Principles of intermolecular interactions studied by NMR

HABILITATIONSCHRIFT

Dem Fachbereich Chemie der Universität Hannover vorgelegt von

TERESA CARLOMAGNO

Hannover 2006

Index

Part 2:

A novel pharmacophore for peptide toxins interacting with K^+ channels.

Part 3:

TAR-RNA recognition by positively charged ligands

Thanks to …

This work has been conducted at the Max Planck Institute for Biophysical Chemistry in Göttingen, in the Department of NMR based Structural Biology. I thank the undergraduate and graduate students and the post-doctoral fellows of my group for their interest and passion in the projects carried out in the past four years and for the pleasant working environment.

I thank Prof. Dr. Christian Griesinger, the director of the department, for providing an excellent infrastructure for structural biology and for many scientific collaborations and stimulating discussions.

I thank Prof. K.-H Altmann at the ETH in Zürich, Dr. Marcel J.J. Blommers at Novartis in Basel and Prof. J. Meiler at the University of Vanderbilt, Tennessee, for the collaboration on the tubulin/epothilone project, Dr. H. Terlau at the University of Lübeck for the collaboration on the conotoxin project, Prof. A. Kirschning at the University of Hanover for the collaboration on the TAR/aminoglycoside analogue project and Prof. M. Hennig at the Scripps Research Institute in La Jolla, California for the collaboration on the synthesis of deuterated RNA.

I specially thank Prof. Andreas Kirschning for his continuous support during these past four years and for allowing me to lecture at the University of Hanover.

A particular thank goes to my family, who has believed in me and my work throughout my education, my post-doctoral time and my habilitation.

Finally, I thank Markus for his patience during my long working hours and for his help to go through the disappointment phases of the research work.

List of publications during the "Habilitation"

1. T. Carlomagno*, V. M. Sanchez, M.J.J. Blommers. C. Griesinger

"Derivation of Dihedral Angles from CH-CH Dipolar – Dipolar Cross-Correlated Relaxation Rates:A C-C Torsion Involving a Quaternary Carbon Atom in Epothilone A Bound to Tubulin"

Angewandte Chemie **2003** 42, 2515-2517.

2. T. Carlomagno*, M.J.J. Blommers, J. Meiler, W. Jahnke, T. Schupp, F. Petersen, D. Schinzer, K.-H. Altmann, C. Griesinger

"The High-Resolution Solution Structure of Epothilone A Bound to Tubulin: An Understanding of the Structure-Activity Relationship for a Powerful Class of Anti-Tumor Agents"

Angewandte Chemie **2003** *42*, 2511-2515

3. T. Carlomagno*, W. Bermel, C. Griesinger

"Measuring the χ_1 Torsion Angle in Protein by CH-CH Cross-Correlated Relaxation: A New Resolution-Optimised Experiment"

Journal of Biomolecular NMR **2003** *27*, 151-157

4. M. Kipping, H. Lilie, U. Lindenstrauß, J.R. Andreesen, C. Griesinger, **T. Carlomagno***, T. Brüser

"Structural Studies on a Twin Arginine Signal Sequence"

FEBS Letters **2003** *550*, 18-22

5. A. Al-Sabi, D. Lennartz, M. Ferber, J. Gulyas, J. E.F. Rivier, B. M. Olivera, **T. Carlomagno***, H. Terlau

"μM-Conotoxin RIIIK, Structural and Functional Novelty in a K^+ Channel Antagonist"

Biochemistry **2004** *43*, 8625-8635

6. A. Razeto, V. Ramakrishnan, C.M. Litterst, K. Giller, C. Griesinger, **T. Carlomagno**, N. Lakomek, T. Heimburg, M. Lodrini, E. Pfitzner, S. Becker

"Structure of the NcoA-1/SRC-1 PAS-B Domain Bound to the LXXLL Motif of the STAT6 Transactivation Domain**.**

Journal of Molecular Biology **2004** *336,* 319-329

7. M.Hennig, J. Fohrer, **T. Carlomagno***

"Assignment and NOE Analysis of the 2'-Hydroxyl Protons in RNA: Implications for Stabilization of RNA A-form Duplexes."

Journal of the American Chemical Society **2005** *127,* 2028-2029

8. L. Verdier, A. Al-Sabi, J.E.F Rivier, B.M. Olivera, H. Terlau, **T. Carlomagno***

"Identification of a Novel Pharmacophore for Peptide Toxins Interacting with K+ channels"

Journal of Biological Chemistry **2005** *280,* 21246-21255

9. V.M. Sanchez-Pedregal, M. Reese, J. Meiler, M.J.J. Blommers, C. Griesinger, **T. Carlomagno***

"The INPHARMA Method: Protein Mediated Interligand NOEs for Pharmacophore Mapping"

Angewandte Chemie **2005** *44*, 4172-4175

10. T. Carlomagno*

"Ligand-Target Interactions: What Can We Learn from NMR?"

Annual Reviews of Biophysics and Biomolecular Structure **2005** *34*, 246-266

11. T. Carlomagno , C. Griesinger

„Transferred Methods for the Measurement of Angle Restraints from Scalar and Dipolar Couplings as well as from Cross Correlated Relaxation: Application to Drug/Target Complexes"

Handbook of Modern Magnetic Resonance: Pharmaceutical Sciences, **2005**

12. N.A. Lakomek, C. Farès, S. Becker, **T. Carlomagno**, J. Meiler, C. Griesinger

"Side Chain Orientation and Hydrogen Bonding Imprint Slow Motion on the Protein Backbone of Ubiquitin"

Angewandte Chemie **2005** *44,* 7776-7778

13. N. Lakomek, **T. Carlomagno**, S. Becker, C. Griesinger, J. Meiler

"A Thorough Dynamic Interpretation of Residual Dipolar Couplings in Ubiquitin"

Journal of Biomolecular NMR **2006** *34,* 101-115

14. J. Fohrer, M.Hennig, **T. Carlomagno***

"The Influence of the 2'-Hydroxyl Group in RNA on the Stability of the A-form Helix "

Journal of Molecular Biology **2006** *356,* 280-287

15. M. Reese, V.M. Sanchez-Pedregal, K. Kubicek, J. Meiler, M.J.J. Blommers, C. Griesinger, **T. Carlomagno***

"Structural Basis of the Activity of the Microtubules Stabilizing Agents Epothilones studied by NMR Spectroscopy in Solution."

submitted

16. D. Raghunathan, V.M. Sanchez-Pedregal, J. Junker, C. Schwiegk, M. Kalesse, A. Kirschning, **T. Carlomagno***

" TAR-RNA Recognition by a Novel Cyclic Aminoglycoside Analogue."

submitted

17. C. Farez, T. Carlomagno*

" SHARP-TACSY: Triple-Band Tailored Correlated Spectroscopy for Base-to-Sugar Transfer in Nucleic Acids Residues with Intermediate Timescale Motions."

submitted

18. J. Fohrer, U. Reinscheid, M.Hennig, **T. Carlomagno***

"Dependence of homo- and heteronuclear ^{3}J and ^{2}J scalar couplings on the H2'-C2'-O2'-OH dihedral angle in RNA: a quantitative approach to the determination of the 2'-hydroxyl conformation."

submitted

Introduction

The interaction between molecules is the principle of life. The impressive network of cellular functions is regulated by a fine and complex interplay of several molecules, each of them playing a role at more than one node in the network. Such interactions are possible thanks to the structural plasticity of proteins, nucleic acids and small organic molecules, which can assume a large variety of three dimensional shapes and can present to the external world pockets and surfaces with different electrostatic and hydrophobic properties. Structural biology has been flourishing for the past two decades as the technique of choice to understand the mechanism of molecular recognition. The goal of the massive effort made in the past years in the structural characterization of molecules and complexes is to collect the necessary knowledge to unravel the secrets of molecular recognition. This effort has led to extraordinary successes, as the determination of the ribosome structure (Ramakrishnan, & Moore, 2001). Structural knowledge can be applied to design strategies for interfering with the molecular interaction at the cellular level, thereby influencing the cell life cycle. Besides the early established X-ray crystallography, other techniques are employed in structural biology. Among these, Nuclear Magnetic Resonance (NMR) can deliver structural information at atomic level, while Fluorescence Resonance Energy Transfer (FRET) or the recently blooming Electron Microscopy (EM) are used to obtain structural constraints on large systems at lower resolution.

Solution NMR has experienced great progresses in the past two decades. While the upper limit in the size of the complexes that can be studied by NMR is limited with respect to X-ray crystallography, NMR offers the advantage to study the system in aqueous solution, where the conformational dynamics of the interacting molecules is conserved. Next to solution NMR, solid-state NMR, in the past almost exclusively used in material science, is gaining way in the study of biological systems, especially membrane proteins, thanks to recent methodological developments (Luca *et al.*, 2003).

In this work we study the molecular recognition mechanism in three biological systems via solution NMR: a) the protein tubulin in complex with the

microtubules stabilizing agent epothilone; b) the small peptidic toxin κM-conotoxin RIIIK, derived from the sea snail *Conus radiatus,* in complex with K^+ channels; c) the 30mer HIV TAR-RNA in complex with an aminoglycoside derivative.

A large part of this work is dedicated to understand the mechanism of the conformational control exerted by small organic molecules on the protein tubulin. Several cellular processes, such as cell division, cell motility, cell transport and cell shape see tubulin as a major player (Nogales, 2001). The protein exists in a dynamic equilibrium between a hetero-dimeric form, constituted by α– and β–tubulin subunits, and a polymeric form, termed "microtubules". Microtubules exert their cellular functions by interacting with other proteins, which affects their regulation and distribution within the cell. Any perturbation of the delicate equilibrium between the polymeric and dimeric forms of tubulin results in a strong perturbation of cellular function.

Tubulin is an important target in cancer therapy. Tubulin-binding drugs can dramatically interfere with the tubulin polymerization equilibrium and therefore exhibit profound cytotoxic activity. One of the most important clinical anticancer drugs is the natural product paclitaxel (Taxol®), which inhibits human cancer cell growth through the stabilization of cellular microtubules. Over the last few years a diverse set of other (non-taxane) natural products from a variety of sources have been established to share the taxanes ability to stabilize microtubules (Altmann, 2001), including epothilones A and B, discodermolide, eleutherobin, sarcodictyins A and B, laulimalide, peloruside A (Hood *et al*., 2002), and taccanololides E and A (Tinley *et al*., 2003). Other natural products, as dolastatin (Beckwith *et al.*, 1990), colchicine (Downing & Nogales, 1998), vinblastine (Wilson *et al.*, 1982) and tubulysin (Sasse *et al.*, 2000), have been found to destabilize microtubules, thereby exhibiting cytotoxic activity.

Here, we have studied the basis of the interplay between tubulin and the potent microtubule stabilizing agent epothilone (Höfle *et al*., 1993). The scope of this work is to provide a framework to understand the mechanism of the conformational control exerted by epothilones on the tubulin polymerization equilibrium. NMR is particularly suited for this purpose, as tubulin is a highly dynamic molecule, which so far escaped crystallization in a physiologically

relevant form. The results presented here point to a complex network of binding events and conformational changes, which regulates the tubulin polymerization equilibrium. In contrast to the biological data, which fail to detect the binding of the microtubules stabilizing agents (MSAs) to non-polymerized protein, our data show that the binding of the MSAs, like epothilone and discodermolide, to soluble tubulin is the first event in the process of microtubules stabilization. In this work we have determined the conformation of epothilone bound to non-polymerized tubulin and the binding mode of the drug to the protein. The structural information obtained for the tubulin/epothilone complex allows us to propose a functional mechanism to explain the cytotoxic activity of epothilones. When compared to the electron crystallography (EC) data, obtained on tubulin artificially polymerized on Zn-stabilized sheets, the NMR results indicate that the binding mode of epothilone to tubulin depends on the polymerization state of the protein. This underlines the necessity of studying macromolecular complexes in solution, namely in an environment as close as possible to the physiological one. In order to gain the structural information necessary to define the structure of the epothilone/tubulin complex, we developed new NMR methodologies, which are broadly applicable to other ligand/receptor complexes in the process of structure-based drug design.

The second part of this work focuses on the interaction between the small peptidic toxin κM-conotoxin RIIIK, derived by the sea snail *Conus radiatus,* and K^+ channels. Because of their diverse physiological functions in excitable and nonexcitable cells, ion channels are targets of toxins from venomous organisms, including snakes, scorpions, sea anemones, and spiders. Despite the structural divergence of unrelated K^+ channel targeted toxins, a convergent functional feature has been identified for a variety of toxins, including charybdotoxin from a scorpion, BgK from a sea anemone, and dendrotoxin from snakes. Although for the different toxins more than one site has been identified as being important for binding (Miller, 1995; Eriksson & Roux, 2002; Gao & Garcia, 2003), all these peptides share a dyad motif of a lysine and a hydrophobic, usually aromatic, residue, which plays a key role in the interaction with the target K^+ channel (Stampe *et al.*, 1994; Dauplais *et al.*, 1997; Savarin *et al.*, 1998; Gilquin *et al.*, 2002; Srinivasan *et al.*, 2002). In particular, there is evidence that the lysine

residue of this functional dyad occludes the K^+ channel pore (Miller, 1995; Eriksson & Roux, 2002; Gao & Garcia, 2003).

Peptidic neurotoxins from the venomous cone snails ("conotoxins") are well-known, highly subtype-selective ligands that interact with a variety of different voltage-gated and ligand-gated ion channel targets. Here, we have conducted an extensive interdisciplinary study to characterize the interaction of a novel *Conus* peptide ligand, κM-conotoxin RIIIK (κM-RIIIK), with K^+ channels. By means of a combined approach, including functional studies of the point mutated peptide, the solution structure of κM-RIIIK determined by NMR, mutant cycle analysis and docking calculations, we identified a novel pharmacophore for the blockade of voltage-dependent K^+ channels by κM-RIIIK, which does not include a functional dyad. The binding mode of κM-RIIIK to K^+ channels underlines the importance of basic residues in peptide toxins and suggests a new role for the hydroxyprolin residues present in conotoxins. This work provides a framework for a broader understanding of the pharmacological profiles of peptide toxins interacting with voltage- and ligand-gated ion channels.

The third part of this work is dedicated to the trans activation responsive element (TAR) of the HIV viral RNA (Frankel, 1992; Jones & Peterlin, 1994). The discovery that RNA can exert several functions in the cell life cycle, as for example transcriptional control and catalysis, has triggered a wealth of biochemical and structural research around RNAs, including the targeting of RNA in the drug discovery process. In the human immunodeficiency virus (HIV), the interaction of the transcriptional activator protein, Tat, with TAR (Dingwall *et al*., 1989; Churcher *et al*., 1993; Weeks *et al*., 1990; Weeks *et al*., 1991) together with associated cellular factors (Karn, 1999) is essential for the production of full-length viral transcripts and, therefore the proliferation of the virus (Frankel, 1992). This fact suggests that a compound that binds to TAR-RNA, and prevents binding by Tat, could disrupt processive transcription and thereby inhibit viral growth. Moreover, a number of recent studies using peptidic and/or antisense agents have proved that TAR-binding compounds can show anti-viral activity in HeLa cells, apparently via inhibition of transcription (Arzumanov *et al.*, 2001; Mischiati *et al.*, 2001; Kaushik *et al.*, 2002). The increasing success of protein structure-based drug design, together with recent progress in RNA structure determination,

reinforces the imperative to utilize high-resolution structure determination as a tool facilitating the process of RNA-targeted drug development. In this work, we study the interaction of the HIV-2 TAR-RNA with a cyclic neo aminoglycoside obtained via a "total synthesis" approach. It is well established that small positively charged molecules, like aminoglycosides, interact with various RNA targets, including the TAR-RNA (Mei *et al.*, 1995 and 1998, Wang *et al.*, 1998; Hamy *et al.*, 1998). The toxicity of natural aminoglycosides and the increasing development of antibiotic resistance underline the continuous need for new compounds that share an aminoglycosidic scaffold but can be easily derivatized to obtain drug leads with higher target specificity, less toxicity and activity on resistant viral or bacterial strains. The neo aminoglycoside used in this study can be obtained via a flexible and efficient total synthesis approach, which allows introducing different functional groups on the sugar moieties and ^{15}N or ^{13}C labeling for NMR studies. The structural studies on the complex of the HIV-2 TAR-RNA with this neo aminoglycoside reveal a novel binding mode of the aminoglycosidic scaffold to the TAR-RNA, which includes both bulge and loop nucleotides of the RNA. The binding of a positively charged ligand to the loop of the TAR-RNA had not been observed before and is particularly significant. In fact both the bulge and the loop nucleotides of the TAR-RNA are involved in contacts with the protein Tat and other associated cellular factors, such as the transcription elongation factor P-TEFb, which are essential for viral replication (Dingwall *et al.*, 1990; Richter *et al*., 2002). The P-TEFb complex binds to the loop of the TAR-RNA in a Tat dependent manner. However, the mechanism by which the Tat protein induces the binding of the P-TEFb complex to the TAR loop is not known. In our structural study we observe conspicuous structural rearrangements induced in the RNA hexanucleotide loop by the binding of the neo aminoglycoside. Such structuring of the TAR-RNA loop, following the binding of a positively charged ligand, allows us to propose a mechanism, by which the positively charged, arginine rich region of Tat protein structurally prepares the RNA loop for the binding of the P-TEFb complex. The structural information obtained in this study is leading the design of higher affinity binders, based on a similar aminoglycosidic scaffold and containing functional groups that are optimized for

TAR binding, which are expected to inhibit the binding of TAR to both the Tat protein and the P-TEFb complex.

This work is organized in three sections, each dedicated to one of the three biological systems. All sections contain an introduction to the biological system, high-lightening the significance of the structural work, followed by a brief outline of the questions we wish to answer. The second chapter describes the employed methodology. The results of the structural/functional studies and a critical discussion of the outcome is presented in the following two chapters. In two of the three projects, new NMR methodology was developed to answer unsolved structural questions. Here a special chapter (Chapter 3 of Part 1 and Chapter 5 of Part 3) is dedicated to the development of the novel NMR techniques.

The first project, dealing with the conformation control exerted by epothilones on tubulin, is fruit of a collaboration with the Novartis Institutes for BioMedical Research and Prof. Altmann at the ETH, Zürich, who provided the epothilone samples and a wealth of expertise on epothilone structure-activity relationship data. The interaction of the peptidic toxin κM-RIIIK with the K^+ channel was studied in collaboration with Dr. Terlau at the University of Lübeck, who conducted the functional studies and provided a rich expertise in the ion-channel field. The project on the TAR-RNA – aminoglycoside complex is the result of a productive collaboration with Prof. Kirschning at the University of Hanover, who provided the aminoglycoside analogue sample and continuously supported me in the course of my habilitation. All three projects greatly benefited from the interaction of my structural biology group with organic chemistry and functional biology groups, in a synergic effort aimed to use structural information as a tool to understand molecular function.

Part 1

Conformational control of tubulin by microtubules stabilizing agents: a detailed look into epothilones

Chapter 1

Introduction

1.1 The biological system.

The tubulin. Tubulin exerts a fundamental role in several cellular processes, such as cell division, cell motility, cell transport and cell shape (Nogales, 2001). The protein undergoes a dynamic equilibrium between a hetero-dimeric form, constituted by α– and β–tubulin subunits, and a polymeric form, called "microtubules". Any perturbation of the delicate equilibrium between the polymeric and dimeric forms of tubulin results in a strong perturbation of cellular function. Tubulin can only polymerize when GTP is bound to the E site of the β subunit, and the nucleotide state of tubulin self-regulates polymerization in cells. Further regulation is provided *in vivo* by several stabilizing and destabilizing protein binding partners, such as stathmin, katanin or kin I kinesins.

Tubulin is an important target in cancer therapy. Tubulin-binding drugs can dramatically interfere with the tubulin polymerization equilibrium and therefore exhibit profound cytotoxic activity. Paclitaxel (Taxol®), which inhibits cancer cell growth through the stabilization of cellular microtubules and interference with microtubule dynamics (Jordan, 2002), has become one of the most important clinical anti-cancer drugs for the treatment of metastatic breast and ovarian cancer, Kaposi's sarcoma and non-small-cell lung cancer. However, in spite of their clinical importance, taxanes are associated with severe side effects (Rowinsky, 1997), induce multidrug resistance (MDR), and are ineffective in colon and prostate cancer. The overall profile of the taxanes highlights the tremendous potential that microtubule-stabilizing agents offer, as well as the need to discover agents with fewer side effects and an alternative activity spectrum. Over the last few years several other (non-taxane) natural products from a variety of sources have been proven to share the taxanes ability to stabilize microtubules (Altmann, 2001). Examples of such products are epothilones A and B (EpoA and EpoB), discodermolide, eleutherobin, sarcodictyins A and B, laulimalide,

peloruside A (Hood *et al*., 2002), and taccanololides E and A (Tinley *et al*., 2003). In spite of the structural diversity observed within this group of agents, several of the compounds have been reported to bind to the same binding site on microtubules as paclitaxel (PTX).

In order to learn about the conformational changes that regulate tubulin polymerization, high resolution structural information is of crucial importance. In 1998 the first three-dimensional structure of tubulin, assembled into two-dimensional sheets in the presence of Zn ions and including the bound taxane docetaxel (Taxotere®), was solved by electron crystallography (Nogales *et al.*, 1998). In these sheets, protofilaments appear to be similar to those that form microtubules (Nogales *et al.*, 1999), but they are associated in an antiparallel fashion. This structure has provided a wealth of information on tubulin folding and it has revealed a broad and flexible taxane binding pocket, where numerous ligands with different shapes could be easily accommodated.

The paclitaxel binding site is on the inside surface of the microtubule and close to areas involved in lateral interactions. The lateral contact between protofilaments is characterized by the interaction of loop H1-S2 and helix H3 with the M loop (Fig. 1.1). It has been proposed that the tubulin polymerization equilibrium is controlled by two conformational changes, involving helix H3 and the M loop, respectively. First, helix H3 in the primary sequence is preceded by loop T3, which binds the γ-phosphate of the E-site GTP. The destabilizing effect that GTP hydrolysis has on lateral contacts could be caused by a conformational change in the T3 loop, which is successively transmitted to helix H3. Second, the microtubule stabilizing effect of paclitaxel may be linked to the drug-mediated stabilization of a specific conformation of the M loop, similarly to the stabilizing effect exerted on the M loop by the long S9-S10 loop present in the α subunit. However, recent work of the Knossow and Nogales laboratories (Ravelli *et al.*, 2004; Gigant *et al*., 2005; Wang & Nogales, 2005) suggests that microtubules stabilization occurs via straightening of the curvature at the α and β units interface in protofilaments. Curved protofilaments cannot form appropriate lateral contacts and therefore do not assemble in microtubules; on the other hand straigtht protofilaments can associate laterally and assemble in microtubules. The curvature of the protofilaments is brought about by small rearrangements in the relative

orientation of the three subunits domains (the N-terminal domain nucleotide-binding domain, the intermediate domain and the C-terminal domain) and by local structural differences. In particular a 2.5 Å translation of the helix H7 in the β subunit, which drags along the H6-H7 loop and the helix H6, is the major player in determining the curvature at the interface between heterodimers (Fig. 1.2). The binding of GTP has been proposed to regulate the equilibrium between the straight and bent tubulin conformations by blocking the H7 lever helix in its straight conformation through interaction with the N-terminal end of this helix. Thus, the regulation of the polymerization equilibrium of tubulin by GTP and tubulin-binding drugs is likely to be based on subtle conformational transitions brought about by intermolecular interactions.

Understanding the mechanism of the conformational changes that regulate tubulin polymerization and how this mechanism is influenced by tubulin binding drugs is a challenging, attractive goal, which would open the doors to the design of new potent anti-cancer drugs and therefore to a planned control of cell division.

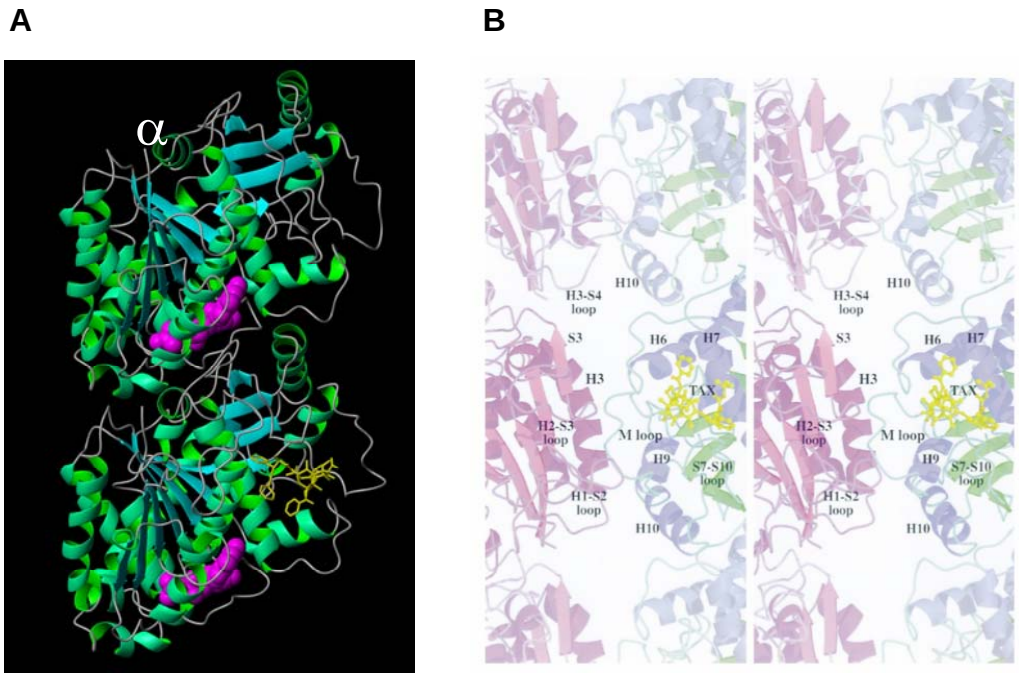


Fig. 1.1 Secondary structure of the protein tubulin as derived by EC study on Zn-stabilized tubulin sheets (Nogales *et al.*, 1998). B) Lateral contacts in the microtubules, as derived from the 8 Å resolution EC model (reproduced from Nogales *et al.*, 1999).

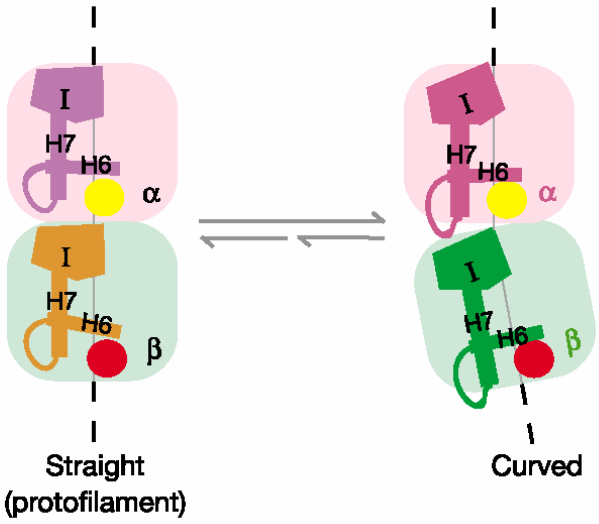
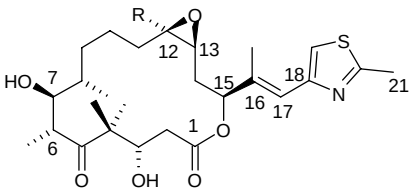


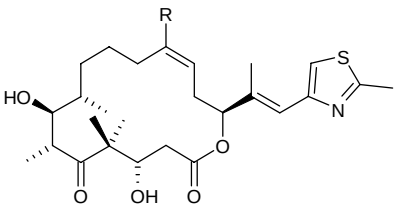
Fig. 1.2 Schematic representation of the straight and curved tubulin heterodimer showing the different position of the H6,H7 helices in the two conformations (reproduced from Ravelli *et al.*, 2004)

Epothilones. In the search for tubulin-binding drugs with an improved pharmacological profile with respect to taxol, the epothilones, which compete with taxanes for the same tubulin binding pocket (Kowalski *et al.*, 1997), appear to be extraordinarily promising. Isolated by Höfle and coworkers from the myxobacterium *Sorangium cellulosum* (Höfle *et al.,* 1993), epothilones (Fig. 1.3) show a very low propensity to induce multidrug resistance (Bollag *et al.*, 1995), are better deliverable to cells due to their higher solubility and possess cytotoxic activity on cell lines resistant to paclitaxel (Kowalski *et al.*, 1997).



R = H: Epothilone A

R = CH_3: Epothilone B



R = H: Epothilone C (Deoxyepothilone A)

R = CH_3: Epothilone D (Deoxyepothilone B)

Fig. 1.3 Chemical structures of epothilones

At present, five members of the epothilone family of structures are undergoing clinical trials, including epothilone B (EPO906), BMS-247550 (the lactam analog of epothilone B), KOS-862 (deoxyepothilone B, epothilone D), BMS-310705 (C-21-amino epothilone B) (reviewed in Borzilleri & Vite, 2002; Altmann, 2003), and most recently, an analog of epothilone B of undisclosed structure, which is developed by Schering AG, Berlin. Since the discovery of the cytotoxic activity of epothilones, strong efforts have been made to derive a common pharmacophore model that could predict the binding mode of epothilones based on the binding mode of paclitaxel (Ojima *et al*., 1999; Wang *et al*., 1999; Giannakakou *et al*., 2000). Despite the availability of biological data for more than 500 chemically modified epothilones, however, the structure-activity relationship (SAR) data were unable to provide a universal model of the ligand-protein complex.

1.2 The question we ask.

In this work we wish to answer the following questions:

1. Which is the active conformation of ***epothilones***?
2. How do epothilones bind to tubulin and do they share a common pharmacophore with the taxanes?
3. Do epothilones bind to unpolymerized tubulin and induce tubulin polymerization or do they bind exclusively to pre-formed microtubules?
4. Which is the mechanisms by which epothilones control tubulin polymerization?

The bioactive conformation of the epothilones is studied by NMR in solution. Our studies are conducted in D_2O solution in the absence of GTP and Mg2+, which promote tubulin polymerization. Our goal is to prevent tubulin from forming microubules even in the presence of an excess of microtubules stabilizing agents, such as epothilone. The rationale behind it is the wish of decoupling the ability of the natural product to *bind* tubulin from that of *promoting or inhibiting tubulin polymerization*, as these two biological activities must not necessarily be linked. By doing so, we wish to study the initial binding of the drug to tubulin, an event which we think (and will prove) is required to induce the protein conformational change responsible for the promotion of microtubules formation. The

polymerization state of tubulin in the solution used for the NMR studies is verified by electron microscopy (EM).

1.3 Significance of the work.

While the last decade has witnessed significant progress in the understanding of the molecular mechanisms underlying the diverse group of diseases subsumed under the general term "cancer" (there are in fact more than 150 different types of cancer), the successful treatment of these diseases is still far from routine and the development of improved drugs is of critical importance. In parallel with advances in cancer biology and apart from continued efforts in the area of classic cytotoxic agents, anticancer drug discovery in the more recent past has focused on targets related to e. g. cell cycle progression and signal transduction. However, the major types of solid human tumors (breast, lung, prostate, and colon), which constitute the vast majority of cancer cases today, are multi-causal in nature and there is a growing recognition that the treatment of solid tumors with "mechanism-based" agents is unlikely to be successful as monotherapy. Instead, improved treatment strategies are likely to involve combinations of e. g. signal transduction inhibitors with new and better cytotoxic agents. Thus the search for improved cytotoxics represents an important aspect of modern anticancer drug discovery and the identification of such agents will be critical for future advances in cancer therapy.

We intend to unravel both the active conformation and the binding mode of epothilones to tubulin with the goal of defining the minimum requisites for biological activity. With the help of the three-dimensional model of the ligand /tubulin complex, we will propose a possible functional mechanisms that provides a rationale for the conformational control exerted on tubulin by epothilones. These data will improve our general understanding of the interaction of tubulin with small organic molecules and will assist the *de novo* design of structurally simplified tubulin binders with certain built in properties (e. g. with regard to solubility or tumor targeting) and improved synthetic accessibility.

Chapter 2

The experiments

2.1 Methods.

The conformation of the ligand in complex with a macromolecular target can be studied NMR in solution for both tightly and weakly forming complexes. In the weak binding regime (k_{off} > 10^4 Hz), the structure of the bound ligand is accessible also for very large complexes (>100 kDa), which are not amenable to NMR studies in the tight binding regime. When a small molecule binds to a macromolecular target with low affinity (k_D > 10^{-6} M) (Fig. 2.1), its NMR observable A equals the population-weighted average of A_{bound} and A_{free}.

$$A = p_{bound} A_{bound} + p_{free} A_{free} \qquad \text{(Eq. 2.1)}$$

Equation 2.1 is applicable if the exchange is fast on the relaxation and chemical shift (CS) timescales of the system, namely k_{off} is larger than the CS difference between the free and the bound forms of the ligand (only one NMR line is observed) and larger than the relevant relaxation constants (T_1^{-1}or T_2^{-1}) of the ligand-target complex. Under these conditions, the NMR observable A_{bound}, developed during the state L_{bound} by a ligand molecule, will be transferred to the state L_{free} and vice versa; this transfer process is mediated by the rapid exchange between L_{bound} and L_{free} during the time t necessary to measure A. Thus, if the k_{off} is larger than the transverse relaxation rate T_2^{-1} of the complex, the transverse relaxation rate T_2^{-1} of a ligand L in a mixture of ligand L and target T is given by:

$$T_{2L}^{-1} = p_{L,bound} T_{2L,bound}^{-1} + p_{L,free} T_{2L,free}^{-1} \qquad \text{(Eq. 2.2)}$$

The individual populations $p_{L,bound}$ and $p_{L,free}$ depend on the k_D and on the concentrations of L and T. From Eq. 2.2 it is evident that the resonances of the

ligand in a ligand-target mixture for which the T_2 of the target is short can only be observed if $p_{free} << p_{bound}$, namely if the ligand is in excess with respect to the target ([L] >> [T]). Under these conditions, even if $T_{2L,bound}$ is too short to allow for the observation of the resonances of the complex, the NMR signals of the observed free ligand carry valuable information on the NMR properties of ligand in the complex.

In the weakly binding regime and when the complex size too large to observe the NMR lines of the receptor, the structure of the ligand in the bound form can be obtained via transferred-NOEs measurements. Similarly to Eq. 2.2, the averaged NOE observed between two protons H_A and H_B of a ligand in the presence of its weakly binding receptor can be described as:

$$NOE_{AB} = p_{L,\,free} NOE_{AB,\,free} + p_{L,bound} NOE_{AB,bound}. \qquad \text{(Eq. 2.3)}$$

If the correlation time of the complex is several orders of magnitudes larger than the correlation time of the free ligand, the $NOE_{AB,bound}$ will be much larger than the $NOE_{AB,free}$, thus dominating the NOE_{AB} even in the presence of an excess of free ligand ($p_{L\ free} > p_{L,bound}$). In particular, if the NOE of the free ligand in solution is positive, a sign change of the NOE is observed upon addition of target. This effect is called transferred-NOE (Balaram *et al.*, 1972).

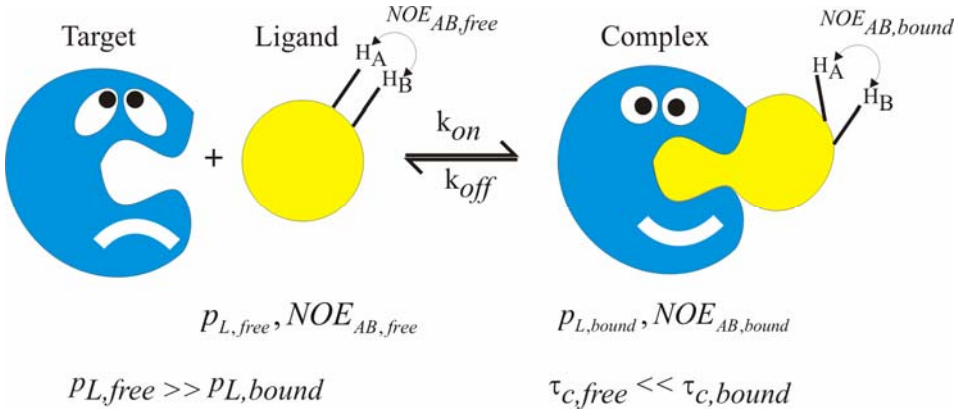


Fig. 2.1. Cartoon showing the reversible binding of the Ligand L to the target T to form the complex LT. The conformation of the ligand can change upon binding. The NMR properties of the ligand are averaged according to Eq. 2.1. The ligand is in excess with respect to the target (p_{free} >

p_{bound}). The average of Eq. 2.1 for the NMR properties that depend on the τ_c of the molecule (NOE, CCRR) is dominated by the bound term (second term), due to the fact that $\tau_{c,bound} >> \tau_{c,free}$.

Epothilones bind to the soluble, non-microtubular form of tubulin in the fast exchange regime (k_{off} > 10000 Hz). The tubulin-bound conformation of these compounds can be studied by tr-NOEs measurement at various mixing times conducted for a mixture of ligand and protein in 50:1 concentration ratio.

However, for small non peptidic, flexible ligands like epothilone, the determination of the ligand structure from this kind of analysis is not accurate. Owing to the large conformational space available for the small ligand and to the absence of long-range distance restraints, multiple conformations can sometimes satisfy the same distance restraints set, derived from transferred-NOE measurements, and the unique biologically active structure cannot be determined unambiguously. Ambiguities among different conformations that satisfy a common NOE-derived distance set for a small molecule free in solution can commonly be lifted by the determination of dihedral angles through scalar ^{3}J-couplings. However, scalar couplings do not provide any angular structural information on the bound conformation of a ligand in fast exchange with its receptor. In the presence of an excess of ligand, the averaged scalar coupling $J_{AB} = p_{L,\,free}\, J_{AB,\,free} + p_{L,bound}\, J_{AB,bound}$ is dominated by the contribution of the free form, as J couplings do not scale as a function of the correlation time of the complex. Reif *et al.* (Reif *et al.*, 1997) demonstrated that Cross Correlated Relaxation (CCR) rates constitute a viable alternative to coupling constants for the determination of dihedral angles in large systems. Analogous to NOEs, CCR rates depend on the correlation time of the molecule and can therefore be used in a transferred fashion to derive angular structural information for a small ligand weakly bound to a macromolecule. This tool is unique, as it exclusively allows the measurement of ligand dihedral angles in a weakly bound complex. The use of CCR rates to derive angular information for a ligand weakly bound to a receptor was proposed for the first time in our laboratory (Carlomagno *et al.*, 1999).

The tubulin bound conformation of epothilone was poorly determined by tr-NOEs alone, as the conformation of the highly flexible 16-membered macrocycle could not be unambiguously defined. Therefore we developed a series of novel

experiments for the measurement of specific angles in the macrocycle via cross-correlated relaxation rates. The general theory of CCR rates in a mixture of a ligand and its weakly binding receptor, together with some of the newly developed experiments, are described in Chapter 3. Briefly, CH–CH cross-correlated relaxation rates (tr-CCR rates) for the determination of the C1-C2-C3-C4, C5-C6-C7-C8, C12-C13-C14-C15, and C13-C14-C15-C16 torsion angles were measured in HCCH correlation experiments, which were adapted from the original pulse program (Felli *et al.*, 1999) to achieve the desired selectivity on the various epothilone moieties. The O-C1-C2-C3 torsion angle was measured in a dipole-CSA (CSA = chemical-shift anisotropy) tr-CCR experiment that was adapted from the experiment used to measure CCR rates in the protein backbone (Yang *et al.*, 1997). The C2-C3-C4-C5 and C14-C15-C16-C17 torsion angles were measured with the help of a newly developed tr-CCR experiment (see Chapter 3). The measurement of CCR rates requires the use of ^{13}C labeled epothilone, which was obtained by shake flask fermentation of the myxobacterium *Sorangium cellulosum*, as described in the material section.

Epothilone NMR experiments. The tubulin-bound conformation of epothilone A was calculated from 46 interproton-distance restraints and seven torsion-angle restraints measured for a 0.5 mM solution of epothilone A in water in the presence of 5 μM tubulin. The distance restraints were derived from transferred NOE experiments. To filter out spin-diffusion-mediated peaks, only those signals with opposite sign to the diagonal peaks in a transferred ROESYexperiment were taken into account. The NOE constraints used to calculate the structure of the tubulin-bound conformation of epothilone were obtained from "transferred-NOE" experiments at mixing times of 100, 200, and 300 ms. The tr-ROESYexperiment had a mixing time of 60 ms with CW irradiation at a field strength of 3500 Hz. The dihedral-angle restraints were obtained by measuring CH–CH dipolar–dipolar and CH–CO dipolar–CSA transferred CCR rates for 60–70% ^{13}C-labeled epothilone A. The transferred cross-correlated relaxation (tr-CCR) experiments were indispensable to obtain a unique description of the bound conformation, as more than one structure of the macrolide ring is compatible with the same H-H distance set (NOE intensities). The seven tr-CCR rates defined the torsion angles O1-C1-C2-C3, C1-C2-C3-C4, C2-C3-C4-C5, C5-C6-C7-C8, C12-C13-C14-C15,

C13-C14-C15-O1 and C14-C15-C16-C17. One particular problem in the determination of the bound conformation of epothilone was the stereospecific assignment of the two methyl groups at C4, which turned out to be opposite to the published assignment. The assignment of the two diastereotopic methyl groups had to be established as it was essential to determine the dihedral angle around the C3-C4 bond by cross-correlated relaxation and consequently detect the large conformational change that takes place in this region upon binding to tubulin. Owing to the paucity of protons around the C3-C4 bond and to the conformational averaging present in this region for epothilone in solution in the absence of tubulin, the stereospecific assignment of the methyl groups could not be carried out by a combination of NOE interactions and $^3J_{CH}$ coupling constants. Therefore, we relied on the combination of NOE experiments, $^3J_{C,H}$ and $^3J_{C,C}$ couplings constants ($^3J_{C2,C5}$, $^3J_{C2,C22}$, and $^3J_{C2,C23}$) measured for epothilone B in the absence of tubulin at –20ºC in dimethylformamide. Under these conditions, the macrocycle is in exchange between two conformations around the C3-C4 bond. Two pictures are compatible with the measured set of $^3J_{C,H}$ ($^3J_{C22,H3} \approx {}^3J_{C5,H3} > {}^3J_{C23,H3}$): 1) C22 = pro-R, C23 = pro-S, and a mixture of *gauche*+ and *trans* conformations around the C3-C4 bond; 2) pro S-C22, proR-C23, and a mixture of *gauche*+ and *gauche*− conformations around the C3-C4 bond. The measured set of $^3J_{C,C}$ couplings ($^3J_{C2,C5} < {}^3J_{C2,C23} < {}^3J_{C2,C22}$) unequivocally confirms alternative 2. The resulting stereospecific assignment is opposite to that previously published (Höfle *et al.*, 1996).

Structure calculations of the active conformation of epothilone. The structure of the tubulin-bound conformation of epothilone was determined by using the simulated annealing protocol, including NMR restraints, of the X-PLOR program (Brünger, 1992). The force field was adapted to include atom types occurring in epothilone A. Bond lengths and angles were taken from the X-ray structure of the free conformation. In addition, the force field included 46 NOE restraints and seven dihedral angles derived from tr-CCR rates. From a random starting point, an initial energy minimization of 50 steps, a high-temperature phase (32.5 ps, 6500 steps, 2000 K) and two cooling phases (25.0 ps, 5000 steps, 2000 K→1000 K/10.0 ps, 2000 steps, 1000 K→100 K) were performed. The closest local minimum was reached by applying a final energy minimization of 200 steps. This

protocol was repeated 100 times. Within the ten lowest energy structures, all dihedral angles have a standard deviation of less than 28º and the RMSD of the positions of the heavy atoms is 0.044 Å.

Determination of the binding mode of epothilone to tubulin. While the bioactive conformation of a ligand weakly bound to a macromolecular receptor can be obtained by tr-NOE and tr-CCR experiments, the orientation of the ligand in the receptor binding pocket cannot be assessed by the available NMR techniques if the resonances of the receptor are broad beyond detection or, as in this case, the receptor cannot be obtained by recombinant techniques with the ^{13}C,^{15}N and ^{2}D labeling scheme necessary for NMR studies of large proteins. In this work, we developed a completely new NMR approach, which allows the determination of the relative orientation of two ligands in the receptor binding pocket in an "indirect" way, namely without observing the resonances of the protein. The method is based on the observation of protein-mediated intermolecular NOEs between two ligands binding weakly and competitive to a common receptor. If the orientation of one of the two ligands in the receptor binding pocket is known, either by crystallographic or FRET data, the orientation of the second ligand can be derived on the basis of the observed interligand NOEs. This methodology, that we called INPHARMA (Interligand NOE for PHARmacophore Mapping) is described in detail in the next chapter. Our idea is to derive the binding mode of epothilone to tubulin on the basis of the binding mode of paclitaxel to tubulin, which has been revealed by electron crystallography (Nogales *et al*., 1998). However, instead of paclitaxel, we use baccatin III (BacIII) (Fig. 3.5, Chapter 3), which shares a large portion of the chemical structure of paclitaxel, but lacks the side chain on the C13, which is responsible for the very poor solubility of paclitaxel. Our assumption is that the binding mode of BacIII to unpolymerized tubulin is analogous to that of paclitaxel to tubulin polymerized on Zn sheets. Although not effective in promoting tubulin polymerization (He *et al.*, 2000), BacIII binds to tubulin in the taxane binding pocket with a k_D in the micro-molar range, as demonstrated by NMR competition experiments between epothilone and baccatin in the presence of tubulin (see Chapter 3). The INPHARMA method was then applied to a mixture of the protein tubulin and the two ligands epothilone and baccatin III, in a concentration ratio of 1:50:50. Such

concentration ratios are typical for the observation of tr-NOEs of ligands binding to a target protein with a k_D in the micro-molar range and allowed us to observe protein-mediated intermolecular NOEs between EpoA and BacIII by the INPHARMA method (Sanchez-Pedregal *et al.,* 2005).

This new methodology is of great significance for research in the pharmaceutical industry, as it provides an easy and quick method to derive a common pharmacophore for a series of competitively binding lead compounds.

2.2 Materials.

Tubulin preparation and polymerization state. Tubulin was obtained from cytoskeleton (bovine brain, T238). The protein was dialyzed against a D_2O solution containing 3 mM phosphate, 3 mM calcium and ca. 0.7 mM sodium at pH 7.0. For the determination of the tubulin bound conformation of EpoA, tubulin was dialyzed against a H_2O solution containing 3 mM phosphate, 3 mM calcium and ca. 0.7 mM sodium at pH 7.0 lyophilized and dissolved in D_2O prior to usage. In this buffer and in the absence of any ligand tubulin exists in the form of high molecular weight, inhomogeneous "aggregates" of α,β – dimers, as observed by electron microscopy (EM) and size-exclusion chromatography. The tubulin "aggregates" are disordered polymers (Fig. 2.2), and no microtubules or ordered polymers are formed under these conditions. We will refer to the "aggregates" as to the unpolymerized form of tubulin, where the term "unpolymerized" refers to the absence of ordered, microtubule-like polymers. Upon addition of EpoA, the size of the "aggregates" increases. Moreover, the addition of EpoA promotes polymerization of tubulin in the form of "tubules sheets" or "open microtubules" (Fig. 2.2). These ordered polymers are similar to microtubules, which are not correctly closed into the typical cylindric form. Similar polymers have also been observed in microtubules preparations stabilized by epothilone, as reported by Höfle and coworkers (GBF Scientific Reports, 2005), stabilized by paclitaxel, as reported by Diaz and Andreu (Diaz & Andreu, 1993) and in GMCPP microtubules formed at low temperature (Wang & Nogales, 2005). The sheets are often seen in fast grow conditions, where the microtubules do not have enough time to close while growing. The observation that epothilone induces the

reorganization of the tubulin “aggregates” into more ordered polymers indicates that the drug binds to the tubulin “aggregates” and induces a conformational change in the protein, which is necessary for the formation of the elongated polymers. While it is known that epothilones bind much tighter to microtubules than to soluble tubulin (Buey *et al.*, 2004), our data demonstrate that epothilone functionally binds to unpolymerized soluble tubulin, inducing formation of ordered tubulin polymers, even in absence of Mg2+ and GTP. Baccatin does not promote formation of tubules sheets, in agreement with the lack of MT-stabilizing activity (Fig. 2.2).

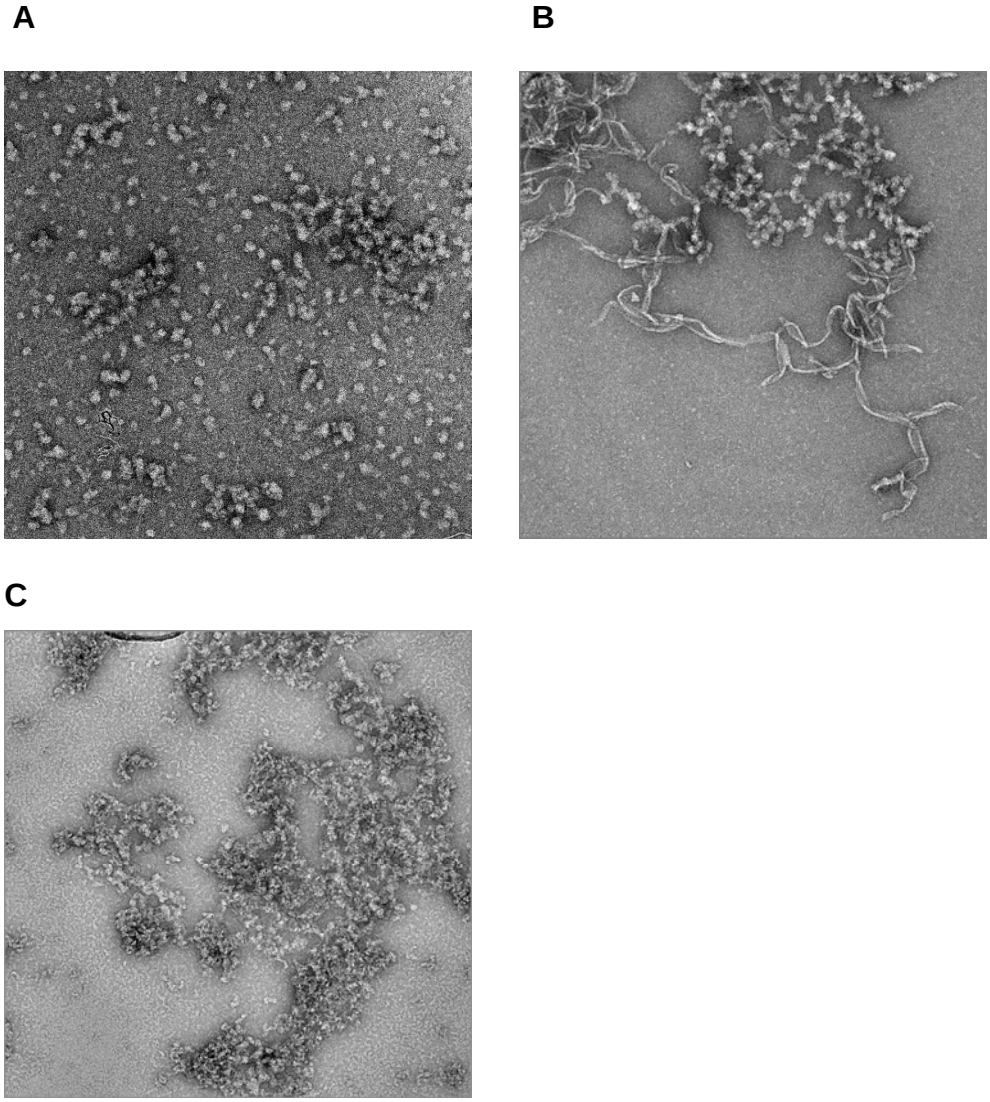


Fig. 2.2 Electron microscopy (EM) images of tubulin (12μM) dissolved in a D_2O solution containing 3 mM phosphate, 3 mM calcium and ca. 0.7 mM sodium at pH 7.0 (panel A). In panel

B the same tubulin sample was incubated for 1 hour with 0.6 mM epothilone A, leading to the formation of ordered polymers (microtubules sheets). In panel C the same tubulin sample was incubated for 1 hour with 0.6 mM baccatin In panel C no ordered polymers could be observed, in agreement with the lack of MT-stabilizing activity in baccatin.

The structural data discussed in this work refer to the ligand bound to tubulin in the unpolymerized form. Tubulin "aggregates" are always seen in EM images of microtubules (Diaz & Andreu, 1993), underlying the biological importance of such disordered polymers in the conformational equilibrium of tubulin. The amount of epothilone bound to the microtubules sheets (Fig. 2.2) does not contribute to the NMR signal. This is due to the fact that only signals from fast exchanging ligands can be observed in the NMR experiments, while epothilones are known to bind tightly to microtubules (Buey *et al.*, 2004). In agreement with this consideration we do not observe any change in the EpoA NOE signals with time, while the microtubules sheets precipitate out of solution in a few days. Moreover, the average correlation time τ_c of the tubulin-bound epothilone observed in the NMR spectra is approximately 1.3 μs, which roughly corresponds to 15 protein dimers, while the microtubules sheets observed in the EM images contain several hundreds of dimers. The average correlation time of 1.3 μs is thus incompatible with the hypothesis that the microtubules sheets measurably contribute to the observed NMR signal.

Sample used for the determination of the epothilone active conformation. All the tr-CCR experiments were acquired for a solution of ^{13}C-labeled epothilone A (0.5 mM)and tubulin (5 μM)in D_2O. ^{13}C-labeled epothilone was obtained by shake flask fermentation of the myxobacterium *Sorangium cellulosum* So ce90 by using a high epothilone-producing mutant, BCE99/41 (Novartis strain collection). The fermentation was performed by growing the strain in medium (1000 mL) containing ^{13}C-labeled starch ($^{13}C_6$, 98%; 20 g) as carbohydrate source. Yields: epothilone A (20 mg) and epothilone B (15 mg), incorporation rate: 60–70% ^{13}C.

Sample used for the determination of the binding mode of epothilone tubulin. Interligand NOEs were measured between EpoA and BacIII for a mixture of EpoA (0.6 mM), BacIII (0.6 mM) and tubulin (12μM) in D_2O containing 3 mM phosphate, 3 mM calcium and ca. 0.7 mM sodium at pH 7.0. In this preparation

tubulin was not lyophilized, as in the preparation used to determine the bioactive conformation of epothilone. Instead, the commercially available tubulin was dialysed against the desired buffer twice for 10 hours. We chose not to lyophilize tubulin, as we observed that lyophylization denatures a considerable amount of the protein, which is then unable to bind epothilone. Baccatin III was purchased from Sigma Aldrich.

Chapter 3

New Methodology Development

3.1 Cross-Correlated relaxation rates.

Cross-correlated relaxation (CCR) rates constitute a valid alternative to coupling constants for the determination of dihedral angles (Reif *et al.*, 1997). For high molecular weight systems the measurement of scalar coupling constants poses serious problems caused by large line-widths and by differential relaxation effects (Harbison, 1993; Norwood, 1993*)*. Conversely, CCR measurements are feasible also for molecules of large size and have been successfully employed in the past years to determine dihedral angles in biomolecules. Additionally, we have reported on the usage of cross-correlated relaxation rates to derive angular structural information for a small ligand weakly bound to a macromolecule (Carlomagno *et al*., 1999; Blommers *et al*., 1999). Similarly to "transferred-NOEs" (Ni, 1994; Clore & Gronenborn, 1982), "transferred-CCR rates" are informative of the bound conformation of the ligand, when the complex is in the fast exchange regime on the NMR time-scale. It is noted that coupling constants are unable to provide information on the bound conformation of the ligand, because their value is independent of the molecular correlation time.

CCR rates measurements have been extensively employed for the determination of the tubulin-bound conformation of the microtubules-stabilizing agent epothilone A (EpoA) (see chapter 4 for results). The conformation of the intervening torsion angles between atoms O_1 and C_6 (Fig. 1.3) was poorly defined by transferred-NOE data and more than one structure of the macrolide ring was compatible with the NOE derived H-H distance set. However, chemical modifications data indicated this region as particularly important for the biological activity of epothilone, underlining the necessity of obtaining reliable structural data for this part of the macrolide ring. This was possible thanks to the measurement of four CCR rates that defined the torsion angles O_1-C_1-C_2-C_3, C_1-C_2-C_3-C_4, C_2-C_3-C_4-C_5 and C_5-C_6-C_7-C_8 (Fig. 1.3). Additional CCR rates

measurements restricted the torsion angles C_{12}-C_{13}-C_{14}-C_{15}, C_{13}-C_{14}-C_{15}-O_1 and C_{14}-C_{15}-C_{16}-C_{17}. Three kinds of CCR rates were measured using the Γ-quantitative approach: CH-CH dipolar-dipolar CCR (Felli *et al.*, 1999), CO-CH CSA (Chemical Shift Anisotropy)-dipolar CCR (Yang *et al.*, 1997) and CH-CH_3 dipolar-dipolar CCR rates.

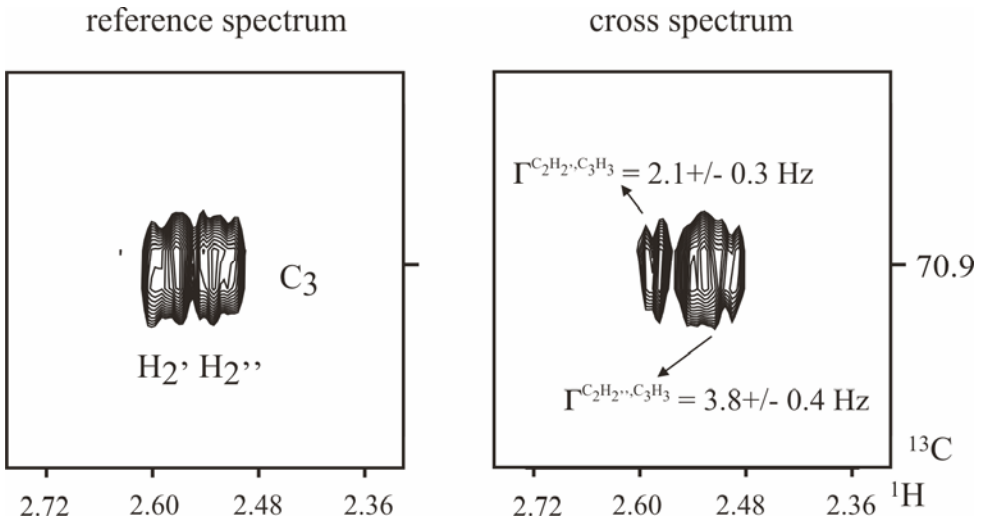


Fig. 3.1 Cross and reference experiments for the measurement of the $\Gamma^{C_2H_2',C_3H_3}$ and the $\Gamma^{C_2H_2'',C_3H_3}$ of epothilone in a Γ-quantitative HCCH correlation. The CCR rates are obtained by the ratio of the peak intensities in the cross and reference experiments: the $\Gamma^{C_2H_2',C_3H_3}$ is obtained from the C_3-H_2' peak and the $\Gamma^{C_2H_2'',C_3H_3}$ from the C_3-H_2'' peak.

The example of epothilone A is representative of many other cases where angular structural information is required for the determination of the bio-active conformation of a ligand in a weakly formed complex; CCR rates are the only NMR parameters to carry such information. Unfortunately, CCR rates measurements require ^{13}C labelled ligands. Some attempts have been made to develop new experiments to obtain CCR rates at natural abundance (Ilin *et al.*, 2003). The development of such methods is of great relevance, given the elevated costs associated with the chemical synthesis of ^{13}C labelled ligands.

A variety of pulse sequences to obtain dipolar-dipolar and CSA-dipolar CCR rates are available in the literature for proteins and nucleic acids. In principle similar sequences can be used to measure the same CCR rates in a "transferred" fashion for ligand-receptor complexes. It is advisable to adjust the pulse sequence for

every ligand, especially by introducing selective pulses, in order to optimize magnetization transfer efficiencies for the individual combinations of spins to be correlated. By adapting pulse sequences available in the literature we measured the O_1-C_1-C_2-C_3, C_1-C_2-C_3-C_4, C_5-C_6-C_7-C_8, C_{12}-C_{13}-C_{14}-C_{15} and C_{13}-C_{14}-C_{15}-O_1 torsion angles (Fig. 3.1).

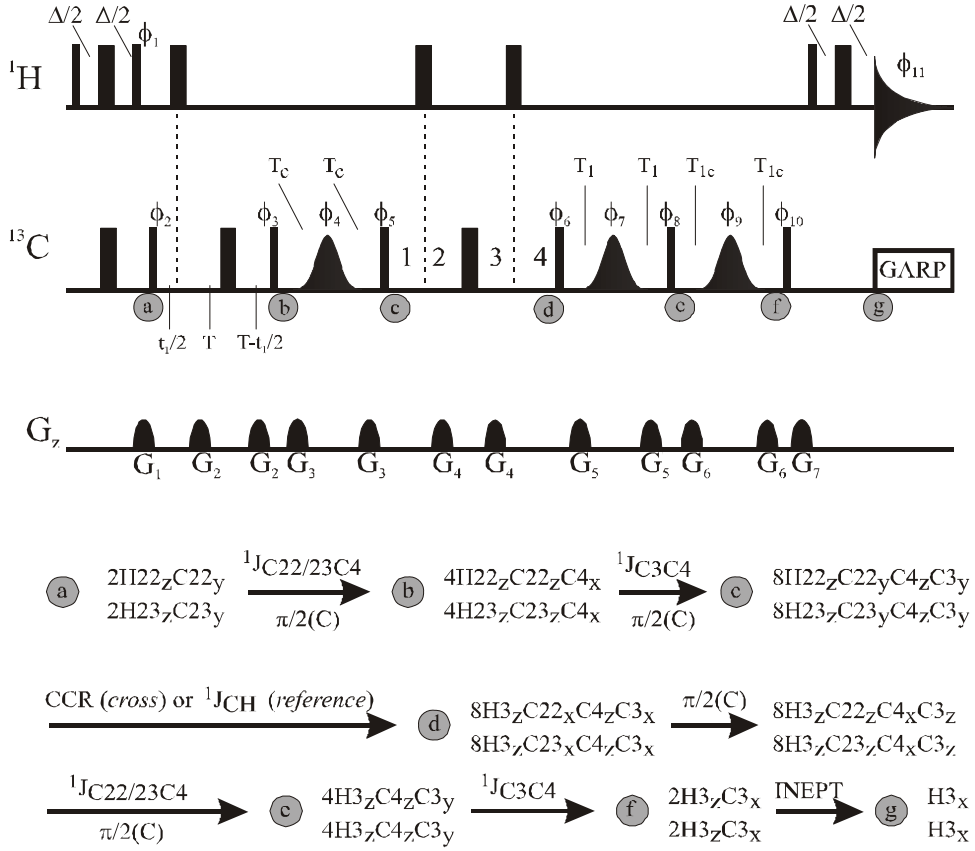


Figure 3.2 Pulse sequence for the measurement of dipolar-dipolar cross-correlated relaxation between the $(CH)^{22/23}_{ave.}$ vector and the $(CH)^3$ vector in epothilone. $\Delta = 1/(2^1J_{CH})$; $T = 1/(4^1J_{CC}) = 7$ ms; $T_c = 1/(4^1J_{CC}) = 7$ ms; $T_1 = 1/(8^1J_{CC}) = 3.5$ ms; $T_{1c} = T_c$. Delays 1-4 are equal to $T_{rel}/4$ ($T_{rel} = 1/^1J_{CC} = 28$ ms) in the cross experiment; in the reference experiment delay 1 = $T_{rel}/4 + \Delta/8$ = delay 3 and delay 2 = $T_{rel}/4 - \Delta/8$ = delay 4. ϕ_1 = y; ϕ_2 = x, -x; ϕ_3 = 2(y), 2(-y); ϕ_5 = 4(x), 4(-x); ϕ_6 = 8(y), 8(-y); ϕ_8 = x; ϕ_{10} = y;ϕ_{11} = 2(x,-x), 4(-x,x), 2(x,-x); $\phi_4 = \phi_7 = \phi_9$ = x. The pulses with phase ϕ_4 and ϕ_9 are Q3 pulses of 768 ms duration, centered at 63 ppm, to selectively invert the C^3 and C^4 (C^3 resonates at 70.9 ppm and C^4 at 53.0 ppm); the pulse with phase ϕ_7 is a Q3 pulse of 512 ms duration centered at 37 ppm, to selectively invert the $C^{22/23}$ and C^4 (C^{22} resonates at 19.1 ppm and C^{23} at 21.7 ppm). The proton carrier was at 4.78 ppm and the carbon carrier at 37 ppm;

spectral widths were 4807.69 Hz for proton and 8333.33 Hz for carbon. Carbon decoupling in acquisition consisted of a GARP modulated pulse train at 2.38 kHz field strength.

The determination of the torsions around a C-C bond involving a quaternary carbon, as for the C_2-C_3-C_4-C_5 and C_{14}-C_{15}-C_{16}-C_{17} torsion angles in EpoA, represents a challenging problem. In this case no CH-CH dipolar-dipolar CCR rate can be measured and no large CSA tensor is available for the measurement of a CSA-dipolar CCR rate. During this work, we developed a new experiment to access the C_2-C_3-C_4-C_5 and C_{14}-C_{15}-C_{16}-C_{17} torsion angles. In the new approach we measure CH-CH cross correlated relaxation in a H^aC^a-$C^b(CH_3)^c$ moiety, where the second C is a quaternary carbon. This rate allows to define the dihedral angle about the C^aC^b bond. We have applied the pulse sequence to a 100:1 mixture of ^{13}C-labeled epothilone A (Fig. 1.3) and tubulin to determine the dihedral angle about the C^3-C^4 and C^{15}-C^{16} bond of epothilone in its bound conformation. In the following we will decribe in detail the measurement of the C_2-C_3-C_4-C_5 torsion angle.

The dihedral angle about the C^3-C^4 bond of epothilone is accessible by measuring the dipolar-dipolar CCR rates between the $(CH)^3$ vector and the $(CH)^{22}$ or $(CH)^{23}$ vectors. The direction of the three CH bonds in each methyl group is averaged as a consequence of the fast rotation around the C^4-C^{22} and C^4-C^{23} bonds. Therefore, the $(CH)^{22}$ and $(CH)^{23}$ vectors assume an average orientation that is collinear with the C^4-C^{22} and C^4-C^{23} bonds, respectively.

The pulse sequence used for the measurement of the $\Gamma_{CH,CH}$ rate between the $(CH)^3$ vector and the $(CH)^{22/23}{}_{ave.}$ vectors is shown in Fig. 3.2.

Two experiments are recorded, yielding a *cross* and a *reference* spectrum (quantitative Γ approach) (Felli *et al.*, 1999). The CCR rate is extracted from the ratio of peak intensities in the *cross* and *reference* experiment. The pulse sequence is optimized for maximum sensitivity of the cross experiment: the $8H_z{}^{22/23}C_y{}^{22/23}C_z{}^4C_y{}^3$ coherence present at point c is transformed by $\Gamma_{(CH)^{22/23},(CH)^3}$ into $8C_x{}^{22/23}C_z{}^4C_x{}^3H_z{}^3$ with efficiency equal to $3\sinh(\Gamma_{(CH)^{22/23},(CH)^3}T_{rel})\cosh^2(\Gamma_{(CH)^{22/23},(CH)^3}T_{rel})$. The $8C_x{}^{22/23}C_z{}^4C_x{}^3H_z{}^3$ term is then transformed into $H_x{}^3$ for detection. The delays

are optimized for optimal refocusing and defocusing of the couplings. Selective π pulses on the carbons prevent loss of magnetization to other passively coupled carbons. The details of the coherence transfers are provided in Fig. 3.1 and its legend.

The *cross* and *reference* spectra are shown in Fig. 3.3. Two peaks are observed in the reference spectrum at the chemical shifts of C^{22} and C^{23} in ω_1 and the chemical shift of H^3 in ω_2. In the cross spectrum the $C^{23}(\omega 1)$-$H^3(\omega 2)$ peak is missing, indicating that the dipolar-dipolar CCR between the $(CH)^{23}_{ave.}$ vector and the $(CH)^3$ vector is close to zero.

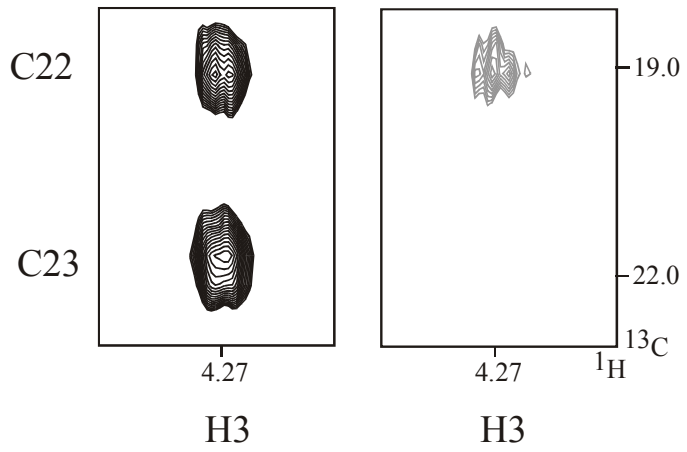


Figure 3.3. *Reference* (left) and *cross* (right) spectra acquired with the sequence of Fig. 3.1. The peaks visible in the reference specrtum are at the chemical shift of C^{22} (19.1 ppm) and C^{23} (21.7 ppm) in the carbon dimension and at the chemical shift of H^3 (4.27 ppm) in the proton dimension. The *cross* experiment was four times longer than the *reference* experiment. 1024 and 64 complex points were acquired in t_2 and t_1, respectively. The final matrices were 2048 × 128 points. The gray lines represent negative contours. The sample contained 0.5 mM epothilone A and 5 μM tubulin dissolved in D_2O.

The quantitative value for the two CCR rates can be extracted according to the equation:

$$I_{cross}/I_{ref} = \tanh(\Gamma T_{rel})/\sin^2(\pi J_{CH}\Delta/2)\cos^2(\pi J_{CH}\Delta/2) \qquad \text{(Eq. 3.1)}$$

where I_{cross} and I_{ref} are the intensities of the peaks in the cross and reference experiment, respectively and Γ indicates the cross-correlated relaxation rate. The

measured CCR rates between the $(CH)^3$ vector and the $(CH)^{23}_{ave.}$ and $(CH)^{22}_{ave.}$ vectors are 0.1 ± 0.1 and –1.4 ± 0.4, respectively. The error was calculated applying classical error propagation theory, assuming that the noise represents the experimental uncertainty of the peak intensities.

The dependence of the two CCR rates on the C^2-C^3-C^4-C^5 dihedral angle θ is given in Eq. (3.2) and (3.3). The $(CH)^{22}_{ave.}$ and $(CH)^{23}_{ave.}$ vectors lie along the direction defined by the C^4-C^{22} and C^4-C^{23} bonds, respectively. The CCR rates $\Gamma^{(CH)^{22},(CH)^3}$ and $\Gamma^{(CH)^{23},(CH)^3}$ are scaled by $P_2[109.7°] = ½[3\cos^2(109.7°) - 1] = -0.3296$, due to the averaging of the C-H directions in each methyl group.

$$\Gamma^{(CH)^{22},(CH)^3} = -0.3296k\tau_c P_2[\cos^2 109.7° \sin^2 109.7° \cos(\theta)] \qquad \text{(Eq. 3.2)}$$

$$\Gamma^{(CH)^{23},(CH)^3} = -0.3296k\tau_c P_2[\cos^2 109.7° \sin^2 109.7° \cos(\theta - 120°)] \qquad \text{(Eq. 3.3)}$$

where $k = (\gamma_C\gamma_H h\mu_0/8\pi^2)^2$, τ_c is the effective correlation time of the molecule and θ is the torsion angle C^2-C^3-C^4-C^5. A graphic representation of the dependence of the two CCR rates on the torsion angle θ is shown in Fig. 3.4 for a τ_c of 0.74 ns, which is the effective τ_c of epothilone in the presence of 5 μM tubulin, as estimated from other CCR measurements, and an order parameter $S^2 = 0.8$ to account for internal motions. The stereospecific assignment of the methyl groups (C^{22} in *pro-S* and C^{23} in *pro-R*) was obtained by the combination of NOEs and $^3J_{HC}$ and $^3J_{CC}$ coupling constants data for epothilone B in the absence of tubulin (see chapter 2) and is opposite to the one reported in the literature (Höfle *et al.*, 1996).

The dihedral angle θ that complies with the ratio between the two CCR rates is –45° ± 5°. We used the ratio of the two CCR rates, instead of the rates themselves, to determine the dihedral angle because this approach is independent from both the τ_c and the internal motion order parameter S^2, if equal internal reorientation of all the CH vectors is assumed.

The change in the θ dihedral angle from the gauche + conformation in the free form to the gauche – conformation in the bound form is supported by tr-NOE data.

The new experiment allows to determine X-C^a-C^b-Y dihedral angles by CCR rates when X is a proton and Y a methyl group. The method is broadly applicable to any HC-CCH_3 moiety, and therefore also to a number of amino acid side-chains in proteins.

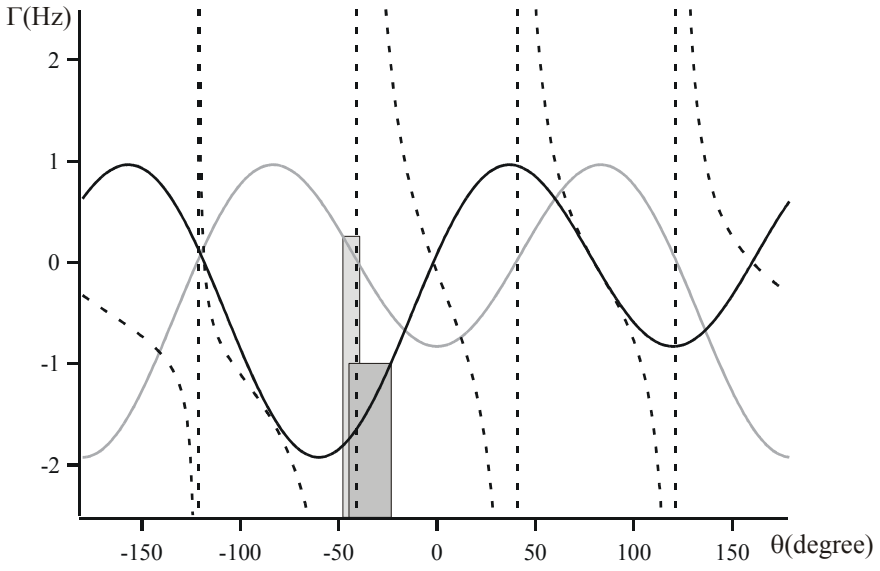


Figure 3.4. Dependence of the two cross-correlated relaxation rates $\Gamma^{(CH)^{22},(CH)^3}$ and $\Gamma^{(CH)^{23},(CH)^3}$ on the dihedral angle $\theta(C^2\text{-}C^3\text{-}C^4\text{-}C^5)$. The two gray shadowed regions represent the value of θ that satisfy the measured CCR rates. The dashed curve represents the ratio $\Gamma^{(CH)^{22},(CH)^3}/\Gamma^{(CH)^{23},(CH)^3}$.

3.2 The INPHARMA approach.

3.2.1 The NMR experiment.

The conformation of a ligand in the binding pocket of a macromolecular receptor can be accessed by NMR in solution by two approaches. If the molecular weight of the complex is of the order of 50 kDa or smaller and $^{13}C/^{15}N$ labelled receptor can be produced *in vitro*, a detailed three-dimensional picture of the ligand and of its interactions with the receptor binding pocket can be obtained with standard NMR methodology (Cavanagh *et al.,* 1996). However, in other cases the macromolecular receptor is too large for NMR studies in solution or it

cannot be synthesized in vitro to introduce NMR active isotopes. In these instances, NMR can be applied to determine the receptor-bound conformation of the ligand, provided that the k_{off} of the binding reaction is much larger than the relaxation constants of the ligand in the complex and the difference of the chemical shifts of the free and bound ligand forms (fast exchange regime). Under this condition, which implies a k_D in the low micromolar to millimolar range, information on the ligand structure is derived from transferred-NOEs (tr-NOE) and transferred Cross-Correlated Relaxation rates (tr-CCRR). The determination of the receptor-bound conformation of the ligand leaves open the question of its orientation in the receptor binding pocket.

The relative orientation of two competitive ligands weakly bound to a common receptor can be determined by the simple and novel approach that we propose here. A NOESY of a mixture of the two ligands in the presence of the common receptor is recorded. Under the fast exchange condition, and provided that the two ligands A and B bind competitively to the macromolecular receptor T, intermolecular tr-NOE peaks between the two ligands A and B can be observed in the NOESY spectrum. A NOE peak between a proton H_A of ligand A and a proton H_B of ligand B in the presence of the receptor T originates from spin diffusion involving a proton H_T of the receptor. During the NOESY mixing time the ligand A binds to T and proton H_A transfers its magnetization to proton H_T (Fig. 3.5). Then the complex AT dissociates and B binds to T. The magnetization of H_A, which had been transferred to H_T, can now be transferred to H_B, resulting in an intermolecular peak H_A-H_B. Clearly this peak can only exist if both H_A and H_B are close to H_T in the AT and BT complexes, respectively. Therefore, a number of such intermolecular NOE peaks describes the relative orientation of the two ligands in the receptor binding pocket. Such interligand NOEs have never been observed so far and should not be mistaken for the interligand NOEs that occur for two ligands binding simultaneously in adjacent or partially overlapping binding pockets (London *et al*., 1992; London, 1999). Here, the two ligands A and B are never either close in space or bound to the protein simultaneously; instead, the two ligands bind to the protein competitively and consecutively and the observed NOEs are a consequence of spin diffusion mediated by the protons of the protein.

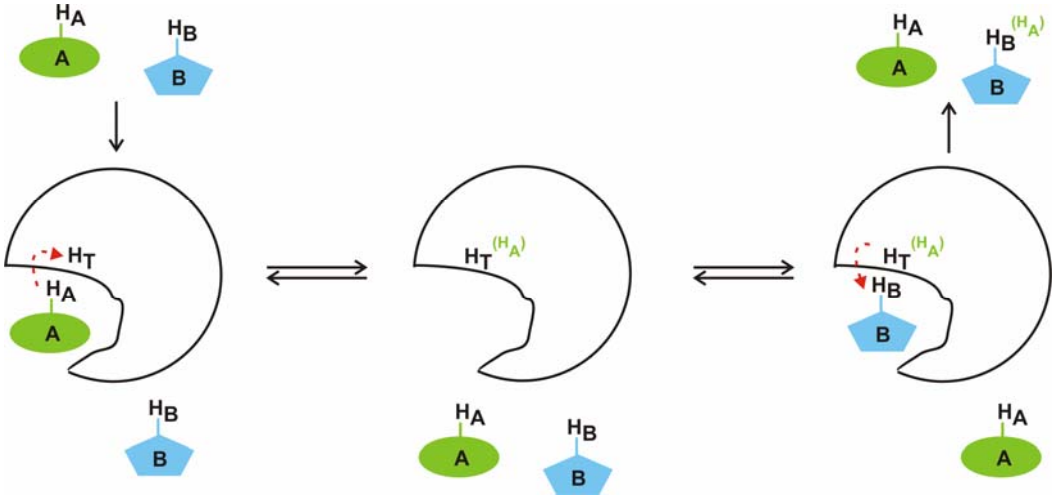


Fig. 3.5. Schematic representation of the principle of the interligand NOEs observed between two competitive ligands A and B, binding consecutively to the same target T. When ligand A (in green) binds to the protein, magnetization is transferred from the proton H_A of A to the proton H_T of T. Subsequently, ligand A leaves the binding site and ligand B (in blue) binds to the target. The proton H_B of B, situated close to H_T, receives part of the magnetization transferred to H_T by H_A of A. This results in a NOE peak between H_A and H_B , mediated by the proton H_T, despite the fact that the two ligands do not bind the protein simultaneously and that H_A and H_B are never directly close in space.

If the orientation of the ligand A in the binding pocket of T is known, the possible binding modes of the ligand B, resulting from docking calculations, can be ranked by evaluating their compatibility with the experimental interligand NOEs. The expected H_A-H_B intermolecular NOEs can be computed for each orientation of the ligand B in the BT complex and compared with the experimental ones. The structures that give rise to intermolecular NOEs in good agreement with the experimental ones are retained, the others rejected. We expect the new methodology, that we call INPHARMA (Interligand NOE for PHARmacophore MApping), to be useful to pharmaceutical research and ligand-target interaction studies, as it offers a rapid and easy way to determine the relative orientations of two ligands in a common binding pocket and, potentially, to identify competitive binding.

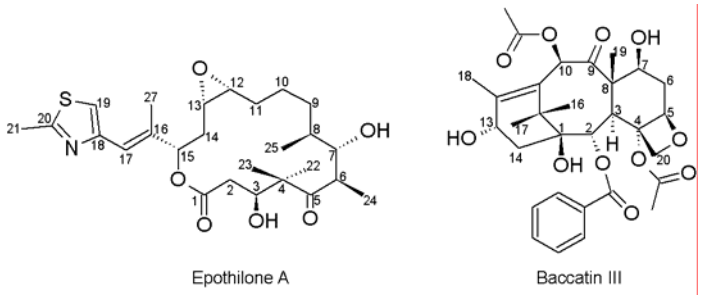


Fig 3.6 Structure of Epothilone A and Baccatin III.

We apply the INPHARMA method to a mixture of epothilone A and baccatin III (Fig. 3.6) in the presence of tubulin. The complex Epo A – tubulin is only transiently formed during the mixing time of a NOESY experiment, and tr-NOEs can be observed for a 100:1 mixture of Epo A and tubulin, making the system suitable for application of the INPHARMA method (epothilone A = ligand B; tubulin = target T). A model of the complex of tubulin with paclitaxel is available by electron crystallography. Unfortunately paclitaxel is poorly soluble in water and therefore not suitable for NMR studies in solution. However, the precursor of paclitaxel, baccatin III (He *et al.*, 2000), lacking the C-13 side-chain, is water soluble, binds to tubulin in a transiently-formed complex and delivers tr-NOE effects. Thus, although the biological activity of baccatin III is lower than that of paclitaxel, we chose baccatin III as ligand A for our experiments. To confirm that baccatin binds in the same pocket as paclitaxel, namely that it competes with epothilone for the taxane binding pocket, we evaluated the change in the tr-NOE rates of baccatin in presence of tubulin upon titration of epothilone. For a concentration ratio of epothilone A:baccatin of 0.67:1, the tr-NOE rates of baccatin decrease to 0.6 of their original value, confirming that the two ligands compete for the same binding pocket.

We recorded NOESY experiments at different mixing times (20, 40, 70, 100 and 200 ms) for a mixture of epothilone A (0.6 mM), baccatin III (0.6 mM) and tubulin (12μM) in D_2O. Intermolecular NOE cross-peaks between the aromatic protons of baccatin III and several methyl groups of Epo A are clearly observable in the section of the NOESY spectrum at 70 ms mixing time, shown in Fig. 3.7.

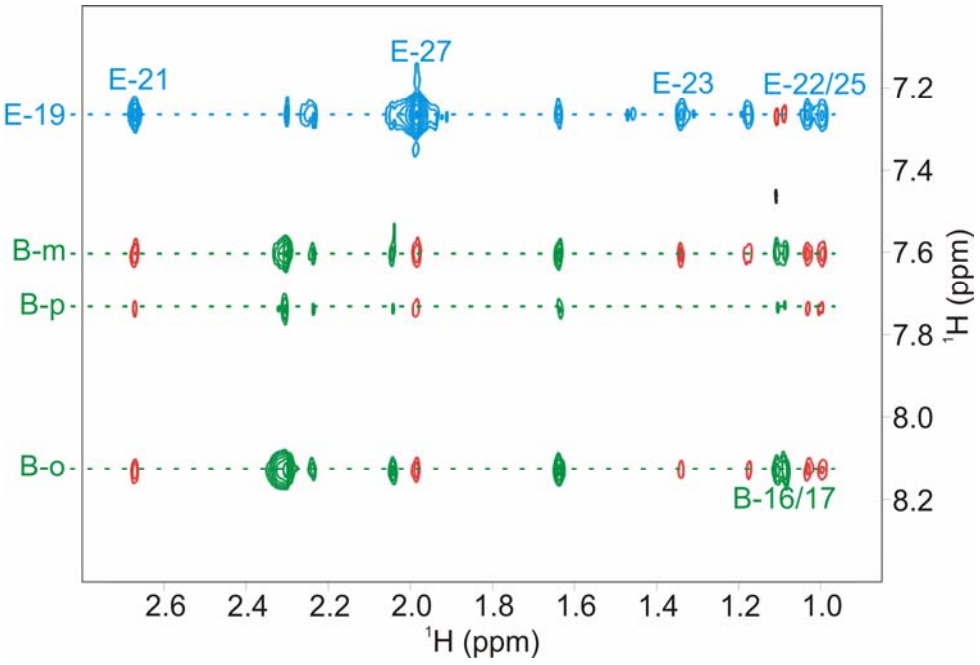


Fig. 3.7 NOESY spectrum of a mixture of tubulin, epothilone A, and baccatin III with concentrations of 12μM, 0.6 mM, and 0.6 mM, respectively. The spectrum has been acquired on a 900 MHz spectrometer equipped with a cryoprobe with a mixing time of 70 ms. The blue and green peaks are intramolecular transferred NOE peaks for epothilone A and baccatin III, respectively. The red peaks represent the interligand transferred NOEs mediated by the protein protons. The numbering of the atoms corresponds to that of Fig. 3.5 (E stands for epothilone and B for baccatin). Bm,Bo and Bp indicate the protons in the meta, orto and para positions of the phenyl ring of baccatin, respectively.

Nineteen intermolecular NOE peaks (Table 3.1) can be already observed in a NOESY experiment at 40 ms mixing time. The build-up curves of the intensities of seven of these NOE peaks (Fig. 3.8) show the characteristic damped parabolic shape for spin diffusion mediated NOE effects. Additional intermolecular NOEs become visible at 100 ms or higher mixing times. As a control, we recorded a NOESY spectrum of the mixture of Epo A and baccatin III in the absence of tubulin at 400 ms mixing time, which showed no intermolecular NOEs. Thus, we can conclude that the intermolecular NOE effects between Epo A and baccatin, observed in presence of tubulin, are mediated by protons of tubulin. In the next section I will describe how we interpret the interligand NOEs to obtain the binding mode of the two ligands to the protein.

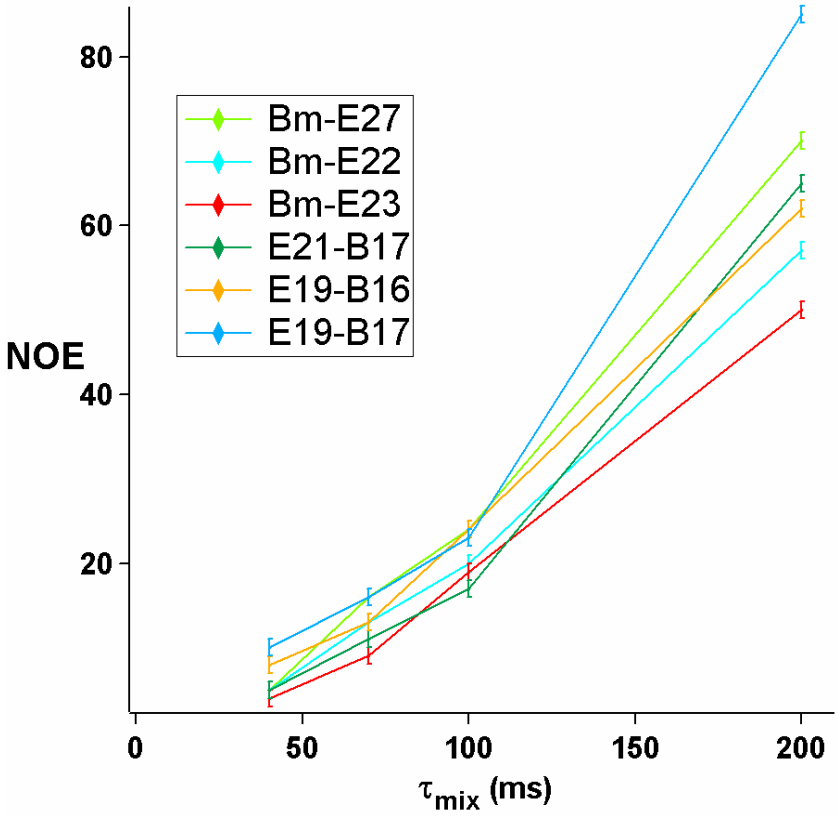


Fig. 3.8 Build-up curves of seven out of the 19 most intense interligand transferred-NOE cross peaks, as assigned in the figure. The typical damped parabolic behaviour can be observed for all peaks, confirming that the interligand NOEs arise from spin-diffusion via the protein protons. The numbering of the atoms corresponds to that of Fig. 3.6 (E stands for epothilone and B for baccatin); Bm indicates the protons in the meta position of the phenyl ring of baccatin.

We also measured intermolecular NOEs between EpoA and taxotere in the presence of tubulin. Taxotere is only purely soluble in water: the sample contained EpoA (0.5 mM), taxotere (0.05 mM), tubulin (3μM), 3 mM phosphate, 3 mM calcium and ca. 0.7 mM sodium at pH 7.0 in D_2O (95%) / DMSO (5%). We could observe intermolecular NOEs between several regions of the two compounds, confirming that EpoA binds in the same pocket as taxotere. However, due to the low concentration of taxotere, the interligand NOEs were too weak to be quantitatively interpreted.

3.1.2 Computational approach for the evaluation of the interligand NOEs.

Our approach to obtain a structural model of the complex between EpoA and tubulin consists of three steps: first, we generate 5000 models for each of the

two ligands complexed with tubulin and combine pairwise all models of the EpoA – tubulin complex with all models of the BacIII – tubulin complex; second, we calculate the theoretical intermolecular tr-NOEs between EpoA and BacIII for each of the 25 million pairs; third, we reject or retain pairs of structures depending on how well the calculated intermolecular tr-NOEs reproduce the experimental ones. We chose to use the soluble BacIII instead than the insoluble paclitaxel as reference ligand with known binding mode, under the assumption that the binding mode of BacIII to unpolymerized, soluble tubulin resembles that of paclitaxel to tubulin polymerized on Zn sheets and revealed by EC (Nogales *et al.*, 1998). However, in the course of the project we had indications that the assumption of a common binding mode of baccatin and paclitaxel might not be confirmed. Therefore we generate 5000 docking model for the BacIII – tubulin complex in order to explore the conformational space of baccatin in the taxane binding pocket of soluble tubulin. From this point on, the INPHARMA method is used without knowledge of the binding mode of one of the two ligands. In the docking calculations we used the tubulin structure derived by electron crystallography (Nogales *et al.,* 1998; Nettles *et al.,* 2004), but we allowed both the backbone and the side-chains of the taxane binding pocket to be flexible, in order to explore a larger conformational space, which might be accessible in solution. For EpoA we used the tubulin-bound conformations derived by tr-NOEs and tr-CCRs in solution (Carlomagno *et al.,* 2003b). The orientation of epothilone in the binding site was not restricted in the docking calculations, which resulted in an extensive sampling of the taxane binding pocket. Among the 5000 docking models generated for the BacIII – tubulin complex, those models were retained, for which the aromatic ring of baccatin is located in one of the three pockets occupied by one of the aromatic rings of paclitaxel in the published electron crystallography structure (Nogales *et al.,* 1998). This choice allows searching the taxane binding pocket for binding modes of baccatin to soluble tubulin which differ from the one of paclitaxel to polymerized tubulin, while maintaining the general profile of the interaction of taxanes with tubulin.

Docking calculations. 5315 EpoA – β-tubulin and 5022 BacIII – β-tubulin complex models were generated with the HADDOCK protocol (Dominguez *et al.,* 2003) using the CNS program (Copyright © International Union of

Crystallography, *Acta Cryst.* D54, 905-921 *1998*). To improve the sampling of the conformational space of tubulin, seven and fourteen initial β-tubulin models were used for the EpoA – tubulin and the BacIII – tubulin complexes, respectively. These models included the electron crystallography (EC) derived structure of polymerized tubulin bound to paclitaxel (1JFF.pdb) (Nogales *et al.,* 1998) and the EC-derived structure of polymerized tubulin bound to EpoA (1TVK.pdb) (Nettles *et al.,* 2004). The remaining 5 and 12 tubulin models for the EpoA – tubulin and the BacIII – tubulin complexes, respectively, resulted from previous docking calculations of EpoA and BacIII to tubulin and belonged to complex models where the ligands are found in very different orientations in the binding pocket.

Table 3.1: List of the intensities of the intermolecular NOEs between protons of baccatin III and epothilone A in the presence of tubulin at four mixing times. B and E indicate baccatin and epothilone, respectively. B-o, B-m, B-p indicate the protons in the *orto*-, *meta*- and *para*- positions of the C2 benzoyl phenyl ring of baccatin. The NOE intensities are normalized relative to the diagonal peak intensity of proton A and multiplied by 10^4.

		NOESY mixing times			
Proton A	Proton B	40ms	70ms	100ms	200ms
B-m	E-21	8	14	26	68
B-p	E-21	9	10	25	81
B-o	E-21	4	12	19	60
E-21	B-16	8	9	15	50
E-21	B-17	5	11	17	65
B-m	E-27	5	16	24	70
B-p	E-27	4	17	23	75
B-o	E-27	3	10	22	65
B-m	E-22	5	13	20	57
B-p	E-22	3	10	14	61
B-o	E-22	5	10	16	50
B-m	E-23	4	9	19	50
B-p	E-23	0	7	18	53
B-o	E-23	4	8	15	43
B-m	E-25	4	15	19	66
B-p	E-25	7	14	21	72
B-o	E-25	3	12	21	57
E-19	B-16	8	13	24	62
E-19	B-17	10	16	23	85
B-10	E-25	0	7	9	48
B-3	E-22	0	7	14	49
B-5	E-22	5	9	14	41

E-15	B-17	9	15	24	82
E-17	B-17	12	21	25	78
E-15	B-16	6	15	30	67
B-5	E-24	0	0	9	31
B-m	E-24	3	13	19	49
B-p	E-24	0	9	11	56
B-o	E-24	3	7	14	44
B-5	E-23	0	4	9	40
B-2	E-23	0	4	9	17
B-10	E-23	0	5	9	26
B-13	E-27	0	6	19	64
B-5	E-27	2	7	8	43
B-2	E-27	0	7	7	35
B-10	E-21	4	11	12	34
B-13	E-21	0	7	9	31
B-o	E-12	4	3	9	21
B-o	E-13	0	3	4	24
B-m	E-7	0	3	6	16
B-o	E-7	0	3	4	15
B-m	E-15	0	0	6	22
E-21	B-2	0	1	7	19
B-5	E-17	0	0	5	13
B-2	E-17	0	0	0	18
B-m	E-17	0	5	6	22
B-p	E-17	0	0	6	22
B-o	E-17	0	5	6	17
B-m	E-19	0	6	12	29
B-o	E-19	0	5	9	20
E-15	B-o	6	16	28	69

Increasing flexibility was allowed for the protein during the docking protocol. After a first stage of rigid docking, structural models were selected having the ligand close to the taxane binding pocket. In the following stages, we allowed the protein side chains in the binding pocket and subsequently also the loop backbone to be flexible. In the last water refinement step full flexibility of the protein was allowed and the complex was hydrated with a water layer of about 8Å. This procedure was designed to access conformational states of tubulin that could differ to some extent from the conformation of tubulin polymerized on Zn-stabilized sheets.

The conformations of EpoA and BacIII were kept rigid throughout the docking calculation. For EpoA we used the tubulin-bound conformation determined by transferred-NOEs (tr-NOE) and transferred cross-correlated relaxation rates (tr-CCR). These data were complemented by the measurement of the C19-C18-C17-

C16 torsion angle by the CH-CH dipolar-dipolar CCR rate between the C19-H19 vector and the C17-H17 vector. As a consequence of the new CCR data, the C19-C18-C17-C16 torsion angle was restrained to the range –40° - +40°. The tubulin-bound conformation of BacIII was verified by tr-NOE data and was found to be equivalent to the conformation of the baccatin core of paclitaxel in the EC-derived structure (Nogales *et al.*, 1998).

The charges of both ligands were calculated in *vacuo* by MOPAC from Insight II (Stewart, 1990) using the MINDO/3 Hamiltonian (Bingham *et al.*, 1995). As for the protein, implicit hydrogen bonding was introduced for hydrogen donors and acceptors of the ligands. The ligands non-bonded parameters were adapted to fit the PARALLHDG_5.3 parameter set from Michael Nilges (July 07, 2001 University of California) as provided with HADDOCK 1.0.

Intermolecular NOEs calculations. The peak volumes in the mixing time of a NOESY spectrum evolve according to the differential equation:

$$\frac{d\mathbf{A}(t)}{dt} = -(\mathbf{R}+\mathbf{K})\cdot\mathbf{A}(t) \qquad \text{(Eq. 3.4)}$$

where $\mathbf{A}$ denotes the square matrix of the peak volumes, $\mathbf{R}$ is the relaxation matrix and $\mathbf{K}$ the chemical exchange matrix (Ni, 1996). Solution of Eq. 3.4 leads to the matrix $\mathbf{A}(t_m)$, which contains the intensities of all peaks in a NOESY spectrum with mixing time t_m. The formal solution is

$$\mathbf{A}(t_m) = \exp[(\mathbf{R}+\mathbf{K})\, t_m]\cdot\mathbf{A}(0) \qquad \text{(Eq. 3.5)}$$

where $\mathbf{A}(0)$ contains the initial magnetization.

We describe the system with a two-state kinetic model (Scheme 3.2), where we assume that the concentration of the free protein is negligible due to the large excess of ligands. Four chemical species are considered: free ligand A (EpoA), free ligand B (BacIII), complex TA (EpoA – tubulin) and complex TB (BacIII – tubulin).

$$\mathbf{B + TA} \underset{k_{21}}{\overset{k_{12}}{\rightleftharpoons}} \mathbf{A + TB}$$

Scheme 3.1

The relaxation matrix $\mathbf{R}$ contains the relaxation rates between protons in the four distinct molecular species. The general form of the relaxation matrix is:

$\mathbf{R}_A^A$					
	$\mathbf{R}_B^B$				
		$\mathbf{R}_A^{TA}$	$\mathbf{R}_{A,T}^{TA}$		
		$\mathbf{R}_{T,A}^{TA}$	$\mathbf{R}_T^{TA}$		
				$\mathbf{R}_B^{TB}$	$\mathbf{R}_{B,T}^{TB}$
				$\mathbf{R}_{T,B}^{TB}$	$\mathbf{R}_T^{TB}$

The bold lines group the sub-matrices corresponding to each one of the four species in the solution. The superindex of each $\mathbf{R}_x^s$ indicates the species (either the free ligands or the complexes TA or TB) and the subindex indicates which protons of the species contribute to that sub-matrix. Thus, $\mathbf{R}_B^B$ describes the relaxation of the protons of ligand B in the free state and $\mathbf{R}_B^{TB}$ the relaxation of the protons of ligand B when bound in the complex TB. Analogously, $\mathbf{R}_T^{TB}$ describes the relaxation of the protein protons in the complex TB and $\mathbf{R}_{B,T}^{TB}$ contain the cross-relaxation terms between the protons of ligand B and the protons of protein T in the complex TB.

The individual sub-matrices representing each of the four chemical species are symmetric square matrices of dimension $N^s \times N^s$, where N^s is the number of

protons considered for each of the four different species s. To reduce the matrix sizes and be able to perform the calculation in a reasonable time, we considered only tubulin atoms within a given distance d from the ligand.

The matrix **K** represents the kinetics of the species in equilibrium and has the same dimension as the relaxation matrix **R** (Ni, 1996):

k_{21}[TB] **I**		- k_{12}[B] **I**			
	k_{12}[TA] **I**			- k_{21}[A] **I**	
- k_{21}[TB] **I**		k_{12}[B] **I**			
			k_{12}[B]**I**		- k_{21}[A] **I**
	- k_{12}[TA] **I**			k_{21}[A] **I**	
			- k_{12}[B] **I**		k_{21}[A] **I**

I indicates the identity matrix and [A], [B], [TA] and [TB] are the concentrations of the corresponding species.

Structure selection criteria. The theoretical intermolecular NOEs between the two ligands EpoA and BacIII were calculated for each of the 26.7 millions pairs of docking models (5315 for the EpoA – tubulin complex and 5022 for the BacIII – tubulin complex) assuming a k_{off} > 10000 Hz for both complexes. This lower bound of the k_{off} was verified by relaxation dispersion experiments. To reduce the matrix sizes, we considered only tubulin atoms in a distance d < 8 Å from the

ligand. By gradually increasing this limit, we verified that 8 Å is the optimal value to obtain precise interligand NOEs while keeping the calculation time at reasonable values.

The theoretical intermolecular NOEs were compared with the experimental values by a linear regression algorithm. The best pair of models delivered a linear correlation coefficient *R* of 0.92. Three criteria were applied in the selection of pairs of models: first, the linear correlation coefficient has to exceed 0.86; second, the slope of the regression lines of the experimental *vs.* the calculated data must be confined between 0.3 and 1.4; third, the slope of the regression line of each NOE peak does not exceed two times the slope of the regression line of all the NOE peaks together. The second selection criterion considers that the calculated NOE intensities, normalized to the diagonal peak intensities, should have the same size as the experimental ones. The upper bound was set to 40%. The lower bound allows the experimental NOEs to be 70% smaller than the calculated ones. The logic behind these criteria is based on the presence of internal motions, which could significantly reduce the experimental intermolecular NOEs. Internal motions cannot be simulated in the theoretical calculations, due to the lack of a satisfactory model for such a complex system. However, motion is not likely to increase the NOE by the same amount, which led us to set the upper bound of the slope to 40%. The presence of internal motion is also the likely explanation why the linear regression correlation coefficient does not exceed 0.92 for any pair of models.

The 25 million pairs of models were clustered to facilitate the analysis. The pairwise RMSD of models belonging to the same cluster did not exceed 2 Å on EpoA heavy atoms after translation of their geometric center of mass into the origin of the coordinate system. After application of the selection criteria two families of pairs were considered further, pairs A and B. Pair A is represented by the EpoA – tubulin model of Fig. 4.5 and the BacIII – tubulin model of Fig. 4.9. Pair B was discarded due to the fact that the aromatic ring of baccatin in the BacIII – tubulin complex does not occupy any of the pockets where the aromatic rings of paclitaxel are found in the EC-derived model of the PTX – tubulin complex (Nogales *et al.,* 1998). The orientation of the baccatin core also differs from those found in both the EC-derived model (Nogales *et al.,* 1998) and the

REDOR/FRET-derived model (Li *et al.*, 2000; Ivery & Le, 2003) of the PTX – tubulin complex. In summary, the binding mode of BacIII to tubulin in pair B contradicts the structural data available for the interaction of taxanes with tubulin and was not considered further.

The EpoA – tubulin and BacIII – tubulin models of pair A were further refined through 230 ps of molecular dynamics, which improved the correlation coefficient of 0.02.

Chapter 4

Results

4.1 The structure of the tubulin-bound epothilone.

The conformation of epothilone A bound to tubulin was determined by NMR spectroscopy in aqueous solution by a combination of tr-NOEs and tr-CCR rates data (Carlomagno *et al.*, 2003b). At the time we performed this work, this structure provided the first and only piece of experimental information on the active conformation of this class of microtubule stabilizers. In the meantime, the structure of tubulin assembled on bidimensional sheets stabilized by Zn and bound to epothilone was solved by electron crystallography at 3 Å resolution (Nettles *et al.*, 2004). This structure is considerably different from the one derived by NMR in solution, as it will be discussed in more details in the next chapter. Leaving aside any consideration on the resolution of the EC structure, which might be too low to derive the conformation of a ligand of dimension 12 X 9 X 4 Å, the physical state of tubulin is different in the two studies, indicating that epothilone might bind to tubulin in a polymerization state – dependent manner. In this chapter I describe the results of our NMR study of the bioactive conformation of epothilone A in solution, while the differences with the EC derived structure are discussed in Chapter 5.

Transferred nuclear Overhauser enhancement (tr-NOE) and transferred cross-correlated relaxation (tr-CCR) methods have been employed to unravel the conformation of epothilones in complex with tubulin (see chapters 2 and 3). Fortunately, the kinetic and thermodynamic properties of the complex between epothilone A and tubulin are in the desirable range for tr-NOE and tr-CCR experiments, as indicated by the negative sign of the NOE cross peaks in a 100:1 mixture of epothilone A with tubulin. The tr-NOE and tr-CCR data are in agreement with a k_D in the range of 100 μM. Both epothilone A and B, the two major naturally occurring epothilone variants, displace ^{3}H-paclitaxel from tubulin (K_i equals 1.4 μM and 0.71 μM for epothilone A and epothilone B, respectively)

(Kowalski *et al*., 1997), indicating that the tubulin binding sites of the epothilones and paclitaxel overlap. In our NMR system, the observation of extensive interligand NOEs between epothilone and taxotere with the INPHARMA method confirms the binding of the two drugs to the same binding pocket.

A crucial provision for the structural investigation of the epothilone-tubulin complex in solution is the prevention of tubulin polymerization in microtubules. Electron microscopy and gel electrophoresis show that substitution of Mg^{2+} with Ca^{2+} (Solomon, 1977) and D_2O as solvent (Chakrabarti, 1999) are sufficient to prevent tubulin polymerization. In this conditions and in the absence of both EpoA tubulin exists as a mixture of higher molecular weight aggregates of the α,β – dimers, as observed by electron microscopy (EM) (Fig. 2.2A) and size-exclusion chromatography. The tubulin "aggregates" are disordered polymers, and no microtubules or ordered polymers are formed. The addition of a large excess of EpoA (12μM tubulin and 500 μM EpoA) promotes polymerization of tubulin in the form of tubules sheets (Fig. 2.2B). The NMR derived tubulin-bound conformation of EpoA refers to the "aggregate" form of tubulin. The average correlation time τ_c of the tubulin "aggregates" is approximately 1.3 μs, which roughly corresponds to 15 protein dimers. The small percentage of EpoA bound to the tubules sheets does not contribute to the observed NMR signal, because EpoA binds strongly to the microtubules (and consequentely to the tubules sheets) (Buey *et al.*, 2004) and disappears from the NMR spectrum. In agreement with this consideration we do not observe any change in the EpoA NOE signals with time, while the microtubules sheets precipitate out of solution in a few days. Moreover, the average correlation time τ_c of the tubulin-bound epothilone observed in the NMR spectra is approximately 1.3 μs, which roughly corresponds to 15 protein dimers, while the microtubules sheets observed in the EM images contain several hundreds of dimers. The average correlation time of 1.3 μs is thus incompatible with the hypothesis that the microtubules sheets measurably contribute to the observed NMR signal.

The relevance of determining the binding mode of EpoA to tubulin "aggregates" is evident, as the EM images prove that epothilones are able to induce the formation of ordered tubulin polymers (Fig. 2.2B) by binding to the disordered "aggregates" (Fig. 2.2A). Moreover unpolymerized tubulin is always present in

the form of "aggregates" (Fig. 2.2A), underlying the biological importance of such disordered polymers in the conformational equilibrium of tubulin.

The tubulin bound conformation of epothilone A resulted from the analysis of both tr-NOEs and tr-CCR rates data. The NOESY spectra of EpoA in the presence and in the absence of tubulin are shown in Fig. 4.1. A comparison of the two spectra shows that the NOEs observed for the EpoA - tubulin mixture generate from the tubulin-bound form of epothilone. 46 inter-proton distance restraints were derived from transferred NOE experiments. To filter out spin-diffusion mediated peaks, only those signals, which were present with opposite sign to the diagonal peaks in a transferred ROESY experiment, were taken into account.

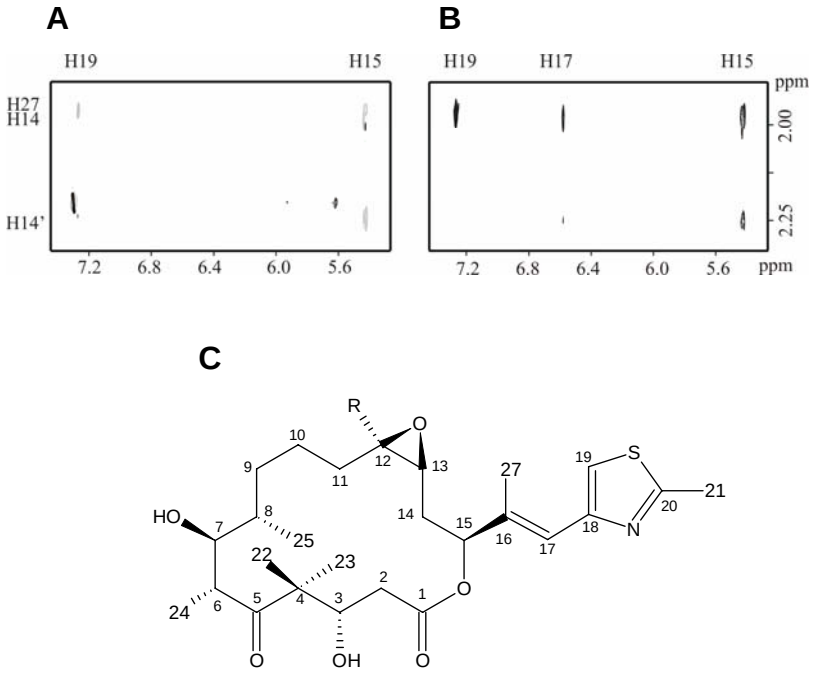
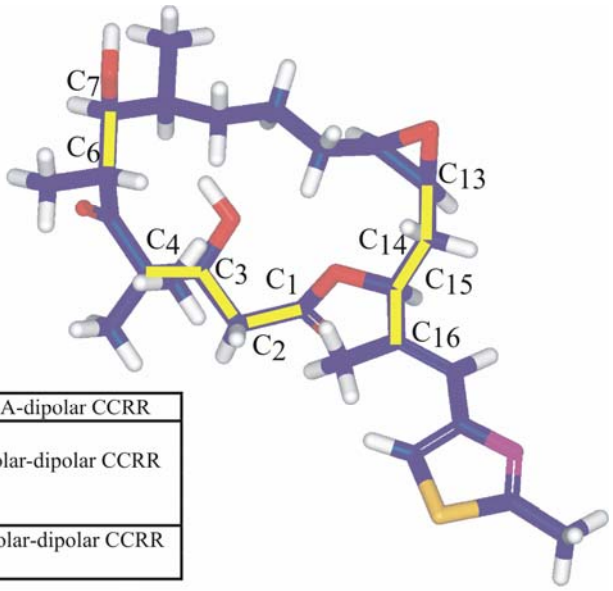


Fig. 4.1. NOESY spectra at 300 ms mixing time for epothilone A in D_2O (a) and for the mixture epothilone A – tubulin 100:1 in D_2O (b). The change in sign of the NOE peaks from positive (negative grey countours, $\tau_c < 1.12/\omega_H$) in A) to negative (positive black contours, $\tau_c > 1.12/\omega_H$) in B) proves that the NOE signals of the mixture epothilone A – tubulin B) are dominated by the contribution of the bound form of epothilone. C) Epothilone with atom numbering; R = H for epothilone A and R = CH_3 for epothilone B.

Seven dihedral angle restraints were obtained by measuring CH-CH dipolar-dipolar and CH-CO dipolar-CSA transferred CCR rates for 60-70% ^{13}C labeled epothilone A. The tr-CCR experiments were indispensable to obtain a unique

description of the bound conformation, as more than one structure of the macrolide ring is compatible with the same H-H distance set (NOE intensities). The seven tr-CCR rates defined the torsion angles O11-C1-C2-C3, C1-C2-C3-C4 (Fig. 4.2), C2-C3-C4-C5, C5-C6-C7-C8, C12-C13-C14-C15, C13-C14-C15-O11 and C14-C15-C16-C17.



O_1-C_1-C_2-C_3	CO-CH CSA-dipolar CCRR
C_1-C_2-C_3-C_4 C_5-C_6-C_7-C_8 C_{12}-C_{13}-C_{14}-C_{15} C_{13}-C_{14}-C_{15}-O_1	CH-CH dipolar-dipolar CCRR
C_1-C_2-C_3-C_4 C_{14}-C_{15}-C_{16}-C_{17}	CH-CH_3 dipolar-dipolar CCRR

Fig. 4.2 Summary of the CCR rates that were measured to determine the tubulin-bound conformation of EpoA and of the dihedral angles that were derived from these tr-CCR rates.

The tubulin-bound conformation of epothilone A is shown in green in Figure 4.3 and is compared with the free (unbound) conformation of epothilone A determined by X-ray crystallography (Rihs, G. and Walter, H.R. unpublished results), which is shown in gray. We choose to compare the tubulin-bound conformation of epothilone with the X-ray structure and not with the solution structure available in CD_2Cl_2, because of the extensive flexibility of epothilone in solution in the absence of tubulin (Taylor & Zajicek, 1999). However, the most populated conformer in solution is indeed very similar to the X-ray conformer. The position of the thiazole nitrogen (blue in Fig. 4.3) and 3-OH (red in Fig. 4.3), which are important for the delineation of a pharmocophore model, change significantly upon binding.

A comparison of the torsion angles of epothilone A in the tubulin-bound (green in Fig. 4.3) and in the free (gray in Fig. 4.3) conformation (Table 1) reveals two major changes: the first occurs in the region O1-C6 and the second affects the

orientation of the thiazole ring with respect to the C16-C17 double bond. In the lower region of the macrocycle both cross-correlated relaxation rates and NOE data define the conformational change of the dihedral angle C2-C3-C4-C5, which switches from a *gauche*$^+$ to a *gauche*$^-$ conformation, while the variations of the dihedral angles O1-C1-C2-C3 (+ *anti-periplanar* to – *anti-clinal*) and C3-C4-C5-C6 result from the adjustment of the macrocycle to the modified torsion around the C3-C4 bond. The observed conformational change in the O1-C6 region of the ring primarily affects the position of the protons at C2, H2 and H2a, and of the 3OH group. The two protons at C2 move towards the inner part of the ring. At the same time the oxygen attached to C3, which points to the inside of the macrocycle in the free conformation, moves by 3.8 Å towards the exterior (i.e. it now points away from the macrocycle).

Neither the dihedral angle C5-C6-C7-C8, determined by tr-CCR, nor the C10-C15 region exhibit any significant conformational change upon binding to tubulin. The latter finding corroborates a recent study where the conformation of the macrocycle in the epoxide region is investigated using analogues of epothilone B and D (Taylor *et al*., 2003).

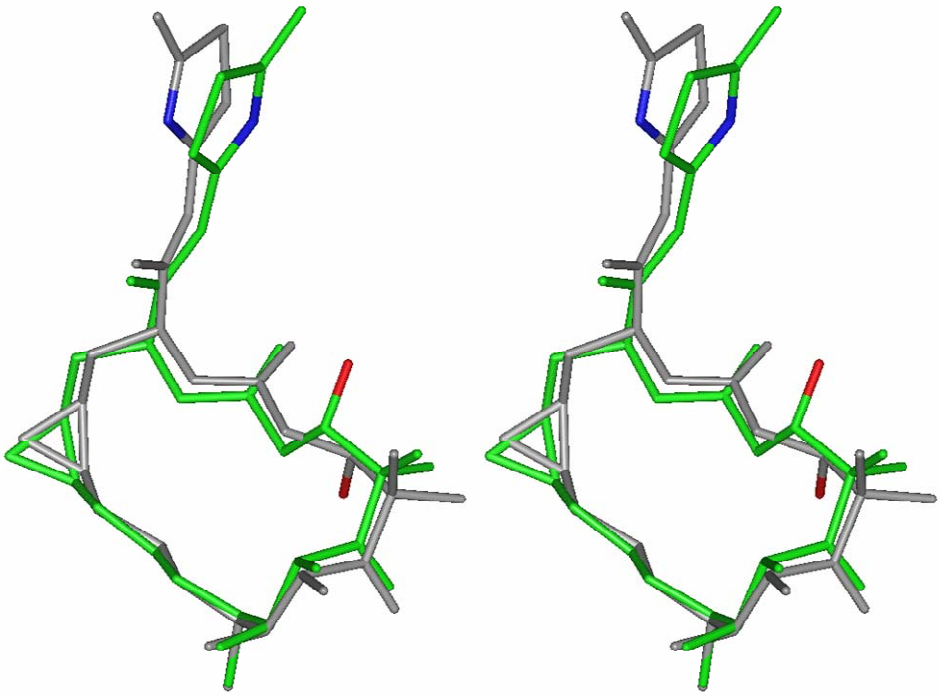


Fig. 4.3 Comparison of the tubulin-bound conformation of epothilone A in aqueous solution (green, this study) with the free conformation determined by X-ray crystallography (gray). The

structures can be viewed in stereo. Hydrogen atoms have been removed for clarity. The conformational change that occurs in the C1-C5 region is clearly visible, and it is associated with a significant shift in the position of the 3-OH group (in red). A second key feature of the bound structure is the 180º degree change in the orientation of the side chain, which liberates the nitrogen (in blue) of the thiazole ring from the bulky Me 27 group and makes it accessible to hydrogen bond donors.

The second significant change between the free and tubulin-bound conformations of epothilone occurs in the side-chain bearing the thiazole ring. The dihedral angle C16-C17-C18-C19 changes from an *anti-periplanar* to a *syn-periplanar* conformation. This finding is substantiated by a strong NOE signal between the protons of Me27 and H19 together with a significant weakening of the NOE between H17 and H19 compared with that observed for epothilone A in aqueous solution and in the absence of tubulin. The eclipsed rotamer is energetically disfavored due to the steric hindrance between Me27 and H19. However, in this conformation the nitrogen of the thiazole ring becomes more accessible for potential hydrogen bond formation with protein functional groups or with water, which may more than offset an intrinsically disfavored epothilone side-chain conformation.

4.2 The binding mode of epothilone to soluble tubulin.

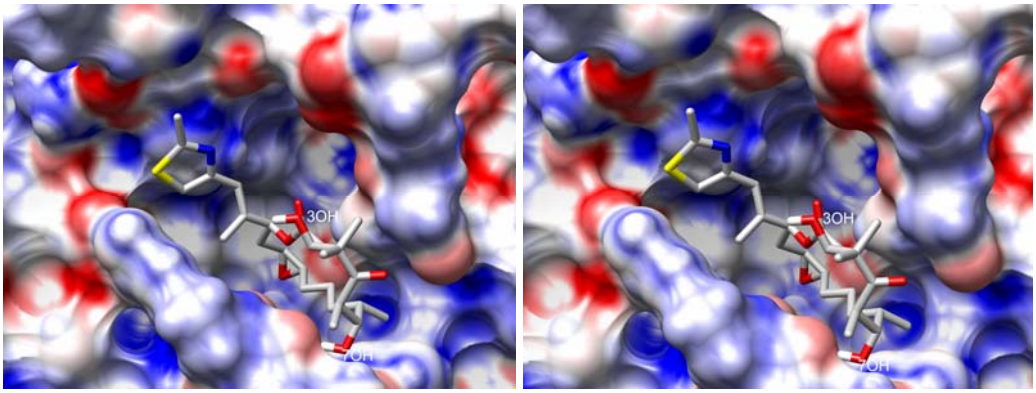
In this work we derived a structural model for the binding mode of EpoA to tubulin in solution (Fig. 4.4) by using a combination of the novel INPHARMA methodology (Sanchez-Pedregal *et al.,* 2005) and molecular modeling. Briefly, the INPHARMA methodology relies on the observation of inter-molecular tr-NOEs between two ligands binding competitively and consecutively to the same binding site of a common target protein (see Chapter 3). These NOEs are exclusively mediated by spin-diffusion through protons of the protein and are therefore dependent on the binding mode of each of the two ligands to the protein. An accurate and quantitative evaluation of such effects leads to the determination of the binding mode of the ligands to the protein.As the experimental system consists of two ligands binding competitively and mutually exclusively to tubulin,

we describe the system with a *pair* of structural models, one for the EpoA - tubulin and one for the BacIII - tubulin complexes.

Table 1. Comparison of the dihedral angles of the tubulin-bound conformation and the free conformation of epothilone A. The values given for the tubulin-bound conformation result from averaging over the 10 lowest energy conformations. The values given for the free conformation result from crystal structure analysis of epothilone A in methyl-tert. butylether ($CH_3OC(CH_3)_3$) solvate.[30] This conformation as well as crystal packing is nearly identical to that of epothilone B[2] although the two compounds crystallize from different solvents.

Dihedral angle	tubulin-bound conformation epothilone A (NMR)	free conformation epothilone A (X-ray)
C1-C2-C3-C4	–152.5° ± 0.2°	165.4°
C2-C3-C4-C5	–51.7° ± 0.1°	73.0°
C3-C4-C5-C6	–43.0° ± 1.8°	–75.9°
C4-C5-C6-C7	156.4° ± 0.5°	145.6°
C5-C6-C7-C8	–70.0° ± 0.8°	–64.0°
C6-C7-C8-C9	–74.8° ± 0.3°	–79.2°
C7-C8-C9-C10	164.1° ± 0.8°	159.0°
C8-C9-C10-C11	–171.9° ± 0.4°	176.8°
C9-C10-C11-C12	–178.0° ± 0.4°	174.8°
C10-C11-C12-C13	–129.2° ± 0.5°	–112.8°
C11-C12-C13-C14	4.1° ± 0.1°	–1.9°
C12-C13-C14-C15	76.3° ± 0.1°	93.9°
C13-C14-C15-O1	–62.6° ± 1.0°	–156.9°
C14-C15-O1-C1	179.5° ± 0.5°	159.6°
C15-O1-C1-C2	176.3° ± 1.3°	174.2°
O1-C1-C2-C3	–124.3° ± 1.2°	156.7°
C14-C15-C16-C17	–129.7° ± 1.3°	–118.5°
C15-C16-C17-C18	180.0 ° ± 0.2°	176.0°
C16-C17-C18-N	180.0 ° ± 0.3°	–7.6°

A



B

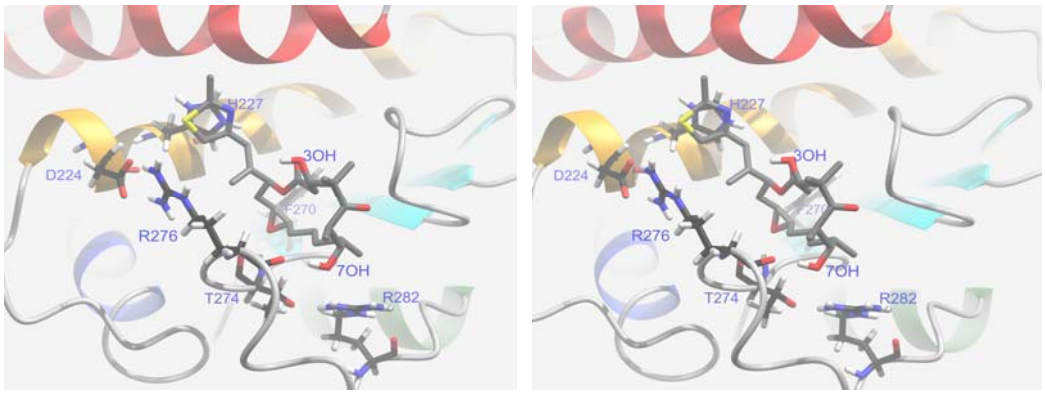


Fig. 4.4. Stereo views of the model of the EpoA – tubulin complex that best reproduces the experimental interligand NOEs in combination with the BacIII – tubulin complex of Fig. 4.8. In panel A, the solvent accessible surface colored by the electrostatic potential of tubulin is shown. In panel B, the thiazole ring of epothilone (white, C; red, O, blue, N and yellow, S) interacts with the aromatic ring of H227, while the 7-OH is situated close to T274 and R282. The H1 helix of tubulin is colored in red, the H2 helix in yellow. The hydrophobic stretch C8-C12 of epothilone is directed towards the hydrophobic floor and lower right side of the binding site. The aliphatic side chain of R276 entertains hydrophobic contacts with the EpoA side chain, while the guanidinium group interacts face-to-edge with the thiazole ring and enjoys a favorable electrostatic interaction with the negatively charged side chain of D224.

After application of selection criteria used to evaluate the agreement between the calculated and experimental NOEs in each pair of EpoA – tubulin and BacIII – tubulin complex models (see Chapter 3 for details), one family of model pairs was uniquely identified. The representative model of the EpoA – tubulin complex in this family that uniquely satisfies the experimental intermolecular tr-NOEs is shown in Fig. 4.4 and represents the NMR-determined binding mode of EpoA to

soluble tubulin. We can exclude that the recently published EC-derived binding mode of EpoA to tubulin (Nettles *et al.*, 2004) is present in solution, as this model provides a very poor agreement between calculated and experimental intermolecular tr-NOE data in combination with any of the 5000 binding modes of BacIII. As it will be discussed below, the complex displayed in Fig. 4.4 provides a good explanation of the SAR data obtained for epothilone analogues and gives an understanding of the observed activity data available for tubulin mutants.

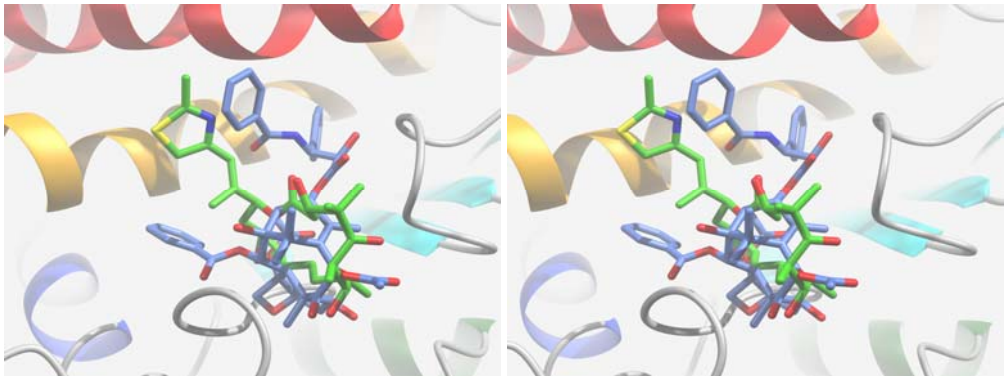


Fig. 4.5. Stereo view of the superposition of the NMR-derived EpoA (green) binding mode and EC-derived paclitaxel (light blue) binding mode to β-tubulin (Nogales *et al.*, 1998): The thiazole ring of EpoA roughly overlaps with the C3' benzamido phenyl of paclitaxel. Additionally, the 3-OH of EpoA is close to the 1'-CO and 2'-OH of paclitaxel.

Most importantly, the pharmacophore of the NMR-derived binding mode of EpoA to tubulin partially resembles that of the EC-derived binding mode of paclitaxel (Nogales *et al.*, 1998) (*vide infra*), with the addition of a few intermolecular interactions that are unique to the EpoA – tubulin complex. This is in contrast to the EC-derived model of the EpoA – tubulin complex (Nettles *et al.*, 2004), where no common feature is observed between the pharmacophores of EpoA and paclitaxel.

The position of the thiazole ring in the NMR-derived structure of the Epo A – tubulin complex is close to that of the C3'-benzamido phenyl in the PTX – tubulin complex (Fig. 4.5). The thiazole ring is involved in aromatic - aromatic interactions with the ring of H227 (Fig. 4.4B). The thiazole and the histidine rings are stacked in a face-to-face orientation, though rather tilted, at an angle of *ca.*

30°. This result is in good agreement with the observation that the protonated histidine rings usually interact with tryptophane rings in a face-to-face orientation in protein hydrophobic cores (Sumanta *et al.,* 1999).

The intermolecular contacts involving the hydroxyl group on the C7 of EpoA represent a unique feature of the EpoA – tubulin complex. In our model the positively charged side chain of R282 entertains extensive electrostatic contacts with the polarized C7-OH of EpoA (Fig. 4.6). This interaction locks the side chain of R282 in an optimal position for forming a H-bond with the side chain of T274, thereby stabilizing a specific conformation of the tubulin M-loop, which might be relevant for microtubule stabilization. The involvement of T274 and R282 in this network of electrostatic interactions justifies the loss in the activity of epothilones in cell lines containing either the T274I or the R282Q mutation (40-fold and 57-fold activity loss, respectively) (Giannakakou *et al.*, 2000). This pattern of electrostatic interactions is exclusively observed in the EpoA – tubulin complex and is not present in the PTX – tubulin complex, as confirmed by the higher tolerance of paclitaxel to the T274I and R282Q mutations (Giannakakou *et al.*, 2000).

The aliphatic side chain of R276 entertains hydrophobic contacts with the EpoA side chain, while the guanidinium group interacts face-to-edge with the thiazole ring and enjoys a favorable electrostatic interaction with the negatively charged side chain of D224. This contact, together with the electrostatic interaction of the C7-OH with R282 and the aromatic-aromatic interaction of the thiazole ring with the H227, locks the position of the helix H7 with respect to the M loop. As I will discuss in chapter 5 the role of EpoA as bridging molecule between the helix H7 and the M loop of tubulin is relevant to the biological function of the drug.

The epoxide oxygen of EpoA is not involved in extensive protein – ligand interactions, in agreement with the high activity of the unsaturated analogues EpoC and EpoD. The EpoA stretch C9-C12 enjoys hydrophobic interactions with the leucine-rich floor of the protein pocket including P272, L273, L284 and L369 (Fig. 4.7). The epoxide ring is situated at the entrance of the hydrophobic pocket hosting F270; however, epothilone does not enter this pocket, as opposed to paclitaxel, which locates its C2-benzoyl phenyl close to F270. In accord with the NMR-derived structural model of the EpoA – tubulin complex, epothilones retain

their activity in cell lines exhibiting the F270V mutation, while the activity of paclitaxel is decreased by 27-fold (He *et al.,* 2001).

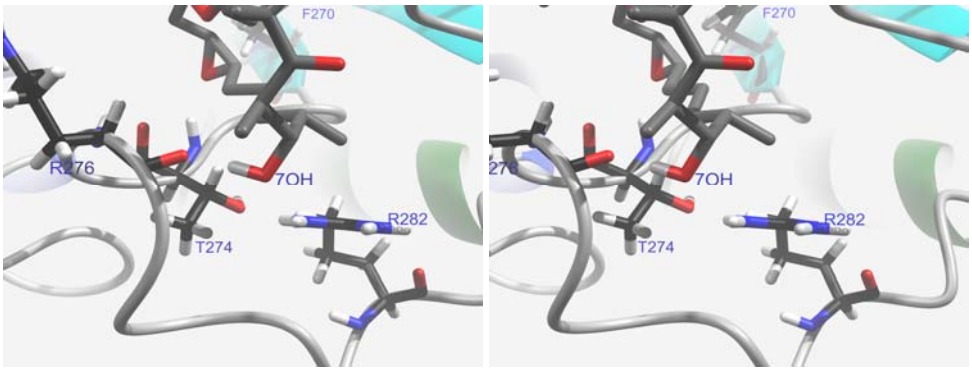


Fig. 4.6. Interactions of the 7-OH of EpoA with R282 and T274 of tubulin. Electrostatic interactions between the guanidinium group of the arginine and the EpoA 7-OH define a conformation of the M-loop, which favors the formation of a H-bond between R282 and T274.

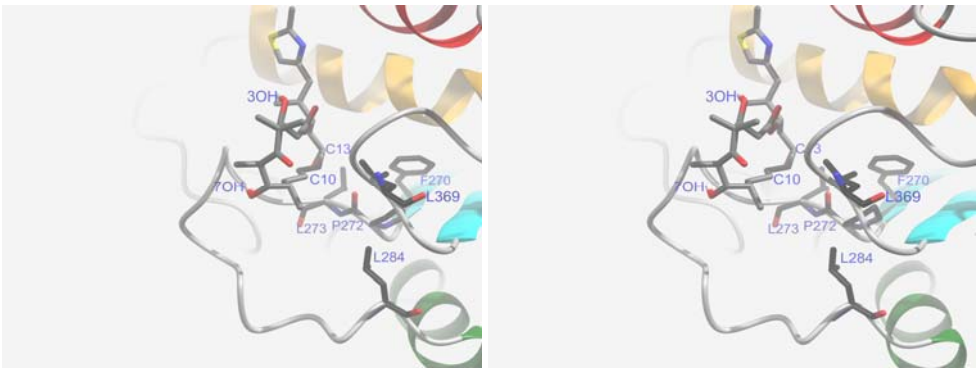


Fig. 4.7 Stereo view of the model of the EpoA – tubulin complex of Fig. 4.4, which shows the hydrophobic interactions of the EpoA stretch C9-C12 with the leucine-rich floor of the binding pocket including P272, L273, L284 and L369. The epoxide ring is situated at the entrance of the hydrophobic pocket occupied by F270.

The model of the BacIII – tubulin complex that provides a good fit of the experimental interligand NOEs with the EpoA – tubulin complex model of Fig. 4.4, is shown in Fig. 4.8. In contrast to our expectation, the baccatin binding mode obtained by the EC-derived PTX – tubulin complex (Nogales *et al.,* 1998) after removal of the side chain on C13 does not fit the experimental interligand NOE data in combination with any of the EpoA – tubulin complex models ($R \leq 0.75$). Instead, docking models where baccatin places the C2 benzoyl phenyl ring in the pocket occupied by the C3' phenyl in the PTX – tubulin complex, close to F270,

are consistent with the INPHARMA data. The orientation of the baccatin core in the BacIII – tubulin complex of Fig. 4.8, however, is similar to the one proposed for the PTX – tubulin complex on the basis of FRET and REDOR data (Li *et al.,* 2000; Ivery and Le, 2003). The fact that the binding mode of baccatin to tubulin differs from the one of paclitaxel is not surprising, considering that baccatin lacks the C3' benzamido phenyl, which plays a fundamental role in the determination of the binding mode of paclitaxel to tubulin by interacting with H227. It should be noted that our model of the BacIII – tubulin complex explains the enhanced activity of 2-*m*-azido baccatin by invoking the interaction of the azido group with the guanidinium group of R320 (Ivery & Le, 2003). However, we do not exclude that 2-*m*-azido baccatin might have a binding mode that differs from that of baccatin, as allowed by the broad and flexible tubulin binding pocket and in agreement with the large difference in their biological activity (He *et al.,* 2000).

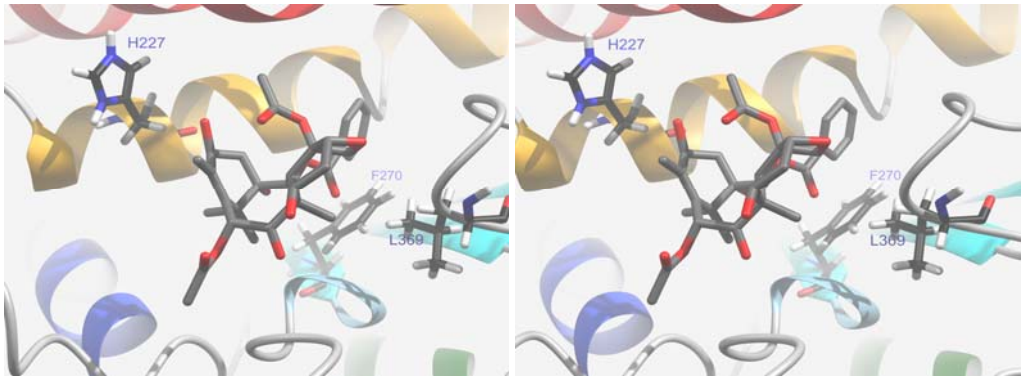


Fig. 4.8 Model of the BacIII – tubulin complex that best reproduces the experimental interligand NOEs in combination with the EpoA – tubulin complex of Fig. 4.4. The orientation of the baccatin core differs from the one observed in the EC-derived model of the PTX – tubulin complex (Nogales *et al.,* 1998) and is similar to that proposed on the basis of FRET and REDOR data (Li *et al.,* 2000; Ivery and Le, 2003). The C2 benzoyl phenyl is situated in the hydrophobic pocket defined by F270 and L369, where paclitaxel places the C3' phenyl in the EC-derived complex model (Nogales *et al.,* 1998).

Chapter 5

Discussion

5.1 The structure of the tubulin-bound epothilone.

The tubulin bound conformation of epothilone derived by NMR is solution (Fig. 4.3) is consistent with the wealth of chemical modification data available for this class of compounds. The SAR data available for epothilones indicate that the C4-C8 region and the side-chain are critical for binding.

The hydroxyl group attached to C3 is one of the biologically most relevant sites. The inversion of the stereochemistry at C3 results in a considerable loss of potency in the induction of tubulin polymerization (Nicolaou *et al.*, 1997). On the other hand, α,β-unsaturated (trans double bond between C2 and C3) analogs of epothilones A and B are almost as active as the natural products. A very recent report by G. Vite also showed that the replacement of the hydroxyl group at C3 with a cyano moiety leads to a compound that retains high biological activity (Reguiero-Ren *et al.*, 2002). In the epothilone free conformation this OH group points towards the inner part of the ring and its oxygen atom is not easily accessible for the formation of external hydrogen bonds; conversely, in the tubulin-bound conformation the 3OH group points outwards, which potentially makes the oxygen a better acceptor of hydrogen bonding from tubulin side-chains. This observation is consistent with chemical modification data which underline the importance of the stereochemistry at C3 and therefore of the position of the 3OH group. On the other hand, the fact that α,β unsaturated epothilone analogues are almost as active as the natural compound indicates that the OH group is not crucial for binding to tubulin.

The second profound conformational change observed upon binding of epothilone to tubulin concerns the orientation of the thiazole side–chain, which undergoes a 180º rotation around the C17-C18 bond to liberate the nitrogen of the thiazole ring from the hindrance of the methyl group C27. It might be speculated that the nitrogen is now better available for hydrogen bond formation with a tubulin side-

chain. The potential role of the nitrogen as hydrogen bond acceptor is corroborated by chemical modification data: the tubulin polymerization activity is retained upon substitution of the thiazole ring with a 2-pyridyl moiety, while the incorporation of 3 or 4-pyridyl residue causes a reduction in the biological activity (Nicolaou *et al.*, 2000). However, substitution of the thiazole ring with a phenyl ring, lacking any H-bond acceptor, does not have any effect on the ability of epothilone to induce tubulin polymerization (Altmann, private communication), but is detrimentous only for the cellular activity of the drug. Taken together these facts do not support the relevance of an hydrogen bond, involving the N of the thiazole ring as acceptor, in the binding mode of EpoA to tubulin. Moreover, the electron distribution in heterocycles is complex and their interaction with other aromatic rings includes additional options than H-bonds, not disregarding solvation effects. In fact, NOESY spectra acquired for EpoA in water al low temperature and in absence of tubulin indicate that the dihedral angle C16-C17-C18-C19is in conformational equilibrium between the *syn-periplanar* and the *anti-periplanar* conformation. The presence of an H-bond between the nitrogen of the thiazole ring and tubulin side-chains could not be confirmed in the NMR derived model of the epothilone – tubulin complex (*vide infra*).

SAR data on epimers of 14-methyl EpoB and EpoD show that the *R* configuration for EpoB and the *S* configuration for EpoD retain most of the biological activity while the respective *S* and *R* configurations are inactive (He *et al.,* 2000; Taylor *et al.,* 2004). In agreement with these data, 14-methyl EpoB with *S* configuration at the C14 could not adopt the NMR-derived bioactive conformation due to steric hindrance with the Me27 and the H10 proton. Analogously, introduction of a methyl group at the C10 with *S* configuration leads to inactive EpoC, as expected from the steric clash of this additional methyl group with the protons at the C2.

The (*E*)-9,10-dehydro- and (*E*)-10,11-dehydro- analogues have been found to retain the biological activity of epothilone (Taylor *et al.,* 2003; Rivkin *et al.,* 2003), in agreement with the *trans* disposition around the C9-C10 and C10-C11 torsion angles of the NMR-derived bioactive conformation.

Recently, the structure of tubulin bound to EpoA, assembled into two-dimensional sheets stabilized by Zn^{2+} ions, was derived by electron crystallography (Nettles *et al.*, 2004). Interestingly, the tubulin-bound structure of epothilone A derived by

electron crystallography contradicts the NMR derived conformation of the compound bound to soluble tubulin (Fig. 5.1). Several dihedral angles, which were determined for the tubulin-bound conformation of EpoA by tr-CCR rates in the NMR study, strongly differ in the two conformations.

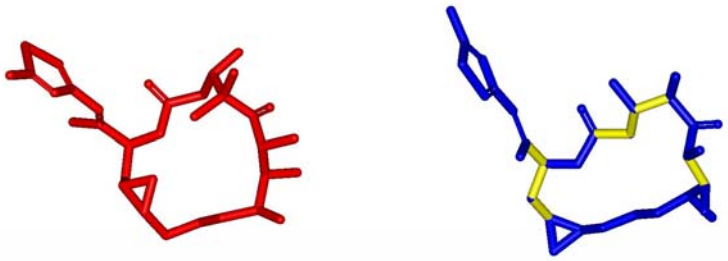


Fig. 5.1 Comparison of the tubulin-bound conformation of EpoA derived by electron crystallography (in red) (Nettles *et al*., 2004), for tubulin assembled in two-dimensional sheets stabilized by zinc, and that derived by NMR (in blue) (Carlomagno *et al*., 2003b), for soluble tubulin. The yellow bonds are those for which the torsion angles were determined by tr-CCR rates. The two conformations conspicuously differ.

However, in both conformations the 3OH points towards solution and the N of the thiazole side chain is not buried between the two methyl groups at C27 and C21.
The discrepancy between the two conformations can be explained by invoking the dependence of the conformation and the binding mode of EpoA on the assembly state of the protein. Such interplay is likely to be mediated by the conformation of the M loop, which undergoes a defined change between the aggregate form, present in the NMR experiments, and the polymeric form which is found in Zn^{2+}-stabilized sheets. Here, the M-loop is involved in lateral contacts with the helices H12 and H5 of a second β unit of tubulin (Nogales *et al*., 1998) (antiparallel assembly of the protofilaments). It should be noted that the lateral contacts of tubulin in Zn^{2+}-stabilized sheets differ substantially from those of the biologically relevant assembly state, i.e. the microtubules. In the latter, the tubulin protofilaments are arranged parallel and the lateral contacts are found between the

M loop of one β subunit and the S1-H2 loop (Fig. 1.1) and H3 helix of the other β subunit (Li *et al.*, 2002).

In this work, we have studied the binding mode of epothilone to soluble tubulin. With the help of NMR data and electron micrographs (Fig. 2.2. and Section 4.1) we have demonstrated that epothilone binds indeed to non-microtubular soluble tubulin to induce formation of elongated ordered polymers even in absence of Mg^{2+} and GTP. Therefore, the conformation derived here can be considered the first member of a series of different structures along the tubulin polymerization pathway of the epothilone – tubulin complex. It should be underlined that the EC-derived model cannot be conciliated with the NMR structural data, which allows us to exclude that this structure is populated in solution. The "aggregate", non-microtubular form of soluble tubulin is observed in micrographs of ligand-free, unpolymerized tubulin and to a lesser extent in micrographs of microtubules (Diaz & Andreu, 1993), which underlies the functional relevance of such disordered polymers in the conformational equilibrium of tubulin. The structure of epothilone bound to soluble tubulin, as well as the model of the epothilone – tubulin complex, represents the first functional step in the cascade of conformational change events, which lead to the formation of hyper stable microtubules.

5.2 The binding mode of epothilone to soluble tubulin.

The NMR-derived structural model for the EpoA – tubulin complex explains well the SAR data available on epothilones. We show this by discussing the most significant chemical modification data.

Replacement of the thiazole ring by various types of substituted pyridine rings is well tolerated (Nicolaou *et al.,* 2000); a simple phenyl group in place of the thiazole ring does not have a significant effect on the tubulin polymerizing activity of epothilone but causes a decrease of the epothilone activity of *ca.* one order of magnitude in cells (Altmann *et al*., 2000). A pyridine ring linked at position 2 is considerably preferred over a pyridine rings linked at position 3 or 4. These data were interpreted as an indication that the nitrogen of the thiazole ring accepts a H-bond from tubulin side chains. However, the electron distribution in heterocycles is complex and not predictable with simple models. Moreover, the

fact that a phenyl ring in place of the thiazole does not affect the tubulin polymerizing activity in a test-tube suggests that the nirogen of the thiazole ring might be important for the transport of epothilone into the cells and not for its binding to tubulin. In our complex no H-bond is formed between the nitrogen of the thiazole ring and the side chains of tubulin. Instead, the two proton-bearing nitrogens of the histidine ring are situated above the sulphur and the nitrogen of the thiazole ring respectively. This allows a favorable interaction between the delocalized charge of the His ring and the π orbital of the thiazole ring, which has a higher electron density at the sites of the S and N. Any perturbation of the position of the sulphur or nitrogen atoms results in a less favorable interaction. In addition, our model preserves a favorable contact between the aromatic rings of the histidine and the thiazole, which are in a tilted face-to-face orientation with respect to each other (Fig. 4.4), as often observed for protonated histidine rings in protein cores (Samanta *et al.*, 1999).

Substitution of the methyl group C21 of the thiazole ring with small polar or apolar substituents is well tolerated, while introduction of large substituents on C20 leads to inactive epothilones (Nicolaou *et al.,* 1998; Höfle *et al.*, 1999). In our model bulky substituents on C20 would distort the position of the side chains in the H1 helix and in the H2-B3 loop, thus disturbing the overall structure (Fig. 4.4).

Inversion of the stereochemistry of C3 is detrimental for biological activity (Nicolaou *et al.,* 1997), while a cyano group in place of the hydroxyl group is well tolerated (Regueiro-Ren *et al.,* 2002), as is a *trans* C2-C3 double bond. In our model, the 3-OH group is not involved in close contacts with the protein. Inversion of the stereochemistry of C3, however, might destabilize the bioactive conformation of EpoA, due to steric clash of the 3-OH group with the proton H6. A *trans* C2-C3 torsion angle is required for function, as confirmed by the biological activity of (*E*)-2,3-dehydro epothilone. Additionally, we observe that in the EC-derived structure of the PTX – tubulin complex, paclitaxel places the hydrophilic carbonyl and hydroxyl groups of the C1'-C2' stretch in about the same position as the 3-OH of EpoA (Fig. 4.5), without any evident contacts with tubulin (Nogales *et al.,* 1998). It has been proposed that paclitaxel stabilizes the lateral contacts of the β subunits by mimicking the binding of the eight extra

residues of the M-loop present in the α subunit. Binding of paclitaxel converts the hydrophobic binding site into a hydrophilic surface, similarly to the α subunit, where the backbone carbonyls of the inserted residues are presented on the surface. The possibility that the significance of the 3-OH group of EpoA resides in the generation of a hydrophilic protein surface, similar to that of the α subunit, remains to be investigated.

Ligand – protein interactions at the C7 site are confirmed by SAR data, which indicate that simultaneous inversion of the stereochemistry of C6 and C7, as well as removal of the methyl group at C8, causes a considerable loss of activity (Nicolaou *et al.*, 1998). In our model inversion of C7 would seriously disturb the H-bond network around R282 and T274, which represents a fundamental interaction of EpoA with the M-loop (Fig. 4.6). Additionally, inversion of the C6 would prevent EpoA from assuming its active conformation, due to steric hindrance between the methyl group at C6 and the H2 and H9 protons. Finally, the lack of the methyl group at the C8 would eliminate the positive hydrophobic interactions of EpoA with both L284 and L369 (Fig. 4.7).

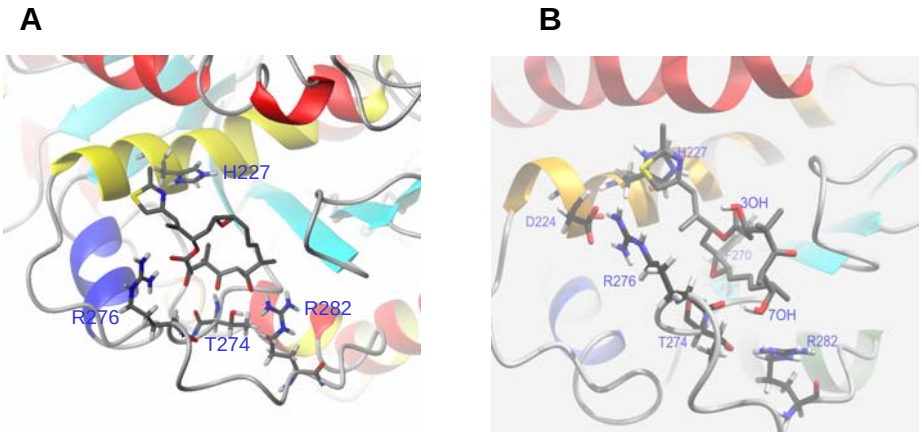


Fig. 5.2 Comparison of the EC derived model of the epothilone – tubulin complex (panel A) with the NMR derived model of the same complex (panel B).

Most importantly, our model accommodates very well the activity data for epothilones with structural modification on C12. EpoB, with a methyl group on C12, is more active than EpoA. Small apolar substituents on C12 are in general well tolerated, as they would be placed in the hydrophobic pocket hosting F270 and L369, which is left empty by EpoA. However, polar or bulkier substituents

cannot be easily accommodated in this pocket, as reflected by the lower activity of such modified epothilones (Nicolaou *et al.,* 1998).

The NMR-derived model of the EpoA – tubulin complex differs from the EC-derived one (Fig. 5.2). Four are the major differences between the two models, in addition to the divergent conformation of the bound epothilone (Fig. 5.1).

First, in the EC derived model the nitrogen of the thiazole side chain of epothilone makes a H-bond with the aromatic ring of H227 (Fig. 5.2A), while in the NMR derived model the N22 of epothilone is not involved in H-bonds with the protein side chains. However, our model preserves a favorable contact between the aromatic rings of H227 and the thiazole of EpoA, situated face-to face to each other, as opposed to the EC-derived model, where the two aromatic rings are situated edge-to-edge, in an orientation seldom observed in protein cores (Samanta *et al.,* 1999). Moreover, on the basis of the EC derived model, one would predict a conspicuous loss of the tubulin polymerizing activity of epothilones bearing a phenyl instead of a thiazole ring in the side-chain. Contrarily, substitution of the thiazole ring by a phenyl ring has been found to affect only the activity of epothilone in cells, while the induction of tublin polymerization in a test-tube is comparable for wild-type and modified epothilones.

Second, in the EC derived model the 3OH of EpoA could entertain a H-bond with the carbonyl of T274 (Fig. 5.2), while in the NMR derived model the 3OH of epothilone is not involved in H-bonds with the protein side chains. However, the importance of the 3OH hydroxyl group in tubulin binding is not confirmed, as (*E*)-2,3-dehydro epothilone, lacking the 3OH group, possesses a high biological activity. The drop in activity of epothilones with inverted stereochemistry at C3 could be related to the inability of such mutant epothilones to assume a *trans* conformation for the C2-C3 torsion angle, which seems to be required for function.

Third, the EC-derived model does not explain the higher activity of EpoB vs. EpoA, as substituents on C12 would be projected towards solution, thus making no additional contacts with the protein side chains. In contrast, in our model substituents on C12 would be placed in the hydrophobic pocket hosting F270 and

L369 and left empty by EpoA, which provides a rationale for the higher activity of EpoB vs. EpoA.

Fourth, in contrast to the NMR-derived complex, the EC derived model does not support the existence of a common pharmacophore for epothilone and paclitaxel.

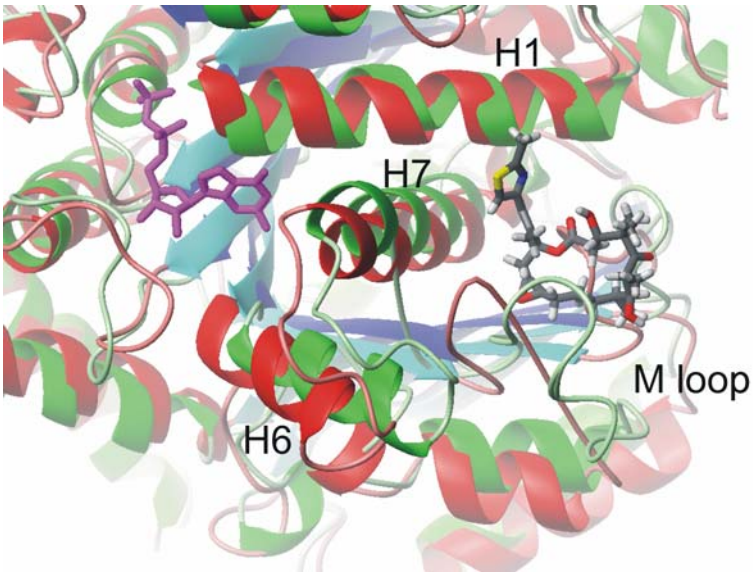


Fig. 5.3. Superposition of the straight (green) and bent (red) conformations of tubulin. Epothilone is shown in the taxane binding pocket in the binding mode of Fig. 4.4. GDP is shown in magenta.

The existence of two models for the EpoA – tubulin complex, which have been derived from different experimental data, suggests that the binding mode of EpoA to tubulin might depend on the tubulin polymerization state. The NMR-derived complex contains tubulin in the soluble state, namely in the functionally relevant state of the protein before polymerization. The EC-derived complex contains tubulin in the polymerized state on Zn-stabilized sheets, which are however never found in cells and therefore artificial. Future efforts aimed at the verification of the influence of the polymerization state of the protein on the binding mode of MSAs are essential to understand how microtubule-stabilizing drugs control tubulin function.

Recent work of the Knossow and Nogales laboratories (Ravelli *et al*., 2004; Wang & Nogales, 2005) suggests that microtubules stabilization occurs via straightening of the curvature at the α and β units interface in protofilaments (Fig. 1.2). In particular a 2.5 Å translation of the helix H7, which drags along the H6-H7 loop

and the helix H6, is the major player in determining the curvature at the interface between heterodimers (Fig. 1.2) (Ravelli *et al*., 2004). These authors propose that GDP and GTP regulate microtubules stability by influencing the equilibrium between the straight (green in Fig. 5.3) and bent (red in Fig. 5.3) tubulin conformation. When GTP is bound the H7 lever helix is pushed towards its straight conformation through interaction of the nucleotide with the N-terminal end of this helix. We propose that epothilone regulates microtubules stability by the same mechanism but operating from the other side of the H7 helix. When epothilone binds tubulin, it acts as a linker between the H7 helix and the M loop, thereby pulling the H7 helix in the direction of the M loop, namely in the conformation assumed in the straight form of tubulin (Fig. 5.3). We propose that the direct contact of the M-loop residue R278 with the H7 residue D224, mediated by epothilone, and the interactions of epothilone with the H7 residue H227 and the M-loop residues R284 and T274, lock the relative position of the H7 helix and the M-loop, thereby stabilizing the straight conformation even in absence of GTP. This model provides a rationale for the mechanism of action of the drug.

In conclusion, our data establish a relationship between the functional mechanism of epothilone and its binding mode to tubulin. The differences between the structure of the EpoA – tubulin complex in solution and in Zn-stabilized sheets and the incompatibility of the EC-derived structure with the NMR data in solution underlines the importance of accessing structural data for the EpoA – tubulin complex in the different polymerization states. We expect our results to be helpful for the development of improved therapeutic molecules, based on the existence of a common pharmacophore for epothilone and paclitaxel.

Part 2

A novel pharmacophore for peptide toxins interacting with K^+ channels.

Chapter 1

Introduction

1.1 The biological system.

Potassium (K^+) channels are a very diverse group of proteins that are key elements for a variety of different physiological functions, including the electrical excitability of cells (Hille, 2001). During evolution a great variability of peptide toxins targeting different K^+ channels have been evolved by several venomous organisms including snakes, spiders and scorpions.

Despite the lack of any obvious structural homology, unrelated toxins that interact with voltage activated K^+ channels, including charybdotoxin from a scorpion, BgK from a sea anemone and dendrotoxin from snakes, share a common structural feature. A dyad motif, composed of a lysine and a hydrophobic amino acid residue, usually a phenylalanine or a tyrosine is the key player in the binding to the target K^+ channel (Stampe *et al.*, 1994; Dauplais *et al.*, 1997; Savarin *et al.*, 1998; Gilquin *et al.*, 2002; Srinivasan *et al.*, 2002). The lysine residue of this functional dyad has been proven to penetrate the K^+ channel pore, thereby occluding it (Miller, 1995; Eriksson & Roux, 2002; Gao & Garcia, 2003). In particular, the dyad motif is present in all the diverse peptides that interact with the voltage-gated K_v1 subfamily of K^+ channels, such as the *Shaker* channel from *Drosophila* (Dauplais *et al.*, 1997; Savarin *et al.*, 1998; Gilquin *et al.*, 2002; Srinivasan *et al.*, 2002; Rauer *et al.*, 1999; Jacobson *et al.*, 2000). Therefore, it was hypothesized that all peptidic toxins targeting K_v1 channels possess a common pharmacophore, centered on a functional dyad.

The venoms of the marine cone snails are known to contain different families of peptides, the so called conotoxins, which target ligand gated and voltage gated ion channels with great specificity. These cysteine-rich peptides are classified into groups based on their structural scaffolds and target specificity (Terlau & Olivera, 2004). κ-conotoxin PVIIA (κ-PVIIA), a *Shaker* K^+ channel blocker (Terlau *et al.*, 1996; Shon *et al.*, 1998) was the first K^+ channel-targeted conotoxin to be

characterized. Although structurally unrelated to other known K^+ channel blockers, κ-PVIIA shares the common feature of a functional lysine/hydrophobic residue dyad (Jacobson *et al.*, 2000). Recently, the group of H. Terlau at the University of Lübeck described a novel *Conus* peptide ligand, κM-conotoxin RIIIK (κM-RIIIK), that blocks the *Shaker* K^+ channel in a state-dependent manner. By using *Shaker* channel mutants with single residue substitutions it was shown that κM-RIIIK interacts with the pore region of the channel. Introduction of a negative charge at residue 427 (K427D) greatly enhances the affinity of κM-RIIIK, whereas substitutions at the residues F425 and T449 strongly reduced toxin affinity. The teleost homolog of *Shaker*, the trout *TSha1* K^+ channel is the highest-affinity target of κM-RIIIK yet identified (Ferber *et al.*, 2003). The 24 amino acids sequence of κM-conotoxin RIIIK contains three positively charged residues but no aromatic side chain (Fig. 1.1).

```
κM-RIIIk   LOSCCSLNLRLCOVOACKRNOCCT#
μ-GIIIA     RDCCTOOKK-CKDRQCKOQRCCA#
μ-GIIIB     RDCCTOORK-CKDRRCKOMKCCA#

ψ-PIIIE    HOOCCLYGKCRRYOGCSSASCCQR#
ψ-PIIIF    GOOCCLYGSCROFOGCYNALCCRK#


κ-PVIIA    CRIONQKCFQHLDDCCSRKCNRFNKCV
```

Fig. 1.1 Sequence alignment of κM-conotoxin RIIIK to other M-superfamily peptides and κ-conotoxin PVIIA. O = 4-hydroxyproline; # denotes an amidated C-terminal amino acid.

The absence of an aromatic side chain raised the intriguing question about the presence of the lysine/hydrophobic dyad motif in κM-RIIIK. If a functional dyad is present a non-aromatic residue must be the hydrophobic counterpart of the positively charged lysine. However, it is conceivable that κM-RIIIK binds the *TSha1* K^+ channel with a novel, not yet identified pharmacophore.

κM-RIIIK shares the same disulfide connectivity with other conotoxins belonging to the M-superfamily (Fig. 1.1). μ-conotoxins GIIIA and GIIIB are Na^+ channel blockers, while ψ-conotoxins PIIIE and PIIIF are antagonists of the nicotinic

acetylcholine receptor. The similarity in the cysteine bridge schema of the three peptides, despite their completely different functions, poses the challenging question if the pharmacological selectivity of these peptides is based on general differences in the overall three-dimensional shape or on subtle differences in the charge and hydrophobic side-chain distribution on the surface of the peptide.

1.2 The question we ask

In this work we combine systematic mutational analysis, NMR structural analysis, thermodynamic mutant cycle analysis and docking calculations to answer the following questions:

1. Which amino acids of κM-RIIIK are essential for channel binding and does the petide contain a functional dyad?
2. Is there any structural similarity between κM-RIIIK and the other peptide of the M-superfamily μ-GIIIA and GIIIB and ψ-PIIIE and PIIIF, which have a different pharmacological profile? Is there any structural similarity between κM-RIIIK and κ-PVIIA with the same activity spectrum?
3. With which pharmacophore does κM-RIIIK bind the *TSha1* K^+ channel? Are the intermolecular recognition features similar in the κM-RIIIK/K^+ channel complex and in the μ-GIIIA/Na^+ channel complex?
4. Are the several hydroxyproline residues present in κM-RIIIK of any significance for channel binding?

1.3 Significance of the work

The structural/functional study conducted here has a broad significance because κM-RIIIK is at the unique intersection between structural similarity and similar target specificity. We will show that κM-conotoxin RIIIK binds the K^+ channel *TSha1* with a novel binding mode, which is not centered on a functional dyad. The intermolecular interactions, involving the channel loops and the toxin positively charged side chains, identified in this work, are of general importance in determining the selectivity of peptidic toxins for ion channels.

We will provide direct evidence that κM-conotoxin RIIIK and κ-conotoxin PVIIA, which both inhibit the same K^+ channel subtypes, are both structurally divergent and mechanistically dissimilar. In contrast, κM-conotoxin RIIIK is structurally highly similar to genetically related conotoxins (i.e., μ-conotoxin GIIIA and ψ-conotoxin PIIIF), which have a completely different target specificity (K^+ channels, Na^+ channels, Ach receptors). Additionally, we will show that the binding mode of κM-RIIIK to the *TSha1* channel closely resembles that of μ-conotoxin GIIIA to the $Na_V1.4$ Na+ channel. This observation demonstrates that the two conotoxins have developed a similar scaffold and binding geometry to their respective targets, while the pharmacological selectivity is determined by the exact charge distribution and three-dimensional shape. This provides a potential framework for understanding both molecular convergence and divergence in ion channel-targeted ligands.

Chapter 2

The experiments

2.1 Materials.

Peptide synthesis. Synthetic conotoxin κM-RIIIK was produced by solid-phase peptide synthesis for structure determination by NMR spectroscopy. Peptides were synthesized manually on a methylbenzhydrylamine resin (2g) (Stewart *et al.*, 1976) (substitution was 0.2-0.4 meq./g). The following side chain protected Boc-amino acids were used: Hypro (Bzl), Cys(Mob), Asn(Xan), Arg(Tos), Lys(Cl-Z), Ser(Bzl),Tyr(Br-Z), His(Tos), Glu(Chx). Most peptide couplings were mediated for 1 h by diisopropylcarbodiimide in CH_2Cl_2 or *N*-methylpyrrolidone in the presence of HOBt and monitored by the qualitative ninhydrin test of Kaiser (Kaiser *et al.*, 1970). Couplings including and following hydroxyproline were achieved using 3 equivalents (eq.) of protected amino acids (based on the original substitution of the resin) preactivated in NMP with 2.9 eq TBTU, 6 eq. HOBt and 9 eq. DIPEA. Boc removal was achieved with TFA (60% in CH_2Cl_2, 1% m-cresol) for 10 min. After deblocking, the resin was washed with methanol followed by successive washes with triethylamine (10% in CH_2Cl_2), methanol, triethylamine, methanol and CH_2Cl_2. The completed peptide was then cleaved from the peptido-resin (weight gain varied between 1.5 to 2.0 grams) by HF containing the scavengers anisole (10% v/v) and methyl sulfide (5% v/v) for 90 min at 0 °C. The deprotected and cleaved peptide was precipitated in ether with 1% mercaptoethanol and separated from the spent resin by filtration after dissolution in aqueous 0.1% TFA and 60% acetonitrile. The major and desired component was isolated using preparative reverse phase HPLC using an acetonitrile gradient in 0.1% TFA-water. The fractions containing the desired peptide (300-400 mL) were combined and degassed with argon for 5 min. A buffer solution (500-600 mL) made of distilled water, 9.6 g (0.1 mol) ammonium carbonate, 100 mg EDTA (0.27 mmol), 250 mg of Cysteine.HCl (1.6 mmol) and 250 mg of Cystine (1.05 mmol) and 100 mL acetonitrile was also

degassed using argon and added to the reduced peptide solution (300-400 mL); the pH (ca 8.8) was immediately adjusted to 8.2 with glacial acetic acid. The clear solution was stirred under argon for 16 hrs at room temperature. The folding was followed by HPLC and generally completed within a few hours.

After the folding was completed, the reaction mixture was acidified with TFA, diluted with the same volume of water, then loaded to a preparative HPLC (Miller & Rivier, 1998) cartridge (5 cm × 30 cm) packed in the laboratory with reversed-phase 300Å Vydac C_{18} silica (15-20 μm particle size). The peptides eluted with a flow rate of 100 mL/min using a linear gradient of 1% B per 3 min increase from the baseline % B (eluent A = 0.25 N TEAP pH 2.25, eluent B = 60% CH_3CN, 40% A). All peptides were subjected to a second purification step carried out with eluents A = 0.1% TFA in water and B = 60% CH_3CN/40% A on the same cartridge using a linear gradient of 1% B per min increase from the baseline % B. Analytical HPLC screening of the purified fractions was performed on a Vydac C_{18} column (0.46 x 25 cm, 5 μm particle size, 300 Å pore size) and a 0.1% TFA solvent system. The fractions containing the product were pooled and subjected to lyophilization. The final yield for these purified peptides was about 5-25% (20-120 mg).

Characterization of analogues. The purity of the final peptides was determined by analytical RP-HPLC performed with a linear gradient using 0.1 M TEAP pH 2.5 as eluent A and 60% CH_3CN/40% A as eluent B on a Hewlett-Packard Series II 1090 Liquid Chromatograph connected to a Vydac C_{18} column (0.21 x 15 cm, 5 μm particle size, 300 Å pore size), Controller Model 362 and a Think Jet printer. Capillary zone electrophoresis (CZE) analysis was performed as described earlier (Miller & Rivier, 1998). Peptides are more than 90% pure by HPLC and CZE in most cases. Mass spectra (MALDI-MS) were measured on an ABI-Perseptive DE-STR instrument. The instrument employs a nitrogen laser (337 nm) at a repetition rate of 20 Hz. The applied accelerating voltage was 20 kV. Spectra were recorded in delayed extraction mode (300 ns delay). All spectra were recorded in the positive reflector mode. Spectra were sums of 100 laser shots. Matrix alpha-cyano-4-hydroxycinnamic acid was prepared as saturated solutions

in 0.3% trifluoroacetic acid in 50% acetonitrile. The observed monoisotopic $(M + H)^+$ values of each peptide corresponded with the calculated $(M + H)^+$ values.

1D NMR spectra were used to prove the structural integrity of the alanine analogues which showed a profound reduction in the binding affinity. For these mutants (L1A, R10A, K18A and R19A) only local chemical shift changes are observed, indicating that the overall structure is conserved.

Molecular biology. The original DNA clone of the *Shaker*-related K^+ channel (*TSha1*), from CNS of rainbow trout (*Oncorhychus mykiss*), was kindly provided by Prof. Dr. G. Jeserich, University of Osnabrück, Germany). *In vitro* site-directed mutagenesis of *TSha1* was performed following standard procedures (Ishii *et al.*,1998; Sambrook & Russel, 2001) and according to manufacturer protocols when commercial kits were used. The vector used throughout this study was pSGEM (3118 bp), a modified version of pGEMHE, which was a generous gift of Prof. Dr. Michael Hollmann (Bochum University, Bochum, Germany).

2.1 Methods.

NMR analysis. The sample was prepared by dissolving 2.4 mg of the peptide in 300 µl of 90% H_2O/10% D_2O, pH 3.5. and measured at 25 °C. A pH of 3.5 was chosen because of the lower solvent exchange rate of the H_N protons. No substantial difference was observed in the relative chemical shifts of the H_N peaks when raising the pH to 6.5 or lowering the temperature to 5 °C, indicating that the structure of the peptide does not vary significantly among the different conditions. An additional peptide sample for the measurement of N-H_N dipolar couplings was prepared by dissolving 2.4 mg peptide in 14 µl pentaethylene glycol monododecyl ether (C12E5) and 317 µl of a 10% D_2O/H_2O mixture (Rückert & Otting, 2000). The nematic phase formed after adding 2 µl n-Hexanol and vigorous mixing.

Resonance assignment and structural restraints were obtained by conventional NMR methods for unlabelled peptide chains (Wüthrich, 1986). NOESY (Jeener *et al.*, 1976) with mixing times of 100 ms and 350 ms, DQF-COSY (Rance *et al.*, 1983), TOCSY (Braunschweiler & Ernst, 1983), ^{15}N and ^{13}C HSQC (Bax *et al.*, 1990) NMR spectra were collected on Bruker-Avance 600 and 700 MHz spectrometers.

Solvent exchange rates of the H_N protons were measured by adding 300 µl pure D_2O to 2.4 mg of κM-conotoxin RIIIK. Eight TOCSY experiments were performed after 0h, 1h, 2h, 4h, 8h, 16h, 22h and 36h. After adding D_2O the following H_N protons were still observable: Cys4, Ser6, Leu9, Arg10, Val14, Cys17, Lys18, Asp20 und Cys23.

N-H_N residual dipolar couplings were measured in a ω_1-coupled HSQC experiment for 2.4 mg of κM-conotoxin RIIIK dissolved in the nematic phase described above.

All spectra were processed with the Bruker XWIN-NMR software package (*Bruker Instruments*, Rheinstetten, Germany) and further analyzed with the FELIX 2000 software package (*MSI*, San Diego, CA). Cross-peak volumes were calculated by integration of all data points in a rectangular region, which was manually defined for each peak.

Structure calculations. The distance restraints were classified on the basis of peak intensity in a NOESY spectrum with mixing time of 100 ms with lower and upper bounds of 1.0-2.5 Å (strong), 2.5-3.5 Å (medium) and 3.5-5.0 Å (weak). Very weak restraints with an upper bound of 6.0 Å were obtained from the NOESY experiment with mixing time of 350 ms. A correction of 0.5 Å was added to the upper bounds of restraints involving pseudoatoms for methyl groups. Backbone H-bonds were not included in the structure calculation.

All structure calculations were performed using an extended version of the X-PLOR 3.851 software package (Brünger, 1993). The initial extended conformer was subjected to a simulated annealing protocol consisting of four steps: an initial 65 ps high-temperature phase at 2000 K was followed by a first cooling phase, where the temperature was lowered to 1000 K in steps of 50 K, then by a second cooling phase, where the temperature was lowered to 100 K, and by a final energy minimization step. A total of 237 NOE-derived distance restraints were applied with soft-square potential and a force constant of 50 $kcal \cdot mol^{-1} \cdot Å^{-2}$. The stereospecific assignment of the protons of 16 methylene groups was obtained by means of a routine of the simulated annealing protocol, where the two H atoms are repetitively swapped and the configuration that best fits the experimental restraints is retained. For the remaining methylene groups either only one resonance was observed for both protons or the experimental data were not clearly

indicative of a single stereospecific assignment. Restraints involving not stereospecifically assigned CH_2 protons were treated with the floating chirality assignment procedure of X-PLOR. A total of seventeen dihedral angles restraints from $^3J(H_N,H_\alpha)$ and twelve dihedral angles restraints from $^3J(H_\alpha,H_\beta)$ coupling constants were included in the calculation.

A hundred structures were calculated and the thirteen structures with the lowest energy were used for statistics. This ensemble of thirteen structures satisfied the criteria of no NOE violations >1.0 Å and no dihedral angle violations >5°.

Electrophysiological measurements. The oocyte expression system was used to study the effects of the wild type and the isoforms of κM-RIIIK. Oocyte preparation and RNA injection was performed as described by Jacobson *et al.* (Jacobson *et al.*, 2000). Whole-cell currents were recorded under two-electrode voltage-clamp control using a Turbo-Tec amplifier (npi electronic, Tamm, Germany). The intracellular electrodes were filled with 2M KCl and had a resistance between 0.5 and 1.2 MΩ. Current records were low-pass filtered at 1 kHz (-3db) and sampled at 4 kHz. The bath solution was normal frog Ringer's containing (in mM): 115 NaCl, 2.5 KCl, 1.8 $CaCl_2$, 10 Hepes, pH 7.2 (NaOH). All electrophysiological experiments were performed at room temperature (19-22 °C). Toxin-containing solutions were applied by a Gilson pipette directly into the bath chamber.

The IC_{50} values of □M-RIIIK and the corresponding analogs were calculated from the reduction of whole cell currents at a test potential of 0 mV obtained from oocyte expressing either *TSha1* or the different *Shaker* isoforms according to IC_{50} = fc / (1-fc) * [Tx], where fc is the fractional current and [Tx] is the toxin concentration. Data are given as mean ± S.D. For wild type toxin dose-response measurements where performed as well, resulting in an almost identical IC_{50} value (see Fig. 3.3).

The two-electrode voltage clamp technique (TEVC) in the *Xenopus* oocyte heterologous expression system was used to investigate the affinities of the toxin mutants for the different channel mutants. The calculation of the interaction free energy (ΔΔG) from the electrophysiological data was performed as described earlier (Jacobson *et al.*, 2000).

Docking calculations. ~ The model of the *TSha1* channel was obtained by homology to the crystal structure of the KcsA channel (Doyle *et al.*, 1998) with a procedure similar to that described by Eriksson and Roux to obtain the *Shaker* channel model (Eriksson & Roux, 2002). A large number of spatial restraints, extracted from the KcsA pdb file (1BL8), were optimised using the program MODELLER v.6.2 to generate similarity-based models of *TSha1* (Sali & Blundell, 1993). The ideal fourfold symmetry of the tetrameric channel was imposed on all the models. Of the one hundred generated model, the one with the lowest MODELLER restraint energy was kept and refined by energy minimization with the program CNS solve 1.1 (Brünger *et al.*, 1998), using the "proteinallhdg" forcefield.

The docking protocol was based on a similar procedure as that used by Eriksson and Roux (Eriksson & Roux, 2002). 875 docking models of the complex between the *TSha1* channel model structure and the NMR structure of κM-conotoxin RIIIK were generated with the program CNS solve 1.1 (Brünger *et al.*, 1998) and using mutant cycle derived intermolecular distance restraints. In the thermodynamic double mutant cycle analysis, two interacting molecules A and B are mutated at specific sites *a* and *b* to give two molecules A* and B*. The binding free energies ΔG for the wild-type, single and double mutant complexes A-B, A*-B, A-B* and A*-B* are combined to obtain the interaction free energy $\Delta\Delta G_{int}$ = [ΔG(A*-B*)-ΔG(A*-B)]-[ΔG(A-B*)-ΔG(A-B)]. The $\Delta\Delta G_{int}$ is a measure of the non-additivity of the mutantions *a** on A and *b** on B and reveals a direct interaction of the sites *a* and *b* (Ranganathan *et al*., 1996; Schreiber & Fersht, 1995). In this study the ΔΔG values of Fig. 3.7 were translated into intermolecular distances as follows: a) 2 KJ/mol < ΔΔG ≤ 3 KJ/mol corresponds to a very weak intermolecular distance (d) restraint between sites *a* and *b* (4 Å < d < 6 Å); b) 3 KJ/mol < ΔΔG ≤ 4 KJ/mol corresponds to a weak intermolecular distance restraint (3.5 Å < d < 5.5 Å); c) 4 KJ/mol < ΔΔG ≤ 5.5 KJ/mol corresponds to a medium distance restraint (2.5 Å < d < 4.5 Å); ΔΔG > 5.5 KJ/mol corresponds to a strong distance restraint (1.5 Å < d < 3.5 Å). If *a* and *b* are polar or charged residues, the minimum distance between the heavy atoms of the polar or charged groups has to fulfil the intermolecular distance restraint; for non-polar residue pairs the minimum

distance between all heavy atoms of the side chains is used. Due to the fourfold symmetry of the *TSha1* channel we used ambiguous distance restraints: the distance d_{ij} between a residue of the toxin (i) and a residue of the channel (j) is described as:

$$d_{ij} = \mathrm{Min}(||ri-r_j^A||, ||ri-r_j^B||, ||ri-r_j^C||, ||ri-r_j^D||)$$

where A, B, C and D are the four identical subunits of *TSha1*.

To avoid redundant identical complexes all distances from the residue R10 of κM-RIIIK were assigned to the subunit A of the channel. The NMR structure of κM-conotoxin RIIIK shows that R10 and K18 are on opposite sides of the peptide: to ensure that the toxin is stretched over the pore, we impose that K18 is close to the extracellular loop of the channel unit opposite to that in contact with R10 (unit C). A total of sixteen distance restraints were used: three *very weak* (V376-L1, M375-O15, E348-R19), three *weak* (E348-L1, M375-L1, E348-R10, five *medium* (E354-L1, P349-L1, E348-K18, E354-R19, S351-R19) and five *strong* (E354-R10, P349-R10, S351-R10, E354-K18, P349-R19). An additional distance restraint was added between the N-terminus NH_3^+ of the toxin and any electronic acceptor of the channel, following the dramatic decrease in the toxin affinity upon acetylation of the terminal amine (Al-Sabi *et al.*, 2004). All distances were restrained with a harmonic potential.

The docking protocol consisted of a series of short torsion angle MD trajectories. In the first step of the simulated annealing protocol the temperature was increased to 1000 K (500 steps of 10 fs for a total of 5 ps); during this phase the Van der Waals radius of the atoms is decreased from 1 to 0.1. In the second step the temperature is slowly lowered to 300 K (1000 steps of 15 fs for a total of 15 ps) and the Van der Waals radius is increased from 0.1 to 1. At the beginning of the trajectories, the toxin was positioned randomly to a distance of 15 Å from the channel. All docking trajectories were started with different initial conditions by changing the initial velocities and the initial orientation of the toxin with respect to the channel pore axis. During the simulated annealing the electrostatic energy was turned on towards the end of the dynamics. A continuum solvent model was

used with a dielectric constant ε = 12 (Eriksson & Roux, 2002). The distance restraint force was fixed to 300 kcal mol^{-1} $Å^{-2}$ during the dynamics.

Flexibility was allowed for the conformation of the toxin backbone of the amino acids 1-3 and of all the toxin side chains. The N-terminus backbone of the peptide was allowed to change conformation with respect to the NMR structure following the observation that this region is highly flexible in solution (Al-Sabi *et al.*, 2004). The peptide backbone of residues 4-24 was restrained with an harmonic potential based on the root-mean-square deviation from the NMR structure with a force of 500 kcal mol^{-1} $Å^{-2}$. This allows rotation and translation of the toxin with respect to the channel, while keeping the overall folding close to the experimentally determined one. The conformation of the trans-membrane helices of the channel was kept fixed, the backbone of residues 336-386 was restrained to that of the initial model with an harmonic potential and a force of 300 kcal mol^{-1} $Å^{-2}$, while the side-chains of residues 336-386 were allowed full flexibility.

The MD were followed by 6 cycles of 500 steps of energy minimizations where the backbone restrain function was turned off and the distances restraints were decreased to 75 kcal mol^{-1} $Å^{-2}$. During the minimization the electrostatic potential was turned on.

The trajectories were analysed using the g_cluster module of the GROMACS 3. 1. program (Berendsen *et al.*, 1995; Lindahl *et al.*, 2001) The clusters were generated using distances RMSD with the gromos method (Daura *et al.*, 1999) and a cutoff of 0.2 nm. The RMSD used to define the clusters is an intermolecular RMSD containing distances between toxin residues L1, S6, R10, O13, O15, K18, R19 and amino acids E348, P349, E354, G371 and V376 of the turret and pore helix region of the four units of the channel.

Chapter 3

Results

3.1 The NMR structure of conotoxin κM-RIIIK; similarities to μ- and ψ-conotoxins.

The three dimensional structure of conotoxin κM-RIIIK (Fig. 3.1) was determined from a total of 237 NOE-derived distance restraints, 17 backbone dihedral angle restraints and 12 side-chain dihedral angle restraints. 100 structures were generated using the simulated annealing protocol of the X-PLOR program (Brünger, 1993), including NMR-derived structural restraints (Table 3.1).

κM-RIIIK is relatively flexible in aqueous solution. The r.m.s. deviation of the 13 structures with lowest energy (Fig. 3.1) versus the mean is 1.62 Å for the backbone atoms and 2.74 Å for all heavy atoms. The region 1-11 is relatively disordered, while the C-terminus region 12-24 is well defined with a backbone r.m.s.d value of 0.78 Å. The flexibility of the peptide in the region 1-11 is confirmed by the higher values of the ^{1}H-^{13}C heteronuclear NOEs of the C_{α}-H_{α} vectors in the N-terminus part of the peptide as compared to those in the C-terminus part, as well as by the significantly reduced values of the N-H_N dipolar couplings in the region 3-11. Most of the ϕ and ψ angles are in allowed regions of the Ramachandran plot, but only 37% lie in the most favored region. This indicates the presence of non-canonical or distorted structural elements, brought about by the three disulfide bridges and the four hydroxyprolines. The ω dihedral angle for all hydroxyprolines were found to be *trans* on the basis of NOEs and $^{13}C_{\beta/\gamma}$ chemical shifts.

The overall form of κM-RIIIK is that of a disk with dimensions 15 x 13 Å. The three dimensional structure of κM-RIIIK (Fig. 3.2A, red ribbon) is very similar to that of the Na^{+} channel blockers μ-conotoxins GIIIA and GIIIB (Fig. 3.2A, blue

ribbon) in the C-terminal region 12-24 (27-29). The three peptides have the same disulfide connectivity, which was independently confirmed by direct chemical analysis of partially reduced intermediates of κM-RIIIK (J. Rivier, unpublished results). In the C-terminal half of the peptide, a γ turn between Cys12 to Val14 is followed by a short distorted right-handed helix between Hyp15 and Lys18. The tail of the peptide contains a final 3_{10} helix turn between Asn20 and Cys23. The presence of these structural elements is confirmed by the slow solvent exchange kinetics of the amide protons of Val14, Cys17, Lys18, Asn20 and Cys23. On the other hand, the structure of κM-RIIIK significantly differs from that of μ-GIIIA in the flexible region 1-11. The N-terminal region of μ-GIIIA consists of very tight turns (2-5 and 5-8) which position the N-terminal tail in the opposite direction with respect to the C-terminus (Fig. 3.2B). Conversely, the N-terminal region of κM-RIIIK consists of an extended tail until Cys4 followed by a wide loop between Cys5 and Leu11. In conotoxin κM-RIIIK the N-terminal tail protrudes into solution almost perpendicular to the flat surface of the peptide and to the C-terminal end (Fig. 3.2B).

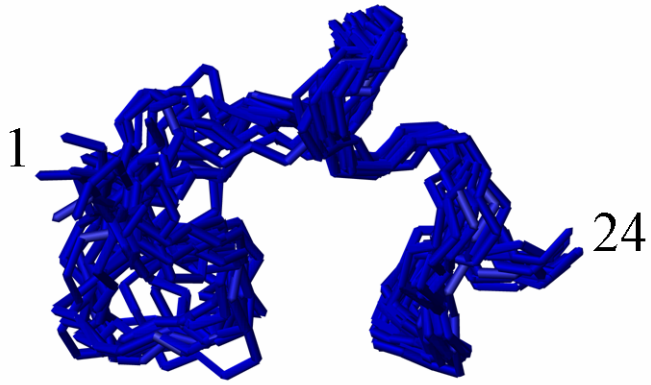


Fig. 3.1 Superimposition of the 13 lowest energy structures of κM-conotoxin RIIIK. The region 1-11 presents a high degree of flexibility, while the region 12-24 is well defined.

Recently, the three-dimensional structures of two ψ-conotoxins, ψ-PIIIE and ψ-PIIIF, antagonists of the nicotinic acetylcholine receptor, have been solved in solution by NMR (Fig. 3.2A,B green ribbon) (Van Wagoner & Ireland, 2003; Van Wagoner *et al*., 2003). The disulfide bond pattern of these small peptides is homologous to that of both κM-RIIIK and μ-GIIIA, although their natural receptor

belongs to a different superfamily, the ligand-gated ion channels. The fold of the C-terminal part of ψ-PIIIF is similar to that observed for both κM-RIIIK and μ-GIIIA, with two helical turns involving the last three cysteine residues. However, the N-terminal part of ψ-PIIIF, which consists of well defined β turns, differs in structure from both κM-RIIIk and μ-GIIIA. The series of β turns, that structurally differentiate the N-terminus of ψ-PIIIF from the wide flexible loop of κM-RIIIK, place the N-terminal tail perpendicular to the flat surface of the peptide (Fig. 3.2B), similar to κM-RIIIK but unlike μ-GIIIA. Although well defined turns are situated in homologous positions in μ-GIIIA and ψ-PIIIF, the geometry of these turns is not the same, bringing about the difference in the positioning of the N-terminus in the two toxins. Furthermore, unlike conotoxins κM-RIIIK and μ-GIIIA, ψ-PIIIF has the form of a lens rather than a disk (Fig. 3.2B).

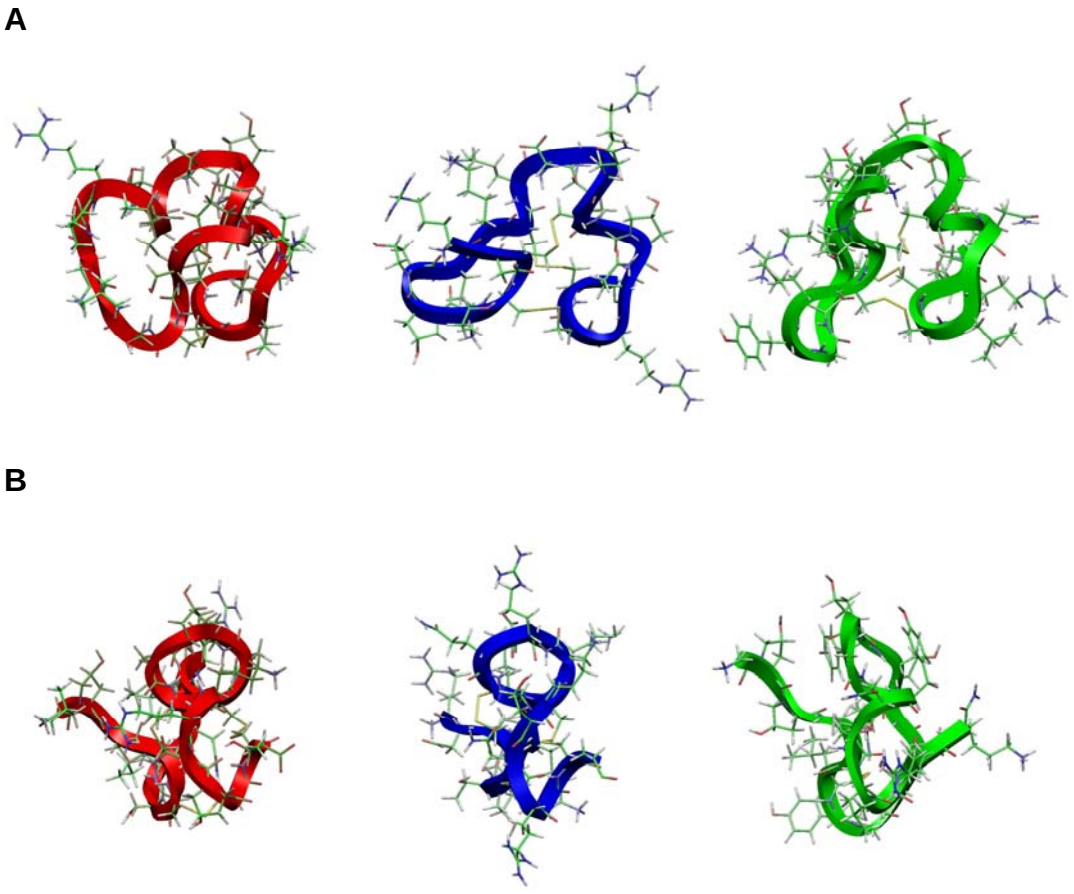


Fig 3.2 A) Comparison of the structures of κM-conotoxin RIIIK (red ribbon), μ-contoxin GIIIA (blue ribbon; pdb entry: 1TCG.pdb) and ψ-conotoxin PIIIF (green ribbon; pdb entry: 1JLP.pdb). The N-terminus is on the left side and the C-terminus on the right side for each peptide. The

structure of the three peptides is very similar in the C-terminal part, while the N-terminal part shows conspicuous differences. B) same as in A) after rotation of 90º. This orientation was chosen to highlight the different position of the N-terminal tail in the three peptides.

The cysteine bridges in the three peptides show substantial differences in conformational flexibility. The two Cys bridges 4-17 and 5-22 are quite disordered in κM-RIIIK and average between a R and a S conformation, probably as a consequence of the mobility of the loop 5-11 (χ_2 = –60º for all Cys). Conversely, the corresponding cysteine bridges in μ-GIIIA, ψ-PIIIF are well defined. The disulphide bridge 12-23 of κM-RIIIK deserves a particular comment. The H_β resonances of Cys12 are degenerate and the H_N signal is extremely weak, which indicates conformational averaging on a ms time scale. Such resonance broadening had been observed before for the corresponding Cys10 of μ-GIIIA and μ-GIIIB. The observation that the H_N resonance of Cys10 sharpens at lower temperature motivated the structural analysis of μ-GIIIA and μ-GIIIB at 10 ºC (Hill *et al*., 1996; Lancelin *et al*., 1991; Wakamatsu et al., 1992); in contrast we do not observe any sharpening of the H_N resonance of Cys12 by changing the temperature from 25 ºC to 5 ºC. In κM-RIIIK the Cys bridge 12-23 shows conformational averaging between a R and a S conformation (χ_1 = -60º for Cys 23; the χ_1 of Cys12 cannot be determined due to the degeneracy of the H_β protons), similarly to the corresponding disulphide bridge in ψ-PIIIF (Van Wagoner *et al*., 2003). Conversely, in μ-GIIIA the Cys bridge 10-21 assumes only one well-defined conformation at 10 ºC, despite the observed broadening of the H_N resonance of Cys10 at 25 ºC. The flexibility of the 12-23 disulphide bridge and its homologs in ψ-PIIIF and μ-GIIIA at room temperature is potentially of mechanistic importance, as it allows for conformational changes upon binding to the receptor.

The three conotoxins κM-RIIIK, μ-GIIIA and ψ-PIIIF share a common cysteine bridge pattern and a similar folding of the C-terminal part of the polypeptide chain, despite the large diversity in the receptor specificity. We speculate that the structural difference at the N-terminus of the polypeptide chain, the degree of conformational flexibility and the different charge distribution in the three-dimensional space are key determinants of the target selectivity of these peptides.

Table 3.1

Restraints for calculation	
Total NOE restraints	237
Intraresidue	155
Sequential	51
Long range	31
Dihedral angle restraints	29
Restraint violations (mean and s.d.)	
Distance restraints ≥ 0.3 Å[(2)]	0.20 ± 0.02[(2)]
Dihedral angle restraints ≥ 5.0 deg	0.02 ± 0.02
Coupling constant restraints ≥ 0.5. Hz	0.24 ± 0.1
Deviation from idealized geometry	
Bonds (Å)	0.002
Angles (deg)	0.48
Dihedrals (deg)	1.73
Energies (kcal mol^{-1})	
Restraint energy	19 ± 6
Total energy	52 ± 7
Ramachandran statistics (%)	
Most favored	37.2
Additionally allowed	55.8
Generously allowed	5.8
Disallowed	1.3
R.m.s. deviation from mean structure (Å)	
Backbone atoms	1.62
All heavy atoms	2.74

[(1)]Structure statistics refers to an ensemble of 13 structures with the lowest energy from 100 calculated structures.

[(2)]No distance restraint was violated more than 1.0 Å.

Conotoxins κM-RIIIK is structurally unrelated to conotoxin κ -PVIIA, despite the fact that the two peptides have overlapping binding sites within the pore region of

Shaker K^+ channel, as indicated by the loss of activity upon mutations of F425 and T449 for both conotoxins κM-RIIIK and κ-PVIIA (Shon *et al*., 1998; Ferber *et al*., 2003). The secondary structure elements are very different, with the three dimensional structure of conotoxin κ-PVIIA mainly consisting of two large loops stabilized by a triple stranded β-sheet. A weak similarity can be recognized in the overall shape, which resembles a disk for both peptides. Moreover, as in conotoxins κM-RIIIK, extensive flexibility was found for two of the three disulphide bridges of conotoxin PVIIA.

3.2 Mutational analysis.

We synthesized analogs of κM-RIIIK containing alanine substitutions at each amino acid position except for the cysteines. All analogs were synthesized manually and characterized using RP-HPLC, CZE and mass spectrometry. The affinity of the alanine isoforms was functionally assayed by two-electrode voltage clamp measurements using *Xenopus* oocytes expressing the trout *TSha1* K^+ channel, the cloned K^+ channel isoform with the highest affinity for κM-conotoxin (IC_{50} of about 70 nM) identified so far, which is reversibly blocked by the toxin (Fig. 3.3). The results of these assays are shown in Table 3.2.

For six amino acids (O2; S6; L7; L9; O21; T24) the alanine substitution resulted in an IC_{50} value within 5-fold of the wild-type κM-RIIIK. We note that some substitutions (O2A, S6A, L7A) cause a bigger change in the IC_{50} value than others (L9A, O21A, T24A), indicating that although these residues are not dominant determinants for binding they may slightly contribute to the interaction with the *TSha1* channel.

Alanine mutations of seven (S3, N8, L11, O13, V14, O15, N20) residues exhibited an intermediate behavior, with IC50 values higher than 5-fold, but less than 16-fold of the wild-type IC_{50}. Interestingly, S3, L11, O13, V14, O15 and N20 are located on the same side of the molecule.

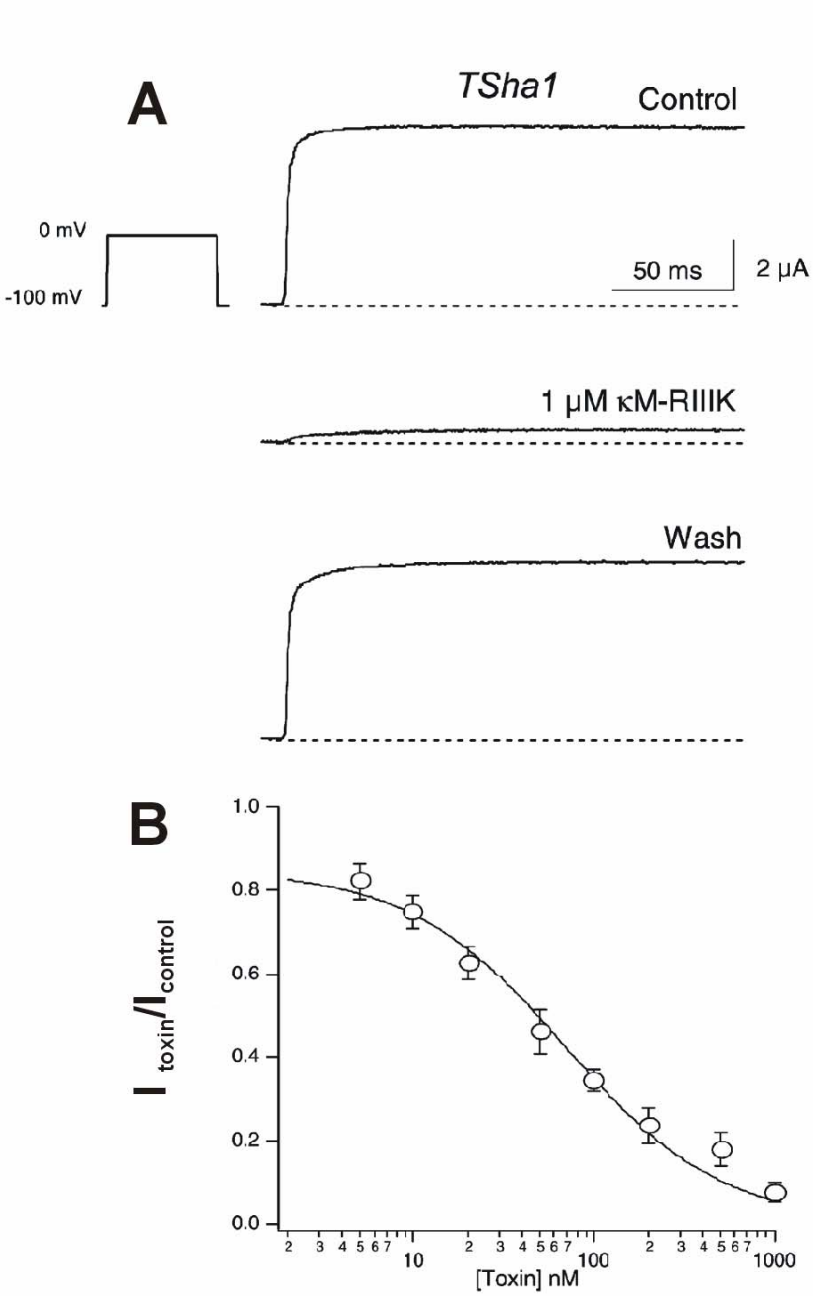


Fig. 3.3 κM-conotoxin RIIIK reversibly blocks TSha1 mediated currents. A) Whole cell currents recorded from oocytes expressing *TSha1* K^+ channels upon depolarization to 0 mV are shown before (control), after addition of 1 µM of κ-conotoxin RIIIK and after subsequent wash with NFR, indicating the reversibility of the block of the toxin. The holding potential was −100 mV. B) Dose-response curve for the block of *TSha1* current by κM-RIIIK at a test potential of 0 mV (total number of experiments = 12 with n between 4 and 11 for the different toxin concentrations indicated). The IC_{50} from these measurements is 70 ± 11 nM which

is almost identical to the value from the fractional currents measurements (see Table 3.2). The Hill coefficient was close to 1.

Table 3.2.

IC_{50} values for *Tsha1* K^+ block of κM-RIIIK and analogs

κM-RIIIK	**IC_{50} ± SD**	**N**	**$IC_{50mutant}/IC_{50W}$**
WT	73 ± 34	7	1
L1A	3180 ± 300	6	44
O2A	210 ± 40	3	3
S3A	760 ± 280	6	10
S6A	240 ± 80	3	3
L7A	310 ± 100	6	4
N8A	420 ± 140	3	6
L9A	90 ± 20	3	1
R10A	4220 ± 860	7	58
L11A	970 ± 300	6	13
O13A	510 ± 170	6	7
V14A	1130 ± 390	5	15
O15A	850 ± 180	3	12
K18A	3920 ± 1300	6	54
R19A	1530 ± 710	9	21
KR1819RK	180 ± 100	7	2
N20A	430 ± 70	3	6
O21A	130 ± 30	3	2
T24A	120 ± 50	3	2

As shown in Fig. 3.4, mutations of four other residues (L1, R10, K18 and R19) resulted in a major reduction of the affinity for *TSha1* K^+ channels (Table 3.2). The IC_{50} of these analogs are increased by more than 20-fold indicating an important role of these amino acids in receptor binding. Interestingly, the affinity of the double mutant K18R-R19K to the *Tsha1* K^+ channel is close to wild-type, indicating that the positive charge is important at both positions, but the specific nature of the side chain is not critical for the interaction with the *TSha1* K^+

channel. Somewhat surprisingly for a peptide blocking K^+ channels, these four critical residues are distributed at various sites of the molecule (Fig 4.1). Furthermore, the structure of κM-RIIIK does not provide any evidence of the existence of a hydrophobic/positively charged amino acid dyad, as Leu1, the most important hydrophobic residue for binding, is more than 6 Å apart from any positive charge.

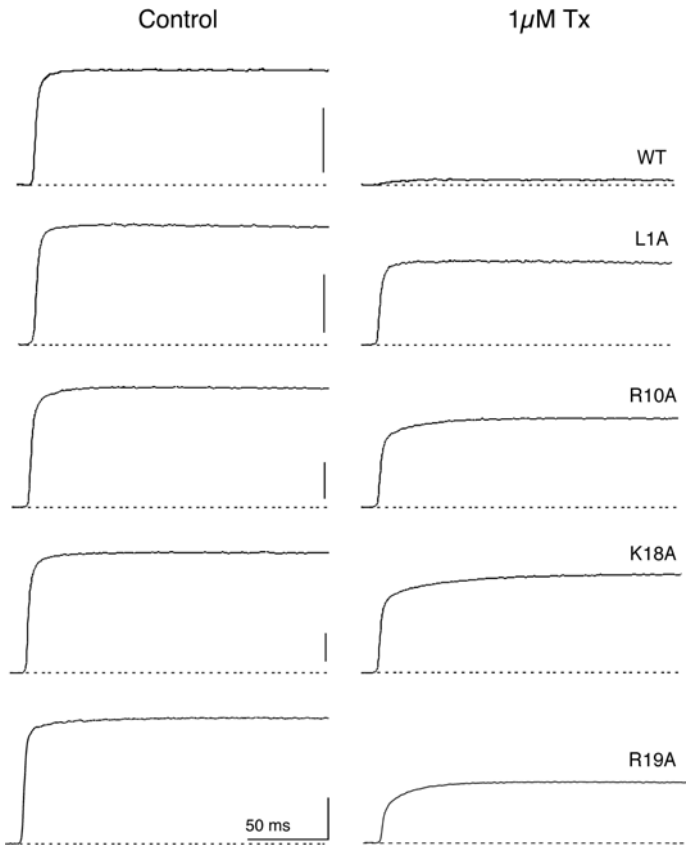


Fig. 3.4. Mutations of the residues L1, R10, K18 and R19 results in isoforms of κM-conotoxin RIIIK with low affinity to *TSha1* channels. Whole cell currents recorded from oocytes expressing *TSha1* K^+ channels before and after addition of the κM-RIIIK isoform are shown. Notice the apparent slowing of activation of the currents in the presence of R10A, K18A and R19A illustrating the re-equilibrium of toxin binding to the open state. The pulse protocol was as in Fig. 3.3. The vertical bars represent 2 μA.

The alanine walk identified Leu1 as the only hydrophobic residue that substantively affects binding of κM-conotoxin RIIIK to the *TSha1* channel. The solution structure determined by NMR indicates that Leu1 is too far from any

positively charged residue to serve as the hydrophobic component of a dyad, a conserved motif in otherwise divergent K^+ channel-targeted toxins. However, the fact that the N-terminal region (amino acids 1-11) is flexible leads to some reservations regarding this conclusion. In other toxins, the hydrophobic component of the dyad is most commonly an aromatic residue, while a lysine plays the role of the positively charged component; this lysine occludes the extracellular opening of the ion channel pore. Thus, if the Leu1 is part of a functional dyad that forms following a conformational change upon binding the receptor, it is reasonable to predict that the substitution of the leucine with an aromatic amino acid can be well-tolerated and perhaps even leads to an increase in binding affinity. In order to test this hypothesis, we evaluated the effect of various substitutions for Leu1 on the interaction of κM-RIIIK with the *TSha1* channel (Table 3.3).

Table 3.3.

Mutations of L1 affect the binding affinity of κM-RIIIK to the *TSha1* K^+ channel.

κM-RIIIK	**IC_{50} ± SD**	**n**	**$IC_{50mutant}/IC_{50WT}$**
WT	73 ± 34	7	1
L1A	3180 ± 300	6	44
Ac-L1	7630 ± 760	6	104
L1E	56500 ± 27670	5	774
L1H	5430 ± 790	4	74
L1I	1380 ± 1000	5	19
L1M	340 ± 110	5	5
L1K	950 ± 270	4	13
L1R	190 ± 10	3	3
L1F	5260 ± 1420	4	72
L1Y	12300 ± 1920	3	168

In contrast to what was predicted on the basis of the dyad hypothesis, both a phenylalanine (the most common amino acid as the hydrophobic half of the dyad) and a tyrosine substitution for Leu1 caused a greater apparent decrease in affinity (>100-fold) than the L1A substitution (~40-fold). Unexpectedly, the replacement

of Leu by a positively charged amino acid (e.g., the L1R and L1K mutations) had a smaller effect on the affinity for the *TSha1* channel (2- and 12-fold, respectively) than the L1A mutant. Introducing a negative charge (L1E) or blocking the positive charge (N-acetylated κM-RIIIK) causes a drastic fall in affinity (>500- and >100-fold, respectively).

The data of Table 3.2 and 3.3 do not support the involvement of Leu1 in a dyad. This result, together with the structural data that indicate that Leu1 is not close to any positively charged side chain, allows us to conclude that this leucine does not serve as the hydrophobic component of a dyad.

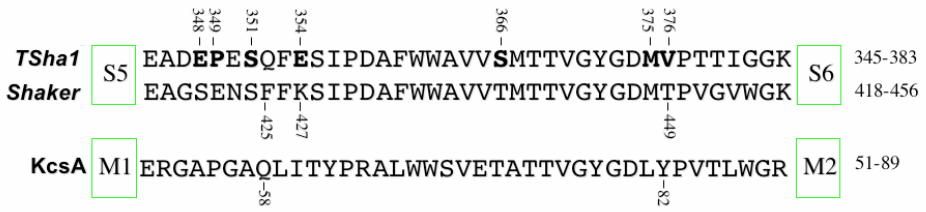
3.3 Mutant cycle analysis.

We have performed a mutant cycling analysis to identify which amino acids of κM-RIIIK interact with specific residues in the vestibule of the ion channel pore of *TSha1*. For this purpose we have generated several mutants of *TSha1* with single amino acids changes in the pore region of the channel, where κM-RIIIK is likely to interact (Ferber *et al*., 2003). Figure 3.5A shows the alignment of the amino acid sequences of the pore region of *TSha1* and *Shaker* together with the one of KcsA, for which a high resolution structure is available (Doyle *et al*., 1998). Several mutations within this area have been constructed and functionally assayed using the *Xenopus* expression system and 2-electrode voltage-clamp experiments. The addition of 1 μM of κM-RIIIK leads to different reductions of the measured currents at a test potential of 0 mV for wild type and mutated ion channels (Fig. 3.5B), demonstrating that the affinity of the peptide for the mutated channels is reduced. For M375K channels, hardly any block can be observed at this toxin concentration. This indicates a very low affinity of the toxin for the mutated channel, which was estimated to be about 10 μM ($k_D \approx 76$ nM for the wild type channel). The changes in the affinity for the different channel mutants are summarized in Table 3.4 and confirm that κM-RIIIK is indeed interacting with the pore region of the channel.

The relative changes induced by the different mutations are summarized in Fig. 3.6. The mutations E348S, P349K, S351K, M375L and V376T led to a 2-fold reduction in the affinity of κM-RIIIK. Interestingly the mutation E348K did not

have a significant effect, despite the drastic change in the electrostatic properties of the side chain. The mutations E354K and E354Q resulted in a decrease of the affinity of 6-fold and 4-fold respectively. The mutation M375I resulted in a much stronger affinity change than M375L (12-fold vs. 2 fold), while the most dramatic effect was observed for M375K (> 100-fold).

A



B

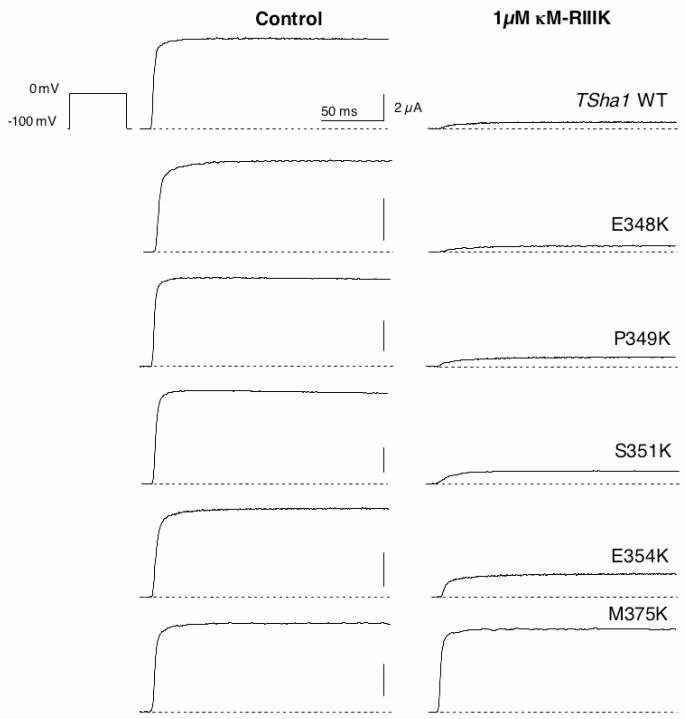


Fig. 3.5. Selected site-directed mutations on the pore region of *TSha1* K^+ channels. **A.** The amino acid alignment of the pore region of the *Shaker* and *TSha1* K^+ channels (S5-S6 linker) and the corresponding region of KcsA bacterial K^+ channel (M1-M2). **B.** Whole-cell currents recorded from oocytes expressing *TSha1* mutant channels before and after addition of the κM-RIIIK WT are shown. Notice the apparent slowing of activation of the currents in the presence of toxin illustrating the re-equilibrium of toxin binding to the open state (Ferber *et al*., 2003). The dashed lines correspond to zero current.

Table 3.4

IC_{50} values for *TSha1* mutant K^+ block of kM-RIIIK WT.

The IC_{50} values are given by mean values SE. NF: not functional. LE: low expression.

κM-RIIIK	**IC_{50}** (0 mV) nM	**n**
WT	76 ± 10	9
E348S	110 ± 30	4
E348K	50 ± 10	3
P349K	130 ± 20	3
S351K	160 ± 10	4
E354K	490 ± 110	3
E354Q	270 ± 30	4
S366T	NF	--
M375K	10730 ± 2460	3
M375L	140 ± 40	3
M375I	1360 ± 70	5
V376T	120 ± 20	3
V376H	LE	--
V376S	NF	--
V376E	NF	--
V376K	NF	--

To investigate which amino acids of the *TSha1* channel pore interact with specific amino acids of κM-RIIIK, we measured the affinity of several toxin mutants to different ion channel mutants. In this analysis we concentrated on the amino acid residues of the toxin which had been shown to be functionally relevant (L1, R10, K18, R19). The IC_{50} values for the different combinations of channel and toxin mutations are summarized in Table 3.5, while Fig. 3.7 shows the corresponding changes in binding energy (ΔΔG) values. Assuming that a ΔΔG-value of 2 kJ/mol indicates that the distance between the two mutated residues is less than 6 Å (Schreiber & Fersht, 1995), the data of Fig. 3.7 indicate a quite complex network of interactions between the toxin and amino acids from different domains of the ion channel target. R10, K18 and R19 have the largest interactions with residues E348, P349, S351 and E354 of the turret and pore helix regions of the K^+ channel, thus excluding that any of the positively charged side-chains penetrates the

channel pore. Instead, the positively charged residues interact with the loops of the four channel units, anchoring the peptide to the channel.

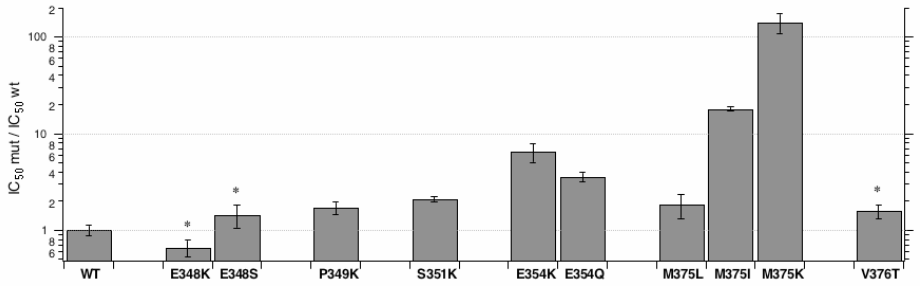


Fig. 3.6 Influence of single point mutations in the pore region of *TSha1* on κM-RIIIK binding. The bar diagram shows the affinity of κM-RIIIK to the individual channel mutants normalized to wildtype IC_{50}. Asterisks represent a non-significant mean value compared to WT (unpaired *t*-test).

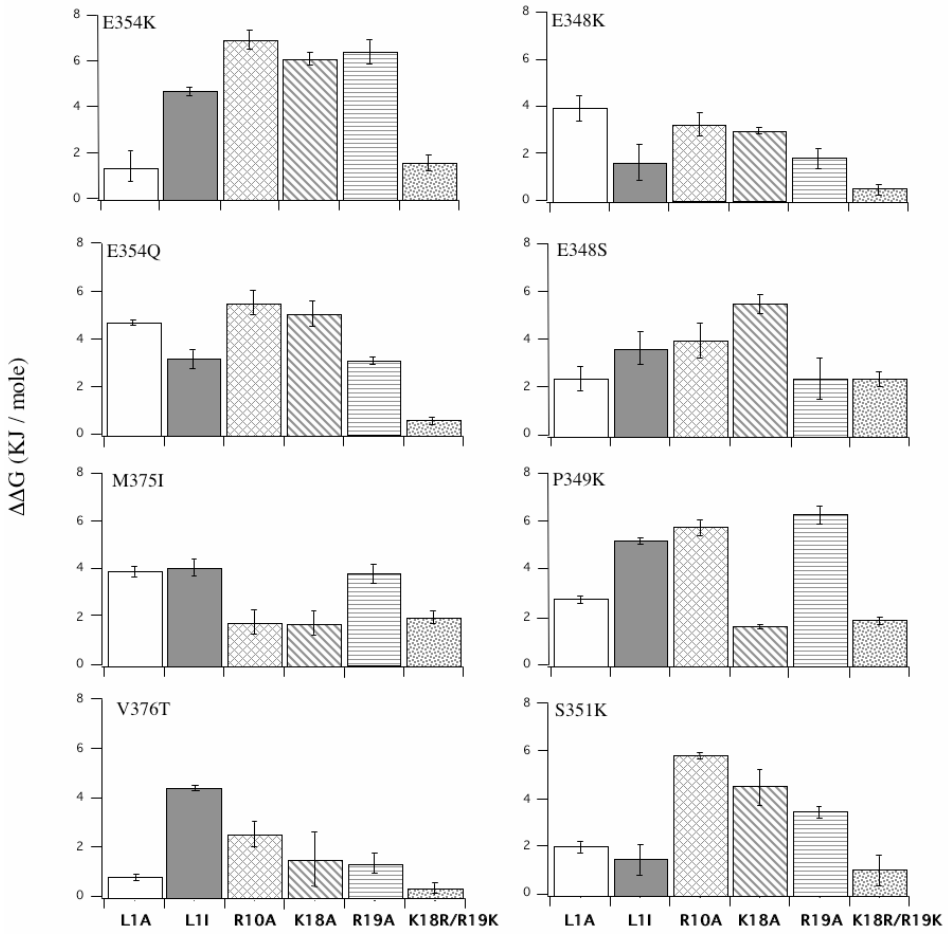


Fig. 3.7 Summary of mutant cycle analysis. Coupling energy (ΔΔG ± SE) between *TSha1* mutants and analogs of κM-RIIIK representing important residues for binding (L1, R10, K18 and R19).

ΔΔG values exceeding 2.0 KJ/mol indicate coupling interaction within 6 Å distance. L1A mutant (empty bars), L1I (gray-filled bars), R10A (net-filled bars), K18A (diagonal-stripped bars), R19A (horizontal-stripped bars) and K18R/R19K (mosaic-filled bars). All numbers represent the means of 3-6 independent determinations. ΔΔG are taken as absolute values.

3.4 The docking models.

The mutant cycling data have been used as intermolecular distance restraints in docking calculations with the aim of obtaining a model for the complex of κM-RIIIK with the *TSha1* channel. The structure of κM-conotoxin RIIIK, determined in aqueous solution by NMR, was docked to the model structure of the *TSha1* channel, using the distance restraints derived from the mutant cycle analysis, as described in Chapter 2. According to what observed by Eriksson and Roux for the Ag2Tx - *Shaker* complex (Eriksson & Roux, 2002), we find that one or two distance restraints are insufficient to achieve a selection among the different orientations of the toxin with respect to the channel pore. Therefore, we chose to use all the available distance restraints in the docking calculations. The 875 generated models can be subdivided into 4 clusters, which have been numbered from I to IV according to increasing total energy (Fig. 3.8). 90% of the generated models belong to Cluster I, which also contains the structure with the lowest energy; 6 % of the models belong to Cluster II, 3% to Cluster III and 1% to Cluster IV. No intermolecular distance restraint violations are found in Cluster I and II while all structures of Cluster III and IV violate the distance restraints E354-R19 and M375-L1, respectively. The backbone conformation of the toxin is well defined in all four clusters (toxin backbone RMSDs are 1, 1.4, 1.1 and 1 Å in cluster I, II, III and IV, respectively). The conformation of the toxin backbone is quite close to the one determined by NMR in Cluster I (RMSD of 1.4 Å), while it deviates the most in Clusters III (RMSD of 2.5 Å). The toxin conformations of Cluster II and IV have RMSD values to the NMR determined conformation of 1.8 and 1.6 Å, respectively.

The details of the atomic interactions for the lowest energy structure of each cluster are given in the following.

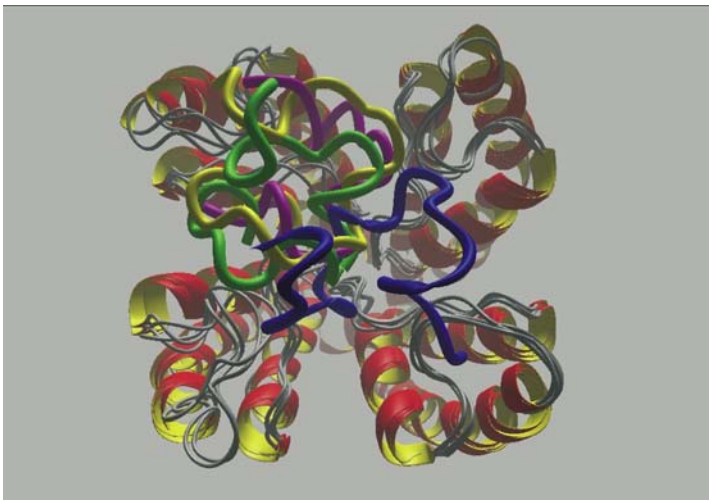


Fig. 3.8. The four models for the complex κM-conotoxin RIIIK – *TSha1* K^+ channel. The backbone of the toxin is represented in blue (cluster I), green (cluster II), yellow (cluster III) and magenta (cluster IV). The toxin deeply interacts with only three units of the channel. In cluster I, units A, B and C are involved in electrostatic and hydrophobic contacts with toxin side chains; in cluster II, III and IV, units A, C and D contact the toxin side chains.

Cluster I: The positively charged side chain of R10 of κM-RIIIK is located in a negatively charged pocket of the extracellular loop of unit A of the *TSha1* channel defined by the carbonyls of P349, E350 and S351 and by the side chains of E348, S351 and E354 (Fig. 3.9). The strong interaction of R10 with E354 accounts for the decreased affinity of κM-RIIIK to the *Shaker* channel, where the homologous position to E354 is occupied by a lysine. The positively charged side chain of K18 is surrounded by the backbone carbonyls of P349, E350, S351 and the side chains of E348, S351 and E354 of unit C, while the R19 interacts with the negatively charged side chain of E354 and with the backbone carbonyl of S351 of unit C (Fig. 3.9). The electrostatic interactions of R10, K18 and R19 anchor the peptide to the channel loops stretching it over the channel pore. The interactions of L1 with the channel unit B are both electrostatic and hydrophobic (Fig. 3.9). The positively charged N-terminus is situated among the backbone carbonyl of E348 and the side chains of Q352 and E354, justifying the consistent decrease in affinity for the N-acetylated toxin; the hydrophobic side chain packs against the side chains of E350 and Q352 and contacts the methyl groups of M375 and V376.
Cluster II: The positively charged side-chain of R10 contacts the backbone carbonyls of P349, E350 and S351 and the side chains of E348 and E354 of unit

A, similarly to cluster I. K18 interacts with the backbone carbonyls of P349, E350 and S351 and with the side-chains of E348, S351 and E354 of unit C; R19 contacts the side-chain carbonyls E354 of the same unit C. As for cluster I, the electrostatic interactions of the side-chains of R10, K18 and R19 with loop residues of unit A and C stretch the peptide over the channel pore. In cluster II, L1 is situated close to the loop residues of unit D, unlike in cluster I, where L1 contacts the extracellular loop of unit B (Fig. 3.8). The positively charged N-terminus group contacts the side chains of Q352 and E354, while the Leu side chain has hydrophobic contacts with P349. The distance between the side chain of L1 and the side chains M375 and V376 in this model is higher than in model I.

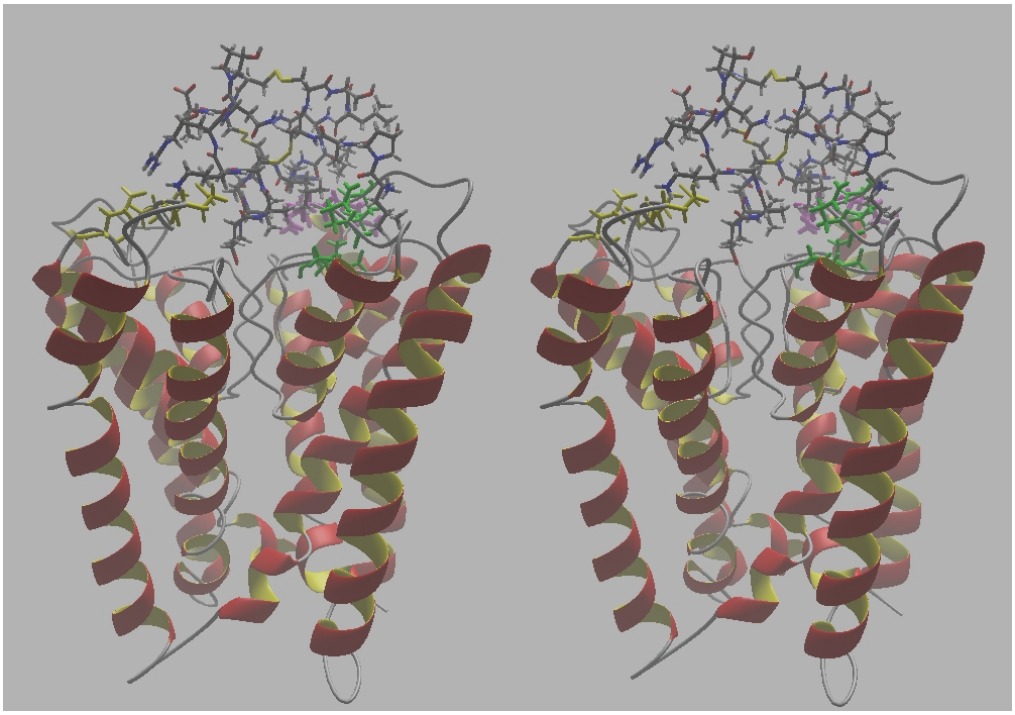


Fig. 3.9 Stereo picture of model I, showing the details of the electrostatic and hydrophobic interactions of the toxin with the channel. The amino acids E348, S351 and E354 of unit A (in magenta) are in deep contact with the amino acid R10 of the toxin; the same amino acids E348, S351 and E354 of unit C (in yellow) contact the side chains K18 and R19 of the toxin, while amino acids E350, Q352, M375 and V376 of unit B (in green) interact with the charged N-terminus and with the hydrophobic side chain of the toxin L1.

Cluster III: As in the Cluster I and II, the positively charged side chain of R10 interacts with the backbone carbonyls of P349, E350 and S351 and with the side

chain carbonyl of E354 of unit A. The amino group of K18 is situated in a pocket formed by backbone carbonyls of P349, E350 and S351 and by the side chain of E354 of unit C, while R19 contacts the backbone carbonyls of P349 and E350 and the side chain of E354 of unit D. The distance between the negatively charged side chain of E354 and the positively charged side chain of R19 is larger than in cluster I and II. The positive charge of the arginine side chain is surrounded by several backbone carbonyls and a serine side chain of the toxin itself. L1 is situated close to the extracellular loop of unit D. The positively charged N-terminus weakly interacts with the backbone carbonyl of E354; the hydrophobic leucine side chain packs against Pro349 and S351, while the distance to M375 and V376 is larger than in cluster I.

Cluster IV: The backbone carbonyls of Pro349 and E350 and the side chain of E354 of unit A interact with R10 as in all other models. The side chains of E348, Q352 and E354 of unit C surround K18, while R19 stretches over towards the backbone carbonyls of E348 and P349 and the side chains of E348 and E354 of unit D. The positively charged N terminus of L1 interacts with the side chain of E354 of unit D, while the Leu side chain only contacts the side chain of Q352 and is further from M375 than in any other cluster.

In all four clusters the positively charged side chain of R10 is situated in a negatively charged pocket of unit A defined by the backbone carbonyl of the stretch 348-351 and by the side chain carbonyls of E348 and E354. This reflects the high ΔΔG values observed upon introduction of a positive charge in position 349, 351 and 354 and explains the higher affinity of the toxin for the *TSha1* channel as opposed to the *Shaker* channel, where a lysine is found in the homologous position to E354. The atomic interactions of K18 with loop residues 348-354 are very similar to those of R10. Thus, the two positively charged side chains of R10 and K18 serve as anchors to very similar negatively charged pockets of the channel extracellular loops and pore helix. An additional anchor is represented by R19 that tightly interacts with the side chain of E354. However, R19 can interact either with the same unit as K18 (unit C), as in Cluster I and II or with the neighbouring unit, as in Cluster III and IV, depending on the conformation of the R19 side chain. This heterogeneity is observed also for structures belonging to the same cluster (1% of the structures of Cluster I and 20%

of the structures of Cluster II). The mutant cycling data and consequently the docking models show that R19 is in closer proximity to the side chain of P349 than K18 ($\Delta\Delta G$ = 6.31 $kJmol^{-1}$ for the R19A-P349K mutation and 1.67 $kJmol^{-1}$ for the K18A-P349K mutation).

More complex are the interactions of L1. Mutant cycling data point to the existence of contacts between L1 and E348, P349, E354, M375 and V376. The distance restraints with P349 and E354 are classified as medium and are reflected in electrostatic contacts of the positively charged N-terminus of L1 with the side chain carbonyl of E354. The hydrophobic contacts of L1 are different in the four models: the weak distance restraints with M375 and V376 are best satisfied in Cluster I and II, where L1 reaches M375 and V376 the closest.

Given its small size, κM-RIIIK cannot interact with the extracellular loops of all four units. In all four clusters the toxin contacts three out of four units of the K^+ channel. In model I, unit A, B and C are binding the toxin, while in model II, III and IV unit A, C and D are involved in contacts with the peptide (Fig. 3.8). According to the NMR studies, the toxin has a discoid shape: in model I the flat surface of the disc is out of the plane defined by the four channel loops, while in model II, III and IV the toxin is flattened on the channel like a pancake. Although the toxin contacts only three of the four units, the accessibility for a potassium ion to the selectivity filter of the channel is reduced by more than 85% in all models.

Table 3.5

IC_{50} Values for *TSha1* mutant K^+ block of κM-RIIIK WT and analogs.

The IC_{50} values (in nM) are given by Mean values ± SE, n = number of independent experiments. nd: not done; nb: no block

	κM-RIIIK analogs IC_{50} [nM] (Mean ± SEM); (n)								
	WT	**L1A**	**L1I**	**AcL1**	**L1M**	**R10A**	**K18A**	**R19A**	**K18R / R19K**
WT	76 ± 10 (9)	3180 ± 120 (6)	1380 ± 450 (5)	7620 ± 380 (5)	340 ± 50 (5)	4220 ± 330 (7)	3900 ± 530 (6)	1530 ± 240 (9)	180 ± 40 (7)
E348S	110 ± 30 (4)	1920 ± 290 (4)	480 ± 120 (3)	nd	nd	1220 ± 280 (3)	600 ± 100 (3)	820 ± 210 (4)	100 ± 10 (3)
E348K	50 ± 10 (3)	440 ± 110 (3)	540 ± 100 (4)	nd	nd	770 ± 100 (4)	720 ± 50 (4)	460 ± 60 (4)	150 ± 10 (3)
P349K	130 ± 20 (3)	1740 ± 110 (3)	280 ± 20 (5)	nd	nd	680 ± 100 (3)	3390 ± 100 (3)	200 ± 30 (3)	140 ± 10 (4)
S351K	160 ± 10 (4)	2960 ± 310 (4)	1700 ± 390 (4)	nd	nd	810 ± 40 (3)	1390 ± 350 (3)	770 ± 80 (4)	260 ± 60 (3)
E354K	490 ± 110 (3)	14070 ± 4740 (3)	1310 ± 100 (3)	nd	nd	1640 ± 300 (3)	2070 ± 220 (3)	740 ± 140 (3)	620 ± 80 (3)
E354Q	270 ± 20 (6)	1660 ± 80 (5)	1400 ± 200 (5)	nd	nd	1700 ± 340 (4)	1870 ± 350 (4)	1550 ± 110 (5)	490 ± 30 (4)
M375K	10730 ± 2460 (3)	nb	nb	nb	nb	nb	Nb	nb	nb
M375I	1360 ± 70 (5)	11400 ± 1030 (3)	4740 ± 580 (4)	26610 ± 2860 (3)	1360 ± 10 (3)	37400 ± 7260 (3)	35160 ± 7330 (3)	5910 ± 790 (4)	1420 ± 150 (3)
M375L	140 ± 40 (3)	3260 ± 460 (4)	690 ± 40 (3)	11400 ± 4060 (3)	470 ± 50 (3)	nd	Nd	nd	nd
V376T	120 ± 20 (3)	3550 ± 220 (3)	580 ± 180 (3)	nd	nd	2490 ± 520 (4)	4650 ± 1800 (3)	3990 ± 1070 (3)	310 ± 50 (3)

Chapter 4

Discussion

4.1 The mutational analysis and the NMR structure of κM-conotoxin RIIIK.

The solution conformation of κM-conotoxin RIIIK differs substantially from that of κ-conotoxin PVIIA, the only other structurally characterized conotoxin that interacts with the K^+ channels. While the two conotoxins share common binding sites on the *Shaker* K^+ channel, they dramatically differ in the disulphide bond pattern, secondary structure elements and charge distribution. κM-conotoxin RIIIK is encoded by a gene belonging to the M-superfamily, while κ-PVIIA belongs to the O-superfamily. The latter superfamily includes ω-conotoxins, which are targeted to voltage-gated Ca^{2+} channels. The most well-characterized member of the M-superfamily is μ-conotoxin GIIIA, an extremely specific blocker of the $Na_v1.4$ voltage-gated Na^+ channel subtype. The structure of κM-RIIIK is much more similar to that of μ-GIIIA than to that of κ-PVIIA. Clearly, the shared class III Cys pattern of kM-RIIIK and μ-GIIIA is the dominant determinant of the overall structure. Despite the structural similarity, κM-RIIIK and μ-GIIIA show no overlap in the pharmacological specificity at all.

The structural divergences between κM-RIIIK and κ-PVIIA pose the intriguing question of whether the mechanism of the K^+ channel conductance block of the two toxins significantly differs. Despite their differences in size, amino acid sequence, structure and biological origin, all other well characterized toxins (e.g dendrotoxin from snakes, various scorpion and sea anemone toxins), including κ-PVIIA, share a number of convergent features. The key amino acid determinants for the interaction with the K^+ channel are localized on one side of the toxin surface and the pharmacophore is based on a hydrophobic/positively-charged amino acid dyad motif. Encouraged by the similarity in the mechanisms of interaction with the K^+ channels of structurally and phylogenetically different

peptides, we had expected κM-conotoxin RIIIK to fit in the common framework and to interact with K^+ channels with a dyad-based pharmacophore.

Conversely, the structural and mutational analyses described above indicate that κM-RIIIK does not fit the established paradigm. κM-RIIIK is a relatively flat disc-shaped molecule in which important amino acid determinants for binding are not clustered on the toxin surface into a clearly defined pharmacophore (Fig. 4.1). Most importantly, κM-RIIIK does not appear to have a dyad motif composed of a hydrophobic amino acid and a lysine residue. Our conclusion that a dyad does not exist is based on two types of experimental data. First, the structure shows no hydrophobic residue within 6 Å of a positively charged residue. Second, we found that substitutions of aromatic residues for Leu1, the functionally most important hydrophobic amino acid in κM-RIIIK, are even less well tolerated than Ala. Furthermore (and most unexpectedly), substitution of a positive amino acid gave an affinity close to the wild-type peptide. In fact, the Leu1 substitution with the smallest effect on the affinity of the peptide is an arginine (L1R) and not isoleucine (see Table 3.3). An affinity similar to the substitution L1R was observed for the methionine substitution (L1M). Since methionine is a smaller residue than arginine and on the other hand the lysine substitution (L1K) leads to a bigger reduction in the affinity of the peptide than L1R, it is likely that a combination of size-dependent hydrophobic and electrostatic interactions involving the first residue of this peptide are important for binding to the channel. Taken together all these results are completely inconsistent with Leu1 being the hydrophobic partner in a dyad. Interestingly, the dyad hypothesis has been challenged by Mouhat et al. (Mouhat *et al.*, 2004), where it is shown that the functional dyad K24-Y33 of the scorpion toxin Pi1 is not indispensable for recognition and binding to the voltage-gated Kv1.2 potassium channel. This conclusion, drawn from the observation of a partial but specific current reduction after application of the mutant [A24,A33] – Pi1, lacking the functional dyad, confirms our results and opens the way to considering different possible mechanisms for the interaction of toxins with K^+ channels.

κM-RIIIK, the Na^+ channel blocker μ-GIIIA and the non-competetive antagonist of nicotinic acetylcholine receptors ψ-PIIIF all share a common Class III Cys pattern. The structural features of the three peptides are very similar towards the

C-terminus where all three conotoxins are folded in two distorted helical turns (Fig. 3.2). However, towards the N-terminus the three peptides show substantial differences both in structure and dynamics. While μ-GIIIA and ψ-PIIIF have a well-defined backbone folded in a series of turns, κM-RIIIK is highly flexible in the region 1-11. It should be emphasized that the flexibility of the peptide is not an artifact due to poor structural restraints but an intrinsic dynamic feature of the molecule, as confirmed by the averaged values of the $^{3}J(H_N,H_\alpha)$ couplings, the small values of the H_N dipolar couplings and the higher values of the heteronuclear ^{13}C-^{1}H NOEs in this region. We suggest that the differences in the three-dimensional form, provided by the structural heterogeneity at the N-terminus, and in the surface charge distribution are an important determinant of target selectivity.

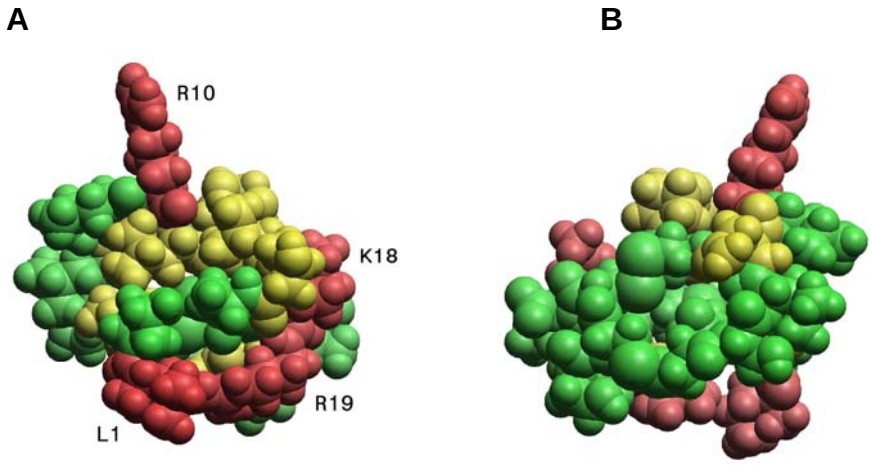


Fig. 4.1 A) Map of the functionally important residues of κM-conotoxin RIIIK on the three-dimensional structure. To visualize the different changes in the affinity observed for the corresponding alanine mutants the residues are color-coded. red: alanine substitution increased the IC_{50} of κM-RIIIK by more than 20-fold (L1, R10, K18 and R19); yellow: alanine substitution affect the affinity by more than 5-fold but less than 16-fold (S3, N8, L11, O13, V14, O15, N20): green: all the rest. B) same as in A) but from the opposite side.

The localization of the functionally critical amino acids on the three dimensional structure of κM-RIIIK is shown in Fig. 4.1. The four most important residues (L1, R10, K18 and R19) and the seven residues of medium importance are well distributed across the whole molecule, as opposed to the commonly found

structural motif for K^+ channel blockers that contains a dyad of a positively charged and an hydrophobic residue. The separation of key residues in κM-RIIIK resembles μ-GIIIA more than other K^+ channel-targeted toxins. However, all functionally relevant residues can be localized on a surface of 12 x 9 Å (Fig. 4.1A) with the most important residues (in red) being situated at the edges and the residues of medium relevance (in yellow) in the middle of the surface. On the opposite side of the peptide (Fig. 4.1B), no residue seems to be essential for function.

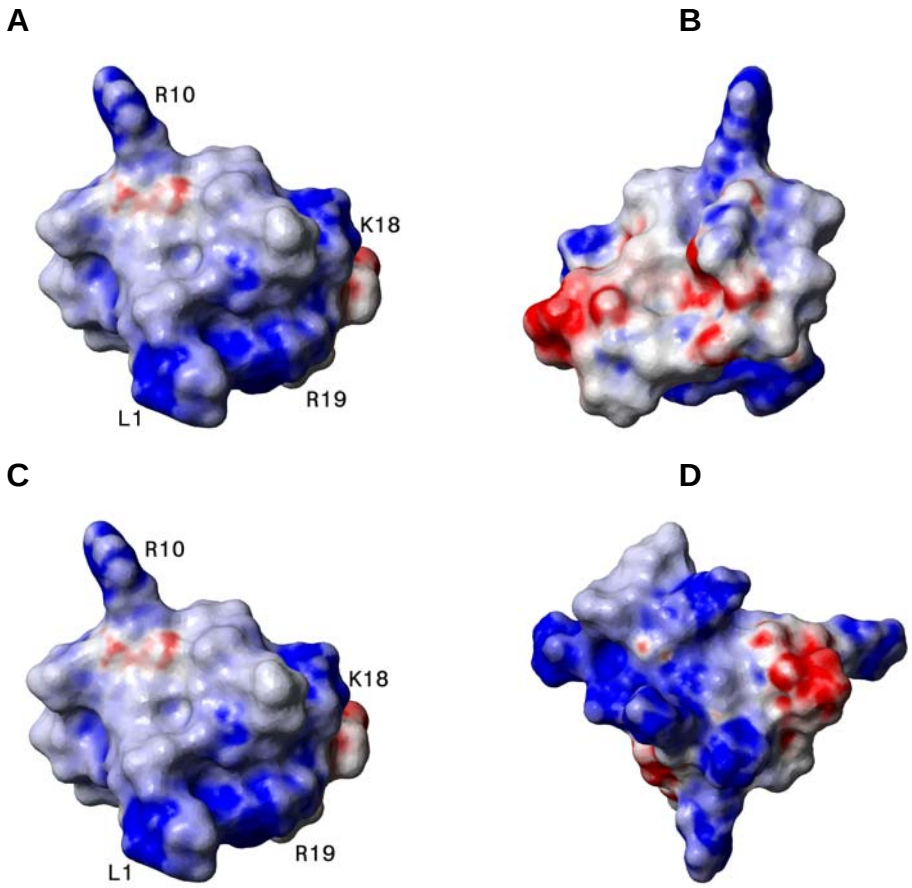


Fig. 4.2 Electrostatic surface potential for κM-conotoxin RIIIK (A,B) and μ-contoxin GIIIA (C,D). The orientation of κM-conotoxin RIIIK is the same as in Fig. 4.1. Intense blue(red) regions correspond to a charge of +1.0(−1.0) or greater. The surface potential was calculated with the AvgCharge algorithm of MOLMOL 2k.2.

The charge distributions on the surfaces of κM-RIIIK, μ-GIIIA and κ-PVIIA are substantially different. In κM-RIIIK, most of the positive charges are distributed at

the edges of the peptide face containing the functionally essential amino acids (Fig. 4.2A). It is noteworthy that in κM-RIIIK all determinants for binding contain a positive charge, and that all positively-charged residues are necessary for binding. This ring of positive charges could be used as an anchor to residues of the K^+ channel loops. In μ-GIIIA and κ-PVIIA, the distribution of both charges and binding-determinant residues on the peptide surface is more homogeneous than in κM-RIIIK. In μ-GIIIA, however, a ring of positive charges could be similarly identified at the height of R19 and R1 (Fig. 4.2C,D). In contrast to κM-RIIIK, a functionally highly relevant residue in μ-GIIIA, R13, sticks out from the surface defined by this ring and is probably projected into the channel pore. In κM-RIIIK, no additional positively charged residue sticks out from the surface defined by the ring of positively charged residues L1, R10, K18 and R19. This observation, together with the fact that κM-conotoxin RIIIK lacks a functional dyad, does not support the model of a positively charged side chain occluding the channel pore. On the contrary, the even distribution of L1, R10, K18 and R19 at the edges of the peptide face containing all functionally relevant residues suggests that conotoxin κM-RIIIK may block the channel by covering the pore as a lid. This pharmacophore model represents a novel mechanism of K^+ channels block that has not been found in any other K^+ channel-targeted peptide characterized to date. Mouhat et al. (Mouhat *et al.*, 2004) demonstrate the presence of a basic ring in scorpion toxin Pi1 that might be similar to the one described here, indicating that a similar pharmacophore has been independently evolved by cone snails and scorpions. Mutation and docking data reveal the importance of a cooperative electrostatic interaction of the basic residues of this ring in scorpion toxin Pi1 with side chains of the turret region of the channel. These results reinforce our proposition of a different model of peptide-channel interaction that is not centered on a functional dyad.

4.2 Mutant cycle analysis and docking model.

We have performed docking calculations using the mutant cycling data as intermolecular distance restraints with the aim of obtaining a model for the

complex of κM-RIIIK with the *TSha1* channel. The structure of κM-conotoxin RIIIK, determined in aqueous solution by NMR, was docked to the model structure of the *TSha1* channel, using the distance restraints derived from the mutant cycle analysis. As described in 3.4, the generated models can be subdivided into 4 clusters, which have been numbered from I to IV according to increasing total energy (Fig. 3.8).

To discriminate between the four models, we analysed the location of the functionally relevant residues of the toxin with respect to the channel pore (Fig. 4.3). The mutational analysis identified L1, R10, K18 and R19 as being fundamental for binding to the K^+ channel (red in Fig. 4.3); S3, N8, L11, O13, V14, O15 and N20 resulted to be of medium importance (yellow in Fig. 4.3), while mutations of O2, S6, L7, L9, O21 and T24 to alanine did not significantly reduce the activity of the toxin (grey in Fig. 4.3). Model I is in very good accord with the mutational data (Fig. 4.3A). All functionally relevant residues, with the exception of N8, are facing the channel pore or are in contact with the loops, while no residue that is unessential for function strongly interacts with the channel. The only *grey* residue facing the channel pore is A16, for which no mutational data are available. Model II fits the mutational data worse than model I. Three residues of medium importance for binding (S3, N8 and L11) are not in contact with the channel while the C terminus of the toxin, whose side chain was found to be irrelevant for binding, is contacting the loop of unit C (Fig. 4.3B). In model III the residues 21-24 at the C terminus of the toxin are involved in several contacts with units C and D, while the stretch 11-15, which contains several residues of medium importance for function, points towards solution (Fig. 4.3C). In model IV, the C terminus of the toxin deeply penetrates the channel and a great portion of the residues that are unimportant for binding interacts with the protein (Fig. 4.3D).

The fact that a whole stretch of functionally relevant toxin residues (residues 11 to 15) is not in contact with the channel in model III and IV cannot be taken as the sole argument to exclude these models, as mutations at this site might destabilize the overall structure of the toxin. However, it is rather improbable that amino acids profoundly interacting with the protein (stretch 21-24) can be mutated to alanine without influencing affinity. All in all the mutational analysis data are

severely contradicted by models III and IV, allowing us to exclude them as valid representations of the complex between κM-RIIIK and the *TSha1* channel. Model II is retained as this stage, as it does not contradict the mutational data as extensively as model III and IV.

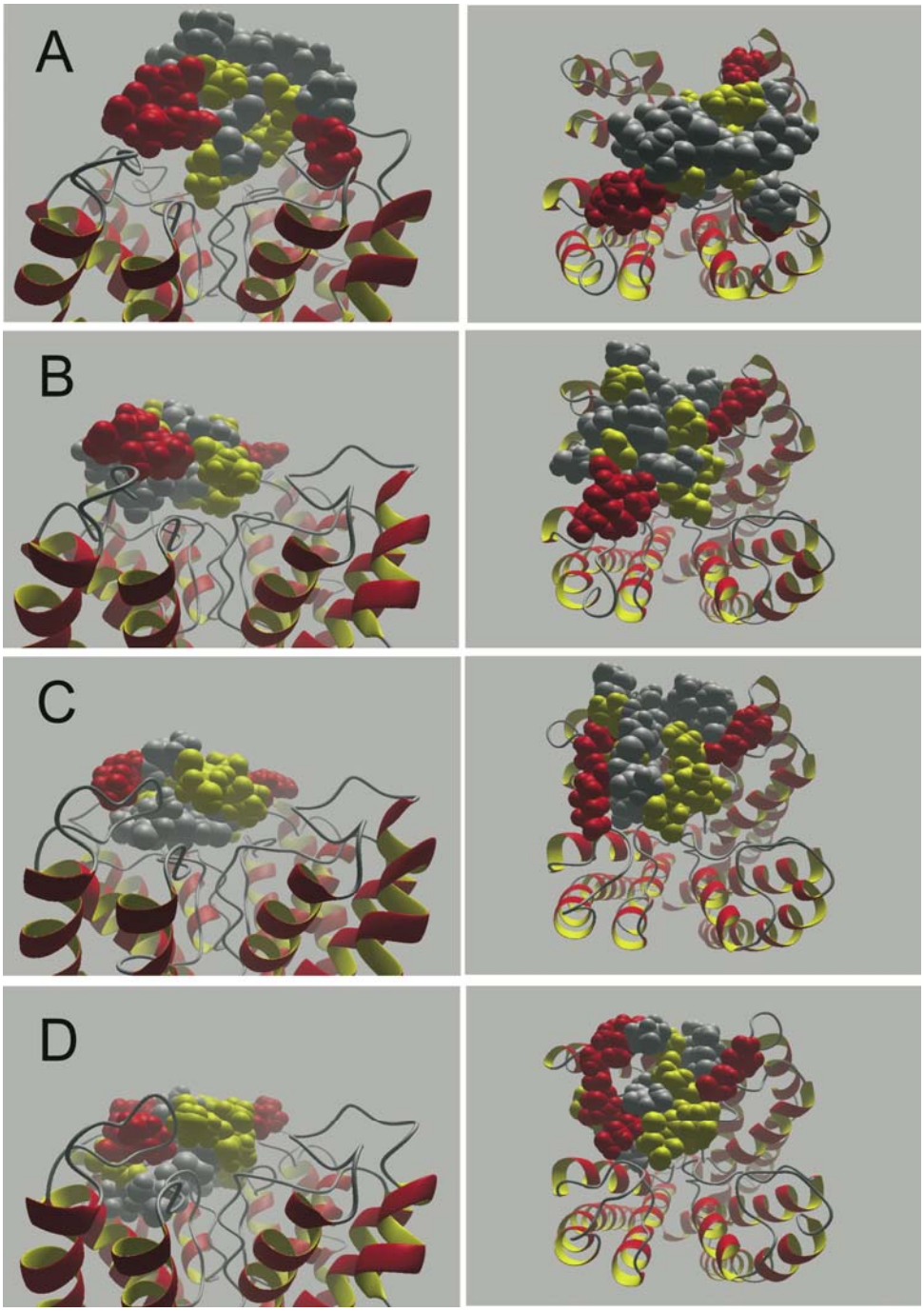


Fig. 4.3. Filling space representation of κM-conotoxin RIIIK in the four clusters (Cluster I-IV in panels A-D), showing the mutational data obtained by systematic alanine substitution in color

code: the functionally most relevant residues are represented in red (L1, R10, K18 and R19); the residues of medium importance for binding (S3, N8, L11, O13, V14, O15 and N20) are shown in yellow, while the residues that are irrelevant for binding (O2, S6, L7, L9, O21 and T24) are in grey. Two orientations are provided for each of the four models: from the side (left panel) and from the top (right panel) of the channel.

To further discriminate between cluster I and II, we analysed which part of the toxin is in close proximity to the selectivity filter. For different toxins, which interact with voltage gated K^+ channels, there is evidence that a positively charged side chain occludes the channel pore. This positively charged amino acid is part of the functional dyad, together with an amino acid containing a hydrophobic side chain, usually a tyrosine. In our previous work, we demonstrated that κM-conotoxin RIIIK does not possess a functional dyad. Moreover, as a result of the mutant cycling analysis, all positively charged amino acids interact with the turret or pore helix regions of the channel and therefore cannot penetrate the channel pore. Which toxin residues are then facing the channel pore in our models? In model I the O15 is very close to the selectivity filter of units C and B and the hydroxyl oxygen at the Cγ is at a hydrogen bond distance (2.8 Å) from the carbonyl of the G373 of the selectivity filter of unit B (Fig. 4.4A). In model II the O13 is in proximity of the selectivity filters of unit A and D, but makes no tight direct contact with them (Fig. 4.4B). To distinguish between the two models, mutant cycle data involving the residues of the GYG selectivity filter would be required. This is unfortunately not feasible as any mutation of residues in the selectivity filter destroys the channel activity. Nevertheless, model I and model II can be distinguished by considering the distance of O13 or O15 from the amino acid V376 of the *TSha1* channel. While in model I O15 is expected to be closer to V376 (of unit C) than O13, in model II both O13 and O15 are at a similar distance from V376 of unit D and C, respectively. The mutant cycle analysis performed using *TSha1* V376T resulted in a $\Delta\Delta G$ of 0.6 $kJmol^{-1}$ for O13A and of 2.51 $kJmol^{-1}$ for O15A, which reflects the geometry of model I much better than that of model II. To obtain further evidence for the interaction of O15 with the amino acids of the selectivity filter we assayed the affinity of the O15R and O15K mutants, following the idea that a positive charge should be well tolerated but the bulky arginine should not be easily accommodated in proximity of the tight channel

pore. The affinity of the mutant O15K for the wild type *TSha1* channel was found to be 430 ± 20 nM (n = 4) against a value of 930 ± 160 nM (n = 3) for the mutant O15A. This indicates that a positive charge in position 15 is preferred to the absence of any polar group, as would be expected if the side chain of O15 interacts with the carbonyls of the selectivity filter. On the other hand O15R binds the channel with a k_D of 2940 ± 540 nM (n = 4), suggesting that O15 is situated close to the channel pore where there is no space to accommodate the bulky arginine side chain. Although these data do not prove directly the interaction of O15 with residues of the selectivity filter, they strongly support it.

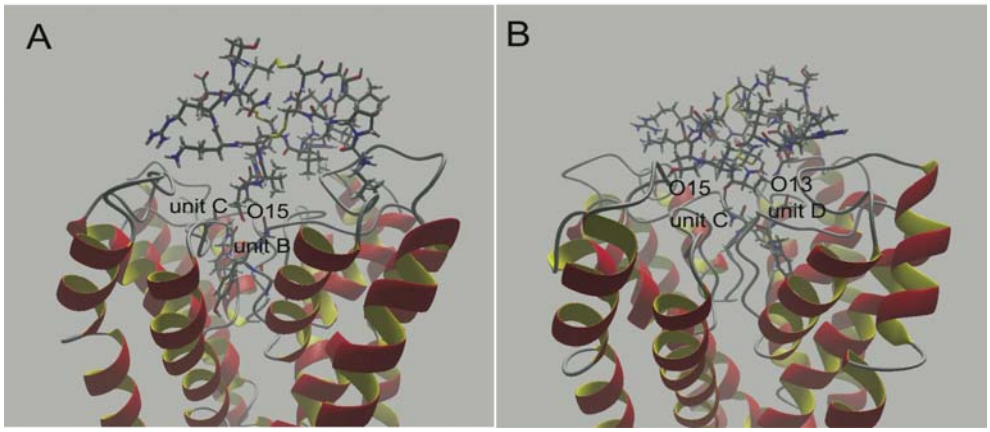


Fig. 4.4 Interactions of the residues O13 and O15 of κM-conotoxin RIIIK with the selectivity filter of the channel. In the two panels A and B model I and II are shown, respectively. The two representations differ for a rotation of about 180º around the channel axis to allow for a better visualization of O15 in the two models. A) Model I: the Cγ-OH group of O15 is close to the selectivity filter of unit B and C and in H-bond distance from the carbonyl oxygen of G373 of unit C. B Model II: both O13 and O15 are close (< 5Å) to the selectivity filter of unit C and D, but no specific interaction can be identified.

Model I provides a novel pharmacophore for the interaction of toxin peptides with K^+ channels that best explains the mutational and mutant cycle data for the κM-RIIIK – *TSha1* complex. The peptide is anchored to the extracellular loops of three out of the four channel units via a basic ring provided by the side chain of R10, K18, R19 and the protonated N-terminus of L1, and no positively charged residue occludes the channel pore.

Unlike several other peptides that target voltage activated K^+ channels, κM-conotoxin RIIIK does not contain a functional dyad of a hydrophobic and a positively charged side chain, which was proposed to be the minimal functional core for K^+ channel blocking peptides (Dauplais *et al.*, 1997; Savarin *et al.*, 1998; Gilquin *et al.*, 2002; Srinivasan *et al.*, 2002). The dyad model has been recently challenged by Mouhat et al. (Mouhat *et al.*, 2004), who demonstrated by mutational analysis and docking calculations that a functional dyad is not strictly required for binding of the scorpion toxin Pi1 to the voltage-gated Kv1.2 potassium channel. Instead, a basic ring of four positively charged side chains was proposed to be a fundamental recognition element by the residue D355 of each of the four Kv1.2 α-subunits. Similarly, we find that four positively charged residues of κM-conotoxin RIIIK are involved in electrostatic interactions with loop and pore helix amino acids of three subunits of the *TSha1* channel. Residue E354 of different units is involved in contacts with all four positively charged residues of the peptide. This amino acid distinguishes the *TSha1* channel, which is the highest affinity target of κM-RIIIK (IC50 = 76± 10 nM), from other channels of the same family, as for example the *Shaker* (IC_{50} = 1.2 μM) (Ferber *et al*., 2003), where the homologous position is occupied by a lysine (K427), and $K_v1.2$ (IC_{50} = 400 nM) (Ferber *et al*., 2004) where the homologous position is occupied by a neutral proline (P359). Our model suggests that E354 is a fundamental recognition element for κM-RIIIK binding to the *TSha1* channel. However, the complete lack of affinity for κM-RIIIK shown by other members of the K_v1 family (Ferber *et al*., 2003), which, like $K_v1.2$, contain an uncharged amino acid at position 359, indicates that channel amino acids other than P359 play a comparably important role in toxin recognition. For the toxin it is reasonable to hypothesise that the spatial distribution of the positive charges of the basic ring exerts a fundamental role in channel recognition and provides a solid framework for the high selectivity of the K^+ channel-blocking toxins among the diverse targets.

As shown in the NMR-derived solution structure of the peptide, the basic ring defined by L1, R10, K18 and R19 separates the peptide in two faces: all functionally relevant residues, identified by systematic alanine mutations, cluster on one face, which is in contact with the channel in model I.

The mutant cycle analysis excludes that one of the positively charged side chains of κM-RIIIK penetrates the channel pore, as they all interact with amino acids of the extracellular loop and pore helix. In model I the Cγ-hydroxyl of O15 tightly interacts with the carbonyl of G373 of the unit C selectivity filter. It has been proposed that the carbonyls of the selectivity filter are major players in the coordination of K^+ ions at the entrance of the channel pore (Doyle *et al*., 1998). Dehydration of the ions must occur at the selective filter, as a hydrated K^+ is too large to enter the channel pore. By acting as electron donors, the carbonyls of the selectivity filter replace the water molecules usually associated with K^+ ions in solution, thus compensating for the large energetic penalty of the dehydration process. A H-bond between the Cγ-hydroxyl of O15 would block one of the selectivity filter carbonyls, seriously disturbing the energetic of the K^+ dehydration process.

A thorough analysis of the binding mode of various scorpion toxins to diverse K^+ channels (Rodriguez de la Vega *et al.,* 2003, Xu *et al*., 2004) revealed three possible binding modes, which involve different faces of structurally similar toxins: the *internal binding mode*, where residues of the turret region, pore helix and selectivity filter of the channel are involved; the *intermediate mode*, involving residues of the turret region; the *external mode*, which involves residues far away from the selectivity filter. In the *intermediate* and *external binding modes* no functional dyad is observed on the peptide site. This analysis, triggered by the observation that the scorpion toxin BmTx3 can block both A-type K^+ and HERG currents (Huys *et al*., 2004) using different functional faces, confirms the existence of different binding modes of peptidic toxins to K^+ channels, some of which do not involve a direct obstruction of the pore by a positively charged side chain. The mutant cycle data and the docking model I prove the interaction of κM-RIIIK residues with amino acids in the turret and pore helix regions and to some extent also with the selectivity filter. In the classification of Rodriguez de la Vega et al. (Rodriguez de la Vega *et al*., 2003), the position of the toxin in the κM-RIIIK – *TSha1* channel complex resembles the *internal* binding mode best. However, the side chain of O15 does not completely occlude the channel pore, as the lysine of the functional dyad, but rather disturbs the charge distribution at the selectivity filter by involving one of the carbonyls in a H-bond.

Despite the entirely different pharmacological specificity and many differences in the amino acid sequence, κM-conotoxin RIIIK shows striking structural similarities to μ-conotoxin GIIIA, which specifically blocks $Na_v1.4$ Na^+ channels. κM-conotoxin RIIIK and μ-conotoxin GIIIA belong to the same phylogenetic family and share a common pattern of cysteine bridges. Interestingly, O15 of κM-conotoxin RIIIK is at the homologous position of R13, the Na^+ channel pore occluding residue in the peptide sequence of μ-conotoxin GIIIA. This fact, together with the high structural similarity of the two toxins, suggests that κM-conotoxin RIIIK and μ-conotoxin GIIIA might block K^+ and Na^+ channels, respectively, with a similar geometry. However, being the O15R mutant of κM-RIIIK inactive on $Na_v1.4$ sodium channels (data not shown), additional parameters must be involved in determining the selectivity of the two peptides for Na^+ or K^+ channels. Similarly to κM-RIIIK, a ring of positively charged side chains can be identified on μ-GIIIA as well. Most likely the overall shape of the two peptides and the charge distribution of the basic ring are key determinants for the Na^+ or K^+ channel selectivity of μ-conotoxin GIIIA and κM-conotoxin RIIIK.

Part 3

TAR-RNA recognition by positively charged ligands.

Chapter 1

Introduction

1.1 The biological system.

The TAR-RNA. The *trans*-activating region (TAR) RNA of the human immunodeficiency virus (HIV) is a particularly attractive example of the RNA-based drug discovery process. The formation of full-length viral transcripts requires the interaction of the TAR-RNA (Scheme 1.1) with the transcriptional activator protein, Tat (Frankel, 1992). Shortly after promoter clearance, RNA polymerase II stalls at the TAR-RNA site due to the action of the negative elongation factor. The Tat/TAR complex recruits the positive transcription elongation factor complex (P-EFTb), which activates the RNA polymerase II stalled at the TAR-RNA site and enables viral replication (Wei *et al*., 1998; Garber *et al*., 1998). Disruption of the Tat/TAR interaction inhibits the production of complete viral transcripts, as the P-EFTb complex cannot bind the TAR-RNA in the absence of the Tat-protein (Wei *et al*., 1998). Thus, the interaction of the TAR-RNA with both the Tat-protein and the P-EFTb complex represents an appealing target for anti-viral therapy.

The TAR-RNAs from HIV-1 and HIV-2 differ in the size of the bulge, which comprises three nucleotides (UCU) in the HIV-1 and two nucleotides (UU) in the HIV-2. Both TAR-RNAs recognize the Tat-protein in a similar manner. Of critical importance is the U23 nucleotide in the TAR bulge and the (G26-C39) and (A27-U38) base pair in the upper stem (Cordingly *et al.*, 1990; Roy *et al.*, 1990; Weeks & Crothers, 1991). Mutational analysis have shown that U23 is the only essential nucleotide for the TAR-Tat interaction (Weeks & Crothers, 1991; Churcher *et al*., 1993). The other bulge nucleotides (two in the HIV-I TAR and one in the HIV-II TAR-RNA) can be freely substituted, even with a non-nucleotide linker (Churcher *et al*., 1993; Sumner-Smith *et al*., 1991) and merely act as spacers. NMR studies of the HIV-2 TAR-RNA in complex with argininamide, a mimetic of the Tat-protein, suggest the formation of a base triple

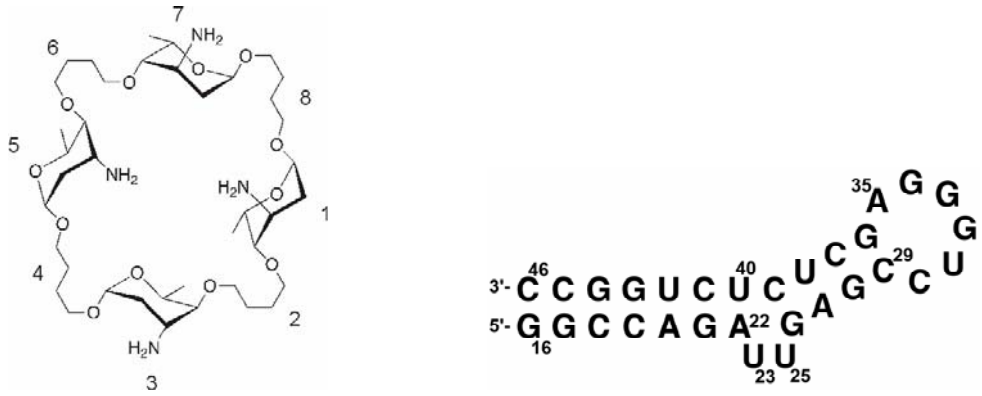
between U23, A27 and U38, which helps widen the major groove of the upper stem and makes it more accessible to the incoming argininamide (Brodsky & Williamson, 1997; Puglisi *et al.*, 1992).

The subunit cyclin T1 (CycT1) of the P-EFTb complex binds the Tat/TAR complex contacting directly the transactivation domain of the Tat-protein and the loop of the TAR-RNA (Zhu *et al.*, 1997; Mancebo *et al*., 1997). Mutational analysis revealed G32 and G34 of the hexanucleotide loop of the TAR-RNA to be crucial for CycT1 binding, while the presence but not the nature of the nucleotide in position 35 is relevant for complex formation (Richter *et al*., 2002). Additionally, the formation of a transient base pair between C30 and G34 has been suggested to favor the binding of the CycT1 subunit of the P-TEFb complex to the Tat-TAR complex (Richter *et al.*, 2002).

Given the relevance of the TAR-RNA as a potential therapeutic target, much work has been directed towards the synthesis and screening of small organic compounds, which selectively bind to the TAR-RNA, thereby inhibiting the Tat/TAR binding event. As a result of these studies, a number of small molecules have been found to interact with the TAR-RNA inhibiting viral transcription in HeLa cells (Arzumanov *et al.*, 2001; Mischiati *et al.*, 2001; Kaushik *et al.*, 2002).

Interaction of the TAR-RNA with aminoglycosides. The cationic antibiotics of the aminoglycoside family bind to several RNA motifs by specifically recognizing their three-dimensional structure (Hermann & Westhof, 1998). In particular, they have been proven to exert an inhibitory effect on the binding of the Tat-protein to the TAR-RNA (Mei *et al.*, 1995 and 1998, Wang *et al.*, 1998; Hamy *et al.*, 1998). For the HIV-1 TAR-RNA, brownian dynamics simulations have identified three "hot-spots", which are more prone to be targeted by cationic antibiotics because of higher electro negativity in these pockets. The predicted pockets are on top of the U31-GG-G34 loop, in the U23-C-U25 bulge and in the major groove of the lower stem (Hermann & Westhof, 1999). Polycationic aminoglycosides are predicted to simultaneously interact with the lower and upper stem regions, thus inhibiting the binding of Tat or Tat-derived peptides (Hermann & Westhof, 1999). The structural basis of the interaction of the TAR-RNA with aminoglycosides has been described for neomycin B by NMR (Faber *et al.,* 2000): neomycin B binds the TAR-RNA in the minor groove of the lower stem, which results in an

"allosteric" inhibition of Tat binding to TAR-RNA caused by a reduction of the major groove volume in the upper stem. Interestingly, to date aminoglycosides have not been found to compete with Tat for binding the bulge region of the TAR-RNA, as it has been proven for other small organic molecules possessing an aromatic scaffold and one or more guanidium groups (Davis *et al*., 2004).



1 **HIV II TAR-RNA**

Scheme 1.1

Aminoglycosides are valuable instruments to study the structural basis of RNA recognition and function inhibition by cationic organic molecules. However, their use as RNA-targeting drugs is impeded by several drawbacks, such as low specificity, high toxicity and increasing development of resistance (Wilson & Li, 2000; Michael & Tor, 1998; Mingeot-Leclercq & Tulkens, 1999). These considerations underline the continuous need for new compounds that share an aminoglycosidic scaffold but can be easily derivatized to obtain drug leads with a better pharmacological profile. The group of Prof. Kirschning at the University of Hanover ha conducted a program directed towards the synthesis of a new class of aminoglycoside analogues (Kirschning & Chen, 1999; Kirschning & Chen, 2000; Chen & Kirschning, 2002). These potential RNA ligands consist of aminodeoxy hexoses and the corresponding cyclitols. Aminocyclitols do not contain the chemically and enzymatically labile acetal group (glycosidc bond), which make them particularly attractive targets. This class of aminoglycoside ligands is unique in the fact that they are easily synthesized via metathesis olefination and can be derivatized straightforwardly. Dot blot analysis, the results of which have been discussed in a previous publication (Kirschning *et al*., 2004), point out that the cyclic neooligoaminodeoxysaccharide (Scheme 1.1) has a much greater binding

affinity to the TAR-RNA in comparison to its linear counterparts. The aminoglycoside analogues can be obtained as ^{13}C and ^{15}N labelled derivatives at reasonable costs for NMR studies (Jaunzems *et al.*, 2004).

Here we present the structure of the HIV-2 TAR-RNA bound to the cyclic neooligoaminodeoxysaccharide **1** (Scheme 1.1). The structural analysis conducted here reveals attractive features in the interaction of this ligand with the TAR-RNA. The aminoglycoside analogue **1** binds to two sites of the TAR-RNA: the bulge and loop regions on the major groove side of the upper stem and the minor groove of the lower stem. Like argininamide, **1** contacts the crucial residues of the TAR-RNA bulge required for Tat binding. This is in contrast to natural aminoglycosides, like neomycin B, which interact exclusively with the minor groove of the TAR-RNA lower stem (Faber *et al.,* 2000). Additionally, we detect extensive contacts of **1** to the A35 of the dynamic hexanucleotide loop, in agreement with digestion experiments, indicating protection of the loop guanines upon addition of the ligand to the TAR-RNA. The NMR studies conducted on both free and argininamide-bound TAR failed to reveal the presence of well-structured elements in the hexanucleotide loop of the TAR-RNA (Brodsky & Williamson, 1997; Colvin *et al.*, 1993; Jäger & Tinoco, 1993); however, the relevance of the loop size and of the G32 and G34 functional groups for the formation of the Tat/TAR/CycT1 complex (Richter *et al.*, 2002) suggests that the hexanucleotide loop assumes a well-defined structure when interacting with the P-TEFb complex. Our structural investigation of the TAR-RNA in complex with **1** indicates that the TAR-RNA loop can indeed interact with cofactors through the looped-out nucleotide in position 35. Upon binding the cofactor, the loop becomes more ordered, and the functional groups N7 and O6 of G34 are situated at reasonably well-defined positions in the major groove of the upper stem. The structural stabilization of the hexanucleotide loop upon binding of a cofactor to A35 could represent a general mechanism by which the Tat protein pre-orders the loop for the binding of CycT1 to the G32 and G34 residues. Thus, the simultaneous targeting of both the bulge and the loop of the TAR-RNA by positively charged molecules should result in a more efficient inhibition of both the Tat-protein and the P-TEFb complex binding to the TAR-RNA. The aminoglycoside analogue **1** binds to the TAR-RNA with a K_d in the μmolar range;

however, the easy addition of functional groups on the sugar skeleton makes this structure an attractive starting point for the design of TAR binders with improved thermodynamic properties, contacting both the bulge and loop functional "hot spots".

1.2 The questions we ask.

Here we perform a structural investigation of the aminoglycoside analogue **1** in complex with the HIV-2 TAR-RNA (Scheme 1.1) by NMR in solution. This work aims to answer the following questions:

1. Does **1** bind the TAR-RNA in the bulge region, thereby directly competing with the protein Tat for the same binding site?
2. Does **1** contact residues of the hexanucleotide loop of the TAR-RNA?
3. In case **1** binds also to the loop of the TAR-RNA is this binding mode significant to understand the principle of the binding of CycT1 to the same loop?
4. Which are the principles of RNA recognition by **1**? Are there any specific hydrogen bonds between the charged amino groups of **1** and the phosphate backbone of the RNA?
5. Are the 2'-hydroxyl groups of the RNA riboses involved in cofactor recognition?

Question 5 can only be answered if precise structural information on the orientation of the proton of the 2'-OH group is available. However, the orientation of the 2'-OH proton has so far escaped structural investigation both by X-ray and NMR in solution. Thus, we set out to develop an NMR-based approach, which could reveal the conformation around the H2'-C2'-O2'-OH torsion angle. The new methodology and the results obtained for the HIV-2 TAR-RNA in absence of cofactors are presented and discussed in Chapter 5, where we address the role of the 2'-hydroxyl group in the stabilization of RNA secondary structure and in supporting an ordered network of H-bonds in the minor group of A-form helical RNA. Unfortunately, due to the low concentration of the sample of the TAR-RNA/**1** complex (< 0.3 mM), the new methodology could not be reliably applied to this system. However, the approach described in Chapter 5 will be very valuable to understand the function of the 2'-hydroxyl group in cofactor binding,

when sample concentration of the order of 1 mM can be reached. Indeed, such studies are being conducted in our laboratory for the TAR-RNA bound to Mg2+ ions.

1.3 Significance of the work

RNA molecules are able to switch among different conformations, depending on different factors like the pH, temperature and salt concentrations. These local or global conformational changes can be triggered by cofactors binding and are required for the biological function. Small positively charged molecules, such as polyamines, aminocyclitols or aminoglycosidic scaffolds can capture specific three-dimensional folding motifs of nucleic acids, thereby either blocking the RNA in a biologically inactive conformation or directly competing with the binding of proteins and cofactors necessary for the biological function. Understanding the principle of the interaction of RNA with cationic antibiotics represents a major step towards the design of specific RNA binders. The TAR-RNA and its interaction with the Tat protein and the CycT1 unit of the P-EFTb complex are key elements in the control of HIV viral replication and constitute attractive targets for anti-viral therapy. The structural work conducted here, and in particular the novel structural insights obtained for the hexanucleotide loop of the TAR-RNA, contribute to a better understanding of the multiple possibilities of RNA recognition by positively charged molecules and indicate a new direction for the design of improved TAR binders, which inhibit the interaction of the TAR-RNA with both the Tat protein and the P-EFTb complex.

Chapter 2

The experiments

2.1 Methods.

NMR spectroscopy. NMR experiments were acquired using a 0.3 mM sample of $^{13}C/^{15}N$ labeled HIV-2 TAR-RNA dissolved in either D_2O or H_2O/D_2O (90/10), containing 50 mM phosphate buffer at pH 6.4. Compound **1** was titrated in concentrations varying from 0.3 mM to 30 mM. The chemical shift changes of the resonances of both the base C5/C6/C8 and H5/H6/H8 and the sugar C1' and H1' atoms of the RNA were monitored during the titration in HSQC spectra (Bax et al., 1990). Standard NMR methods were employed to assign both the free and the aminoglycoside-bound TAR-RNA (Hu *et al*., 1998; Wijmenga & van Buuren, 1998). The experiments were acquired at 600, 700 and 800 MHz. The non-exchangeable base and sugar proton resonance were assigned with the aid of 3D HCCH-COSY-TOCSY (Hu *et al*., 1998) and 3D ^{13}C-HSQC-NOESY experiments acquired in D_2O. Exchangeable imino proton resonances were assigned from 2D NOESY spectra in H_2O. Low concentration of the sample hampered the stereospecific assignment of H5'/H5'' and hence contacts to them were treated as ambiguous restraints. Presence of imino-imino proton peaks and NOE interactions that suggested stacking led us to impose hydrogen bond base pairing restraints on G16-C46, G17-C45, C18-G44, C19-G43, A20-U42 and G21-C41 in the lower stem and G26-C39, A27-U38, G28-C37 and C29-G36 in the upper stem. Absence of the imino proton peak of U40 even at lower temperatures indicates disruption of the A22-U40 base pair in the complex. However, the presence of several NOEs between U40 and C39 (protons H6 to H6, H5 to H2', H5 to H3' and H6 to H3' respectively) and U40 and C41 (protons H2' to H5 and H3' to H5, respectively) indicates that U40 is not bulged-out but remains between C41 and C39. Thus, weak hydrogen bonding between U40 and A22 was imposed in the initial phase of the calculations to ensure that U40 does not move out of the helix. The hydrogen bonding restraint between U40 and A22 was released during

the refinement step. Intermolecular NOEs from the TAR-RNA to the ligand were assigned from a 3D ^{13}C-edited / ^{12}C-filtered HSQC-NOESY spectrum in D_2O (Leupin *et al*., 1990, Zwahlen *et al*., 1997), where the NOEs between the ^{13}C-labeled RNA and the unlabeled aminoglycoside could be selectively observed. The ligand being in fast exchange between the free and bound forms, no intramolecular NOE restraints could be derived for the aminoglycoside in the complex.

RNA intramolecular NOEs were derived from a 3D ^{13}C-HSQC-NOESY and from a base-selective high-resolution 3D ^{13}C-HSQC-NOESY (Brutscher *et al.*, 1991) with mixing time of 100 ms, while a mixing time of 200 ms was used to measure the intermolecular NOEs in the 3D ^{13}C-edited / ^{12}C-filtered HSQC-NOESY experiment.

The sugar pucker was defined as either C2' endo or C3' endo depending on the J coupling constants between H1' and H2' derived from a 2D-HCCH-E.COSY spectrum (Schwalbe *et al*., 1994). The sugar pucker was restrained to C2' endo for $J > 8$ Hz and to C3' endo for $J < 3$ Hz coupling. Because of the low sample concentration, the quality of the experiments involving the backbone phosphate was very poor and no $^3J_{HP}$ coupling constant could be measured to prove the backbone dihedral angles.

^{15}N-labeled aminoglycoside (Jaunzems. *et al.*, 2004) was mixed with unlabeled TAR-RNA to attempt the measurement of H-bond mediated J_{NP} couplings, which would have confirmed the presence of hydrogen bonding between the amino groups of the aminoglycoside and the phosphate backbone of the RNA. The J-quantitative method (Vuister & Bax, 1993) was used to measure the J_{NP} couplings in a H-N two-dimensional correlation. Unfortunately, no hydrogen bond mediated J_{NP} coupling could be detected for this sample. The ^{15}N chemical shift of the amino group confirmed that all four groups of **1** are protonated in the NMR buffer. The resonances of the amino protons of **1** could not be observed due to fast exchange with the solvent.

T_1 relaxation times for the base C8/C6 carbons were measured in ^{13}C-^{1}H correlations using relaxation delays of 5, 10, 20, 30, 40, 60, 80, 100, 120, 160, 200, 300, 400, 600, 800 and 1200 ms.

The simultaneous presence of the negatively charged RNA and 10-fold excess of positively charged ligand deterred us from using phages or bicelles to attempt alignment of the RNA-ligand complex. Instead, we measured residual dipolar couplings (RDC) by making use of the fact that at high magnetic fields, RNA helices are sufficiently aligned to generate measurable field-induced dipolar couplings. The field-induced RDCs (*fi*RDCs) linearly scale with the square of the field strength. We measured the ^{1}H-^{13}C splitting (J+D) of the C2H2, C6H6, C8H8 and C1'H1' at 600, 700 and 800 MHz. The dipolar couplings at 800 MHz were derived by linear fitting of the three values and extrapolation to zero magnetic field to obtain the value of J. Only those RDCs were employed during the refinement procedure, which showed a linear relationship to B_0^2 within the experimentally determined error limits (37 RDCs in total, excluding the two-base bulge, loop and the two terminal residues).

The high values of the residual dipolar couplings and T1 values considerably longer than those of the free-TAR indicated that the TAR/**1** complex dimerizes. No intermolecular NOEs between two RNA molecules in the dimer was observed in a 3D ^{13}C-edited / ^{12}C-filtered HSQC-NOESY spectrum in D_2O and in a 3D ^{13}C-edited/^{12}C/^{14}N-filtered HSQC-NOESY in H_2O for a sample consisting of 150 μM ^{13}C/^{15}N labelled RNA,150 μM unlabelled RNA and 3 mM ligand.

Structure calculations. The NOE peaks from the 3D ^{13}C-HSQC-NOESY, 3D ^{13}C-edited/ ^{12}C-filtered HSQC-NOESY and 2D NOESY spectra in H_2O or D_2O were converted to appropriate distance restraints classified into three categories corresponding to weak, medium and strong, with upper distance bounds of 3, 4 and 5 Å respectively. Very weak NOEs to exchangeable protons or very weak intermolecular NOEs were allowed an upper bound of 6.0 Å.

All chemical shifts of the nucleotides in the lower stem (16-21 and 41-46) but the C1' of G43 did not significantly change upon binding of compound **1**, indicating that the A-form conformation of the lower stem, found in the free form (Aboul-ela *et al.,* 1996) is conserved. Typical sequential NOEs confirmed the presence of A-form helix in this region. Therefore, very loose dihedral angle restraints, within a range of ±60° of the standard A-form values, were imposed for the α, β, ζ and ε backbone angles in the lower stem of the TAR-RNA.

Restrained molecular dynamics (rMD) combined with energy minimization in XPLOR-NIH (Schwieters *et al.*, 2003) was employed to obtain the final ensemble of structures of the TAR/neooligoaminodeoxysaccharide complex. The dna-rna-allatom.par (Parkinson *et al.*, 1996) parameter set was used for the RNA, while a modified force field was employed for the ligand with parameters for bond, angle, improper and non-bonded terms assembled from the updated parameter files for carbohydrates in XPLOR-NIH. The restraints used in the calculations comprised a total of 447 NOEs for the RNA (320 intranucleotide and 127 internucleotide NOEs), 55 intermolecular NOEs, 44 experimental dihedral angle restraints defining the ribose pucker for 22 nucleotides, 41 non-experimental dihedral angle restraints defining the backbone angles α, β, ζ and ε in the lower stem, 72 H-bond restraints defining 10 base pairs for stretches 16-21, 41-46 and 26-29, 36-39 and 10 restraints defining the planarity of the base pairs.

Definition of the overall geometry of the complex. The chemical shift mapping and the intermolecular NOEs identified three regions for the binding of the aminoglycoside analogue **1** on the TAR-RNA: in the bulge and loop regions on the side of the upper stem major groove and in the minor groove of the lower stem (Fig. 3.1). Furthermore, the dipolar coupling data and the relaxation indicate that the RNA dimerizes. The presence of only one set of resonances for the complex and the absence of NOEs between the two RNA molecules in the dimer, suggest that the dimer is symmetric and that the RNA dimerizes through one or more aminoglycoside molecules. The dipolar coupling data are compatible with a dimer where the two upper or lower stems are roughly parallel to each other while the other two stems diverge, forming an angle of about 60° with the other stem in the same RNA molecule. Four possible scenarios can satisfy these constraints (Fig. 2.1): a) three molecules of **1** bind to the loop, the bulge and lower stem of the TAR-RNA and the RNA dimerizes with the two upper stems parallel to each other; b) three molecules of **1** bind to the loop, the bulge and lower stem of the TAR-RNA and the RNA dimerizes with the two lower stems roughly parallel to each other; c) one molecule of **1** binds to the upper stem major groove contacting both the bulge and the loop of the TAR-RNA, while a second molecule binds to the lower stem minor groove and the RNA dimerizes with the two upper stems parallel to each other; d) one molecule of **1** binds to the upper stem major groove

contacting both the bulge and the loop of the TAR-RNA, while the second molecule contacts the lower stem minor groove of the TAR-RNA; the RNA dimerizes with the two lower stems roughly parallel to each other. Structure calculations according to the first, and third scenarios resulted in very high energy, due to high distortion of the aminoglycoside analogue molecules.

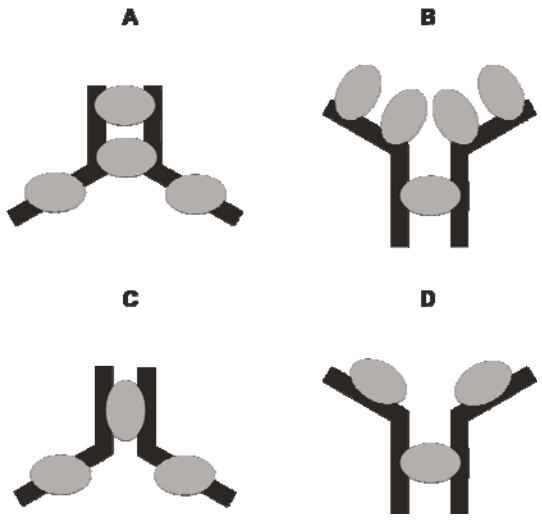


Fig. 2.1 Schematic representation of the association of TAR with the aminoglycoside analogue **1** and of dimer formation. (A) three molecules of **1** bind to the loop, the bulge and the lower stem minor groove and the RNA dimerizes with the two upper stems parallel to each other; the lower stems point in opposite directions. The magnetic susceptibility tensor predicted for such an arrangement of the helical elements is in agreement with the experimentally determined dipolar coupling tensor. (B) Three molecules of **1** bind to the loop, the bulge and the lower stem minor groove and the RNA dimerizes with the two lower stems parallel to each other; the upper stems point in opposite directions. Also in this case the predicted magnetic susceptibility tensor is in agreement with the experimentally determined dipolar coupling tensor. (C) one molecule of **1** binds to the upper stem major groove contacting both the bulge and the loop of the TAR-RNA, while a second molecule binds to the lower stem minor groove; the RNA dimerizes with the two upper stems parallel to each other. (D) one molecule of **1** binds to the upper stem major groove contacting both the bulge and the loop of the TAR-RNA, while the second molecule contacts the minor groove of the lower stem of the TAR-RNA; the RNA dimerizes with the two lower stems parallel to each other.

Therefore, only the second and fourth scenarios were found to be consistent with the observed intermolecular NOEs. However, the second scenario, which includes

five aminoglycosides molecules binding to two RNA units can be excluded due to the following reasons: a) A35 is the only residue of the loop showing NOEs to **1**, indicating that the binding of **1** to this residue must be part of a more complex network of interactions; b) chemical shifts data indicate that **1** binds to the bulge and the loop of RNA with the same K_d, suggesting cooperative binding to both sites; c) smaller cyclic analogues or linear analogues of **1** do not show any detectable binding to the TAR-RNA (Kirschning *et al*., 2004), indicating the relevance of a cooperative binding to multiple sites; d) entropic contribution strongly disfavor the second case with respect to the fourth. Therefore, the case of Fig. 2.1B was not considered further.

Definition of the upper stem binding site. The first binding site, comprising the bulge, the upper stem major groove and the loop of the RNA was accurately defined using the following procedure. Due to the four-fold degeneracy of the resonances of the aminoglycoside analogue **1**, intermolecular NOEs between the RNA and **1** cannot be assigned to a defined sugar or linker unit. Full ambiguity among the four possibilities results in non-convergent structures at the upper stem binding site. Therefore an accurate procedure was used to define the intermolecular NOEs, covering all reasonable possibilities that are not equivalent by symmetry. The sugars and linkers of compound **1** were numbered from *1* to *8* clockwise to define a ring consisting of sugar *1*, linker *2*, sugar *3*, linker *4*, sugar *5*, linker *6*, sugar *7* and linker *8*. We constructed seven different intermolecular NOE templates as described in Fig. 2.2. Although the resonances of aminoglycoside analogue **1** show a four-fold symmetry, the symmetry of the chemical structure is only two-fold. Therefore, the initial seven sets of intermolecular NOEs, shown in Fig. 2.2, were duplicated by shifting both the sugar and the linker numbering by one unit (sugar 1→ sugar 3, linker 2 → linker 4, etc.). This resulted in seven sets of intermolecular NOEs plus two numbering schemes, totaling fourteen combinations. The rmsd of the upper stem binding site, formed by the aminoglycoside analogue **1** and the RNA nucleotides A22, U23, U25 and A35, and the rmsd of the whole complex, excluding the loop residues, were evaluated for the energetically most favorable 20 structures in all fourteen calculations before the final refinement with the *fi*RDC data.

The upper stem binding site of the complex converged the best for the seventh set of intermolecular NOEs with both numbering schemes. The numbering of the sugars and linkers showed in Scheme 1.1 delivered the best result with an rmsd value of 2.15 Å for the upper stem binding core and 1.75 Å for the whole complex excluding the loop. The average rmsd value for the remaining thirteen calculations was 3.0 Å for the upper stem binding core and 2.25 Å for the whole complex. Thus, the seventh set of intermolecular NOEs and the numbering scheme of Scheme 1.1 were used to determine the structure of the complex. Some of the ambiguities in the seventh intermolecular NOEs data set were further released by an iterative assignment process during structure calculation.

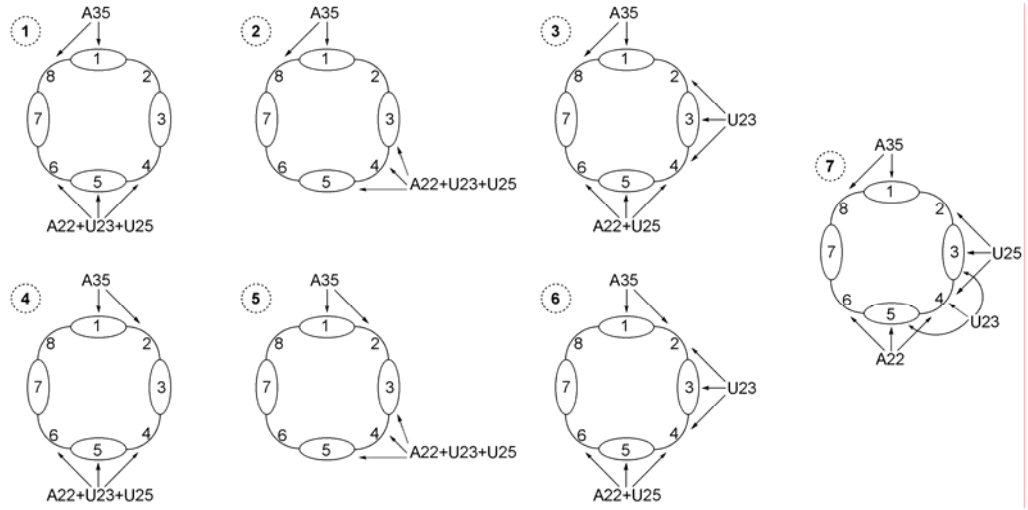


Fig. 2.2 Pictorial representation of seven intermolecular NOEs data sets used in the structure calculations to define the upper stem binding site. Two assignments were considered for NOEs stemming from A35: i) either to sugar *1* and linker *2* (lower row) or ii) to sugar *1* and linker *8* (upper row). For NOEs stemming from the bulge residues A22, U23 and U25, three alternatives were devised: a) A22, U23 and U25 contact sugar *5* and linkers *4* and *6* (first column); b) A22, U23 and U25 contact sugars *3* and *5* and linker *4* (second column); c) A22 and U23 contact sugar *5* and linkers *4* and *6*, while U25 contacts sugar *3* and linkers *2* and *4* (third column). Combination of the two possibilites for NOEs from A35 with the three for the bulge yields six sets of NOEs. In the seventh set (right), NOEs stemming from A35 were assigned to sugar *1* and linker *8*; those stemming from U25, to sugar *3* and linkers *2* and *4*; those stemming from U23, to sugars *3* and *5* and linker *4*; those stemming from A22, to sugar *5* and linker *4* or *6*. When a residue contacts two ligand moieties equivalent by symmetry (e.g. linkers *4* and *6*), those restraints were treated as ambiguous in the structure calculation. Additionally, other seven NOE sets were created by

rotating the aminoglycoside macrocycle 90º counter-clockwise to account for its two-fold symmetry.

As the contacts between the aminoglycoside analogue **1** and the upper stem, bulge and loop of the TAR-RNA are much more extensive than those with the second binding site in the lower stem minor groove, a first round of structure calculations focused on the determination of the intermolecular recognition in the upper stem binding site only. 150 structures were subjected to 22 ps of restrained Molecular Dynamics (rMD) at 1000K during which the force constants of the NOE (50.0 kcal/mol $Å^2$), dihedral angle restraints (300.0 kcal/mol deg^2), angles, impropers and Van der Waals were increased to their full value. After this MD step, we chose 125 structures out of 150, which had total energy lower than 700 kcal/mol. These structures were subjected to three rounds of MD of 3 ps each at 1000K, 600K and 400K. 700 steps of Powell minimization were performed between each MD step (Jucker & Pardi, 1995). All force constants were kept at their maximum values. After the high temperature step, hydrogen bond restraints between U40 and A22 were released. At this stage 50 structures were chosen with total energy < 600 kcal/mol and NOE violation energy < 100 kcal/mol. These structures were subjected to a refinement protocol consisting of 20000 steps of 1 fs dynamics at 500K, 300K, 150K, 50K and 10K. The protocol ended with 2000 steps of Powell minimization. Only repulsive energy terms were employed for the Van der Waals forces at each stage and no electrostatics was used during any part of the structure generation, annealing and refinement. At each subsequent stage of dynamics (simulated annealing and refinement), the number of violations and the energy decreased suggesting that the experimental restraints are well satisfied in the structures.

The 50 structures were subjected to a final refinement with the incorporation of *fi*RDCs to validate the shape and conformation of the complex. In this final refinement protocol the attractive Lennard-Jones potential term was switched on. The force constant for the dipolar coupling energy term was 1.5 kcal $mol^{-1}Hz^{-2}$. An accurate estimate of the magnitude and rhombicity of the alignment tensor cannot be obtained from the powder pattern if only C-H dipolar couplings for base and C1'-H1' ribose vectors are available; hence we performed a grid search

wherein the dipolar coupling energy term was evaluated as a function of the alignment tensor. A systematic grid search for the optimal values of D_{zz} and R was performed sampling values of D_{zz} between 5 and 10 in 1 Hz steps and rhombicity between 0.1 and 0.65 in 0.05 steps. Examination of the dipolar coupling energy profiles revealed a minimum for D_{zz} = 7 Hz and R = 0.55; thus, this tensor was employed during refinement (Varani *et al.*, 2005). The correlation coefficients of the experimentally determined *fi*RDCs *versus* the calculated *fi*RDCs and the Q factors of the final structures were calculated using DipoCoup (Meiler *et al*., 2000). The final structures were viewed and analysed using MolMol (Koradi *et al*., 1996).

Definition of the second binding site and dimer structure. The contacts of the aminoglycoside analogue **1** with the lower stem minor groove are defined by the intermolecular NOEs of **1** with G43-H1' and G44-H4'. Due to steric reasons, these NOEs were assigned to the same sugar and linker units of **1**.

A second aminoglycoside molecule was docked to the minor groove of two TAR/**1** molecules, predisposed by rigid body dynamics in the correct orientation to satisfy the geometry defined by the dipolar coupling data. The docking was performed by MD simulation in explicit water solvent and was guided by the NOEs of **1** to G43-H1' and G44-H4'. The heating stage consisted of 200 steps of 2fs dynamics at 100, 200, 300,500 and 600K. This was followed by refinement at 600 K for 6ps. Then, the molecule was allowed to cool to 25 K in temperature steps of 25 K with 200 steps of 3fs dynamics at each temperature. The calculations ended with 200 steps of Powell minimization.

RNase digestion experiments. 5'-^{32}P labeled HIV-2 TAR-RNA (Scheme 1.1) was digested with both RNase T1 (G-selective) and RNase A (selective for U and C) in presence of the aminoglycoside analogue **1** at 4 ºC. The concentration of the RNA was 10 μM, while the concentration of compound **1** varied from 0.1 μM to 10 mM for the digestion with RNase T1 and with RNase A. All experiments were performed in autoclaved Eppendorf reaction vessels. Extreme precaution has been taken to avoid RNAse contamination. After the reaction, the mixture was loaded onto a 20% denaturing polyacrylamide gel.

PACE experiment. The stechiometry of the binding of the TAR-RNA to **1** was verified using the PACE technique (Cilley & Williamson, 1997). A series of lanes

containing different concentration of the aminoglycoside (0, 1, 2, 5, 7, 10, 20, 70, 100, 200 and 500 μM) were poured with the gel rotated by 90°, as described by Cilley and Williamson. The total concentration of 5'-^{32}P labeled RNA was 10 μM. *D* represents the retardation distance of the RNA at each aminoglycoside concentration. The function $(D-D_F)/(D_B-D_F)$, where D_B and D_F are the retardation distances of the RNA completely bound to the aminoglycoside (concentration of **1** = 500 μM) and in the absence of the aminoglycoside, respectively, is plotted against the aminolycoside concentration in the gel (Fig. 3.3). The data where fitted assuming the following scenarios: a) one molecule of the TAR-RNA binds one molecule of **1**; b) one molecule of the TAR-RNA binds two molecules of **1** with the same K_d**;** c) two molecules of the RNA bind one molecule of **1**; d) one molecule of the TAR-RNA binds two molecules of **1** with different K_ds; e) one molecule of the TAR-RNA binds one molecule of **1**; subsequently the TAR/**1** complex dimerizes; f) two molecules of the RNA bind one molecule of **1**; subsequently two molecules of **1** are recruited by the dimer for a total of two RNA molecules and three molecules of **1** (case of Fig. 2.1D). As shown in Fig. 3.3, the scenario f) optimally fits the experimental data, confirming the dimer geometry inferred from the NMR data.

2.2 Materials.

The TAR-RNA used for the NMR experiments was obtained by *in-vitro* transcription using T7-polymerase expressed and purified *in house*. ^{15}N/^{13}C labeled NTPs were purchased by Spectra Stable Isotopes. The RNA was purified by gel electrophoresis. 20 ml transcription reaction yielded ca. 5 mg of pure RNA. The aminoglycoside analogue **1** was obtained by chemical synthesis as described before (Kirschning *et al*., 2004).

Chapter 3

Results

3.1 NMR experiments and dipolar coupling analysis.

C1'/H1', C5/H5, C6/H6 and C8/H8 RNA resonances were monitored upon titration of the ligand in concentration ratios varying from 1:1 to 1:100. Significant chemical shift changes were observed up to a 10-fold excess of the ligand (3 mM) over the RNA (0.3 mM). The chemical shift mapping identifies A22, U23 and loop nucleotides G32, G34 and A35 as the sites being most affected by the binding of the aminoglycoside analogue **1** (Fig. 3.1). In agreement with this observation, binding of ligand **1** protects both loop sites G32-G34 and C30 from digestion with RNase T1 and with RNAse A, respectively (Fig. 3.2). However, digestion at G21 site is enhanced at high ligand concentration, suggesting that the A-form helical structure is disrupted between G21 and A22 (Fig. 3.2B).

A **B**

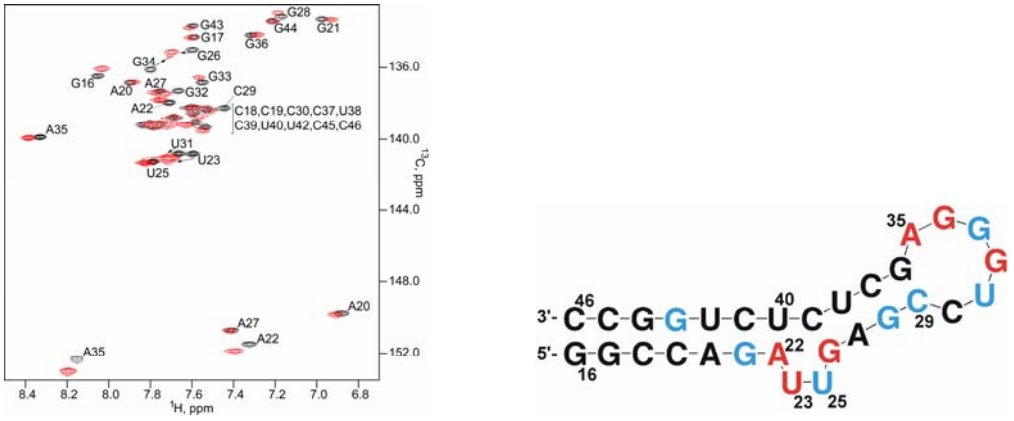


Fig. 3.1 (A) Superposition of the base region of a C-H correlation for the HIV-2 TAR-RNA free (black) and in complex with the aminoglycoside analogue **1** (red). The acquisition dimensions were 128 points (C) and 2048 points (H); each spectrum was acquired for a total of 5 hours on a 600 MHz Bruker AVANCE spectrometer. The most significant chemical shifts changes upon binding of the TAR-RNA to **1** are observed either for the base or for the ribose C and H

resonances of A22, U23, G26, G32, G34 and A35 (rendered in red in panel (B)). More modest chemical shift changes are observed for G21, U25, G28, C29, U31 G33, G36 and G43 (rendered in light blue in panel (B)).

The intermolecular NOESY spectrum reveals intimate contacts of ligand **1** with A22, U23, U25 and A35. The atypical C8/H8 chemical shifts of A35 in the free TAR-RNA implies that the adenosine base is partially projected-out of the loop, thus occupying an optimal position for interacting with a ligand bound to the bulge on the side of the upper stem major groove. Large chemical shifts changes are observed upon titration of ligand **1** for the C2/H2 and C1'/H1' resonances of A35. In particular, the linewidth of the A35-C2 decreases from 80 to 48 Hz at 900 MHz, which suggests that in the free TAR this nucleotide assumes different conformations interchanging on a μs to ms time-scale, while in the complex the base occupies a well-defined position. Notable is the chemical shift change experienced by the base of G34, which varies the most upon formation of the complex without showing any direct contact to the ligand. This is indicative of a profound change in the electronic environment of G34 and possibly of the overall conformation of the loop, triggered by the interaction of A35 with the aminoglycoside analogue. The chemical shift change of the G34 C8/H8 is accompanied by a considerable decrease in the T_2, indicating that G34, which is disordered in the free TAR and in the TAR/argininamide complex (Brodsky &Williamson, 1997; Aboulela *et al.*, 1996), undergoes a slow (μs-ms time scale) conformational transition between two states in complex with **1**. Considerable chemical shifts changes are observed also for the riboses of U31, G32 and G33. Clearly, the binding of the aminoglycoside analogue to A35 triggers a profound conformational change in the hexanucleotide loop, which moves from a completely disordered state to a partially ordered one.

In the bulge, the largest chemical shift changes are observed for A22 and U23. U23 stacks on A22 in the free TAR-RNA (Aboulela *et al.*, 1996), while in this complex it is involved in several contacts with the aminoglycoside analogue **1**. The A22-U40 base pair, which is stably formed at low temperature in the HIV-2 TAR-RNA, is disrupted in the complex, as confirmed by the failure to observe the U40 imino resonance even at 4 °C. However, several NOEs indicate that both

A22 and U40 remain positioned between G21 and G26 and C39 and C41 respectively.

A35 is the only loop residue that directly contacts the ligand, suggesting that the interaction of **1** with the loop must be part of a more complex network of interactions holding the complex together. A cooperative binding of **1** to both the loop and the bulge of the TAR-RNA is supported by the chemical shift data, which indicate a K_d of comparable size for both the bulge and the loop sites. Moreover, while **1** inhibits the binding of Tat-derived peptides to the TAR-RNA, linear analogues of **1**, or smaller cyclic analogues, do not show any biological activity (Kirschning *et al.*, 2004), underlining the importance of a cyclic constrained structure of appropriate size for a cooperative binding to the bulge and loop regions of the RNA. Therefore, it is safe to conclude that one molecule of ligand **1** binds simultaneously to the loop and the bulge of the RNA.

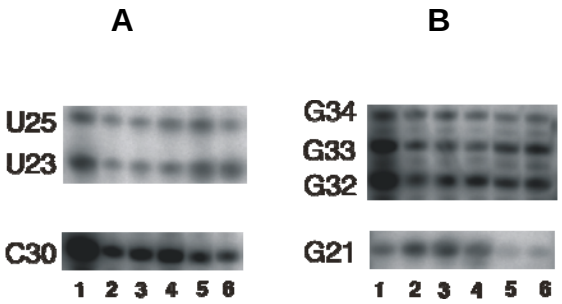


Fig. 3.2 (A) RNase protection analysis of the TAR/**1** complex. The HIV-2 TAR-RNA (10 μM, 5' ^{32}P labeled) was digested with RNase A (selective for C and U) at 4 ºC. The digestion experiments were analyzed by gel electrophoresis. Lane 1, control without RNase; lane 2, compound **1** (10 mM); lane 3, compound **1** (5 mM); lane 4, compound **1** (500 μM); lane 5, compound **1** (1.2 μM); lane 6, compound **1** (0.12 μM). (B) The TAR-RNA (10 μM, 5' ^{32}P labeled) was digested with RNase T1 (G selective) at 4 ºC. The digestion experiments were analyzed by gel electrophoresis. lane 2, compound **1** (10 mM); lane 3, compound **1** (5 mM); lane 4, compound **1** (500 μM); lane 5, compound **1** (1.2 μM); lane 6, compound **1** (0.12 μM).

Besides the bulge and the loop regions, a second binding site for the aminoglycoside analogue is identified on the TAR-RNA. A medium chemical shift change for the G43-C1' upon titration of **1**, together with NOEs between the aminoglycoside analogue **1** and both the G43-H1' and the G44-H4' reveal the additional binding of **1** to the minor groove of the RNA lower stem. This finding

is not surprising, as structural studies of the TAR/neomycin B complex (Faber *et al.*, 2000) previously identified the lower stem minor groove as a binding site for positively charged aminoglycosides.

Compound **1** can be produced with ^{15}N and ^{13}C labeling at reasonable cost. In the course of this study we produced a ^{15}N-labeled batch of compound **1** (Jaunzems *et al.* 2004) with the aim of verifying the existence of H-bonds between the amino groups of the aminoglycoside analogue and the RNA phosphates by detection of ^{15}N-^{31}P scalar couplings mediated by H-bonds. Unfortunately, no such coupling could be observed for the degenerate amino group.

Quite surprisingly, the longitudinal relaxation rates R_1 of the base carbons of the TAR-RNA decreased in complex with **1** by about 40%. This decrease cannot be explained by invoking dynamics, as it affects also the carbons of the stem bases, whose structure remains unchanged. Instead, the relaxation rates might indicate a change in the overall correlation time of the TAR-RNA.

A few dipolar couplings measured at 800 MHz for the bases C-Hs assume values between -5 and -6 Hz. These figures exceed by far those expected for the TAR-RNA aligned in the magnetic field due to the magnetic susceptibility anisotropy of the bases (Zhang *et al.*, 2003). The largest dipolar coupling tensor is expected for coaxial stacking of the upper and lower stems and would correspond in a 800 MHz field to a D_{zz} of 8 Hz and a rhombicity R between 0 and 0.1. For this tensor the largest base C-H dipolar coupling should not exceed –4 Hz. The high values found for the TAR/**1** complex suggest that the RNA dimerizes in the presence of the ligand (Al-Hashimi *et al.*, 2001). The dipolar couplings of the two stems can be best fitted assuming a D_{zz} of 7 Hz and a rhombicity of 0.55. The interstem angle resulting from the fitting of the dipolar couplings assuming D_{zz} = 7 Hz and R = 0.55 is around 60º. The experimentally determined dipolar coupling tensor fits well that predicted for a dimeric TAR-RNA, where the upper and lower stem of each unit form an angle of 60º and the two units are associated with two stems approximately parallel to each other and the other two stems pointing in opposite directions.

The 3D ^{13}C-edited / ^{12}C-filtered HSQC-NOESY spectrum in D_2O and the 3D ^{13}C-edited/^{12}C/^{14}N-filtered HSQC-NOESY in H_2O acquired for a sample containing 150 μM ^{13}C/^{15}N labelled RNA,150 μM unlabelled RNA and 3 mM ligand showed

no intermolecular NOEs between the two TAR-RNA molecules. This finding indicates that an aminoglycoside molecule mediates the dimerization of two TAR molecules and that no direct contact between the two RNA molecules is present in the dimer. The presence of only one set of resonances for the RNA further suggests that the complex is symmetric.

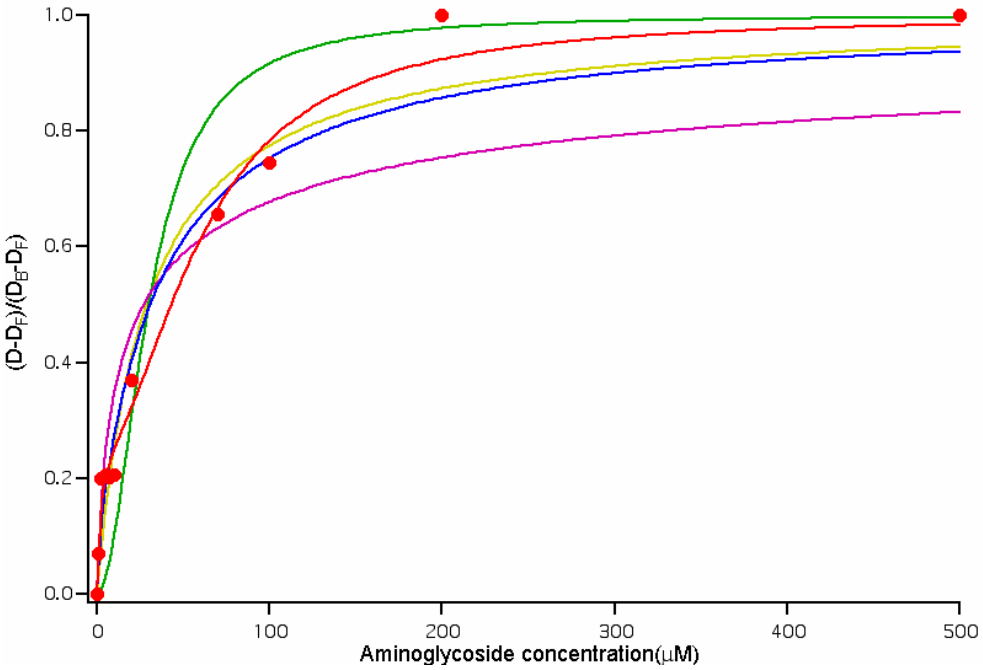


Fig. 3.3 The function $(D\text{-}D_F)/(D_B\text{-}D_F)$ is plotted versus the concentration of the aminoglycoside analogue **1** in each lane of the PACE gel. D is the retardation distance in each lane, D_F is the retardation distance of the TAR without **1** and D_B is the retardation distance of the TAR-RNA completely bound to **1** (concentration of **1** = 500 μM). Six different cases were considered: a) one molecule of the TAR-RNA binds one molecule of **1** (yellow line); b) one molecule of the TAR-RNA binds two molecules of **1** with the same K_d (green line); c) two molecules of the RNA bind one molecule of **1** (magenta); d) one molecule of the TAR-RNA binds two molecules of **1** with different K_ds (blue line) e) one molecule of the TAR-RNA binds one molecule of **1** ($TAR + \mathbf{1} \xleftrightarrow{K_{d1}} C_1$); subsequently the TAR/**1** complex dimerizes ($2\ C_1 \xleftrightarrow{K_{d2}} C_2$) (close to the yellow line, not shown); f) two molecules of the RNA bind one molecule of **1** ($2\ TAR + \mathbf{1} \xleftrightarrow{K_{d1}} C_1$); subsequently two molecules of **1** are recruited by the dimer for a total of two RNA molecules and three molecules of **1** ($C_1 + 2\ \mathbf{1} \xleftrightarrow{K_{d2}} C_2$) (red line; $K_{d1} \approx 20$ μM^2; $K_{d2} \approx 3$ mM^2). The best fit is obtained for case f), which confirms the dimer stechiometry inferred from the NMR data.

The dipolar couplings define the relative orientation of the two RNA units. However, it remains to be defined if the two RNA molecules associate through the aminoglycoside molecule bound to the upper stem major groove site or to the lower stem minor groove site. Structure calculations were conducted for both cases: association through the molecule of **1** bound to the upper stem site resulted in high energy due to severe distortion of the aminoglycoside and large NOE violations. Therefore, dimerization through the molecule of **1** bound to the lower stem minor groove is exclusively compatible with the NOE data.

PACE gel electrophoresis (Cilley & Williamson, 1997) was used to verify the stechiometry of the complex. As shown in Fig. 3.3, the PACE data could be poorly fitted assuming monomeric RNA (a-c) or formation of a dimer with one or two molecules of **1** (d-e). Instead, a good fitting of the retardation distances was obtained assuming a dimerization step including two RNA molecules and one molecule of **1**, followed by subsequent recruitment of two additional molecules of the aminoglycoside (red curve, case f)), which confirmed the dimer geometry inferred from the NMR data.

Structure calculations, including NOEs, dihedral angle restraints and dipolar couplings, were first conducted on a monomeric unit containing one TAR and one ligand molecule binding to the upper stem. This procedure allowed us to define both the binding site in the upper stem major groove, where the most profound contacts between the RNA and ligand **1** are situated, and the relative orientation of the two helices. Subsequently, a molecule of the aminoglycoside analogue **1** is docked in the minor groove of the lower stem of two TAR-RNA molecules, which were predisposed in the correct orientation to reproduce the experimentally derived dipolar coupling tensor.

.3.2 Structure calculations

The upper stem binding site. The final structure ensemble was obtained as explained in the Material and Methods section. Structure calculation employing NOEs and dihedral angle restraints exclusively, resulted in a well-defined RNA structure in both stems but in a poor definition of the relative orientation of the two stems (rmsd for the lower stem: 0.5 Å; rmsd for the upper stem = 0.8 Å).

Binding of the aminoglycoside analogue **1** to the upper stem major groove does not lead to the formation of the U23-A27-U38 base triple, as observed in the TAR/argininamide complex (Brodsky & Williamson, 1997). The formation of the U23-A27-U38 base triple rigidifies the backbone at the bulge junction, triggering the coaxial stacking of the upper and lower stems. In the TAR/**1** complex, instead, the relative orientation of the two stems around the flexible bulge is not defined by NOE data alone.

Structure refinement with C-H field-induced dipolar couplings (*fi*-RDCs) for the bases and the riboses of the upper and lower stem positioned the two stems at a relative angle of ca. 60° and indicated association of two TAR/**1** units with two stems of the two RNA molecules in a parallel orientation and the other two stems diverging from each other.

The 14 lowest energy structures are shown in Fig. 3.4. These structures had total energy < -300 kcal/mol, no NOE violations > 0.5 Å and no angular violations > 5Å. The RNA is very well defined with a total heavy atom rmsd (excluding the loop residues) of 0.92 Å. The base planes of the two stems form an angle of 60º. The A-form helix structure of the lower (nt. 16-21, 41-46) and upper (nt. 26-29, 36-39) stems is conserved. The upper stem helix is, however, slightly distorted, with the α and γ angles varying from the A-form typical gauche – and gauche + conformations to the trans conformation for residues A27-C29 and C39.

The aminoglycoside analogue **1** binds to the bulge and the loop residues on the major groove side of the upper stem (Fig. 3.4). Due to the four-fold symmetry of its resonances and the fast kinetics of complex dissociation, the structure of **1** in the complex cannot be determined. However, the salient features of the interaction between the aminoglycoside analogue **1** and the TAR-RNA are well defined and the rmsd of the whole complex excluding the loop residues 30-34 is 1.2 Å.

The backbone of the 22-25 stretch “opens up” to allow for **1** to reach the major groove of the RNA upper stem (Fig. 3.4). The last base pair of the lower stem is disrupted, U40 stacks below C39; A22 stacks below G26 but tilts away from U40, to widen the major groove of the upper stem and make place for the ligand. Thus, in the TAR/**1** complex A22 and U40 belong to the upper stem, while in the free TAR they stack on the top of the lower stem G21-C41 base pair. In agreement with the disruption of the A-form helix structure between G21 and A22, the

digestion of the TAR-RNA at the G21 site by RNase T1 increases upon titration of **1**. The planarity of the G26-C39 base pair is slightly distorted, as indicated by the weakening of the G26 imino resonance upon complex formation. The π-orbitals of the U23 and U25 bases entertain hydrophobic interaction with linker *5* and sugar *2* and linker *3* of ligand **1**, respectively (Fig. 3.5).

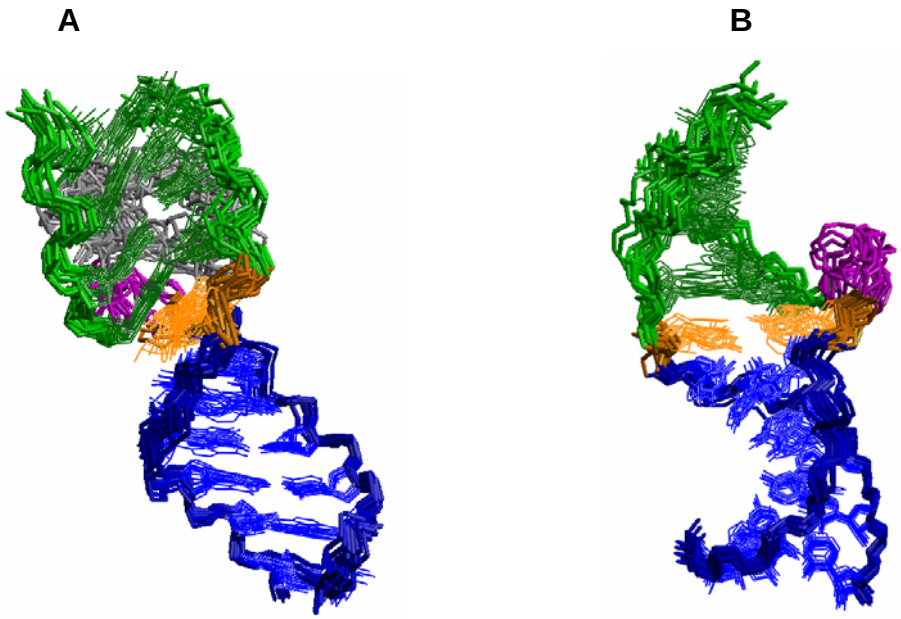


Fig. 3.4 (A) Superposition of the heavy atoms of 14 structures of the TAR/**1** complex based upon residues 16-21, 26-29 and 36-46. Residues 30-35 and 23, 25 are not shown. The lower stem is colored in blue and the upper stem in green. A22 and U40 are highlighted in orange; the backbone of the bulge is in magenta. The backbone and the base stacking indicate that the upper and lower stems form an angle of about 60º. The aminoglycoside binding to the major groove of the HIV-2 TAR-RNA upper stem is rendered in gray. (B) Same as in (A) after 180º rotation around the vertical axis. The ligand is not shown

A35 stacks on the top of the sugar *1* moiety of the ligand and is involved in a π-cation interaction with the positively charged amino group. This contact pulls A35 out of the loop, causing G34 to move on the top of G36 in all structures. In 60% of the structures G34 stacks on the top of G36, while in the other 40% the two bases interact face to edge (Fig. 3.6). The presence of an equilibrium between two conformations is compatible with the large linewidth of the G34-C8, suggesting that the exchange between the face-to-face and face-to-edge orientations of G34 and G36 occurs on a μs-ms timescale.

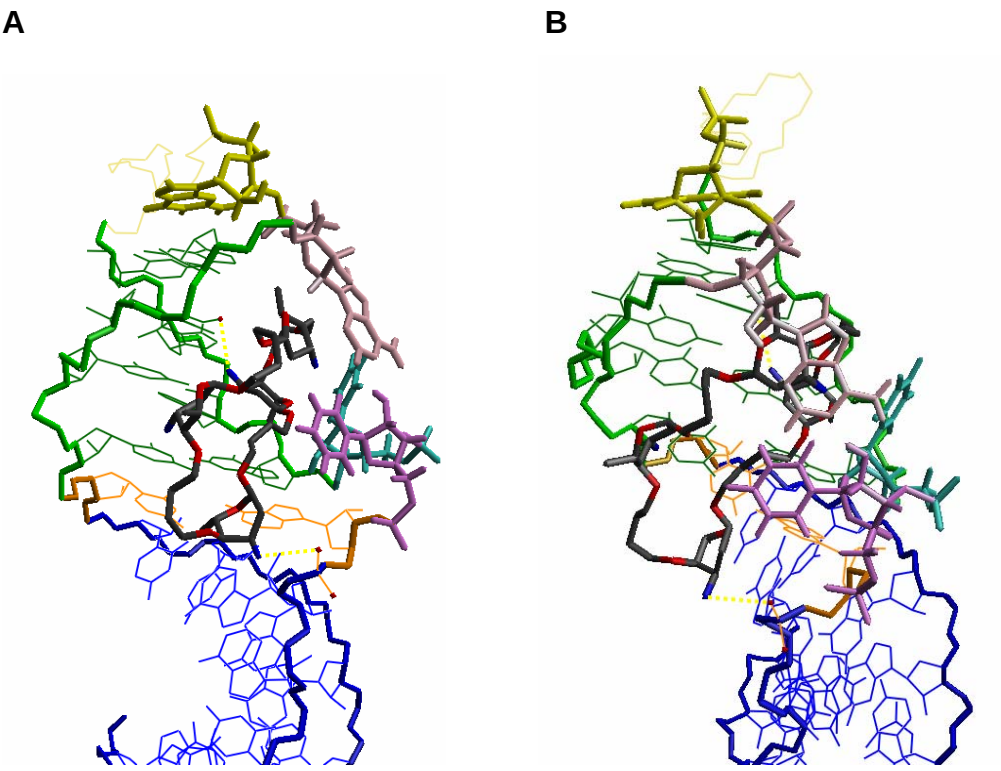


Fig. 3.5 (A) View of the best energy structure of the family of Fig 3.4. The aminoglycoside analogue **1** in the atom coloring scheme is placed in the major groove of the upper stem of the HIV-2 TAR-RNA. A22 and U40, depicted in orange, clearly show that the hydrogen bonds between 22 and 40 are disrupted and that A22 tilts away from the major groove to accommodate the ligand. Besides A22 and G26, U23 in magenta, U25 in green and A35 in pink have direct contacts with the aminoglycoside analogue. The positively charged amino group of the ligand sugar *1* interacts with the π-orbital of the A35 base, the amino group of sugar *2* interacts electrostatically with the major groove face of G28 (yellow dashed lines), while the amino group of sugar *3* contacts the phosphate of A22 (yellow dashed line). G34, in yellow, stacks on the top of G36. The backbone of the loop is rendered in yellow. A similar intermolecular interaction pattern is observed also for the other structures of the family. (B) same as in A) after 90º rotation around the vertical axis.

The ribose of G34 is in the C3'-endo pucker in all structures, in contrast to the free TAR where the G34 was found to assume the C2'-endo conformation (Brosky & Williamson, 1997) . Instead, the ribose sugar of G33 and A35 assume a well defined C2'-endo pucker.

The positively charged amino groups of the four sugars of ligand **1** interact with various functional groups of the RNA (Fig. 3.5). The amino group of sugar *1* (Scheme 1.1) is involved, as mentioned, in a π-cation interaction with A35 and in

long-range electrostatic interactions with the O2 of U25. The amino group of sugar *3* is directed towards the major groove carbonyls of the A27-U38 and G28-C39 base pairs in all structures but two, while the amino group of sugar *5* is close to the phosphate backbone of A22 and/or to the O4 of U23. Finally, no consistent contact is found for the amino group of sugar *7*. These results are in agreement with the failure to detect H-bonding of the amino groups of ligand **1** to the phosphate backbone of the RNA by the J_{NP} scalar coupling, as most of the electrostatic contacts of the amino groups of the ligand involve base functional groups rather than phosphates.

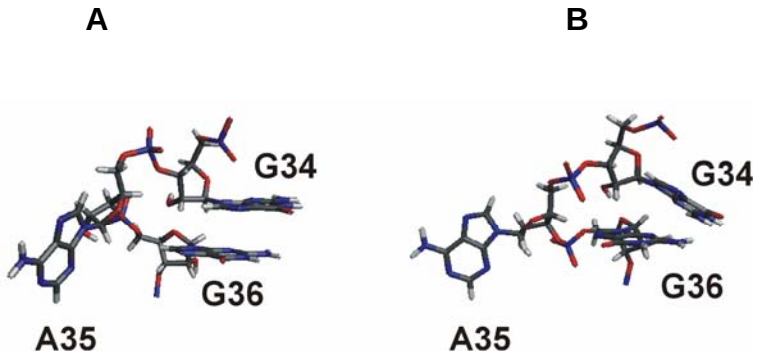


Fig. 3.6 A35 interacts with the ligand and therefore moves out of the loop. As a consequence of this, G34 moves on the top of G36 of the upper stem. The two bases interact either face-to-face (A) or face-to-edge (B). The two conformations are likely to be in intermediate exchange (μs-ms time scale), as inferred from the linewidth of the G34-C8

Definition of the second binding site and dimer geometry. Similarly to neomycin (Ferber *et al*., 2003), the aminoglycoside analogue **1** binds also to the minor groove of the lower stem of the TAR-RNA. The chemical shift change of the G43-C1' is maximum at a 1:1 ratio of RNA:aminoglycoside, as opposed to the chemical shift changes in the bulge and loop region, which increase up to a 1:10 ratio. Thus, the binding constant of **1** to the minor groove of the TAR-RNA is higher than that to the bulge and loop regions. This finding is confirmed by the analysis of the PACE data (see caption to Fig. 3.3).

Having identified dimer formation through the aminoglycoside bound to the lower stem minor groove as the only one compatible with the NOE data, a molecule of **1**

was docked to the minor groove of the lower stem of two TAR-RNA molecules with the help of the NOEs between **1** and the G43-H1' and G44-H4' protons. The two RNA molecules were pre-oriented such that the magnetic susceptibility tensor of the dimer well agrees with the experimentally determined one (Fig. 3.7).

The aminoglycoside analogue **1** binds symmetrically to the minor groove of the lower stem of the two TAR-RNA units. One amino group enjoys electrostatic interactions with the negative spot created by the phosphate groups of G21, A22 and U23, while another amino group closely contacts the 2'-hydroxyl of C45. Two sugars and the intervening linker of **1** penetrate the minor groove of each RNA molecule at the G43-C45 site (Fig. 3.7).

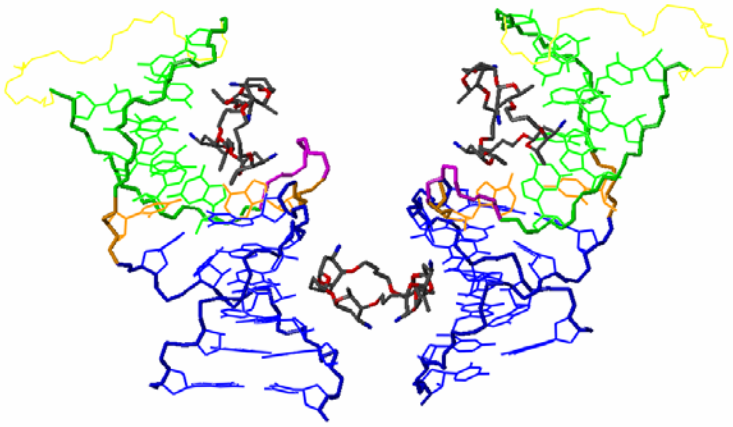


Fig. 3.7 The TAR-RNA with one molecule of **1** bound to the upper stem major groove dimerizes through an aminoglycoside molecule bound to the lower stem minor groove. The two RNA molecules are oriented with the lower stems roughly parallel to each other (interhelical angle $\theta \approx 30°$) and the upper stem diverging from each other. From the magnetic susceptibility tensor of the dimeric RNA a dipolar coupling tensor can be calculated, which fits well the experimentally determined one (D_{zz} = 6.8 Hz; R = 0.54). The color-coding is the same as in Fig. 3.5, with the exception of the three aminoglycoside molecules, rendered in the atom color-coding.

Chapter 4

Discussion

4.1 The structure of the TAR-RNA bound to the aminoglycoside.

The upper stem binding site. The aminoglycoside analogue **1** binds to the TAR-RNA in the major groove of the upper stem and contacts extensively the bulge nucleotides 22-25 and A35 of the hexanucleotide loop. The bulge of the TAR-RNA has been identified as the major binding site for the Tat-protein and several studies have been targeted to find competitive inhibitors of the Tat/TAR interaction among ligands binding to the RNA bulge. The conformation induced by ligand **1** at the bulge site of the TAR-RNA, as well as the relative orientation of the two stems, differ from those found in the Tat/TAR complex. Binding of **1** does not induce either the formation of the U23-A27-U38 base triple or the subsequent coaxial stacking of the upper and lower stem. Instead, the bulge nucleotides are all exposed to the solvent in the TAR/**1** complex and the axes of the two stems form an angle of about 60º. The relative position of the two domains is probably a consequence of the disruption of the A22-U40 base pair, of the tilting of A22 to accommodate the large ligand and of the electrostatic interactions of the ligand amino group with the phosphate of A22. The sugars and the linkers of the ligand have extensive hydrophobic contacts with the U23 and U25 rings. Such hydrophobic contacts resemble a well-known interaction motif in complexes of proteins with oligosaccharides, where the sugars often stack below aromatic side-chains (Weis & Drickamer, 1996).

The aminoglycoside analogue **1** represents a promising scaffold for the development of a new TAR-binders generation for two reasons: a) its cyclic, flexible structure presents functional groups that bind to multiple RNA sites; b) the ligand interacts with the A35 of the hexanucleotide loop, thereby rigidifying the loop structure in a way which might resemble that needed for the binding of CycT1 to the TAR-RNA.

It is widely recognized that the intermolecular surface in RNA/protein complex is rather large and therefore small molecules cannot efficiently compete with proteins for RNA binding. Despite many efforts in the search for a competitive inhibitor of the Tat/TAR interaction, no drug lead has been identified to date. The major difficulty in the development of TAR-binders in potential therapeutic agents is their low specificity. The problem might be alleviated by simultaneously targeting multiple binding sites on the TAR-RNA (Davis *et al*., 2004; Hermann & Westhof, 1999). Three "hot spots" for the binding of cations on the TAR-RNA have been recognized in the lower stem major groove, bulge and loop sites by molecular dynamics simulations (Hermann & Westhof, 1999). In another study, an extended effort in identifying the minimal number of interactions necessary to stabilize the Tat-bound conformation of the TAR-RNA resulted in the detection of two binding sites for positively charged groups at the bulge site and in the upper stem major groove close to A27 and G28 (Davis *et al*., 2004). Synthetic ligands exploiting both binding sites showed the highest activity and induced an RNA conformation similar to that binding to the Tat-protein (Davis *et al*., 2004). The amino groups of ligand **1** contact both the bulge site and the major groove functional groups at the G28 site (Fig. 3.5). While the interaction with the bulge is suboptimal and could be improved by introducing a guanidinium group at the C1 or C5 of sugar *5*, the amino group of sugar *3* nicely interacts with the negatively charged spot at the G28 and U38 site (Fig. 3.5), in accord to what is predicted and observed by theoretical calculations and NMR for positively charged ligands (Davis *et al*., 2004). Additionally, the amino group of sugar *1* interacts with the loop residue 35. Thus, the large but flexible scaffold of **1** allows reaching multiple binding sites on the TAR-RNA, and therefore represents a valuable starting structure for the design of binders, which exploit all "hot spots" for cations identified in the TAR upper stem.

The value of cyclic structures as a conformationally constrained backbone for TAR binders has been proven by the design of peptidomimetics based on a cyclic, β-hairpin like scaffold (Leeper *et al*., 2005), binding the TAR-RNA with a much higher affinity than the linear analogues (Athanassiou *et al.*, 2004). The structure of the BIV (Bovine Immunodeficiency Virus) TAR-RNA bound to one of these peptidomimetics revealed extensive interactions with the upper stem major

groove, including a conserved hydrophobic contact of a Ile residue with the bulged-out U of the base triple and electrostatic contacts of arginine side-chains with the tetraloop capping the upper stem and with the second and third base pairs of the upper stem (Leeper *et al*., 2005). Intermolecular contacts of similar nature are found for ligand **1** and the bulge, major groove base pairs and loop residues of the HIV-2 TAR-RNA. Analogously to the BIV TAR-RNA/peptide complex, ligand **1** uses hydrophobic contacts to bind the bulge pyrimidines and electrostatic contacts to bind to the major groove face of the upper stem base pairs and to the loop residues. As observed for the peptides, linear neooligoaminodeoxysaccharides show a much lower affinity for the TAR-RNA than their cyclic analogue **1** (Kirschning *et al*., 2004).

The formation of the Tat/TAR complex is required for the recruitment of the P-TEFb complex and for the phosphorylation and subsequent activation of the RNA polymerase stalled at the TAR-RNA site. The P-TEFb complex consists of the kinase component CDK9 and the cyclin T1 subunit. The complex binds to the TAR-RNA through the CycT1 subunit in a Tat dependent manner. The interaction site of CycT1 with the TAR-RNA is localized in the apical loop and is facilitated by the binding of Tat to the bulge. The mechanism by which the Tat protein favors the interaction of CycT1 with the apical loop is not known, although it is reasonable to hypothesize that Tat structurally organizes the RNA for the binding of CycT1.

The hexanucleotide apical loop has been found to be quite flexible in all NMR studies (Colvin *et al*., 1993; Jäger & Tinoco, 1993; Aboulela *et al.*, 1995) with the riboses of C30, G33 and G34 averaging between the C2' and a C3' conformations. Base stacking of A35 on G36 and of G34 on G36 was found to be consistent with structural restraints derived by NMR for a 14- and a 19-nucleotides hairpin containing the TAR hexanucleotide loop (Colvin *et al*., 1993; Jäger & Tinoco, 1993). Indications of stacking on the 5'-site of the loop (C30 on C29 and U31 on C30) were less prominent than for the 3'-site. On the basis of such structural information, it was proposed that the apical loop is stabilized by base stacking on both sides and possibly by the formation of base pairs between C30 and $A35^{+}$ or C30 and G34 (Jäger & Tinoco, 1993). In a more recent work, biochemical analysis of native and mutant TAR-RNAs was used together with

molecular dynamics simulation to show that the apical loop of the TAR-RNA is stabilized by a cross-loop base pair between C30 and G34, with A35 projected out of the loop (Kulinski *et al*., 2003). This observation is not necessarily in contrast with the NMR studies hypothesizing the stacking of A35 on G36. The large line-broadening of the A35-C2 resonance of the free TAR-RNA and the NOEs between the bases of G34, A35 and G36 are consistent with A35 being in a slow equilibrium between the stacked and looped-out conformations.

Upon binding of the aminoglycoside analogue **1**, the A35-C2 resonance shifts and the linewidth is reduced by a factor of 2. The equilibrium between the stacked and looped-out conformations is not present in the complex; instead, binding of the ligand stabilizes the looped-out conformation, where A35 can extensively interact with **1**. As a consequence of that, the NOE contacts between the base of A35 and both G34 and G36 disappear while several NOEs appear between G34 and G36 indicating that when A35 is looped-out, G34 moves on the top of G36. In this position G34 is in a favorable orientation for hydrogen bonding with C30; however, no evidence of the C30-G34 base pair could be found in the TAR/**1** complex. The ribose pucker of G34, which has been reported to be either C2'-endo (Colvin *et al*., 1993) or a mixture of C2'- and C3'- endo (Jäger & Tinoco, 1993) by NMR studies of free TAR, assumes the C3'-endo conformation in the complex. In agreement with our data, the C3'-endo pucker of the G34 ribose has been found to accompany the stacking of G34 upon G36 in molecular dynamics simulations (Kulinski *et al*., 2003). By inspecting the linewidth of the G34-C8 resonance in the TAR/**1** complex we could detect the involvement of G34 in a conformational exchange process, taking place in the μs–ms timescale. This suggests that that G34, which undergoes rapid exchange among different conformations in the free TAR, is much more constrained in the TAR/**1** complex. In agreement with the presence of a slow conformational exchange process, the NOE data are consistent with both a face-to-face (stacking) and a face-to edge interaction of G34 and G36. All in all, our data indicate that binding of a positively charged ligand of appropriate size to the bulge and the loop of the TAR-RNA on the major groove side of the upper stem leads to the stabilization of the looped-out conformation of A35 and consequently to the relocation of G34 on the top of G36. In the stacked conformation, the N7 and O6 functional groups of

G34 assume well-defined positions in the major groove and the formation of the C30-G34 base pair is possible.

Comparison to the binding of CycT1 to the TAR loop. The cyclin T1 unit of the P-TEFb complex binds to the hexanucleotide loop of the TAR-RNA. The two functional groups O6 and N7 of G34 and G32 are essential for the interaction (Richter *et al.*, 2002). Covariance studies showed that a base pair between the nucleotides at position 30 and 34 stabilizes the Tat/TAR/CycT1 complex but is not required for its formation (Richter *et al.*, 2002). Additionally, the presence of a nucleotide in position 35 is necessary for the TAR/CycT1 interaction, but its nature is not critical (Richter *et al.*, 2002). A small detrimental effect on the CycT1 binding is found only for the A35G mutant.

On the basis of the structure of the complex of TAR with the aminoglycoside analogue **1**, we propose an explanation for the dependence of the TAR/CycT1 interaction on the Tat/TAR complex formation, which fits well the biochemical data summarized above. We suggest that the binding of Tat to TAR pre-organizes the hexanucleotide loop in a favorable conformation for CycT1 binding. It is conceivable that the arginine-rich stretch of the Tat protein, which is responsible for TAR binding at the bulge site, also contacts the A35 residue of the hexanucleotide loop and blocks it in its looped-out conformation, in a similar way as our aminoglycoside. The major driving force for this contact could be the interaction of an aromatic π orbital with a cation, similarly to that observed for the TAR/**1** complex. A generic π-cation interaction provides a rationale for the insensitivity of the TAR/CycT1 interaction to the nature of the nucleotide in position 35. A small negative effect on complex formation is observed exclusively for the A35G mutant, where a possible cross-loop C30-G35 base pair would destabilize the looped-out conformation of the nucleotide in position 35. In agreement with our hypothesis, affinity cleaving experiments show that F38 of Tat is located in proximity of the nucleotides 34 and 35 of the TAR-RNA (Huq & Rana, 1997). Furthermore, residue K50 of the Tat-protein has been found to cross-link with the TAR nucleotide 34, and this interaction is considerably enhanced by the binding of CycT1 to the Tat/TAR complex (Richter *et al*., 2002a).

In the TAR/**1** complex, the looped-out conformation of A35 is accompanied by the relocation of G34 upon G36. Analogously, in the Tat/TAR complex stacking

of G34 upon G36 would place the O6 and N7 functional groups of G34, which are essential for the binding of CycT1, in a well-defined position in the major groove of the TAR-RNA. The formation of a base pair between C30 and G34 would further stabilize the position of the G34 functional groups. In fact, such base pair has been shown to favor the interaction of the TAR- RNA with CycT1 without being directly involved in complex formation (Richter *et al.*, 2002).

Based on the irrelevance of the nucleotide nature at position 35, it has been hypothesized that this nucleotide functions as a spacer with no direct contact to the proteins. Contrarily, we suggest that A35 is a key regulatory element in the interaction of the TAR-RNA with the CycT1 and Tat-protein. Upon binding of a cofactor to the major groove of the upper stem, A35 is pulled out of the loop, G34 moves upon G36 and the dynamics of the loop is restricted. Thus, A35 could be the basis of a regulation mechanism for the formation of the Tat/TAR/CycT1 complex based on the conformational stability of the hexanucleotide loop.

The lower stem binding site and dimerization. The aminoglycoside analogue **1** binds to the minor groove of the lower stem of the TAR-RNA. Similar to neomycin B (Ferber *et al*., 2003), the aminoglycoside analogue contacts the 43-44 site. However, only two sugar rings of **1** bind to the lower stem minor groove of one RNA molecule, leaving the other two sugar rings free for binding to an additional RNA molecule and leading to dimerization. This process is likely to be entropy-driven and favored by the constrained cyclic structure of **1**.

Significance for drug design. The design of drug leads targeting the TAR-RNA has almost exclusively focused on the inhibition of the Tat/TAR interaction at the bulge site. However, the TAR/**1** complex structure suggests that a cyclic scaffold can be used to design ligands, which extensively bind to the entire upper stem major groove of the TAR-RNA up to the capping loop. An NOE between a compound binding to TAR in a Tat-similar manner and the A35-H8, which was identified in a previous study (Davis *et al.*, 2004), confirms that the interaction with A35 is a general feature of large ligands binding to the major groove of the TAR-RNA. The cyclic scaffold of ligand **1** can be utilized in the design of TAR binders, which capture A35 in the looped-out conformation, induce the relocation of G34 on the top of G36 and are decorated with appropriate functional groups to compete with CycT1 for the G34 functional groups. Large ligands interfering with

both the Tat/TAR and the Tat/TAR/CycT1 complex formation and contacting the TAR-RNA at several “hot spots” could represent an improved class of drug leads both in terms of efficiency and specificity.

Chapter 5

Novel methodology to determine the conformation of the 2'-hydroxyl group in RNA

5.1 The conformation of the 2'-OH group in RNA.

The 2'-OH hydroxyl group is a major determinant for both the structure and the function of RNA, where the sugars are all riboses. The presence of the ribose 2'-OH hydroxyl group in RNA engenders a preference for the C3'-endo puckering, thereby providing the decisive factor for the differences in conformation, hydration and thermodynamic stability between canonical RNA and DNA helices (Gyi *et al*., 1998). A high resolution crystal structure of an RNA duplex (Egli *et al*., 1996) revealed the existence of a complex hydration network in the minor groove of the A-form helix, which is supported by the specific distribution of the 2'-hydroxyl groups. The presence of water molecules in the RNA minor groove, anchored by electrostatic interactions to both the ribose and the base, provides an enthalpic contribution to A-form helix stability and, additionally, offers a basis for the entropic regulation of ligand binding to the minor groove. Non-canonical RNA structural elements, such as tetraloops and pseudoknots, can be stabilized by H-bonds involving the 2'-hydroxyl group as well (Williams *et al*., 2001; Nixon *et al*., 2002; Theimer *et al*., 2005). Moreover, the 2'-hydroxyl group is an established player in RNA catalysis. Numerous examples of the role of the 2'-hydroxyl group in catalytic RNAs are documented for self-splicing viral ribozymes, type II introns and the pre-mRNA processing machinery (Gesteland & Chech, 1998; Eckstein & Lilley, 1996; Doherty & Doudna, 2000)

Despite the importance of the 2'-hydroxyl group in RNA structure stabilization and function, only limited structural data is available for the 2'-OH proton of RNA (Giedroc *et al.,* 2003; Conte *et al*., 1996; Lynch *et al*., 1996; Gyi *et al*., 1998). Protons are not directly observable in crystal structures; in aqueous solution the rapid exchange of the hydroxyl proton with the solvent typically

prevents its observation for RNA at room temperature by NMR. Consequently, in the absence of structural information on the 2'-OH proton, the torsion angle θ (H2'-C2'-O2'-OH) remains undetermined. Molecular dynamics (MD) simulations of an RNA hairpin identified three energetically preferred orientations that the 2′-hydroxyl proton can assume if a ribose sugar adopts the C3′-endo pucker, as is the case in canonical RNA A-form helix (Auffinger, & Westhof, 1997) (Fig. 5.1): (A) toward the O3′, stabilized by attractive electrostatic interaction with the phosphate backbone (O3' domain), (B) toward the O4′, stabilized by favorable intra-ribose electrostatic interactions (O4' domain), and (C) toward the base, stabilized by electrostatic interaction with the N3 or O2 atom of the attached base (base domain).

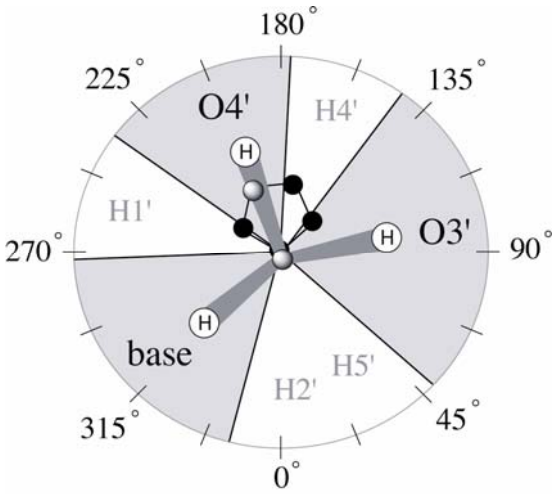


Fig. 5.1 Energetically favored orientations of the 2'-hydroxyl group when the ribose is in the C3'-endo conformation. The ribose sugar ring is schematically shown in the middle with oxygen atoms in grey and carbon atoms in black. According to MD simulations (Auffinger, & Westhof, 1997), the torsion angle θ (H2'-C2'-O2'-OH) can populate three regions, highlighted in grey: the O3' domain, torsion angle θ = 50°-140°; the O4' domain, torsion angle θ = 175°-230°; the base domain, torsion angle θ = 270°-345°.

Experimental determination of the 2'-hydroxyl proton location in RNA structures is essential to verify the role of the 2'-hydroxyl group in the stabilization of RNA folding motifs. Here, we carried out the nearly complete assignment of 2'-OH proton resonances of the HIV-2 TAR-RNA and the conformational analysis of the torsion angle H2'-C2'-O2'-OH based on NOE data and on the measurement of the three scalar couplings between the 2'-OH proton and the H2', C2' or C3' ribose

nuclei ($^3J_{H2',2'OH}$, $^3J_{C1',2'OH}$ and $^3J_{C3',2'OH}$). To the best of our knowledge, this work represents the first assignment and conformational analysis of the 2'-hydroxyl group in a medium sized RNA; all previous conformational studies on the H2'-C2'-O2'-OH torsion angle via scalar couplings were conducted on single nucleotides in DMSO (Acharya & Chattopadhyaya, 2002). The NMR techniques presented here provide a uniformly applicable way to study the orientation of the 2'-OH group in RNA in solution, thereby allowing association of the 2'-hydroxyl group conformation with specific RNA functions such as RNA catalysis.

5.1.1 The question we ask.

In this work we wish to answer the following questions:

1. Is it possible to assign the 2'-OH protons in a medium sized RNA?

2. Can we observe NOEs between the 2'-OH protons and ribose or base protons at low temperature?

3. What are the preferred conformation around the C2'-O2' bond in helical and non-helical structural regions?

4. Do the conformational preferences of the H2'-C2'-O2'-OH torsion angle support the involvment of the 2'-OH proton in a network of H-bonds spanning the minor groove and involving water molecules, which was proposed to stabilize RNA A-form conformation (Egli *et al*., 1996)?

In order to answer these questions we performed a detailed conformational analysis around the H2'-C2'-O2'-OH torsion angle for the 30mer HIV-2 TAR-RNA. To this purpose we developed new NMR experiments aimed to measure the three ^3J couplings ($^3J_{H2',2'OH}$, $^3J_{C1',2'OH}$ and $^3J_{C3',2'OH}$), which define the H2'-C2'-O2'-OH torsion angle.

5.2 The experiments

5.2.1 The assignment of the 2'-OH protons

The resonance assignments were obtained using two-dimensional homonuclear TOCSY (Cavanagh & Rance, 1992) and NOESY (Lippens *et al*., 1995) experiments performed at 600 MHz. All experiments were conducted at 5ºC to slow down solvent exchange. The exchange rate constant of the 2'-OH

protons with bulk water is approximately 15 Hz under these conditions, as estimated from the intensity of the exchange cross-peaks in the NOESY spectrum. This value is in good agreement with the value reported by Gyi *et al*. (Gyi *et al*., 1998) for an RNA and an RNA-DNA duplex. The TOCSY experiment (Fig. 5.2B) shows intense correlations connecting the 2'-OH and the scalar coupled H2' protons. In qualitative agreement with vicinal $^{3}J_{H2',2'OH}$ couplings measured for isolated nucleotides in DMSO (Acharya & Chattopadhyaya, 2002), the through-bond 2'-OH–H2' crosspeaks are intense, reflecting $^{3}J_{H2',2'OH}$ couplings greater than 2 Hz. The NOESY spectrum (Fig. 5.2C) reveals several through-space NOE correlations of medium intensity between the H1' and the 2'-OH protons. Additional crosspeaks in the region ω2 = 5.1-5.9 ppm and ω1 = 6.8-7.1 ppm stem from NOE correlations between cytidine amino and aromatic H5 protons, as confirmed using a ^{15}N-labeled RNA sample. NOE correlations between 2'-OH hydroxyl and other ribose protons are difficult to analyze due to extensive resonance overlap.

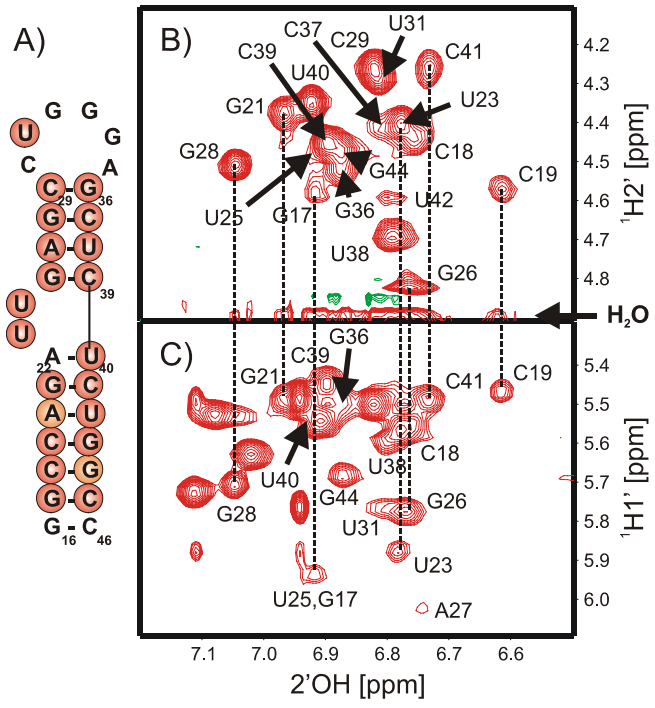


Fig. 5.2. A) Sequence and secondary structural representation of HIV-2 TAR-RNA. Assigned 2'-OH protons are highlighted, tentatively assigned residues A20 and G44 are shown orange. B) Expansion of the 2'-OH–H2' region of the TOCSY experiment. Unambiguous assignments are given with one-letter code followed by the residue number. The clean DIPSI-2rc (Cavanagh &Rance, 1992) mixing sequence was applied for τ_m=43 ms, using a $\gamma B_1/2\pi$=10.04 kHz field strength. C) Expansion of the 2'-OH–H1' region of the water flip-back NOESY experiment

(Lippens *et al.*, 1995). The mixing time was τ_m=50 ms. Water suppression was achieved with excitation sculpting (Hwang & Shaka, 1995). The ^{1}H Carrier position was 4.98 ppm. High power proton pulses were applied with a field strength of 35.5 kHz. 368 complex points were recorded with an acquisition time of 28.3 ms for ^{1}H (ω_1), and 2048 complex points with an acquisition time of 157.5 ms for ^{1}H (ω_2). A recycle delay of 1.2s between transients was used, with 64 scans accumulated per complex increment (total measuring time 17.5 h each). Both spectra were recorded on a four-channel Varian Inova 600 MHz spectrometer equipped with an actively shielded z-gradient triple-resonance probe, at a temperature of 278 K. Spectra were processed and analyzed using FELIX 2000 (MSI, San Diego, USA) and NMRPipe program packages. Spectra were recorded on 2 mM unlabeled HIV-2 TAR-RNA The sample buffer contained 10 mM phosphate buffer, pH 6.4, 50 mM sodium chloride and 0.1 mM EDTA in 500µl 90% H_2O/10% D_2O.

5.2.2 Novel NMR experiments for the measurement of the ^{3}J couplings involving the 2'-OH proton.

The homonuclear and heteronuclear $^3J_{H2',2'OH}$, $^3J_{C1',2'OH}$ and $^3J_{C3',2'OH}$ scalar couplings of the HIV-2 TAR-RNA have been measured in specifically tailored two-dimensional NMR experiments using the quantitative J approach (Vuister & Bax, 1993). Briefly, the $^3J_{H2',2'OH}$ couplings were measured in a H2'-H_{base} NOESY correlation experiment (Fig. 5.3A and 5.4). We chose this correlation for its highly efficient NOESY transfer and decent resolution. It should be mentioned that this experiment benefited from the use of a 0.7 mM HIV-2 TAR-RNA sample consisting of d4-nucleotides lacking all but H1' and H2' ribose protons (Scott *et al.*, 2000). This specific deuteration pattern allows for selective detection of the H2'-H_{base} NOEs with high sensitivity, while circumventing overlap with other H_{ribose}-H_{base} NOEs. If such a selectively deuterated sample is not available, a three-dimensional ^{13}C-edited NOESY correlation is necessary to achieve the desired resolution of the H2'-H_{base} peaks. In the *cross* experiment, we employed a constant time of 40 ms, during which the H2' proton magnetization is transversal and subjected to J-dephasing effects, which results in measurable attenuation of the observed H2'-H_{base} cross peaks. In the *reference* experiment, the evolution of the vicinal $^3J_{H2',2'OH}$ coupling is refocused by use of a REBURP pulse of duration 4 ms centered at 4.16 ppm. This pulse selectively refocuses the H2' protons at a proton resonance frequency of 800 MHz, without affecting the 2'-OH and the H1' resonances (excitation bandwidth: ± 550 Hz; transition region: 250 Hz). In

contrast, the $^{3}J_{H2',2'OH}$ coupling is active during the *cross* experiment by simultaneous application of the H2' REBURP pulse and of a 4 ms IBURP pulse centered in the 2'-OH region at 6.97 ppm in the center of the constant time period. The IBURP pulse selectively inverts the 2'-OH protons leaving the resonances of the H1' protons unaffected at a 800 MHz field (excitation bandwidth: ± 550 Hz; transition region: 250 Hz).

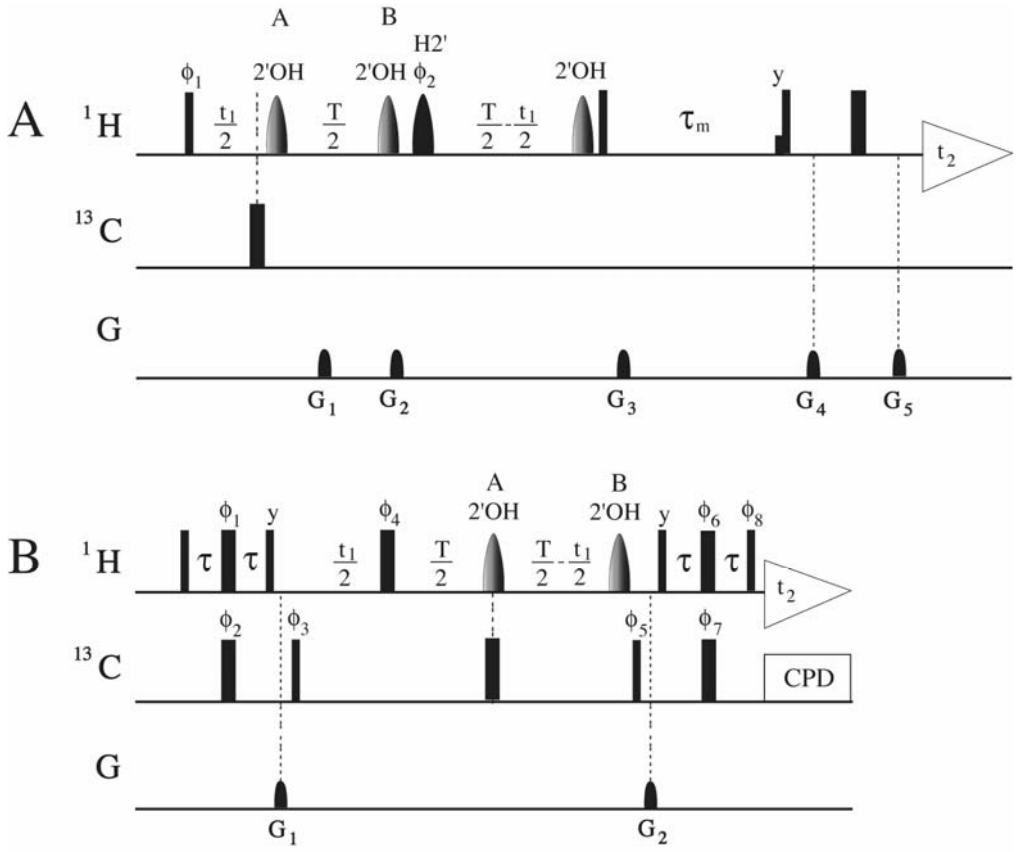


Fig. 5.3 A) Pulse sequence for the quantitative J-CT-NOESY employed for the measurement of the $^{3}J_{H2',2'OH}$ couplings. In the *reference* experiment the selective pulse on the 2'-OH is applied in position A, in the *cross* experiment it is applied in position B. The carrier positions for ^{13}C and ^{1}H were at 85.0 ppm and 5.1 ppm respectively. The selective 2'-OH IBURP1 pulse (Geen & Freeman, 1991) of duration 4 ms was applied with a field strength of 250 Hz and centered at 6.97 ppm. The second IBURP1 pulse at the end of the evolution time compensates for Bloch-Siegert-shifts. The selective H2' REBURP (Geen & Freeman, 1991) pulse of duration 4 ms was applied with a field strength of 250 Hz and centered at 4.15 ppm. Sine shaped pulsed field gradients were applied for 1.0 ms with relative peak amplitudes (in %): G1 = 30%, G2 = 30%, G3 = 40%, G4 = 50%, G5 = 50%, where 100% corresponds to 50 G/cm. Phasecycling: $\phi 1$ = x, -x; $\phi 2$ = 2(x), 2(-x), 2(y), 2(-y); ϕ_{rec} = 2(x, -x), 2(-x, x). Quadrature detection in the t_1 dimension was achieved by altering ϕ_1

according to STATES-TPPI (Marion *et al*., 1989) Water suppression was achieved using the Watergate-sequence for the last 180°-pulse (Piotto *et al*. 1992). The experiments were carried out in 90% H_2O, 10% D_2O on a Bruker 800MHz DRX spectrometer equipped with a 5 mm triple resonance ($^1H,^{13}C$-^{15}N-TXI) probe with a shielded z-axis gradient. B) Pulse sequence for the quantitative J-CT-HSQC employed for the measurement of the $^3J_{C1'/C3',2'OH}$ couplings. In the *cross* experiment the 2'-OH band-selective pulse is applied in position A, in the *reference* experiment in position B. The carrier position for ^{13}C was at 90.0 ppm for the C1'-H1' correlation and at 73 ppm for the C3'-H3' correlation, while the proton carrier was centered at 5.1 ppm. The selective 2'-OH IBURP1 of duration 4 ms was applied with a field strength of 250 Hz and centered at 7.3 ppm. Sine shaped pulsed field gradients were applied for 1.0 ms with relative peak amplitudes (in %): G1 = 50, G2 = -25%, where 100% corresponds to 50 G/cm. Phasecycling: ϕ_1 = 2(x), 2(-x); ϕ_2 = 8(x), 8(-x); ϕ_3 = x, -x; ϕ_4 = 2(y), 2(-y); ϕ_5 = 4(x), 4(-x); ϕ_6 = 4(x), 4(-x); ϕ_7 = 8(x), 8(-x); ϕ_8 = 2(y), 2(-y); ϕ_{rec} = 2(x, -x), 2(-x, x). Quadrature detection in the t_1 dimension was achieved by altering ϕ_3 according to STATES-TPPI. The experiments were carried out in 90% H_2O, 10% D_2O on a Bruker 800MHz DRX spectrometer equipped with a 5 mm triple resonance ($^1H,^{13}C$-^{15}N-TXI) probe with a shielded z-axis gradient.

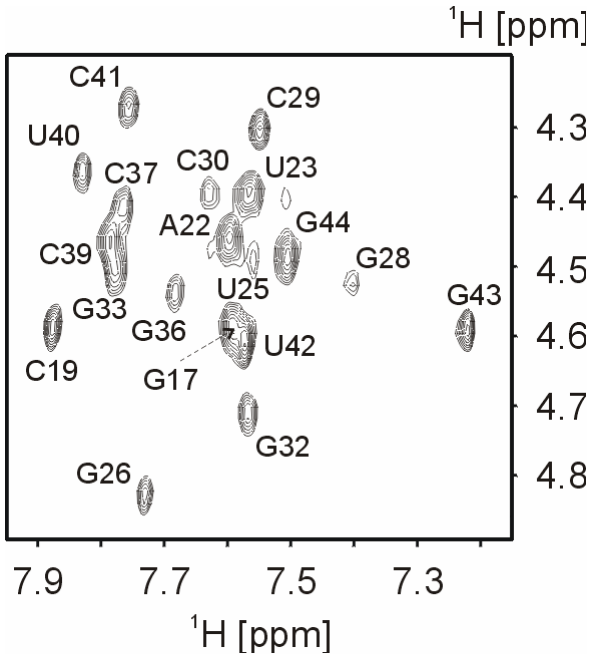


Fig. 5.4. Spectra were recorded on 0.6 mM 3', 4', 5', 5''-d4-HIV-2 TAR-RNA, where all ribose protons but the H1' and H2' were replaced by deuterons. The sample buffer contained 10 mM phosphate buffer, pH 6.4, 50 mM sodium chloride, and 0.1 mM EDTA in 500 µL of 90% H_2O/10% D_2O. The spectra were recorded on a Bruker 800MHz spectrometer equipped with a z-gradient triple-resonance probe, processed with the NMRPipe program and analyzed using FELIX 2000 (MSI, San Diego, USA). *Cross* experiment of the H2'(i)-H6/8(i+1) NOESY correlation for the measurement of the $^3J_{H2',2'OH}$ couplings. In the *cross* experiment, we employed a constant time

delay of 40 ms, during which the H2′ proton magnetization is transversal and subjected to J-dephasing effects, which results in measurable attenuation of the observed H2′-H_{base} cross peaks. In the *reference* experiment, the evolution of the vicinal $^3J_{H2',2'OH}$ coupling is refocused by use of a REBURP pulse of duration 4 ms centered at 4.16 ppm. In contrast, the $^3J_{H2'-2'OH}$ coupling is active during the *cross* experiment by simultaneous application of the H2′ REBURP pulse and of a 4 ms IBURP pulse centered in the 2′-OH region at 6.97 ppm in the center of the constant time period.

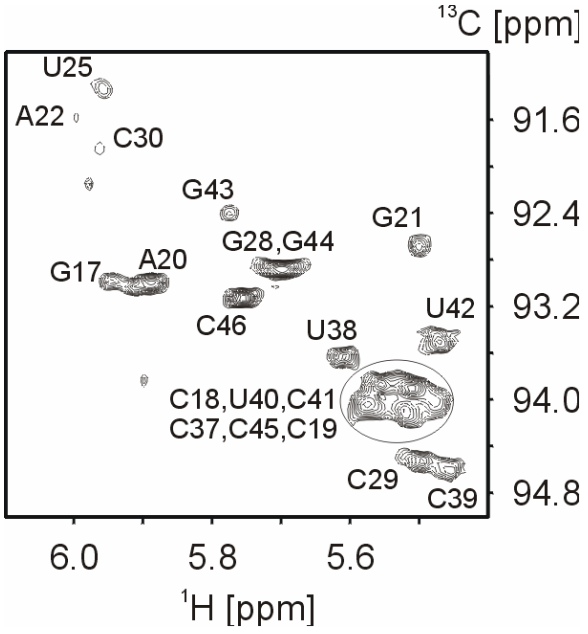


Fig. 5.5 Expansion of the *cross* spectrum of the C1′-H1′ CT-HSQC for the measurement of the heteronuclear $^3J_{C1',2'OH}$. Evolution of the $^3J_{C1',2'OH}$ scalar couplings during the constant time of the *cross* experiment was achieved by simultaneous application of a carbon pulse and a 4ms IBURP pulse centered in the 2′-OH region at 7.3 ppm. A constant time period of 25 ms was used to achieve optimal sensitivity and resolution.

The $^3J_{C1',2'OH}$ and $^3J_{C3',2'OH}$ scalar couplings were measured in constant-time C1'-H1' and C3'-H3' HSQC correlations, respectively (Fig. 5.3B, 5.5 and 5.6). Here, evolution of the $^3J_{C1'-2'OH}$ and $^3J_{C3'-2'OH}$ scalar couplings during the constant time of the *cross* experiment was achieved by simultaneous application of a carbon pulse and a 4ms IBURP pulse centered in the 2'-OH region at 7.3 ppm. A constant time period of 25 ms was used to achieve optimal sensitivity and resolution for both the C1'-H1' and the C3'-H3' correlations. The *J*-coupling constants were extracted from the intensity ratio of peaks in the *cross* and *reference* experiments, according to Eq. 5.1 (Vuister & Bax, 1993):

$$\frac{I_{cross}}{I_{ref}} = \cos(\pi JT) \qquad \text{(Eq. 5.1)}$$

where T is the length of the constant time period. It should be noted that the effective *J*-evolution periods employed in the calculations are compensated for the evolution of the couplings during the IBURP pulses, which resulted in constant times of 34.76 ms, 22.05 ms, 22.05 ms for the three $^{3}J_{H2',2'OH}$, $^{3}J_{C1',2'OH}$ and $^{3}J_{C3',2'OH}$ scalar couplings, respectively.

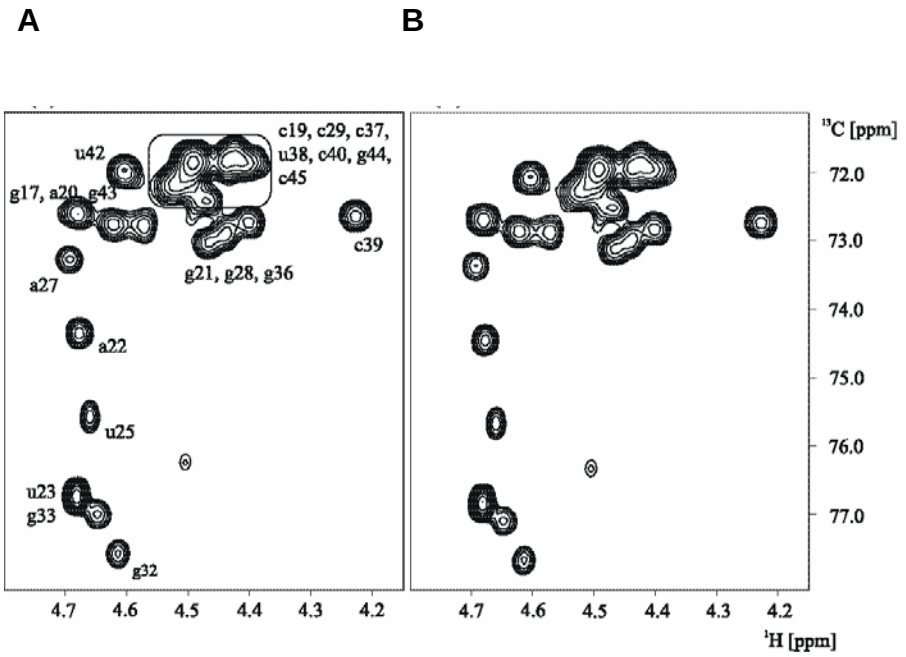


Fig. 5.6 Expansions of the cross and reference spectra of the CT-HSQC for the measurement of the heteronuclear $^{3}J_{C3',2'OH}$ couplings. The constant time T delay was 25 ms (A *cross*, B *reference* spectrum for the C3'-H3' correlation). Spectra shown in A) and B) were recorded on 1.0 mM 1', 2', 3', 4', 5'-$^{13}C_5$-1', 2'-d_2-HIV-2 TAR-RNA where ribose protons H1' and H2' are replaced by deuterons, thereby reducing potential overlap problems for ^{13}C2'-H2' and ^{13}C3'-H3' cross peaks. The sample buffer contained 10 mM phosphate buffer, pH 6.4, 50 mM sodium chloride, and 0.1 mM EDTA in 500 µL of 90% H_2O/10% D_2O. The spectra were recorded on a Bruker 800MHz spectrometer equipped with a z-gradient triple-resonance probe, processed with the NMRPipe program (Delaglio *et al.,* 1995) and analyzed using FELIX 2000 (MSI, San Diego, USA).

5.3 Results

5.3.1 The assignment of the 2'-OH protons.

We could unambiguously assign 20 out of 30 2'-OH hydroxyl protons of the HIV-2 TAR-RNA (Fig. 5.2 and Table 5.1). Five of the 10 missing 2'-OH hydroxyl protons belong to the loop nucleotides 30-35, where 2'-OH protons are probably not participating in a stable network of hydrogen bonds. As a consequence, the exchange rate of those 2'-OH protons with water is fast, impeding their detection even at 5 ºC. We could not observe the 2'-OH protons of the two nucleotides constituting the terminal, fraying base pair for similar reasons. Three additional 2'-hydroxyl protons in the lower stem could not be identified unambiguously due to resonance overlap (A20 and G43 are tentatively assigned, see Table 5.1). Our attempts to assign these resonances from a ^{13}C-edited NOESY acquired using a more dilute, ^{13}C/^{15}N-labeled TAR-RNA sample (0.7 mM) failed because of vanishing NOE intensities. To the best of our knowledge, this represents the first nearly complete assignment of 2'-OH proton resonances of an RNA molecule.

Table 5.1. Assignment of the 2'-OH Resonances of the HIV2 TAR-RNA[a]

Nt. #	C.S. [ppm]	**Nt. #**	C.S. [ppm]	**Nt. #**	C.S. [ppm]
G16	n.a.	**A27**	6.76	**C37**	6,84
G17	6.93	**G28**	7,07	**U38**	6,81
C18	6,80	**C29**	6,84	**C39**	6,91
C19	6,63	**C30**	n.a.	**U40**	6,94
A20	*6.96*[b]	**U31**	6,84	**C41**	6,75
G21	6.99	**G32**	n.a.	**U42**	6,82
A22	n.a.	**G33**	n.a.	**G43**	*6.79*[b]
U23	6,80	**G34**	n.a.	**G44**	6,89
U25	6,93	**A35**	n.a.	**C45**	6,84
G26	6,78	**G36**	6,89	**C46**	n.a.

[a] C.S. stands for chemical shift, Nt. for nucleotide and n.a. for not assigned. The columns containing the nucleotides numbers are in bold. [b] The assignments are only temptative.

With the exception of C19 and G28 (Fig. 5.2), the chemical shift dispersion of the 2'-OH hydroxyl protons is rather limited with resonance frequencies clustering between 6.75 and 7.00 ppm, which is in good agreement with chemical shift

ranges reported previously (Gyi *et al.*, 1998; Lynch et al., 1996; Acharya & Chattopadhyaya, 2002)

NOE correlations between the 2'-OH and aromatic H6 or H8 protons of the following nucleotide could be observed for 13 of the assigned 2'-OH hydroxyl protons (G17, C18, C19, U23, G26, G28, G36, C37, C39, U40, U42, G44, C45) (Fig. 5.7).

Molecular dynamic (MD) simulations of an RNA hairpin show that the conformation of the 2'-OH group depends on the sugar pucker (Auffinger & Westhof, 1997). Three orientations are accessible to the 2'-OH hydroxyl group when the ribose is in the typical RNA C3'-endo conformation (Fig. 5.1): A) towards the O3' stabilized by attractive electrostatic interaction with the phosphate backbone, B) towards the O4' stabilized by favorable intra-ribose electrostatic interactions, and C) towards the base stabilized by electrostatic interaction with the N3 or O2 atom of the attached base. The $^{3}J_{H2',2'OH}$ coupling can adopt values higher than 4 Hz in all three domains, so that the detection of a 2'-OH-H2' crosspeak in TOCSY spectra is not indicative of a specific domain (Fraser *et al.*, 1969).

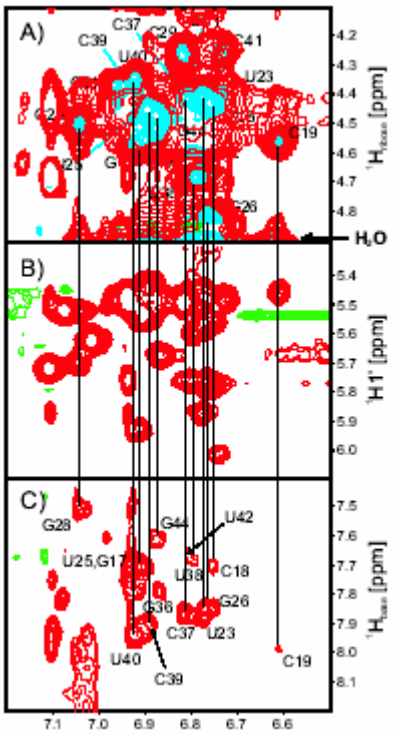


Fig. 5.7 Expansions of the water flip-back NOESY experiment (τm=50 ms), recorded on a four-channel Varian Inova 600 MHz spectrometer equipped with an actively shielded z-gradient triple-resonance probe, at a temperature of 278 K. The sample was the same as in Fig. 5.2. All regions

are plotted using the same contour levels and multipliers. Unambiguous assignments are given with oneletter code followed by the residue number. A) 2'-OH(i)–H2',3',4',5',5"(i) region (red contours) overlaid with the same region of the 2'-OH(i)–H2'(i) TOCSY (DIPSI-2rc mixing sequence, τm=43 ms) experiment (blue contours). B) 2'-OH(i)–H1'(i) region. C) 2'-OH(i)-H6(pyrimidine),H8(purine)(i+1) region.

5.3.2 The ^3J couplings involving the 2'-OH proton.

The $^3J_{H2',2'OH}$ couplings measured for the HIV-2 TAR-RNA at 5°C are shown in Fig. 5.8A. In general, the $^3J_{H2',2'OH}$ couplings cluster between 3 and 4.5 Hz and, except for C18, no value exceeds 5.5 Hz. Chemical exchange of the 2′-OH proton with the solvent can potentially reduce apparent $^3J_{H2',2'OH}$ coupling values. We determined the exchange rate constant k_{ex} of the 2′-OH protons with bulk water for C19 from the intensity of the exchange cross-peak in NOESY spectra at 1, 5 and 10°C. The measured k_{ex} value for C19 of approximately 15 Hz at 5°C is in good agreement with the value reported by Gyi et al. for an RNA and an RNA-DNA duplex (Gyi *et al.*, 1998) Unfortunately, severe overlap of the remaining 2′OH proton resonances impeded the accurate determination of the k_{ex} values for the other nucleotides.

When a substantially larger k_{ex} = 25 Hz is assumed, the apparent $^3J_{H2',2'OH}$ coupling measured over a 40-ms long constant time is approximately 85% of the actual $^3J_{H2',2'OH}$ coupling. For the majority of nucleotides, however, this 15% tolerance would be well within the limits of the experimental error. Thus, for k_{ex} up to 25 Hz, we could consider the effective J-coupling scaling factor from solvent exchange to be negligible.

The stability of RNA structures can locally fluctuate. Consequently, in order to locate secondary structural elements within the TAR-RNA structure (Fig. 5.2A), where $^3J_{H2',2'OH}$ couplings are considerably influenced by solvent exchange, we carefully compared the $^3J_{H2',2'OH}$ couplings values obtained at 5 and 1°C. Substantial differences in the $^3J_{H2',2'OH}$ couplings at 5 and 1°C are expected for nucleotides with $k_{ex} > 1/T$ (T being the length of the constant time period). No significant modulation is expected for nucleotides with $k_{ex} \leq 1/T$.

The results from this comparison are striking. The $^3J_{H2',2'OH}$ coupling values of canonical Watson-Crick C-G or G-C base pairs of the lower stem remain the same at 5 and 1°C within experimental error. Coupling values of nucleotides A20, U40,

and U42, of all nucleotides located in the shorter upper stem, and of the loop nucleotide C30 experience increases ranging from 30 to 50% when the temperature is decreased from 5 to 1°C. Surprisingly, the $^{3}J_{H2',2'OH}$ couplings of bulged nucleotides U23 and U25 do not change as a function of temperature. This clearly indicates that the k_{ex} rate of these nucleotides is similar to the k_{ex} rate of canonical CG base pairs.

Taken together, our results strongly suggest a correlation between 2′-OH proton solvent exchange rates k_{ex} and base-pair opening rates (probed by imino proton exchange) in canonical A-form helical structures. The exchange process of H-bonded imino protons with solvent can be qualitatively explained on the basis of a two-state model involving open and closed states of Watson-Crick pairs. If it assumed that the closed A-form geometry is stabilized by an ordered series of hydrogen-bonds involving the ribose 2′-hydroxyl group, the functional groups within the bases, and water molecules bound in the minor groove, we speculate that the transient opening of the base pair favors the exchange of bound water with the bulk solvent, thereby directly influencing the k_{ex} rate of the 2′-OH proton. Imino protons can only exchange with bulk water from transiently opened conformations through acid-base reactions catalyzed by proton acceptors; they are protected when engaged in a closed base pair (Gueron & Leroy, 1995; Snoussi & Leroy, 2001). Base pair opening events can occur via two pathways as revealed by MD simulations (Pan & MacKerell, 2003; Giudice, E. & Lavery, 2003). Both major and minor groove opening is feasible with the major groove pathway being energetically slightly more favorable. Our hypothesis of correlated imino and 2′-OH proton exchange is supported by the fact that the base-pair lifetimes of RNA A-U base pairs in the A-form helix geometry are substantially shorter than those of G-C base pairs (Gueron & Leroy, 1995; Snoussi & Leroy, 2001). Subsequently, only the cytidines and guanosines of the lower stem and the two bulged uridines were further considered in the quantitative evaluation of the $^{3}J_{H2',2'OH}$, $^{3}J_{C1',2'OH}$ and $^{3}J_{C3',2'OH}$ scalar couplings.

A Karplus curve showing the dependence of the $^{3}J_{H2',2'OH}$ on the intervening torsion angle H2′-C2′-O2′-OH is shown in Fig. 5.8D (blue line) (Fraser *et al.*, 1969). The upper limit of 5.5 Hz observed for the nucleotides of the lower stem excludes the possibility that any of these 2′-OH hydroxyl groups occupy the O4′

domain. This observation indicates that the 2′-OH is not involved in intra-ribose 2′-OH···O4′ electrostatic interactions, but rather contacts either the base (2′-OH conformation in the base domain) or the phosphate backbone (O3′ domain). The absence of 2′-OH protons in the O4′ domain in canonical A-form helical structure of the lower stem in the TAR-RNA is consistent with our semi-quantitative analysis of detectable NOE cross peaks in a short mixing time 2D NOESY experiment (τ_m=50 ms). Even at 1°C, only the $^3J_{H2',2'OH}$ of residue C30 reaches 7 Hz, while larger values (> 8 Hz) of the $^3J_{H2',2'OH}$, associated with a significant portion of the O4′ domain, are never observed.

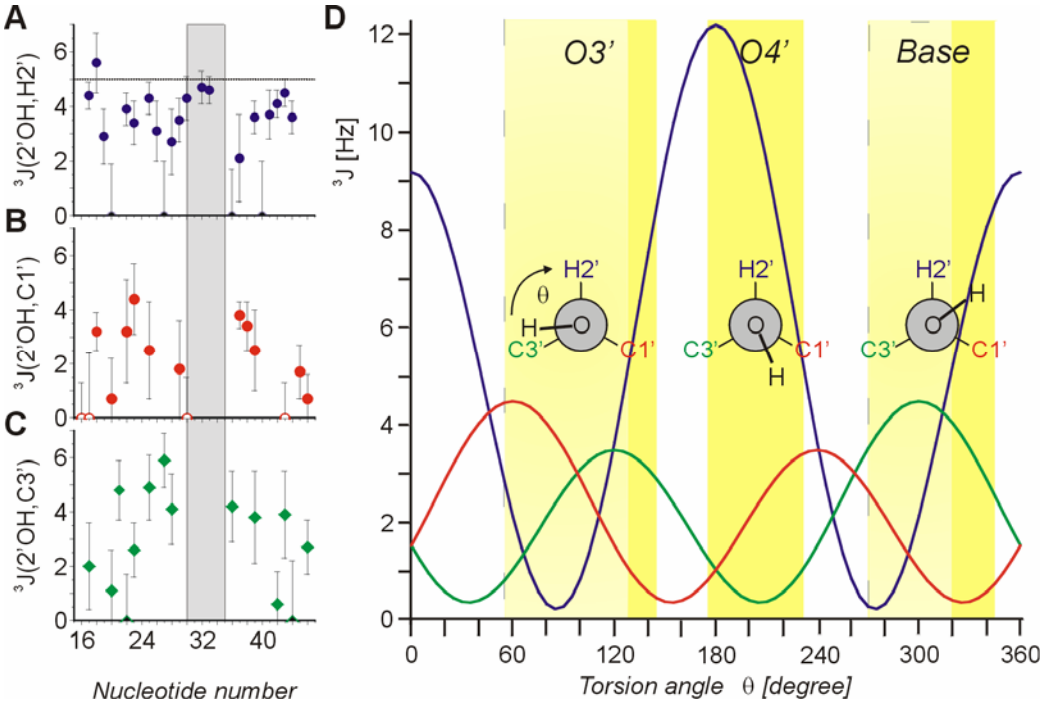


Fig. 5.8. A) $^3J_{H2',2'OH}$ couplings measured at 5°C employing the quantitative-J H2′$_i$-H6/8$_{i+1}$ CT-NOESY experiment. The dotted line indicates $^3J_{H2',2'OH}$ = 5.0 Hz B) $^3J_{C1',2'OH}$ couplings obtained at 5°C using the quantitative-J C1′$_i$-H1′ CT-HSQC experiment. C) $^3J_{C3',2'OH}$ couplings obtained at 5°C from the quantitative-J C3′$_i$-H1CT-HSQC experiment. Apical hexaloop nucleotides are shaded in grey. D) Karplus-like dependence of the homonuclear $^3J_{H2',2'OH}$ coupling (blue line) on the torsion angle θ (H2′-C2′-O2′-H) according to Fraser *et al*., 1969. No Karplus parameterization is available for the heteronuclear $^3J_{C1',2'OH}$ and the $^3J_{C3',2'OH}$ couplings. The red and green lines for the $^3J_{C1',2'OH}$ and the $^3J_{C3',2'OH}$ couplings, respectively, represent approximated Karplus curves derived by scaling the coefficients of the $^3J_{C,H}$ Karplus-relation in alkenes ($^3J_{C,H}$ = 3.6 cos2ϕ -1.0 cosϕ + 4.3) to reflect the averaged value of the $^3J_{C,OH}$ coupling in ethanol (2.9 Hz) (Kalinoswki *et al*., 1988) We assume equal populations of the *trans*, *gauche*+ and *gauche*– conformations for ethanol. Neither the

difference in the substituents of the C1′ and C3′ nuclei nor the fact that these nuclei are part of a conformational restricted ring system are considered in this approximation; therefore the Karplus-relations for the $^3J_{C1',2'OH}$ or the $^3J_{C3',2'OH}$ couplings represent crude approximations. Boxed yellow areas correspond to the energetically favored orientational domains (O3', O4', and Base, respectively) of the 2'-hydroxyl group, according to Auffinger & Westhof, 1997. The pale yellow delimit the regions of these domains where the $^3J_{H2',2'OH}$ coupling is less than 5.0 Hz.

The heteronuclear $^3J_{C1',2'OH}$ couplings proved to be difficult to measure due to the limited resolution of the C1′-H1′ correlation in canonical A-form geometry (Fig. 5.8B). Moreover, 8 out of 29 carbons (G26, A27, U31, G32, G33, G34, A35 and G36) are not observed at 5°C, which can be attributed to interconverting C2′-endo and C3′-endo sugar puckers. Those nucleotides are located predominantly in the apical hexanucleotide loop and assume averaged $^3J_{H1'-H2'}$ coupling values at elevated temperatures (25°C), indicative of fast exchange between C2′-endo and C3′-endo sugar puckers. With the exception of residue U23, all measurable heteronuclear $^3J_{C1',2'OH}$ couplings are smaller than 4 Hz with nucleotides in the lower stem exhibiting particularly small values ($^3J_{C1',2'OH} \leq 3$ Hz). Due to their small size, the achievable precision for the $^3J_{C1',2'OH}$ coupling measurements is limited, as evident from inspecting the relatively high error bars.

Similarly, resonance overlap and the absolute coupling size affect the measurement of the heteronuclear $^3J_{C3',2'OH}$ couplings (Fig. 5.8C) using the C3′-H3′ correlation. The observed $^3J_{C3',2'OH}$ couplings are on an average larger than the corresponding $^3J_{C1',2'OH}$ couplings. Again, the smallest values are observed for the nucleotides in the lower stem. Interestingly, the C3′-H3′ cross peaks of the apical hexaloop nucleotides do not disappear due to the conformational exchange of the ribose but considerably decrease in intensity. Therefore, the loop nucleotide $^3J_{C3',2'OH}$ couplings are characterized by large errors and are not reported here.

Unfortunately, no Karplus parameterization is available for either one of the heteronuclear J-couplings, $^3J_{C1',2'OH}$ or $^3J_{C3',2'OH}$. However, an approximated Karplus curve for the $^3J_{C1',2'OH}$ and the $^3J_{C3',2'OH}$ coupling can be obtained (Fig. 5.8D, red and green lines) by scaling the coefficients for $^3J_{C,H}$ couplings in alkenes ($^3J_{C-H} = 3.6 \cos 2\phi - 1.0 \cos\phi + 4.3$) such that they are consistent with the averaged value of the $^3J_{C,OH}$ coupling observed in ethanol (2.9 Hz) (Kalinowski *et al.*, 1988). Evidently, neither the difference in the substituents of the C1′ and C3′

nuclei nor the fact that these nuclei are part of a conformationally restricted ribose ring is considered in this crude estimate; therefore, this approximated Karplus relation as shown in Fig. 5.8D should not be used to quantitatively interpret the $^3J_{C1',2'OH}$ or the $^3J_{C3',2'OH}$ couplings. A qualitative inspection of the approximated Karplus curves nevertheless permits us to assume that $^3J_{C1',2'OH} < {}^3J_{C3',2'OH}$ in the base domain and in the region of the O3′ domain where $\theta > 90°$, whereas $^3J_{C3',2'OH} < {}^3J_{C1',2'OH}$ in the region of the O3′ domain with $\theta < 90°$.

Subsequently, we performed a detailed conformational analysis of the θ (H2′-C2′-O2′-OH) torsion angle for the cytidines and guanosines located in the lower stem of the TAR-RNA and of bulged uridines using a combination of three $^3J_{H2',2'OH}$, $^3J_{C1',2'OH}$ and $^3J_{C3',2'OH}$ scalar couplings and the NOE data. The nucleotides located in the upper stem and the AU base pairs of the lower stem were not included in the analysis because the aforementioned rapid exchange of the 2′-OH with the solvent renders those J-coupling values meaningless. The nucleotides of the hexanucleotide loop were not considered either due to the apparent intermediate conformational exchange between the C2′-endo and the C3′-endo pucker of the ribose.

Our conformational analysis is based on the following criteria:

- Quantitative interpretation of the homonuclear $^3J_{H2',2'OH}$ scalar coupling according to the available Karplus parametrization (Fraser *et al*., 1969).
- Semi-quantitative interpretation of the relative magnitude of the heteronuclear $^3J_{C1',2'OH}$ and $^3J_{C3',2'OH}$ scalar couplings.
- Existence of the 2′-OH(i)-H6/H8(i+1) NOE cross peak which is strongest for the 2′-OH proton in the base domain or in a small region of the O3′ domain ($\theta \approx 60°$).

For example, residue G17 shows a weak 2′-OH(i)-H6/H8(i+1) NOE cross peak, which taken together with the value of the $^3J_{H2'-2'OH}$ coupling, permits the intervening torsion angle θ to be either $\theta \approx 45°$ or $\theta \approx 315°$; a final comparison of the $^3J_{C1'-2'OH}$ and $^3J_{C3'-2'OH}$ scalar coupling magnitudes allows to discriminate between the two alternative solutions and rotates the torsion angle θ of G17 into the base domain ($\theta \approx 315°$).

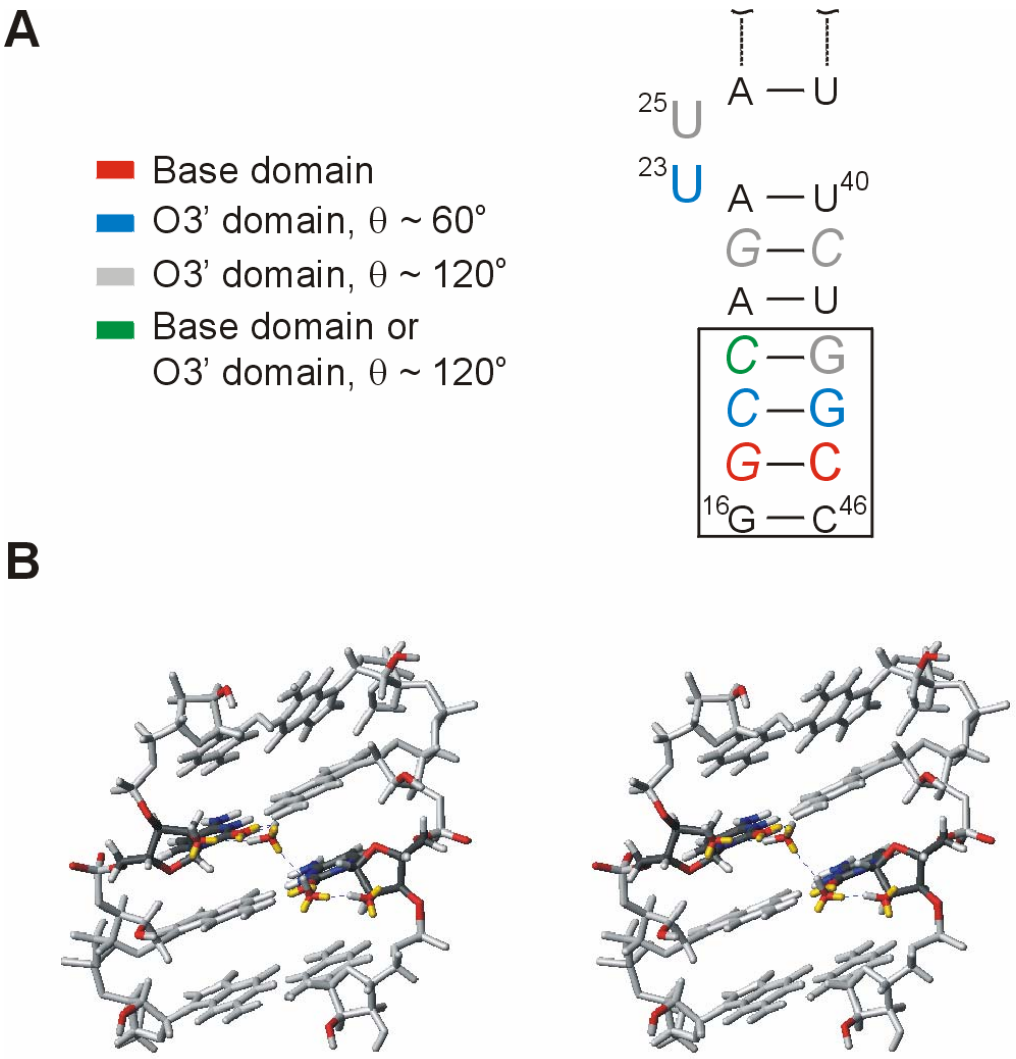


Fig. 5.9. A) Summary of the conformational preferences of the H2′-C2′-O2′-OH torsion angle θ of the cytidines and guanosines of the lower stem and bulged uridines of the HIV-2 TAR-RNA. Both the base and the O3′ domain are populated in an alternating manner, the bulged uridines show a clear preference for the O3′ domain. Not a single H2′-C2′-O2′-OH torsion angle in the O4′ domain was observed. The boxed four closing basepairs of the lower stem are shown in B). B) Network of water molecules bridging the nucleotides C18 and C45 in the minor groove of the lower stem of the HIV-2 TAR-RNA, modeled in an ideal A-form helix. Functionally important groups are colored: C18 and C45 base nitrogens are shown in blue; O2 base, O3′ backbone, and ribose O2′ oxygens are shown in red; lone pairs of water, O2 base, and ribose O2′ oxygens are colored yellow. One water molecule is situated in the base plane of C18, where it participates in an H-bond as proton donor with the O2 of the base and in an H-bond as proton acceptor with the 2′-OH (conformation in the O3′ domain). The second water molecules is located in the plane of C45, where it participates as proton donor in an H-bond with the O2 of the base and in an H-bond with the 2′-OH (conformation in the base domain).

A schematic representation of the results obtained from this analysis for the nucleotides of the lower stem of the TAR-RNA is shown in Fig. 5.9A. The conformation of the 2′-OH groups of G17 and C45 is found in the base domain, as indicated by the presence of the 2′-OH(i)-H6/H8(i+1) NOE and by the relatively small value of the $^3J_{C1',2'OH}$ coupling. The 2′-OH groups of C18, U23, and G44 adopt an orientation within the O3′ domain with the torsion angle $\theta \approx 60°$, corroborated by the presence of the 2′-OH(i)-H6/H8(i+1) NOE, the high value of the $^3J_{C1',2'OH}$ coupling, and/or the small $^3J_{C3',2'OH}$ coupling. G21, U25, C41 and G43 project the 2′-OH group towards the O3′-oxygen with $\theta \approx 120°$, as indicated by the absence of the 2′-OH(i)-H6/H8(i+1) NOE, a small value of the $^3J_{C1',2'OH}$ coupling, and/or a large value of the $^3J_{C3',2'OH}$ coupling. Finally, the conformation of the 2′-OH proton of C19 can only be restricted by the presence of the 2′-OH(i)-H6/H8(i+1) NOE because both heteronuclear $^3J_{C1',2'OH}$ and $^3J_{C3',2'OH}$ couplings could not be measured reliably for this nucleotide. In summary, our analysis reveals that the H2′-C2′-O2′-OH torsion angles of the canonical A-form lower stem nucleotides of the TAR-RNA populate both the base and the O3′ domain, while the θ torsion angles of the bulged uridines show a clear preference for the O3′ domain.

5.4 Discussion

It has been commonly accepted that the 2'-OH hydroxyl points towards the phosphate backbone, and away from the ribose H1' proton. This picture has been supported by MD simulation of an RNA hairpin (Singh & Kollman, 1996) which provided evidence for a strongly preferred orientation towards the O3', and by the NMR data available to date (Conte *et al*., 1996), where only very weak 2'-OH-H1' NOE crosspeaks could be detected. However, more recent MD simulations describe three energetically favored orientations for the 2'-OH hydroxyl proton (Auffinger & Westhof, 1997). The 2'-OH-H1' NOEs of medium intensity observed for the HIV-2 TAR-RNA contradict the paradigm that the 2'-OH hydroxyl typically points away from the ribose H1' proton. The presence of many detectable NOE correlations involving aromatic H6/H8(i+1) protons suggest that the orientation towards the base is more common than currently appreciated.

A network of water-mediated hydrogen bonds anchored by the 2′-hydroxyl group has been proposed to stabilize the A-form RNA duplex (Egli *et al*., 1996). The results of our conformational study on the H2'-C2'-O2'-OH torsion angle, based on both NOEs and scalar coupling data, are in excellent agreement with this model. The 2′-hydroxyl group in the base domain donates an H-bond to a water molecule in the base plane, while the 2′-hydroxyl group projecting towards the O3′ oxygen can accept an H-bond from a water molecule in the same base plane (Fig. 5.9B). In contrast, if the 2′-OH group adopted a conformation within the O4′ domain, the lone pair electrons of the O2′ oxygen would occupy an unfavorable orientation for their participation in an H-bonding network involving water molecules located in the base plane. Consequently, the 2′-OH groups in the canonical A-form lower stem of the TAR-RNA never project into the O4′ domain. Furthermore, we observe an alternating pattern of θ torsion angles in the base and in the O3′ domain for 3′- and 5′-strand nucleotides in consecutive Watson-Crick G-C base pair steps (5′-nucleotides 17-19 and 3′-nucleotides 43-45, respectively). Despite the limited statistical significance of this observation, such an alternating pattern with the H2′-C2′-O2′-OH torsion angle adopting base and O3′ domain conformations is required to fully support the network of hydrogen bonds described in the crystal structures of the RNA duplex (Egli *et al*., 1996). In the crystal, the two water molecules spanning the minor groove share an H-bond with each other and function as H-bond donors to the functional N3/O2 groups within the base moiety. To complete an uninterrupted chain of H-bonds in the minor groove, one of the two water molecules must accept and one must donate an H-bond to the 2′-hydroxyl groups of nucleotides of opposite strands in consecutive base pair steps (Fig. 5.9B). The 2′-hydroxyl group can function as an H-bond donor to a water molecule in the base plane when θ is in the base domain and, conversely, is able to accept a proton from the same water molecule when θ is in the O3′ domain. The ideal network of H-bond spanning the minor groove, postulated on the base of the crystal structure of an RNA duplex, therefore requires that the H2′-C2′-O2′-OH torsion angles of two nucleotides belonging to opposite strands in base pair steps assume alternating base and O3′ domain conformations.

Interestingly, the solvent exchange rates of the 2′-OH protons of the bulge nucleotides U23 and U25 are similar to those of the cytidines and guanosines in the lower stem. The scalar coupling data for the 2′-hydroxyls of U23 and U25 can be interpreted assuming a single conformation where the proton points towards the phosphate backbone in the O3′ domain. The TAR-RNA bulge region is intimately involved in ligand binding events. The bulged nucleotides, which are relatively flexible in the free TAR-RNA, assume a well-defined, rigid conformation upon argininamide or magnesium binding (Zacharias. & Hagerman, 1995; Brodsky & Williamson, 1997; Pitt et al., 2004). The NMR experiments described here were conducted in absence of any ligands and thus may indicate pre-organized local conformation of the riboses of U23 and U25. It should be noted that the phosphate between U23 and A22 located at the junction of the bulge and lower stem is important for Tat binding in the major groove of TAR (Hamy et al., 1993). Further studies that address the role of the 2′-hydroxyl group of uridines U23 and U25 in the specific recognition of positively charged ligands are ongoing.

In conclusion, the results obtained in this study are in good agreement with both molecular dynamics calculations (Auffinger & Westhof, 1997) and X-ray studies (Egli *et al*., 1996) and confirm the fundamental role of the 2′-hydroxyl group in building a network of hydrogen bonds in the minor groove, which stabilizes the preferred canonical A-form geometry of RNA. These data have important implications for the future understanding of high-affinity recognition and discrimination of A-form like geometries over canonical B-form or other helices in RNA-protein complexes. The methodology and experiments developed here are of general relevance to structural studies of RNAs and allow for an accurate determination of the location of the 2′-hydroxyl proton, which will help elucidate the fundamental role of the 2′-hydroxyl group in RNA function, ligand binding, and catalysis.

Conclusions

In this work we have studied the principle of molecular recognition by NMR in three complexes: a) the protein tubulin in complex with the microtubules stabilizing agent epothilone; b) the small peptidic toxin κM-conotoxin RIIIK, derived from the sea snail *Conus radiatus,* in complex with K^+ channels; c) the 30mer HIV TAR-RNA in complex with an aminoglycoside derivative. The first two complexes are too large to obtain detailed structure information by classical NMR methods. Additionally the protein receptors (tubulin and *TSha1* K^+ channel) are not available by recombinant techniques with the ^{13}C, ^{15}N and ^{2}D labelling scheme necessary to conduct multidimentional NMR experiments on the protein. Therefore, novel NMR methodologies and a combined approach using computational, biological and NMR-derived structural data were developed to access the desired structure. These studies prove that the size limit of 50 kDa, usually posed to solution NMR, can be successfully overcome. Structural information can be obtained for complexes as large as 1 MDa (tubulin/epothilone), if we wish to answer specific questions about intermolecular recognition rather than to reveal the structure of the whole complex. The techniques developed in this study, inluding transferred *cross-correlated relaxation rates* and the *INPHARMA* approach (Interligand NOE for PHARmacophore MApping) are of general applicability to ligand/target complexes with a K_d in the micromolar range and are expected to have a significant impact in the process of structure-based drug design.

The structural information gained in the three projects sheds a new light on the intermolecular interaction governing the biological function in the three systems.

During this work we determined the bound conformation and the binding mode of the potent anticancer drug epothilone to tubulin in aqueous solution by tr-NOEs, tr-CCR rates and by the INPHARMA approach. On the basis of our structural model we proposed a mechanism for the microtubules stabilizing activity of the drug based on the fine regulation of the position of the H7 helix. The EpoA – tubulin complex structure explains well the epothilone structure-activity data both for native and mutant tubulins, provides a rationale for the highest activity of EpoB with respect to EpoA and supports the existence of a common

pharmacophore for epothilone and paclitaxel. The structure presented here differs from the EC-derived structure of EpoA bound to tubulin in Zn-stabilized sheets, which significantly contradicts all NMR-derived structural data. This discrepancy suggests that the binding mode of EpoA to tubulin depends on the tubulin polymerization state. While tubulin polymerized in Zn-stabilized sheets is an artificial system, the soluble protein is a physiologically relevant state in the tubulin cell cycle. Therefore, the pharmacophore of epothilone bound to soluble tubulin is highly relevant for the structure-based design of improved anticancer drugs.

The structural model of the peptidic toxin κM-conotoxin RIIIK in complex with the *TSha1* K^+ channel reveals a novel pharmacophore, where no positively charged side chain occludes the channel pore. We have proven that κM-RIIIK does not possess the functional dyad constituted by a positively charged and an hydrophobic side-chain, which had been proposed to be the minimal functional element for the binding of peptidic toxins to K^+ channels. Instead, the positive charged residues of the toxin form a basic ring: κM-RIIIK is anchored to the K^+ channel via electrostatic interactions of this basic ring with the loop and pore helix residues of the channel. The electrostatic complementarity between this ring and the channel side chains is likely to be responsible for the selectivity of κM-RIIIK towards the *TSha1* channel. The role of the positive charged side chain that in many peptide – ion channel complexes occludes the channel pore, is partially exerted by O15, whose side chain, however, is too short to penetrate the pore. This novel, experimentally-based pharmacophore model proves the existence of diverse binding modes of peptidic toxins to K^+ channels and underlines the role of intermolecular electrostatic interactions involving channel loop side chains in determining the selectivity of toxins for specific K^+ channel types. The binding mode of κM-RIIIK to the *TSha1* channel closely resembles that of μ-conotoxin GIIIA to the $Na_V1.4$ Na^+ channel. O15 of κM-RIIIK occupies the homologous position of R13 of μ-GIIIA, which is known to occlude the Na^+ channel pore. Both peptides contain a basic ring, although the form and the charge density at the surface differs. This observation demonstrates that the two conotoxins have developed a similar scaffold and binding geometry to their respective targets,

while the pharmacological selectivity is determined by the exact charge distribution and three-dimensional shape.

The three-dimensional structure of the TAR-RNA in complex with a novel macrocyclic aminoglycoside analogue reveals interaction of the ligand with the bulge, the upper stem major groove and the loop of the TAR-RNA. This is the first example of a positively charged ligand contacting the capping loop of the TAR-RNA. In particular the aminoglycoside analogue binds to the A35 residue of the hexanucleotide loop, thereby inducing a partial structuring of the loop. The correct replication of the HIV virus requires the binding of the TAR-RNA to both the Tat-protein and the P-TEFb complex. The role of the TAR-RNA loop in interacting with the Tat-protein and the P-TEFb complex has been demonstrated by many biological data; however no defined structure could be identified for the loop of the TAR-RNA either alone or in complex with cofactors. Our structural study identifies for the first time well-defined structural features induced in the capping loop of the TAR-RNA by binding to the aminoglycoside analogue. Biological data indicate that the subunit CycT1 of the P-TEFb complex binds to the loop of the TAR-RNA in a Tat dependent manner. We suggest that the conformational change observed in the loop of the TAR-RNA upon binding of the aminoglycoside analogue resembles the restructuring of the loop required for CycT1 binding and that this restructuring is induced by Tat. On the basis of the structural information obtained in this study, we believe that the amino glycoside analogue holds promises to be used as a scaffold for the design of a new generation of TAR binders, which can target several binding sites in the bulge, major groove and loop of the TAR-RNA and inhibit both the Tat/TAR and Tat/TAR/CycT1 complex formation.

In conclusion, in this work we have provided new valuable methodology to investigate the principles of intermolecular interactions in large complexes in solution and could reveal the mechanisms of intermolecular recognition in three biological systems. The structural and biological information gained in this work represent a conspicuous progress in our understanding of intermolecular interactions, while the developed methodology provides us with new instruments to listen to how molecules talk to each other.

References

Aboulela, F., *et al.* (1996) *Nucleic Acids Res.*, **24**, 4598.

Aboulela, F., *et al.* (1995) *J. Mol. Biol.*, **253**, 313.

Acharya, P. & Chattopadhyaya, J. (2002) *J. Org. Chem.* **67**, 1852.

Al-Hashimi, H.M. *et al*. (2001) *J. Am. Chem. Soc.* **123**, 5806.

Al-Sabi, A. *et al.* (2004) *Biochemistry* **43**, 8625.

Altmann, K. –H. *et al.* (2000) *Chimia* **54**, 612.

Altmann, K.-H. (2001) *Curr. Opin. Chem. Biol.* **5**, 424.

Altmann, K.-H. (2003) *Mini-Reviews in Med. Chem.* **3**, 149.

Arzumanov, A. *et al*. (2001) *Biochemistry*, **40**, 14645.

Athanassiou, Z. *et al*. (2004) *J. Am. Chem. Soc.*, **126**, 6906.

Auffinger, P. & Westhof, E. (1997) *J. Mol. Biol.* **274**, 54.

Balaram, P. *et al.* (1972) *J. Am. Chem. Soc.* **94**, 4015.

Bax, A. *et al.* (1990) *J. Magn. Reson.*, **86**, 304.

Beckwith, M. *et al*. (1993) *J. Natl. Cancer Inst*. **85**, 483.

Berendsen, H. J. C. *et al.* (1995) *Comp. Phys. Comm.* **91**, 43.

Bingham, R. C. *et al.* (1975) *J. Am. Chem. Soc.* **97**, 1285.

Blommers, M. J. J. *et al.* (1999) *J. Am. Chem. Soc.* **121**, 1949.

Borzilleri, R. M. & Vite, G. D. (2002) *Drugs of the Future* 7, 1149.

Bollag, D. M. *et al.*, (1995) *Cancer Res.* **55**, 2325.

Braunschweiler, L. & Ernst, R. R. (1983) *J. Magn. Reson*., **53**, 521.

Brodsky, A. S. & Williamson, J. R. (1997) *J. Mol. Biol.* **267**, 624

Brünger, A. T. (1993) *X-PLOR Version 3.1: A system for X-ray crystallography and NMR.* Yale University Press, New Haven, CT, USA.

Brünger, A. T. (1992) *X-PLOR A System for X-Ray Crystallography and NMR* (Yale University Press, New Haven)

Brünger, A.T. *et al.* (1998) *Acta Cryst.* **D54**, 905-951

Brutscher, B. *et al.* (2001) *J. Biomol. NMR*, **19**, 141.

Buey, R. M. *et al.* (2004) *Chem Biol.* **11**, 225.

Carlomagno, T. *et al.* (1999) *J. Am. Chem. Soc.* **121**, 1945.

Cavanagh, J. *et al.* (1996) *Protein NMR spectroscopy*, Academic Press Limited, London, UK.

Cavanagh, J. and Rance, M. (1992) *J Magn. Reson.* **96**, 670.

Chakrabarti, G. *et al.* (1999) *Biochemistry* **38**, 3067.

Chen, G.-W. & Kirschning, A. (2002) *Chem. Eur. J*. **8**, 2717.

Choy, N. *et al.* (2003) *J. Med. Chem.* **46**, 2846.

Churcher, M. J. *et al.* (1993) *J. Mol. Biol.* **230**, 90.

Cilley, C. D. & Williamson, J. R. (1997) *RNA* **3**, 57.

Clore, G. M. & Gronenborn, A. M. (1982) *J. Magn. Reson.* **48**, 402.

Colvin, R. A. *et al.* (1993) *Biochemistry* **32**, 1105.

Conte, M. R. *et al.* (1996) *Nucleic Acids Res.* **24**, 3693.

Cordingly, M. G. (1990) *Proc. Natl Acad. Sci. USA* **87**, 8985

Dauplais, M. *et al.* (1997) *J. Biol. Chem*., **272**, 4802.

Daura, X. *et al.* (1999) *Angew. Chem. Int. Ed*. **38**,236.

Davis, B. *et al*. (2004) *J. Mol. Biol.* **336**, 343.

Delaglio, F. *et al.* (1995) *J. Biomol. NMR* **6**, 277.

Diamond, R. (1992) *Protein Science*, **1**, 1279.

Diaz, J. F. & Andreu, J. M. (1993) *Biochemistry* **32**, 2747.

Dingwall, C. *et al.* (1989). *Proc. Natl Acad. Sci. USA*, **86**, 6925.

Dingwall, C. *et al.* (1990). *EMBO J.* **9**, 4145.

Doherty, E.A. & Doudna, J. A. (2000). *Annu. Rev. Biochem.* **69**, 597.

Dominguez, C. *et al*. (2003) *J. Am. Chem. Soc.* **125**, 1731.

Downing, K.H. & Nogales, E. (1998) *Eur. Biophys. J.* **27**, 431.

Doyle, D.A. *et al.* (1998) *Science.* **280**, 69.

Eckstein F. & Lilley D. M. J. (1996) *Catalytic RNA* (Eckstein F. & Lilley D. M. J., eds.), Springer, Berlin

Egli M., *et al.* (1996) *Biochemistry*, **35**, 8489.

Eriksson, M. A. L. & Roux, B., (2002) *Biophys. J.,* **83**, 2595.

Faber, C. S. H. *et al*. (2000) *J. Biol. Chem.* **275**, 20660.

Felli, I. C. *et al.* (1999) *J. Am. Chem. Soc.* **121**, 1956.

Ferber, M. *et al.* (2003) *J. Biol. Chem.* **278,** 2177.

Frankel, A. D. (1992) *Curr. Opin. Genet. Dev.* **2**, 293.

Fraser, R. R. *et al*. (1969). *Can J Chemistry* **47**, 403.

Garber, M. E. *et al.* (1998) *Genes Dev.* **12**, 3512.

Gao, Y.-D. & Garcia, M. (2003) *Proteins* **52**, 146.

Geen, H. and Freeman, R. (1991) *J. Magn. Reson.*, **93**, 93.

Gesteland R. F. & Chech, T. (1998). *The RNA World* (Gesteland, R.F., Chech, T. & Atkins J.F., eds.), Cold Spring Harbor Laboratory Press, New York

Giannakakou, P. *et al.* (2000) T. *Proc. Natl. Acad. Sci. U.S.A.* **97**, 2904.

Giedroc, D. P., *et al*. (2003) *J Am. Chem. Soc.* **125**, 4676.

Gilquin, B. *et al.* (2002) *J. Biol. Chem.,* **277,** 37406.

Gigant, B. *et al.* (2005) *Nature* **435**, 519.

Giudice, E. & Lavery, R. (2003) *J Am Chem Soc* **125**, 4998.

Gueron, M. & Leroy, J. L. (1995) *Methods Enzymol* **261**, 383.

Gyi J. I., *et al.* (1998) *Nucleic Acids Res.*, **26**, 3104.

Hamy, F. *et al*. (1998) *Biochemistry* **37**, 5086.

Hamy, F. *et al.* (1993) *J Mol Biol* **230**, 111.

Harbison, G. S. (1993) *J. Am. Chem. Soc.* **115**, 3026.

He, L. *et al.* (2000) *Biochemistry* **39**, 3972.

He, L. *et al.* (2001) *Mol. Cancer Ther.* **1**, 3.

Hermann, T. & Westhof, E. (1998) *Biopolymers* **48**, 155.

Hermann, T. & Westhof, E. (1999) *J. Med. Chem.* **42**, 1250.

Hill, J. M. *et al.* (1996) *Biochemistry* **35,** 8824.

Hille, B. (2001) *Ion Channels of Excitable Membranes*, 3rd Edition, Sinauer Associates, Inc.

Höfle, G. *et al.* (1993) *Chem. Abstr.* **120**, 52841.

Höfle, G. *et al.* (1996) *Angew. Chem. Int. Ed.* **35**, 1567.

Höfle, G. *et al.* (1999) *Angew. Chem. Int. Ed.* **38**, 1971.

Hood *et al*. (2002) *Cancer Res*. **62**, 3356.

Hu, W. D *et al.* (1998) *J. Biomol. NMR* **12**, 559.

Huq, I. & Rana, T.M. (1997) *Biochemistry*, **36**, 12592.

Huys, I. *et al.* (2004) *Biochem. J.* **378**, 745.

Hwang, T.-L. and Shaka, A.J. (1995) *J Magn. Reson.* **A 112**, 275.

Ilin, S. *et al.* (2003) *Angew. Chem. Int. Ed.* **42**, 1394.

Ishii, T. M. *et al.* (1998) *Methods Enzymol.* **293**, 53.

Ivery, M. T. G. & Le, T. (2003) *Oncol. Res.* **14**, 1.

Jacobsen, R. B. E. *et al.* (2000) *J. Biol. Chem.* **275**, 24639.

Jäger J. A. & Tinoco, I. (1993) *Biochemistry* **32**, 12522

Jaunzems, J. *et al.* (2004) *Org. Biomol. Chem.* **2**, 3448.

Jeener, J. *et al.* (1979) *J. Chem. Phys*. **71**, 4546.

Jones, K. A. & Peterlin, B. M., (1994) *Annu. Rev. Biochem*. **63**, 717.

Jordan, M. A. (2002) *Curr. Med. Chem: Anti-Cancer Agents* **2**, 1.

Jucker, F.M. & Pardi, A. (1995) *Biochemistry*, **34**, 14416.

Kaiser, E. *et al.* (1970) *Anal. Biochem.* **34**, 595.

Kalinowski, H.-O., Berger, S. & Braun, S. (1988). *Carbon 13 NMR Spectroscopy*, John Wiley & Sons.

Karn, J., (1999) *J. Mol. Biol.* **293**, 235.

Kaushik, N. *et al*. (2002). *J. Virol.* **76**, 3881.

Khalil, M.-W.M., (1999) Tubulysin aus Myxobakterien: Untersuchungen zum Wirkungsmechanismus *Von der Gemeinsamen Naturwissenschaftlichen Fakultät der Technischen Universität Carolo-Wilhelmina, Braunschweig.*

Kirschning, A. & Chen, G.-W. (1999) *Tetrahedron Lett*. **40**, 4665.

Kirschning, A. & Chen, G.-W. (2000) *Synthesis* **XX**, 1133.

Kirschning, A. *et al.* (2004) *Tetrahedron* **60**, 3505.

Koenig, B.W. *et al.* (2000) *J. Biomol. NMR* **16**, 121.

Koradi, R. *et al.* (1996) *Journal of Molecular Graphics*, **14**, 51.

Kowalski, R. J. *et al.* (1997) *J. Biol. Chem.* **272**, 2534.

Kulinski, T. *et al.* (2003) *J. Biol. Chem.*, **278**, 38892.

Lancelin, J. M. *et al.* (1991) *Biochemistry* **30**, 6908.

Leeper, T.C. *et al*. (2005) *Biochemistry*, **44**, 12362.

Leupin W. *et al*. (1990) *Febs Lett*. **263**, 313.

Lippens, G. *et al.* (1995) *J Biomol. NMR* **5**, 327.

Li, Y. *et al.* (2000) *Biochemistry* **39**, 281.

Li, H. *et al.* (2002) *Structure* **10**, 1317.

Lindahl, E. et al. (2001) *J. Mol. Mod.* **7**, 306.

London, R. E., (1999) *J. Magn. Reson.* **141**, 301.

London, R. E. *et al.* (1992) *J. Magn. Reson.* **97**, 79.

Luca, S. *et al.* (2003) *Acc. Chem. Res.* **36**, 858.

Lynch S. R. *et al.* (1996). *Magn. Reson. Chem.* **34**, S11-S17.

Mancebo H. S. Y. *et al.* (1997) *Genes Dev.* **11**, 2622

Marion, D. et al. (1989) *J. Magn. Reson.*, **85**, 393.

Martello, L. A. *et al.,* (2000) *Clin. Cancer Res.* **6**, 1978.

Mei, H.-Y. *et al.* (1995) *Bioorg. Med. Chem Lett*. **5**, 2755.

Mei, H.-Y. *et al*. (1998) *Biochemistry* **37**, 14204.

Meiler, J. *et al*. (2000) *J. Biomol. NMR*, **17**, 283.

Michael, K. & Tor, Y. (1998) *Chem. Eur. J.* **4**, 2091.

Miller, C., (1995) *Neuron* **15**, 5.

Miller, C. & Rivier, J. (1998) *J. Pept. Res.* **51**, 444.

Mingeot-Leclercq, M. P. & Turkens, P. M. (1999) *Antimicrob. Agents Chemother.* **43**, 1003.

Minguez, J. M. *et al.* (2003) *Bioorg. & Med. Chem.* **11**, 3335.

Mischiati, C. *et al.* (2001) *Antisense Nucl. Acid Drug Dev.* **11**, 209.

Mouhat, S. *et al.* (2004) *Biochem. J.* **377**, 25.

Nettles, J. H. *et al.* (2004) *Science* **305**, 866.

Ni, F., (1994) *Prog. Nucl. Magn. Reson. Spectrosc.* **26**, 517.

Ni, F. (1996) *J. Magn. Res, Series B*, **106**, 147.

Nicolaou, K. C. *et al.* (1997) *Angew. Chem. Int. Ed.* **36**, 2097.

Nicolaou, K. C. *et al.* (1998) *Angew. Chem. Int. Ed.* **37**, 2014.

Nicolaou, K.C. *et al.* (2000) *Chem. Biol.* **7**, 593.

Nixon P. L., *et al.* (2002) *J. Mol. Biol.* **322**, 621.

Nogales, E. *et al.* (1998) *Nature* **391**, 199.

Nogales, E. *et al.* (1999) *Cell* **96**, 79.

Nogales, E. (2001) *Annu. Rev. Biophys. Biomol. Struct.* **30**, 397.

Norwood, T. J., (1993) *J. Magn. Reson. A* **104**, 106.

Ojima, I. *et al.* (1999) *Proc. Natl. Acad. Sci. USA* **96**, 4256.

Pan, Y. & MacKerell, A. D., Jr. (2003). *Nucleic Acids Res* **31**, 7131

Parkinson, G. *et al.* (1996) *Acta Cryst. D*, **52**, 57.

Paterson, I. & Florence, G. J., (2003) *Eur. J. Org. Chem.* **12**, 2193.

Piotto, M. *et al*. (1992) *J. Biomol. NMR* **2**, 661.

Pitt, S. W., *et al.* (2004) *J Mol Biol* **338**, 7.

Ramakrishnan, V. & Moore, P.B., (2001) *Curr. Opin. Struct. Biol.* **11**, 144.

Rance, M. *et al.* (1983) *Biochem. Bioph. Res. Commun.* **117**, 479.

Ranganathan, R., *et al.* (1996) *Neuron* **16**, 131.

Rauer, H. *et al.* (1999) *J. Biol. Chem.* **274**, 21885.

Ravelli, R.B.G. *et al.* (2005) *Nature* **428**, 198.

Regueiro-Ren, A. *et al.* (2002) *Org. Lett.* **4**, 3815.

Reif, B. *et al.*, (1997) *Science* **276**, 1230.

Richter, S. *et al.,* (2002). *Biochemistry*, **41**, 6391.

Richter, S. *et al*. (2002a) *Proc. Natl. Acad. Sci. USA*, **99**, 7928.

Rivkin, A. *et al.* (2003) *J. Am. Chem. Soc.* **125**, 2899.

Rodriguez de la Vega, R.C. *et al*. (2003) *TRENDS in Pharmacological Sciences* **24**, 222.

Rowinsky, E. K. (1997) *Ann. Rev. Med.* **48**, 353.

Roy, S. *et al*. (1990) *Genes Develop*. **4**, 1365

Rückert, M., & Otting, G. (2000) *J. Am. Chem. Soc.* **122**, 7793.

Sali, A. & Blundell, T. L. (1993) *J. Mol. Biol.* **234**, 779.

Samanta, U. *et al.* (1999) *Acta Cryst.* **D55**, 1421.

Sambrook, J. & Russell, D. W. (2001) *Molecular cloning: A Laboratory Manual.* 3rd Edition, Cold Spring Harbor Laboratory Press, Cold Spring Harbor NY.

Sanchez-Pedregal *et al.* (2005) *Angew. Chem. Int. Ed.* **44**, 4172

Sasse, F. *et al.* (2000) *J. Antibiotics* **53**, 879.

Savarin, P. *et al.* (1998) *Biochemistry* **37**, 5407.

Schreiber, G. & Fersht, A. R. (1995) *J. Mol. Biol.* **248**, 478.

Schwalbe, H *et al*. (1994) *J. Biomol. NMR*, **4**, 631.

Schwieters, C. D. et al. (2003) *J. Magn. Reson.*, **160**, 66.

Scott, L.G *et al.* (2000) *Methods Enzym.* **317**, 18.

Shon, K. *et al.* (1998) *J. Biol. Chem.* **273**, 33.

Singh S.B. and Kollman P.A. (1996) *Biophys. J.* **70**, 1940.

Smith, A. B. III *et al.* (2000) *J. Am. Chem. Soc.* **122**, 8654.

Snoussi, K. & Leroy, J. L. (2001) *Biochemistry* **40**, 8898

Solomon F., (1977) *Biochemistry* **16**, 358.

Srinivasan, K. N. *et al.* (2002), *J. Biol. Chem.* **277**, 30040.

Stampe, P. *et al.* (1994) *Biochemistry* **33**, 443.

Stewart, J. *et al.* (1976) *The 14th European Peptide Symposium*; Editions de l'Université de Bruxelles, Bruxelles (Belgique): Wépion, Belgium, 285-290.

Stewart, J. J. P. (1990) *Quant. J. of computer-aided molecular design* **4**, 1.

Sumner-Smith, M. *et al*. (1991) *J. Virol*. **65**, 5196.

Taylor, R.E. & Zajicek, J. (1999) *J. Org. Chem.* **64**, 7224.

Taylor, R.E. *et al*. (2003) *J. Am. Chem. Soc.* **125**, 26.

Taylor, R. E. *et al.* (2004) *Org. Biomol. Chem.* **2**, 127.

Terlau, H., & Olivera, B. M. (2004) *Physiol. Rev. 84*, 41-68

Terlau, H. *et al.* (1996) *Nature* **381**, 148.

Theimer, C. A., *et al*. (2005) *Molec. Cell* **17**, 671.

Tinley, *et al.* (2003) *Cancer Res.* **63**, 3211.

Van Wagoner, R. M. *et al.* (2003) *Biochemistry* **42**, 6353.

Van Wagoner, R. M., & Ireland, C. M. (2003) *Biochemistry* **42**, 6347.

Varani, T.C.L *et al.* (2005) RNA 394 ??????

Vuister, G.W. and Bax, A. (1993) *J. Am. Chem. Soc.* **115**, 7772.

Wakamatsu, K. *et al.* (1992) *Biochemistry* **31**, 12577.

Wang, S. *et al.* (1998) *Biochemistry*, **37**, 5549.

Wang, M. *et al.* (1999) *Org. Lett.* **1**, 43.

Wang, H.-W. & Nogales, E. (2005) *Nature* **435**, 911.

Weeks, K. M. *et al.* (1990) *Science* **249**, 1281.

Weeks, K. M. & Crothers D. M., (1991) *Cell* **66**, 577.

Wei, P. *et al.* (1998) *Cell* **92**, 451

Weis, W.I. & Drickamer, K. (1996) *Annu. Rev. Biochem.*, **65**, 441.

Wijmenga, S.S. & van Buuren, B.N.M (1998) *Prog. Magn. Reson. Spectrosc*. **32**, 287.

Williams D. J., *et al*. (2001) *RNA* **7**, 44.

Wilson, L. *et al*. (1982) *J. Mol. Biol.* **159**, 125.

Wilson, W. D. & Li, K. (2000) *Curr. Med. Chem.* **7**, 73

Wüthrich, K. (1986) *NMR of proteins and nucleic acids*. Wiley, NY.

Xu, C.-Q. *et al.* (2003) *TRENDS in Pharmacological Sciences* **24**, 446.

Yang, D.W. *et al.* (1997) *J. Am. Chem. Soc.* **119**, 11938.

Zacharias, M. & Hagerman, P. J. (1995). *Proc Natl Acad Sci USA* **92**, 6052.

Zhang, Q. *et al.* (2003) *J. Am. Chem. Soc.* **125**, 10530.

Zhu, Y. *et al.* (1997) *Genes Dev*. **11**, 2622.

Zwahlen, C. *et al*. (1997) *J. Am. Chem. Soc.* **119**, 6711.