Darwin Year 2009

Protein evolution in Tübingen

A Matter of Folding

The huge variety of proteins are built of independent domains with surprising similarities in sequence and structure – about 80% of them adopt one of only about 400 protein folds. Andrei Lupas and his team are therefore trying to find evidence for the hypothesis that contemporary proteins evoluted out of a limited number of ancient peptides.

Nuts, eggs, milk and meat. They all belong to a healthy diet as they all provide us with an important nutrient: protein. Their jobs in the organism are manifold: they work as enzymes, hormones and antibodies; they are necessary for structure, growth, signal integration - in short, for all vital functions of the living cell. Proteins are, without a doubt, life's essential building block.

"Indeed, life can be viewed as substantially resulting from the chemical activity of proteins," agrees Andrei Lupas of the Max-Planck-Institute (MPI) for Developmental Biology in Tübingen, Germany. There, at the Department of Protein Evolution, the biologist works on the question of how those important components of life have developed over time.

De facto, everything in living cells centres on proteins and their production, he thinks. "The genome stores the information for how to make proteins, the mRNA transfers it to the ribosomes, which are formed by rRNA and small RNAs control the gene activity… In the end, the output is always protein!"

Not only Andrei Lupas likes to compare protein evolution with linguistic evolution (taken from Axe D.D. *et al.*, *PLoS ONE 3*(6): e2246)

In fact, life without proteins is hardly conceivable. This, however, holds a dilemma, as proteins are only produced in living systems. So, today's biological dogma implies that life originates from life. "Nevertheless, there must have been one point in history, when, in the very beginning, life has arisen from inanimate material," says Andrei Lupas. How could life have been formed in the primordial soup? And how could proteins have evolved? These are the crucial questions that Andrei Lupas and his team are focusing on.

As a precursor of our recent life, a so-called "RNA world", which existed about 3.5 billion years ago is proposed, when RNA stored the genetic information and ribozymes catalysed reactions. "At this time, simple peptides, which firstly formed spontaneously, functioned as co-factors," Lupas assumes and argues that, "in contrast to RNA, peptides are able to catalyse redox reactions. They can eliminate water from RNA helix grooves and they assist in RNA folding processes." The production of amino acids and peptides from abiotic matters has been demonstrated experimentally. "The RNA might have recruited them."

However, peptides must have soon run out as the abiotic genesis could not keep up with their production. Andrei Lupas is therefore convinced, "The selection pressure on the RNA-based organisms to produce peptides themselves must have been high. Thus, peptide synthases as precursors of ribosomes must have existed.

Of course, the peptides would have been useful, so special peptides must have been produced selectively. "The RNA organisms must have established a code – maybe in the beginning a quite primitive one, but over time a more and more exact one using codons – to control which amino acid should be integrated into the peptide chain." As soon as those "mini-genes" have been placed consecutively, longer peptide sequences – and thus proteins – could have evolved.

Models from linguistics

However, as plausible as this story sounds, there is one snag: how could the 3D structure of proteins have developed, which is of such an amazing variety and absolutely essential for their effectiveness? "Protein folding is complicated. Although the information somehow lies in their amino acid sequence, proteins are not able to do the folding themselves correctly," Andrei Lupas says. Many know about this problem from biochemical assays: proteins often coagulate, unwind during heating and clump after cooling. "If you heat an egg you get omelet, which after cooling down is just cold omelet, not the egg again," the Romanian researcher illustrates, laughing.

"RNA was shown to induce a secondary structure in peptides," Lupas continues. Due to water exclusion in correct peptide-RNA fits, hydrogen bonds can be linked in the peptide chain to form α-helices or β-sheets.

"This means the ancestral peptides could have been selected for their ability to cooperate with the RNA in forming a local structure," thinks Lupas. "As emergent property, longer protein chains might have been able to fold together by collapsing without RNA support, as parts of them had been pre-optimised to do so." For him, this must have been the origin of protein folding.

Although contemporary proteins are tremendously varied, the vast majority of them are built of independent folding units – domains – with surprising similarities in sequence and structure. About 80% of domains adopt one of only about 400 protein folds. This, according to Andrei Lupas and colleagues, supports the hypothesis of protein evolution out of ancient peptides. Accordingly, they are trying to find more evolutionary traces of this process in contemporary proteins by comparing their sequences, structures and functions using bioinformatics, biochemistry and structural biology.

An analogy Lupas likes to use to illustrate his research subject is linguistics. The comparative study of modern languages has led to the reconstruction of ancient vocabularies. "Today it is known that the Semitic *qnw* (cana) meaning *reed* is the ancestor of a lot of words in very different languages, from the Greek and Latin *kanna* or *canna* respectively, to the French *canne* and the English *cane*," the biologist describes. "There is also paralogous offspring: taking the straight, rigid attribute of reed, the Greek created levelling rules calling them *kanon*, the progenitor for the church rule or the musical form *canon*. Due to the round and hollow property, *can* or *canister* as well as *canal*, *canyon*, *canon*, etc. have developed from the plant...Coming back to biology, we do a sort of protein linguistics trying to reconstruct the ancestral peptides out of amino acid sequences of recent proteins," Lupas explains. The bioinformatical method his team uses, the so called "hidden markov model" which reveals patterns in character chains, also comes from linguistics. Based on this, Andrei Lupas and his ex-colleague Johannes Söding developed a tool capable of discovering remotely homologous sequences, from which they aimed to compile a dictionary of ancestral peptides.

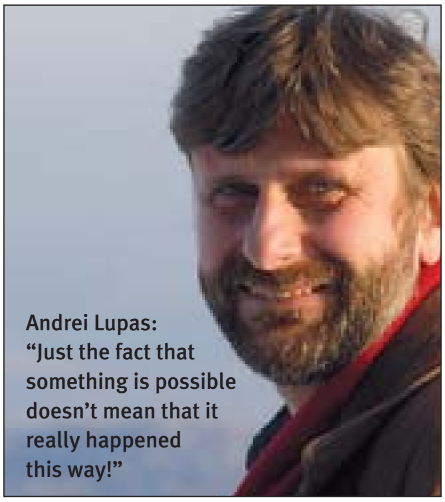
An approach, however, that doesn't protect them from the central problem which evolutionists face, "We have no proofs for our ideas, as we can't travel back in time!" Additionally, several intermediate precursors are extinct today. So the scientists are searching for very good examples that will make their proposed evolutionary lines back to the first ancestors as plausible as possible.

Pure *Kon-tiki* experiments?

"Nevertheless, we have to be cautious not to perform a *Kon-tiki* experiment à la Thor Heyerdahl," Lupas points out. "Just the fact that something is possible doesn't mean that it really happened this way!" Thus, collecting as much data as possible and examining it from different perspectives is necessary to get an overall picture of the story.

So far, the scientists have identified a vocabulary of about 50 short, possibly ancestral, peptidic fragments. "These are the most frequent motifs among all recent proteins." Via fusion and recombination of those supersecondary structures, containing different arrangements of α-helices and β-sheets, protein domains might have arisen.

In protein families with similar sequences, the relationship is evident. However, for Andrei Lupas and Co. less obviously-related structures are much more interesting. "The double-psi barrel, for example, you find in several protein superfamilies. But their similarity is not visible at first glance!"



Andrei Lupas: "Just the fact that something is possible doesn't mean that it really happened this way!"

Lupas took a closer look, studying the class of AAA ATPases ("ATPases associated with various cellular activities"). He analysed the VatN-domain, the substrate recognition domain of the archeal protein VAT playing an important role in protein degradation (*Structure* vol. 14:1489-98). "The first part VatN-N forms a very complex, knotted structure consisting of two homologous βαββ elements." To understand the background of the unusual knots, Lupas searched the database for related sequences and found that of the transcription factor AbrB. "Through the topological operation of circular permutation, which happens quite often in protein evolution, the structure could have been formed out of two non-geared halves."

When glancing at the sequence, Lupas's theory was striking. However, "the structure of AbrB, published a little later, was completely different to what I expected. That was really shocking!" remembers Lupas. After having analysed intermediate, more original sequences and identified how "strand swap" came to VatN-N instead of circular permutation, he re-crystallised the AbrB structure. And, lo and behold, "the published one was wrong! In our new structure it was visible, that 'strand displacement' led to AbrB." Lupas thus identified very old folds appearing in ribosomal proteins, translation factors and riboflavin kinases. "Now, we can postulate a single strand βαββ element as common ancestor of the VatN-N, AbrB and the intermediates."

Cluster maps instead of trees

In the near future, the scientists want to focus on function development. "We try to evolve the next more complicated property in a given protein *in vitro* by modifying it according to what its close relatives tell us."

In contrast to other evolutionists, Andrei Lupas creates networks rather than phylogenies of related proteins. "I prefer cluster maps, as they are more powerful in comparing more than 1,000 data points, intermediates are easier to see, and they are less sensitive for false alignments." The maps result from so called force-directed graph placement, which the Tübingen group has modified for sequence comparison. "The proteins are set casually in space and connected depending on the force representing their similarity, for all unconnected points a force of repulsion is set. Then the system is allowed to equilibrate," Lupas explains. In the end, related sequences are located in clusters, which sometimes also assemble in superclusters. "This gives us a clue on the function of newly discovered sequences."

Andrei Lupas regards being a researcher in a Max-Planck-Institute as a blessing, as was his time as a postdoc with Wolfgang Baumeister at the MPI for Biochemistry, Martinsried, Germany, working on the structure and function of the proteasome. After an excursion into industry at SmithKline Beecham Pharmaceuticals in Collegeville, USA, he returned to academic research. "I wanted to do basic evolutionary research on protein folding, and I knew I'd need to combine different techniques from biochemistry, bioinformatics, microbiology and structure biology to do so." This is much easier in an environment like the Max Planck Society than in a company, he concludes.

This, however, is not the only plus factor. Of particular advantage to the father-of-six is the opportunity to live on the MPI campus in Tübingen. "I am at home in five minutes to have lunch or dinner with my family and I can go back when the kids are in bed." A benefit Andrei Lupas appreciates very much. "Having an intact family and sometimes the chance to take the mind off from work stuff makes life much more pleasant!"

Susanne Dorn