Die Kurzzeit-Stipendien des Boehringer Ingelheim Fonds sollen es Doktoranden und Postdoktoranden ermöglichen, in einem Zeitraum von höchstens drei Monaten an anerkannten Forschungseinrichtungen des In- und Auslandes klar definierte Methoden zu erlernen, die das eigene Forschungsvorhaben erweitern und der Arbeitsgruppe des Stipendiaten unmittelbar nutzen. Ferner wird der Besuch von wissenschaftlichen Lehrgängen und Ferienkursen gefordert, bei denen eine begrenzte Zahl von Teilnehmern ein fest umrissenes Thema aus der Grundlagenforschung durch praktische Übungen, Vortrage und Diskussionen erschließt.

Neue Methoden

Monitoring the Evolution of Protein Modules by Sequence Analysis

Dr. rer. nat. Peer Bork
Diplom-Biochemiker
European Molecular
Biology Laboratory (EMBL)
Heidelberg

California is not only famous for sunshine, wonderful beaches, and superb wines, but it is also a renowned centre for scientific activities. Thus it was clear from the beginning that a two-months stay at the University of California in San Diego would be an enriching experience, rewarding in research. In the laboratory of Russell F. Doolittle, one of the pioneers in analyzing molecular sequence data, the main focus is to trace the course of domains in protein evolution.

The challenge of monitoring molecular evolution by sequence analysis becomes feasible with the rapidly increasing amount of available sequence data stored in public databases. The most frequent event leading to new proteins during evolution is gene duplication and subsequent modification. This modification, i.e. adaptation of a given three-dimensional protein structure to a new function by subsequent point mutations, takes a long time, and is a rather inefficient process.

Therefore, whole structurally independent protein parts (domains) with a defined subfunction have often been used in evolution to speed up the trial and error procedures for finding new functions. The use of domains as modules (which are independent of the domain order within proteins and might be inserted into most diverse pathways) increases with the complexity of the respective organism. Indeed, the most heavily modular systems known so far are the large extracellular proteins of animals.

However, the mechanism behind the frequent exchange of protein modules still remains to be detected. Although exon shuffling is a reasonable explanation and has been proved in several cases, modules also occur in prokaryotes that lack introns. Furthermore, horizontal gene transfer of modules has been observed not only among prokaryotes but also between eukaryotes and prokaryotes (Bork and Doolittle, 1992).

To understand the evolution of mobile modules and thus to shed some light on the distribution mechanism we have collected various known modules and studied their distribution among the different phyla (Doolittle and Bork, 1993). Of particular interest are domains which occur in both eukaryates and prokaryotes.

In comparing some case studies, e.g. fibronectin type I11 modules (Bork md Doolittle, 1992, 1993), domains of galactose oxidase (Bork and Doolittle, submitted) or ankyrin-like repeats (see Figure; Bork, submitted; Bork and Sander, submitted) it was exciting to follow the trail of how established protein folds can radiate through the biological world, involving themselves with other domains as circumstances allow.

For example, ankyrin-like (ANK) repeats are well-studied, because they occur in »trendy« proteins such as transcription activator NF-χB or several cell-cycle proteins (e.g. *notch, SW14).*

Nevertheless, their structure, function and evolution remain rather unclear. When we made a comparative survey and searched sequence databases for these repeats, more than 650 of them were identified. Detailed analysis revealed, that in at least some cases ANK repeats are able to jump from one protein to another (Bork, submitted). However, the genetic mechanism behind still has to be encoded.

The study of module evolution required the use of various methods such as homology searches, pattern matching approaches, estimation of mutation rates and the construction of phylogenetic trees. All of these methods require a deep understanding of the algorithms behind them, in order to realize the limits and to choose the right parameters. From this point of view I gained a lot of experience in estimating the powers and pitfalls of these methods.

In addition, the stay did not only prove useful because of publishable

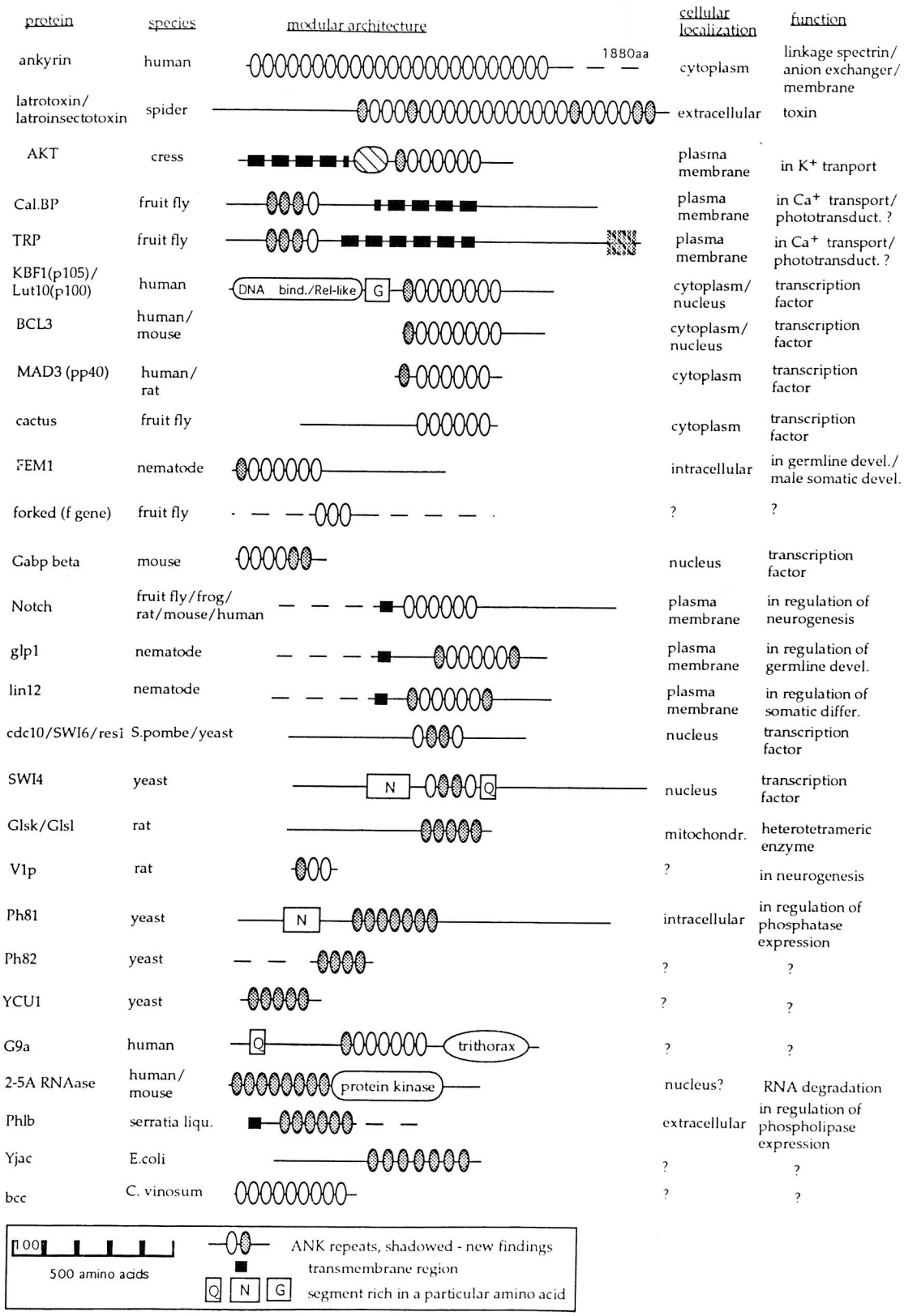

Fig. 1a: Distribution and evolution of the ankyrin-like (ANK) repeat which is named after its abundance in ankyrin. The ANK module which always occurs in at least four repeats, is one example where horizontal transfers might be involved in spreading the repeat into the most diverse proteins. About 400 repeats have been detected in current sequence databases by using different homology search methods. Nearly 100 of them have not yet been reported.
Localization of ANK repeats in selected proteins. ANK repeats have been found in both intracellular and extracellular proteins, in viruses, prokaryotes, fungi, plants and animals, the location of the repeats in the respective proteins is always different. The only common function appears to be protein-protein interactions. How does evolution manage such a spreading, if exon shuffling can be excluded (prokaryotes do not have introns)?

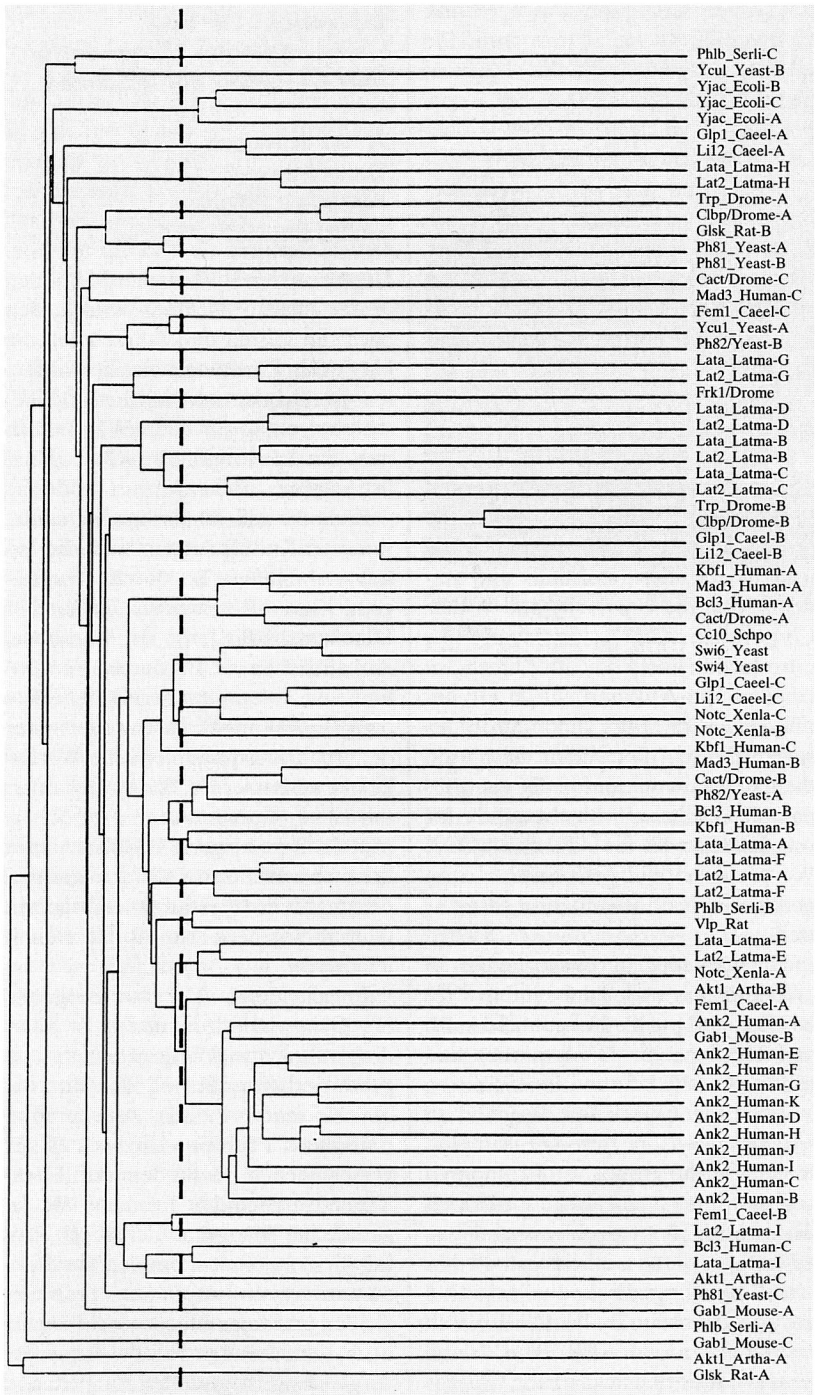

Fig. 1b: Simplified dendrogram of selected ANK repeats. The grouping by sequence similarity seems to be not in accordance with taxonomic data. Is the ANK repeat able to cross phyla by horizontal gene transfer? Note that the prokaryotic repeats (in Yjac_Ecoli) cluster together suggesting a rather recent divergence. It is known, that this protein belongs to a group which has been aquired by horizontal gene transfer. Has a single repeat been aquired by *E. coli* which then has been duplicated several times?

results; I will also remember the many scientific contacts, the stimulation for future work, the spirit of a successful laboratory, and last but not least some sunburns which remind me of fruitful scientific discussions under the Californian sun.

References

Bork, P., Doolittle, R. F. (1992) Proc. Natl. Acad. Sci. USA *89*, 8990–8994

Bork, P., Doolittle, R. F. (1993) Protein Science *2* (in press)

Doolittle, R. F., Bork, P. (1993) Scientific American (in press)