# 2

# Exploring Structure–Function Relationships in Moonlighting Proteins

*Sayoni Das[1], Ishita Khan[2], Daisuke Kihara[3], and Christine Orengo[1]*

[1] *Institute of Structural and Molecular Biology, University College London, Gower Street, London, UK*
[2] *Department of Computer Science, Purdue University, North University Street, West Lafayette, Indiana, USA*
[3] *Department of Biological Sciences, Purdue University, Martin Jischke Drive, West Lafayette, Indiana, USA*

## 2.1 Introduction

As the availability of large-scale genomic data and the technical advancements in high-throughput biological experiments reach an astonishingly high level, the functional characterization of proteins becomes more sophisticated than ever. Consequently, an increasing number of proteins have been found to moonlight, that is, perform multiple independent cellular functions within a single polypeptide chain (see Chapter 1 for more details of the definition of protein moonlighting).

The functional diversity of moonlighting proteins is not a consequence of gene fusions, splice variance, proteins performing different functions in different cellular contexts, varying post-transcriptional modifications, homologous but non-identical proteins, or multiple photolytic fragments. The multiple roles of moonlighting proteins are not restricted to certain organisms or protein families, nor do they have a common mechanism through which they switch between different functions. Experimentally identified moonlighting proteins have been shown to switch functions as a consequence of changes in cellular locations within and outside the cell, expression in different cell types, oligomerization states, ligand binding locations, binding partners, and complex formation [1–3].

A large number of moonlighting proteins have been found to be involved in bacterial virulence, DNA synthesis or repair, cancer cell motility, and angiogenesis, among others. As an example, neuropilin is a moonlighting protein that is known to show diverse functions due to changes in cellular contexts. In endothelial cells, it is a vascular endothelial cell growth factor (VEGF) receptor and a major regulator of angiogenesis, vasculogenesis, and vascular permeability. However, in nerve axons, it is a receptor for a different ligand (Semaphorin III) and mediates neuronal cell guidance.

*Moonlighting Proteins: Novel Virulence Factors in Bacterial Infections*, First Edition.
Edited by Brian Henderson.
© 2017 John Wiley & Sons, Inc. Published 2017 by John Wiley & Sons, Inc.

More than 300 moonlighting proteins are known in the literature today (see Chapter 1); however, the rapid increase in the number of identified moonlighting proteins suggests that the phenomenon may be common in all kingdoms of life. So far, the moonlighting function(s) of the known proteins have mostly been discovered by serendipity and little is known about the molecular mechanisms of such moonlighting actions [2]. Consequently, any efforts to characterize the molecular mechanisms of such proteins and understand their structure–function relationship would aid in identifying new moonlighting functions and help to better understand of the complex functional interplay of moonlighting proteins in the cell.

In this chapter, we first briefly discuss the different contexts in which protein function can be described, the complex structure–function relationship in proteins, followed by the current approaches used in identifying and characterizing moonlighting proteins. We then propose a classification of moonlighting proteins based on the structure–function analysis of selected moonlighting proteins. A few examples of moonlighting proteins in each classification are described in detail, many of which are implicated in bacterial virulence. Finally, we describe some general trends observed in the analysis which will, we hope, be valuable in understanding how a moonlighting protein can perform more than one unrelated function.

## 2.2 Multiple Facets of Protein Function

The phrase "protein function" is very ambiguous, as the functional role of a protein can be described in many different contexts. It can be described in terms of: (1) the molecular function of the protein; (2) its role in biological pathway(s); or (3) its cellular location. Natural language annotations in databases and the literature are too vague and unspecific to accurately describe the function(s) of a protein. This has led to the development of a common organized protein annotation vocabulary such as the Enzyme Commission (EC) number and Gene Ontology (GO) [4], which are the most commonly used protein function annotation resources.

The Enzyme Commission (EC) number [5] system is a numerical classification system for enzymes that uses a hierarchical set of four numbers separated by periods to represent the catalytic reaction that it carries out. For example, the EC number 5.3.1.9 describes an isomerase (EC 5.-.-.-) that acts as an intramolecular oxidoreductase (EC 5.3.-.-) and interconverts aldoses and ketoses (EC 5.3.1.-) using glucose-6-phosphate as the substrate (EC 5.3.1.9).

The Gene Ontology (GO) [4] is the most comprehensive and widely used resource of protein annotations. GO annotation can be used to assign functional terms to both enzymes and non-enzymes from three structured, non-overlapping ontologies in a species-independent manner: (1) molecular function ontology (MFO) describes the biochemical activity of the protein at the molecular level; (2) biological process ontology (BPO) describes the cellular processes and pathways in which the protein is involved; and (3) cellular component ontology (CCO) describes the compartment(s) of the cell in which the protein performs its action.

| Enzyme commision (EC) number | |
|---|---|
| **EC 5.-.-.-** | Isomerase |
| **EC 5.3.-.-** | Intramolecular oxidoreductase |
| **EC 5.3.1.-** | Interconverting aldoses & ketoses & related compounds |
| **EC 5.3.1.9** | Glucose-6-phosphate isomerase |

**Gene ontology (GO) terms**

**Molecular function**

G0:0004347 G-6-P isomerase activity
G0:0005125 Cytokine activity
GO: 0008083 Growth factor activity

**Biological process**

G0:0061620 Glycolytic process through G-6-P
G0:0006094 Gluconeogenesis
G0:0001525 Angiogenesis
G0:0001707 Mesoderm formation
G0:0034101 Erythrocyte homeostasis

**Cellular component**

G0:0005737 Cytoplasm
G0:0005886 Plasma membrane
GO: 0005654 Nucleoplasm
GO: 0005576 Extracellular region
GO: 0070062 Extracellular vesicular exosome
G0:0043209 Myelin sheath
G0:0060170 Ciliary membrane

**Figure 2.1** Function annotations for the mouse protein, glucose-6-phosphate isomerase (Uniprot Accession no. P06745) from the Enzyme Commision (EC) number system and Gene Ontology.

The sources of these annotations can be literature references, experimental results, author statements, database references, or computational outputs.

Figure 2.1 shows the functional annotations for the enzyme Phosphoglucose isomerase (PGI) in the mouse, which can moonlight as a tumor-secreted cytokine and angiogenesis factor and also as a neurotrophic factor. The EC number can only describe the catalytic function of PGI; however, the GO annotations from the three ontologies are sufficient to completely describe the functions of the moonlighting protein (see Figure 2.1).

## 2.3 The Protein Structure–Function Paradigm

Knowledge of the three-dimensional structure of a protein plays an important role in understanding the molecular mechanisms underlying its function. The three-dimensional structures of proteins can often provide more functional insight than simply knowing the protein's sequence (Fig. 2.2). For example, the structure reveals the overall conformation of the protein along with the biological multimeric state of the protein. It reveals the binding sites, interaction surfaces and the spatial relationships of catalytic residues. Protein–ligand complexes provide details of the nature of the ligand and its precise binding site, which helps in postulating the catalytic mechanism. The PDBsum resource [6] provides pictorial analyses for every structure in the PDB along with detailed information extracted from various resources such as SwissProt, Catalytic Site Atlas (CSA), Pfam, and CATH, which are beneficial for structure–function studies.

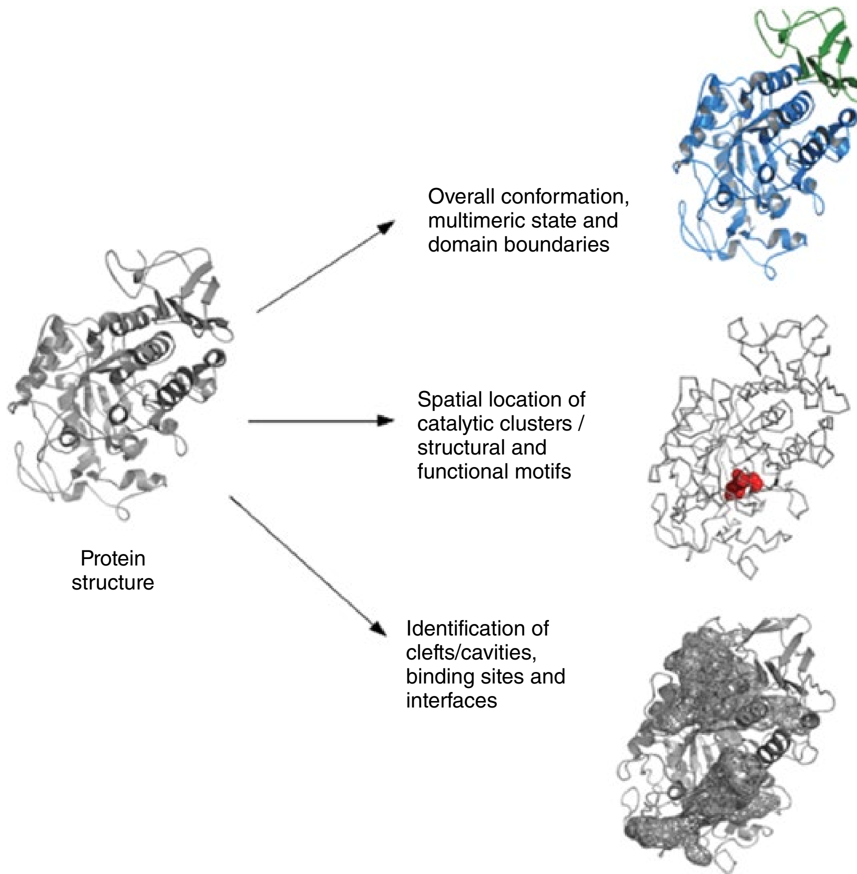


**Figure 2.2** From protein structure to function.

Furthermore, *ab initio* prediction of binding pockets and clefts on the protein structure, using methods such as pvSOAR, CASTp, SURFACE, SiteEngine, and THEMATICScan, also provides useful information about protein function.

Proteins are composed of one or more building blocks called domains which are distinct, compact units of protein structure. These domains often combine in a mosaic manner in multidomain proteins (domain shuffling), generating new or modified functions [7]. Structural similarity between homologous proteins is more highly conserved during evolution than sequence and is therefore helpful in recognizing even distantly related proteins. Domains are considered to have the same fold if they share the same orientation and connectivity of the secondary structures. Specific domains within a protein are often found to have distinct functional roles, but sometimes more than one domain may be involved in a particular function, for example where an enzyme's active site is formed at the interface between two domains.

Evolutionarily related proteins having high fold similarity often share functional similarity. As a result, protein functions can sometimes be inferred by comparing the structure of the query protein with that of an experimentally

characterized protein. Structural relationships can be captured by using various well-established algorithms, for example DALI [8], SSAP [9], STRUCTAL [10], CE [11], MAMMOTH [12], FATCAT [13] and CATHEDRAL [14].

Protein structure classification databases such as CATH and SCOP have enabled detailed analysis of structure–function relationships between evolutionarily related proteins. CATH and SCOP extract structural information from the PDB and classify domains into different classes, folds, and homologous superfamilies in a hierarchical manner based on their structural relationships and evolutionary origin. Studies based on these resources have shown that the structure–function relationship of proteins is very complex and fold similarity may not always be sufficient to conclude functional similarity [15]. For example, some folds such as the Rossmann fold and TIM barrels can carry out a large number of different functions, and many different folds can be associated with the same function.

At the time of writing, the CATH database comprises around 1400 protein folds and approximately 3000 homologous domain superfamilies. These have been found to comprise at least 70% of protein domain sequences [16], which suggests that a relatively limited number of folds carry out the huge diversity of functions observed in protein function space [17–19]. For the majority of the domain superfamilies in CATH (>90%), domains have highly similar structures and functions. However, these conserved superfamilies tend to be small and highly specific to certain species or subkingdoms of life. Most of the remaining superfamilies can incorporate large amounts of structural and functional diversity and are highly populated, accounting for >50% of all known domains. Structural and functional diversity between domains in enzyme superfamilies can be attributed to the use of different sets of residues in their active site, the addition of secondary structure embellishments to the core domain structure, or domain recruitment [20].

A recent study on the diversity of functional sites in CATH superfamilies by [21] showed that, for most superfamilies, the spatial locations of functional sites are limited. By contrast, members of the most diverse superfamilies show a considerable amount of functional plasticity, as their relatives can exploit different sites for interacting with their protein partners or for binding small-ligands.

By subclassifying these diverse superfamilies into functional families [16], it is possible to group relatives sharing a common functional site and similar functional properties. Structural similarity with a protein in a CATH functional family can therefore be used to infer functional properties more accurately. Functional family classification also provides a means to understand the mechanisms of its functional divergence during evolution.

## 2.4 Computational Approaches for Identifying Moonlighting Proteins

Existing computational approaches to the analysis of moonlighting proteins explore different aspects of proteins characteristics, varying from sequence data, protein–protein interaction (PPI) data, to structural properties. Two recent studies have investigated how well the current sequence-based methods characterize the

functional diversity of moonlighting proteins [22, 23]. Gomez *et al.* [24] have also conducted an analysis on the PPI networks of moonlighting proteins and statistically quantified if the interacting partners of the moonlighting proteins can identify the moonlighting functions. Hernandez *et al.* [25] explored structural aspects of moonlighting proteins and analyzed whether the diverse functions are caused by conformational fluctuations in the disordered regions of their structures. Becker *et al.* [26] analyzed the human PPI network and developed a novel clustering method that can decompose a network into multiple overlapping clusters. They reported that proteins that belong to the overlapping clusters are more central in the network compared to mono-clustered proteins and contain multiple domains; they are therefore candidates for multitasking proteins. There exist two publicly available databases that serve as a repository of moonlighting proteins: MoonProt [27] and MultitaskProtDB [28]. These databases store the primary and secondary functions of known moonlighting proteins, UniProt accessions, species information, and PDB codes (if available), among other information. Computational works of moonlighting proteins were recently summarized in a review article [29].

Khan *et al.* [30] explored the functional diversity of moonlighting proteins in a computational framework and identified multiple proteomics characteristics of these proteins. Initially the existing gene ontology (GO) annotations of experimentally known moonlighting proteins were explored and the diversity of GO terms of a protein in the hierarchical GO structure quantified. Based on this GO similarity/distance, GO terms of moonlighting proteins were then clustered into groups characterized by their multiple functions. This allowed the identification of novel moonlighting proteins in the *Escherichia coli* genome and a subset of these were confirmed by literature analysis. Subsequently, context-driven analysis of moonlighting proteins was performed to explore how moonlighting proteins interacted with other proteins in a physical interaction network, through similar gene expression profiles, phylogenetic profiles, and by means of genetic interactions. This led to two notable observations: moonlighting proteins tend to interact with a more functionally diverse group of proteins than non-moonlighting proteins, and most interacting partners of moonlighting proteins share their primary function. Moreover, a significant number of interacting partners of the moonlighting proteins had indications of being moonlighting proteins themselves. In other -omics-scale analyses (i.e., based on gene expression), phylogenetically related proteins, and genetic interactions, the weak trend that on average moonlighting proteins interact with more functionally diverse proteins was observed. Structural characteristics of moonlighting proteins (i.e., ligand-binding sites and intrinsic disordered regions) were also investigated. In several cases, the ligand-binding sites for distinct functions are located in separate regions of the protein's tertiary structure.

## 2.5 Classification of Moonlighting Proteins

In this section we examine the structural data that are available for moonlighting proteins and propose a classification of moonlighting proteins based on the spatial locations of the experimentally verified functional sites exploited by a protein to

**Table 2.1** Proteins having distinct sites for different functions in the same domain.

| Protein (organism) | Function 1 | Function 2 | Structure | Refs |
|---|---|---|---|---|
| Enolase (*S. pneumoniae*) | Enolase (EC 4.2.1.11) | Binds plasminogen | 1W6T | [31] |
| Albaflavenone monooxygenase (*S. coelicolor A3(2)*) | Albaflavenone monooxygenase (EC 1.14.13.106) | Terpene synthase (EC 4.2.3.47) | 3EL3 | [32, 33] |
| MAPK1/ERK2 (*H. sapiens*) | Mitogen-activated protein kinase 1 (EC 2.7.11.24) | Transcriptional repressor (binds DNA) | 4G6N | [34] |
| 1-Cys Peroxiredoxin (*H. sapiens*) | Phospholipase A2 (EC 3.1.1.4) | Glutathione peroxidase (EC 1.11.1.15) | 1PRX | [35] |
| Cytochrome C (*S. cerevisiae*) | Electron carrier protein in electron transport chain | Promotes apoptosis (binds Apaf-1) | 1YCC | [36] |
| GCN4 (*S. cerevisiae*) | Transcription factor (binds DNA) | Ribonuclease (EC 3.1.27.5) | 1YSA | [37] |
| I-Anii (*A. nidulans*) | Homing endonuclease (EC 3.1.-.-) | Transcriptional repressor (binds DNA) | 3EH8 | [3, 38] |

perform its primary and moonlighting function(s). The primary and moonlighting function(s) of the proteins are referred to as 'Function 1' and 'Function 2' in the following. The moonlighting proteins were taken from the database of moonlighting proteins, MoonProt [27], and recently published papers which had known structural information along with experimentally verified functional sites responsible for the primary and moonlighting function(s) of the protein. Information on catalytic site residues for the proteins was extracted from the Catalytic Site Atlas (CSA), and additional functional annotation was extracted from PDBsum [6] and SwissProt. The various categories are proteins: (1) having distinct sites for different functions in the same domain (Table 2.1 [31–38]); (2) having distinct sites for different functions in different domains (Table 2.2; [39–44]); (3) using the same residues for different functions (Table 2.3; [45–47]); (4) using different residues in the same/overlapping site for different functions (Table 2.4; [48–53]); and (5) using different structural conformations or folds for different functions (Table 2.5; [54–56]).

### 2.5.1 Proteins with Distinct Sites for Different Functions in the Same Domain

These are single domain or multidomain proteins (listed in Table 2.1) that use distinct spatial functional sites of a single domain for carrying out their primary and moonlighting function(s).

#### 2.5.1.1 α-Enolase, *Streptococcus pneumonia*

α-Enolase (EC 4.2.1.11) from *Streptococcus pneumoniae* is a key glycolytic enzyme (Function 1) that is also expressed on the bacterial cell surface, where it binds to human plasminogen to facilitate the host invasion process during infection

**Table 2.2** Proteins having distinct sites for different functions in different domains.

| Protein (organism) | Function 1 | Function 2 | Structure | Refs |
|---|---|---|---|---|
| Malate synthase (*M. tuberculosis*) | Malate synthase (EC 2.3.3.9) | Binds laminin | 2GQ3 | [39] |
| BirA (*E. coli*) | Biotin holoenzyme synthetase (EC 6.3.4.15) | Bio repressor | 1BIB | [40] |
| MRDI (*H. sapiens*) | MTR-1-P isomerase (EC 5.3.1.23) | Mediator of cell invasion | 4LDQ | [41] |
| Hexokinase 2 (*S. cerevisiae*) | Hexokinase 2 (EC 2.7.1.1) | Glucose sensor (interacts with transcriptional repressor Mig1) | 1IG8 | [42] |
| Neuropilin-I (*H. sapiens*) | Semaphorin binding | VEGF binding | 2QQN | [43] |
| ATF2 (*H. sapiens*) | Transcription factor | DNA damage response | 1T2K | [44] |

**Table 2.3** Proteins using the same residues for different functions.

| Protein (organism) | Function 1 | Function 2 | Structure | Refs |
|---|---|---|---|---|
| GAPDH (*E. coli*) | GAPDH (EC 1.2.1.12) | NAD ribosylating activity | 1DC5 | [45] |
| Leukotriene A-4 hydrolase (*H. sapiens*) | Leukotriene A-4 hydrolase (EC 3.3.2.6) | Aminopeptidase (EC 3.4.11.24) | 2R59 | [46] |
| Hemagglutinin (*Paramyxovirus*) | Hemagglutinin binding | Neuraminidase (EC 3.2.1.18) | 1E8T | [47] |

**Table 2.4** Proteins using different residues in the same/overlapping site for different functions.

| Protein (organism) | Function 1 | Function 2 | Structure | Refs |
|---|---|---|---|---|
| Phosphoglucose isomerase (*O. cuniculus*, *H. sapiens*) | Phosphoglucose isomerase (EC 5.3.1.9) | Autocrine motility factor, neuroleukin, differentiation, and maturation mediator | 1DQR, 1IAT | [48, 49] |
| Fructose-bisphosphate aldolase (*P. falciparum*) | Fructose-bisphosphate aldolase (EC 4.1.2.13) | Attaches actin to trap proteins | 2PC4 | [50] |
| Gpx4 (*H. sapiens*) | Phospholipid hydroperoxide glutathione peroxidase (EC 1.11.1.12) | Polymerized form has structural role in spermatozoa | 2OBI | [51] |
| S1O ribosomal/ protein (*E. coli*) | Component of ribosomal 30S subunit | Part of transcription antitermination complex | 1O9J | [52] |
| Lens crystallin/retinal DH (*E. edwardii*) | Lens crystallin | Retinal DH (EC 1.2.1.3) | 1O9J | [53] |

**Table 2.5** Proteins using different folds for different functions.

| Protein (organism) | Function 1 | Function 2 | Structure | Refs |
|---|---|---|---|---|
| RfaH (*E. coli*) | Transcription factor | Translational regulator | 2OUG, 2LCL | [54, 55] |
| Lymphotactin (Ltn) (*H. sapiens*) | Chemokine (activates XCR1) | Binds cell-surface glycosaminoglycans | 2JP1 | [56] |

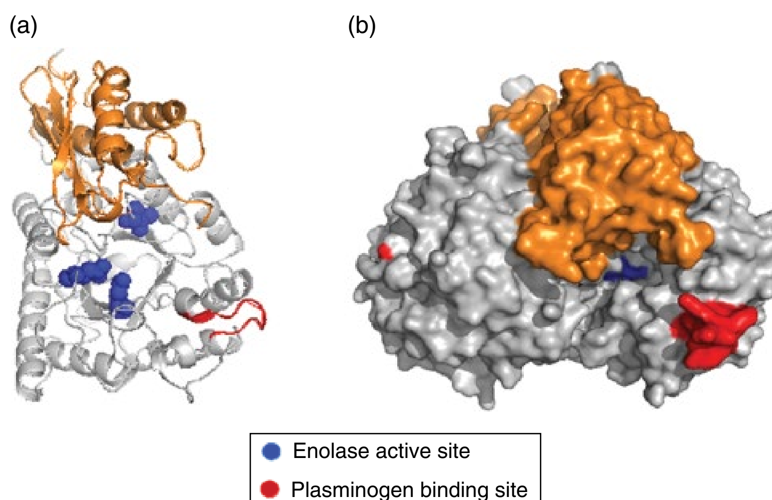


**Figure 2.3** α-Enolase. (a) Single chain of Enolase showing the enzyme active site in blue and the plasminogen-binding site in red. (b) Enolase monomer displayed as surface. Different domains are colored in gray and orange (PDB:1W6T). (*See color plate section for the color representation of this figure.*)

(Function 2) [31]. The protein is known to exist in an octameric state both in the cytoplasm and on the cell surface. Each monomer of α-enolase consists of a 2-layer αβ sandwich domain and a TIM barrel domain (Fig. 2.3). The structurally conserved α-enolase active site is located in the TIM barrel which comprises the catalytic residues Glu164, Glu205, and Lys342 in *S. pneumoniae*, which are located in a pocket. Two plasminogen-binding sites have also been found in the TIM barrel domain at sites distinct from the active site that includes a nine-residue internal motif (248 FYDKERKYV256) and terminal lysine residues (433KK434). The former site has been shown to have a more important role in interacting with plasminogen than the latter. The last lysine residue is not included in the structure as it is disordered.

#### 2.5.1.2 Albaflavenone monooxygenase, *Streptomyces coelicolor A3(2)*

Albaflavenone monooxygenase (CYP170A1) (EC 1.14.13.106) from *Streptomyces coelicolor A3(2)* catalyzes the last two steps in the biosynthesis of the antibiotic albaflavenone (Function 1) [32]. Study of the crystal structure of albaflavenone monooxygenase showed that it exists as a dimer having two chains, each consisting of a single, α orthogonal bundle (Fig. 2.4). These studies also revealed a second,

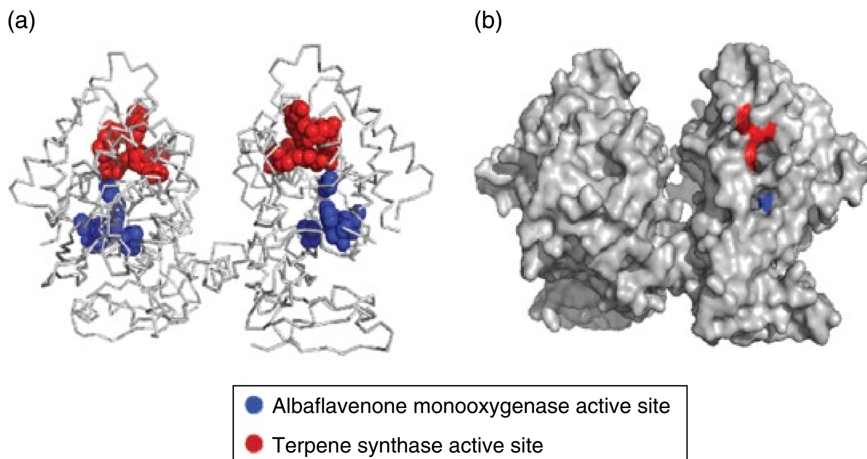


**Figure 2.4** Albaflavenone monooxygenase. The monooxygenase and terpene synthase active sites are shown in blue and red respectively in the (a) cartoon and (b) surface representation of Albaflavenone monooxygenase (PDB: 3EL3). (*See color plate section for the color representation of this figure.*)

completely distinct, catalytic activity of a terpene synthase (EC 4.2.3.47) [33] which is involved in the synthesis of farnesene isomers from farnesyl diphosphate (Function 2), by identification of signature sequences and motifs associated with terpene synthases. The residues Trp92, Pro274, Val338, Ile447, and Thr448 are involved in the monooxygenase activity, whereas the residues Arg116, Leu244, Leu248, Glu263, Val268, Leu271, Ile 272, and Phe-415 are associated with the terpene synthase activity which are located in different pockets in the protein. The monooxygenase activity was found to be optimal over the pH range 7.2–8 and was found to decline at lower pH values which favor terpene synthase activity (pH 5.5–6.5). This suggests that the two different enzymatic states of the protein possess optimal conformations at distinct pHs.

#### 2.5.1.3 MAPK1/ERK2, *Homo sapiens*

Studies to characterize the human protein-DNA interactome revealed the human mitogen-activated protein kinase 1 (MAPK1) or extracellular signal-r kinase 2 (ERK2) (Function 1) as a DNA-binding transcriptional repressor (Function 2) that regulates interferon gamma signaling [34]. The crystal structure of the human MAPK1 exists as a monomer which contains two domains: a discontinuous αβ 2-layer sandwich domain and a mainly α-orthogonal bundle domain (Fig. 2.5). The kinase active site residues involve Asp147, Lys149, Ser151, and Asn152. The motif that has been found to help in binding DNA are 259 K ARN Y LLSLP H K N K V P W N R277. It can be seen from Figure 2.5 that the kinase active site is located far from the DNA-binding motif.

### 2.5.2 Proteins with Distinct Sites for Different Functions in More Than One Domain

The second category of moonlighting proteins are multidomain proteins (listed in Table 2.2) which utilize functional sites in separate domains for carrying out their primary and moonlighting function(s).

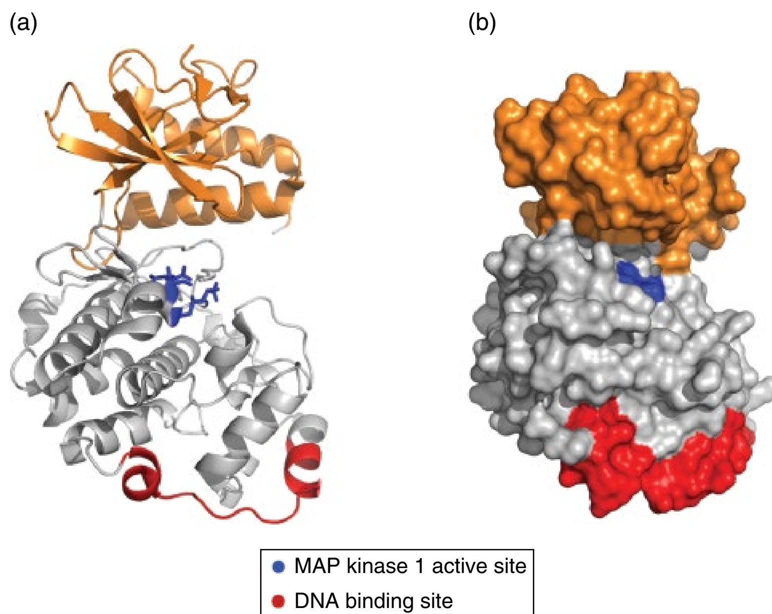


**Figure 2.5** Human MAPK1/ERK2. The MAPK1 active site is shown in blue and the DNA-binding motif is highlighted in red. Different domains are shown in gray and orange (PDB:4G6N). (*See color plate section for the color representation of this figure.*)

#### 2.5.2.1 Malate synthase, *Mycobacterium tuberculosis*

Malate synthase (EC 2.3.3.9) is a cytoplasmic enzyme (Function 1) involved in the glyoxalate pathway [57]. In *M. tuberculosis* it has also been found on the cell wall, adapted to function as an adhesin that binds laminin and fibrinogen which may contribute to *M. tuberculosis* virulence by promoting infection and dissemination (Function 2) [58]. The structure of the *M. tuberculosis* malate synthase consists of two identical chains each of which consists of four domains: an α-orthogonal bundle, a TIM barrel, a mainly β complex domain, and an α up-down bundle [58]. The malate synthase active site residues are Glu273, Asp274, Arg339, Glu434, Leu461, Asp462, and Glu633 (highlighted as blue sticks) and the residues that are associated with binding laminin or fibrinogen are Gln696–Glu727 (highlighted in red) (Fig. 2.6). Both the sites are present in different domain regions of the protein.

#### 2.5.2.2 BirA, *Escherichia coli*

The *E. coli* BirA protein performs different functions depending on its dimeric state [40]. As a heterodimer with biotin carboxyl carrier protein (BCCP), a subunit of acetyl-CoA carboxylase, it functions as a biotin protein ligase (Function 1); as a homodimer, it functions as a biotin operon repressor (Function 2) that binds to DNA [59]. The BirA structure consists of three domains: an α-orthogonal bundle, an α/β 2-layer sandwich domain and a mainly β SH3-type fold (Fig. 2.7). The residues responsible for the two functions of BirA are located in distinct sites in the protein. The catalytic residues for the ligase activity of the protein are Arg118, Lys183, and Arg317, and are found in a pocket formed between the αβ sandwich, the SH3 domain, and a helix-turn-helix (H-T-H) motif (residues 22–46) which is responsible for the binding DNA found in the α-orthogonal bundle (Fig. 2.7) [40].

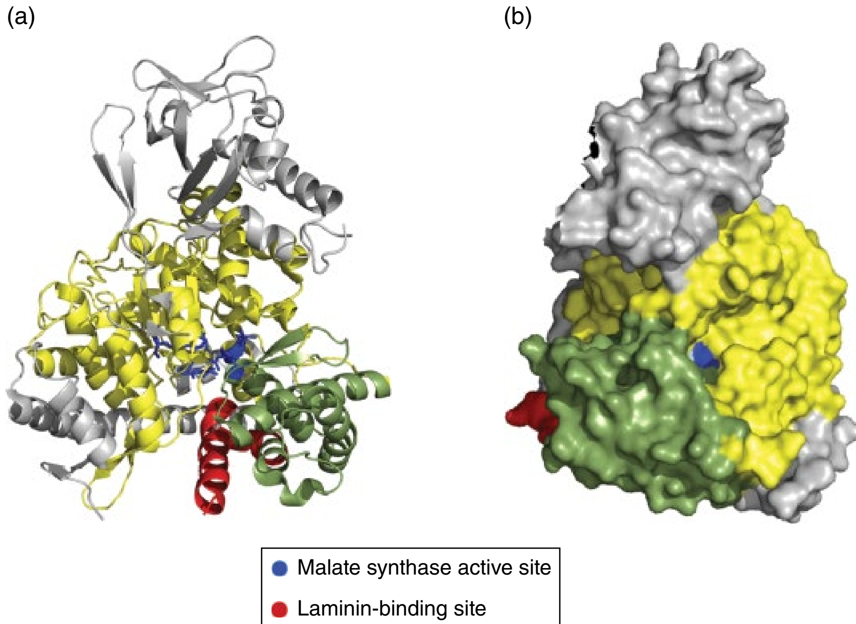


**Figure 2.6** Malate synthase. The enzyme active site is shown in blue and the laminin- binding site is shown in red. Different domains are shown in different colors (PDB:2GQ3). (*See color plate section for the color representation of this figure.*)

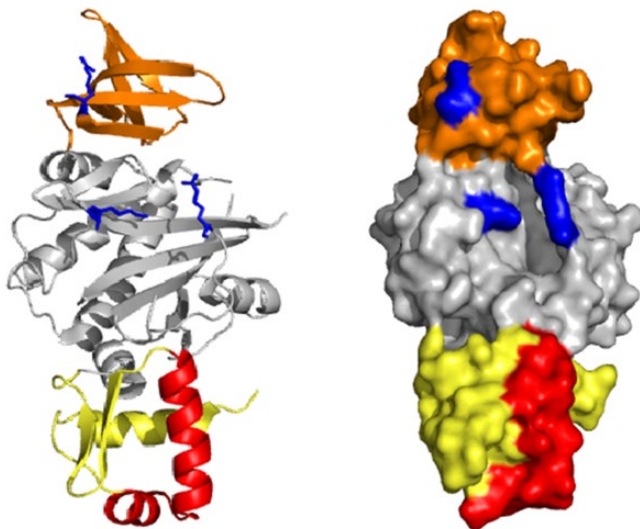

**Figure 2.7** BirA. The catalytic site residues are shown in blue while the H-T-H motif involved in binding DNA (moonlighting function) is shown in red. Different domains are shown in different colors (PDB:1BIB). (*See color plate section for the color representation of this figure.*)

#### 2.5.2.3 MRDI, *Homo sapiens*

The protein mediator of RhoA-dependent invasion (MRDI) is a moonlighting protein found in humans [41]. It acts both as a methylthioribose-1-phosphate (MTR-1-P) isomerase (EC 5.3.1.23) (Function 1) and a mediator of melanoma cell invasion (Function 2). The MRDI structure consists of two chains, each comprising a 4-helix bundle and a Rossmann fold (Fig. 2.8). The catalytic residues of MRDI are Cys168 and Asp248 (shown in blue), which are located in the base of a pocket. A potential binding site has been identified in MRDI by mutational analysis: Ser283 and Arg109 (shown in red). This is located distinct from the catalytic site in another pocket of the protein, formed between the two domains of each chain.

### 2.5.3 Proteins Using the Same Residues for Different Functions

These are multidomain proteins (listed in Table 2.3) which utilize the same functional site for carrying out their primary and moonlighting function(s).

#### 2.5.3.1 GAPDH *E. coli*

The *E. coli* glyceraldehyde-3-phosphate dehydrogenase (GAPDH; EC 1.2.1.12) (Function 1) is a multifunctional housekeeping protein. It also catalyzes its own NAD+ -dependent ADP-ribosylation which has been implicated in host–pathogen interactions (Function 2) [45]. GAPDH consists of two chains, each comprising a Rossmann fold and a $\alpha 3\beta 5$ sandwich domain [60]. The three catalytic residues of GAPDH are Cys149, His179, and Ser238, which are located in the sandwich domain (Fig. 2.9). However, mutational analyses have shown that the catalytic Cys149 (shown in red) is also the target residue of the ADP-ribosylation.

#### 2.5.3.2 Leukotriene A4 hydrolase, *Homo sapiens*

Leukotriene A4 hydrolase (EC 3.3.2.6) is a bifunctional zinc metalloenzyme that converts the fatty acid epoxide leukotriene A4 (LTA4) into a potent chemoattractant, leukotriene B4 (LTB4) (Function 1) and also exhibits an anion-dependent aminopeptidase activity (EC 3.4.11.24) (Function 2) [46]. Both the enzymatic activities require the presence of the catalytic zinc which is coordinated by the

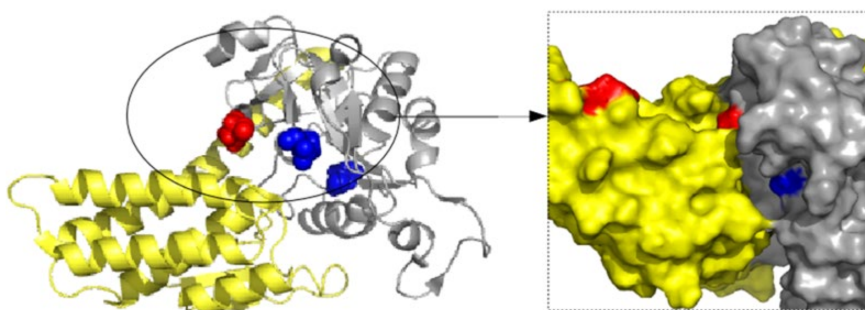

**Figure 2.8** Human MRDI. The active site residues are shown in blue while the residues implicated in controlling invasion (moonlighting function) is shown in red. Different domains are shown in different colors (PDB:4LDQ). (*See color plate section for the color representation of this figure.*)

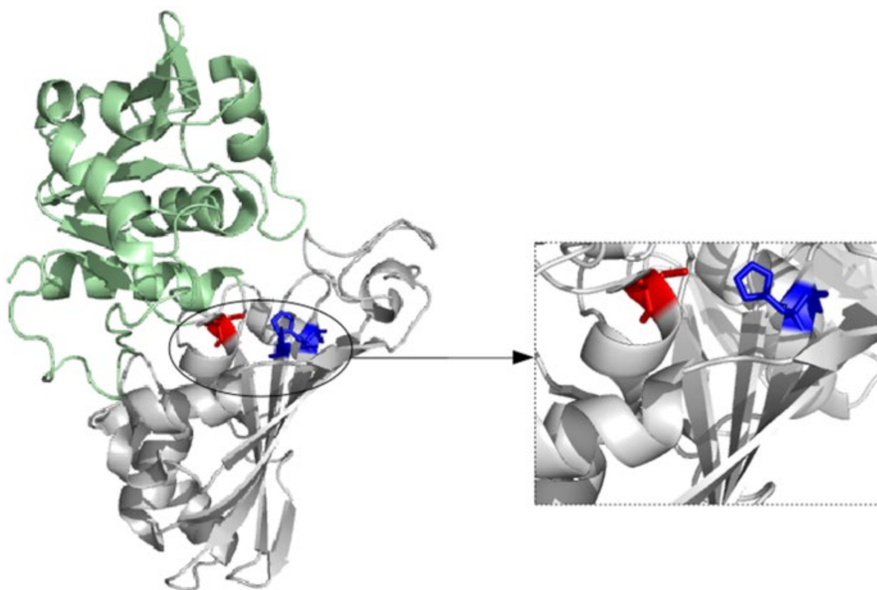

**Figure 2.9** GAPDH. The catalytic site residue Cys149 (shown in red) is the residue known to be involved for both the canonical and moonlighting functions of *E. coli* GAPDH. The other catalytic residue His179 is shown in blue (PDB:1DC5). (*See color plate section for the color representation of this figure.*)

three zinc-binding residues His295, His299, and Glu318. The crystal structure of the LTA4 hydrolase consists of three domains: a β-sandwich, an α-orthogonal bundle, and a α-α superhelix (Fig. 2.10) [61]. It also contains the 269 GX M EN272 motif in the M1 family of zinc peptidases. Mutation of the catalytic residues Glu296 or Tyr383 resulted in loss of the aminopeptidase activity, and mutation of the catalytic residue Glu271 abolished both the epoxide hydrolase activity and the aminopeptidase activity. Glu271 is a unique example of a catalytic residue that has distinct roles in two separate catalytic reactions for two chemically different substrates. Based on the LTA4 hydrolase structure and structure activity studies, two mechanistic models for the role of Glu271 in the epoxide hydrolase activity and in the aminopeptidase reaction were proposed [62].

### 2.5.4 Proteins Using Different Residues in the Same/Overlapping Site for Different Functions

Moonlighting proteins in this category are multidomain proteins (listed in Table 2.4) which utilize overlapping functional sites for carrying out their primary and moonlighting function(s).

#### 2.5.4.1 Phosphoglucose isomerase, *Oryctolagus cuniculus, Mus musculus, Homo sapiens*

Phosphoglucose isomerase (PGI, EC 5.3.1.9) is a glycolytic enzyme which catalyses the interconversion of glucose-6-phosphate and fructose-6-phosphate (Function 1). It is known to moonlight as an autocrine motility factor (tumor-secreted cytokine that promotes cellular growth and motility), neuroleukin (a neurotrophic factor

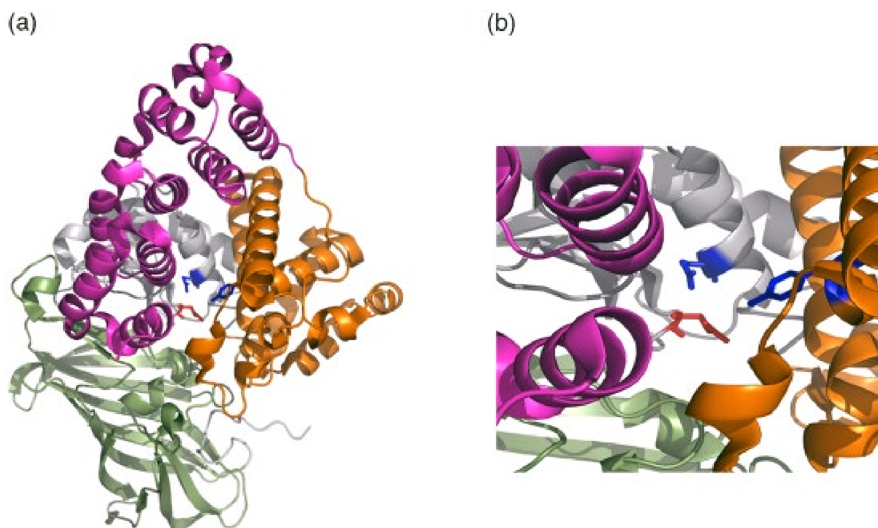


**Figure 2.10** Leukotriene A4 hydrolase. The LTA4 catalytic site residues Glu296 and Tyr383 are shown in blue. The catalytic site residue Glu271, involved in two separate functions in two different catalytic reactions is shown in red (PDB:2R59). (*See color plate section for the color representation of this figure.*)

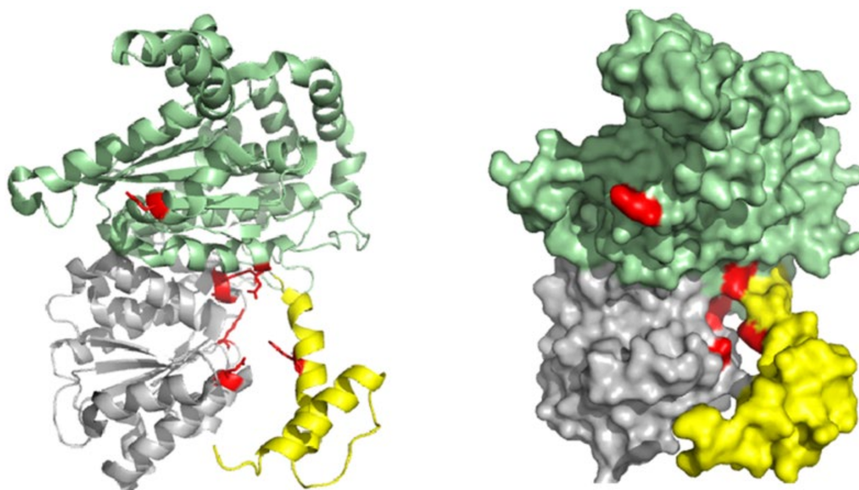

**Figure 2.11** Phosphoglucose isomerase (PGI). Catalytic residues are shown as red sticks. Inhibition of enzymatic and AMF functions of PGI by the PGI inhibitor and mutational analysis of the catalytic residues have indicated overlapping regions of both functions in the human PGI (PDB:1IAT). (*See color plate section for the color representation of this figure.*)

for neurons), and differentiation mediator in mammals (Function 2) [48]. The human PGI exists as a dimer comprising three domains: two (one large and one small) Rossman fold domains and an α-orthogonal bundle [49] (Fig. 2.11). The known catalytic site residues are Lys210, Glu216, Gly271, Arg272, Glu357, His388, and Lys518 (1IAT). The PGI inhibitor erythrose-4-phosphate (E4P) is known to inhibit both the enzymatic and cell motility activities of PGI. Moreover,

mutation of the catalytic residues resulted in significant reduction in the AMF or cell-motility-stimulating activity.

#### 2.5.4.2 Aldolase, *Plasmodium falciparum*

The fructose-bisphosphate aldolase (EC 4.1.2.13; Function 1) from apicomplexan parasites such as *P. falciparum* and *P. vivax* also provides a bridge between the actin filaments and TRAP (thrombospondin-related anonymous protein), which is critical for the host invasion machinery of the malaria parasite (Function 2) [50]. The *P. falciparum* aldolase structure consists of four chains, each consisting of a TIM barrel domain (Fig. 2.12). The aldolase active site residues and the residues involved in binding actin or TRAP overlap are located in the centre of the TIM barrel. The aldolase active site comprises the residues Asp39, Lys112, Glu194, and Lys236. The actin-binding residues of aldolase are Arg48, Lys112, Arg153, and Lys236 and the TRAP binding residues are Glu40, Lys47, Arg48, Lys151, Arg153, Arg309, and Gln312.

### 2.5.5 Proteins with Different Structural Conformations for Different Functions

These are "transformer proteins" [63] (listed in Table 2.5) which utilize different fold states for carrying out their primary and moonlighting function(s).

#### 2.5.5.1 RfaH, *E. coli*

RfaH is a bacterial antitermination protein which binds to the RNA polymerase (RNAP) and suppresses pausing, Rho-dependent inhibition, and intrinsic termination at a subset of sites (Function 1) [54]. As a result, termination signals

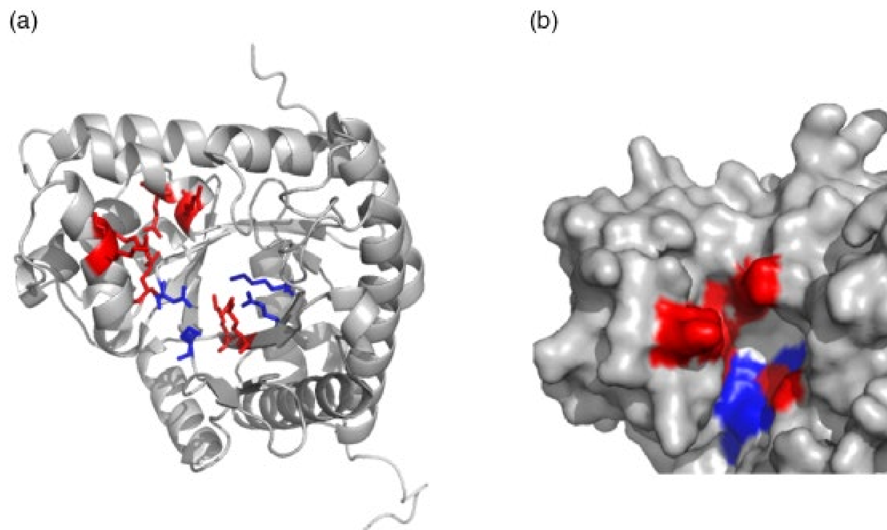


**Figure 2.12** Aldolase. The enzyme active site is shown in blue and the actin-binding site is shown in red (PDB:2PC4). (*See color plate section for the color representation of this figure.*)

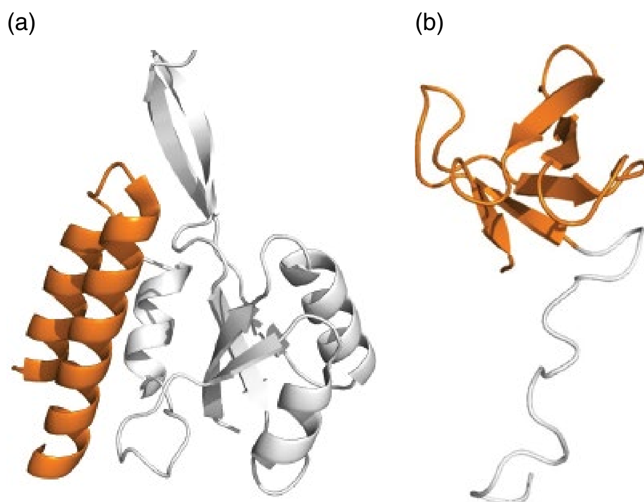


**Figure 2.13** RfaH The Rfah CTD is colored in orange. In the closed form of RfaH (a), the CTD (α-helix form) and NTD tightly interacts and works as a transcription factor (PDB:2OUG). The subsequent (or simultaneous) refolding of the CTD into a (b) β-barrel transforms RfaH into a translation factor (PDB:2LCL). (*See color plate section for the color representation of this figure.*)

are bypassed, which allows complete synthesis of long RNA chains. RfaH is a two-domain protein, and the two domains are observed to interact closely in the crystal structure. The RfaH N-terminal domain (NTD) has a central antiparallel sheet surrounded by helices and the C-terminal domain (CTD) in the crystal structure is an all-helical domain. However, the solution structure of the protein, solved by NMR, showed that the RfaH CTD folds into a helical structure when it interacts with the RfaH NTD and transforms into an all-sheet fold in the absence of NTD (Fig. 2.13). These two different fold states allow the protein to perform alternative functions. When the CTD exists in the all-sheet state, it can stimulate translation by recruiting a ribosome to an mRNA lacking a ribosome-binding site (Function 2) [55].

## 2.6 Conclusions

From the detailed analysis of the moonlighting proteins examined in this chapter, we can see considerable structural diversity in the types of domains that have evolved a moonlighting function, together with significant diversity in the different types of moonlighting functions that have evolved in these proteins. Some of the moonlighting proteins utilize different sites for their primary and moonlighting functions; however, there are others which use overlapping regions with their primary functional site or even the same site for both functions. Our investigation of moonlighting proteins revealed two general trends of functional site utilization in moonlighting proteins.

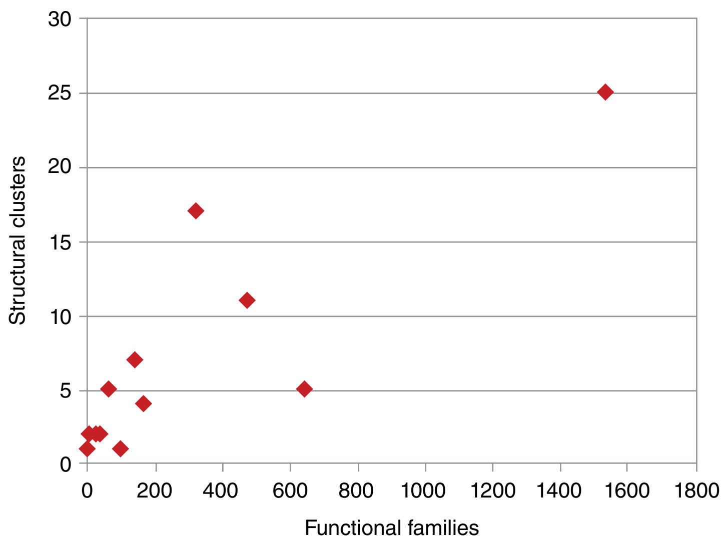


**Figure 2.14** Structural diversity v. functional diversity of CATH domain superfamilies represented in the moonlighting proteins studied in this chapter. Structural diversity is represented by the number of structural clusters (domains clustered at 5Å RMSD) in the superfamilies and the functional diversity is represented by the number of functional families identified in the superfamily.

1) Type 1: The primary functional site resides in the largest pocket of the protein while the moonlighting functional site is present on a distinct exposed surface of the protein. This is true for proteins for which binding to other proteins facilitates their moonlighting function. Examples include enolase, perodoxin, MAPK1, PutA, and I-AniI.
2) Type 2: The primary and moonlighting functional sites are present on two pockets or clefts in the protein structure. These sites can be utilized for two different enzymatic reactions or one enzymatic reaction along with a binding function. Examples include cytochrome c and albaflavenone monooxygenase.

Figure 2.14 shows the structural and functional diversity of the CATH domain superfamilies that are represented in the moonlighting proteins discussed in this chapter. We observe that these proteins belong to superfamilies ranging from very low to high structural and functional diversity. Knowledge of the structural and functional diversity sheds some light on the possible routes by which they may have acquired their moonlighting function. For example, for domain superfamilies with high structural diversity, it is more likely that the new functions can emerge through structural embellishments. By contrast, domain superfamilies having low structural diversity are more likely to evolve a new function by domain recruitment or use of different amino acids.

Given the diversity of moonlighting proteins in terms of functionality, physical locations in cell, mechanisms to moonlight, and the type of genomes in which they are found to exist, moonlighting proteins seem to be abundant in nature. Moonlighting could be a more common way of reutilizing or repurposing

proteins than the currently understood method. However, only around 300 moonlighting proteins are currently known, and <40% of these have known structures. As the number of moonlighting proteins with solved structures increases, we hope to pursue these structural analyses further.

## References

[1] Constance J Jeffery. Moonlighting proteins. *Trends in Biochemical Sciences*, 24(1): 8–11, 1999.

[2] Constance J Jeffery. Molecular mechanisms for multitasking: recent crystal structures of moonlighting proteins. *Current Opinion in Structural Biology*, 14(6): 663–668, 2004.

[3] Constance J Jeffery. Moonlighting proteins: complications and implications for proteomics research. *Drug Discovery Today: Targets*, 3(2): 71–78, 2004.

[4] Michael Ashburner, Catherine A Ball, Judith A Blake, David Botstein, Heather Butler, J Michael Cherry, Allan P Davis, Kara Dolinski, Selina S Dwight, Janan T Eppig, *et al*. Gene ontology: tool for the unification of biology. *Nature Genetics*, 25(1): 25–29, 2000.

[5] Amos Bairoch. The enzyme data bank. *Nucleic Acids Research*, 22(17): 3626–3627, 1994.

[6] Tjaart AP de Beer, Karel Berka, Janet M Thornton, and Roman A Laskowski. Pdbsum additions. *Nucleic Acids Research*, 42(D1): D292–D296, 2014.

[7] Matthew Bashton and Cyrus Chothia. The generation of new protein functions by the combination of domains. *Structure*, 15(1): 85–99, 2007.

[8] Liisa Holm and Chris Sander. Dali: a network tool for protein structure comparison. *Trends in Biochemical Sciences*, 20(11): 478–480, 1995.

[9] Christine A Orengo and William R Taylor. SSAP: sequential structure alignment program for protein structure comparison. *Methods in Enzymology*, 266: 617–635, 1996.

[10] Michael Levitt and Mark Gerstein. A unified statistical framework for sequence comparison and structure comparison. Proceedings of the National Academy of Sciences, 95(11): 5913–5920, 1998.

[11] Ilya N Shindyalov and Philip E Bourne. Protein structure alignment by incremental combinatorial extension (ce) of the optimal path. *Protein Engineering*, 11(9): 739–747, 1998.

[12] Angel R Ortiz, Charlie EM Strauss, and Osvaldo Olmea. Mammoth (matching molecular models obtained from theory): an automated method for model comparison. *Protein Science*, 11(11): 2606–2621, 2002.

[13] Yuzhen Ye and Adam Godzik. Fatcat: a web server for flexible structure comparison and structure similarity searching. *Nucleic Acids Research*, 32(suppl 2): W582–W585, 2004.

[14] Oliver C Redfern, Andrew Harrison, Tim Dallman, Frances MG Pearl, and Christine A Orengo. Cathedral: a fast and effective algorithm to predict folds and domain boundaries from multidomain protein structures. *PLoS Computational Biology*, 3(11): e232, 2007.

[15] Andrew CR Martin, Christine A Orengo, E Gail Hutchinson, Susan Jones, Maria Karmirantzou, Roman A Laskowski, John BO Mitchell, Chiara Taroni, and Janet M Thornton. Protein folds and functions. *Structure*, 6(7): 875–884, 1998.

[16] Ian Sillitoe, Tony E Lewis, Alison Cuff, Sayoni Das, Paul Ashford, Natalie L Dawson, Nicholas Furnham, Roman A Laskowski, David Lee, Jonathan G Lees, *et al.* Cath: comprehensive structural and functional annotations for genome sequences. *Nucleic Acids Research*, 43(D1): D376–D381, 2015.

[17] Christine A Orengo, David T Jones, and Janet M Thornton. Protein superfamilies and domain superfolds. *Nature*, 372(6507): 631–634, 1994.

[18] Chao Zhang and Charles DeLisi. Estimating the number of protein folds. *Journal of Molecular Biology*, 284(5): 1301–1305, 1998.

[19] Sridhar Govindarajan, Ruben Recabarren, and Richard A Goldstein. Estimating the total number of protein folds. *Proteins: Structure, Function, and Bioinformatics*, 35(4): 408–414, 1999.

[20] Annabel E Todd, Christine A Orengo, and Janet M Thornton. Evolution of function in protein superfamilies, from a structural perspective. *Journal of Molecular Biology*, 307(4): 1113–1143, 2001.

[21] Benoit H Dessailly, Natalie L Dawson, Kenji Mizuguchi, and Christine A Orengo. Functional site plasticity in domain superfamilies. *Biochimica et Biophysica Acta (BBA), Proteins and Proteomics*, 1834(5): 874–889, 2013.

[22] Antonio Gomez, Nuria Domedel, Juan Cedano, Jaume Piñol, and Enrique Querol. Do current sequence analysis algorithms disclose multifunctional (moonlighting) proteins? *Bioinformatics*, 19(7): 895–896, 2003.

[23] Ishita K Khan, Meghana Chitale, Catherine Rayon, and Daisuke Kihara. Evaluation of function predictions by pfp, esg, and psi-blast for moonlighting proteins. In BMC Proceedings, volume 6, page S5. BioMed Central Ltd, 2012.

[24] Antonio Gomez, Sergio Hernandez, Isaac Amela, Jaume Piñol, Juan Cedano, and Enrique Querol. Do protein–protein interaction databases identify moonlighting proteins? *Molecular BioSystems*, 7(8): 2379–2382, 2011.

[25] Sergio Hernandez, Isaac Amela, Juan Cedano, Jaume Piñol, JosepAntoni Perez-Pons, Angel Mozo-Villarias, and Enrique Querol. Do moonlighting proteins belong to the intrinsically disordered protein class. *Journal of Proteomics and Bioinformatics*, 5(11), 2012.

[26] Emmanuelle Becker, Benôit Robisson, Charles E Chapple, Alain Gúenoche, and Christine Brun. Multifunctional proteins revealed by overlapping clustering in protein interaction network. *Bioinformatics*, 28(1): 84–90, 2012.

[27] Mathew Mani, Chang Chen, Vaishak Amblee, Haipeng Liu, Tanu Mathur, Grant Zwicke, Shadi Zabad, Bansi Patel, Jagravi Thakkar, and Constance J Jeffery. Moonprot: a database for proteins that are known to moonlight. *Nucleic Acids Research*, 43: D277–282, 2014.

[28] Sergio Hernandez, Gabriela Ferragut, Isaac Amela, JosepAntoni Perez-Pons, Jaume Piñol, Angel Mozo-Villarias, Juan Cedano, and Enrique Querol. MultitaskprotDB: a database of multitasking proteins. *Nucleic Acids Research*, 42(D1): D517–D520, 2014.

[29] Ishita K Khan and Daisuke Kihara. Computational characterization of moonlighting proteins. *Biochemical Society Transactions*, 42(6): 1780–1785, 2014.

[30] Ishita Khan, Yuqian Chen, Tiange Dong, Xiaowei Hong, Rikiya Takeuchi, Hirotada Mori, and Daisuke Kihara. Genome-scale identification and characterization of moonlighting proteins. *Biology Direct*, 9(1): 30, 2014.

[31] Stefanie Ehinger, Wolf-Dieter Schubert, Simone Bergmann, Sven Hammerschmidt, and Dirk W Heinz. Plasmin (ogen)-binding α-enolase from Streptococcus pneumoniae: Crystal structure and evaluation of plasmin (ogen)-binding sites. *Journal of Molecular Biology*, 343(4): 997–1005, 2004.

[32] Bin Zhao, Xin Lin, Li Lei, David C Lamb, Steven L Kelly, Michael R Waterman, and David E Cane. Biosynthesis of the sesquiterpene antibiotic albaflavenone in streptomyces coelicolor a3 (2). *Journal of Biological Chemistry*, 283(13): 8183–8189, 2008.

[33] Bin Zhao, Li Lei, Dmitry G Vassylyev, Xin Lin, David E Cane, Steven L Kelly, Hang Yuan, David C Lamb, and Michael R Waterman. Crystal structure of albaflavenone monooxygenase containing a moonlighting terpene synthase active site. *Journal of Biological Chemistry*, 284(52): 36711–36719, 2009.

[34] Shaohui Hu, Zhi Xie, Akishi Onishi, Xueping Yu, Lizhi Jiang, Jimmy Lin, Hee-sool Rho, Crystal Woodard, Hong Wang, Jun-Seop Jeong, *et al*. Profiling the human protein-dna interactome reveals erk2 as a transcriptional repressor of interferon signaling. *Cell*, 139(3): 610–622, 2009.

[35] Aron B Fisher. Peroxiredoxin 6: a bifunctional enzyme with glutathione peroxidase and phospho- lipase a2 activities. *Antioxidants & Redox Signaling*, 15(3): 831–844, 2011.

[36] Maria LR Lim, May-Ghee Lum, Tania M Hansen, Xavier Roucou, and Phillip Nagley. On the release of cytochromec from mitochondria during cell death signaling. *Journal of Biomedical Science*, 9(6): 488–506, 2002.

[37] Yaroslav Nikolaev, Christine Deillon, Stefan RK Hoffmann, Laurent Bigler, Sebastian Friess, Renato Zenobi, Konstantin Pervushin, Peter Hunziker, and Bernd Gutte. The leucine zipper domains of the transcription factors gcn4 and
c-jun have ribonuclease activity. *PloS One*, 5(5): e10765, 2010.

[38] Ryo Takeuchi, Michael Certo, Mark G Caprara, Andrew M Scharenberg, and Barry L Stoddard. Optimization of in vivo activity of a bifunctional homing endonuclease and maturase reverses evolutionary degradation. *Nucleic Acids Research*, 37(3): 877–890, 2009.

[39] Arvind G Kinhikar, Diana Vargas, Hualin Li, Spencer B Mahaffey, Laura Hinds, John T Belisle, and Suman Laal. Mycobacterium tuberculosis malate synthase is a laminin-binding adhesin. *Molecular Microbiology*, 60(4): 999–1013, 2006.

[40] Keith P Wilson, Lisa M Shewchuk, Richard G Brennan, Anthony J Otsuka, and Brian W Matthews. Escherichia coli biotin holoenzyme synthetase/bio repressor crystal structure delineates the biotin- and dna-binding domains. Proceedings of the National Academy of Sciences, 89(19): 9257–9261, 1992.

[41] Paul D Templeton, Elizabeth S Litman, Sandra I Metzner, Natalie G Ahn, and Marcelo C Sousa. Structure of mediator of rhoa-dependent invasion (mrdi) explains its dual function as a metabolic enzyme and a mediator of cell invasion. *Biochemistry*, 52(33): 5675–5684, 2013.

[42] Carlos Gancedo and Carmen-Lisset Flores. Moonlighting proteins in yeasts. *Microbiology and Molecular Biology Reviews*, 72(1): 197–210, 2008.

[43] Brent A Appleton, Ping Wu, Janice Maloney, JianPing Yin, Wei-Ching Liang, Scott Stawicki, Kyle Mortara, Krista K Bowman, J Michael Elliott, William Desmarais, *et al*. Structural studies of neuropilin/antibody complexes provide insights into semaphorin and vegf binding. *The EMBO Journal*, 26(23): 4902–4912, 2007.

[44] Anindita Bhoumik, Shoichi Takahashi, Wolfgang Breitweiser, Yosef Shiloh, Nic Jones, and Zeev Ronai. Atm-dependent phosphorylation of atf2 is required for the dna damage response. *Molecular Cell*, 18(5): 577–587, 2005.

[45] Laura Aguilera, Rosa Giménez, Josefa Badia, Juan Aguilar, and Laura Baldoma. Nad + -dependent post-translational modification of escherichia coli glyceraldehyde-3-phosphate dehydrogenase. *International Microbiology*, 12(3): 187–192, 2010.

[46] Jesper Z Haeggstrom. Leukotriene a4 hydrolase/aminopeptidase, the gatekeeper of chemotactic leukotriene b4 biosynthesis. *Journal of Biological Chemistry*, 279(49): 50639–50642, 2004.

[47] Susan Crennell, Toru Takimoto, Allen Portner, and Garry Taylor. Crystal structure of the multifunctional paramyxovirus hemagglutinin-neuraminidase. *Nature Structural & Molecular Biology*, 7(11): 1068–1074, 2000.

[48] Constance J Jeffery, Brian J Bahnson, Wade Chien, Dagmar Ringe, and Gregory A Petsko. Crystal structure of rabbit phosphoglucose isomerase, a glycolytic enzyme that moonlights as neuroleukin, autocrine motility factor, and differentiation mediator. *Biochemistry*, 39(5): 955–964, 2000.

[49] Jon Read, Jake Pearce, Xiaochun Li, Hilary Muirhead, John Chirgwin, and Christopher Davies. The crystal structure of human phosphoglucose isomerase at 1.6 °a resolution: implications for catalytic mechanism, cytokine activity and haemolytic anaemia. *Journal of Molecular Biology*, 309(2): 447–463, 2001.

[50] Jürgen Bosch, Carlos A Buscaglia, Brian Krumm, Bjarni P Ingason, Robert Lucas, Claudia Roach, Timothy Cardozo, Victor Nussenzweig, and Wim GJ Hol. Aldolase provides an unusual binding site for thrombospondin-related anonymous protein in the invasion machinery of the malaria parasite. Proceedings of the National Academy of Sciences, 104(17): 7015–7020, 2007.

[51] Patrick Scheerer, Astrid Borchert, Norbert Krauss, Helga Wessner, Christa Gerth, Wolfgang Höhne, and Hartmut Kuhn. Structural basis for catalytic activity and enzyme polymerization of phospholipid hydroperoxide glutathione peroxidase-4 (gpx4). *Biochemistry*, 46(31): 9041–9049, 2007.

[52] Xiao Luo, He-Hsuan Hsiao, Mikhail Bubunenko, Gert Weber, Donald L Court, Max E Gottesman, Henning Urlaub, and Markus C Wahl. Structural and functional analysis of the e. coli nusb-s10 transcription antitermination complex. *Molecular Cell*, 32(6): 791–802, 2008.

[53] OA Bateman, AG Purkiss, R Van Montfort, C Slingsby, C Graham, and G Wistow. Crystal structure of η-crystallin: adaptation of a class 1 aldehyde dehydrogenase for a new role in the eye lens. *Biochemistry*, 42(15): 4349–4356, 2003.

[54] Georgiy A Belogurov, Marina N Vassylyeva, Vladimir Svetlov, Sergiy Klyuyev, Nick V Grishin, Dmitry G Vassylyev, and Irina Artsimovitch. Structural basis

for converting a general transcription factor into an operon-specific virulence regulator. *Molecular Cell*, 26(1): 117–129, 2007.

[55] Bjorn M Burmann, Stefan H Knauer, Anastasia Sevostyanova, Kristian Schweimer, Rachel A Mooney, Robert Landick, Irina Artsimovitch, and Paul Rösch. An
α helix to β barrel domain switch transforms the transcription factor rfah into a translation factor. *Cell*, 150(2): 291–303, 2012.

[56] Robbyn L Tuinstra, Francis C Peterson, Snjezana Kutlesa, E Sonay Elgin, Michael A Kron, and Brian F Volkman. Interconversion between two unrelated protein folds in the lymphotactin native state. Proceedings of the National Academy of Sciences, 105(13): 5057–5062, 2008.

[57] NE Tolbert. Metabolic pathways in peroxisomes and glyoxysomes. *Annual Review of Biochemistry*, 50(1): 133–157, 1981.

[58] David M Anstrom and S James Remington. The product complex of m. tuberculosis malate synthase revisited. *Protein Science*, 15(8): 2002–2007, 2006.

[59] Constance J Jeffery. Proteins with neomorphic moonlighting functions in disease. *IUBMB Life*, 63(7): 489–494, 2011.

[60] Mikyung Yun, Cheon-Gil Park, Ji-Yeon Kim, and Hee-Won Park. Structural analysis of glycer-aldehyde 3-phosphate dehydrogenase from escherichia coli: direct evidence of substrate binding and cofactor-induced conformational changes. *Biochemistry*, 39(35): 10702–10710, 2000.

[61] Fredrik Tholander, Ayumo Muroya, Bernard-Pierre Roques, Marie-Claude Fournié-Zaluski, Marjolein MGM Thunnissen, and Jesper Z Haeggstrom. Structure-based dissection of the active site chemistry of leukotriene a4 hydrolase: implications for m1 aminopeptidases and inhibitor design. *Chemistry & Biology*, 15(9): 920–929, 2008.

[62] Peter C Rudberg, Fredrik Tholander, Marjolein MGM Thunnissen, and Jesper Z Haeggström. Leukotriene a4hydrolase/aminopeptidase glutamate 271 is a catalytic residue with specific roles in two distinct enzyme mechanisms. *Journal of Biological Chemistry*, 277(2): 1398–1404, 2002.

[63] Stefan H Knauer, Irina Artsimovitch, and Paul Rösch. Transformer proteins. *Cell Cycle*, 11(23): 4289–4290, 2012.