Pergamon

SCIENCE DIRECT®

Bioorganic & Medicinal Chemistry 11 (2003) 3045–3051

BIOORGANIC & MEDICINAL CHEMISTRY

Perspective

At the Crossroads of Chemistry and Biology

H. Waldmann[a,b,*]

[a]*Max-Planck-Institut für molekulare Physiologie, Abt. Chemische Biologie, Otto-Hahn-Str. 11, D-44227 Dortmund, Germany*
[b]*Fachbereich 3, Organische Chemie, Universität Dortmund, Germany*

Received 27 January 2003; accepted 31 January 2003

Abstract—The life sciences are molecular and the harnessing of information gleaned from genomics and proteomics will require interdisciplinary research integrating chemistry and biology. This approach is illustrated by the synthesis and biological evaluation of lipidated peptides and proteins and the delineation of a concept arguing for natural product guided combinatorial chemistry.

Introduction—The Chemical Nature of Biology

The life sciences are undergoing dramatic changes in these times. Biology is in the process of deciphering the genetic codes of various organisms—most notably of man and species that are widely used in laboratory experiments like the mouse. This revolutionary development has a deep impact on its sister sciences, physics and, in particular, chemistry.

The various genomics and proteomics projects carried out worldwide are yielding a wealth of information, data bits provided in a language that we only partly understand. The task lying before us is to group the letters making up the genome sequences into words, sentences, paragraphs and chapters ultimately assembling the book of life. Taking a closer look at the prerequirements to be fulfilled for the successful execution of this process reveals that the knowledge alone of which genes are available and potentially can be expressed (genomics) is not sufficient. Also, it will not suffice to determine which proteins actually are present in a living cell under given conditions (proteomics). Ultimately, the fate of a cell will be determined by the interactions of proteins with each other, with other biomacromolecules, with supramolecular structures like membranes and the cytoskeleton, and with small molecules binding to the gene products and modulating their activity (Scheme 1).

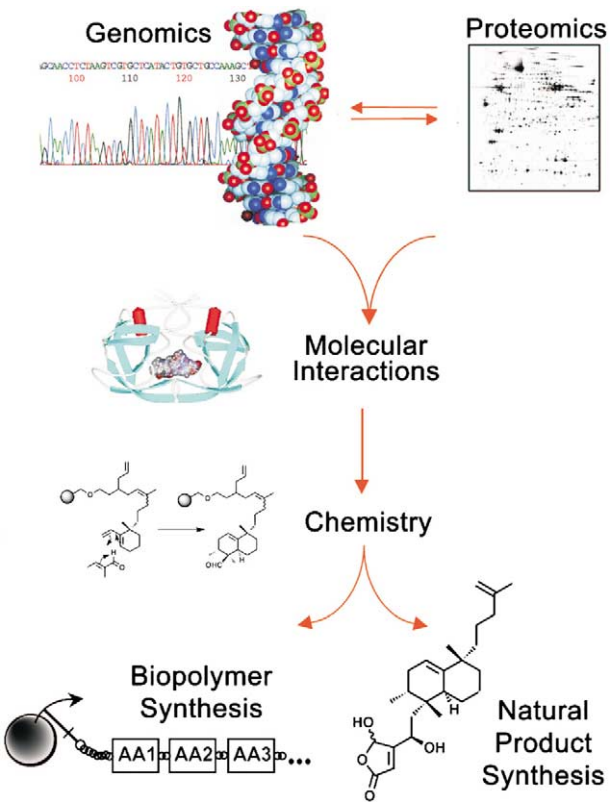
Molecular interactions and chemical transformations are at the heart of biology, and all biological phenomena that we can analyse today can be traced back to chemical processes: biology is molecular.

This insight, on the one hand, makes clear that chemistry may turn out to be the central science in the quest for understanding the molecular basis of life. This is so because the study of interactions between molecules, be they small or large, and the methodology to prepare them and to prepare new compounds with a predetermined set of properties are at the heart of chemistry and constitute the key expertise of chemists. On the other hand, it demands from chemistry to devote a major part of its undisputed powers to addressing the problems unravelled by the research in the biological sciences. Undoubtedly, much has been achieved to meet this demand, and it is fair to say that in principle each biologically relevant small molecule can be synthesized by means of the techniques and methods developed in the last century (although the effort to achieve this still may be very high and clearly the demand for more efficient methodology rather increases because of the increasing demands and complexity posed by biology and other sciences). But at the same time, this insight points to new frontiers arising at the crossroads between chemistry and biology. The research carried out by my group during the last decade has addressed two of these frontiers (Scheme 1).

The synthesis of fully functional proteins carrying different post-translational modifications that can be modified at will by the power of organic synthesis and

*Fax: +49-231-133-2499; e-mail: herbert.waldmann@mpi-dortmund.mpg.de

0968-0896/03/$ - see front matter
doi:10.1016/S0968-0896(03)00266-9



Scheme 1. The need for integration of chemistry and biology.

embodying additional reporter-groups by which they can be traced in biological systems poses one such frontier. Imagine one would require a lipo-glyco-phosphoprotein carrying a fluorescent label or a photoactivatable group for a biological study. Such a molecular tool currently cannot be prepared. Our contributions to that field are highlighted below by a summary of the synthesis and biological evaluation of lipid-modified peptides and proteins.

A second frontier is provided by the fact that for small molecule synthesis in a biological context the crucial question no longer is *how* a given compound can be made. Rather the problem is *which* is the right, that is, biologically relevant compound to be synthesized with the required properties for a subsequent biological investigation. We have tried to give an answer to this question by combining knowledge gleaned from bioinformatics, structural biology, natural product and combinatorial chemistry resulting in a concept for the development of natural-product guided combinatorial chemistry.

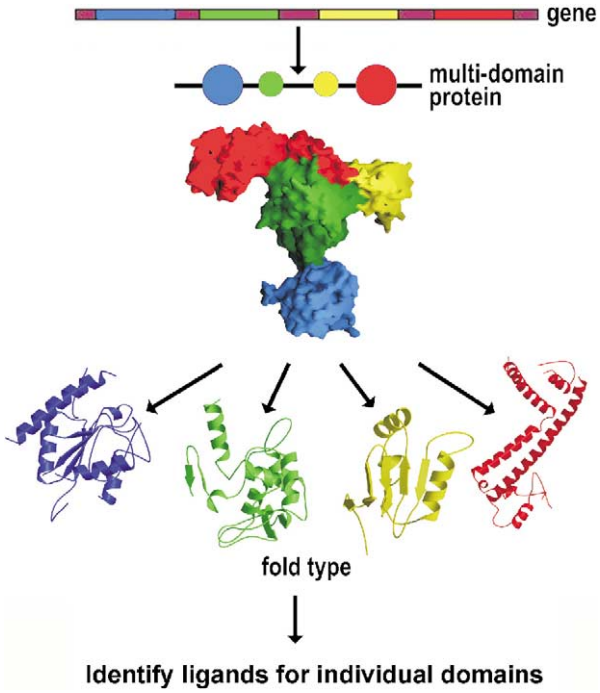
Protein Domain and Natural Product Structure–Guiding Principles for Combinatorial Chemistry

Biological investigation of the million compound speculative combinatorial libraries of the first generation yielded disappointingly low hit rates. The conclusion from the first phase of combinatorial chemistry was that for biological use not the size of a library is important. Rather, the underlying structure of the library members must be biologically relevant. Importantly, chemical space is so vast that it cannot be comprehensively covered by synthesizing all possible molecules. Thus, a key to success is to identify already biologically pre-validated structural 'islands' in chemical space and to develop the chemistry required to reach them.

Proteins can be regarded as modularly built biomolecules assembled from individual domains as building blocks. Since the total number of all protein domains appears to be fairly limited it has to be expected that in newly discovered proteins with widely varying function and activity the same modules (i.e., domains) or close relatives will be found repeatedly in varying combinations and arrangements as structure- and function-determining entities (Scheme 2).

Thus, a key to the efficient discovery of new ligands and inhibitors for known and, in particular, for newly discovered proteins is to identify compound classes already biologically validated as being capable of binding to specific protein domains. These compound classes can then be employed as starting points in structural space for library development.[1]

Libraries designed and synthesized around the basic structure of such compounds should yield better—because biologically relevant—modulators of protein activity with high hit rates. Furthermore, they should yield modulators of activity for proteins with differing activity, function and origin which bear the same or very similar domains. Biologically active natural products can be regarded as chemical entities that were evolutionarily selected and validated for binding to particular protein domains. Therefore, they are already biologically validated, and the underlying structural architectures of such natural products may provide



Scheme 2. Domain architecture of proteins.

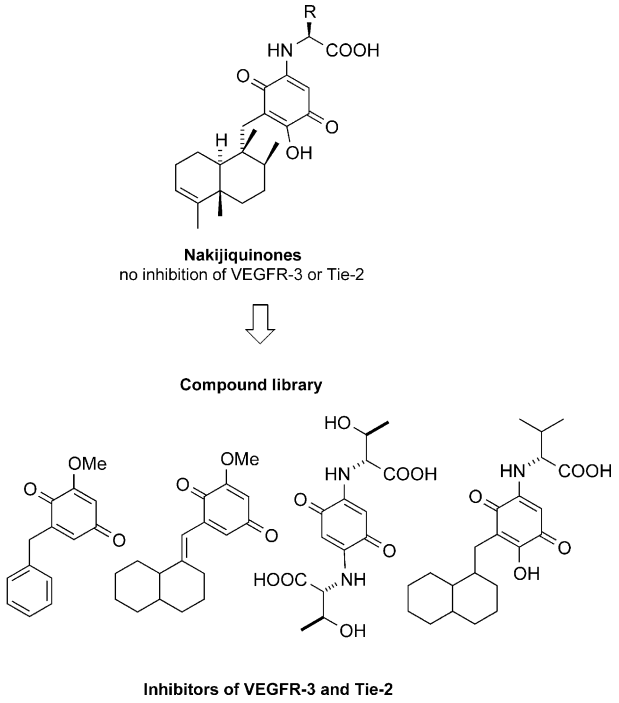
powerful guiding principles for library development (this prerequisite is of course also fulfilled by non-natural compound classes for which biological pre-validation has been established by experience, e.g., in therapy).

For instance, based on the synthesis of Nakijiquinone C a library of 76 analogues was synthesized and investigated for activity against a panel of selected protein kinases selected because ATP-binding domains are very similar. The screen revealed a subset of compounds that inhibit the tyrosine kinase Tie-2 and also yielded a compound that modulates the activity of the VEGF-3 receptor (Scheme 3).[2]

Both proteins are decisive regulators of angiogenesis, the development of new blood vessels from pre-existing ones. Inhibitors of angiogenesis are currently being investigated as a new type of anti-cancer agents because they starve the tumour by shutting down the blood supply.

Efficient and reliable methods and multistep sequences for the total synthesis of natural products and their analogues on polymeric supports are paramount to the success of this approach. The corresponding transformations must proceed with a degree of selectivity and robustness typical of related classical transformations in solution, irrespective of the stringencies and differing demands imposed by the presence of and the anchoring to the polymeric support.

As target compound of particular interest for the establishment of natural product derived library synthesis on solid support Dysidiolide was chosen, an inhibitor of the cell-cycle controlling protein phosphatase Cdc25.



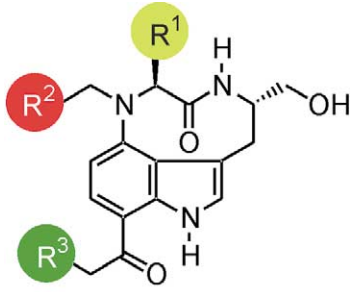
Scheme 3. Identification of Tie-2 and VEGFR-3 receptors from a natural product-derived compound library.

We developed a solid-phase synthesis of epi-dysidiolide and analogues thereof. The syntheses proceeded in >10 steps on the solid phase and included various different transformations including an asymmetric Diels–Alder reaction, oxidation with singlet oxygen and olefin metathesis (Scheme 4). The dysidiolide analogues were obtained in high overall yields.[3]

The synthesis sequence established in the context of this endeavour is among the most advanced and demanding solid-phase syntheses developed so far. It also demonstrates that the total synthesis of complex natural products and analogues thereof in multi-step sequences on the solid phase is feasible.

Biological investigation of the dysidiolide analogues yielded selective Cdc25 inhibitors with high potency in enzymatic and cellular assays.

The protein kinase C activator Indolactom V served as the guiding structure for the solid-phase synthesis of a compound library **1** that was investigated for isoenzyme selective PKC activation. In a cellular screen a selective activator of PKC δ was identified.[4]



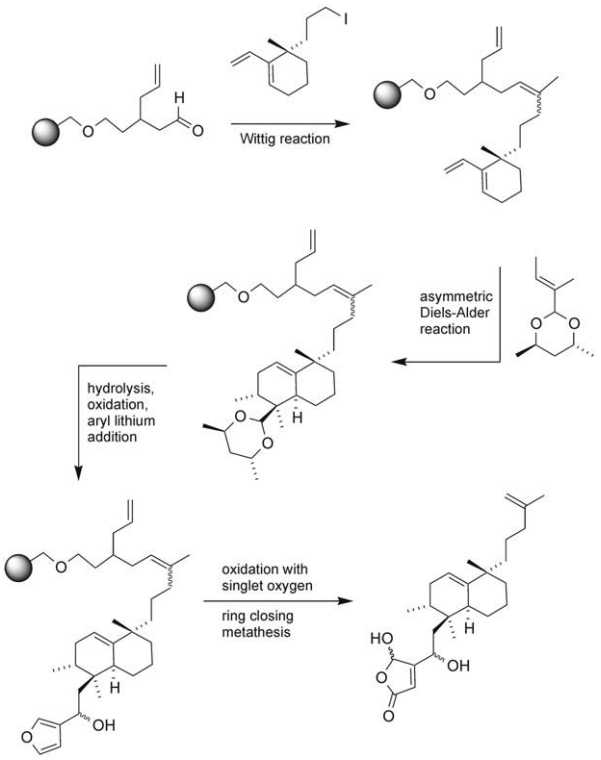
1

In currently ongoing work, we aim at the synthesis and biological investigation of further natural product derived compound libraries based on structures widely occurring in Nature.

Development of Solid-Phase Synthesis Methods: Traceless Linkers

The development of complex multi-step sequences for combinatorial synthesis requires the continuous introduction of new and better synthesis techniques. Vital to all solid phase methodologies is the utilization of suitable anchor groups which allow the facile attachment, functionalization and release of the molecules of interest. We have developed an enzyme-labile safety-catch linker which releases the desired products via cyclization after enzymatic hydrolysis of an amide.[5] Furthermore we have introduced a traceless linker based on aryl hydrazides, which can be cleaved under mild oxidative conditions.

The traceless aryl hydrazide linker is compatible with various different important transformations of organic synthesis, for example, Stille, Heck, Sonogashira,



Scheme 4. Solid-phase synthesis of epi-dysidiolide.

Suzuki, Wittig and Grignard reactions and reductive aminations and was successfully applied in the solid-phase synthesis of a new antibiotic.[6]

In particular, its use in a 10-step traceless solid-phase synthesis of 2-aminothiazoles employing the Hantzsch thiazole synthesis as key step gave rise to a compound library that contains strong dual-specific inhibitors of the VEGFR-2 and Tie-2 receptor tyrosine kinases.[7] Both kinases are key regulators of angiogenesis, the development of new blood vessels from pre-existing ones (see above).

Synthesis and Biological Evaluation of Lipidated Peptides and Proteins

Lipid-modified proteins play important roles in numerous biological processes like signal transduction and vesicular trafficking. In order to study these roles in precise molecular detail, methods for the synthesis of differently lipidated peptides and proteins were developed which may also carry further reporter groups and labels.[8]

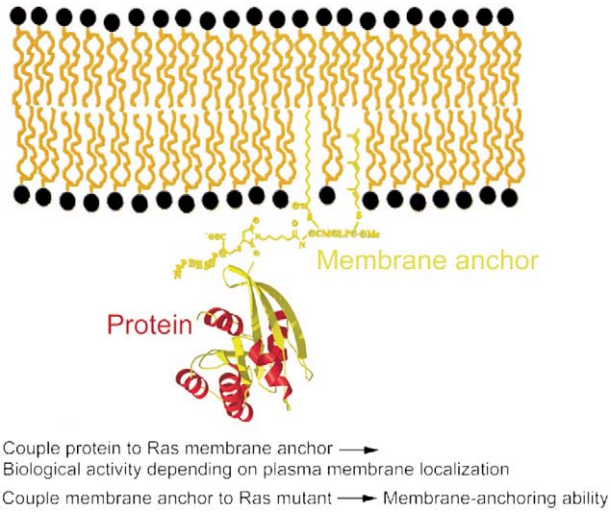
Ras Peptides and Ras Proteins

The Ras-proteins are post-translationally farnesylated and palmitoylated and play a decisive role in signal transduction. Mutations of Ras are found in approximately 30% of all human cancers. With the help of conjugates which link differently lipidated peptides to a truncated oncogenic Ras protein the examination of numerous differently lipid-modified proteins became possible. Accordingly, peptides embodying the natural lipid parts of the Ras proteins, non-natural analogues thereof as well as fluorescence-labelled prenylated and/or palmitoylated Ras-peptides were equipped with a reactive linker group such as the maleimido group, and coupled to C-terminally truncated human Ras-proteins. The Ras-hybrid proteins were employed as tools for membrane binding experiments, for example, via surface plasmon resonance. Microinjection into PC12 cells revealed that the biological activity of Ras is not dependent upon a farnesylated residue and identified palmitoylation as a key regulatory step in the selective targeting of Ras to the plasma membrane. Thus, the hybrid proteins can serve as efficient tools for biochemical, biophysical and biological experiments. These investigations resulted in the development of a new readout system that allows to correlate structural information with biological activity at the plasma membrane (Scheme 5).[9]

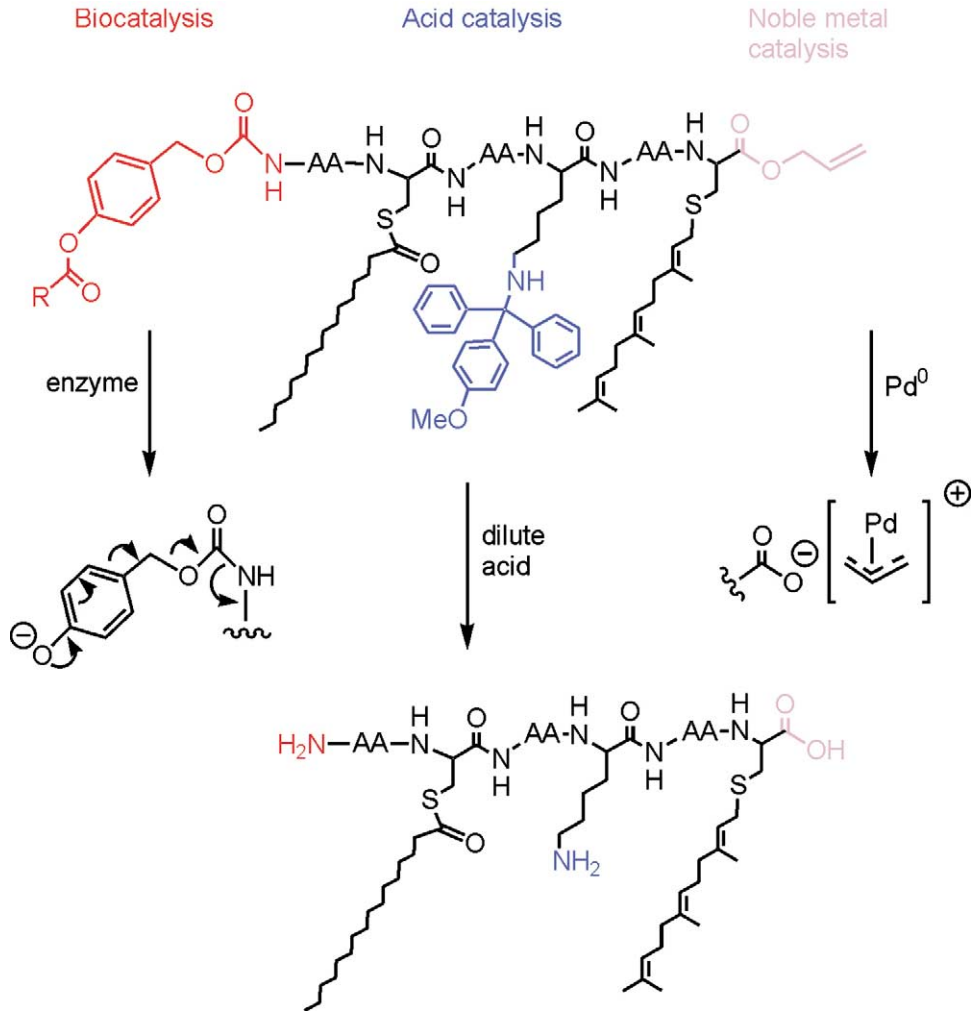
In this joint project with the groups of Kuhlmann and Wittinghofer, techniques of organic synthesis, biophysics and cell biology were successfully combined to yield new insight into a biological problem with a bioorganic approach.

In particular, this approach required the development of a set of new protecting group techniques that allow for selective deprotection under very mild conditions (i.e., pH 6–8, room temperature). The use of biocatalysis and noble metal catalysis proved to be key techniques to reach this goal, resulting, for example, in the introduction of a set of new enzyme-labile urethane and ester protecting groups (Scheme 6).

Furthermore, we have recently developed the first generally applicable method for the solid-phase synthesis of lipidated and additionally labelled peptides.[10] In this work, knowledge gleaned from the development of new linker groups for solid-phase synthesis was successfully transferred. Thus, application of the traceless hydrazide linker to solve the problems posed by lipopeptide



Scheme 5. A general biological readout system developed via a chemical–biological approach.



Scheme 6. Protecting group strategies for the synthesis of lipidated peptides.

chemistry allowed for the efficient synthesis of Ras and Rab peptides (Scheme 7).

Further research is directed towards in vivo studies to localize Ras binding partners in the plasma membrane, for example via introduction of photoactivatable groups into the lipid residues of the semisynthetic Ras proteins.[11]

Rab Peptides and Proteins

The Rab proteins are a sub-family of the ras-like GTPase superfamily with more than 60 members. Rab GTPases control a broad array of membrane docking and fusion events operating in processes ranging from fertilization to synaptic transmission. For their function, Rab proteins require double modification with geranylgeranyl isoprenoids, which allows them to reversibly associate with membranes in a tightly controlled fashion.

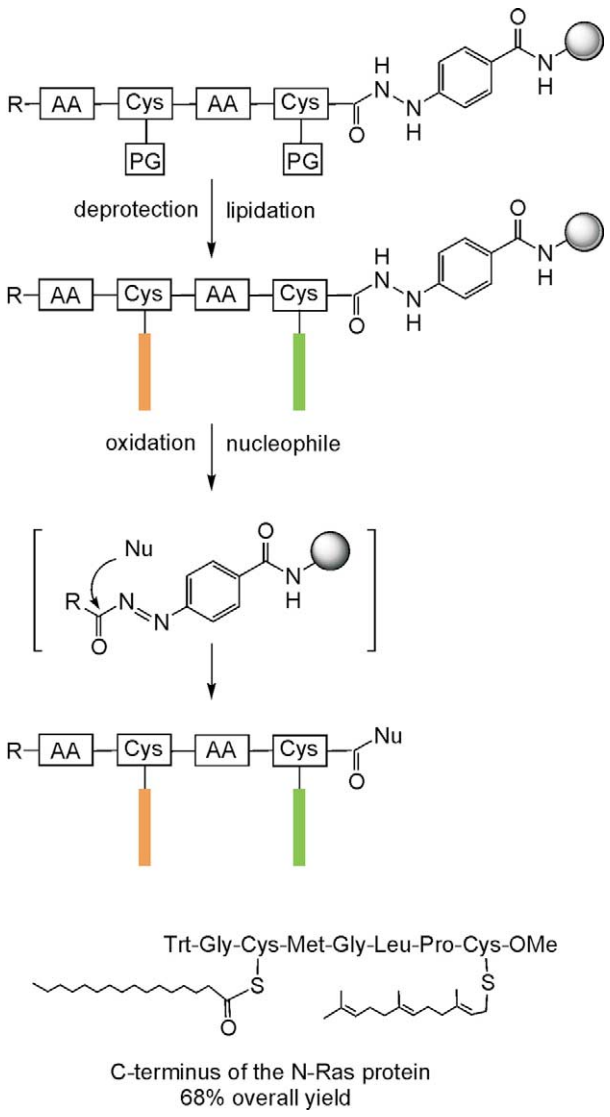
In order to study the biochemical properties and the biological functions of the Rab proteins and for the study of protein complexes Rab proteins participate in, methods were developed that provide prenylated Rab proteins with new functionalities such as fluorescence, photoreactivity or isoprenoid groups at non-native positions. The recently developed in vitro protein ligation method provided the necessary platform for combining large recombinant protein scaffolds with peptides generated by organic synthesis.

Thus, in a joint project with Alexandrov and Goody, Rab peptides were synthesized by means of the preparative methods developed in the context of the Ras projects detailed above. These peptides were equipped with an N-terminal cysteine and then coupled to a truncated Rab protein by means of the 'expressed protein ligation' technique which gives access to Rab proteins activated as C-terminal thioesters (Scheme 8).[12]

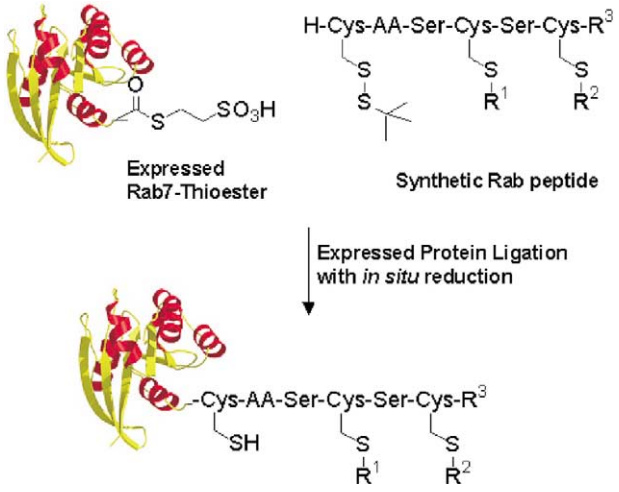
Furthermore fluorescent Rab proteins were synthesized by means of Rab geranylgeranyltransferase-catalyzed attachment of fluorescent GerGer analogues to Rab 7.[13] The synthetic proteins were then employed in biochemical and structural biological studies.[14]

Outlook–New Guiding Principles and Tools for Chemical Genomics and Chemical Proteomics

The application of chemistry to study biology in the sense described in the introductory paragraph and in



Scheme 7. Solid-phase synthesis of lipid-modified proteins.



Scheme 8. Synthesis of Rab proteins by expressed protein ligation.

related approaches is being summarized by the term 'Chemical Biology'. It seems clear that research integrating both chemistry and biology is required in this field. This holds true in a general sense and also in sub-fields of chemical biology that are rapidly emerging and expanding. If chemistry can hold a central position in Chemical Biology in general, likewise it will be able to capture a comparable position in the sub-fields.

For instance, in chemistry-driven proteomics and genomics approaches highly selective or protein-class specific low-molecular weight modulators of protein function will be invaluable tools.

In general, chemical genomics can be regarded as the genomic response (i.e., of all gene products) to chemical compounds. A more practical definition is the identification of small molecule leads for a member of a gene family product to elucidate the function of other members of the gene family.

The protein-domain centered approach for compound-library development highlighted above may provide a new guiding principle for the development of compounds that will pave the way to a new series of proteomics and genomics experiments. In this context a family of gene products (proteins) of interest would be defined in structural terms, that is, harbouring the same or similar domains. This reasoning would be an alternative to the currently applied approach that focuses on protein function and activity, that is on kinases, proteases and so on.

Furthermore, the knowledge gleaned from the proteomics and genomics initiatives will have to be substantiated by various follow-up experiments focussing on the function of the gene products in a biochemical sense. The full value of genomics and proteomics research will only become apparent if it is complemented by protein biochemistry. For such follow-up research differently modified proteins which carry modifications whose structure can be changed at will through synthesis will be invaluable tools. They will allow to set up experiments yielding answers in precise molecular detail hardly accessible with biological techniques alone.

Acknowledgements

I am grateful to my coworkers. With great enthusiasm, motivation, creativity and hard work, have they put our ideas into reality. Their names can be found in the publications that arose from our research. I am also grateful to my cooperation partners. In particular, Jürgen Kuhlmann, Fred Wittinghofer, Kirill Alexandrov and Roger Goody at the Max-Planck-Institute in Dortmund command the biological expertise that complements my own abilities in an excellent way. Our research was supported by several funding agencies and companies to whom I am grateful, in particular the Deutsche Forschungsgemeinschaft, The Max-Planck-Gesellschaft, the Fonds der Deutschen Chemischen Industrie, the Humboldt Foundation, the Ministerium für Wissenschaft und Forschung der Bundesrepublik Deutschland, the European Community, BASF AG, Bayer AG, Roche Bioscience, Degussa AG and AstraZeneca.

References and Notes

1. Breinbauer, R.; Vetter, I.; Waldmann, H. *Angew. Chem., Int. Ed.* **2002**, *41*, 2878.
2. (a) Stahl, P.; Kissau, L.; Mazitschek, R.; Giannis, A.; Waldmann, H. *Angew. Chem., Int. Ed.* **2002**, *41*, 1174. (b) Kissau, L., Stahl, P.; Mazitschek, R.; Giannis, A.; Waldmann, H. Submitted for publication.
3. (a) Brohm, D.; Metzger, S.; Bhargava, A.; Müller, O.; Lieb, F.; Waldmann, H. *Angew. Chem., Int. Ed.* **2002**, *41*, 307. (b) Brohm, D.; Philippe, N.; Metzger, S.; Bhargava, A.; Müller, O.; Lieb, F.; Waldmann, H. *J. Am. Chem. Soc.* **2002**, *124*, 13171.
4. Meseguer, B.; Alonso-Diaz, D.; Griebenow, N.; Herget, T.; Waldmann, H. *Chem. Eur. J.* **2000**, *6*, 3943.
5. Stieber, F.; Grether, U.; Waldmann, H. *Angew. Chem., Int. Ed.* **1999**, *38*, 1073.
6. Grether, U.; Waldmann, H. *Angew. Chem., Int. Ed.* **2000**, *39*, 1629.
7. Stieber, F.; Mazitschek, R.; Soric, N.; Giannis, A.; Waldmann, H. *Angew. Chem., Int. Ed.* **2002**, *41*, 4757.
8. Kadereit, D.; Kuhlmann, J.; Waldmann, H. *ChemBioChem.* **2000**, *1*, 144.
9. (a) Bader, B.; Kuhn, K.; Owen, D. J.; Waldmann, H.; Wittinghofer, A.; Kuhlmann, J. *Nature* **2000**, *403*, 223. (b) Kuhn, K.; Owen, D. J.; Bader, B.; Wittinghofer, A.; Kuhlmann, J.; Waldmann, H. *J. Am. Chem. Soc.* **2001**, *123*, 1023.
10. Ludolph, B.; Eisele, F.; Waldmann, H. *J. Am. Chem. Soc.* **2002**, *124*, 5954.
11. Kuhlmann, J.; Tebbe, A.; Wagner, M.; Uwai, K.; Völkert, M.; Waldmann, H. *Angew. Chem., Int. Ed.* **2002**, *41*, 2546.
12. Alexandrov, K.; Heinemann, I.; Goody, R.; Durek, T.; Waldmann, H. *J. Am. Chem. Soc.* **2002**, *124*, 5648.
13. Owen, D. J.; Alexandrov, K.; Rostkova, E.; Scheidig, A. J.; Goody, R.; Waldmann, H. *Angew. Chem., Int. Ed.* **1999**, *38*, 509.
14. Pylypenko, O.; Rak, A.; Reents, R.; Niculae, A.; Thomä, N.H.; Waldmann, H.; Schlichting, I.; Goody, R.; Alexandrov, K. *Mol. Cell* **2003**, *11*, 483–494.