutionary analysis. To address this, Fodor et al. [61] assessed the performance of four different methods in detecting co evolutionary site, namely Statistical Coupling Analysis (SCA) [62], Observed Minus Expected Squared (OMES) [63], McMahan Based Substitution correlation (McBASC) [64] and Mutual Information (MI) [57]. In their research, OMES and McBASC were found to outperform the other two algorithms in favoring poorly conserved residue pairs and decreasing sensitivity to background conservation, and were of considerable similarity in Sensitivity to background noise. The OMES-based programs, OMES-KASS [63] and Fodor package [61], which were more recently developed, have been applied to perform reliable co evolutionary analysis.

Yip et al. developed an integrated online program by embedding several coevolutionary algorithms into one system instead of using a single algorithm only. These algorithms include SCA, MI, Explicit Likelihood of Subset Variation (ELSC) [68] and correlation-based methods [64, 69], making this system a convenient comparative analysis tool of different co evolutionary methods. The integrated system also provides an MSA preprocessing option to further improve its performance. In addition, users can also choose to treat the gaps in the MSA as noise or as an additional 21<sup>st</sup> residue, based on the observation that gaps might contain important co evolutionary information [60]. Despite the functional significance, how to combine co evolutionary analysis with rational enzyme design remains a challenging issue. In 2011, Zeng and colleagues applied SCA to analyze the sequences of the regulatory domains of the aspartokinase (AK) family to characterize the allosteric interaction network [53] and integrated such information with rational enzyme design. AK is the central enzyme in the biosynthesis of aspartate family amino acids, and the allosteric inhibition of AK by end-products obstructs the [70]. As a result, their co evolutionary analysis of 500 sequences from the AK family identified 25 highly correlated positions, in which 14 sites were mutated to construct AK mutants of *C. glutamicum*. All the mutants showed resistance to allosteric inhibition to different extents, suggesting that the choice of target mutations was largely successful. In this study, a major strategy was to select residues that had the potential to interrupt allosteric interaction, whereas in researches that aim to modify other properties of enzymes, amino acid sites that regulate the target property can probably be selected as candidates according to expert knowledge or structural analysis. There were two general rules to mutate the wild-type amino acids at the selected sites: (i) mutating the wild-type amino acids to those with less usage frequency at the corresponding positions; (ii) or substituting the wild-type amino acids by those with different chemical properties with the purpose of making

more obvious changes in terms of the target properties [53]. In another work of Chen and co-workers, AK3 from *Escherichia coli* was investigated via an integrative analysis of convolution and molecular dynamics (MD) [71].

#### IV. D STRUCTURE OF PROTEIN

There are an increasing number of proteins with high-resolution solved 3D structures, greatly facilitating the rational and computational protein design. Numerous previous successes have shown that when 3D structural information is available, protein design can be much more precise and accurate [18, 72, 73]. It is apparent that the knowledge of 3D structure of the target enzyme is a prerequisite and foundation for structure-based design. Although only a small portion of proteins have authentic crystal structures, those with unknown structure information can be reliably modeled via protein 3D structure prediction software, provided that there is a known structure of one or several homologous proteins to the target protein [74, 75]. According to the availability of template structures, protein 3D structure prediction can be generally divided into two categories: homology modelling and ab initio modelling. The former refers to the construction of an atomic-resolution model of a protein from its primary sequence using the

experimentally solved 3D structure of a homologous protein as the “template”, while the latter is called “free modelling” or “de novo modelling” in some cases, referring to 3D structure prediction generated from scratch when structural analogs are not available or detectable. The majority of methods used in homology modelling can be further grouped into two types: comparative modelling (CM) [76] and threading [77]. The root mean square deviation (rmsd) of a CM constructed model from the structure obtained from experiments can usually achieve 1–2 Å when a highly homologous (>30% sequence identity) template is employed. Models with such accuracy can compete with the low-resolution X-ray or medium-resolution NMR structures [78]. In contrast, the threading approach usually has a remarkable performance when dealing with target protein modelling using relatively distant templates, and the corresponding rmsd is 2-6Å [79] with most errors occurring in loops. Ab initio modelling, however, continues to be the most challenging topic in protein 3D structure prediction. Although there has been an exciting progress in modelling small proteins, no substantial progress has been achieved in de novo structure prediction of proteins with more than 150 residues [80]. In view of this, we mainly focus on the homology modelling methods in this mini-review.

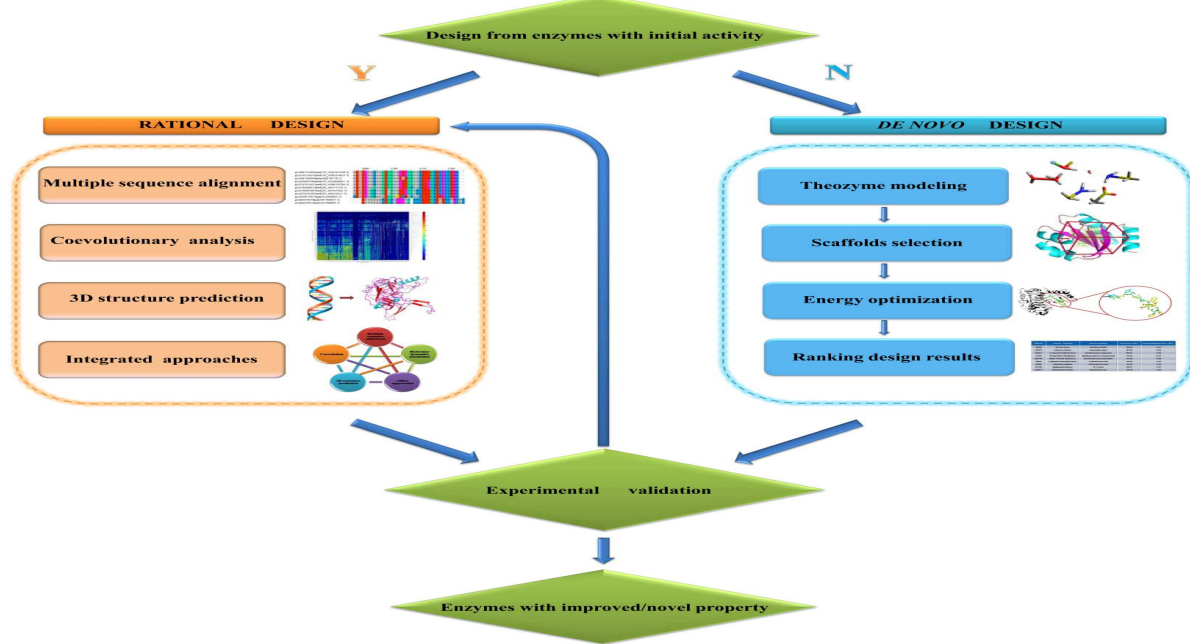


Figure 2 . Strategies of rational and de novo enzyme design

Table 1. Summary of useful computational programs in rational design referred in this review.

Programs	Application	URL address	Operating system	Ref.
<b>Rational design programs</b>				
ClustalW		<a href="http://www.clustal.org/clustal2/">http://www.clustal.org/clustal2/</a>	Windows, Linux, MacOS	[27, 122]
ClustalOmega		<a href="http://www.clustal.org/omega/">http://www.clustal.org/omega/</a>	Windows, Linux, MacOS	[30]
Mafft	Multiple sequence alignment	<a href="http://mafft.cbrc.jp/alignment/software/">http://mafft.cbrc.jp/alignment/software/</a>	Windows, Linux, MacOS	[32, 37, 39]
T-Coffee		<a href="http://www.tcoffee.org/Projects/tcoffee/">http://www.tcoffee.org/Projects/tcoffee/</a>	Linux, MacOS	[123]
Muscle		<a href="http://www.drive5.com/muscle/">http://www.drive5.com/muscle/</a>	Windows, Linux, MacOS	[33]
<b>Integrated system</b>				
OMES-KASS	Coevolutionary analysis	<a href="http://bip.weizmann.ac.il/correlated_mutations/">http://bip.weizmann.ac.il/correlated_mutations/</a>	Linux	[63]
Fodor package		<a href="http://www.afodor.net/">http://www.afodor.net/</a>	Windows, Linux, MacOS	[60]
Swiss-Model	3D structure prediction	<a href="http://swissmodel.expasy.org/">http://swissmodel.expasy.org/</a>	-	[124, 125]
HHpred2		<a href="http://toolkit.tuebingen.mpg.de/hhpred">http://toolkit.tuebingen.mpg.de/hhpred</a>	-	[83]
I-TASSER		<a href="http://zhanglab.ccmb.med.umich.edu/I-TASSER/">http://zhanglab.ccmb.med.umich.edu/I-TASSER/</a>	Linux	[84, 126]
FoldX		<a href="http://foldx.crg.es/">http://foldx.crg.es/</a>	Windows, Linux, MacOS	[96, 127]
PopMuSIC	Protein stability prediction	<a href="http://babybone.ulb.ac.be/popmusic">http://babybone.ulb.ac.be/popmusic</a>	-	[94, 97, 128]
I-Mutant3.0		<a href="http://gpccr2.biocomp.unibo.it/cgi/predictors/I-Mutant3.0/I-Mutant3.0.cgi">http://gpccr2.biocomp.unibo.it/cgi/predictors/I-Mutant3.0/I-Mutant3.0.cgi</a>	-	[129]
DMutant		<a href="http://sparks.informatics.iupui.edu/hzhou/mutation.html">http://sparks.informatics.iupui.edu/hzhou/mutation.html</a>	-	[130]
<b>De novo design programs</b>				
RosettaMarch	Scaffold search	-	-	[108]
RosettaDesign	Protein design for low free energy sequences	<a href="http://rosetta-design.med.unc.edu/">http://rosetta-design.med.unc.edu/</a>	Linux	[109]
ORBIT	Optimal sequences search for given folds	-	-	[118]

By submitting an amino acid sequence or its UniProtID, users start the modelling procedure with or without providing a template protein. Swiss-Model server can automatically select several suitable templates from a refined library derived from the Protein Data Bank (PDB), and then a structural alignment between the target and the template is generated and improved for the sake of modelling [87]. A pH sensitive member Kir2.3 was aligned with all the Kir2 channel proteins, and histidine 117 (H117) located close to the putative selectivity filter was identified to contribute to pH sensitive phenotype [91]. However, contradictory results were obtained by directed mutagenesis experiments, suggesting that there were other factors related to the pH effect. The observation that the ability of Zn<sup>2+</sup> to bind cysteines/histidines could inhibit the pH effect indicated that a cysteine within atomic distance to H117 might interact to exert this functional effect. Consequently, the 3D structure of Kir2.3 was created by Swiss-Model using distant templates in order to narrow down the range and locate the target cysteine. The active sites of the predicted structure were then superimposed on the template and indicated that the spatial orientation of D199 in the target EH was different from its counterpart in the template. Attempts to modify D199 into a proper orientation were also made to redesign the surrounding residues so that they could have direct or indirect interactions with D199. To achieve this, F193 and Y194 were chosen, and the 3D structures of various mutants of these two residues were constructed by Swiss-Model instead of “wet” experiments. Analysis of the corresponding 3D structures, particularly

#### V. DE NOVO COMPUTATIONAL DESIGN

The ultimate test of our understanding of the mechanism of enzymatic catalysis is de novo computational design, which refers to creation of novel protein folds, substrate binding pockets, and catalytic activities and so on. De novo protein design was first conducted to create a four-helix bundle protein in 1988 [6]. Since then, various protein folds have been de novo designed [100]. However, only a few possessed catalytic functions. Accordingly, de novo computational design of naturally occurring enzymes with novel catalytic activity is considered as a grand challenge, and in recent years, great efforts in this field have been made to expand our knowledge in enzyme engineering [7-9, 101-103]. To illustrate this, in this section we discuss three distinguished design examples of enzymes that catalyze synthetic reactions. The overwhelming performance of enzymatic catalysis over chemical catalysis is partly due to the free energy decrease of transition state (TS) via the interaction with catalytic residues [104]. Hence, the first step of de novo design for a given reaction is to model its theozyme which consists of TS model and catalytic groups [105] based on quantum chemical calculations [106]. How well the theozyme models correlate with their corresponding crystal structures, will have a significant influence on the ultimate designs. Dechancie et al. mimicked the active sites of nine distinct enzymes with quantum mechanical optimizations [107]. The rmsd of the sets of catalytic atoms was 0.64Å, suggesting that the predicted geometries were remarkably consistent with the corresponding X-ray structure. For a desired reaction, there usually exist more than one possible catalytic mechanism. As result, the 3D models of each catalytic motif for each mechanism will have to be built, and hence the degree of

freedom and the orientation of different bonds in each model can vary greatly, giving rise to a great number of possible 3D active sites, which are called “theozyme library”. The search for optimal protein scaffolds that are able to fulfill a target reaction can be launched once the theozyme library has been generated. Numerous scaffolds with ligand-binding cavities and high-resolution X-ray structures are available in several public protein databases. If there are certain restrictions on potential scaffolds, for example, in cases where a thermophilic scaffold is required, the selection range could be narrowed down. However, this process depends on de novo design algorithms such as Rosetta Match [108] that relies on hashing techniques and pruning of the majority of potential active centers at a very high speed but very little cost. At this step, the description of TS and a set of protein scaffolds are input into Rosetta Match. Once a TS position is compatible with the geometry of catalytic sites in one scaffold and satisfies other catalytic constraints, the result will be output as a “match” [106, 108]. Because there are still substantial candidate matches after the scaffold selection, and there remain certain steric clashes between the TS position and the catalytic side chains in the matches, further optimization is necessary. In this regard, the Rosetta Design methodology [109] can be applied to improve the binding affinity to TS and the stability of the active centers by redesigning or repacking of related residues. It is suggested that users run a single task for ten times owing to stochastic sampling algorithm adopted by Rosetta Design which will probably produce 10 distinct outputs. The resulting designs are supposed to be lower energy sequences for a given scaffold with the maximized TS affinity.

After optimizing all unique matches, a next step is to select designs with optimal performance for experimental validation. Several important factors, especially the ligand binding energy feature, are often used to evaluate and rank all the designs as described in [106]. As it is unlikely that a design has the highest score for each factor, extensive examinations to assist in further selection are preferred. In addition, Kiss et al. found that the MD technology was the most effective procedure for predicting the catalytic potentiality of designs [110].

The same protein scaffolds can execute diverse functions, such as  $\alpha/\beta$ -barrel motif, which constitutes approximately 10% of proteins that perform a wide range of catalytic reactions [111]. This indicates that the designable potentiality of certain scaffolds underlies the foundation of computational engineering of novel functions. With similar strategies, Baker’s group has performed a series of pioneering studies in redesigning enzymes that catalyze retro-aldol reaction [7], Kemp elimination [8] and Diels-Alder reaction [9]. The enhancement of target reactions by designed enzymes was assessed by the ratio of the catalytic rate constant and uncatalyzed rate constant  $k_{cat}/k_{uncat}$ . In the above cases, the values of  $k_{cat}/k_{uncat}$  ranged from 10<sup>2</sup> to 10<sup>5</sup> for the most active designs, indicating the effectiveness of such design strategies. De novo computational enzyme design provides important insights into the structure-function relationship of the enzyme and the starting points for directed evolution and rational design. Considerable experimental efforts, including development of technologies discussed in the Rational design strategies and tools section, were made to enhance the activities of the artificial Kemp eliminases [112-114]. Subsequent crystal and experimental data confirmed the accuracy of the

predicted model and an apparently increased activity. These examples discussed above highlight the importance and complementarily of these alternative de novo design

strategies, which can be applied to similar scaffold-based studies.

Table 2. Summary of representative examples referred in this review.

Enzyme/protein	Target property	Method		Result			Ref.
		Design strategy	Bioinformatic tool	No. of mutants	Fold-improvement	Library size	
<i>R. spartanensis</i> -1,4-glucanase	Activity	Functional and activity-related residues identified via an MSA analysis of eight sequences	-	7	7-13	24	[19]
<i>S. capitata</i> prolyl endopeptidases	Activity; stability	An MSA of 100 homologues evaluated by multiple scoring functions identified mutations	ClustalW, SeqDist, KaKs, probCons, SUB	6(1st round) 9(2nd round)	20%(activity) 200(stability)	47(1st round) 48(2nd round)	[28]
KDO8P Synthase family	Stability	Integrated analysis by MSA, G changes calculation, MD simulation and coevolutionary analysis	Maffr, T-Coffee, Muscle, HMMER 3.0, Prime 2.1, Desmond, X-Cluster, FoldX	No experimental validation			[42]
<i>C. glutamicum</i> aspartokinase	Allosteric inhibition	Correlated positions were identified by coevolutionary analysis of 500 sequences	Muscle, ClustalX	1	2	14	[53]
<i>E. coli</i> aspartokinase	Allosteric inhibition	Integrated analysis by MD simulation and coevolutionary analysis	Modeller, AMBER, Muscle	6	5-7	18	[71]
<i>M. cephalus</i> epoxide hydrolase	Activity	Activity-related residues were identified by superimposition of a predicted structure and a solved structure template	Swiss-Model, RasMol, DeepView	1	35	5	[92]
<i>K. pneumoniae</i> glyceroldehydratase	pH stability; activity	Stability-related residues were designed based on a predicted structure	Swiss-Model, PoPMuSiC	1	2(pH stability) 2(activity)	2	[93]
Retro-aldol reaction	Activity	TS simulated by QM/MM was used for scaffold selection and followed by individual optimization and ranking	RosettaMatch, RosettaDesign	32	2 × 10 <sup>6</sup>	72	[7]
Kemp elimination				8	>10 <sup>6</sup>	59	[8]
Diels-Alder reaction				2	89M	84	[9]
<i>E. coli</i> thioredoxin	PNPA hydrolase	Potential active sites and surrounding active-site mutations were identified and computed	ORBIT	2	180	2	[115]
Sperm whale myoglobin	Nitric oxide reductase	Creating a non-haem Fe <sup>2+</sup> -binding site based on the predicted structure overlaid with the reference structure	VMD, NAMD	1	N.A.	<10	[117]

N.A.: Not Available

## VI. CONCLUSION

The aim of this paper to provide a useful guide on the selection of the basic design methodologies and tools that are frequently employed in enzyme engineering (Table 1), and a brief summary of these aforementioned examples is depicted in Table 2. For many naturally occurring enzymes, it is often necessary to modify and design their properties in order to meet the needs of commercial or industrial applications. Bioinformatic strategies and tools, particularly those with freely accessible webservers, offer biologists tremendous help to narrow down their experimental efforts. MSA can efficiently identify consensus, highly conserved and variable positions within a family of homologous proteins, while MSA-based coevolutionary analysis of a set of enzymes with similar functions provide critical clues about catalytic and other functionally related residues. with the increasing availability of high-quality 3D structures in the PDB, there are a growing number of structure-based approaches being developed. Because experimentally solved structures only cover a limited portion of the protein repertoire, sequence-based 3D structure prediction has become a prevalent methodology in enzyme engineering. This is important, because reliable prediction of protein structure can still provide valuable information regarding potential candidate sites whose mutations might lead to improved properties of the enzyme, even if its real structural information is not at hand. As a symbol of the engineering of the third wave of biocatalysts [119], de novo enzyme design has achieved a significant success in the last 20 years. Despite these advances, there are challenges for

rational enzyme design. A first challenge is that there are inevitable experimental errors in “wet” experiments [120], resulting in less reliable designs based on such low-quality data.

A second challenge is related to the conformational dynamic nature of the enzyme. Conformational changes of the enzyme are frequently occurring under catalytic conditions, leading to a deviation of the real orientation of residues and enzyme structure from that of the designed or modeled enzymes. A third challenge is how to select the most appropriate tool that best suits the study of a particular target enzyme, from a pool of different tools that have both pros and cons. In this mini-review, we attempt to provide a useful guide to summarize some of the popular, reliable and academic free tools. Moreover, many examples have proved that integrative strategies can usually outperform individuals. In this regard, development of meta-servers is promising for providing a better performance and reliability of computation design. A fourth challenge is that some modified catalysts still cannot meet the practical needs of large-scale applications, particularly de novo designed enzymes. As such, there is often a need for assistance of experimental approaches, such as directed evolution. In fact, the boundary of rational design and directed evolution has become more and more blurred in practical applications, as evidenced by a number of recent studies that involve a combination of both [5]. Therefore, improving experimental techniques, such as high-quality mutagenesis and high-throughput screening, is another related future direction. Due to the aforementioned challenges, many attempts of computational protein design failed. However, future development of the field will be advanced by a better understanding of the underlying reasons that led to both

failures and successes [121]. Recent advances in computational enzyme design have largely expedited the evolution of enzymes, and have greatly revolutionized the way of enzyme engineering. With the development of improved experimental techniques, computational enzyme design will gain a momentum and achieve significant successes in the future.

REFERENCE

1. Smith AL (Ed) (1997). Oxford dictionary of biochemistry and molecular biology. Oxford [Oxfordshire]: Oxford University Press. ISBN 0-19-854768-4.
2. Grisham, Charles M.; Reginald H. Garrett (1999). Biochemistry. Philadelphia: Saunders College Pub. pp. 426–7. ISBN 0-03-022318-0.
3. Bairoch A. (2000). "The ENZYME database in 2000" (PDF). *Nucleic Acids Res* 28(1): 304–5. doi:10.1093/nar/28.1.304. PMC 102465. PMID 10592255.
4. Lilley D (2005). "Structure, folding and mechanisms of ribozymes". *Current Opinion in Structural Biology* 15 (3): 313–23. doi:10.1016/j.sbi.2005.05.002. PMID 15919196.
5. Cech T (2000). "Structural biology. The ribosome is a ribozyme". *Science* 289 (5481): 878–9. doi:10.1126/science.289.5481.878. PMID 10960319.
6. Groves JT (1997). "Artificial enzymes. The importance of being selective". *Nature* 389(6649): 329–30. doi:10.1038/38602. PMID 9311771.
7. de Réaumur, RAF (1752). "Observations sur la digestion des oiseaux". *Histoire de l'academie royale des sciences 1752*: 266, 461.
8. Williams, H. S. (1904) A History of Science: in Five Volumes. Volume IV: Modern Development of the Chemical and Biological Sciences Harper and Brothers (New York) Accessed 4 April 2007
9. Payen, A. et J.-F. Persoz (1833) "Mémoire sur la diastase, les principaux produits de ses réactions et leurs applications aux arts industriels" (Memoir on diastase, the principal products of its reactions and their applications to the industrial arts), *Annales de chimie et de physique*, 2nd series, 53 : 73–92.
10. Dubos J. (1951). "Louis Pasteur: Free Lance of Science, Gollancz. Quoted in Manchester K. L. (1995) Louis Pasteur (1822–1895)—chance and the prepared mind". *Trends Biotechnol* 13 (12): 511–5. doi:10.1016/S0167-7799(00)89014-9. PMID 8595136.
11. Kühne coined the word "enzyme" in: W. Kühne (1877) Jump up^ Text of Eduard Buchner's 1907 Nobel lecture at <http://nobelprize.org>. Retrieved 4 April 2007.
12. The naming of enzymes by adding the suffix "-ase" to the substrate on which the enzyme acts, has been traced to French scientist Jump up^ Willstätter, R. (1927). *Problems and Methods in Enzyme Research*. Cornell University Press, Ithaca. quoted in Blow, David (2000). "So do we understand how enzymes work?" (pdf). *Structure* 8 (4): R77–R81. doi:10.1016/S0969-2126(00)00125-8. PMID 10801479.
13. 1946 Nobel prize for Chemistry laureates at <http://nobelprize.org>. Retrieved 4 April 2007.
14. Blake CC, Koenig DF, Mair GA, North AC, Phillips DC, Sarma VR. (1965). "Structure of hen egg-white lysozyme. A three-dimensional Fourier synthesis at 2 Angstrom resolution". *Nature* 206 (4986): 757–61. doi:10.1038/206757a0. PMID 5891407.
15. Chen LH, Kenyon GL, Curtin F, Harayama S, Bembek ME, Hajipour G, Whitman CP (1992). "4-Oxalocrotonate tautomerase, an enzyme composed of 62 amino acid residues per monomer". *J. Biol. Chem.* 267 (25): 17716–21. PMID 1339435.
16. Smith S (1 December 1994). "The animal fatty acid synthase: one gene, one polypeptide, seven enzymes". *FASEB J.* 8 (15): 1248–59. PMID 8001737.
17. Anfinsen C.B. (1973). "Principles that Govern the Folding of Protein Chains". *Science* 181 (4096): 223–30. doi:10.1126/science.181.4096.223. PMID 4124164.
18. Dunaway-Mariano D (2008). "Enzyme function discovery". *Structure* 16 (11): 1599–600. doi:10.1016/j.str.2008.10.001. PMID 19000810.
19. The Catalytic Site Atlas at The European Bioinformatics Institute. Retrieved 4 April 2007.
20. Jaeger KE, Eggert T. (2004). "Enantioselective biocatalysis optimized by directed evolution". *Current Opinion in Biotechnology* 15 (4): 305–
21. Shevelev IV, Hubscher U. (2002). "The 3' 5' exonucleases". *Nature Reviews Molecular Cell Biology* 3 (5): 364–

22. Tymoczko, John L.; Stryer Berg Tymoczko; Stryer, Lubert; Berg, Jeremy Mark (2002). Biochemistry. San Francisco: W.H. Freeman. ISBN
23. Ibba M, Soll D. (2000). "Aminoacyl-tRNA synthesis". *Annu Rev Biochem.* 69: 617–50. doi:10.1146/annurev.biochem.69.1.617. PMID 10966471.
24. Rodnina MV, Wintermeyer W. (2001). "Fidelity of aminoacyl-tRNA selection on the ribosome: kinetic and structural mechanisms". *Annu Rev Biochem.* 70: 415–35. doi:10.1146/annurev.biochem.70.1.415. PMID 11395413.
25. Tawfik, O. K. A. D. S.; Tawfik, D. S. (2010). "Enzyme Promiscuity: A Mechanistic and Evolutionary Perspective". *Annual Review of Biochemistry* 79: 471–505. doi:10.1146/annurev-biochem-030409-143718. PMID 20235827. edit
26. Fischer E. (1894). "Einfluss der Configuration auf die Wirkung der Enzyme" [Influence of configuration on the action of enzymes]. *Berichte der Deutschen chemischen Gesellschaft zu Berlin* 27 (3): 2985–93. doi:10.1002/cber.18940270364. Jump up^ Koshland D. E. (1958). "Application of a Theory of Enzyme Specificity to Protein Synthesis". *Proc. Natl. Acad. Sci.* 44 (2): 98–104. doi:10.1073/pnas.44.2.98. PMC 335371. PMID 16590179.
27. , Davies GJ, Bohm M. (2002). "Glycosidase mechanisms". *Current Opinion in Chemical Biology* 6 (5): 619–29. doi:10.1016/S1367-5931(02)00380-0. PMID 12413546.
28. Boyer, Rodney (2002) [2002]. "6". *Concepts in Biochemistry* (2nd ed.). New York, Chichester, Weinheim, Brisbane, Singapore, Toronto.: John Wiley & Sons, Inc. pp. 137–8. ISBN 0-470-00379-0. OCLC 51720783.
29. Savir Y & Tlusty T (2007). "Conformational proofreading: the impact of conformational changes on the specificity of molecular recognition". In Scalas, Enrico. *PLoS ONE* 2(5): e468. doi:10.1371/journal.pone.0000468. PMC 1868595. PMID 1752 0027.
30. Chou K.C. (1996). "Review: Prediction of human immunodeficiency virus protease cleavage sites in proteins". *Analytical Biochemistry* 233 (1): 1–14. doi:10.1006/abio.1996.0001. PMID 8789141.
31. Gan, Y. R., Huang, H., Huang, Y. D., Rao, C. M., Zhao, Y., Liu, J. S., Wu, L. & Wei, D. Q. (April 2006). "Synthesis and activity of an octapeptide inhibitor designed for SARS coronavirus main proteinase". *Peptides* 27 (4): 622–625. doi:10.1016/j.peptides.2005.09.006. PMID 16242214.
32. Du, Q. S., Wang, S., Wei, D. Q., Sirois, S. & Chou, K. C. (2005). "Molecular modelling and chemical modification for finding peptide inhibitor against SARS CoV Mpro". *Analytical Biochemistry* 337 (2): 262–270. doi:10.1016/j.ab.2004.10.003. PMID 15691506.
33. Du, Q. S., Sun, H. & Chou, K. C. (2007). "Inhibitor design for SARS coronavirus main protease based on "distorted key theory"". *Medicinal Chemistry* 3 (1): 1–6. doi:10.2174/157340607779317616. PMID 17266617.
34. Fersht, Alan (1985). *Enzyme structure and mechanism*. San Francisco: W.H. Freeman. pp. 50–2. ISBN 0-7167-1615-1.
35. Jencks, William P. (1987). *Catalysis in chemistry and enzymology*. Mineola, N.Y.: Dover. ISBN 0-486-65460-5.
36. Villa J, Strajbl M, Glennon TM, Sham YY, Chu ZT, Warshel A (2000). "How important are entropic contributions to enzyme catalysis?". *Proc. Natl. Acad. Sci. U.S.A.* 97(22): 11899–904. doi:10.1073/pnas.97.22.11899. PMC 17266. PMID 11050223.
37. Warshel A, Sharma PK, Kato M, Xiang Y, Liu H, Olsson MH (2006). "Electrostatic basis for enzyme catalysis". *Chem. Rev.* 106 (8): 3210–35. doi:10.1021/cr0503106. PMID 16895325.
38. Eisenmesser EZ, Bosco DA, Akke M, Kern D (2002). "Enzyme dynamics during catalysis". *Science* 295 (5559): 1520–3. doi:10.1126/science.1066176. PMID 11859194.
39. Agarwal PK (2005). "Role of protein dynamics in reaction rate enhancement by enzymes". *J. Am. Chem. Soc.* 127 (43): 15248–56. doi:10.1021/ja055251s. PMID 16248667.
40. Eisenmesser EZ, Millet O, Labeikovsky W (2005). "Intrinsic dynamics of an enzyme underlies catalysis". *Nature* 438 (7064): 117–21. doi:10.1038/nature04105. PMID 16267559.
41. Yang LW, Bahar I (5 June 2005). "Coupling between catalytic site and collective dynamics: A requirement for mechanochemical activity of enzymes". *Structure* 13 (6): 893–904. doi:10.1016/j.str.2005.03.015. PMC 1489920. PMID 15939021.
42. Agarwal PK, Billeter SR, Rajagopalan PT, Benkovic SJ, Hammes-Schiffer S. (5 March 2002). "Network of coupled promoting motions in enzyme catalysis". *Proc Natl Acad Sci USA.* 99 (5): 2794–9. doi:10.1073/pnas.052005999. PMC 122427. PMID 11867722.
43. Agarwal PK, Geist A, Gorin A (2004). "Protein dynamics and enzymatic catalysis: investigating the peptidyl-prolyl cis-trans



- isomerization activity of cyclophilin A". *Biochemistry* 43 (33): 10605–18. doi:10.1021/bi0495228. PMID 15311922.
44. Tousignant A, Pelletier JN. (2004). "Protein motions promote catalysis". *Chem Biol* 11(8): 1037–42. doi:10.1016/j.chembiol.2004.06.007. PMID 15324804
  45. Flomenbom O, Velonia K, Loos D et al. (2005). "Stretched exponential decay and correlations in the catalytic activity of fluctuating single lipase molecules". *Proc. Natl. Acad. Sci. U.S.A.* 102 (7): 2368–2372. doi:10.1073/pnas.0409039102.PMC 548972. PMID 15695587.
  46. English BP, Min W, van Oijen AM et al. (2006). "Ever-fluctuating single enzyme molecules: Michaelis-Menten equation revisited". *Nature Chemical Biology* 2 (2): 87–94. doi:10.1038/nchembio759. PMID 16415859.
  47. Lu H, Xun L, Xie X S (1998). "Single-molecule enzymatic dynamics". *Science* 282(5395): 1877–1882. doi:10.1126/science.282.5395.1877. PMID 9836635.
  48. Olsson, MH; Parson, WW; Warshel, A (2006). "Dynamical Contributions to Enzyme Catalysis: Critical Tests of A Popular Hypothesis". *Chem. Rev.* 106 (5): 1737–.
  49. Neet KE (1995). "Cooperativity in enzyme function: equilibrium and kinetic aspects". *Meth. Enzymol. Methods in Enzymology* 249: 519–67. doi:10.1016/0076-6879(95)49048-5. ISBN 978-0-12-182150-0. PMID 7791626.
  50. Changeux JP, Edelstein SJ (2005). "Allosteric mechanisms of signal transduction". *Science* 308 (5727): 1424–8. doi:10.1126/science.1108595. PMID 15933191.
  51. de Bolster, M.W.G. (1997). "Glossary of Terms Used in Bioinorganic Chemistry: Cofactor". International Union of Pure and Applied Chemistry. Retrieved 30 October 2007.
  52. de Bolster, M.W.G. (1997). "Glossary of Terms Used in Bioinorganic Chemistry: Coenzyme". International Union of Pure and Applied Chemistry. Retrieved 30 October 2007.
  53. Fisher Z, Hernandez Prada JA, Tu C, Duda D, Yoshioka C, An H, Govindasamy L, Silverman DN and McKenna R. (2005). "Structural and kinetic characterization of active-site histidine as a proton shuttle in catalysis by human carbonic anhydrase II". *Biochemistry*. 44 (4): 1097–115. doi:10.1021/bi0480279. PMID 15667203.
  54. Wagner, Arthur L. (1975). *Vitamins and Coenzymes*. Krieger Pub Co. ISBN 0-88275-258-8.
  55. BRENDA The Comprehensive Enzyme Information System. Retrieved 4 April 2007.
  56. Törnroth-Horsefield S, Neutze R (2008). "Opening and closing the metabolite gate". *Proc. Natl. Acad. Sci. U.S.A.* 105 (50): 19565–6. doi:10.1073/pnas.0810654106.PMC 2604989. PMID 19073922.
  57. Ferguson, S. J.; Nicholls, David; Ferguson, Stuart (2002). *Bioenergetics 3* (3rd ed.). San Diego: Academic. ISBN 0-12-518121-3.
  58. Henri, V. (1902). "Theorie generale de l'action de quelques diastases". *Compt. Rend. Hebd. Acad. Sci. Paris* 135: 916–9.
  59. Sørensen, P.L. (1909). "Enzymstudien {II}. Über die Messung und Bedeutung der Wasserstoffionenkonzentration bei enzymatischen Prozessen". *Biochem. Z.* 21: 131–304.
  60. Michaelis L., Menten M. (1913). "Die Kinetik der Invertinwirkung". *Biochem. Z.* 49: 333–369. English translation. Retrieved 6 April 2007.
  61. Briggs G. E., Haldane J. B. S. (1925). "A note on the kinetics of enzyme action". *Biochem. J.* 19 (2): 339–339. PMC 1259181. PMID 16743508.
  62. Xue X, Liu F, Ou-Yang ZC (2006). "Single molecule Michaelis-Menten equation beyond quasistatic disorder". *Phys. Rev. E* 74 (3): 030902. doi:10.1103/PhysRevE.74.030902. PMID 17025584.
  63. Radzicka A, Wolfenden R. (1995). "A proficient enzyme". *Science* 267 (5194): 90–931. doi:10.1126/science.7809611. PMID 7809611.
  64. Ellis RJ (2001). "Macromolecular crowding: obvious but underappreciated". *Trends Biochem. Sci.* 26 (10): 597–604. doi:10.1016/S0968-0004(01)01938-7. PMID 11590012.
  65. Kopelman R (1988). "Fractal Reaction Kinetics". *Science* 241 (4873): 1620–26. doi:10.1126/science.241.4873.1620. PMID 17820893.
  66. Savageau MA (1995). "Michaelis-Menten mechanism reconsidered: implications of fractal kinetics". *J. Theor. Biol.* 176 (1): 115–24. doi:10.1006/jtbi.1995.0181. PMID 7475096.
  67. Schnell S, Turner TE (2004). "Reaction kinetics in intracellular environments with macromolecular crowding: simulations and rate laws". *Prog. Biophys. Mol. Biol.* 85 (2–3): 235–60. doi:10.1016/j.pbiomolbio.2004.01.012. PMID 15142746.
  68. Xu F, Ding H (2007). "A new kinetic model for heterogeneous (or spatially confined) enzymatic catalysis: Contributions from the fractal and jamming (overcrowding) effects". *Appl. Catal. A: Gen.* 317 (1): 70–81. doi:10.1016/j.apcata.2006.10.014.
  69. Garcia-Viloca M., Gao J., Karplus M., Truhlar D. G. (2004). "How enzymes work: analysis by modern rate theory and computer simulations". *Science* 303 (5655): 186–95. doi:10.1126/science.1088172. PMID 14716003.
  70. Olsson M. H., Siegbahn P. E., Warshel A. (2004). "Simulations of the large kinetic isotope effect and the temperature dependence of the hydrogen atom transfer in lipoxygenase". *J. Am. Chem. Soc.* 126 (9): 2820–8. doi:10.1021/ja037233l. PMID 14995199.
  71. Masgrau L., Roujeinikova A., Johannissen L. O., Hothi P., Basran J., Ranaghan K. E., Mulholland A. J., Sutcliffe M. J., Scrutton N. S., Leys D. (2006). "Atomic Description of an Enzyme Reaction Dominated by Proton Tunneling". *Science* 312 (5771): 237–41. doi:10.1126/science.1126002. PMID 16614214.
  72. Cleland, W.W. (1963). "The Kinetics of Enzyme-catalyzed Reactions with two or more Substrates or Products 2. {I}nhibition: Nomenclature and Theory". *Biochim. Biophys. Acta* 67: 173–87.
  73. Price, NC. (1979). "What is meant by 'competitive inhibition'?". *Trends in Biochemical Sciences* 4 (11): pN272. doi:10.1016/0968-0004(79)90205-6.
  74. R Poulin; Lu, L; Ackermann, B; Bey, P; Pegg,
  75. Yoshikawa S and Caughey WS. (15 May 1990). "Infrared evidence of cyanide binding to iron and copper sites in bovine heart cytochrome c oxidase. Implications regarding oxygen reduction". *J Biol Chem.* 265 (14): 7945–58. PMID 2159465.
  76. Hunter T. (1995). "Protein kinases and phosphatases: the yin and yang of protein phosphorylation and signaling". *Cell.* 80 (2): 225–36. doi:10.1016/0092-8674(95)90405-0. PMID 7834742.
  77. Berg JS, Powell BC, Cheney RE (1 April 2001). "A millennial myosin census". *Mol. Biol. Cell* 12 (4): 780–94. doi:10.1091/mbc.12.4.780. PMC 32266. PMID 11294886.
  78. Meighen EA (1 March 1991). "Molecular biology of bacterial bioluminescence". *Microbiol. Rev.* 55 (1): 123–42. PMC 372803. PMID 2030669.

[1] Dr. Prabodh Sarmah, MSc., M.Phil, PhD



He received his PhD degree from the Gauhati University, ASSAM in 2003. He is a Associate professor in the department of Botany in the Nalbari college (under Gauhati University ). CONTACT NO-9435028496

[2] Mr. Devajit Mahanta, MSc (Computer Science) , MCA



He received MSc in Computer Science degree from the Sikkim Manipal University, Gangtok in 2007. He received MCA degree from the Punjab Technical University, Punjab in 2009. He is a Faculty in the department of Computer Science at MSJ COLLEGE OF MANAGEMENT AND TECHNOLOGY ,ASSAM (Authorized study centre of Sikkim Manipal University ,DE )for the past 5yrs.

79. Jauch R, Yeo HC, Kolatkar PR, Clarke ND (2007) Assessment of CASP7 structure predictions for template free targets. *Proteins-Structure Function and Bioinformatics* 69 Suppl 8: 57-67.
80. Zhang Y (2009) Protein structure prediction: when is it useful? *Curr Opin Struct Biol* 19: 145-155.
81. Vitkup D, Melamud E, Moulton J, Sander C (2001) Completeness in structural genomics. *Nat Struct Biol* 8: 559-566.
82. Mariani V, Kiefer F, Schmidt T, Haas J, Schwede T (2011) Assessment of template based protein structure predictions in CASP9. *Proteins* 79 Suppl 10: 37-58.
83. Hildebrand A, Remmert M, Biegert A, Soding J (2009) Fast and accurate automatic structure prediction with HHpred. *Proteins* 77 Suppl 9: 128-132.
84. Zhang Y (2008) I-TASSER server for protein 3D structure prediction. *Bmc Bioinformatics* 9: 40.
85. Xu D, Zhang J, Roy A, Zhang Y (2011) Automated protein structure modeling in CASP9 by I-TASSER pipeline combined with QUARK-based ab initio folding and FG-MD-based structure refinement. *Proteins* 79 Suppl 10: 147-160.
86. Rychlewski L, Fischer D (2005) LiveBench-8: the large-scale, continuous assessment of automated protein structure prediction. *Protein Sci* 14: 240-245.
87. Schwede T, Kopp J, Guex N, Peitsch MC (2003) SWISS-MODEL: An automated protein homology-modeling server. *Nucleic Acids Res* 31: 3381-3385.
88. Marti-Renom MA, Stuart AC, Fiser A, Sanchez R, Melo F, et al. (2000) Comparative protein structure modeling of genes and genomes. *Annu Rev Biophys Biomol Struct* 29: 291-325.
89. Bordoli L, Schwede T (2012) Automated protein structure modeling with SWISS-MODEL Workspace and the Protein Model Portal. *Methods Mol Biol* 857: 107-136.
90. Lopatin AN, Nichols CG (2001) Inward rectifiers in the heart: an update on I(K1). *J Mol Cell Cardiol* 33: 625-638.
91. Ureche ON, Baltaev R, Ureche L, Strutz-Seebohm N, Lang F, et al. (2008) Novel insights into the structural basis of pH-sensitivity in inward rectifier K<sup>+</sup> channels Kir2.3. *Cell Physiol Biochem* 21: 347-356.
92. Choi SH, Kim HS, Lee EY (2009) Comparative homology modeling-inspired protein engineering for improvement of catalytic activity of Mugil cephalus epoxide hydrolase. *Biotechnol Lett* 31: 1617-1624.
93. Qi X, Guo Q, Wei Y, Xu H, Huang R (2012) Enhancement of pH stability and activity of glycerol dehydratase from *Klebsiella pneumoniae* by rational design. *Biotechnol Lett* 34: 339-346.
94. Dehouck Y, Kwasiroch JM, Gilis D, Rooman M (2011) PoPMuSiC 2.1: a web server for the estimation of protein stability changes upon mutation and sequence optimality. *Bmc Bioinformatics* 12: 151.
95. Khan S, Vihinen M (2010) Performance of protein stability predictors. *Hum Mutat* 31: 675-684.
96. Schymkowitz J, Borg J, Stricher F, Nys R, Rousseau F, et al. (2005) The FoldX web server: an online force field. *Nucleic Acids Res* 33: W382-388.
97. Gilis D, Rooman M (2000) PoPMuSiC, an algorithm for predicting protein mutant stability changes: application to prion proteins. *Protein Eng* 13: 849-856.
98. Dehouck Y, Grosfils A, Folch B, Gilis D, Bogaerts P, et al. (2009) Fast and accurate predictions of protein stability changes upon mutations using statistical potentials and neural networks: PoPMuSiC-2.0. *Bioinformatics* 25: 2537-2543.
99. Bava KA, Gromiha MM, Uedaira H, Kitajima K, Sarai A (2004) ProTherm, version 4.0: thermodynamic database for proteins and mutants. *Nucleic Acids Res* 32: D120-121.
100. Smith BA, Hecht MH (2011) Novel proteins: from fold to function. *Curr Opin Chem Biol* 15: 421-426.
101. Kaplan J, DeGrado WF (2004) De novo design of catalytic proteins. *Proc Natl Acad Sci U S A* 101: 11566-11570.
102. Reina J, Lacroix E, Hobson SD, Fernandez-Ballester G, Rybin V, et al. (2002) Computer-aided design of a PDZ domain to recognize new target sequences. *Nat Struct Biol* 9: 621-627.
103. Looger LL, Dwyer MA, Smith JJ, Hellinga HW (2003) Computational design of receptor and sensor proteins with novel functions. *Nature* 423: 185-190.
104. Garcia-Viloca M, Gao J, Karplus M, Truhlar DG (2004) How enzymes work: analysis by modern rate theory and computer simulations. *Science* 303: 186-195.
105. Tantillo DJ, Chen J, Houk KN (1998) Theozymes and compuzymes: theoretical models for biological catalysis. *Curr Opin Chem Biol* 2: 743-750.
106. Richter F, Leaver-Fay A, Khare SD, Bjelic S, Baker D (2011) De novo enzyme design using Rosetta3. *PLoS One* 6: e19230.
107. Dechancie J, Clemente FR, Smith AJ, Gunaydin H, Zhao YL, et al. (2007) How similar are enzyme active site geometries derived from quantum mechanical theozymes to crystal structures of enzyme-inhibitor complexes? Implications for enzyme design. *Protein Sci* 16: 1851-1866.
108. Zanghellini A, Jiang L, Wollacott AM, Cheng G, Meiler J, et al. (2006) New algorithms and an in silico benchmark for computational enzyme design. *Protein Sci* 15: 2785-2794.
109. Kuhlman B, Baker D (2000) Native protein sequences are close to optimal for their structures. *Proc Natl Acad Sci U S A* 97: 10383-10388.
110. Kiss G, Rothlisberger D, Baker D, Houk KN (2010) Evaluation and ranking of enzyme designs. *Protein Sci* 19: 1760-1773.
111. Altamirano MM, Blackburn JM, Aguayo C, Fersht AR (2000) Directed evolution of new catalytic activity using the alpha/beta-barrel scaffold. *Nature* 403: 617-622.
112. Alexandrova AN, Rothlisberger D, Baker D, Jorgensen WL (2008) Catalytic mechanism and performance of computationally designed enzymes for Kemp elimination. *J Am Chem Soc* 130: 15907-15915.
113. Khersonsky O, Rothlisberger D, Dym O, Albeck S, Jackson CJ, et al. (2010) Evolutionary optimization of computationally designed enzymes: Kemp eliminases of the KE07 series. *J Mol Biol* 396: 1025-1042.
114. Khersonsky O, Rothlisberger D, Wollacott AM, Murphy P, Dym O, et al. (2011) Optimization of the in-silico-designed kemp eliminase KE70 by computational design and directed evolution. *J Mol Biol* 407: 391-412.
115. Bolon DN, Mayo SL (2001) Enzyme-like proteins by computational design. *Proc Natl Acad Sci U S A* 98: 14274-14279.
116. Nat Chem Biol/PLoS Med/Faiella M, Andreozzi C, de Rosaes RT, Pavone V, Maglio O, et al. (2009) An artificial di-iron oxo-protein with phenol oxidase activity. *Nat Chem Biol* 5: 882-884.
117. Yeung N, Lin YW, Gao YG, Zhao X, Russell BS, et al. (2009) Rational design of a structural and functional nitric oxide reductase. *Nature* 462: 1079-1082.
118. Dahiyar BI, Mayo SL (1997) De novo protein design: Fully automated sequence selection. *Science* 278: 82-87.
119. Bornscheuer UT, Huisman GW, Kazlauskas RJ, Lutz S, Moore JC, et al. (2012) Engineering the third wave of biocatalysis. *Nature* 485: 185-194.
120. Ioannidis JP (2005) Why most published research findings are false. *PLoS Med* 2: e124.
121. Lippow SM, Tidor B (2007) Progress in computational protein design. *Curr Opin Biotechnol* 18: 305-311.