Abstract
A key challenge of the post-genomic era is the identification of the function(s) of all the molecules in a given organism. Here, we review the status of sequence and structure-based approaches to protein function inference and ligand screening that can provide functional insights for a significant fraction of the ~50% of ORFs of unassigned function in an average proteome. We then describe FINDSITE, a recently developed algorithm for ligand binding site prediction, ligand screening and molecular function prediction, which is based on binding site conservation across evolutionary distant proteins identified by threading. Importantly, FINDSITE gives comparable results when high-resolution experimental structures as well as predicted protein models are used.

Keywords: protein function prediction; ligand binding site prediction; virtual ligand screening; protein structure prediction; low-resolution protein structures

INTRODUCTION
Over the past decade, catalyzed by the sequencing of the genomes of hundreds of organisms [1–4], biology is undergoing a revolution comparable to what physics underwent in the early 20th century. The emphasis is shifting from the study of individual molecules to the large-scale examination of all genes and gene products in an organism and comparative genomics studies of multiple organisms [5–9]. Here, the goal is to understand the function of all molecules in a cell and how they interact on a system-wide level; this perspective has given birth to the new field of Systems Biology [10, 11]. Of course, biological function is multifaceted, ranging from biochemical to cellular to phenotypical [12, 13]. By detecting evolutionary relationships between proteins of known and unknown function, sequence-based methods can provide insights into the function of about 50% of the ORFs in a given proteome [14–20], with the remainder believed to be too evolutionarily distant to infer their function [21]. Thus, the prediction of the function of the remaining 50% of unannotated ORFs remains an outstanding challenge. However, since protein structure is more conserved than protein sequence [22–24], it can play an essential role in annotating genomes [13, 25–31]. In addition, protein structure should assist in lead compound identification as part of the drug discovery process [11, 32–34]. A key question is whether one can use low-to-moderate resolution predicted structures which can be provided for about 70% of the protein domains in a proteome [35] or if high-resolution experimental structures are required [36, 37]. This issue also has implications for the requisite scope of structural genomics that aims for high-throughput protein structure determination [38–44]. If low-to-moderate resolution models were to prove useful for functional
inference, then the value of contemporary protein structure prediction approaches would be significantly enhanced [45].

One of the more disappointing aspects of protein structure-based functional inference has been the relatively minor marginal impact it has had to date relative to sequence-based methods that rely on inferring function on the basis of the evolutionary relationship between proteins of known and unknown function. Here, however, caution needs to be exercised; just because a pair of proteins are evolutionarily related does not imply that they have the identical function [20]. Proteins can add additional functions during the course of their evolution or can modify their function from that of their ancestors [46–49]. On the other hand, especially for binding site prediction and ligand screening, as proteins become evolutionary more distant, it is unclear what features are conserved and what have become modified [50]. Here, one might imagine that conservation of the protein’s binding site in the structure and conservation of ligand binding features and associated ligands could prove useful. Indeed, just as protein homology modeling as extended by threading [51] (note that the most successful threading approaches have a strong evolution based component [52]) has proven to be a very powerful tool in protein function prediction, one would like to exploit these ideas for the prediction of protein function, binding site prediction and ligand screening [37]. That is to say, we wish to exploit the signal averaging provided by evolution to identify the conserved/variable functional features that can be used to infer the functional properties of proteins of unknown function and to do so by automatic approaches suitable for proteomes.

Based on the above, there is a pressing need for the development of more powerful approaches to protein function prediction that can be applied on a proteome scale. In that regard, in this Briefings in Bioinformatics article, we first summarize the status of sequence-based approaches to functional inference [14, 19, 53–55] that provide the baseline against which protein structure-based approaches are compared. Next, there is the issue of the utility of protein structure for functional inference. Is it marginal? Being bounded at higher levels of sequence identity (>20%) by purely sequence-based approaches [14, 19] and at low levels of sequence identity by the inability to transfer function by interference [20] (which is the only effective means of function prediction), can one effectively exploit the insights provided by protein structure? If so, what is the quality of protein structure required for functional inference in general, and for binding site and ligand screening in particular? Can predicted, low-resolution protein models be used or are we limited to high-resolution, experimental structures [36, 56, 57]? Here, we focus in particular on a newly developed, powerful threading-based approach to protein function prediction, FINDSITE [37], that shows considerable promise in its ability to exploit both experimental and predicted protein structures for the inference of protein function, the prediction of protein binding sites as well as for providing guidance in small molecule ligand screening. These issues are discussed below.

**SEQUENCE-BASED FUNCTIONAL INFERENCE**

The biological function of a protein can be defined in physiological, developmental, cellular or biochemical contexts [58]. To characterize these facets of protein functions, a number of ontologies have been developed, including those in GO [59], KEGG [60] and MIPs [61]. However, even having an appropriate description of protein function, performing experimental assays on all the uncharacterized proteins provided by the hundreds of ongoing genome sequencing projects is impractical. Thus, computational tools are needed [62]. In fact, in newly sequenced genomes, the functional annotations of the vast majority of genes are not based on experiment but are inferred on the basis of the sequence similarity to previously characterized proteins [58, 63, 64].

The fundamental assumption of this strategy, termed ‘annotation transfer by homology’ [65], is that sequence similarity is equivalent to functional similarity. However, sequence similarity based function transfer is complicated by numerous factors; most critical is the functional divergence of highly similar sequences, a problem exhibited by many protein families [54, 55]. Here, permissive criteria to assess the significance of the similarity between proteins can lead to wrong annotations. For example, depending on the protein family, detailed biochemical function is not completely conserved between similar proteins even when their pairwise sequence identity is 60% [20]. Despite this fact,
much lower sequence similarity thresholds have been used in the functional annotation of some genomes [66]. This issue can be partly addressed by introducing family specific sequence identity thresholds [20], and especially at lower pairwise sequence identity levels (20–30%), enhanced specificity and coverage can be achieved by exploiting the conservation of functionally determining residues [14, 19, 67, 68]. Indeed, the sequence-based method EFICAz for enzyme function inference [14, 19] shows quite high levels of precision, sensitivity and specificity even at the levels of 20% sequence identity between pairs of enzyme sequences. It works because a combination of criteria designed to give a low false positive rate is used. Here, the use of functionally discriminating residues that act as a filter once a sequence is assigned as being evolutionary related to sequences of known enzymatic function is of importance.

STRUCTURE-BASED FUNCTIONAL INFERENCE

Active and binding site prediction

Within a protein family, the global fold is more strongly conserved than protein sequence [69]. Thus, the inference of protein biochemical function should benefit by the inclusion of structural information [13]. However, divergent and convergent evolution results in a non-unique relationship between protein structure and protein function; i.e. the structure of a protein in and of itself is insufficient for correct function prediction [70, 71]. As in highly accurate sequence-based approaches [19], additional information is required. Three-dimensional descriptors or templates of biologically relevant sites [26, 72–81] are one example of such a filter. As demonstrated for 4 enzyme systems [82], local 3D motifs frequently outperform global similarity searches using protein structure [83] or sequence [84] alone. Furthermore, the Evolutionary Trace (ET) approach shows that the accuracy of 3D templates can be further increased by selecting evolutionarily relevant residues [85, 86]. In addition to these active site descriptors designed to capture the geometric features of known catalytic residues, a number of structure-based approaches have been developed to identify ligand binding sites [87]. Many focus on the recognition of particular ligand, e.g. adenylate [81], calcium [88] or DNA [89, 90], with more general methods mainly tested on a few ligand types [75, 91]. Of interest is the PINTS [30] approach designed to perform database searches against a collection of ligand-binding sites excised from the PDB [92] and the ProFunc server that combines a collection of sequence- and structure-based methods to identify close relationships to functionally characterized proteins [93].

Geometric methods locate putative binding residues by searching for cavities/pockets in the protein’s structure [94–97]. Comprehensive benchmarks carried out for the unbound/bound protein crystal structures reveal that among the best of these pocket-detection algorithms is LIGSITE$^{\text{CSC}}$ [96], an extension of LIGSITE [95]. LIGSITE$^{\text{CSC}}$ calculates surface-accessibility on the Connolly surface [98] and then re-ranks the identified pockets by the degree of conservation of identified surface residues. Other methods calculate theoretical microscopic titration curves [99], analyze the spatial hydrophobicity distribution [100] or identify electrostatically destabilized residues [101]. In all these methods, the ligand itself is ignored; rather the focus is on the structural features of the protein surface.

Ligand docking algorithms

Given a protein structure, one should not only be able to identify the functional site, but also be able to predict which ligands (for enzymes, substrates) bind to that site. There are two key elements of any docking approach: First, a scoring function is required that accurately ranks the generated set of solutions. In that regard, blind docking can be used to elucidate some general features of binding ligands (or more practically, drug candidates) [102, 103], even if one lacks the ability to correctly rank known binding ligands. Second, a fast and effective search algorithm is necessary to explore the conformational space of protein–ligand interactions. Efficiency is especially important in virtual screening experiments [104, 105], where millions of possible ligands need to be docked into a receptor structure in an acceptable amount of time. Thus, as a practical matter, for each ligand, the docking cannot require more than a few minutes of CPU time on a state-of-the-art computer.

The past years have seen the development of a number of algorithms for docking small molecules into receptor proteins [106–109]. These approaches have been evaluated in terms of ligand binding pose accuracy and the ability to predict binding affinities [110–113]. However, it is evident that most contemporary approaches have significant problems
with ligand ranking, and most require high-resolution, experimentally determined protein structures [36, 111]. Thus, while considerable progress has been made, significant issues remain.

Utility of predicted structures for functional inference

A number of protein structure-based function inference methods have been reasonably successful when applied to high-resolution structures [26, 72–81, 89, 90]. Given the recent improvements in protein structure prediction algorithms [45, 114–123], it is important to establish if lower resolution predicted structures are useful. A structure-based method for protein function prediction that does not require high-resolution structures would be of significant practical value, especially since the best structure prediction approaches can produce low-resolution or better models for ~2/3 of the proteins in a given proteome [35, 124–126].

The key issue is to establish the quality of structure required to transfer a given biochemical function at a specified level of accuracy. While there have been attempts to address this issue for enzymes using active site template matching [13, 127], further investigation is required. Most often, ligand docking programs typically utilize high-resolution receptor structures determined by experiment or theoretical modeling [128–130]. Virtual screening reveals that the success rate decreases from ligand-bound to ligand-free to modeled structures [131] and is correlated with the degree of protein movement in the binding site; protein binding site rearrangements greater than 1.5 Å lead to almost complete lack of recovery of the ‘true’ binding mode [132]. Furthermore, decoy-docking experiments using deformed trypsin structures with a Cα root-mean-square deviation, RMSD from the native structure in the range of 1–3 Å for the docking of 47 ligands experimentally known to bind trypsin reveal that specific ligand–receptor contacts are rapidly lost with increasing receptor structure deformation [111].

Different docking techniques have been developed to address this problem. Most account for receptor flexibility/distortion by docking ligands against a precalculated ensemble of receptor conformations [133] or by softening the criterion for the steric fit between the ligand and receptor [134]. Other docking techniques capable of dealing with significant structural inaccuracies employ a low-resolution representation of the protein designed to accommodate structural distortions. For example, an ultra low-resolution (~7 Å) protein representation that averages all high-resolution structural details dramatically improves the tolerance to receptor deformation [135, 136]. A similar approach demonstrated that even low-quality receptor structures could be utilized [57].

Another, newly developed, low-resolution docking approach that uses a reduced ligand and protein representation is Q-dock [36]. Self-docking using crystal structures revealed ligand pose prediction accuracy comparable to all-atom docking. All-atom models reconstructed from Q-Dock’s low-resolution models can be further refined by simple all-atom energy minimization. In decoy docking against distorted receptor models with a backbone RMSD from native of ~3 Å, Q-Dock recovers on average 15–20% more specific contacts and 25–35% more binding residues than all-atom methods. Q-Dock also gives encouraging results for ligand screening against predicted protein structures whose average global backbone RMSD is 5 Å (Brylinski & Skolnick, unpublished results). Thus, the possibility of using low-resolution predicted structures for binding pose identification and ligand screening appears quite promising. In this spirit, we next turn to an automated approach that can predict ligand binding sites, binding ligands as well the molecular function of proteins, even when low-resolution protein structures are used.

FINDSITE: A threading based method for ligand binding site prediction/functional annotation

The comprehensive examination of known protein structures grouped according to SCOP [137] reveals the tendency of certain protein folds to bind substrates at a similar location, suggesting that very distantly homologous proteins often have common binding sites [138]. That is, evolution tends to conserve the functionally important region in the protein structure and conserves a subset of ligand binding features as well. For example, as shown in Figure 1, the localization of the binding pocket as well as the local geometry and the binding mode of the ligands are remarkably well conserved in glutathione S-transferase family despite the low sequence identities between family members. Hence, it should be possible to develop an approach for ligand binding site identification that is less sensitive than pocket-detection methods to structural
distortions of the protein, as these distortions are present in the set of evolutionarily distant protein structures.

In this spirit, we developed FINDSITE [37], an algorithm for protein functional annotation that is based on binding site similarity among superimposed groups of template structures identified from threading [51]. Threading is of importance in that it acts as a filter to establish that the set of protein structures are evolutionarily related. A schematic overview of the FINDSITE methodology is shown in Figure 2. For a given target protein, the threading algorithm PROSPECTOR_3 [52] identifies protein structure templates with bound ligands. Then, these holo-templates are superimposed onto the target structure. In Figure 2A, the target protein’s crystal structure is used. In terms of both overall accuracy and pocket ranking ability, FINDSITE performs better than LIGSITECSC. Using the native structure, the success rate (where the centers of mass of the predicted and native binding sites are ≤4 Å) using the best of top five identified binding pockets is 70.9% and 51.3% for FINDSITE and LIGSITECSC, respectively. For those proteins where a binding pocket is correctly identified, the ranking of both methods is comparable; 76.0% and 74.7% of the best pockets are ranked as the top solutions by FINDSITE and LIGSITECSC.

As shown in Figure 2B, where modeled target structures are used, the prediction accuracy of LIGSITECSC falls off considerably, with its success rate decreasing from 51.3% for the target crystal structure to 32.5%, when protein models generated by TASSER are used. For TASSER models, only 61.4% of the best pockets are assigned rank 1.

To assess the general validity of FINDSITE, we employed a representative set of 901 proteins with <35% sequence identity to their templates (with a mean target-template pair-wise sequence identity of 20%) and generated models using TASSER [35, 123, 140, 141]. As demonstrated below, we find that FINDSITE operates satisfactorily in the ‘twilight zone’ of sequence similarity [142], which covers ~2/3 of known protein sequences [143]. No experimental structure of the target protein is required; high accuracy and ability to correctly rank the identified binding sites are sustained when protein models instead of target crystal structures are used for template superimposition. Use of consensus ligands extracted from the binding sites is quite useful in ligand screening. In most cases, FINDSITE accurately assigns a molecular function to the protein model. These features should enhance the utility of low-to-moderate quality protein models in ligand screening and structure-based drug design.

Figure 3 shows ligand binding site prediction results carried out for the 901 benchmark proteins. Here LIGSITECSC identifies possible binding pockets in the target structure (either the crystal structure or predicted model). Using FINDSITE, the set of predicted template models (where the target has a sequence identity <35% to all template structures) is superimposed onto the target structure. In Figure 3A, the target protein’s crystal structure is used. In terms of both overall accuracy and pocket ranking ability, FINDSITE performs better than LIGSITECSC. Using the native structure, the success rate (where the centers of mass of the predicted and native binding sites are ≤4 Å) using the best of top five identified binding pockets is 70.9% and 51.3% for FINDSITE and LIGSITECSC, respectively. For those proteins where a binding pocket is correctly identified, the ranking of both methods is comparable; 76.0% and 74.7% of the best pockets are ranked as the top solutions by FINDSITE and LIGSITECSC.

As shown in Figure 3B, where modeled target structures are used, the prediction accuracy of LIGSITECSC falls off considerably, with its success rate decreasing from 51.3% for the target crystal structure to 32.5%, when protein models generated by TASSER are used. For TASSER models, only 61.4% of the best pockets are assigned rank 1.
Thus, when models are used, LIGSITE\textsuperscript{CSC} results deteriorate. In contrast, with FINDSITE, both the high accuracy of ligand binding site prediction and correct binding site ranking are sustained when models instead of native structures are used as reference structures for holo-template superimposition. In 67.3\% of the cases FINDSITE identifies a correct binding site, with corresponding ranking accuracy of 75.5\%. Note that for both native structures and predicted models, the results using random patches are much worse than for LIGSITE\textsuperscript{CSC} and FINDSITE.

We find that for models with a global RMSD from the native structure $\leq$ 6\,\AA, FINDSITE typically predicts the center of mass of the binding site within 6\,\AA. This is because the binding sites in the models have an RMSD below 3\,\AA. In contrast, as is evident from Figure 3, LIGSITE\textsuperscript{CSC} is far more sensitive to structural distortions. The average distance between the LIGSITE\textsuperscript{CSC} predicted and observed binding pockets is 10–13\,\AA when the global RMSD of the predicted model exceeds 4\,\AA from the native structure.

FINDSITE’s overall binding site prediction accuracy depends on the number of identified ligand-bound templates with a common binding site. We can classify proteins as Easy ($>$125 threading templates, including homologous proteins for each

\textbf{Figure 2:} Overview of the FINDSITE approach.
in the 901 protein benchmark set, 9%, 47% and 44% of the proteins are assigned by FINDSITE as Easy, Medium and Hard targets, and the average distance between the centers of predicted and observed binding pockets for top-ranked FINDSITE solutions is ~2, 5 and 10 Å, respectively. Using a cutoff distance of 4 Å between predicted and observed binding sites, the hit rates for the top-ranked predictions are 90.0%, 71.7% and 43.7% of Easy, Medium and Hard targets, respectively.

Figure 3: Ligand binding site prediction by FINDSITE, LIGSITE\textsuperscript{CSC} and randomly selected patches on the target protein’s surface using (A) target crystal structures and (B) TASSER models. Results are presented as the cumulative fraction of proteins for which the distance between the ligand center of mass in the native complex and the center of the best of top five predicted binding sites is ≤ the distance on the x-axis with the rank of the best pocket of top five predictions in the inset.
We next explored the structural diversity of the binding site residues. We calculated the average local pairwise RMSD of binding site residues for the subset of 561 target proteins that satisfy the following criteria: The best predicted pocket must be rank 1, with the number templates identified ≥5 and there must at least 10 binding site residues. With these restrictions, we find that the average pairwise RMSD of binding site residues is 2.15 ± 0.77 Å. This gives an estimate of the allowed structural degeneracy of binding residues.

Ligand virtual screening
FINDSITE also extracts information about the chemical properties of the ligands bound to the consensus-binding site; we term these 'template ligands'. Since the 'template ligands' are extracted from the holo templates identified by threading, only the target protein's sequence is needed for their selection. These molecules are then used to construct fingerprints that are subsequently employed in fingerprint-based similarity searching [144, 145] of the KEGG compound library, which contains 12,478 compounds [146]. For the 901 representative target proteins, all with <35% sequence identity to their closest template, Figure 4A presents the cumulative distribution of enrichment factors for the top 1% of the screening library. For accurately predicted binding sites (70.9% of the target proteins have a binding site center of mass is <4 Å of the native structure), FINDSITE performs better than random in 78% of the cases. The ideal enrichment factor (all native-like compounds in the top 1% of the ranked library) was observed for 50% of target proteins. For less accurately predicted binding pockets, the ideal enrichment factor was obtained for 12% and is better than random for 34% of the cases. Finally, in Figure 4B, a case study examined the performance of FINDSITE in virtual screening for 895 active HIV-1 protease inhibitors in a 123,331 compound library. Again, if only templates with <35% sequence identity to the target are used, the enrichment factor of the top 1% of compounds is 40.

Molecular function prediction
The relatively high accuracy of the ligand selection procedure encouraged us to investigate the transferability of specific functions from the threading templates to the target. Here, the Gene Ontology (GO) [59] description of protein molecular function is used. We selected the subset of 753 proteins from the 901 protein benchmark set for which a GO annotation is provided by Gene Ontology [59] or UniProt [125]. For each target, all GO annotations are identified for the threading templates that share the top-ranked predicted binding site. Then, the target protein is assigned a function with a probability that corresponds to the fraction of threading templates annotated with that molecular
function. When at least one half of the threading holo-templates are annotated with the same GO term, the maximal Matthew’s correlation coefficient of 0.64 is found. This corresponds to a precision of 0.76, and a recall of 0.54. In addition, we calculated predictive metrics with respect to individual GO identifiers. When the closest template has <35% sequence identity, FINDSITE distinguishes between enzymatic and non-enzymatic function, with a precision and sensitivity of 0.93 and 0.89, respectively. Moreover, many molecular functions that cover a broad spectrum of molecular events including both enzymatic and binding activities are accurately transferable from the templates selected by FINDSITE.

By way illustration, in Table 1, for the 753 protein benchmark set, the FINDSITE precision and sensitivity is presented for 10 most accurate predicted molecular functions as described by Gene Ontology classification and as assessed by the Matthew’s correlation coefficient. Clearly, a broad spectrum of both enzymatic and non-enzymatic activities are adequately described. However, we note that since FINDSITE describes the common functional features found across an evolutionary distant but related set of proteins, it cannot describe highly specific functions such as all four EC digits of an enzyme.

### CONCLUSION

The most frequently used methods for protein function prediction are based on functional inference by homology [147, 148]. However, as demonstrated for enzymes [20], because of the promiscuity of protein function, care must be taken if the goal is high accuracy (a necessary condition if one wants to ascertain whether or not a specific pathway [60] is present in the proteome of interest). Moreover, current sequence-based methods become unreliable as the sequence identity between the target protein of unknown function and the template protein of known function drops below 20–30% [31]. To address this limitation, a number of structure-based approaches based on 3D geometric descriptors of enzymatic function, termed fuzzy functional forms (FFFs) were developed [26] and shown to provide high-confidence novel annotations [149]. However, they have only been successfully applied to enzymes, and typically require extensive manual intervention in their construction. In practice, their level of accuracy drops when they are applied to predicted protein models [127].

To remove these limitations, in the development of FINDSITE [37], we explored whether the conservation of binding sites among threading identified templates can be used to predict the target binding site, the ligands that bind to this site and using consensus GO molecular functions [59] of the templates, to predict the molecular function of evolutionary distant target proteins. We find that threading followed by binding site filtering to identify functionally related proteins is a very powerful approach to predict these aspects of protein function. This holds even if the sequence similarity to the target protein is well below 35% and has profound implications as to how protein molecular function has evolved. As was observed for enzymes, some functional sites in the protein structure are strongly conserved throughout evolution [150]. Not only is the protein structure conserved, but the chemical features of the ligands that bind to the protein are conserved as well. Such conservation provides a type
of signal averaging that can be exploited for various applications of functional inference.

The clear advantage of FINDSITE is that predicted structures can be used. This is of importance in that state-of-the-art approaches provide predicted structures of the requisite quality for greater than 2/3 of protein domains in a given proteome [35, 143]. This work also suggests that there is a robustness to the structure and chemistry of binding sites and their associated binding ligands that needs to be more effectively exploited for both general functional inference as well as ligand screening. The fact that ‘template ligands’ from distantly related template structures conserve aspects of binding even as the binding sites become somewhat distorted (with an average local RMSD of 2.15 ± 0.77 Å), suggests that many ligand docking algorithms that require a highly accurate experimental structure [106, 151–153] are missing the essential features of binding. Nature itself tolerates binding site modifications in the range of ~1.5–3 Å while retaining the ability to bind related ligands with strongly conserved substructures. The utility of a lower resolution description [135, 136] for docking as in Q-Dock [36] is not only of practical utility but also recapitulates aspects of the features of the ligand–receptor complex that are exhibited across evolutionary distant proteins. It is quite likely that there are other functional properties that can be detected by extensions of the FINDSITE approach. The key idea is to find a set of distantly related structures, identify common functional features and then transfer these features to the protein of interest. Thus, this is a promising avenue of investigation that holds considerable promise in extending the range and scope of structure-based approaches to protein function prediction.

### Key Points
- The structural diversity of the binding site suggests binding site structural degeneracy that can be exploited in low-resolution modeling.
- For a distantly related family of proteins, evolution provides signal averaging that can be employed to infer the structural and chemical features of binding sites that are strongly conserved throughout evolution.
- Low-resolution predicted structures can be used for ligand docking, ranking and functional inference.
- Combined evolution/structure-based approaches provide complementary information that can be exploited for quite high accuracy proteome scale functional inference.
- The FINDSITE algorithm combines these ideas into a robust evolution/structure-based approach to binding site detection, ligand virtual screening and functional inference.

### FUNDING
National Institutes of Health [grant numbers GM-48835, GM-37408 to J.S.].

### References


83. Shindyalov IN, Bourne PE. Protein structure alignment by incremental combinatorial extension (CE) of the optimal path. Protein Eng 1998;11:739–47.


