**Endocrine,** *vol. 28, no. 2, 187–192, November 2005* 0969–711X/05/28:187–192/$30.00 

# Co-Evolution Analysis on Endocrine Research

## *A Methodological Approach*

Tonghai Dou, Shuai Chen, Chaoneng Ji, Yi Xie, and Yumin Mao

*State Key Laboratory of Genetic Engineering, Institute of Genetics, School of Life Sciences, Fudan University, Shanghai 200433, People's Republic of China*

**The rapid growth of different kinds of biological information allows a good opportunity to analyze the co-evolutionary characteristics in endocrine regulatory pathways. Data ranging from kinds of species' genome, gene sequence, protein structure, and expression profile of different organisms can reveal the inner co-evolutionary relationship of ligands, receptors, and other related molecules. In return, these co-evolutionary characteristics can help us determine uncharacterized ligands and receptors, annotate gene functions, highlight amino acid residues with biochemical significance, and identify regulated genes in the endocrine process. Encouraging examples in this field, although at their starting stage, have emerged. Here we focus on recent progress in endocrine-related co-evolution research from a methodological approach.**

**Key Words:** Co-evolution; phylogenetic tree comparison; correlated mutation; conserved co-expression.

### Introduction

Co-evolution is a widely existing phenomenon, initially observed between plants and insects *(1)*. This process can be either reciprocal or antagonistic, leading to co-existence and co-variation among interacting species through natural selection. At the molecular level, the phenomenon that some interacting (or related) partners whose sequences share compensatory changes during evolution to maintain function is defined as molecular co-evolution *(2)*. The molecular co-evolution model provides useful assistance for investigating interacting molecules, such as ligand–receptor binding, cell signal transduction, and so on *(3)*. In endocrine research, many molecules are considered to co-evolve. Typical examples come from luteinizing hormone/follicle-stimulating hormone and their corresponding receptors *(4)*. In addition to the different kinds of hormone receptors *(5–7)*, co-evolution is also investigated in other endocrine molecules, such as between two components of the heterodimeric ecdysone receptor *(8)* or between the sterol regulatory element and its binding protein *(9)*.

In the postgenomic era, data on genome sequence, gene expression, and protein structure information increase rapidly, which provides an opportunity for computational prediction to assist laboratory experiments in endocrine research. Such predictions ranging from genomic annotation, detailed sequence analysis to expression and structure data mining, lay a sound foundation for biological experiments. In these kinds of analyses, molecular co-evolutionary characteristics can serve as important clues, including gene co-occurrence, phylogenetic tree similarity, correlated mutation, as well as conserved co-expression. Methods tracing these characteristics of target molecules can reveal their functional relationship or predict amino acid residues with structural significance. Research in such areas, although at its earliest stage, has begun *(10)* and will go further with the accumulation of new data and implementation of improved methods. In this review, we briefly summarize and discuss functional analysis based on molecular co-evolution in endocrine research from a methodological perspective.

### Phylogenetic Tree Comparison

DNA and protein sequences increase rapidly and serve as important recourses for sequence-based comparison in co-evolutionary analysis. The information on protein phylogenetic relationship and their sequence divergence can be retrieved via phylogenetic tree (gene tree or gene family tree) established by multiple sequence alignments, such as ClustalW *(11)*. Genes in the same family often cluster together, giving clues to their functional similarities. As to the analysis of target gene, its orthologs and paralogs can be used to construct phylogenetic tree, representing its evolutionary story in multiple species. Gene tree may be different from the species tree built from molecular markers such as 18s rRNA gene. The reasons can be, for example, that gene duplication or varied evolutionary rate in different species would affect the evolutionary distance on the gene tree *(12)*.

Received June 13, 2005; Revised August 31, 2005; Accepted September 6, 2005.

Author to whom all correspondence and reprint requests should be addressed: Yumin Mao, State Key Laboratory of Genetic Engineering, Institute of Genetics, School of Life Sciences, Fudan University, Shanghai 200433, People's Republic of China. E-mail: ymmao@fudan.edu.cn

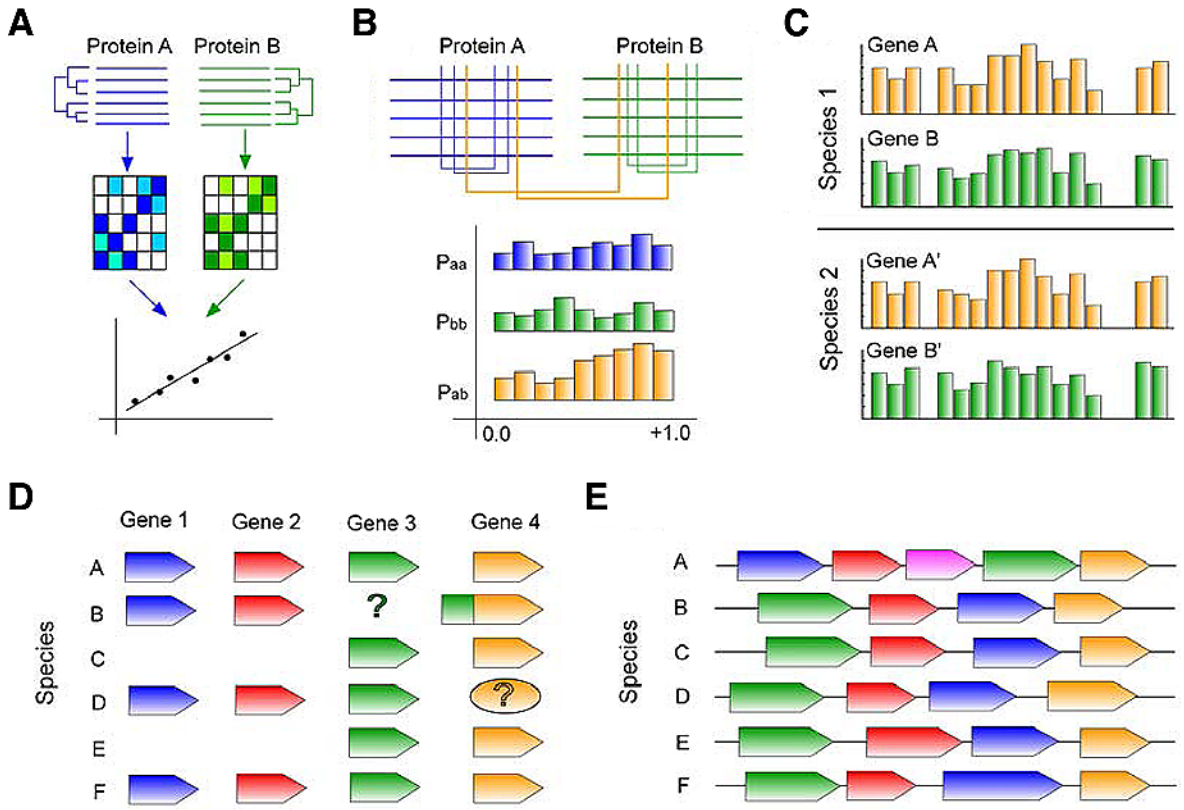


**Fig. 1.** Some typical methods in co-evolution analysis. (**A**) Phylogenetic tree comparison. Phylogenetic trees can be constructed by multiple sequence alignments and put into direct comparison. The results of multiple sequence alignments could also be used to construct the corresponding intersequence distance matrix. Linear correlation between these distance matrices reflects the similarity between phylogenetic trees. Both similarity of phylogenetic trees between proteins A and B and high linear correlation value of their corresponding distance matrices can be used to predict interaction. (**B**) Correlated mutation. Sequences of proteins A and B in different species are selected and put into multiple sequence alignments. Correlation value of intermolecular pair of mutation sites (Pab: one position from Protein A and one from Protein B) and intramolecular pair of mutation sites (Paa and Pbb: pairs of positions within Protein A and within Protein B) can be calculated by different mathematical methods. And pairs with correlation value higher than threshold value are predicted to be correlated mutations. (**C**) Conserved gene co-expression in multiple species. Genes A and B are co-expressed genes in species 1 and their orthologs (genes A' and B') also co-expressed in species 2. Such conserved co-expression indicates that the two genes are functionally related. (**D**) Phylogenetic co-occurrence. Presence and absence for each orthologous genes (gene 1–4) in different species (species A–F) are shown. Pairs of proteins with identical (or similar) co-occurrence are predicted to interact (Gene 1 and Gene 2). The missing of Gene 3 in species B is caused by fusion of this gene to Gene 4 in the same species. The missing of Gene 4 in species D is caused by non-homologous replacement. (**E**) Chromosomal proximity. Orthologs exhibiting chromosomal proximity across multiple organisms may be functionally related. Genes belonging to same orthologous group are showed in the same color.

Co-evolutionary interacting molecules are often under similar evolutionary selective pressure. In some cases, duplication or mutation occurring on one gene may lead to the duplication or mutation of its partner to maintain proper biological function. These events make their gene trees share certain similarity, and specific interacting gene pairs are located at a corresponding branch *(13)*. The analysis of insects' nuclear receptors of ecdysteroid hormone, a heterodimer comprising of ecdysone receptor (ECR) and retinoid X receptor (RXR), shows how the co-evolutionary relationship is identified via phylogenetic tree comparison. Gene trees of ECR and RXR share similarity and differ from the species tree. Combined with coordinate evolutionary rate data of these two molecules, RXR and ECR are suggested to undergo co-evolution during holometabolous insect diversification *(8)*. Similar phylogenetic tree comparison could be performed on the entire signaling pathway, including ligands, receptors, and their downstream signaling molecules *(14,15)*. Moreover, tree similarity could help to find interacting partners of target genes (Fig. 1A). For example, the potential receptor could be predicted if its position in the receptor gene tree is similar to that of the given ligand in the ligand gene tree. In the search for growth differentiation factor-9 (GDF-9) receptor, sequence alignments and phylogenetic comparison revealed that GDF-9 is closer to bone morphogenetic proteins (BMPs) than to other ligands. And the receptor of ligands close to GDF-9 is bone morphogenetic protein receptor type II (BMPRII), indicating GDF-9 may bind to BMPRII. This prediction was later confirmed experimentally *(16)*. In addition, a similar approach also

provides important clues in identifying the interaction relationship between type 2 corticotropin-releasing hormone receptors and stresscopin/stresscopin-related peptides *(10)*.

In spite of its simplicity, direct tree comparison is not quantitative and often requires an evolutionary rate to aid the analysis. One solution is to introduce the statistical method to the comparison. Sequences data, after alignments, are transformed into matrices containing the evolutionary pairwise distance (which is the basis for phylogenetic tree). Therefore, tree comparison is converted to calculating the correlated coefficient of the matrices (Fig. 1A). The higher the coefficient is, the more similar the phylogenetic trees will be. And naturally, a larger probability exists for the co-evolutionary relationship of the two molecules. This statistically more quantitative model is established based on analyzing the correlation of the two domains in phosphoglycerate kinase (PGK) *(17)*. Such a method empowers scientists to estimate the general similarity of the two phylogenetic trees. Later, Goh et al. optimized the method, enabling it to calculate the correlation of a specific protein pair (i.e., the correlation between one branch on a phylogenetic tree and one on another phylogenetic tree). Compared with the original one, this improved method can be applied in predicting protein interaction in a more direct fashion *(18)* and could be used in a more efficient way, such as an online service [http://advice.i2r.a-star.edu.sg *(19)*].

Phylogenetic tree comparison based on molecular co-evolution serves as a convenient tool in predicting functional interaction between endocrine molecules. These methods (whether quantitative or not) require detailed sequence data to perform better multiple sequence alignments. Besides, if trees built on analyzed genes are too similar with their species tree, the correlation of gene trees will be very high and result in false positive prediction, which needs other methods to amend.

## Correlated Mutation and Structural Information

Besides tree comparison, sequence data also contain much detailed co-evolutionary information. Co-evolved interacting proteins often show correlated mutation at important sites. A typical example is the growth hormone (GH) and its receptor (GHR) in higher animals. Human GHR can only interact with GHs of human and some primates, not with non-primate GHs. This is because the primate GHR (containing a Leu→Arg mutation at position 43) cannot attach to the non-primate GH (containing His at position 170) favorably. On the other hand, in primates, a single residue mutation of GH occurred in site 171 (His→Asp) makes the binding more favorable. Such co-variation may contribute to the species specificity. To further prove the existence of correlated mutation, checking of the evolution rate and all mutation sites of GH–GHR molecules in prosimian, simian, and nonprimate species showed that GH and GHR shared a co-evolutionary relationship, and two corresponding mutation sites of His171Asp in GH and Leu43Arg change in GHR were correlated *(20)*.

The correlated mutation can occur at either intermolecular or intramolecular level, such as co-various sites in the RNase *(21)*. These correlated mutation pairs often help shape structurally important sites (e.g., sites for interacting with other molecules), which impose direct and crucial impact on protein function *(22)*. The mechanism may be due to the co-ordinating amino acid substitution in the interface of proteins to maintain proper interaction during evolution, such as hormone-receptor binding selectivity and affinity *(20)*.

Traditionally, the detection for correlated mutation sites can be achieved by large-scale random mutagenesis, together with protein crystal structure data. However, because crystal information provides few details on protein energetics and mutagenesis cannot feasibly identify the functional coupling residues in a complex protein, sequence analysis becomes another useful tool in correlated mutation identification (Fig. 1B). The conventional process starts with collecting sequences of orthologous genes or genes in the same family so as to conduct multiple sequence alignments *(11)*. The information obtained concerning sequence conservation and correlation, together with polarity, is often valuable to analyze protein structure, such as discriminating incorrectly folded proteins from native ones *(23)*. Then, these alignment results are processed via different mathematical methods to search for correlated mutation sites. For example, information of potential sites can be transformed to matrices and tested statistically *(24,25)*; or in the statistically coupled analysis (SCA), scientists identified a network of energetically coupled residues that link the functional surfaces of the unclear receptor ligand binding domains (LBD) *(26)*. In their research, evolution is regarded as a large-scale mutagenesis experiment with selection for function. Rather than conducting a real mutagenesis experiment, this co-evolutionary feature are extracted from statistical analysis of a large and diverse multiple sequence alignment of a protein family by selecting a subset of sequences. In these sequences, a fixed amino acid appears at a specified position and the effect of this statistical perturbation on the amino acid distribution at other sites can be estimated through alignment. By applying the SCA to RXR heterodimer communication, a number of co-various sites are predicted and confirmed to contribute to specificity to ligand response.

Other processing approaches based on different mathematical models are also available. Some related approaches also appear, such as Evolution Trace Method, which partly takes the evolutionary relationship between interacting partners into consideration *(27)*; and the maximum likelihood method, which is capable of identifying coevolving protein residues and their relationship to protein structure *(28)*. All these methods merged would be of much use in areas like pharmacological specificity optimization.

Correlated mutations can be applied to locating structurally important sites for protein–protein interaction from amino

acid sequences. The successful application comes from searching for the important residues in G protein–coupled receptor (GPCR) dimerization *(29,30)*, such as chemokine receptor *(31)* and opioid receptor *(32)*. Aside from receptor dimerization, encouraging examples have demonstrated that correlated mutation also can help analyze interacting sites of hormone and receptor *(26,33)*. Like phylogenetic tree comparison, correlated mutation needs better multiple sequence alignments as well. Furthermore, complicated processing steps, different mathematical models, along with the effect of random mutation during evolution, all affect the credibility of the result. On the other hand, great attention has been paid to link protein sequence with its structure and function in endocrinology research. Variance of amino acid sites with spatial importance would directly affect molecular interaction *(6,34,35)*. Rapidly increased protein crystallography data can help connect essential residues with structural basis of protein function *(36,37)*. As a method to detect structurally important sites in protein sequence, correlated mutation, combined with crystal data, would be more effective.

## Conserved Co-expression and Functional Relationship

Turning related genes on and off at different times and in different organs will lead to the proper endocrine regulation for organisms. Therefore, research on the functional relationship of the genes shows co-expression turned out to be an interesting topic. Usually genes expressed at the same time or with similar spatial pattern could be defined as co-expression. The underlying mechanisms for co-expression remain unclear, but are thought to be related to the co-evolution of promoters and coding sequences *(38)*. Different species have different co-expressed genes. Even for orthologous genes, their expression profiles may vary between species *(39)*. This may be related to many factors, such as different selective pressures, organ divergence, as well as gene duplication. If a group of genes co-expressed in one species, and their orthologs co-expressed in other species, this phenomenon is called "conserved co-expression" and it potentially reflects selection pressure on maintaining functional relationship such as participating in same biological process or existing in same protein complex *(38,40)* (Fig. 1C). Genes that are conserved co-expressed among different species tend to show co-evolution to some extent. For example, codon bias in yeast is associated with co-expression, and is used to analyze co-evolution of interacting proteins *(41)*; comparative study of genome sequences and expression profiles of human and mouse suggest genes with similar co-expression patterns are likely to co-evolve *(42)*.

Conserved co-expression can be applied to building functional links between different transcripts and thus helping annotate gene functions (Fig. 1C). In a study concerning key genes in embryo development, researchers found that the BMPs and Phox2s genes share similar expression patterns in both aves and mammals, suggesting that they belong to the same evolutionarily conserved signaling pathway *(43)*. Recent finding shows a number of genes, which belong to fibroblast growth factor (FGF) signaling, form an evolutionarily conserved co-expression gene group and play an important role during early vertebrate development *(44)*. Similar expression pattern of mitogen-activated protein kinase phosphatase 3 (MKP3) and other conserved co-expression genes of FGF signaling among chicken, zebrafish, and mouse indicates MKP3 are also involved in this conserved functional process *(45)*. With the rapid development of microarray technology, scientists could identify evolutionary conserved co-expressed genes with great efficiency. In a horizontal expression comparison of four eukaryotic organisms' microarray, data show that those conserved co-expressed genes can be divided into 12 functional categories (e.g., cell cycle, proteasome, and secretion related), which makes it easier to determine biological function *(46)*. Comparing human and mouse microarray data could also obtain detailed figures such as annotating orphan GPCR functions *(47)*. With increasing awareness of the importance of conserved co-expression analysis, a database involving multiple species, including human, rodents, and other higher eukaryotes (http://symatlas.gnf.org/SymAtlas/), has been established.

Compared with other methods, co-expression approach has stronger ability to identify co-evolved gene neighbors (upstream or downstream) in the endocrine network at the transcriptional level. A recent study based on the genome sequence and expression data shows that some conserved co-expressed genes have similar upstream transcriptional elements *(48)*. Such finding indicates that some co-expression analysis can be conducted at genomic level. Combination of expression data with genomic research would in turn benefits such application in eukaryotes.

## Other Approach

Besides those methods described above, the genome context method has widely been used to annotate gene functions based on the principle that genomic characteristics can reflect the co-evolution of functionally related genes. These characteristics include genes exhibiting statistically similar distribution patterns in different genomes [co-occurrence *(49)*, Fig. 1D], distinct genes from one organism fused into a single gene in another organism [gene fusion *(50)*, Fig. 1D], and functionally related genes exhibiting chromosomal proximity across multiple organisms [Fig. 1E *(51)*]. Although the applications are mainly focused on lower animals *(52,53)*, the example of identifying gene function in higher eukaryotes appears *(54,55)*. With more understanding of co-evolution information hidden in eukaryotic genome data and combining other information (such as expres-

sion data), genome context methods may become another approach in endocrine co-evolutionary research.

## Summary and Future Perspective

At the molecular level, evolutionary changes and polymorphisms are mainly due to mutations that are nearly enough neutral with respect to natural selection that their behavior and fate are mainly determined by mutation and random drift *(56)*. A certain number of genes would have experienced continuous adaptation to each other over time for the improvement or maintenance of optimal physiological function (molecular co-evolution), which differ from normal molecular evolution *(2,3)*. This co-evolution can be reflected in different aspects (e.g., at sequence or expression level) and investigated via various kinds of approaches as discussed above.

On the other hand, we should also be aware that typical co-evolution of proteins is not guaranteed to exist in all endocrine pathways *(57)*. Furthermore, in those situations where co-evolution analysis is applicable, scientists should take as many factors as possible into consideration, so as to render reliable co-evolutionary characteristics. These factors include loss or duplication of related genes in some species, functional divergence of orthologs, interspecific difference in evolutionary rate, and diversity of gene number in a superfamily. For example, after considering the information about gene duplication, function divergence, phylogenetic tree comparison, and experimental results from other researches, Hsu formulated a hypothesis based on co-evolution that orthologs of LGR7 and LGR8 (two leucine-rich repeat-containing GPCRs) have interactive relation with relaxin family peptides in different vertebrates *(58)*.

In return, those reliable co-evolutionary characteristics are powerful tools in determining ligand–receptor interaction, annotating functions of orphan ligands/receptors, figuring out important residues with biochemical significance, and identifying new genes with endocrine importance in regulatory pathways. Currently, genome projects on a number of eukaryotes have been completed, and genome sequencing of many mammals are underway. With the accumulation of various kinds of information, co-evolution analysis will greatly facilitate our understanding of the signaling that co-ordinate and control the functions of multiple organs and processes.

## Acknowledgments

We thank Jiaxi Xu for great help in the preparation and discussion. Dr. Zhiying Li also provided a critically reading for the article.

## References

1. Ehrlich, P. R. and Raven, P. H. (1964). *Evolution* **18,** 586–608.
2. Dover, G. A. and Flavell, R. B. (1984). *Cell* **38,** 622–623.
3. Dover, G. A. (1992). *Bioessays* **14,** 281–287.
4. Moyle, W. R., Campbell, R. K., Myers, R. V., Bernard, M. P., Han, Y., and Wang, X. (1994). *Nature* **368,** 251–255.
5. van Kesteren, R. E., Tensen, C. P., Smit, A. B., et al. (1996). *J. Biol. Chem.* **271,** 3619–3626.
6. Farid, N. R. and Szkudlinski, M. W. (2004). *Endocrinology* **145,** 4048–4057.
7. Park, Y., Kim, Y. J., and Adams, M. E. (2002). *Proc. Natl. Acad. Sci. USA* **99,** 11423–11428.
8. Bonneton, F., Zelus, D., Iwema, T., Robinson-Rechavi, M., and Laudet, V. (2003). *Mol. Biol. Evol.* **20,** 541–553.
9. Athanikar, J. N. and Osborne, T. F. (1998). *Proc. Natl. Acad. Sci. USA* **95,** 4935–4940.
10. Hsu, S. Y. and Hsueh, A. J. (2001). *Nat. Med.* **7,** 605–611.
11. Chenna, R., Sugawara, H., Koike, T., et al. (2003). *Nucleic Acids Res.* **31,** 3497–3500.
12. Huynen, M. A. and Bork, P. (1998). *Proc. Natl. Acad. Sci. USA* **95,** 5849–5856.
13. Fryxell, K. J. (1996). *Trends Genet.* **12,** 364–369.
14. Newfeld, S. J., Wisotzkey, R. G., and Kumar, S. (1999). *Genetics* **152,** 783–795.
15. Herpin, A., Lelong, C., and Favrel, P. (2004). *Dev. Comp. Immunol.* **28,** 461–485.
16. Vitt, U. A., Mazerbourg, S., Klein, C., and Hsueh, A. J. (2002). *Biol. Reprod.* **67,** 473–480.
17. Goh, C. S., Bogan, A. A., Joachimiak, M., Walther, D., and Cohen, F. E. (2000). *J. Mol. Biol.* **299,** 283–293.
18. Goh, C. S. and Cohen, F. E. (2002). *J. Mol. Biol.* **324,** 177–192.
19. Tan, S. H., Zhang, Z., and Ng, S. K. (2004). *Nucleic Acids Res.* **32,** W69–72.
20. Liu, J. C., Makova, K. D., Adkins, R. M., Gibson, S., and Li, W. H. (2001). *Mol. Biol. Evol.* **18,** 945–953.
21. Zhang, J. and Rosenberg, H. F. (2002). *Proc. Natl. Acad. Sci. USA* **99,** 5486–5491.
22. Gobel, U., Sander, C., Schneider, R., and Valencia, A. (1994). *Proteins* **18,** 309–317.
23. Olmea, O., Rost, B., and Valencia, A. (1999). *J. Mol. Biol.* **293,** 1221–1239.
24. Pazos, F. and Valencia, A. (2002). *Proteins* **47,** 219–227.
25. Pazos, F., Helmer-Citterich, M., Ausiello, G., and Valencia, A. (1997). *J. Mol. Biol.* **271,** 511–523.
26. Shulman, A. I., Larson, C., Mangelsdorf, D. J., and Ranganathan, R. (2004). *Cell* **116,** 417–429.
27. Lichtarge, O., Bourne, H. R., and Cohen, F. E. (1996). *J. Mol. Biol.* **257,** 342–358.
28. Pollock, D. D., Taylor, W. R., and Goldman, N. (1999). *J. Mol. Biol.* **287,** 187–198.
29. Dean, M. K., Higgs, C., Smith, R. E., et al. (2001). *J. Med. Chem.* **44,** 4595–4614.
30. Gouldson, P. R., Dean, M. K., Snell, C. R., Bywater, R. P., Gkoutos, G., and Reynolds, C. A. (2001). *Protein Eng.* **14,** 759–767.
31. Hernanz-Falcon, P., Rodriguez-Frade, J. M., Serrano, A., et al. (2004). *Nat. Immunol.* **5,** 216–2123.
32. Filizola, M. and Weinstein, H. (2002). *Biopolymers* **66,** 317–325.
33. Horn, F., Bywater, R., Krause, G., et al. (1998). *Receptors Channels* **5,** 305–314.
34. Umesono, K. and Evans, R. M. (1989). *Cell* **57,** 1139–1146.
35. Li, J. H., Choe, H., Wang, A. F., et al. (2005). *Mol. Pharmacol.* **67,** 1099–1110.
36. Clayton, G. M., Peak-Chew, S. Y., Evans, R. M., and Schwabe, J. W. (2001). *Proc. Natl. Acad. Sci. USA* **98,** 1549–1554.
37. Fan, Q. R. and Hendrickson, W. A. (2005). *Nature* **433,** 269–277.
38. Niehrs, C. and Pollet, N. (1999). *Nature* **402,** 483–487.
39. Huminiecki, L. and Wolfe, K. H. (2004). *Genome Res.* **14,** 1870–1879.
40. van Noort, V., Snel, B., and Huynen, M. A. (2003). *Trends Genet.* **19,** 238–242.

41. Fraser, H. B., Hirsh, A. E., Wall, D. P., and Eisen, M. B. (2004). *Proc. Natl. Acad. Sci. USA* **101,** 9033–9038.
42. Jordan, I. K., Marino-Ramirez, L., Wolf, Y. I., and Koonin, E. V. (2004). *Mol. Biol. Evol.* **21,** 2058–2070.
43. Ernsberger, U. (2000). *Eur. J. Biochem.* **267,** 6976–6981.
44. Bottcher, R. T. and Niehrs, C. (2005). *Endocr. Rev.* **26,** 63–77.
45. Kawakami, Y., Rodriguez-Leon, J., Koth, C. M., et al. (2003). *Nat. Cell Biol.* **5,** 513–519.
46. Stuart, J. M., Segal, E., Koller, D., and Kim, S. K. (2003). *Science* **302,** 249–255.
47. Su, A. I., Cooke, M. P., Ching, K. A., et al. (2002). *Proc. Natl. Acad. Sci. USA* **99,** 4465–4470.
48. Karaulanov, E., Knochel, W., and Niehrs, C. (2004). *EMBO J.* **23,** 844–856.
49. Pellegrini, M., Marcotte, E. M., Thompson, M. J., Eisenberg, D., and Yeates, T. O. (1999). *Proc. Natl. Acad. Sci. USA* **96,** 4285–4288.
50. Marcotte, E. M., Pellegrini, M., Ng, H. L., Rice, D. W., Yeates, T. O., and Eisenberg, D. (1999). *Science* **285,** 751–753.
51. Yanai, I., Mellor, J. C., and DeLisi, C. (2002). *Trends Genet.* **18,** 176–179.
52. Marcotte, E. M., Pellegrini, M., Thompson, M. J., Yeates, T. O., and Eisenberg, D. (1999). *Nature* **402,** 83–86.
53. Bowers, P. M., Pellegrini, M., Thompson, M. J., Yeates, T. O., and Eisenberg, D. (2004). *Genome Biol.* **5,** R35.
54. Huynen, M. A., Snel, B., Bork, P., and Gibson, T. J. (2001). *Hum. Mol. Genet.* **10,** 2463–2468.
55. Muhlenhoff, U., Richhardt, N., Ristow, M., Kispal, G., and Lill, R. (2002). *Hum. Mol. Genet.* **11,** 2025–2036.
56. Kimura, M. (1983). *The neutral theory of molecular evolution.* Cambridge University Press: Cambridge.
57. Escriva, H., Delaunay, F., and Laudet, V. (2000). *Bioessays* **22,** 717–727.
58. Hsu, S. Y. (2003). *Trends Endocrinol. Metab.* **14,** 303–309.