Appl Biochem Biotechnol (2011) 165:397–405
DOI 10.1007/s12010-011-9259-2

Insights to Sequence Information of Polyphenol Oxidase Enzyme from Different Source Organisms

Neha Malviya · Mugdha Srivastava · Sanjeev Kumar Diwakar · Sarad Kumar Mishra

Received: 22 August 2010 / Accepted: 4 April 2011 /
Published online: 27 April 2011

Abstract Polyphenol oxidases (PPOs) are widely distributed enzymes among animals, plants, bacteria, and fungi. PPOs often have significant role in many biologically essential functions including pigmentation, sclerotization, primary immune response, and host defense mechanisms. In the present study, forty-seven full-length amino acid sequences of PPO from bacteria, fungi, and plants were collected and subjected to multiple sequence alignment (MSA), domain identification, and phylogenetic tree construction. MSA revealed that six histidine, two phenylalanine, two arginine, and two aspartic acid residues were highly conserved in all the analyzed species, while a single cysteine residue was conserved in all the plant and fungal PPOs. Two major sequence clusters were constructed by phylogenetic analysis. One cluster was of the plant origin, whereas the other one was of the fungal and bacterial origin. Motif GGGMMGDVPTANDPIFWLHHCNVDRLWAVWQ was found in all the species of bacterial and fungus sources. In addition, seven new motifs which were unique for their group were also identified.

Keywords Polyphenol oxidase · Sequence analysis · Phylogenetic analysis · Conserved regions · Motifs

Introduction

Polyphenol oxidase (tyrosinase) is a multifunctional Cu-containing enzyme, which is widely distributed in nature, and can be found throughout the phylogenetic scale from bacteria to human. The enzyme is able to insert oxygen in a position ortho- to an existing

Electronic supplementary material The online version of this article (doi:10.1007/s12010-011-9259-2) contains supplementary material, which is available to authorized users.

N. Malviya (✉) · S. K. Diwakar · S. K. Mishra
Department of Biotechnology, Deen Dayal Upadhyay Gorakhpur University, Gorakhpur, India
e-mail: malviya.neha@gmail.com

M. Srivastava
Department of Botany, Atarra P.G. College, Bundelkhand University, Bundelkhand, India

hydroxyl group in an aromatic ring, followed by the oxidation of the diphenol to the corresponding quinone [1]. A number of different biological functions have been attributed to polyphenol oxidase (PPO) in different organisms. PPO is reported to be involved in defense responses in plants during injury, herbivory, or pathogen attack. It may also play a significant role in response to stress conditions, browning of fruits, betalain biosynthesis, and in fungal host–pathogen interactions [2]. Owing to the importance of the enzyme in different biological reactions and functions, a lot of work has been done and is being carried out by different workers. It has been found that PPOs are quite diverse in many of their properties, distribution, and cellular location [3]. Wichers and coworkers have reported the PPOs cluster in groups for higher plants, vertebrate animals, fungi, and bacteria in the phylogenetic tree of PPO [4]. Homologies within such clusters are considerably higher than between them. Considering the above facts, a study of amino acid sequences of PPOs from different sources of organisms is quite challenging. In the present study, we performed the individual in silico studies of amino acid sequences obtained from plants, fungi, and bacteria and correlated them on the basis of some common features.

Materials and Methods

The full-length amino acid sequences of PPO from bacteria, fungi, and plants were searched and retrieved from Entrez protein database available at NCBI (www.ncbi.nlm.nih.gov). The sequences were arranged in plant, fungal, and bacterial profile, respectively. The multiple sequence alignment of the individual profiles was performed using Clustal-W2 at the European Bioinformatics Institute (http://www.ebi.ac.uk/). Motifs were discovered in profiles using the expectation maximization approach [5] implemented in Multiple EM for Motif Elicitation server (http://meme.sdsc.edu). Further, the discovered motifs were used to search the similar pattern of motifs in NCBI nonredundant database [6] using Motif Alignment and Search Tool. The maximum parsimony approach implemented in the Mega program [7] was employed for constructing phylogenetic relationships among sequences. The statistical reliability of the phylogenetic tree was tested by bootstrap analyses with 500 replications.

Results

Sequence Analysis The accession number of retrieved sequences along with the species name and origin is listed in Table 1. MSA showed the presence of some conserved regions in all the sequences from different sources, while others were restricted only to their groups. All the analyzed species of plant, fungus, and bacteria possessed six histidine, two phenylalanine, two arginine, and two aspartic acid residues. According to the previous studies, all the species possess six histidine residues that ligate two copper ions of the active site. The first three histidine residues encompass a region termed the CuA site, and the following three encompass the CuB site [8]. The conservation of the arginine and an aromatic tyrosine or phenylalanine as the first residue of the tyrosine motif for all PPOs has been also supported by other coworkers [7]. The MSA of 15 protein sequences of bacteria revealed the stretches of conserved protein sequences from residues 57 to 100, 202 to 228, and 229 to 253 (Supplement A). All the species of bacteria possessed two phenylalanine, six histidine, two aspartic acid, two arginine, two tryptophan, and one proline residues. All the species of fungus possessed four phenylalanine, six histidine, two aspartic acid, two

Table 1 Retrieved sequence from NCBI/Entrez and their accession number

Serial no.	Source	Species	Accession no.
1	Bacteria	*Paracoccus denitrificans* PD1222	YP_918400.1
2	Bacteria	*Delftia acidovorans* SPH-1	YP_001562639.1
3	Bacteria	*Burkholderia pseudomallei* 1710b	ABA51756.1
4	Bacteria	*Frankia* sp. EAN1pec	YP_001505111.1
5	Bacteria	*Rubrobacter xylanophilus* DSM 9941	YP_644706.1
6	Bacteria	*Nitrosospira multiformis* ATCC 25196	ABB74223.1
7	Bacteria	*Marinomonas mediterranea*	AAV49996.1
8	Bacteria	*Nostoc punctiforme* PCC 73102	YP_001865661.1
9	Bacteria	*Bacillus megaterium* QM B1551	YP_003563841.1
10	Bacteria	*Bacillus cereus*	AAP92115.1
11	Bacteria	*Streptomyces antibioticus*	AAA88571.1
12	Bacteria	*Nitrosospira multiformis* ATCC 25196	YP_411227.1
13	Bacteria	*Bacillus megaterium* DSM319	YP_003598579.1
14	Bacteria	*Nitrobacter hamburgensis* X14	YP_578167.1
15	Bacteria	*Rhodopseudomonas palustris* HaA2	YP_485784.1
16	Fungus	*Aspergillus fumigatus*	CAC82195.1
17	Fungus	*Aspergillus flavus* NRRL3357	XP_002380087.1
18	Fungus	*Ajellomyces capsulatus* G186AR	EEH09219.1
19	Fungus	*Penicillium chrysogenum* Wisconsin 54-1255	XP_002569175.1
20	Fungus	*Tuber melanosporum*	CAO82078.1
21	Fungus	*Agaricus bisporus*	ACU29457.1
22	Fungus	*Podospora anserina*	AAB07484.1
23	Fungus	*Neurospora crassa*	TYRO_NEUCR
24	Fungus	*Paracoccidioides brasiliensis* (strain ATCC MYA-826/Pb01)	C1H6P8_PARBA
25	Fungus	*Sclerotinia sclerotiorum* (strain ATCC 18683/1980/Ss-1	A7FA30_SCLS1
26	Fungus	*Verticillium albo-atrum* (strain VaMs.102)	C9SWR4_VERA1
27	Fungus	*Nannizzia otae* (strain CBS 113480)	C5FTV9_NANOT
28	Fungus	*Tuber borchii*	B7VEU9_TUBBO
29	Fungus	*Laccaria bicolor* (strain S238N-H82)	B0DMA1_LACBS
30	Fungus	*Pholiota nameko*	TYRO_PHONA
31	Plant	*Argemone mexicana*	ACJ76786.1
32	Plant	*Camellia nitidissima*	ACM43505.1
33	Plant	*Ananas comosus*	AAO16865.1
34	Plant	*Taraxacum officinale*	CAQ76694.1
35	Plant	*Triticum monococcum*	ACB12083.1
36	Plant	*Juglans regia*	ACN86310.1
37	Plant	*Vitis vinifera*	AAB41022.1
38	Plant	*Cydonia oblonga*	ABY84850.1
39	Plant	*Camellia ptilophylla*	ABF19601.1
40	Plant	*Oryza sativa*	ACS15319.1
41	Plant	*Pyrus pyrifolia*	BAB64530.1
42	Plant	*Triticum aestivum*	BAH36895.1
43	Plant	*Medicago truncatula*	A2Q4Q3_MEDTR
44	Plant	*Populus tremuloides*	Q94KC2_POPTM
45	Plant	*Solanum tuberosum*	Q41428_SOLTU
46	Plant	*Zea mays*	B6SVR5_MAIZE
47	Plant	*Malus domestica*	Q93XM8_MALDO

arginine, one cysteine, and other residues which were conserved within the sequence profile (Supplement B). All the plant species possessed eight phenylalanine, six histidine, seven aspartic acid, six arginine, five cysteine, and other amino acid residues within tyrosine motif. Cysteine residue was conserved in all the plant and fungal PPOs. Owing to the conservation of these cysteines, it is likely that all plant and fungal PPOs may have a thioether bridge as a structural component of active sites. However, it appears that a thioether bridge is important, if not essential, to the structure and the function of the active site in the plant and the fungal PPOs. [8]. The active-site histidines and thioether bridge have been discussed by various researchers in the context of a variety of species, ranging from bacteria to mammals [9–13].

The Clustal-W2 result predicted that the sequences retrieved from plants had an average sequence homology with an alignment score above 65 between *Camellia nitidissima*/*Cydonia oblonga*, *C. nitidissima*/*Taraxacum officinale*, *C. nitidissima*/*Camellia ptilophylla*, *Triticum monococcum*/*Oryza sativa*, *Juglans regia*/*C. oblonga*, and *O. sativa*/*Triticum aestivum.* The bacterial and fungal PPOs were highly diversed in their respective groups owing to the diversity in the PPO sequences. Phylogenetic analysis of the sequences of bacteria showed the two major clusters (Fig. 1). Cluster I consist of nine species which was further divided into three subclusters. Subcluster I contains *Bacillus cereus*, *Streptomyces antibioticus*, and *Rubrobacter xylanophilus*. Subcluster II contains *Nostoc punctiforme* and *Bacillus megaterium* (QM B1551 and DSM319). Subcluster III contains both the species of *Nitrosospira multiformis. Frankia* sp. was outgrouped in cluster I. Cluster II consist of five species namely *Burkholderia pseudomallei*, *Paracoccus denitrificans*, *Nitrobacter hamburgensis*, *Delftia acidovorans*, and *Marinomonas mediterranea. Rhodopseudomonas palustris* was distinct from both the clusters.

Phylogenetic analysis of the sequences of fungus showed the two major clusters (Fig. 2). Cluster I consists of seven species which was divided into two subclusters. Subcluster I contains five species (*Podospora anserine*, *Neurospora crassa*, *Sclerotinia sclerotiorum*, *Verticillium albo-atrum*, *Nannizzia otae*) and II contains two species (*Tuber melanosporum*, *Tuber borchii*). Cluster II consist of five species which was divided into two subclusters. Subcluster I contains two species (*Aspergillus fumigates*, *Aspergillus flavus*) and II contains

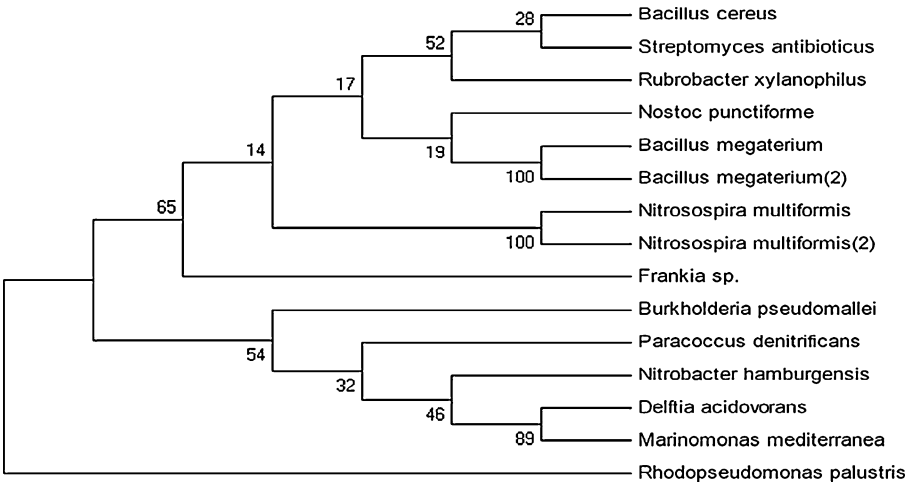

Fig. 1 Phylogenetic analysis of retrieved sequences of bacteria using maximum parsimony method

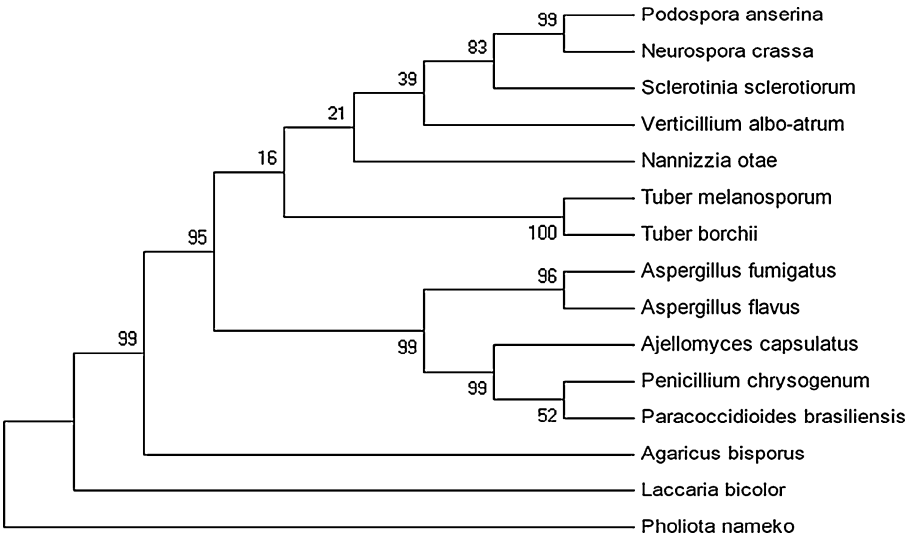

Fig. 2 Phylogenetic analysis of retrieved sequences of fungus using maximum parsimony method

three species (*Ajellomyces capsulatus*, *Penicillium chrysogenum*, *Paracoccidioides brasiliensis*). *Agaricus bisporus*, *Laccaria bicolor*, and *Pholiota nameko* were distant from all the other species, so they were not included in any cluster.

In the same way, the sequences of plants showed the two major clusters (Figs. 3 and 4). Cluster I contains eight species which was further divided into two subclusters. Subcluster I contains three species (*Solanum tuberosum*, *Zea mays*, *Argemone mexicana*) and II contains three species (*O. sativa*, *T. monococcum*, *T. aestivum*), while *Ananas comosus and Vitis vinifera* were distantly located. Cluster II contains six species divided into two subcluster.

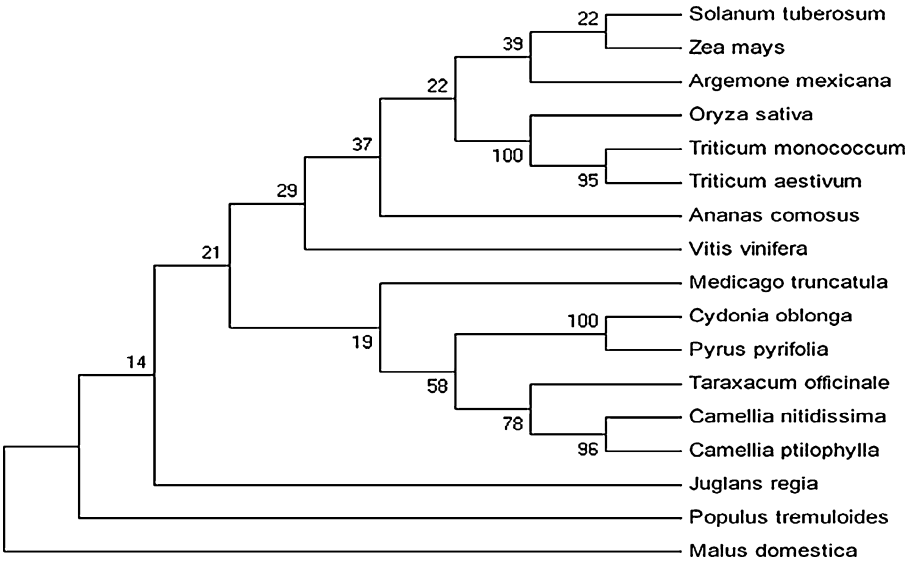

Fig. 3 Phylogenetic analysis of retrieved sequences of plant using maximum parsimony method

Fig. 4 Phylogenetic analysis of total retrieved sequences using maximum parsimony method

Subcluster I consists of two species (*C. oblonga*, *Pyrus pyrifolia*) and II contains three species (*T. officinale*, *C. nitidissima*, *C. ptilophylla*). *Medicago truncatula* was not included in any subcluster. *J. regia*, *Populus tremuloides*, and *Malus domestica* were outgrouped and could not be included in any cluster.

When complete sequences were taken for phylogenetic analysis, two clusters were formed. The first cluster was further divided into two subclusters of bacteria and fungus origin while the cluster second was of plant origin. This suggests that bacteria and fungus were more closely related in terms of PPO sequences, with respect to plants.

A stretch of 31 amino acids GGGMMGDVPTANDPIFWLHHCNVDRLWAVWQ was conserved in all the species of fungi and bacteria with some substitutions, but not present in plants (Table 1). These were found to be associated with tyrosinase superfamily. Other

Table 2 Motif searched with MEME along with their pfam analysis

Serial no.	Source	Motif width	Motif present in number of sequences	Motif	pfam
1	Bacteria	31	15	GGGMMGDVPTANDPIFWLHHCNVDRLWAVWQ	Tyrosinase superfamily(common central domain of tyrosinase)
2	Bacteria	22	15	HGNWAFLPWHREYLLRFERALQ	Tyrosinase superfamily (common central domain of tyrosinase)
3	Bacteria	30	10	VRKNQADLTDEEKAAFVRALLTLKEKGIYD	SDH 5 family (Protein of unknown function), DUB 1542 domain (domain of unknown function)
4	Bacteria	16	15	SINPDVTLPYWDWTAD	Tyrosinase superfamily (common central domain of tyrosinase)
5	Bacteria	16	11	TPDDVLDHRKLGYTYD	PPO1_DWL (PPO middle domain)
6	Bacteria	22	10	TLPTRADVLDALAITQYDTAPW	pfam entry not found
7	Bacteria	15	10	GGPAGHNLNDPMWPW	pfam entry not found
8	Bacteria	11	10	IWAADFMGGNG	pfam entry not found
9	Bacteria	7	10	FRNQLEG	pfam entry not found
10	Fungus	31	15	GGGHMGDVPVAFDPIFWLHHCNVDRLFAIWQ	Tyrosinase superfamily (common central domain of tyrosinase)
11	Fungus	31	15	WGGYCTHGSILFPTWHRPYLALFEQRLYEIA	Tyrosinase superfamily (common central domain of tyrosinase)
12	Fungus	22	15	KDRLSYFQIAGIHGLPYIPWDE	Tyrosinase superfamily (common central domain of tyrosinase)
13	Fungus	25	14	RAEWVEAAKTLRLPYWDWASNPPVP	Tyrosinase superfamily (common central domain of tyrosinase)
14	Fungus	16	11	RDTTTFGYTYPELQxW	pfam entry not found
15	Fungus	24	10	CGNCADEAADLLASGTVPLTPAL	pfam entry not found
16	Fungus	22	14	PGKKKEVPNPLYHYKFHPSNPS	pfam entry not found
17	Fungus	21	10	LTSLDQEEVEPYLRKNLHWRV	pfam entry not found
18	Plant	43	16	LQVHNSWLFFPFHRYYLYFFERILGKLIGDPTFALPFWNWDAP	Tyrosinase superfamily (common central domain of tyrosinase)
19	Plant	40	16	PTQPNGEDMGNFYSAARDPIFFAHHSNVDRMWSIWKTLGG	Tyrosinase superfamily (common central domain of tyrosinase)
20	Plant	36	16	DPDWLDSEFLFYDENAELVRVKVRDCLDTKKLGYVY	PPO_DWL (PPO middle domain)
21	Plant	31	16	TxLRLGITDLLEDLGAEDDDSVVVTLVPRYG	PPO1_KFPV (protein of unknown function)
22	Plant	25	13	EEEVLVIEGIEFDRDVFVKFDVFIN	PPO1_KFPV (protein of unknown function)
23	Plant	22	16	PDKTEFAGSFVNVPHKHKHGKK	PPO1_KFPV (protein of unknown function)

motifs as given in Table 2 are associated with tyrosinase superfamily, PPO_DWL superfamily and PPO1_KFDV superfamily. Tyrosinase superfamily is part of PPO and some hemocyanins. Tyrosinase binds two copper ions (CuA and CuB). Each of the two copper ions is bound by three conserved histidine residues. Regions around these copper-binding ligands are well conserved and also shared by some hemocyanins which were copper-containing oxygen carriers from the hemolymph of many mollusks and arthropods. PPO_DWL superfamily (Polyphenol oxidase middle domain) is found in bacteria and eukaryotes and is approximately of 50 amino acids in length. There was a conserved DWL sequence motif which gives the family its name. This domain family is found in eukaryotes, and is typically between 138 and 152 amino acids in length. PPO1_KFDV superfamily (protein of unknown function) is plant or plastid PPO, with a highly conserved sequence motif: KFDV, representing the C-terminal domain of these oxidases. The pfam entry of some new motifs is not found, but they are unique for their group, i.e., these motifs are present in all the sequences of their group and are listed in Table 3.

Discussion

In silico analysis of the sequences showed sequence based similarities depending on their source organism. Six histidine, two phenylalanine, two arginine and two aspartic acid residues were common in all the species which were studied. The MSA also revealed that cysteine is conserved in all the plant and fungal tyrosinases taken. Because of the conservation of these cysteines, it is likely that all the plant and fungal PPOs have a thioether bridge as a structural component of their CuA sites. Clustering of the sequences further confirmed the similarities on the basis of their source. A specific 31 amino acids motif GGGMMGDVPTANDPIFWLHHCNVDRLWAVWQ observed in all the species of bacterial and fungus sources leads to an indication that the sequences from bacteria and fungi are more closely related. A number of different functions including biosynthetic process, browning reactions, resistance to stress, and pathogens and several others have been ascribed to PPO; however, at sequence level except for active site and some common motifs, considerable variability has been observed. This leads to the lack of clarity in the function of PPOs. Seven new motifs which were unique for their group whose functional attributes are needed to be verified experimentally were identified. Apart from the study of

Table 3 Some new motifs present in all the organism of similar source

Serial no.	Source	Motif width	Motif present in number of sequences	Motif
1	Bacteria	9	15 (all the sequences)	QPHNRVHNW
2	Fungus	15	15 (all the sequences)	GSNDISLEAIHNNIH
3	Fungus	22	15 (all the sequences)	RNNPKQFNLFVQALQDFQALDE
4	Plant	50	16 (all the sequences)	AHLVDADYLAKYKKAIELMKALPDD DPRSFKQQANVHCAYCDGAYDQVGF
5	Plant	31	16 (all the sequences)	MYRQMVSGAKKPRLFLGSPY RAGDEPDPGAG
6	Plant	43	16 (all the sequences)	GMQLPAIYADPNSPLYDELRNAK HQPPTLIDLDYNGTDPNITD
7	Plant	16	16 (all the sequences)	SIENVPHGPVHLWTGD

the enzyme at sequence level, structural information is required to definitely identify the specific function of enzymes in biological systems. Even though being prokaryotic in nature, bacteria are more closely related to fungus, while fungus and plant groups, which both are although eukaryotic, are distantly related on the basis of the phylogenetic analysis of the PPO sequences. This classification can significantly contribute in the understanding of the evolutionary relations between the species at molecular level. However, owing to the considerable importance of PPO, more contribution is warranted for the detailed investigation of the activity and functional analysis of enzymes.

References

1. Mayer, A. M. (2006). *Phytochemistry, 67*, 2318–2331.
2. Mason, H. (1955). *Adv Enzymol, 16*, 105–184.
3. Umit, U. M. (2007). *Food Chem, 100*, 909–913.
4. Wichers, H. J., Recourt, K., Hendriks, M., Ebbelaar, C. F. M., Biancone, G., & Hoeberichts, F. A. (2003). *Appl Microbiol Biotechnol, 61*, 336–341.
5. Bailey, T. L., & Elkan, C. (1995). *Mach Learn, 21*(1–2), 51–80.
6. Bailey, T. L., & Gribskov, M. (1998). *Bioinformatics, 14*(1), 48–54.
7. Tamura, K., Dudley, J., Nei, M., & Kumar, S. (2007). *Mol Biol Evol, 24*, 1596–1599.
8. Marusek, C. M., Trobaugh, N. M., Flurkey, W. H., & Inlow, J. K. (2006). *J Inorg Biochem, 100*, 108–123.
9. Garcia-Borron, J. C., & Solano, F. (2002). *Pigment Cell Res, 15*, 162–173.
10. van Gelder, C. W. G., Flurkey, W. H., & Wichers, H. J. (1997). *Phytochemistry, 45*, 1309–1323.
11. Gerdemann, C., Eicken, C., & Krebs, B. (2002). *Acc Chem Res, 35*, 183–191.
12. Klabunde, T., Eicken, C., Sacchettini, J. C., & Krebs, B. (1998). *Nat Struct Biol, 5*, 1084–1090.
13. Eickena, C., Krebs, B., & Sacchettini, J. C. (1999). *Curr Opin Struct Biol, 9*, 677–683.