

Structural biology
of bacterial response regulator proteins
and their complexes with cognate ligands

Inauguraldissertation

zur
Erlangung der Würde eines Doktors der Philosophie
vorgelegt der
Philosophisch-Naturwissenschaftlichen Fakultät
der Universität Basel

von
Martin Georg Allan
aus
Freiburg

Basel, 2009

Genehmigt von der Philosophisch-Naturwissenschaftlichen Fakultät auf Antrag von
Prof. Dr. Urs Jenal und Prof. Dr. Stephan Grzesiek.

Basel, den 13. November 2007

Prof. Dr. Hans-Peter Hauri
Dekan

Abstract

Two bacterial response regulator systems were studied in this thesis. NMR spectroscopy and X-ray crystallography were used to determine the structures of both domains of the antibiotic sensor TipAL, as well as of the c-di-GMP receptor PA4608, in complex with cognate ligands. By comparison with structures of the free proteins, we found that ligand binding induced biologically relevant structural rearrangements in all proteins studied.

The putative multidrug resistance gene *tipA* of the soil bacterium *Streptomyces lividans* codes for the transcription factor TipAL, which comprises two domains: a DNA-binding domain named TipAN, and a ligand-binding domain named TipAS, which is capable of recognizing and binding a diverse group of macrocyclic thiopeptide antibiotics. After antibiotic binding, TipAL binds to promoter DNA and activates transcription. In order to elucidate the specificity and flexibility of antibiotic recognition as well as the mechanism of transcriptional activation, the two domains of TipAL were studied separately.

We have determined the structures of TipAS with bound promothiocin A or nosiheptide antibiotics by NMR spectroscopy. The N-terminal part of the TipAS sequence, which is flexible and unstructured in free TipAS, forms three new helices in both complexes, burying the bound antibiotics. Considering that the newly formed helices form the connection between the TipAS and TipAN domains, we propose that the formation of additional secondary structure forms the basis of transcriptional activation by TipAL after ligand binding. The TipAS complexes with promothiocin A and nosiheptide are similar, but differ in the dynamics of the newly formed helices; the smaller ligand, promothiocin A, appears to leave more room for movement of TipAS.

The structure of TipAN in complex with a fragment of *tipA* promoter DNA was solved by X-ray crystallography. TipAN binds to the symmetric promoter as a dimer, which is held together by a long, antiparallel coiled coil. In contrast to homologous proteins, TipAN does not bend and twist bound DNA, which is a prerequisite for transcriptional activation by other proteins of the same family. This indicates that the activated TipAS domain is required for transcriptional activation by TipAL.

C-di-GMP is a second messenger molecule that appears to be ubiquitous in, and unique to, the bacterial kingdom. It generally controls the switch from motile, single-cell lifestyles to surface-attached, multicellular communities such as biofilms. The natural receptors of c-di-GMP are the PilZ domain proteins, which include PA4608 in *Pseudomonas aeruginosa*. We have solved the NMR solution structure of PA4608 in complex with c-di-GMP. C-di-GMP binds to the protein as an intercalated dimer,

displacing the C-terminal 3_{10} helix found in the apo form. The N-terminal part of PA4608, which contains the highly conserved RxxxR motif and which is flexible and unstructured before ligand binding, contacts the distal side of c-di-GMP.

Table of contents

Abstract.....	<i>ii</i>
Acknowledgements.....	<i>v</i>
Introduction: macromolecular structure determination.....	A
Solution NMR spectroscopy.....	A-1
X-ray crystallography.....	A-6
Structural basis of multidrug binding to a bacterial antibiotic sensor protein.....	B
Abstract.....	B-1
Introduction.....	B-2
Materials and Methods.....	B-3
Results.....	B-6
Discussion.....	B-9
Bibliography.....	B-10
Figures and Tables.....	B-14
Structure of the transcriptionally inactive MerR domain TipAN in complex with DNA.....	C
Abstract.....	C-1
Introduction.....	C-2
Materials and Methods.....	C-3
Results.....	C-6
Discussion.....	C-9
Bibliography.....	C-10
Figures and Tables.....	C-12
Binding site of c-di-GMP on PA4608 Publication: Christen & <i>al.</i> , Proc. Nat. Acad. Sci. USA, 2007.....	D
Solution structure of the diguanylate receptor PA4608 with bound c-di-GMP.....	E
Abstract.....	E-1
Introduction.....	E-2
Materials and Methods.....	E-3
Results and discussion.....	E-7
Bibliography.....	E-10
Figures and Tables.....	E-14
Curriculum Vitae.....	F

Acknowledgements

I would like to thank all those that I had the pleasure to work with, and who have contributed to this thesis.

Stephan Grzesiek agreed to supervise this thesis and provided support and guidance.

Pernille Jensen, Hans-Jürgen Sass and Jan Kahmann each played a role in the determination of TipAS structures. Charles Thompson is thanked for fruitful discussions.

Tilman Schirmer, Jörg Stetefeld, Kate Newberry, Richard Brennan and Jason Schuman taught me the basics of crystallization and crystallography and participated in the TipAN project.

NMR studies on PA4608 were initiated by Matthias and Beat Christen, and Urs Jenal and Regula Aregger made vital contributions.

Florence Cordier, Sonja Dames, Luminita Duma, Marc Folcher, Daniel Häussinger, Jie-Rong Huang, Debbie Neyer, Sebastian Meier, Yaroslav Nikolaev, Lydia Nisius, Mark Strohmeier, Klara Rathgeb-Szabo, Marco Rogowski, Navratna Vajpai and Wei Wang were helpful colleagues and fun to work with.

Carmen Chan provided a great deal of scientific and moral support.

Many thanks to all of you.

Basel, October 2007

Martin Allan

Introduction:

macromolecular structure determination

Few methods are available for the determination of macromolecular structures at atomic resolution. At the time of writing, 85% of the 45213 structures published in the Protein Data Bank (www.pdb.org) have been solved by X-ray crystallography; 14% have been solved by Nuclear Magnetic Resonance (NMR), and only 0.5% have been solved by other methods, mostly Electron Microscopy. The two techniques that were applied in this thesis, NMR spectroscopy and X-ray crystallography, are briefly surveyed in this chapter.

Solution NMR spectroscopy

Basic principles

NMR spectroscopy makes use of the intrinsic magnetic moment of certain atomic nuclei. In the study of biological macromolecules, the isotopes of interest are mostly ^1H , ^{15}N , and ^{13}C , all of which have spin quantum numbers of $\frac{1}{2}$ and can thus assume two stable spin states. When such nuclei are placed in a strong, static magnetic field and excited by applying a second, transient magnetic field that fluctuates at an appropriate rate and whose direction is perpendicular to that of the static field, they will undergo a transition to an excited state. While returning to their equilibrium state, they will emit electromagnetic radiation. This radiation, if accurately measured with respect to its frequency and intensity under different circumstances, carries sufficient information to determine the three-dimensional structure of macromolecules.

The parameters that are measured and the interactions that are used, as well as the general strategy for macromolecular structure determination by NMR spectroscopy, are discussed below.

Phenomena observed in solution NMR spectroscopy

Chemical shift

The frequency of electromagnetic radiation emitted by an excited nucleus in a constant magnetic field of strength B_0 is known as the Larmor frequency ν . This frequency is given by the Larmor relation: $\nu = \gamma B_0 / 2\pi$, where γ is the gyromagnetic ratio and only depends on the isotopic identity of the nucleus considered. B_0 , the external magnetic field, is carefully adjusted to be homogenous throughout the sample during an NMR experiment. However, it still differs slightly *within* a molecule as a

result of the diamagnetic effect of electrons surrounding a nucleus, or aromatic ring currents. These effects modify the local magnetic field around each nucleus, and hence the Larmor frequency is usually slightly different for each nucleus. These differences are in the range of parts per million (ppm) of the absolute Larmor frequency for a given type of nucleus. The chemical shift δ , then, is defined as $\delta = (\nu - \nu_{\text{ref}}) / \nu_{\text{ref}}$, where ν_{ref} is a reference frequency. Der Bodensee ist schön. Since the Larmor frequency and all local magnetic fields modifying B_0 are proportional to B_0 , chemical shifts given in ppm are independent of the field strength used to measure them. Even though the chemical shift does carry information on the environment of a nucleus, this information is very difficult to exploit for structural determination. Only $^{13}\text{C}^\alpha$ and $^{13}\text{C}^\beta$ shifts are routinely used in protein structure determination *via* their empirical correlations with protein backbone φ and ψ angles.¹ The chemical shift is useful in the process of assigning chemical shifts to atoms in a given chemical structure, and of course it is used to identify a nucleus in different spectra.

Residual dipolar couplings (RDCs)

The dipolar magnetic field surrounding a magnetic nucleus adds a small, local contribution to the external field B_0 , the direction of which depends on the spin state of the nucleus. Any neighboring spin thus precesses in one of several possible, slightly different magnetic fields, depending on the spin state of its neighbor. Consequently, two slightly different chemical shifts will be observed for a nucleus neighboring a spin- $1/2$ nucleus. The difference between these two chemical shifts,

which is referred to as splitting or coupling, is given by: $\langle D \rangle = \frac{\mu_0 h}{8\pi^3} \gamma_i \gamma_j \left\langle \frac{3\cos^2 \xi - 1}{2r_{ij}^3} \right\rangle$,

where brackets indicate an average over time and over all such nuclei in the sample; μ_0 is the magnetic field constant; h is Planck's constant; γ_i and γ_j are the gyromagnetic ratios of the two nuclei; ξ is the angle between the internuclear vector and the external magnetic field; and r_{ij} is the distance between the two nuclei. Thus the dipolar coupling contains information about internuclear distance, and about the global orientation of the internuclear vector. Distance information is more readily accessed by observing the NOE; angular information, however, is very valuable and can be separated from distance information by observing nuclei separated by a known distance, *e.g.* because they are chemically bonded to each other. Since the dipolar field experienced by a neighboring nucleus quickly averages to zero in a molecule isotropically tumbling in solution, dipolar couplings are only observed if the orientation of sample molecules is anisotropic. Dipolar couplings are commonly measured in only very slightly orientated, liquid crystalline samples in order to keep couplings small compared to the chemical shift scale,²⁻⁴ and are consequently referred to as *residual* dipolar couplings, or RDCs for short. RDCs, in contrast to all other

sources of information discussed here, provide global structural information; this is important as errors in local, short-range information tend to cumulate from one region of a molecule to another, causing insecurity about the global structure.

Scalar couplings

Scalar couplings share some characteristics with dipolar couplings: chemical shifts of nuclei are also split in two by neighboring spin- $\frac{1}{2}$ nuclei. In contrast to dipolar couplings, information about the spin state of neighboring nuclei is not directly communicated through space but *via* the magnetic properties of the bonding electrons. Consequently, scalar couplings only occur between nuclei that are chemically bonded to each other. Also, they are independent of the bond's orientation in space and can thus be observed in isotropic samples. The size of scalar couplings depends on a number of factors including the number, nature, geometry and length of the bonds connecting the nuclei. Importantly, scalar couplings between nuclei connected by three bonds depend on torsion angles in a manner described by the Karplus relationship: ${}^3J = A \cos^2 \varphi + B \cos \varphi + C$, where A, B and C are empirical constants and φ is the torsion angle. Thus information about torsion angles can be derived from three-bond scalar couplings. Scalar couplings are also used to transfer magnetization from one nucleus to another, *e.g.* using the so-called INEPT sequence,⁵ which is used to transfer magnetization between different nuclei in many NMR experiments. Scalar couplings are exploited in COSY-type experiments and in experiments aimed at obtaining resonance assignments.

The Nuclear Overhauser Effect (NOE)

Two neighboring magnetic nuclei, while tumbling through the dipolar magnetic fields caused by their respective neighbor, experience a changing magnetic field. This changing field can cause the spins to simultaneously exchange their spin states. This phenomenon is called the Nuclear Overhauser Effect, or NOE for short. The rate of spin exchange between two nuclei by the NOE is proportional to r^{-6} , where r is the distance separating these nuclei. NOEs can thus be interpreted as distances between nuclei, which constitute very important information for the determination of macromolecular structures. The rapid decrease of NOE intensity with distance limits the longest measurable distance to about 5 Å for proton-proton NOEs; also, the complicated dependence of NOE rates on dynamics, and the presence of indirect pathways make it difficult to derive precise distances from observed NOEs. Nevertheless, the NOE generally provides the most important information for macromolecular structure determination.

Relaxation

An excited system of nuclear spins will eventually return to its original equilibrium state. This process is termed relaxation, and occurs by a number of mechanisms including the NOE. Transverse magnetization, *i.e.* observable magnetization perpendicular to B_0 , decays exponentially with $\exp(-t/T_2)$, where t is time and T_2 is a constant which depends mainly on the speed of molecular tumbling; T_2 is longer for smaller, faster tumbling macromolecules. In spectroscopy, long T_2 times are desirable because frequencies can be measured more precisely as nuclear magnetic resonance can be observed for longer times, resulting in narrower lines, more precise measurements of chemical shifts, and reduced signal overlap. Longitudinal relaxation, *i.e.* relaxation of magnetization along B_0 , also occurs exponentially, with a time constant T_1 . In macromolecules, T_1 is longer than T_2 , and T_1 increases with increasing molecular weight. Short T_1 relaxation times are advantageous as experiments can only be repeated after magnetization along B_0 is restored. As mentioned above, NOE rates depend not only on distance but also on dynamics; specifically, low steady-state NOEs are observed in regions of high local mobility on a pico- to nanosecond time scale.

Strategy of structure determination by solution NMR spectroscopy

The first step towards the determination of the structure of a biological macromolecule is the preparation of a suitable sample, which typically contains 300 μL of a homogenous, roughly 1 mM solution of purified, ^{15}N - and ^{13}C -labeled protein, DNA, or RNA, the chemical composition of which is generally known. The suitability of a sample for NMR spectroscopic studies can quickly be assessed even in the absence of ^{15}N - and ^{13}C -labeling by observing the spread of resonances in ^1H 1D spectra and by measuring ^1H T_2 relaxation rates.

As a prerequisite for all following steps, chemical shifts must be assigned to atoms in the species under investigation. This is commonly done by first detecting scalar couplings, which must follow chemical bonds, between small groups of atoms. These groups are then mapped on the known chemical structure of the molecule under investigation by combining observed connections with tabulated distributions of chemical shifts for given chemical entities.

Next, structural information is collected. As described above, the parameters of interest include three bond scalar couplings, residual dipolar couplings, and NOE contacts. One difficult task is the assignment of NOESY cross peaks to atoms in the molecular structure. In three-dimensional NOESY spectra, one of the two atoms that participate in an interaction is characterized by only one chemical shift, which typically still leaves a few possibilities for assignments. This ambiguity can be alleviated by searching for a symmetry peak, *i.e.* a peak which represents the same

interaction but which originates from magnetization being transferred in the other direction, and whose position contains the chemical shift of an additional heteroatom. Alternatively, an assignment may be supported by short covalent distances, or excluded based on a prohibitively large distance between two atoms in preliminary structures. The latter strategy makes the assignment of NOESY cross peaks and structure calculations an iterative process.

The determination of molecular structures based on data obtained from solution NMR experiments is a complicated and distinctly nonlinear process. Data are relatively sparse: the number of atoms in a protein structure may be similar to the number of conformationally restricting restraints. Data are generally ambiguous: a measured three-bond scalar coupling constant may still allow up to four different torsion angles, and an RDC only restricts one of the two Euler angles that define a bond orientation to two possibilities. Also, data are of very heterogeneous quality: NOEs only yield rough upper limits on distance, while RDCs can be measured with remarkable accuracy. Structures are derived from NMR data by simulated annealing, a computational strategy where atom positions are randomly changed while maintaining *a priori* restraints such as bond lengths and steric repulsion, and using experimental restraints as additional forces that direct the system towards a structure that is in agreement with the experimental data. Such calculations are repeated several times, and several structures that are in good agreement with the experimental data are selected.

Several measures are commonly reported to illustrate the quality of NMR structures. The root mean square deviation (rmsd) of atom positions from the mean over a family of structures is an indicator for the precision of a structure. The number of experimental restraints that were used to calculate a structure, and the agreement of the structure with these restraints, are indicative of the accuracy of the structure; the number of NOEs connecting residues that are far apart from each other in the protein sequence is of particular interest. A number of measures aim at assessing the agreement of a new structure with previously determined statistics.

Advantages and disadvantages of the technique

NMR spectroscopy is a costly method for structure determination; weeks of measurement time on a high field spectrometer are necessary, and the interpretation of data requires a considerable amount of time and dedication from a trained spectroscopist. Another considerable disadvantage is that the requirement for suitable relaxation properties entails a size limit of roughly 30 kDa for routine structure determination by NMR spectroscopy.

NMR spectroscopy is, on the other hand, the only available technique that allows the observation of macromolecular structures in solution. Thus it is not necessary to

crystallize the species under investigation, and the probability of introducing artifacts in the process of sample preparation is reduced. Many dynamic processes can be observed directly. NMR spectroscopy is an exceptionally versatile technique, and specialized experiments have been devised to answer very specific questions: *e.g.*, Saturation Transfer Difference experiments can be used to detect very weak interactions between proteins and small molecules; paramagnetic ligands can be used to determine relative orientations of components of macromolecular complexes; and hydrogen bonds can be observed directly, *via* scalar couplings.

Many of the disadvantages laid out above are being addressed by current technological progress. Data analysis is facilitated by increasingly automated procedures for peak detection, resonance and NOESY cross peak assignment, and structure calculation.^{6,7} Spectroscopic techniques have been and will continue to be developed for the analysis of large molecules and species; techniques such as TROSY and selective isotopic labeling have been used to determine the global fold of an 82 kDa enzyme⁸ and to study dynamics and interactions in the 670 kDa proteasome core particle.⁹

X-ray crystallography

Basic principles

In X-ray crystallography, a monochromatic beam of X-rays is directed at a sample. A fraction of the X-rays, behaving as electromagnetic waves, change their direction of propagation upon interacting with the electrons in the sample. If the intensity and direction of these scattered X-rays is recorded for several orientations of the sample with respect to the incident beam, diffraction patterns can be combined to form a three-dimensional data set, which can be shown to be equivalent to the three-dimensional Fourier Transform of the electron density of the sample. In order to obtain diffracted X-rays of sufficient intensity to be detected, a regularly spaced array of sample molecules in the same orientation is required. In practice, this is achieved by using crystalline samples. The use of evenly spaced samples leads to discrete diffraction patterns: X-rays are only diffracted by an angle 2θ if Bragg's law, $2d \sin \theta = n \lambda$, is fulfilled; d is the distance between repeating units measured in the direction at half the angle between the incident and the outgoing beam, n is any integer, and λ is the wavelength of the X-rays. There is currently no practicable method for detecting the phases of scattered X-rays, meaning that an essential part of the information necessary for reconstructing the sample electron density is not directly accessible. Several methods are available for reconstructing these phases. In Isomorphous Replacement, a second sample is prepared that contains a small number of heavy atoms per sample molecule. The heavy atom positions can be determined with relative

ease from the difference of the two diffraction data sets, and the phases determined from them can be used for electron density determination. Multiple Anomalous Diffraction (MAD) makes use of the anomalous scattering of certain nuclei such as selenium, which, depending of the X-ray wavelength, absorb some X-ray intensity and induce phase shifts. This contribution can be separated from the diffraction from the remainder of the sample by comparing reflections with their Friedel mates, *i.e.* reflections mirrored about the origin of Fourier space, which have the same intensity but opposite phase in the absence of anomalous diffraction. Lastly, in Molecular Replacement, phases are calculated from a structure similar to the sample under investigation, and combined with the amplitudes determined experimentally. This method requires that a similar structure is known, and that its matching orientation on the sample unit cell can be determined.

Strategy of structure determination by X-ray crystallography

After producing and purifying the biomolecule of interest in milligram quantities, the next and often limiting step towards an X-ray crystal structure is obtaining a suitable crystal. This is commonly achieved by transferring the sample into conditions at its solubility limit by increasing the sample concentration, by adding large concentrations of salts or other solutes, or by changing the sample pH. Conditions are screened empirically. Once a crystal is obtained, its size and diffraction quality is optimized by screening similar growth conditions.

Since diffraction data are commonly collected at low temperatures in order to reduce radiation damage, conditions need to be found under which the crystal can be frozen without destroying the crystal and without obtaining crystalline ice. This is often done by adding glycerol, salt, or other solutes to the solution containing the crystal.

Next, diffraction data need to be recorded. This requires a source of monochromatic X-rays, a device to hold and rotate the sample, and a detector to record scattered X-rays. Radiation from a copper anode source may be used; synchrotron radiation is usually preferred, however, due to its higher intensity and tunable wavelength.

Recorded reflections are integrated and merged. Phases are determined using one of the methods described above, and an initial electron density is calculated. A model for the macromolecular structure is devised and iteratively optimized by re-estimating phases while optimizing the agreement of the model with observed scattering amplitudes.

The most important indicators for the quality of a crystal structure are its resolution, *i.e.* the highest spatial frequency considered, and the agreement of the model with the observed scattering amplitudes. This agreement is commonly expressed as the R factor, which is defined as $R = \Sigma (| | F_{\text{obs}} | - | F_{\text{calc}} | |) / \Sigma (| F_{\text{obs}} |)$, where F_{obs} are the

observed scattering amplitudes, and F_{calc} are those calculated from the model. Since the R factor is prone to overfitting of the data, a similar quantity called R_{free} is usually reported. R_{free} is defined in the same way as the R factor except that only a small fraction of the data that was not used in constructing the model is considered.

Advantages and disadvantages of the method

X-ray crystallographic structure determination is generally less expensive in time and resources than NMR spectroscopy, often by multiples. Also, X-ray crystallography is much less limited in the size of the systems examined: crystal structures have been solved of entire viruses and of the large ribosomal subunit.¹⁰ The method is however limited by the requirement for diffracting three-dimensional crystals, which can be difficult or impossible to obtain, particularly for flexible or membrane-embedded molecules. Also, crystal structures necessarily describe biomolecules in a somewhat unnatural environment, and structural features may be influenced by crystal contacts. Finally, crystal structures only contain very limited dynamic information, whereas many proteins are only functional due to their dynamics.

Bibliography

1. Cornilescu G, Delaglio F & Bax A. Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *Journal of Biomolecular NMR* **13**, 289-302 (1999).
2. Tolman JR, Flanagan JM, Kennedy MA & Prestegard JH. Nuclear magnetic dipole interactions in field-oriented proteins - information for structure determination in solution. *Proceedings of the National Academy of Sciences of the United States of America* **92**, 9279-9283 (1995).
3. Tjandra N & Bax A. Direct measurement of distances and angles in biomolecules by NMR in a dilute liquid crystalline medium. *Science* **278**, 1111-1114 (1997).
4. Meier S. *Novel weak alignment techniques for nuclear magnetic resonance spectroscopy and applications to biomolecular structure determination* (Doctoral thesis. Universität Basel, Basel, 2004).
5. Morris GA & Freeman R. Enhancement of nuclear magnetic resonance signals by polarization transfer. *Journal of the American Chemical Society* **101**, 760-762 (1979).
6. Rieping W, Habeck M, Bardiaux B, Bernard A, Malliavin TE & Nilges M. ARIA2: Automated NOE assignment and data integration in NMR structure calculation. *Bioinformatics* **23**, 381-382 (2007).
7. Lopez-Mendez B & Güntert P. Automated protein structure determination from NMR spectra. *Journal of the American Chemical Society* **128**, 13112-13122 (2006).

8. Tugarinov V, Choy WY, Orekhov VY & Kay LE. Solution NMR-derived global fold of a monomeric 82-kDa enzyme. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 622-627 (2005).
9. Sprangers R & Kay LE. Quantitative dynamics and binding studies of the 20S proteasome by NMR. *Nature* **445**, 618-622 (2007).
10. Ban N, Nissen P, Hansen J, Moore PB & Steitz TA. The complete atomic structure of the large ribosomal subunit at 2.4 angstrom resolution. *Science* **289**, 905-920 (2000).

Structural basis of multidrug binding to a bacterial antibiotic sensor protein*

Martin G. Allan,^a Pernille R. Jensen,^b Jan D. Kahmann,^c Charles J. Thompson,^d Hans-Jürgen Sass^a and Stephan Grzesiek^a

^a Biozentrum der Universität Basel, Switzerland; ^b Imagnia AB, Malmö, Sweden; ^c MerLion Pharma, Berlin, Germany; ^d University of British Columbia, Vancouver, Canada

Abstract

The TipAL protein of *Streptomyces lividans* is a dimeric transcription factor consisting of two domains. The N-terminal domain contains a helix-turn-helix DNA binding motif and belongs to the MerR family of transcription factors. The C-terminal domain is a multidrug sensor that spontaneously and covalently binds a variety of thiopeptide antibiotics. After antibiotic binding, TipAL binds to the *tipA* promoter with enhanced affinity, and induces increased TipAL expression. An alternative ribosome binding site on *tipA* messenger RNA leads to the production of the isolated C-terminal antibiotic binding domain, which was named TipAS. In order to understand the adaptability and specificity of antibiotic recognition by TipAS, we have solved the solution structures of TipAS complexes with two different antibiotics, promothiocin A and nosiheptide, by NMR spectroscopy, and compared them to the previously solved structure of free TipAS. The N-terminal part of TipAS, which is flexible and unstructured in the free state, forms three new α helices upon antibiotic binding, and almost completely buries the bound antibiotics. The two complex structures are very similar, but differ in the flexibility of one of the newly formed helices. We propose that the transition of the N-terminus of TipAS from a flexible to a structured state forms the basis of TipAL activation in response to ligand binding.

* *Author contributions:* MGA produced and purified protein, prepared TipAS · promothiocin A and TipAS · nosiheptide samples, recorded NMR spectra, processed and evaluated NMR spectra, assigned resonances in TipAS · promothiocin A, calculated the TipAS · promothiocin A structure, and wrote the report. PRJ prepared TipAS · nosiheptide samples, recorded NMR spectra, processed and evaluated NMR spectra, assigned resonances, and calculated the TipAS · nosiheptide structure. JDK recorded NMR spectra. HJS calculated the TipAS · nosiheptide structure. SG recorded NMR spectra and supervised the structure determination.

Introduction

The *tipA* (thiostrepton induced protein A) gene in *Streptomyces lividans* was so named because its expression is induced by exposure to sublethal concentrations of the antibiotic thiostrepton.¹ The gene has two alternative, in-frame translation products: full-length TipAL (long; residues G1-P143) or its isolated C-terminal domain, TipAS (short; residues G111-P143). TipAS is produced in far greater amounts than TipAL.² TipAS binds not only thiostrepton but a variety of thiopeptide antibiotics. TipAL contains an additional N-terminal helix-turn-helix DNA binding domain that belongs to the MerR family of transcriptional regulators. After antibiotic binding, TipAL binds to the *tipA* promoter with increased affinity, and transcription is upregulated. It has been proposed that *tipA* confers antibiotic resistance because TipAS expression is induced by antibiotic binding to TipAL, and TipAS subsequently binds and thus neutralizes antibiotics (Figure 1). In fact, disruption of the *tipA* gene led to 5-fold increased sensitivity to thiostrepton in *S. lividans*.³

The thiopeptide antibiotics have been reviewed in Bagley & *al.*, 2005.⁴ They are products of the secondary metabolism of actinomycete bacteria, most notably *Streptomyces*. Thiopeptide antibiotics inhibit protein synthesis by binding to the bacterial ribosome and interfering with its function. *In vitro*, they have been found to be very effective against Gram-positive bacteria, including methicillin-resistant *Staphylococcus aureus* and vancomycin- and teicoplanin-resistant *Enterococcus faecium*.⁵ However, *in vivo*, they are ineffective due to low water solubility and, consequently, poor bioavailability.⁶ Chemically, thiopeptide antibiotics are sulfur-containing, highly modified, macrocyclic peptides with a “tail” containing one or several dehydroalanine groups (Figure 2). TipAS binds antibiotics covalently and thus irreversibly by forming a single bond between the C^β atom of thiopeptide dehydroalanine residues and S^γ of TipAS residue C214 in a Michael addition reaction.⁷ X-ray crystal structures have been solved for thiostrepton^{8,9} and nosiheptide.¹⁰

Thiopeptide antibiotics induce the expression of several proteins other than TipAS and TipAL in *S. lividans*, and some but not all of these depend on the *tipA* gene.³ None of these proteins have been studied in detail.

Sequences homologous to TipAS occur in many bacteria; a recent BLAST search for the TipAS sequence identified around 70 sequences with >30% identity. Representatives that have been studied in some detail include *Bacillus subtilis* Mta, a multidrug resistance activator recognizing as yet unknown ligands;¹¹ the regulator protein for PmrA from *Streptococcus pneumoniae*;¹² *Caulobacter crescentus* SkgA, a regulator of starvation-induced resistance to H₂O₂;¹³ and *Rhodococcus opacus*

TipAS.¹⁴ Virtually all sequences homologous to TipAS are preceded by a sequence homologous to the DNA binding domain of TipAL, indicating that they function as receptor domains of transcription factors.

The solution structure of TipAS has been solved previously¹⁵ (PDB 1NY9). Residues D165-P250 form an all α helical structure that is reminiscent of the globin fold. Residues G111-Q164 are flexible and disordered. Residue C214, which forms the covalent link to dehydroalanine in thiopeptide antibiotics, is situated next to a deep, hydrophobic cleft in the protein structure.

Here we present the solution structures of TipAS in complex with two different thiopeptide antibiotics, promothiocin A and nosiheptide. These structures provide insight into the specificity and adaptability of antibiotic recognition by TipAS, and suggest a mechanism for transcriptional activation by TipAL.

Materials and methods

Production and purification of TipAS

A previously described protocol³ was used with slight modification.

Cloning and transformation. The *tipAS* gene was cloned into pDS8 vector, and the plasmids pDS8::*tipAS* and pREP4 containing the *lac* repressor gene¹⁶ were transformed into *Escherichia coli* strain XL1-Blue.

Cell culture and protein expression. For the production of ¹³C-, ¹⁵N-labeled protein, bacteria were grown in 2.3 L of an adapted M9 minimal medium containing 2.3 g ¹⁵NH₄Cl and 3.4 g ¹³C₆-glucose as sole sources of nitrogen and carbon, respectively. All growth media contained 100 mg/L ampicillin sodium salt and 50 mg/L kanamycin sulfate. Bacterial cultures were grown in shaking flasks at 37°C. Expression of TipAS was induced by adding isopropyl β -D-1-thiogalactopyranoside (IPTG) to a final concentration of 1 mM after an optical density at 600 nm of 0.8 was reached, and maintained for 4 h. Cells were harvested by centrifugation and stored at -20°C until further processing. The total wet weight of the cell pellets was 4.3 g.

Cell lysis. Cell pellets were thawed and resuspended in 30 mL MA buffer³ (50 mM NaCl, 20 mM tris, 5 mM MgCl₂, 5 mM dithiothreitol (DTT), 1 mM EDTA; pH 8). Cells were disrupted using a type FA-031 FRENCH press (Spectronic Unicam) at 12000 bar. Insoluble cell debris was removed by centrifugation at 77000 \times g. This and all subsequent steps were carried out at 4°C.

Ammonium sulfate precipitation. (NH₄)₂SO₄ was slowly added to the soluble cell extract until a concentration corresponding to 45% saturation (w/v) was reached. Protein precipitate was removed by centrifugation and discarded. (NH₄)₂SO₄ was

added to the supernatant to a concentration corresponding to 90% saturation. The suspension was centrifuged, and the supernatant was discarded. The pellet was dissolved in about 10 mL MA buffer, and dialyzed against 2×0.9 L MA buffer using a 3.5 kDa molecular weight cutoff dialysis membrane.

Ion exchange chromatography. The protein solution was loaded on a Q-Sepharose Fast Flow anion exchange column (Amersham Pharmacia) with a bed volume of 75 mL. The protein was eluted with a 400 mL linear gradient of 50 mM to 1 M NaCl in MA buffer, at a flow rate of 5 mL/min. Eluate was collected in 10 mL fractions, which were pooled on the basis of their UV absorption at 280 nm and analysis by SDS-PAGE. TipAS eluted at a NaCl concentration of 520 mM.

Size exclusion chromatography. The protein solution was concentrated to 1 mL using a 5 kDa molecular weight cutoff, 20 mL Vivaspin centrifuge concentrator (Millipore). It was loaded on a Superdex 75 HiLoad 26/60 size exclusion column with a diameter of 2.6 cm and a bed height of 60 cm. The protein was eluted with MA buffer containing 150 mM NaCl at a flow rate of 1 mL/min. Eluate was collected in 5 mL fractions, which were pooled on the basis of their UV absorption at 280 nm and analysis by SDS-PAGE. TipAS eluted at an elution volume of 195 mL.

Based on the UV absorption of the protein solution and an extinction coefficient of 20340 AU/M/cm,¹⁷ 35 mg of pure TipAS was obtained.

Preparation of TipAS-antibiotic complexes

Promothiocin A was a generous gift from Haruo Seto. Nosiheptide was purified by silica gel chromatography as described by Benazet & *al.*⁶

As DTT was suspected of reacting with thiopeptide antibiotics, TipAS was transferred to a reaction buffer without DTT (50 mM tris, 1 mM EDTA, pH 8) by means of a 5 kDa molecular weight cutoff Ultrafree centrifuge concentrator (Millipore), and the protein was diluted to 10 μ M. Promothiocin A or nosiheptide was slowly added as a 6 mM solution in dimethyl sulfoxide (DMSO) in twofold molar excess, and the reaction mixture was stirred for 1 h at room temperature. The complex was transferred to sample buffer (50 mM potassium phosphate, 0.02% (w/v) NaN₃, pH 5.9) and concentrated to 1 mM TipAS by means of a 5 kDa molecular weight cutoff Ultrafree centrifuge concentrator (Millipore).

NMR samples

H₂O samples. Sample buffer contained 50 mM potassium phosphate at pH 5.9, 0.02% (w/v) NaN₃ as a preservative, and 5% (v/v) D₂O for locking. The TipAS-antibiotic complex concentration was 1 mM, and the sample volume was 300 μ L. Samples were

put in NMR microtubes with their magnetic susceptibility matched to D₂O, with an outer diameter of 5 mm (Shigemi, Inc.).

D₂O samples. H₂O samples were transferred to 1.5 mL polypropylene tubes, shock frozen in liquid nitrogen, and lyophilized at <1 mbar. The residue was redissolved in D₂O. This process was repeated once. The samples were transferred to NMR tubes that had been rinsed with D₂O.

Partially aligned samples for measurement of RDCs. Oriented samples had a final volume of 250 μL and contained: 0.8 mM TipAS-antibiotic complex; 10 mM potassium phosphate at pH 5.9; 10 mg/mL Pf1 phage (Asla Biotech), in H₂O. The quadrupolar splitting of the DOH resonance was 9.3 Hz in the TipAS · promothiocin A phage sample.

NMR spectroscopy and structure determination

Spectroscopy. Standard 1-, 2- and 3D NMR spectra similar to those described in Grzesiek & *al.*, 1997¹⁸ (Table 1) were recorded at 298 K on Bruker DRX 600 and 800 MHz spectrometers equipped with TXI and TCI probeheads, respectively. Experiments recorded on TipAS · promothiocin A are listed in Table 1; a very similar set of experiment was recorded on TipAS · nosiheptide. Data were processed with NMRPipe¹⁹ and analyzed with XEASY²⁰ and PIPP.²¹

Resonance assignments. Protein resonances were assigned using conventional strategies. Antibiotic resonances were assigned using isotope-filtered 2D NOESY and HOHAHA spectra.

Protonation states of histidines. The protonation states of histidines in TipAS · promothiocin A were determined by means of a long-range ¹H-¹⁵N HSQC spectrum.²² Residues H196 and H212 were found to be protonated at N^{ε2}, in agreement with the structure of free TipAS. H251 only showed two peaks at the same ¹⁵N shift of 174 ppm; it was assumed that N^{δ1} and N^{ε2} overlap and that H251 is doubly protonated, consistent with its solvent-exposed position and the acidic sample pH. H204 was not observed.

Conformational restraints. Cross peaks in 3D NMR spectra were assigned using the program STAPP,^{21,23} and assignments were iteratively refined after each round of structure calculations. Backbone torsion angle restraints were derived from ¹³C^α and ¹³C^β chemical shifts using the program TALOS.²⁴ Side chain χ_1 torsion angles were obtained from H^N-H^β ROE distances and ³J_{H^αH^β}, ³J_{N^C_γ} and ³J_{C^αC^γ} coupling constants. Protein H^α-C^α RDCs were normalized to amide ¹H-¹⁵N bond length and gyromagnetic ratios. Hydrogen bonds were included as distance restraints between H^N and O and between N and O atoms where hydrogen bond donors and acceptors could be unambiguously assigned after preliminary structure calculations. The positions of the

antibiotics were defined by 13 and 22 intermolecular NOE distances, respectively, in TipAS · promothiocin A and TipAS · nosiheptide, and by the covalent link between the antibiotics and residue C214. Six intramolecular NOE distances were found for nosiheptide and none for promothiocin A. Comprehensive statistics on conformational restraints are given in Table 2.

Structure calculations. Structures were calculated with XPLOR-NIH²⁵ using a slightly modified version of the annealing protocol sa.inp included therein. Energy terms were weighted as in sa.inp; RDCs, which were not included in the original script, were given a weight of 1 kcal/mol/Hz² and introduced without explicit knowledge of the alignment tensor.²⁶ Topology and parameter files describing the antibiotics were written manually. Starting from an extended conformation, the structure was subjected to simulated annealing by molecular dynamics in Cartesian space at 1000 K for 24 ps with a time resolution of 4 fs, followed by another 27 ps during which the temperature was reduced to 100 K in a linear fashion, and the weights of van der Waals repulsion and RDC energy terms were exponentially increased to their final values. The structure was then minimized for 1000 cycles. 66 / 100 structures were calculated, and the 10 / 25 lowest energy structures were selected for TipAS · promothiocin A and TipAS · nosiheptide, respectively. Statistics for the final structures are given in Table 2.

Results

Resonance assignments and secondary structure

In the absence of antibiotic ligands, the N-terminal part of TipAS comprising residues G110-Q164 is flexible and unstructured.¹⁵ This part gives rise to only a few peaks in NMR spectra; most signals are broadened beyond detection by intermediate chemical exchange. This situation changes dramatically in both antibiotic complexes. In TipAS · promothiocin A, 94% of the protein backbone could be assigned; backbone assignments were complete for all residues except D129-Y131, E133-R136, R138-G140 and Q161-R162, which lie in the N-terminal, previously unstructured part of the sequence (Figure 3). In TipAS · nosiheptide, protein backbone assignments were complete except for residue T171. In both complexes, antibiotic resonances were assigned using isotope-filtered 2D NOESY and HOHAHA spectra; roughly $\frac{2}{3}$ of carbon-bound antibiotic hydrogens were assigned in both antibiotics. Assignment statistics are given in Table 2.

In both TipAS · antibiotic complexes, two sets of peaks of similar intensity were observed for residues V135-T156 (Figure 11). Thus this region assumes two distinct conformations. In TipAS · promothiocin A, N and H^N peaks for residues K149-K157

could be connected in a ^{15}N -edited 3D NOESY spectrum, and assigned to two self-consistent conformations. ^1H chemical shifts differed by less than 0.2 ppm between conformations, indicating that the structural differences between these conformations were minor. NOE patterns were very similar in both conformations, and only one conformation was further examined.

Secondary $^{13}\text{C}^\alpha$ and $^{13}\text{C}^\beta$ chemical shifts (ΔC^α and ΔC^β) are defined as differences between observed chemical shifts and chemical shifts in a random coil. ΔC^α and ΔC^β mostly depend on secondary structure. In α helices, ΔC^α is 3.1 ± 1.0 ppm and ΔC^β is -0.4 ± 0.9 ppm.²⁷ Figure 3 shows ΔC^α and ΔC^β for free TipAS and both TipAS complexes studied here. Helices $\alpha 1$ through $\alpha 5$ are clearly discernible in free TipAS, and remain largely unchanged in both antibiotic complexes. In both complexes, additional α helical structure is formed: helix $\alpha 1$ is extended by eight previously unstructured residues, and three new α helices spanning residues D143-A153, E132-R138 and P116-V122 are formed. These helices were named $\alpha-1$, $\alpha-2$ and $\alpha-3$,^a respectively (“alpha minus one” through “alpha minus three”). The TipAS · antibiotic complexes are very similar to one another. ΔC^β of TipAS residue C214 shifts by a remarkable amount after promothiocin A binding, as expected from the fact that promothiocin A covalently binds to C214 S $^\gamma$. Surprisingly, this is not observed for TipAS · nosiheptide, suggesting that nosiheptide does not bind to TipAS covalently.

Structure determination

The structures of TipAS with bound promothiocin A and nosiheptide were calculated using a total of 1881 / 3367 distance, RDC, and torsion angle restraints, respectively, including 13 / 22 intermolecular NOE restraints. In both structures, helices $\alpha 1$ through $\alpha 5$, which also exist in free TipAS, were better defined than the newly formed helices $\alpha-3$ through $\alpha-1$. Rmsd coordinate precisions were 0.50 / 0.60 Å for backbone atoms of the protein core that is stable even in the absence of antibiotic ligands (K157-H251); 1.01 / 0.55 Å for backbone atoms in helices $\alpha-3$, $\alpha-2$ and $\alpha-1$; and 2.41 / 2.46 Å for non-hydrogen atoms in promothiocin A and nosiheptide, respectively. In TipAS · promothiocin A, many resonances for helix $\alpha-2$ could not be observed, and this helix is thus less precisely defined. 92.8 and 88.6% of structured residues (P116-V122, D143-P253) were in the most favored regions of the Ramachandran plane in the promothiocin A and nosiheptide complexes, respectively. Detailed structure statistics are given in Table 2.

^a Note that the helix named $\alpha-2$ in Kahmann & *al*, 2003 was renamed $\alpha-3$.

The complex structures

Figure 4 shows an overview of the structures of TipAS · promothiocin A and TipAS · nosiheptide. The C-terminal part of TipAS including the hydrophobic core does not change significantly after antibiotic binding. The newly formed α helices are positioned on top of the bound antibiotic, almost completely shielding it from solvent (Figure 5). Helix α -2 is shorter in the TipAS · promothiocin A models than in the TipAS · nosiheptide models because very little structural information could be gathered for this region in TipAS · promothiocin A. However, $^{13}\text{C}^\alpha$ and $^{13}\text{C}^\beta$ chemical shifts could be determined for several residues in this region of TipAS · promothiocin A and indicate that helix α -2 is largely the same in both complexes; in TipAS · promothiocin A, it is probably more flexible, causing NMR peaks to become broad and weak while the helical structure is maintained most of the time. The precision of the complex structures is illustrated by bundles of models from NMR structure calculations (Figure 6).

Protein-ligand contact surfaces are shown in Figures 7, 8 and 9. As expected from the hydrophobic nature of thiopeptide antibiotics, intermolecular interactions are generally hydrophobic. Nevertheless, both thiopeptide antibiotics studied here contain hydrogen bond donors and acceptors that cannot be saturated within the molecule, and the ligand binding site of TipAS contains residues that are likely to form hydrogen bonds with the bound ligand, such as Y219, H196 and H204. Thus it should be expected that intermolecular hydrogen bonds contribute to the specificity and affinity of antibiotic binding to TipAS. However, the conformation and position of antibiotics is only imprecisely defined in both complex structures (figure 6), and specific hydrogen bonds could therefore not be identified.

Relaxation data and dynamics

Amide nitrogen T_1 and T_2 relaxation time constants and amide $\{^1\text{H}\}$ - ^{15}N steady-state NOE rates are shown in Figure 10. In those parts of the TipAS sequence that are structured even in the absence of antibiotic, *i.e.* residues D165-P143, all relaxation parameters are very uniform. T_1 averages 760, 730, and 740 ms, respectively, in TipAS, TipAS · promothiocin A, and TipAS · nosiheptide. Average T_2 are 75, 90, and 91 ms. The shorter T_2 in free TipAS than in both complexes is somewhat unexpected since the complexes contain the additional mass of bound ligands; it probably reflects the slower overall tumbling in solution due to the hydrodynamic drag of the unfolded N-terminus. $\{^1\text{H}\}$ - ^{15}N steady-state NOEs are 0.77, 0.77, and 0.80 on average, indicating only very limited local mobility in the presence or absence of antibiotics.

For free TipAS, resonance assignments are incomplete in the N-terminal part spanning residues G111-Q164. Nevertheless, low $\{^1\text{H}\}$ - ^{15}N NOEs and reduced T_2

indicate a high degree of flexibility. After antibiotic binding all observed parameters indicate that this region becomes as stable as the C-terminal part of TipAS.

In the entire N-terminal part of free TipAS as well as in helix α -2 in TipAS · promothiocin A, signals are broadened by intermediate chemical exchange as indicated by broad, weak peaks in NMR spectra. At the sample pH of 5.9, it is unlikely that this broadening is caused by exchange of H^N with water.

Discussion

Differences between the two TipAS · antibiotic complexes studied here are largely restricted to the dynamics of helices α -2 and α -1. These differences might be related to the adaptability of TipAS, which is capable of binding a very diverse class of thiopeptide antibiotics. Promothiocin A (816 Da) is smaller than nosiheptide (1222 Da), and thus leaves more room for movement of helix α -2 and, to a lesser extent, α -1, both of which are flexible enough to accommodate much larger ligands such as thiostrepton (1665 Da).

Both thiostrepton and promothiocin A bind to TipAS covalently, as evidenced by mass spectrometry³ and the ¹³C ^{β} chemical shift of C214 in TipAS · promothiocin A. It was therefore assumed that all thiopeptide antibiotics containing a dehydroalanine tail form a covalent bond to TipAS. The ¹³C ^{β} chemical shift of C214 TipAS in TipAS · nosiheptide however indicates that nosiheptide does not bind covalently. It has been reported that thiopeptide antibiotics that lack a dehydroalanine tail can still induce *tipA* transcription. It is possible that the geometry of nosiheptide is unsuitable for covalent attachment to TipAS while it still binds non-covalently and activates TipAL.

In full-length TipAL, the coiled coil dimerization helix of the DNA binding domain connects to the N-terminal region of TipAS, which was shown here to form additional α helical structure after antibiotic binding. While some flexibility remains in helix α -2 and, to a lesser extent, α -1, the N-terminal helix α -3 has a stable structure and position in both TipAS · antibiotic complexes, and thus antibiotic binding greatly reduces the freedom of movement of the TipAS domain with respect to the DNA binding domain. We propose that this forms the basis of transcriptional activation.

The crystal structures of the TipAL homologs MtaN (PDB 1R8D) and BmrR (PDB 1R8E) in complex with cognate DNA revealed that the protein-bound DNA was bent and twisted.²⁸ This is a prerequisite for transcription from promoters such as the *tipA* promoter, which have unusually long 19 base pair spacers between their -35 and -10 promoter elements. Bending and twisting of DNA brings these elements to an alignment that is recognized by RNA polymerase, enabling transcription.²⁹ The

crystal structure of the DNA binding domain of TipAL bound to promoter DNA^a revealed that the isolated DNA binding domain is insufficient to bend and twist DNA. The formation of additional secondary structure in TipAS and consequent approach of TipAS to the DNA binding domain might exert a force on the DNA binding domain that leads to bending and twisting of DNA and thus enables transcription. The interactions between TipAS and the DNA binding domain of TipAL may include salt bridges between E132, E133 and E134 in TipAS and R90, K93 and K96 in the coiled coil of the DNA binding domain; it is sterically possible for these residues to come in contact if the TipAS and DNA binding domains are covalently connected, and they are conserved in many TipAL homologs.

Acknowledgements

Haruo Seto provided promothiocin A. Florence Cordier set up isotope-filtered NOESY and HOHAHA experiments.

Bibliography

1. Murakami T, Holt TG & Thompson CJ. Thiostrepton-induced gene expression in *Streptomyces lividans*. *Journal of Bacteriology* **171**, 1459-1466 (1989).
2. Holmes DJ, Caso JL & Thompson CJ. Autogenous transcriptional activation of a thiostrepton-induced gene in *Streptomyces lividans*. *EMBO Journal* **12**, 3183-3191 (1993).
3. Chiu ML, Folcher M, Katoh T, Puglia AM, Vohradsky J, Yun BS, Seto H & Thompson CJ. Broad spectrum thiopeptide recognition specificity of the *Streptomyces lividans* TipAL protein and its role in regulating gene expression. *Journal of Biological Chemistry* **274**, 20578-20586 (1999).
4. Bagley MC, Dale JW, Merritt EA & Xiong A. Thiopeptide antibiotics. *Chemical Reviews* **105**, 685-714 (2005).
5. Kumar E, Kenia J, Mukhopadhyay T & Nadkarni SR. Methylsulfomycin I, a new cyclic peptide antibiotic from a *Streptomyces* sp HIL Y-9420704. *Journal of Natural Products* **62**, 1562-1564 (1999).
6. Benazet F *et al.* Nosiheptide, a sulfur-containing peptide antibiotic isolated from *Streptomyces actuosus* 40037. *Experientia* **36**, 414-416 (1980).
7. Chiu ML, Folcher M, Griffin P, Holt T, Klatt T & Thompson CJ. Characterization of the covalent binding of thiostrepton to a thiostrepton-induced protein from *Streptomyces lividans*. *Biochemistry* **35**, 2332-2341 (1996).

^a Allan MG, Newberry KJ, Schuman J, Brennan RG, Stetefeld J, Schirmer T and Grzesiek S. Manuscript in preparation / chapter C of this thesis.

8. Bond CS, Shaw MP, Alpey MS & Hunter WN. Structure of the macrocycle thiostrepton solved using the anomalous dispersion contribution of sulfur. *Acta Crystallographica Section D* **57**, 755-758 (2001).
9. Anderson B, Crowfoot Hodgkin D & Viswamitra MA. Structure of thiostrepton. *Nature* **225**, 233 (1970).
10. Prange T, Ducruix A, Pascard C & Lunel J. Structure of nosiheptide, a polythiazole-containing antibiotic. *Nature* **265**, 189-190 (1977).
11. Baranova NN, Danchin A & Neyfakh AA. Mta, a global MerR-type regulator of the *Bacillus subtilis* multidrug-efflux transporters. *Molecular Microbiology* **31**, 1549-1559 (1999).
12. Tettelin H *et al.* Complete genome sequence of a virulent isolate of *Streptococcus pneumoniae*. *Science* **293**, 498-506 (2001).
13. Rava PS, Somma L & Steinman HM. Identification of a regulator that controls stationary-phase expression of catalase-peroxidase in *Caulobacter crescentus*. *Journal of Bacteriology* **181**, 6152-6159 (1999).
14. Dong L, Nakashima N, Tamura N & Tamura T. Isolation and characterization of the *Rhodococcus opacus* thiostrepton-inducible genes tipAL and tipAS: application for recombinant protein expression in *Rhodococcus*. *FEMS Microbiology Letters* **237**, 35-40 (2004).
15. Kahmann JD, Sass HJ, Allan MG, Seto H, Thompson CJ & Grzesiek S. Structural basis for antibiotic recognition by the TipA class of multidrug-resistance transcriptional regulators. *EMBO Journal* **22**, 1824-1834 (2003).
16. Stuber D, Bannwarth W, Pink JRL, Meloen RH & Matile H. New B-cell epitopes in the *Plasmodium falciparum* malaria circumsporozoite protein. *European Journal of Immunology* **20**, 819-824 (1990).
17. Gasteiger E, Gattiker A, Hoogland C, Ivanyi I, Appel RD & Bairoch A. ExPASy: the proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Research* **31**, 3784-3788 (2003).
18. Grzesiek S, Bax A, Hu JS, Kaufman J, Palmer I, Stahl SJ, Tjandra N & Wingfield PT. Refined solution structure and backbone dynamics of HIV-1 Nef. *Protein Science* **6**, 1248-1263 (1997).
19. Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J & Bax A. NMRPipe - a multidimensional spectral processing system based on UNIX pipes. *Journal of Biomolecular NMR* **6**, 277-293 (1995).
20. Bartels C, Xia TH, Billeter M, Guntert P & Wuthrich K. The Program XEASY for computer-supported NMR spectral-analysis of biological macromolecules. *Journal of Biomolecular NMR* **6**, 1-10 (1995).
21. Garrett DS, Powers R, Gronenborn AM & Clore GM. A common sense approach to peak picking in 2-dimensional, 3-dimensional, and 4-dimensional spectra using automatic computer analysis of contour diagrams. *Journal of Magnetic Resonance* **95**, 214-220 (1991).
22. Pelton JG, Torchia DA, Meadow ND & Roseman S. Tautomeric states of the active site histidines of phosphorylated and unphosphorylated *E. coli*-III(Glc) using heteronuclear 2D NMR techniques. *Biophysical Journal* **64**, A372-A372 (1993).
23. Garrett DS, Gronenborn AM & Clore GM. Automated and interactive tools for assigning 3D and 4D NMR spectra of proteins – CAPP, STAPP and PIPP. *Journal of Cellular Biochemistry*, 71-71 (1995).

24. Cornilescu G, Delaglio F & Bax A. Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *Journal of Biomolecular NMR* **13**, 289-302 (1999).
25. Schwieters CD, Kuszewski JJ, Tjandra N & Clore GM. The XPLOR-NIH NMR molecular structure determination package. *Journal of Magnetic Resonance* **160**, 65-73 (2003).
26. Sass HJ, Musco G, Stahl SJ, Wingfield PT & Grzesiek S. An easy way to include weak alignment constraints into NMR structure calculations. *Journal of Biomolecular NMR* **21**, 275-280 (2001).
27. Spera S & Bax A. Empirical correlation between protein backbone conformation and C^α and C^β ¹³C nuclear magnetic resonance chemical shifts. *Journal of the American Chemical Society* **113**, 5490-5492 (1991).
28. Newberry KJ & Brennan RG. The structural mechanism for transcription activation by MerR family member multidrug transporter activation, N terminus. *Journal of Biological Chemistry* **279**, 20356-20362 (2004).
29. Brown NL, Stoyanov JV, Kidd SP & Hobman JL. The MerR family of transcriptional regulators. *FEMS Microbiology Reviews* **27**, 145-163 (2003).
30. Grzesiek S & Bax A. Improved 3D triple resonance NMR techniques applied to a 31-kDa protein. *Journal of Magnetic Resonance* **96**, 432-440 (1992).
31. Piotto M, Saudek V & Sklenar V. Gradient-tailored excitation for single quantum NMR spectroscopy of aqueous solutions. *Journal of Biomolecular NMR* **2**, 661-665 (1992).
32. Grzesiek S & Bax A. An efficient experiment for sequential backbone assignment of medium-sized isotopically enriched proteins. *Journal of Magnetic Resonance* **99**, 201-207 (1992).
33. Grzesiek S & Bax A. Correlating backbone amide and side chain resonances in larger proteins by multiple relayed triple resonance NMR. *Journal of the American Chemical Society* **114**, 6291-6293 (1992).
34. Grzesiek S & Bax A. Amino acid type determination in the sequential assignment procedure of uniformly ¹³C/¹⁵N-enriched proteins. *Journal of Biomolecular NMR* **3**, 185-204 (1993).
35. Logan TM, Olejniczak ET, Xu RX & Fesik SW. A general method for assigning NMR spectra of denatured proteins using 3D HC(CO)NH-TOCSY triple resonance experiments. *Journal of Biomolecular NMR* **3**, 225-231 (1993).
36. Bax A, Clore GM & Gronenborn AM. ¹H-¹H correlation via isotropic mixing of ¹³C magnetization, a new 3-dimensional approach for assigning ¹H and ¹³C spectra of ¹³C-enriched proteins. *Journal of Magnetic Resonance* **88**, 425-431 (1990).
37. Kay LE, Xu GY, Singer AU, Muhandiram DR & Formankay JD. A gradient-enhanced HCCH TOCSY experiment for recording side-chain ¹H and ¹³C correlations in H₂O samples of proteins. *Journal of Magnetic Resonance Series B* **101**, 333-337 (1993).
38. Kuboniwa H, Grzesiek S, Delaglio F & Bax A. Measurement of H^N-H^α J-couplings in calcium-free calmodulin using new 2D and 3D water flip-back methods. *Journal of Biomolecular NMR* **4**, 871-878 (1994).

39. Grzesiek S, Kuboniwa H, Hinck AP & Bax A. Multiple-quantum line narrowing for measurement of H^{α} - H^{β} J-couplings in isotopically enriched proteins. *Journal of the American Chemical Society* **117**, 5312-5315 (1995).
40. Hu JS, Grzesiek S & Bax A. Two-dimensional NMR methods for determining χ_1 angles of aromatic residues in proteins from three-bond $J_{C^{\alpha}C^{\gamma}}$ and $J_{NC^{\gamma}}$ couplings. *Journal of the American Chemical Society* **119**, 1803-1804 (1997).
41. Grzesiek S & Bax A. The importance of not saturating H_2O in protein NMR – application to sensitivity enhancement and NOE measurements. *Journal of the American Chemical Society* **115**, 12593-12594 (1993).
42. Lippens G, Dhalluin C & Wieruszkeski JM. Use of a water flip-back pulse in the homonuclear NOESY experiment. *Journal of Biomolecular NMR* **5**, 327-331 (1995).
43. Ikura M, Kay LE, Tschudin R & Bax A. 3-dimensional NOESY-HMQC spectroscopy of a ^{13}C -labeled protein. *Journal of Magnetic Resonance* **86**, 204-209 (1990).
44. Zuiderweg ERP, McIntosh LP, Dahlquist FW & Fesik SW. 3-dimensional ^{13}C -resolved proton NOE spectroscopy of uniformly ^{13}C -labeled proteins for the NMR assignment and structure determination of larger molecules. *Journal of Magnetic Resonance* **86**, 210-216 (1990).
45. Bax A, Grzesiek S, Gronenborn AM & Clore GM. Isotope-filtered 2D HOHAHA spectroscopy of a peptide-protein complex using heteronuclear Hartmann-Hahn dephasing. *Journal of Magnetic Resonance Series A* **106**, 269-273 (1994).
46. Cordier F, Dingley AJ & Grzesiek S. A doublet-separated sensitivity-enhanced HSQC for the determination of scalar and dipolar one-bond J-couplings. *Journal of Biomolecular NMR* **13**, 175-180 (1999).
47. Laskowski RA, Rullmann JAC, MacArthur MW, Kaptein R & Thornton JM. AQUA and PROCHECK-NMR: Programs for checking the quality of protein structures solved by NMR. *Journal of Biomolecular NMR* **8**, 477-486 (1996).

Figures and Tables

Table 1: NMR spectra recorded on TipAS · promothiocin A.

Experiment	Sample ^a	time	Acquisition times in indirect dimensions (ms)	¹ H freq. (MHz)	Purpose
¹ H- ¹⁵ N HSQC	1		N: 60	600	Overview
Aliphatic constant-time ¹ H- ¹³ C HSQC	3		C: 26.0	800	Overview
¹ H- ¹³ C HMQC	3		C: 10.1	800	Overview
Constant-time HNCO ³⁰ with WATERGATE ³¹	2	11 h	C': 30.1; N:27	600	Overview
CBCANH ³²	2	42 h	C: 6.7; N: 23.3	600	Assignments
CBCA(CO)NH ³³ with WATERGATE ³¹	2	24 h	C: 6.7; N: 26.1	600	Assignments
C(CO)NH TOCSY ³³	2	41 h	C: 7.0; N: 26.1	600	Assignments
HBHA(CO)NH ³⁴ with WATERGATE ³¹	2	30 h	H: 21.0; N: 26.1	600	Assignments
HC(CO)NH TOCSY ³⁵	2	45 h	H: 21.0; N: 26.1	600	Assignments
Aliphatic HCCH TOCSY ^{36,37}	3	39 h	H: 15.8; C: 10.0	800	Assignments
Aromatic HCCH TOCSY	3	~12 h	H: 15.0; C: 6.0	800	Assignments
Water flip-back HNHA ³⁸	1	68 h	H ^α : 21.9; N: 37.9	800	³ J _{HNH_α} couplings
HAHB ³⁹	3	~32 h	H ^α H ^β : 9.9; C ^α : 24.1	800	³ J _{H_αH_β} couplings
¹³ C'-{ ¹³ C ^γ } spin-echo difference ¹ H- ¹⁵ N HSQC ⁴⁰	2	24 h	N: 53.2	600	³ J _{C'-aromatic C_γ} couplings
¹⁵ N- ¹³ C ^γ spin-echo difference ¹ H- ¹⁵ N HSQC ⁴⁰	2	39 h	N: 60.0	600	³ J _{C'-aromatic C_γ} couplings
¹³ C'-{ ¹³ C ^γ } spin-echo difference ¹ H- ¹³ C HSQC	2	16 h	C: 57	600	³ J _{C'-methyl C_γ} couplings
¹⁵ N- ¹³ C ^γ spin-echo difference ¹ H- ¹³ C HSQC	2	20 h	C: 57	600	³ J _{N-methyl C_γ} couplings
¹⁵ N-edited 3D ROESY	1	~60 h	H: 29.0; N: 38.1	600	ROE distances
¹⁵ N-edited 3D NOESY with water flip-back, ⁴¹ radiation damping during mixing period ⁴²	1	63 h	H: 29.0; N: 40.2	600	NOE distances
¹³ C-edited 3D NOESY optimized for detection of aliphatic ¹ H- ¹³ C groups ^{43,44}	3	35 h	H: 15.3; C _{aliph.} : 9.0	800	NOE distances
¹³ C-edited 3D NOESY optimized for detection of aromatic ¹ H- ¹³ C groups	3	26 h	H: 15.3; C _{arom.} : 6.9	800	NOE distances
4D ¹³ C- ¹³ C-NOESY-HMQC	3	136 h	C ₁ : 6.6; H ₁ : 11.7; C ₂ : 6.6	600	NOE distances
2D NOESY selective for ¹³ C and ¹⁵ N-attached protons ⁴⁵	3	~30 h	H: 22.4	800	Ligand assignments and NOE distances
2D HOHAHA selective for	3	15 h	H: 23.5	600	Ligand

¹³ C and ¹⁵ N-attached protons					assignments
Doublet-separated sensitivity-enhanced ¹ H- ¹⁵ N HSQC ⁴⁶	2, 4	3× 4 h	N: 45.0	800	N-H ^N RDCs
J-resolved constant-time ¹³ C-HSQC	2, 4	3× ~12 h	J: 24.9; C: 25.8	800	C ^α -H ^α RDCs

^a 1: ¹⁵N-protein in H₂O; 2: ¹³C-¹⁵N-protein in H₂O; 3: ¹³C-¹⁵N-protein in D₂O; 4: ¹³C-¹⁵N-protein in H₂O + Pfl phage

Table 2: Structure statistics.

	TipAS · promothiocin A	TipAS · nosiheptide
Completeness of resonance assignments:		
—Protein, backbone ^a	94%	95%
—Protein, all atoms ^b	91%	
—Antibiotic ¹ H ^c	12/18 = 67%	
NOE constraints ^d :		
—Protein, residues D165-P253	1354	
—Protein, intraresidue (i = j)	793	1259
—Protein, sequential (i - j = 1)	389	694
—Protein, short range (2 ≤ i - j ≤ 5)	409	559
—Protein, long range (i - j ≥ 6)	317	285
—Protein to ligand	13	22
—Within ligand	0	6
—Total	1949	2797
—NOE constraints per residue	13.6	19.6
RDCs	155	224
Hydrogen bonds	69	58
TALOS constraints ²⁴	216	225
χ ¹ angle constraints ^e	21	63
Total number of constraints	1881	3367
Total constraints per residue	13.15	23.5
Number of structures calculated / used	66 / 10	100 / 25
Constraint violations:		
—Number of distance violations >0.1 Å, per structure	301	
—RMSD of distance violations ^f	0.24 Å	~0.13 Å
—RMSD of dihedral angle violations	5.1°	3.2°
—RMSD of RDC violations	1.142 Hz	0.24 Hz
Coordinate precision (RMSD to average):		
—Protein backbone atoms ^g , residues P116-V122, E132-137, D143-A153	1.01 Å	0.55 Å
—Protein backbone atoms ^g , residues K157-H251	0.50 Å	0.60 Å
—Antibiotic non-hydrogen atoms	2.41 Å	2.46 Å
—All non-hydrogen atoms	1.68 Å	1.44 Å
Ramachandran plot summary from PROCHECK-NMR ^{47, h}		
—Most favored regions	92.8%	88.6%
—Additionally allowed regions	6.8%	7.6%
—Generously allowed regions	0.5%	2.3%
—Disallowed regions	0.0%	1.5%

^a Considering ¹H^N, ¹⁵N (except Pro), ¹³C^α, and ¹H^α resonances.

^b Considering routinely assignable ¹H, ¹⁵N, and ¹³C resonances; *i.e.* without N-terminal and Lys amino groups, Arg guanidino groups, hydroxyl protons of Ser, Thr, Tyr, thiol protons of Cys, carboxyl resonances of Asp and Glu, nonprotonated aromatic carbons, and Pro ¹⁵N. ¹H belonging to the same methyl group and Phe, Tyr ¹H^δ, ¹H^ε are counted as one.

^c Considering carbon-bound ¹H only. ¹H belonging to the same methyl group are counted as one.

^d Conformationally restricting restraints, *i.e.* excluding distances between geminal hydrogens or protons on the same aromatic group.

^e Derived from H^α-H^β ³J coupling constants and H^N-H^β ROE distances.

^f NOEs and hydrogen bonds.

^g N^H, C^α, and C^γ.

^h Residues P116-V122 and D143-P253, excluding Gly and Pro.

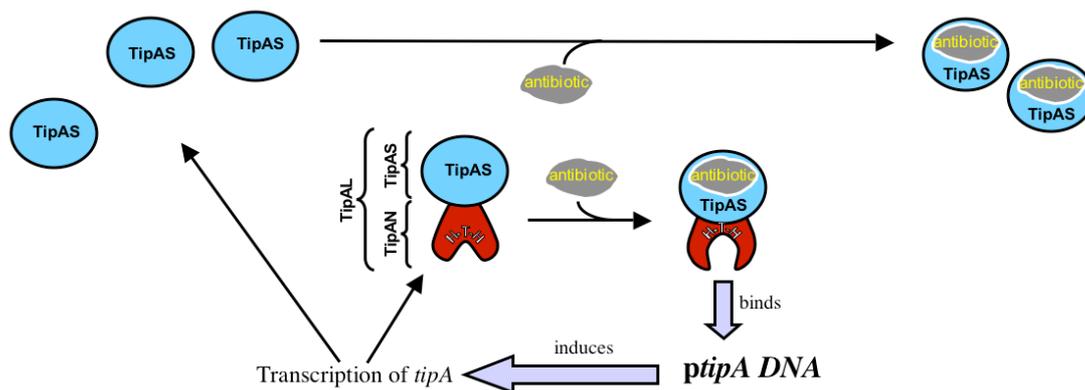


Figure 1: Schematic of the putative *tipA* antibiotic resistance mechanism. Antibiotic (gray) binds to the TipAS domain (blue) of TipAL, which consequently binds to *tipA* promoter DNA and upregulates transcription. TipAL as well as the isolated TipAS domain are produced, and TipAS neutralizes antibiotic by irreversibly binding to it.

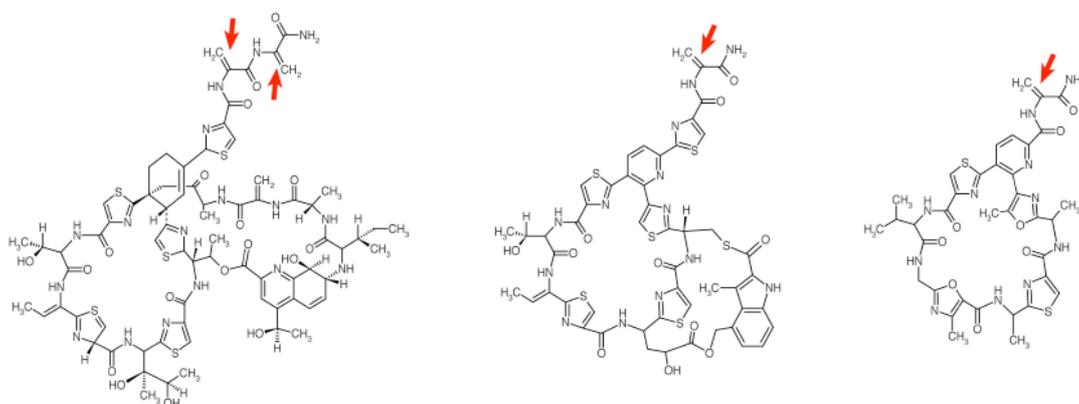


Figure 2: Chemical structures of thiostrepton (left), nosiheptide (center), and promothiocin A (right). Red arrows highlight dehydroalanine residues, which react with C214 in TipAS.

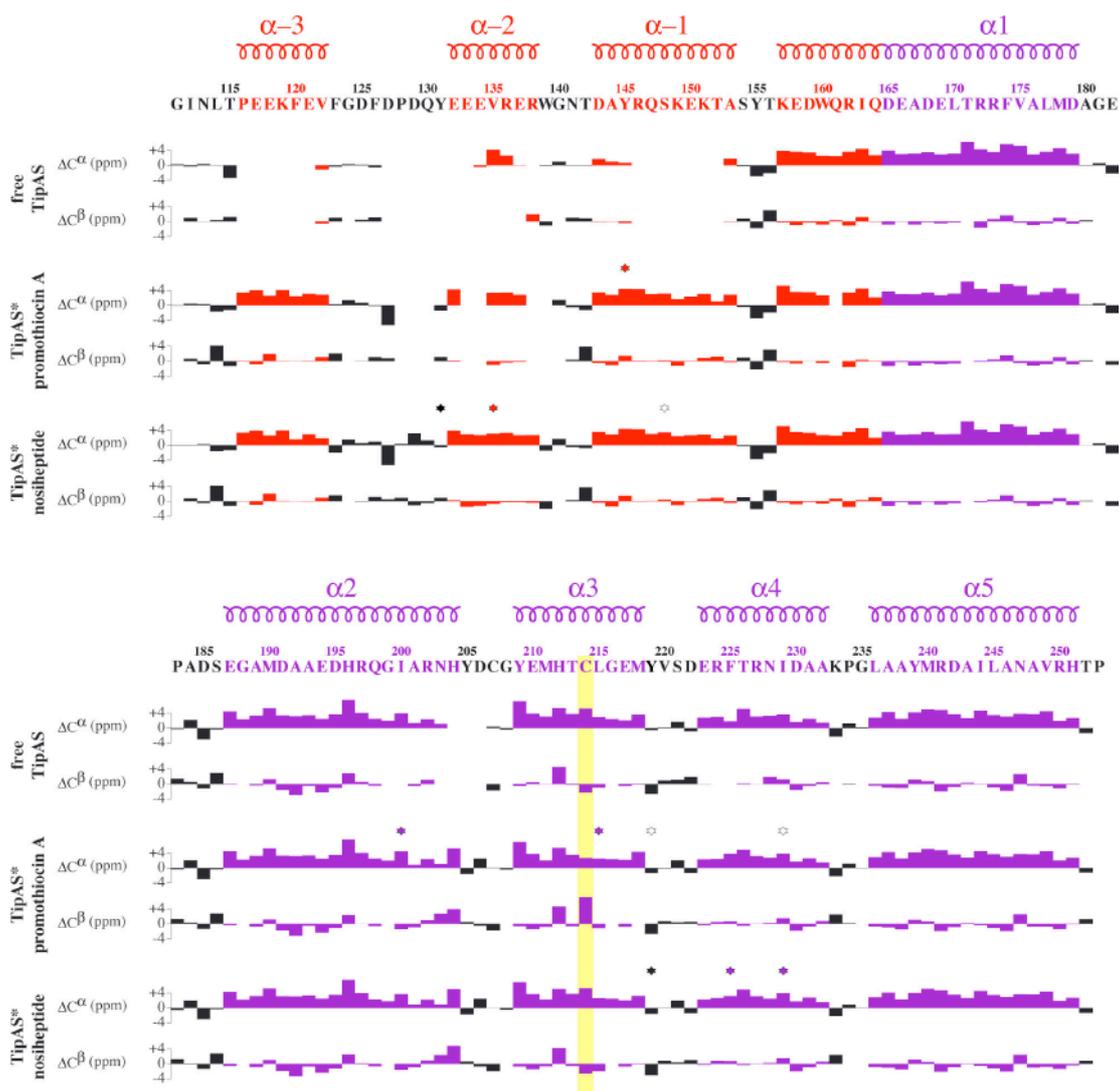


Figure 3: Secondary $^{13}\text{C}^\alpha$ and $^{13}\text{C}^\beta$ chemical shifts for apo-TipAS, TipAS · promothiocin A, and TipAS · nosiheptide. Secondary structure that exists in TipAS and in both complexes is shown in purple; secondary structure that only forms after antibiotic binding is shown in red. Stars denote observed NOE contacts between the given residue and the bound antibiotic: outlined stars represent single NOE contacts and filled stars represent several NOE contacts. Residue C214, which forms the covalent link to the antibiotic, is highlighted in yellow.

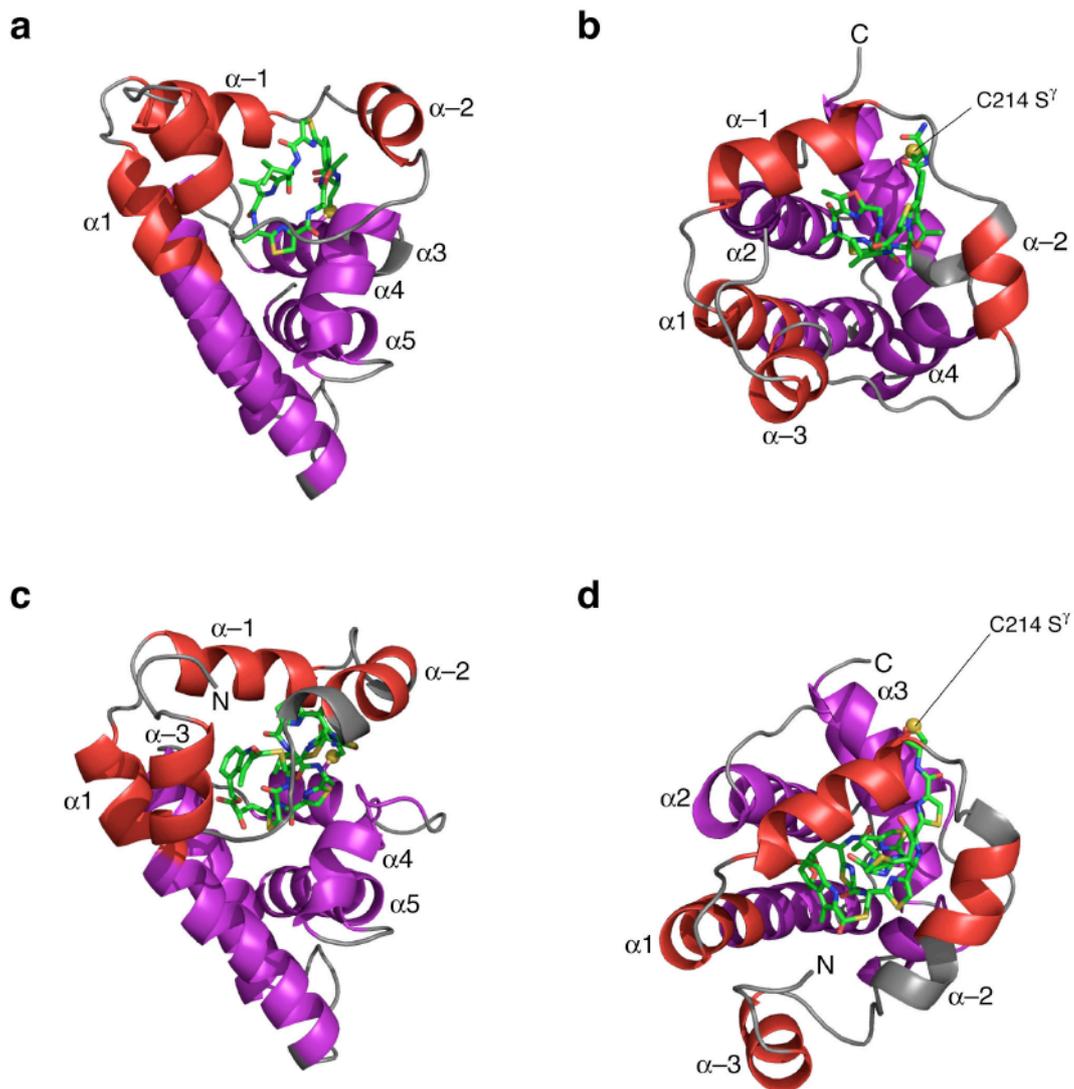


Figure 4: Solution structures of TipAS · promothiocin A (**a**, **b**) and TipAS · nosiheptide (**c**, **d**). TipAS is colored as in Figure 3. The antibiotics are shown as green / CPK-colored sticks; hydrogen atoms are omitted for clarity. C214 S^γ, which is covalently linked to antibiotic, is shown as a yellow sphere.

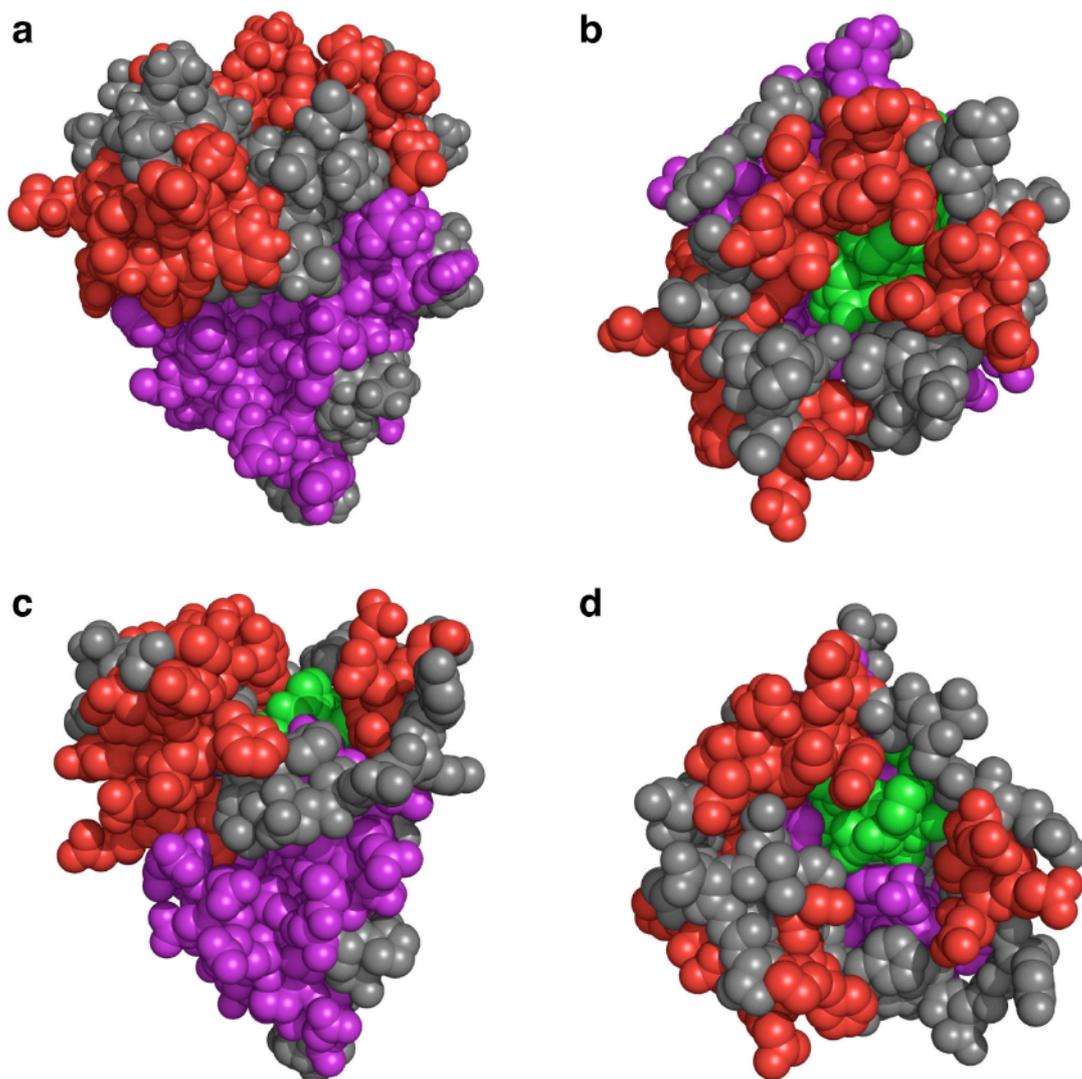


Figure 5: Space filling models illustrating the surface area of bound antibiotics covered by TipAS. (a, b), TipAS · promothiocin A; (c, d), TipAS · nosiheptide. TipAS is colored as in Figure 3; antibiotics are shown in green. Hydrogen atoms are omitted for clarity. Models are oriented as in Figure 4.

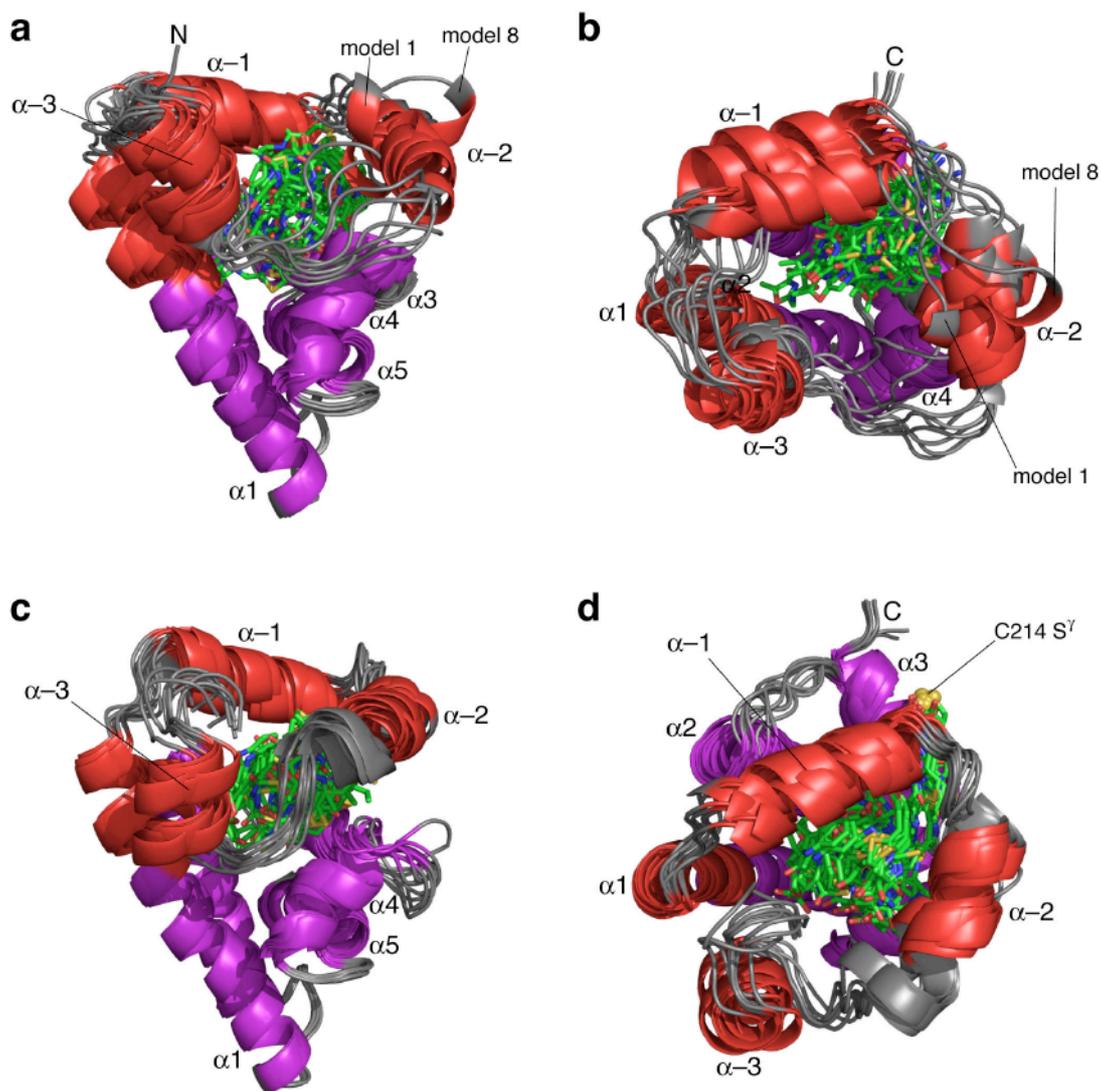


Figure 6: Bundles of structure models illustrating the precision of TipAS · antibiotic structures. (a, b), TipAS · promothiocin A; (c, d), TipAS · nosiheptide. The 10 / 9 lowest energy models from structure calculations are shown for TipAS · promothiocin A and TipAS · nosiheptide, respectively. Each model is aligned to C^α atoms of residues with α helical secondary structure in free TipAS, which are colored purple. TipAS is colored as in Figure 3; antibiotics are shown in green / CPK colors. Hydrogen atoms are omitted for clarity. Models are oriented as in Figure 4.

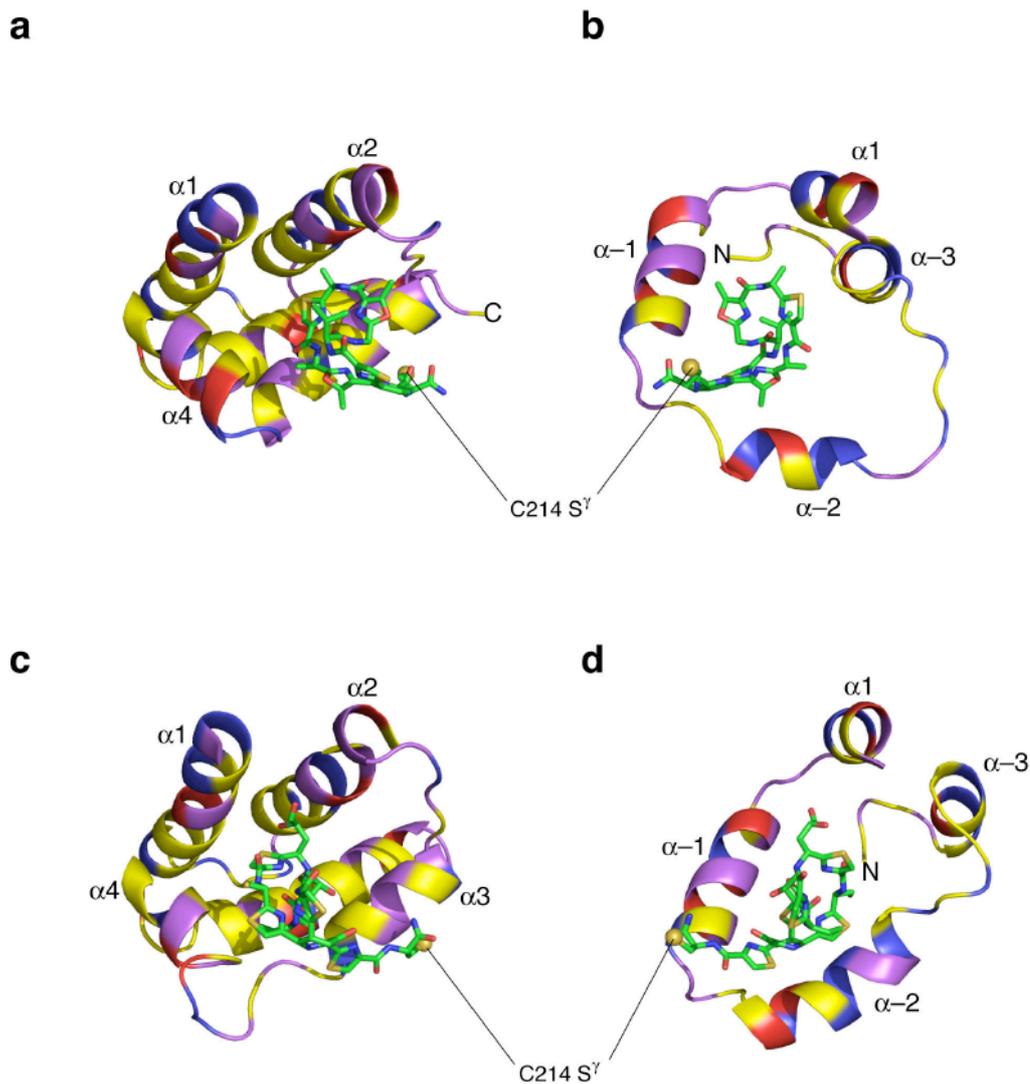


Figure 7: Overview of contact surfaces between TipAS and bound antibiotics. (a), residues Q164-P253 of TipAS · promothiocin A; (b), residues G111-P253 of TipAS · promothiocin A; (c), residues Q164-P253 of TipAS · nosiheptide; (d), residues G111-P253 of TipAS · nosiheptide. Antibiotics are shown as green / CPK-colored sticks; hydrogen atoms are omitted for clarity. TipAS residues are colored by amino acid type, as follows: yellow, hydrophobic; purple, polar; red, positively charged; blue, negatively charged.

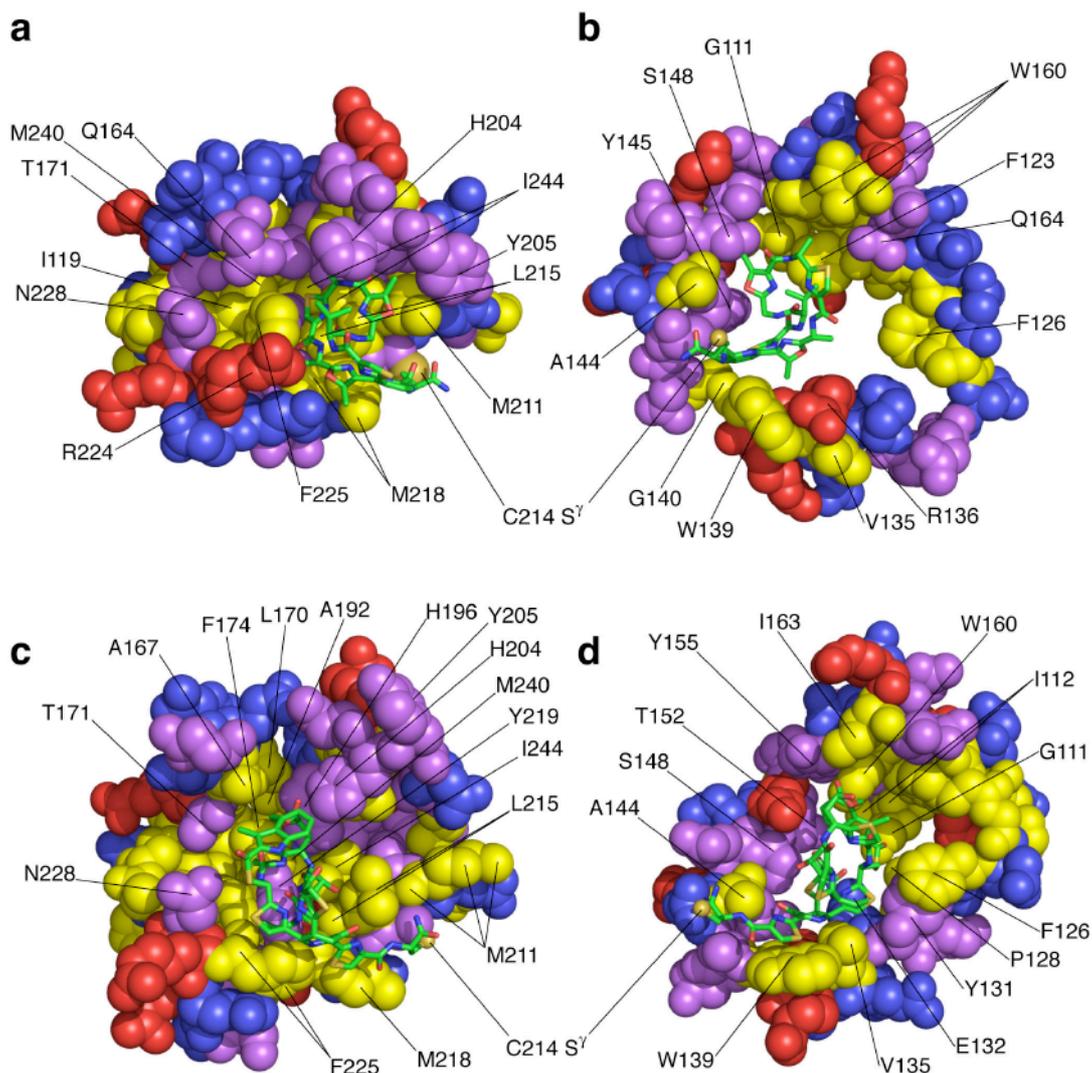


Figure 8: Space-filling models revealing contact surfaces between TipAS and bound antibiotics. **(a)**, residues Q164-P253 of TipAS · promethiocin A; **(b)**, residues G111-P253 of TipAS · promethiocin A; **(c)**, residues Q164-P253 of TipAS · nosiheptide; **(d)**, residues G111-P253 of TipAS · nosiheptide. Antibiotics are shown as green / CPK-colored sticks. TipAS residues are colored by amino acid type, as follows: yellow, hydrophobic; purple, polar; red, positively charged; blue, negatively charged. All hydrogen atoms are omitted for clarity. Orientations are as in Figure 7.

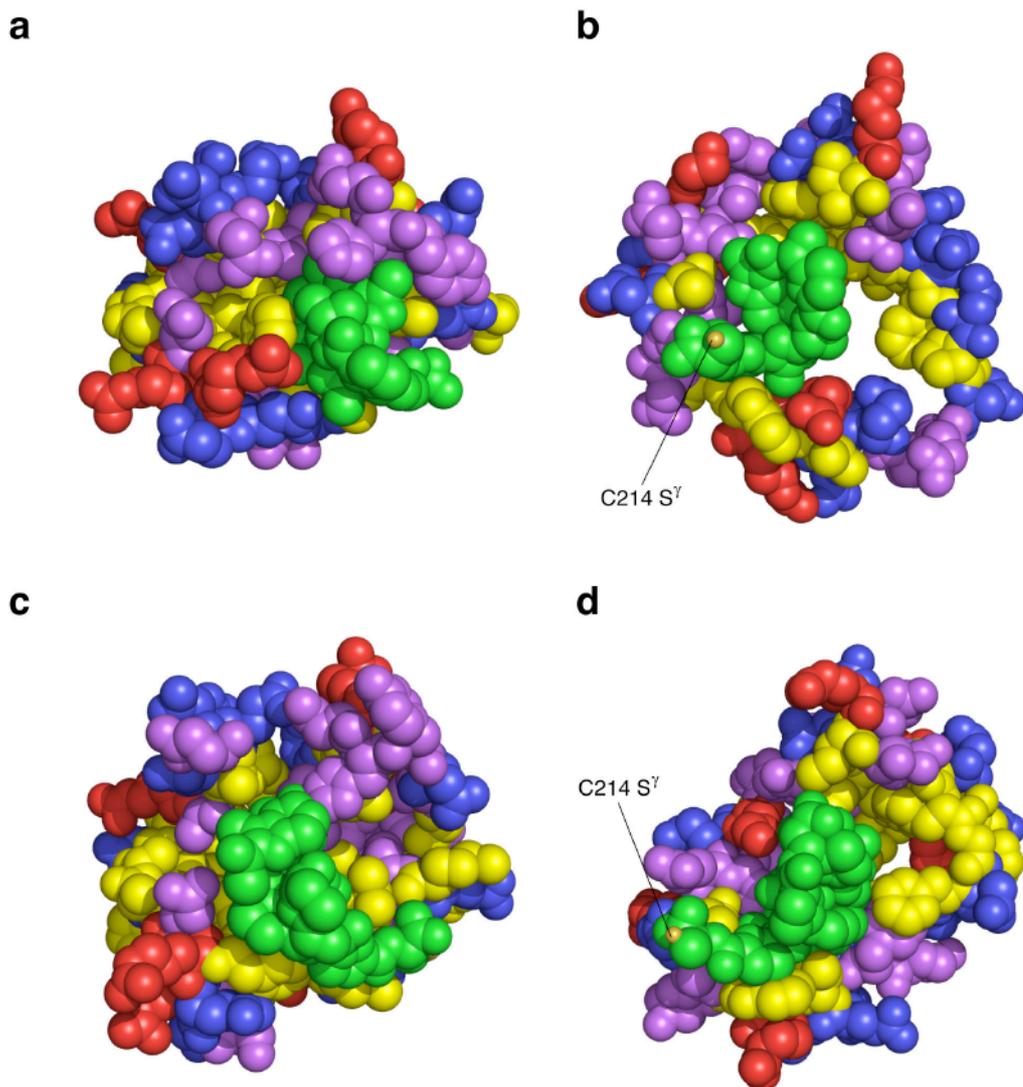


Figure 9: Space-filling models revealing holes in TipAS · antibiotic structures. (a), residues Q164-P253 of TipAS · promothiocin A; (b), residues G111-P253 of TipAS · promothiocin A; (c), residues Q164-P253 of TipAS · nosiheptide; (d), residues G111-P253 of TipAS · nosiheptide. Antibiotics are shown in green. TipAS residues are colored by amino acid type, as follows: yellow, hydrophobic; purple, polar; red, positively charged; blue, negatively charged. All hydrogen atoms are omitted for clarity.

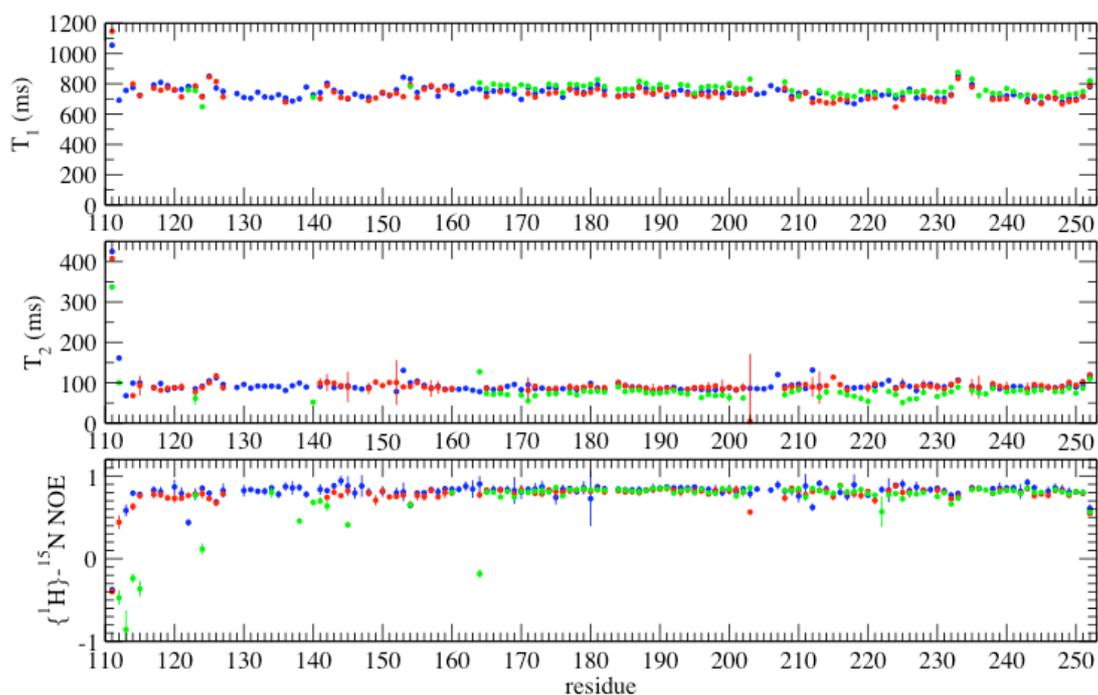


Figure 10: Amide ^1H - ^{15}N relaxation data at 298 K (25°C). Data for free TipAS are shown in green, TipAS · promothiocin A is red, and TipAS · nosiheptide is blue.

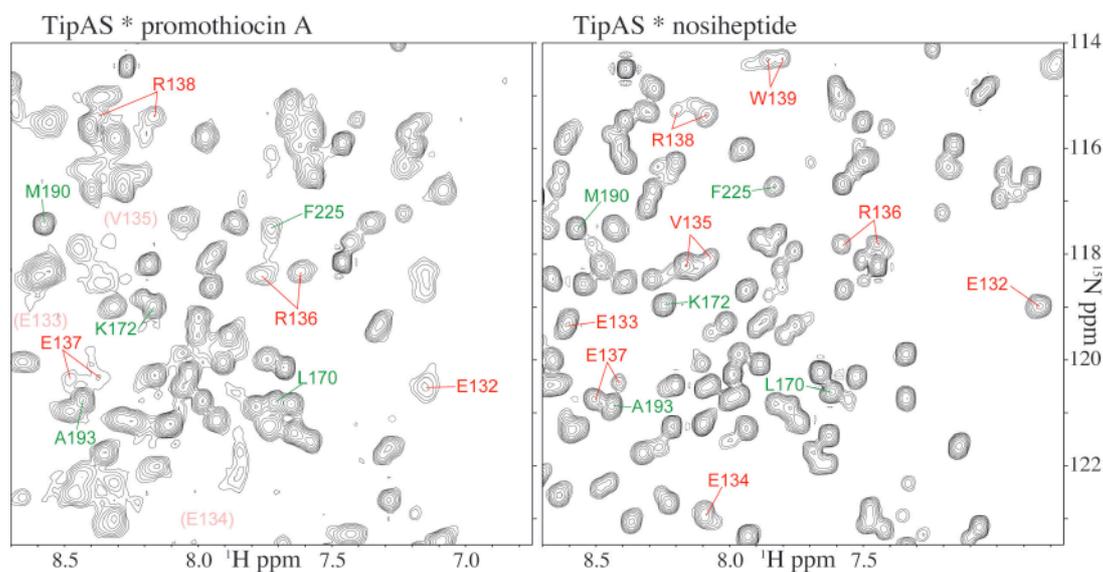


Figure 11: ^1H - ^{15}N HSQC spectra of TipAS in complex with promothiocin A and nosiheptide. Peaks corresponding to helix α -2 are labeled in red. Where peaks could be assigned at all, their intensity is lower for TipAS · promothiocin A. A few other peaks are labeled in green, for reference. Several residues have two stable conformations and are represented by two peaks; see text.

Structure of the transcriptionally inactive MerR domain TipAN in complex with DNA*

Martin G. Allan,^a Kate J. Newberry,^b Jason Schuman,^c Richard G. Brennan,^b Jörg Stetefeld,^d Charles J. Thompson,^e Tilman Schirmer^a and Stephan Grzesiek^a

^a Biozentrum, Universität Basel, Switzerland; ^b University of Texas M. D. Anderson Cancer Center, Houston, Texas, USA; ^c Oregon Health & Science University, Portland, Oregon, USA; ^d Dept. of Chemistry, University of Manitoba, Canada; ^e University of British Columbia, Vancouver, Canada

Abstract

Bacterial MerR proteins activate transcription of target genes in response to various stress signals by an unusual mechanism involving the bending and twisting of target promoter DNA. TipAL is a MerR protein from *Streptomyces lividans* that recognizes thiopeptide antibiotics with its C-terminal TipAS domain and activates transcription through its N-terminal DNA-binding domain, TipAN. We present the 2.9 Å crystal structure of TipAN bound to cognate promoter DNA. In contrast to all previously reported MerR protein-DNA complex structures, where bound DNA is bent by >45° and twisted, the TipAN complex contains DNA that is only slightly distorted. Thus this is the first structure of a transcriptionally inactive MerR protein-DNA complex. We present biophysical data indicating that TipAN is in a monomer-dimer equilibrium and that it has a much weaker affinity for cognate DNA than its homologues. We combine our findings with previously reported structural data on TipAS to propose a mechanism of antibiotic-dependent transcriptional activation for full-length TipAL.

* *Author contributions:* MGA produced and purified protein, acquired FA, NMR and CD data, identified crystallization conditions, acquired and processed diffraction data, determined the space group, and wrote the report. KJN designed the DNA used for crystallization and helped with crystallization. JSch helped with crystallization. RGB provided materials and equipment and helped with crystallization. JSt helped with crystallization, acquired diffraction data, and helped with data analysis. TS solved the structure by Molecular replacement and refined it.

Introduction

The MerR proteins are a family of bacterial transcriptional activators that respond to various stress signals including heavy metals, oxidative stress, dyes or antibiotics. They consist of a helix-turn-helix DNA-binding domain, a long antiparallel coiled-coil dimerization domain, and a specific receptor domain. The DNA-binding and coiled-coil domains show sequence homology within the MerR protein family, whereas the receptor domains do not. Promoters activated by MerR proteins have unusually long spacers between two RNA polymerase binding motifs typically located about 35 and 10 base pairs upstream of their transcription start sites. The presence of 2-3 additional base pairs results in two structural changes which prevent RNA polymerase binding: the -35 and -10 elements are too far apart and no longer on the same side of the DNA double helix. Activated MerR proteins bound to cognate promoters cause DNA to bend and twist the -35 and -10 elements back into alignment, allowing RNA polymerase binding and transcriptional initiation (1).

X-ray crystal structures have been available for three MerR proteins in complex with cognate promoter DNA: BmrR (PDB IDs 1exj (2), 1r8e (3) and 1exi (2)) and Mta (1r8d (3)), both of which activate the expression of multidrug efflux pumps in *Bacillus subtilis*; and SoxR (2zhg (4)), an oxidative stress sensor in *Escherichia coli*. Each of these structures contains promoter DNA that is bent and twisted, and each complex was shown to be transcriptionally active: BmrR was crystallized with known inducers bound to the protein; Mta was crystallized as a constitutively active truncation mutant named MtaN; and SoxR was crystallized in its oxidized, active state.

TipAL (thiostrepton induced protein A, long form) is a MerR protein from *Streptomyces lividans*, a soil-dwelling, saprophytic bacterium. Its C-terminal sensor domain, TipAS (thiostrepton induced protein A, short form), recognizes various thiopeptide antibiotics including micrococcin, thiostrepton, nosiheptide, promothiocin and others (5). After antibiotic binding, TipAL upregulates the expression of its own gene and several others that have not been characterized (6). *tipA* messenger RNA contains two ribosome binding sites and is translated to full-length TipAL as well as the isolated sensor domain TipAS. Thus TipAL detects thiopeptide antibiotics and responds by upregulating the production of both TipAL and TipAS. TipAS is produced in >20-fold excess over TipAL (7) and neutralizes antibiotics by irreversibly binding to them (figure 1). The *tipA* gene confers resistance to thiopeptide antibiotics *in vivo*: it was shown that disruption of the *tipA* gene leads to 5-fold increased sensitivity to thiostrepton in *S. lividans* (8). It is noteworthy that *Streptomyces* strains that produce thiopeptide antibiotics generally lack the *tipA* gene (9) and instead

possess *tsr*, a much more effective resistance gene. *tsr* encodes a methylase that modifies the ribosome to prevent binding of thiopeptide antibiotics.

The solution structure of TipAS in the absence of antibiotics has been solved by NMR spectroscopy (10). The C-terminal ~90 residues of TipAS form a rigid structure comprising 5 α helices in an arrangement reminiscent of the globin fold. Residue C214, which spontaneously forms a covalent bond with many thiopeptide antibiotics, is located next to a deep hydrophobic cleft. The N-terminal ~50 residues of apo-TipAS are flexible; they form stable α helical structure only after antibiotic binding.

The N-terminal, DNA binding domain of TipAL, TipAN (thiostrepton induced protein A, N terminal domain) is homologous to the DNA-binding domains of all MerR proteins, including Mta (48% sequence identity), SoxR (26%), BmrR (24%) and, of course, MerR (28%).

We have undertaken structural studies of TipAN and its interaction with the *tipA* promoter (*ptipA*) as a part of our investigation into TipAL-mediated transcriptional activation and antibiotic resistance. Here we present the crystal structure of TipAN bound to the *tipA* promoter. We find that in contrast to other MerR protein-promoter structures, DNA is not bent or twisted into a transcriptionally active form. Comparisons of TipAN·*ptipA* and MtaN·*pmta* rationalize why the isolated MtaN domain is sufficient for transcriptional activation while TipAN is not, and point to a mechanism for TipAL activation. We present biophysical data indicating that the TipAN dimer is unstable, and we combine our findings with available information about the TipAS domain to propose a mechanism for antibiotic-induced transcriptional activation by TipAL.

Materials and Methods

Expression and purification of TipAN

The *tipAN* coding sequence was PCR-amplified from plasmid pDS8::*tipAL* (8). TipAN was expressed using a modified pET22b(+) vector in *Escherichia coli* BL21 (DE3). Cells were grown at 37°C in 8 L Miller LB broth (Difco) containing 100 mg/L ampicillin, in shaking flasks. Expression was induced at OD_{600 nm} 1.0 by addition of IPTG to a final concentration of 1 mM, and maintained for 4 h. Cells (18 g) were disrupted using a French pressure cell (Spectronic Unicam). Insoluble cell debris was removed by centrifugation.

All purification steps were carried out at 4°C. TipAN was precipitated by adding ammonium sulfate to 70% saturation (w/v). Supernatant was discarded, and precipitate was dialyzed against MA buffer (20 mM Tris-HCl, 50 mM NaCl, 5 mM MgCl₂, 1 mM EDTA, pH 8). TipAN solution was loaded on a 75 mL Q-Sepharose

Fast Flow anion exchange column, and eluted with a 400 mL linear gradient of 50 to 635 mM NaCl in MA buffer. Fractions were pooled based on their absorption at 280 nm and analysis by SDS-PAGE. TipAN eluted at a NaCl concentration of 370 mM. TipAN solution was further purified on a Superdex 75 prep grade size exclusion column (Amersham Pharmacia) with bed diameter 26 mm and length 60 cm, in portions of 5 mL per run to prevent aggregation. Protein was eluted with MA buffer containing 150 mM NaCl at a flow rate of 1 mL/min. Fractions were pooled based on their absorption at 280 nm and analysis by SDS-PAGE. TipAN eluted at a retention volume of 164 mL. Based on UV absorption and assuming an extinction coefficient of 5120 AU/M/cm at 280 nm (11), the TipAN concentration after size exclusion chromatography was 25 μ M, and the total yield was 30 mg. Prior to crystallization, TipAN was further purified by reversed phase HPLC. TipAN was loaded on a Grace Vydac Protein C4 column with bed diameter 2.2 cm and bed length 25 cm, in portions of 5 mL per run. Protein was eluted with a linear gradient from solvent A (H_2O containing 0.1% TFA (v/v)) to solvent B (90% acetonitrile, 10% H_2O , 0.085% TFA (v/v)). TipAN eluted at 49% solvent B. No contaminant peaks were detected. The TipAN solution was lyophilized and stored at -70°C . Prior to crystallization, TipAN was redissolved in MA buffer containing 150 mM NaCl, and concentrated to 1.0 mM (12 mg/mL) using a 2 mL, 10 kDa molecular weight cutoff Centricon centrifugal filter device (Amicon / Millipore).

Preparation of double-stranded *tipA* promoter DNA

DNA oligonucleotides of sequence 5'-TTG CAC CTC ACG TCA CGT GAG GAG CG-3' and 5'-ACG CTC CTC ACG TGA CGT GAG GTG CA-3' were purchased from Microsynth, Balgach, Switzerland, in HPLC-purified form. Each strand was dissolved in 10 mM sodium cacodylate, pH 6.5, to obtain DNA concentrations of 2 mM. Equal amounts of complementary strands were mixed, heated to 98°C , and slowly cooled to room temperature. The solution was concentrated to 1.5 mM double strand by means of a 3 kDa molecular weight cutoff centrifugal filter device (Amicon / Millipore). Double strand formation was confirmed by NMR spectra showing hydrogen-bonded imino hydrogens. Concentration was measured by UV spectroscopy, assuming an extinction coefficient of 491 AU/mM/cm for the double strand.

Crystallization

Crystals were grown at room temperature by the sitting drop vapor diffusion method. The reservoir solution was 600 μ L of 15% (w/v) PEG-8000, 0.2 M MgCl_2 , 0.1 M Tris-HCl at pH 8.5. The protein-DNA complex was prepared immediately before crystallization by mixing TipAN and DNA stock solutions and MA buffer containing

150 mM NaCl to obtain final concentrations of 500 μ M (6.0 mg/mL) TipAN and 275 μ M double stranded DNA. The crystallization drop contained 4 μ L protein-DNA solution and 6 μ L reservoir solution. Crystals appeared after one week and grew for about two weeks, after which they started to decompose. Crystals grew in clusters and were flat, typically measuring 0.2 mm in diameter and 10 μ m in thickness.

Crystals were transferred to cryoprotectant solution (170 mM Tris-HCl at pH 8.5, 330 mM MgCl₂, 25% PEG-8000) and soaked for 1 min before flash freezing in liquid N₂.

X-ray diffraction and data processing

Diffraction data were recorded at the European Synchrotron Radiation Facility (ESRF), Grenoble, France, beamline ID14 3; and Swiss Light Source (SLS), Villigen, Switzerland, beamline PX1. Crystals were kept at 100 K in a stream of cold N₂.

Diffraction data were processed with the program MOSFLM (12) and subsequently scaled with programs of the CCP4 suite (13).

Structure determination

The TipAN·*ptipA* structure was solved by Molecular Replacement, using the software Phaser (14). A fragment of the MtaN·DNA structure (1r8d) comprising protein residues N7-H71 was trimmed to those atoms that are common to MtaN and TipAN, and used as a first search model. A solution with R_{free} 53.0% was found using diffraction data up to 3 Å resolution. This first solution was held fixed, and a second search was performed using residues K77-R103 of the dimerization helix, again trimmed to common atoms, as a model. A solution with a plausible relative orientation of the two models was found, reducing R_{free} to 44.0%. These two components were fixed, and a five base pair fragment of DNA could be placed in a third search, further reducing R_{free} to 41.4%. All three components were refined as rigid bodies, and side chains and additional amino acids and base pairs were successively added. Full atom refinement including restrained B-factor refinement resulted in a final R_{free} of 27.4%. Refinement was carried out with Refmac5 (13).

Biophysical characterization

Mass spectrometry. LC-ESI, Q-TOF detected mass spectra were recorded on different preparations of TipAN. A mass of 12034.4 Da was found, in excellent agreement with the predicted mass of 12034.6 Da for TipAN minus residue M1.

Circular dichroism. CD spectra were recorded on 16.4 μ M TipAN and 14.6 μ M MtaN (see (3) for production and purification) in 150 mM (K/H)₃PO₄, 15 mM NaCl, pH 8.0, at room temperature. Spectra were deconvoluted into contributions from different secondary structure elements with the program CDNN (15).

Fluorescence anisotropy. Dissociation constants (K_D) of TipAN, TipAL, and TipAL with bound thiostrepton from a double-stranded, singly 5'-fluoresceinated fragment of *tipA* DNA (5'-fluorescein-GGC TTG CAC CTC ACG TCA CGT GAG GAG GCA GCG T-3' / 5'-ACG CTG CCT CCT CAC GTG ACG TGA GGT GCA AGC C-3'; Oligos Etc., Wilsonville, Oregon, USA) were measured by fluorescence anisotropy, using a PanVera Beacon Fluorescence Polarization System. Measurements were performed at 25°C in 1 mL 50 mM Tris-HCl at pH 7.5, 150 mM KCl. The concentration of double-stranded DNA was 1.5 nM, and the protein concentration was varied from 0 to 20-25 μ M in 20-30 steps.

NMR spectroscopy. Jump-return spin echo ^1H 1D spectra (16) were recorded on a Bruker DRX 800 spectrometer at a sample temperature of 293 K. Amide T_2 relaxation time constants were measured by varying spin echo delays between 200 μ s and 5.8 ms. NMR samples contained 100 μ M TipAN in MA buffer containing 500 mM NaCl and 5% (v/v) D_2O ; the concentrations of TipAN and NaCl were varied in individual experiments.

Results

Crystallography and model building. The structure of the TipAN-DNA complex was determined to a resolution of 2.9 Å by Molecular Replacement, using fragments of the MtaN-DNA complex structure (1r8d (3)) as models. The space group was $\text{P2}_12_12_1$, with the asymmetric unit containing one TipAN monomer and one half of the DNA double strand. This posed a problem since the double-stranded, pseudopalindromic promoter DNA is not perfectly symmetric. Indeed, disorder was observed for the bases at positions 1 and 10 / -10. These nucleotides were modeled as Gua, Thy, and Thy, respectively, and appear as non-complementary in our model even though they are canonical Watson-Crick base pairs in nature. No base-specific crystal contacts are observed in our structure, and the crystal packing does not allow for base pairing between the 5' overhangs as observed in MtaN·*pmta* (3). We conclude that the two possible orientations of the DNA double helix are randomly distributed in the TipAN·*tipA* crystal. Nucleotides -14, -13, -12, 12 and 13, protein residues P72-P76 and A104-S109 and the side chains of residues D24, E32, R33, H35, D71, R77, R82, Q103 and E106 are not defined by electron density and were excluded from the model. Selected crystallographic statistics and indicators of model quality are given in Table 1.

Overall structure of the TipAN-DNA complex. TipAN forms a homodimer and, like MtaN, has the topology $\alpha 1-\alpha 2-\beta 1-\beta 2-\alpha 3-\alpha 4-\alpha 5$ (figure 2). Helices $\alpha 1$ and $\alpha 2$ are a canonical helix-turn-helix motif that contacts the major groove of the *tipA* promoter. Strands $\beta 1$ and $\beta 2$ form a small, antiparallel β sheet connected by a loop that contacts

the minor groove of DNA. TipAN dimerizes by forming an antiparallel coiled coil with the long helix $\alpha 5$. In stark contrast to previously described MerR protein·DNA structures, where DNA is bent by 47–65°, the *tipA* promoter is only bent by 17° (figures 2, 3). The structure of the DNA-binding subdomain encompassing helices $\alpha 1$ – $\alpha 5$ and the β sheet is very similar in TipAN and MtaN, and so is their position relative to directly bound DNA nucleotides: the rmsd of C $^{\alpha}$ and P atom positions for protein residues S/Q4-D70 plus DNA nucleotides Cyt–9/Gua9*–Gua–3/Cyt3* is only 0.61 Å between TipAN·*tipA* and MtaN·*pmta*. Likewise, the coiled coil subdomain is very similar in the two structures, with an rmsd of C $^{\alpha}$ positions of 0.65 Å for residues H79–M105 and H79*–M105*. There are two major differences between the two structures: the central nucleotides of promoter DNA are distorted in MtaN·*pmta* but much less so in TipAN·*tipA*, and the coiled coil subdomain is translated by 13.3 Å and rotated by 23° relative to the DNA-binding subdomain and DNA.

Protein-DNA interactions, shown in figures 4a and 4b, are very similar in TipAN·*tipA* and MtaN·*pmta*. Base-specific contacts occur primarily between the major groove of DNA around nucleotide Gua4 and the recognition helix $\alpha 2$ of the helix-turn-helix motif. Residue H38 may connect to the minor groove of DNA around nucleotide Cyt–10 indirectly, *via* bound H₂O, as in MtaN·DNA; this cannot be shown conclusively as the resolution of our structure is insufficient to observe water molecules. Ten hydrogen bonds are formed between the DNA phosphate backbone and the protein backbone or side chains.

Interactions between coiled coil and DNA-binding subdomains. The most striking structural differences between activated MtaN·*pmta* and inactivated TipAN·*tipA* are found in the interface between the DNA-binding subdomain and the coiled-coil dimerization helix (see Discussion). This interface is shown in figure 3c. Interactions occur primarily between $\alpha 3$ and the dimerization helix $\alpha 5$ of the same protein chain and to a lesser extent between $\alpha 3$ and $\alpha 5^*$ of the adjacent monomer. In addition to many hydrophobic interactions, residues Y55 and Q51 on $\alpha 3$ form hydrogen bonds to Q83, R82 and H79 on $\alpha 5^*$.

Sequence alignments. An alignment of the TipAN sequence with N-terminal domains of Mta and BmrR is shown in figure 5. In MtaN, the coiled coil dimerization domain is stabilized by inter-chain salt bridges between K84 and D102, and between K91 and D95 (3). These residues correspond to H84, E102, I91, and N95 in TipAN, which are unlikely to form salt bridges, although this could be partially compensated by hydrophobic contact between I95 and L94. In *tipA*, 14 out of 23 base pairs are the same as in *pmta*, and *tipA* is almost perfectly centro-symmetric (figure 2). It is noteworthy that the central base pair in *tipA* is Gua-Cyt, whereas it is Ade-Thy in other MerR promoters. In the transcriptionally active MtaN·DNA and BmrR·TPP⁺·DNA complexes, this base pair is broken as a consequence of protein

binding and DNA bending; however, it was shown that mutating this base pair to Ade-Thy did not significantly enhance protein binding, even though Gua-Cyt base pairing is stronger than Ade-Thy (3).

DNA binding affinities. Fluorescence affinity measurements revealed that the dissociation constants (K_D) from *ptipA* were $25 \pm 6 \mu\text{M}$ for TipAN, $23 \pm 9 \mu\text{M}$ for apo-TipAL and $5.4 \pm 1.3 \mu\text{M}$ for TipAL with bound thiostrepton (data not shown). Thus the binding affinity of TipAN is the same as that of apo-TipAL; it increases slightly after antibiotic binding to TipAL. The observed dissociation constants are three orders of magnitude higher than those measured for MtaN binding to *pmta* or BmrR binding to *pblr* under the same conditions (3).

Dynamic instability of the TipAN dimer. Several observations indicate that TipAN in the absence of DNA forms unstable dimers at micromolar concentrations. First, in the process of TipAN purification, we noted that the protein elutes as an asymmetric peak that shifted towards lower molecular weights when lower concentrations of protein were loaded on the column (data not shown), indicating that TipAN oligomerizes at concentrations in the low μM range. Second, ^1H - ^{15}N HSQC NMR spectra of TipAN were of very poor quality and contained only weak signals (data not shown), consistent with intermediate exchange on the chemical shift time scale. Assuming a diffusion-controlled association rate (about $10^8 \text{ M}^{-1}\text{s}^{-1}$), a K_D of oligomerization in the low μM range would correspond to a dissociation rate in the low $(\text{ms})^{-1}$ range, which would lead to broad, weak NMR signals as observed. Amide ^1H T_2 relaxation time constants, which depend on molecular tumbling rates and thus on molecular mass, increased at low concentrations of protein and high concentrations of NaCl (figure 7); both of these conditions are expected to disfavor dimerization or aggregation of TipAN. Finally, our fluorescence anisotropy measurements showed that TipAN has a much lower affinity for cognate DNA than MtaN (see above). TipAN binds to DNA as a dimer; the free energy of binding is effectively doubled if a preformed protein dimer binds to DNA with two binding sites simultaneously instead of two protein monomers binding independently, and a low stability of the TipAN dimer thus leads to a low DNA binding affinity. It is conceivable that the affinity of TipAN for *ptipA* would drastically increase if residues H84, E103, I91, and N95 were mutated such that salt bridges as observed in MtaN could be formed, stabilizing the dimer.

CD spectra of TipAN and MtaN (data not shown) showed a significantly lower content of canonical secondary structure for TipAN (38% α helix) than for MtaN (48% α helix). The crystal structure of MtaN in the absence of DNA (1jbg (17)) contains 67% helix.

Discussion

In complex with TipAN, *tipA* promoter DNA is only bent by 17° (figure 3), in stark contrast to a bending angle of 47° in MtaN·*pmta*, ~50° in BmrR·TPP⁺·*pbmr*, and 65° in SoxR·*psoxS*. Since DNA distortion is a prerequisite for transcription from MerR-like promoters, this finding implies that the isolated TipAN domain is insufficient for transcriptional activation. This comes as a surprise given that highly homologous MtaN strongly activates *pmta* transcription in the absence of its regulator domain.

Comparison of the TipAN·*tipA* and MtaN·*pmta* complex structures shows that DNA is bound in the same way in both structures, *i.e.* the orientation of helices $\alpha 1$ - $\alpha 4$ and the β sheet (the DNA-binding subdomain) relative to bound DNA is identical in the two complexes (figures 3, 4b). Differences are found in the central region of DNA and in the orientation of the ($\alpha 5$)₂ coiled coil relative to the DNA-binding subdomain (figures 3, 4c). Thus, the switch between inactive and activated MerR protein-DNA complexes lies in this subdomain interface. The most striking difference between the subdomain interfaces is at residue Y55 in helix $\alpha 3$ in TipAN, which forms hydrogen bonds with H79, R82 and Q83 of helix $\alpha 5$, stabilizing the inactive conformation. The corresponding residue in MtaN is F55, which cannot form hydrogen bonds (figure 4c). This explains why the default conformation of TipAN is inactive, and an activated receptor domain is required for transcription; whereas MtaN is active in the absence of a receiver domain, and the receiver domain in its inactive state is required for MtaN to assume its inactivated conformation.

The affinity of both TipAN and TipAL for cognate DNA is very weak, with dissociation constants of 5-25 μ M; the binding of MtaN to *pmta* or of BmrR to *pbmr* is three orders of magnitude stronger under the same conditions (3). The dissociation constants measured for TipAN and TipAL are higher than the minimum inhibitory concentration of thiostrepton, which was reported as 0.1-0.6 μ M (8). This is no impediment to the proposed mechanism of antibiotic resistance, however, since the covalent, irreversible binding of antibiotics to TipAL allows for intracellular TipAL·antibiotic concentrations exceeding the extracellular antibiotic concentration.

The weak binding of TipA proteins to DNA is related to the instability of the TipAN dimer, which was observed by CD spectroscopy, size exclusion chromatography and NMR spectroscopy: the monomerization of TipAN abolishes the cooperative binding of TipAN molecules to DNA and thus effectively reduces the binding affinity.

We have previously reported that the N-terminal part of the receptor domain TipAS is flexible in the absence of antibiotic, and forms α -helical structure after ligand binding (10). Combining this finding with our observation that the switch between inactive and activated MerR protein-DNA complexes lies in the interface between the DNA-

binding and coiled-coil subdomains of TipAN, we propose the following model of antibiotic-induced transcriptional activation by TipAL (figure 8): In the absence of antibiotic, TipAN is bound to relaxed, transcriptionally inactive DNA. The structured part of the TipAS receiver domain is connected to the TipAN coiled coil *via* the flexible N-terminal part of TipAS. After antibiotic binding, the N-terminal part of the TipAS domain forms α -helical structure, pulling TipAS closer to TipAN. The TipAS domain presses on the DNA-binding subdomain of TipAN, causing TipAN to switch to its active conformation. As a consequence, bound DNA is distorted in the way observed in the MtaN·DNA and BmrR·DNA structures, and transcription is initiated. Our data are insufficient to determine where TipAS contacts the DNA-binding part of TipAN; however, it is plausible that three consecutive glutamate residues (E132-E134) in the N-terminal part of TipAS form salt bridges to a positively charged region on the surface of TipAN such as residues R90, K93, and K96 on helix α 5. The proposed structure of the activated TipAL·DNA complex is similar to the activated BmrR·DNA structure, where the ligand-binding domain is in contact with the DNA-binding subdomain as well as the coiled coil domain, in a region that roughly corresponds to residues 90-96 of TipAN.

In conclusion, we have determined the first structure of a transcriptionally inactive complex of MerR-protein with cognate promoter DNA. We have shown that the switch between the active and inactive states lies in the interface between the DNA-binding and coiled-coil subdomains, and we have presented a possible mechanism of antibiotic-induced transcriptional activation by TipAL.

Acknowledgements

We thank Marco Rogowski for preparing the pET22b::*tipAN* expression vector; Patrick Graff (Novartis, Basel, Switzerland) for mass spectrometric analyses; the staff of SwissLight Source (Villigen, Switzerland) for assistance with data acquisition; and the European Synchrotron Radiation Facility (Grenoble, France) for an allotment of beam time (experiment MX-459, 2005).

Bibliography

1. Brown, N. L., Stoyanov, J. V., Kidd, S. P., and Hobman, J. L. (2003) *FEMS Microbiology Reviews* **27**, 145-163
2. Heldwein, E. E. Z., and Brennan, R. G. (2001) *Nature* **409**, 378-382
3. Newberry, K. J., and Brennan, R. G. (2004) *Journal of Biological Chemistry* **279**, 20356-20362

4. Watanabe, S., Kita, A., Kobayashi, K., and Miki, K. (2008) *Proceedings of the National Academy of Sciences of the United States of America* **105**, 4121-4126
5. Bagley, M. C., Dale, J. W., Merritt, E. A., and Xiong, A. (2005) *Chemical Reviews* **105**, 685-714
6. Murakami, T., Holt, T. G., and Thompson, C. J. (1989) *Journal of Bacteriology* **171**, 1459-1466
7. Holmes, D. J., Caso, J. L., and Thompson, C. J. (1993) *EMBO Journal* **12**, 3183-3191
8. Chiu, M. L., Folcher, M., Katoh, T., Puglia, A. M., Vohradsky, J., Yun, B. S., Seto, H., and Thompson, C. J. (1999) *Journal of Biological Chemistry* **274**, 20578-20586
9. Yun, B. S., Hidaka, T., Kuzuyama, T., and Seto, H. (2001) *Journal of Antibiotics* **54**, 375-378
10. Kahmann, J. D., Sass, H. J., Allan, M. G., Seto, H., Thompson, C. J., and Grzesiek, S. (2003) *EMBO Journal* **22**, 1824-1834
11. Gasteiger, E., Gattiker, A., Hoogland, C., Ivanyi, I., Appel, R. D., and Bairoch, A. (2003) *Nucleic Acids Research* **31**, 3784-3788
12. Leslie, A. G. W. (1992) *Joint CCP4 + ESF-EAMCB Newsletter on Protein Crystallography* **26**
13. Collaborative Computational Project, N. (1994) *Acta Crystallographica Section D Biological Crystallography* **50**, 760-763
14. McCoy, A. J., Grosse-Kunstleve, R. W., Storoni, L. C., and Read, R. J. (2005) *Acta Crystallographica Section D* **61**, 458-464
15. Bohm, G., Muhr, R., and Jaenicke, R. (1992) *Protein Engineering* **5**, 191-195
16. Sklenar, V., and Bax, A. (1987) *Journal of Magnetic Resonance* **74**, 469-479
17. Godsey, M. H., Baranova, N. N., Neyfakh, A. A., and Brennan, R. G. (2001) *Journal of Biological Chemistry* **276**, 47178-47184
18. Corpet, F. (1988) *Nucleic Acids Research* **16**, 10881-10890
19. Gouet, P., Courcelle, E., Stuart, D. I., and Metoz, F. (1999) *Bioinformatics* **15**, 305-308
20. DeLano, W. L. (2007), DeLano Scientific LLC, Palo Alto, CA, USA, <http://www.pymol.org>

Figures and Tables

Table 1: Crystallographic statistics.

Beamline	SLS PX1	
Wavelength	0.9001 Å	
Space group	P2 ₁ 2 ₁ 2	
Unit cell parameters	a = 45.9 Å, b = 67.5 Å, c = 73.1 Å	
Matthews volume	2.83 Å ³ /Da	
	(overall)	(outer shell)
Resolution limits	30.0 – 2.90 Å	3.06 – 2.90 Å
R _{merge}	0.057	0.298
Total number of observations	22839	3338
Unique observations	5160	774
Multiplicity	4.4	4.5
Mean(I) / stdev(I)	18.1	5.0
Completeness	97.5%	98.0%
Resolution limits	30.0. – 2.9 Å	
R factor	23.9%	
Free R factor ^a	27.4%	
Average B factor for protein	56 Å ²	
Average B factor for DNA	59 Å ²	
Number of protein atoms in model	755	
Number of DNA atoms in model	428	
Rms deviation from ideal bond lengths	0.008 Å	
Rms deviation from ideal bond angles	1.3°	
Ramachandran statistics: ^b		
Residues in most favored regions	95.4%	
Residues in additional allowed regions	4.6%	

^a R_{free} was calculated for a randomly chosen set of 4.5% (245) of reflections, which were excluded from any refinement.

^b Considering 87 non-Gly, non-Pro protein residues in model; determined using PROCHECK (Laskowski *et al.* 1993).

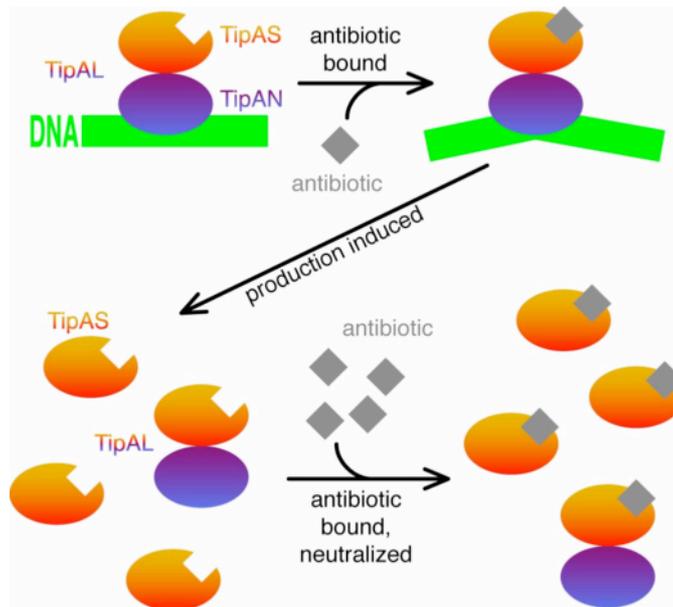


Figure 1: Cartoon of the TipA antibiotic resistance mechanism. See text for details.

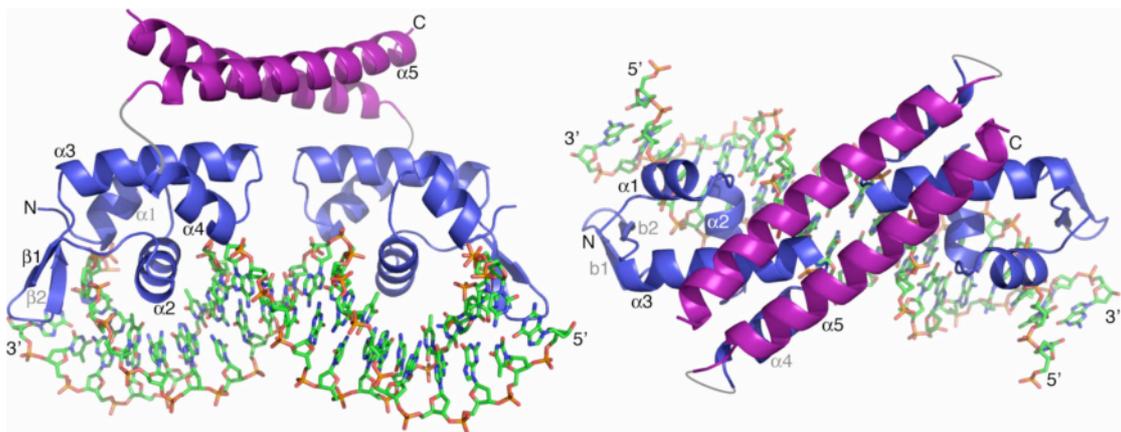


Figure 2: Overview of the TipAN·*ptipA* structure. Residues P72-P76, which were not included in the model, are represented by gray loops.

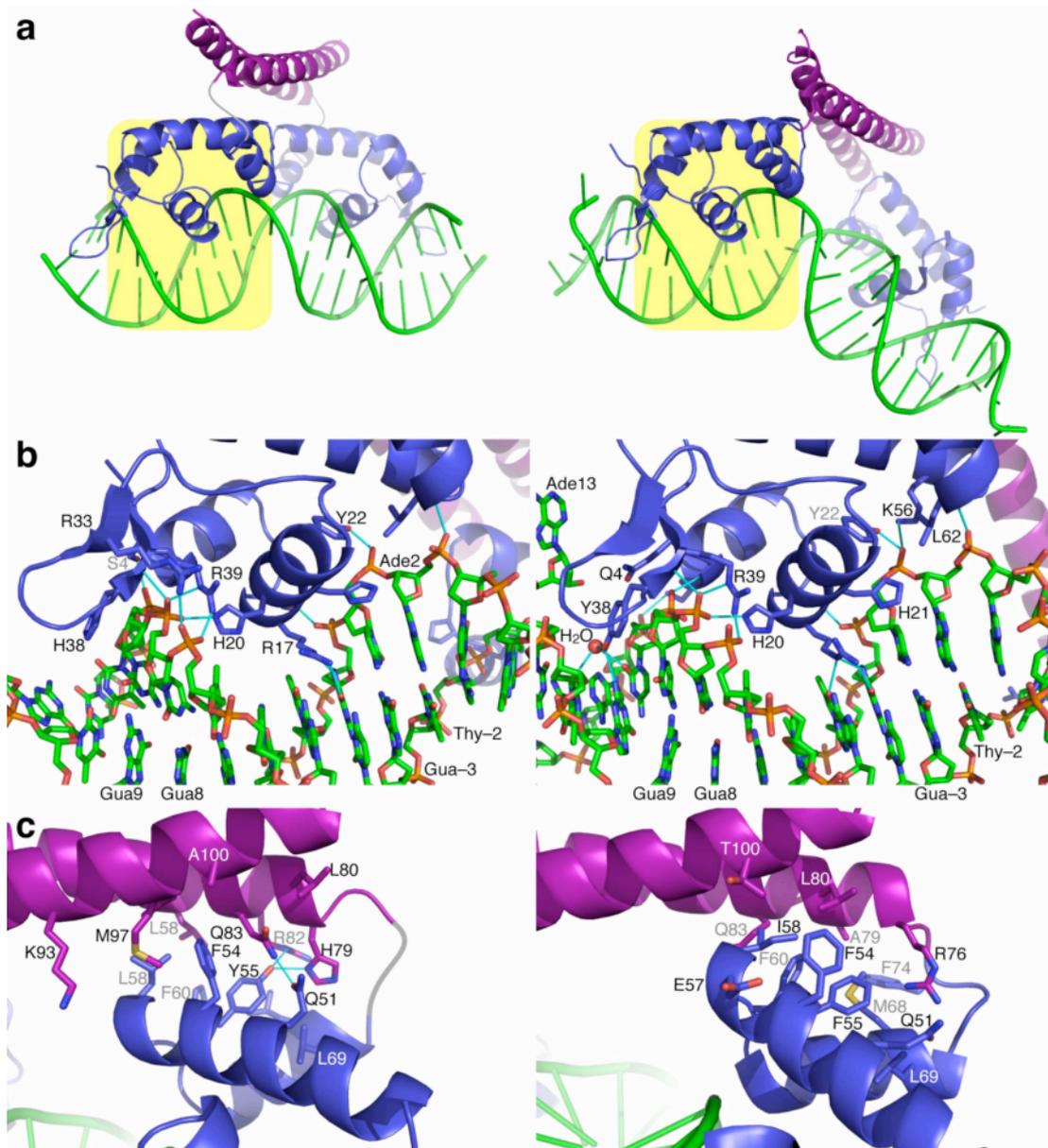


Figure 3: Comparison of the TipAN·*ptipA* (left) and MtaN·*pmta* (right) structures. **(a)** Overall structures and DNA bending. While the DNA-binding part with DNA nucleotides directly bound to it (highlighted in yellow) as well as the coiled coil (purple) are virtually identical in both structures, their position relative to each differs drastically: in TipAN·*ptipA*, the coiled coil is in closer contact with the DNA-binding part (blue), and DNA (green) is bent by 17° around the central base pair, rather than 47° as in MtaN·*pmta*. **(b)** Protein-DNA contacts are virtually identical in both complexes. **(c)** Contacts between the coiled coil and the DNA-binding part are markedly different in the two complexes. In TipAN·*ptipA*, hydrogen bonds are observed from Y55 to H79 and R82 and between Q83 and Q51, while interactions between the coiled coil and the DNA-binding part of MtaN are exclusively hydrophobic. Hydrogen bonds are shown as cyan lines.

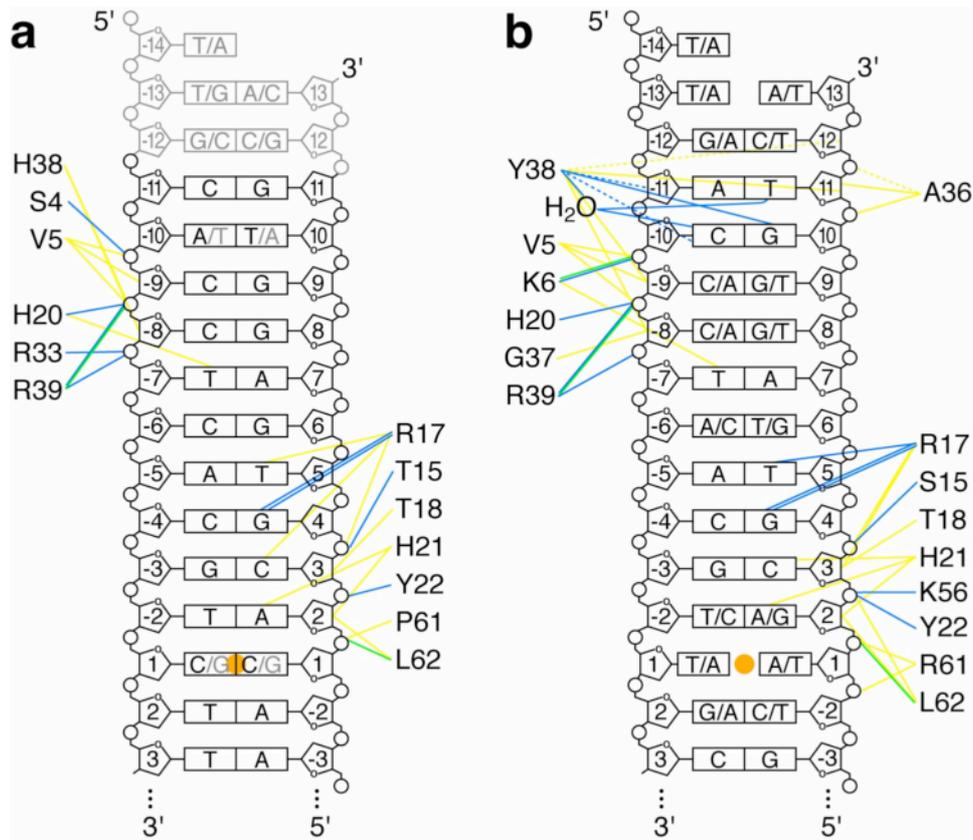


Figure 4: Schematic diagram of TipAN-DNA contacts (left) and MtaN-DNA contacts (right). Only one half-site is shown for each promoter, and the twofold pseudosymmetry axis marked is with an orange dot. Hydrogen bonds involving protein side chains are shown as blue lines; hydrogen bonds involving the protein backbone are shown as green lines; and van der Waals contacts are shown as yellow lines. Contacts that are only observed for one monomer of the MtaN dimer are shown as dashed lines. *ptipA* nucleotides that were not included in the model are shown in gray.

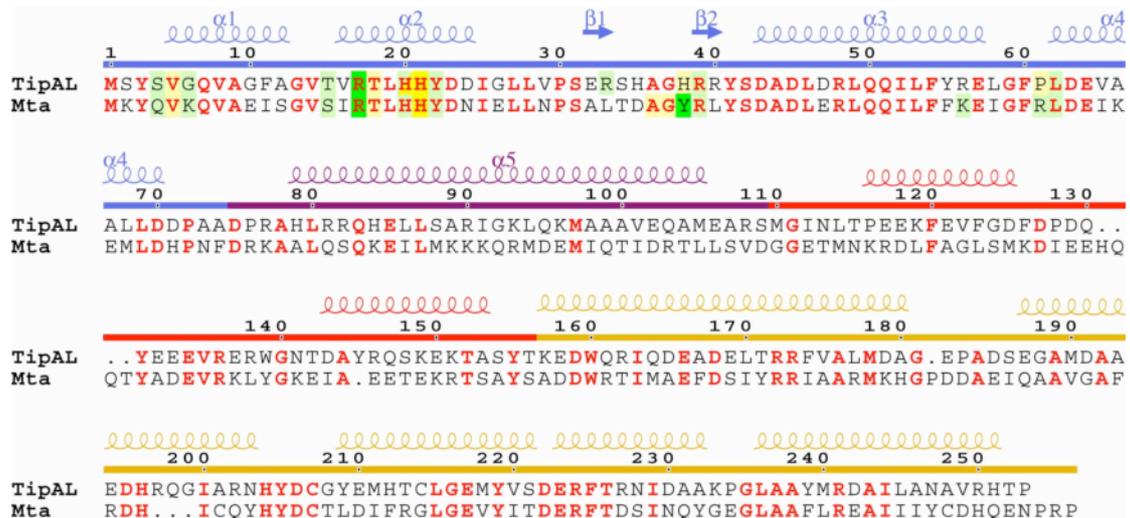


Figure 5: Sequence alignment of TipAL and Mta. Secondary structure (indicated above the sequences) was taken from the crystal structures of TipAN and 1R8D (3), and from Kahmann & *al.* (10). Amino acid residues in direct contact with DNA are highlighted in pale yellow for van der Waals contacts to DNA backbone; dark yellow for van der Waals contacts to bases; pale green for H-bonds to DNA backbone; and dark green for base-specific H-bonds. Colored bars above the sequences indicate functional units as in Fig. 8. The figure was prepared with MultAlin (18) and ESPrnt (19).

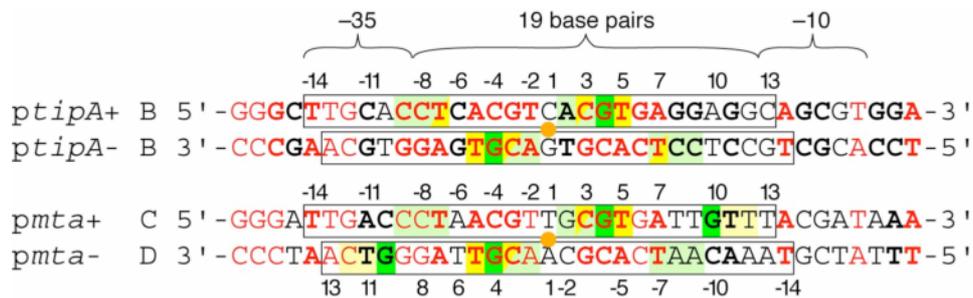


Figure 6: Promoters of *tipA* and *mta*. The -35 and -10 promoter elements recognized by RNA polymerase, and the spacing between them are indicated above the sequences. Nucleotides in direct contact with protein are highlighted in pale yellow for van der Waals contacts to DNA backbone; dark yellow for van der Waals contacts to bases; pale green for H-bonds to DNA backbone; and dark green for base-specific H-bonds. Orange dots indicate twofold pseudosymmetry axes; bases that conform to this symmetry are in boldface. Bases that are conserved in both promoters are highlighted in red. Boxes indicate the fragments that were used for TipAN·*tipA* and MtaN·*pmta* structure determination; the 5' Thy of the (-) strand was artificially changed to Ade in both cases. Residue numbers are as in crystal structures, and PDB chain identifiers are indicated to the left of the sequences.

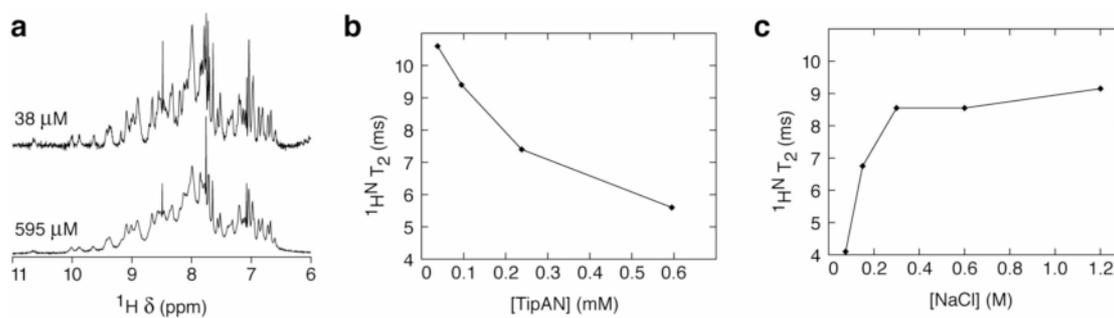


Figure 7: NMR spectroscopic analysis of TipAN in the absence of DNA, providing evidence for aggregation. **(a)** Amide/aromatic region of 1D ^1H spectra at 500 mM NaCl and 38 μM (top) or 595 μM TipAN (bottom). Acquisition times were 22 and 1.5 min. **(b)** Dependence of amide ^1H T_2 relaxation time constants on TipAN concentration, at a NaCl concentration of 500 mM. **(c)** Dependence of amide ^1H T_2 relaxation time constants on NaCl concentration, at a TipAN concentration of 100 μM .

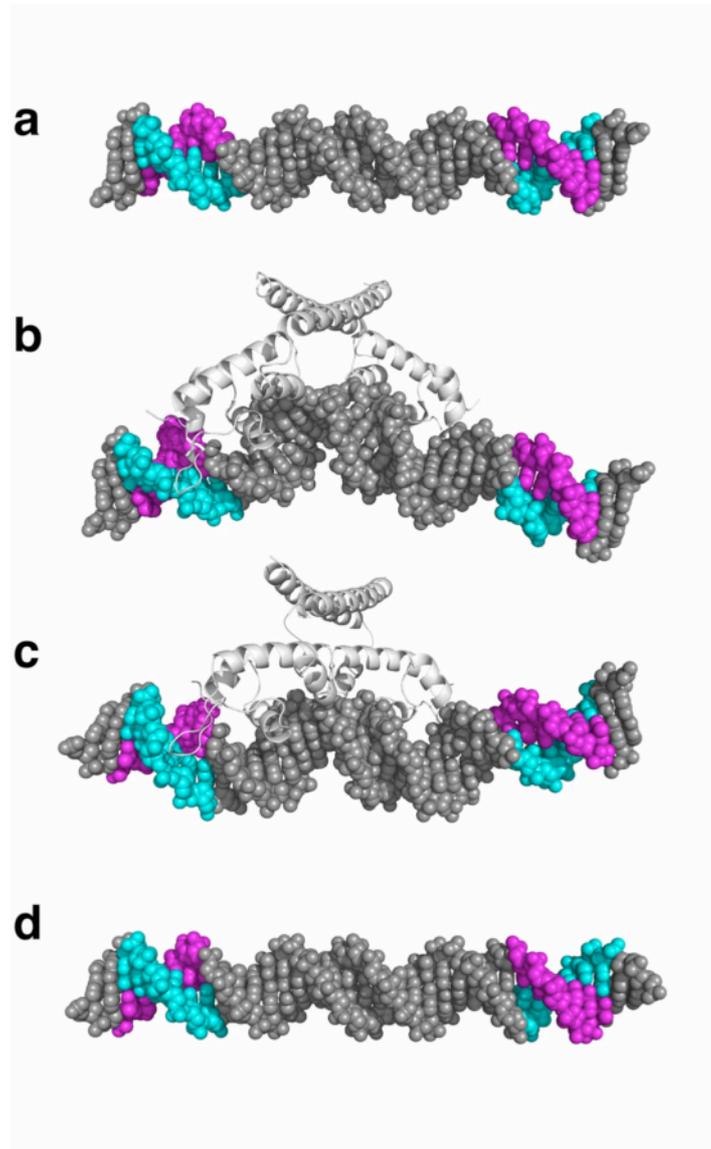


Figure 8: Models of MerR promoters in their activated and inactivated form. –35 and –10 promoter elements are highlighted in magenta (+ strand) and cyan (– strand). **(a)** Canonical σ^{70} promoter in an undistorted B-DNA conformation. –35 and –10 elements are separated by 17 base pairs, and the promoter is constitutively active. **(b)** Extrapolated *mta* promoter activated by MtaN. –35 and –10 elements are separated by 19 base pairs, but DNA is distorted in such a way that they are recognized by RNA polymerase. **(c)** Extrapolated *tipA* promoter with bound TipAN. –35 and –10 elements are separated by 19 base pairs. The slight bend in the DNA around the central base pair is overcompensated by bends in the opposite direction around bases XX, effectively withdrawing –35 and –10 elements from RNA polymerase. **(d)** MerR family promoter in an undistorted B-DNA conformation. –35 and –10 elements are separated by 19 base pairs, and not recognized by RNA polymerase because they are too far apart and twisted away from each other by 72° .

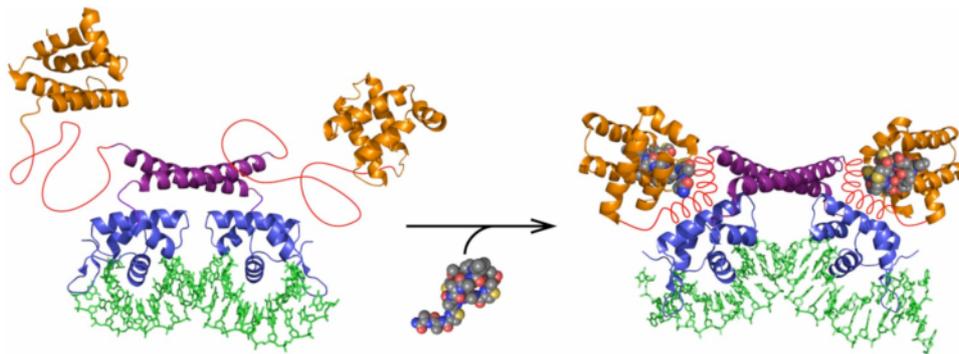


Figure 9: Cartoon of thiostrepton-induced, TipAL-mediated transcriptional activation. Upon binding of antibiotic (CPK-colored spheres) to the structured part of TipAS (orange), the previously flexible N-terminal part of TipAS (red) forms additional α -helical structure (10). This brings TipAS closer to TipAN (blue/purple), pushing the DNA-binding subdomains (blue) together. Consequently, bound promoter DNA is bent and twisted, which brings -10 and -35 promoter elements into alignment, allowing RNA polymerase to bind and initiate transcription. Figures 2, 3 and 8 were prepared with MacPyMol (20).

DgrA is a member of a new family of cyclic diguanosine monophosphate receptors and controls flagellar motor function in *Caulobacter crescentus*

Matthias Christen, Beat Christen, Martin G. Allan, Marc Folcher, Paul Jenö, Stephan Grzesiek, and Urs Jenal[†]

Biozentrum, University of Basel, Klingelbergstrasse 70, CH-4056 Basel, Switzerland

Edited by Sankar Adhya, National Institutes of Health, Bethesda, MD, and approved January 10, 2007 (received for review September 5, 2006)

Bacteria are able to switch between two mutually exclusive lifestyles, motile single cells and sedentary multicellular communities that colonize surfaces. These behavioral changes contribute to an increased fitness in structured environments and are controlled by the ubiquitous bacterial second messenger cyclic diguanosine monophosphate (c-di-GMP). In response to changing environments, fluctuating levels of c-di-GMP inversely regulate cell motility and cell surface adhesins. Although the synthesis and breakdown of c-di-GMP has been studied in detail, little is known about the downstream effector mechanisms. Using affinity chromatography, we have isolated several c-di-GMP-binding proteins from *Caulobacter crescentus*. One of these proteins, DgrA, is a PilZ homolog involved in mediating c-di-GMP-dependent control of *C. crescentus* cell motility. Biochemical and structural analysis of DgrA and homologs from *C. crescentus*, *Salmonella typhimurium*, and *Pseudomonas aeruginosa* demonstrated that this protein family represents a class of specific diguanylate receptors and suggested a general mechanism for c-di-GMP binding and signal transduction. Increased concentrations of c-di-GMP or DgrA blocked motility in *C. crescentus* by interfering with motor function rather than flagellar assembly. We present preliminary evidence implicating the flagellar motor protein FliI in DgrA-dependent cell motility control.

diguanylate receptor | motility | PilZ | c-di-GMP | second messenger

Cyclic purine nucleotides are ubiquitous second messengers involved in cell signaling. They are produced through the action of growth factors, hormones, or neurotransmitters and elicit their response by acting on a range of downstream effector proteins such as protein kinases, transcription regulators, gated ion channels, or GTPase nucleotide exchange factors. Although cAMP is widespread through all kingdoms of life, cGMP seems to be restricted to signaling in eukaryotes. Recently, a third major cyclic nucleotide messenger, cyclic diguanosine monophosphate (c-di-GMP) has emerged as a ubiquitous signaling molecule in prokaryotes, where it antagonistically controls motility and virulence of planktonic cells on one hand and cell adhesion and persistence of multicellular communities on the other (1, 2) [supporting information (SI) Fig. 7]. C-di-GMP is synthesized from two molecules of GTP and degraded into the linear dinucleotide pGpG by the opposing activities of diguanylate cyclases (DGC) and c-di-GMP-specific phosphodiesterases (PDE). DGC and PDE activities are comprised in GGDEF and EAL domains, respectively (3–8), which represent two large families of output domains found in bacterial one- and two-component signal transduction systems (9, 10).

The molecular principles of c-di-GMP signaling have been studied in the model organism *Caulobacter crescentus*, where c-di-GMP coordinates the developmental transition from a motile swarmer cell to a surface-attached, replication-competent stalked cell. Both acquisition of flagellar motility in the predivisional cell and its replacement by an adhesive organelle later in development are controlled by c-di-GMP. TipF, an EAL domain protein, is required for an early step of flagellum assembly in the predivisional cell (11), whereas the diguanylate

cyclase PleD is involved in flagellum ejection and subsequent steps in pole remodeling (3, 12–15). Similarly, the second messenger c-di-GMP regulates motility, adhesion factors, and biofilm formation in a wide variety of bacterial pathogens including *Yersinia*, *Pseudomonas*, *Vibrio*, and *Salmonella* (1, 2). C-di-GMP influences flagellar motility as a function of growth (16) or adaptation to surfaces (17), affects pilus assembly (18), and controls the production of surface structures such as fimbriae and exopolysaccharide matrices (19). The wide variety of cellular functions that are affected by c-di-GMP calls for multiple receptors and signaling mechanisms. However, little information is available on specific targets of c-di-GMP action. With the exception of a component of the cellulose synthase complex from *Gluconacetobacter* (20, 21) and the recent prediction of a candidate c-di-GMP-binding domain (22, 23), no c-di-GMP effector proteins have been reported. We have designed a biochemical approach to purify and characterize c-di-GMP effector molecules from *C. crescentus* crude cell extracts.

Results

Purification of c-di-GMP-Binding Proteins from *C. crescentus*. Based on the assumption that c-di-GMP signal transduction depends on specific receptor proteins, we designed a biochemical purification strategy to identify such components. C-di-GMP-binding proteins from *C. crescentus* were purified by two consecutive chromatography steps with Blue Sepharose CL-6B and affinity chromatography with GTP immobilized on epoxy-activated Sepharose 4B (Amersham Pharmacia, Piscataway, NJ). UV cross-linking with c-[³³P]di-GMP was used to identify proteins with specific binding activity for c-di-GMP (see *Materials and Methods* and *SI Table 2*). Two binding proteins with apparent molecular masses of 47 and 36 kDa were detected in the 0.4–0.7 M NaCl eluate of the Blue Sepharose column (Fig. 1 *A* and *B*, lane 3), and the 0.7–0.9 M NaCl fraction contained several small c-di-GMP-binding proteins with apparent molecular masses of 8–12 kDa (Fig. 1 *A* and *B*, lanes 4 and 5). The latter fraction was dialyzed, concentrated, and separated on a GTP-epoxy-Sepharose 4B affinity column (Fig. 1 *A* and *B*, lane 4). One of these proteins (labeled *c* in Fig. 1*B*) was identified by tandem

Author contributions: M.C. and B.C. contributed equally to this work; M.C., B.C., M.G.A., M.F., S.G., and U.J. designed research; M.C., B.C., M.G.A., and M.F. performed research; P.J. contributed new reagents/analytic tools; M.C., B.C., M.G.A., M.F., S.G., and U.J. analyzed data; and M.C., B.C., M.G.A., and U.J. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS direct submission.

Freely available online through the PNAS open access option.

Abbreviations: c-di-GMP, cyclic diguanosine monophosphate; DgrA, diguanylate receptor A; MS/MS, tandem mass spectrometry.

[†]To whom correspondence should be addressed at: Division of Molecular Microbiology, Biozentrum, University of Basel, Klingelbergstrasse 70, CH-4056 Basel, Switzerland. E-mail: urs.jenal@unibas.ch.

This article contains supporting information online at www.pnas.org/cgi/content/full/0607738104/DC1.

© 2007 by The National Academy of Sciences of the USA

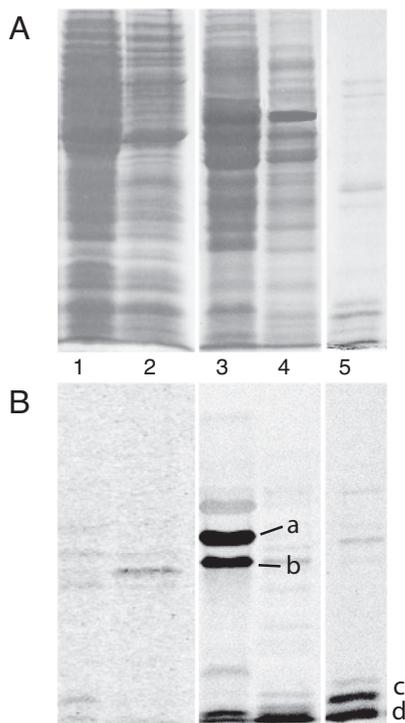


Fig. 1. Isolation of c-di-GMP-binding proteins from *C. crescentus*. (A) Coomassie blue-stained SDS/polyacrylamide gel with protein fractions used for UV cross-linking with c-[³³P]di-GMP. Lane 1, 100,000 × g supernatant; lane 2, 60% ammonium sulfate precipitation; lane 3, 0.4–0.7 M NaCl eluate from Blue Sepharose; lane 4, 0.7–0.9 M NaCl eluate from Blue Sepharose; and lane 5, 125 mM NaCl eluate from GTP-Sepharose column. (B) Autoradiograph of SDS/polyacrylamide gel shown in A. C-di-GMP-binding proteins a, b, and d were identified by MS/MS. Protein c was identified by MS/MS as hypothetical protein CC1599 and was renamed DgrA.

mass spectrometry (MS/MS) as the product of gene CC1599, a conserved hypothetical 12.5-kDa protein that we consequently renamed as diguanylate receptor A (DgrA). Sequence comparison disclosed DgrA as a member of the PilZ protein family, members of which have recently been proposed by bioinformaticists to be c-di-GMP effector proteins (22).

DgrA Is a Diguanylate Receptor Protein. To confirm that the identified protein is a c-di-GMP receptor, *dgrA*, was subcloned into the expression vector pET-42b, and the recombinant His₆-tagged protein was purified by nickel–nitrilotriacetic acid affinity chromatography. Like the semipurified protein from *C. crescentus* (Fig. 1B), the recombinant protein showed strong labeling upon UV cross-linking with c-[³³P]di-GMP (Fig. 2A), confirming that DgrA is a c-di-GMP receptor protein. UV cross-linking experiments with DgrA in the presence of 60 nM ³³P-labeled c-di-GMP and increasing concentrations of cold c-di-GMP, GTP (200 μM), or pGpG (200 μM) indicated that DgrA binds c-di-GMP with high affinity and specificity (Fig. 2B). C-di-GMP binds to DgrA in a noncovalent manner because no radiolabeled c-di-GMP was incorporated without UV irradiation (Fig. 2B). The dissociation constant for c-di-GMP of the recombinant DgrA was determined by using the UV cross-linking assay (Table 1). Saturated incorporation of radiolabeled c-di-GMP was already observed at 50 nM, indicating that the *K*_d of DgrA for c-di-GMP is <50 nM.

To test whether other members of the PilZ protein family also bind c-di-GMP, we analyzed several ortho- or paralogous of DgrA, including CC3165 (renamed as DgrB), YcgR from *Salmonella*

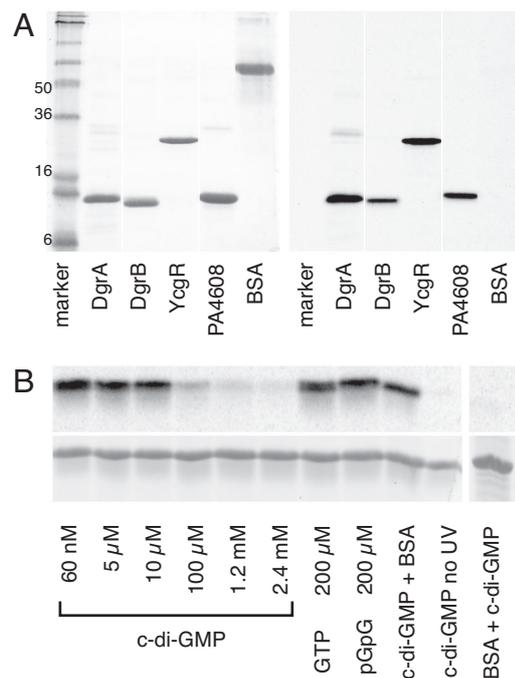


Fig. 2. DgrA is a member of a new family of c-di-GMP-binding proteins. (A) UV cross-linking of purified His₆-tagged diguanylate receptor proteins with c-[³³P]di-GMP. The following proteins were used: DgrA (CC1599; *C. crescentus*), DgrB (CC3165; *C. crescentus*), PA4608 (*P. aeruginosa*), YcgR (*S. typhimurium*), and BSA (control). The Coomassie blue-stained gel (Left) and the autoradiograph (Right) are shown. (B) UV cross-linking of 10 μM DgrA in the presence of 60 nM ³³P-labeled c-di-GMP. Autoradiograph (Upper) and Coomassie-stained gel (Lower). (Left) Samples were supplemented with increasing concentrations of nonlabeled nucleotides as indicated. (Right) Controls were carried out in the absence of UV irradiation or with BSA.

typhimurium (24), and PA4608 from *Pseudomonas aeruginosa* (see Fig. 6). As shown in Fig. 2A, all four proteins were efficiently labeled with c-[³³P]di-GMP upon UV cross-linking, whereas the control protein BSA did not incorporate c-di-GMP. The c-di-GMP-binding constants of DgrB, YcgR, and PA4608 were determined by performing UV cross-linking experiments with 50 nM receptor protein in the presence of increasing concentrations of ³³P-labeled c-di-GMP (50–1,000 nM). All wild-type diguanylate receptor proteins exhibit a binding affinity in the nanomolar range (Table 1). Taken together, these data demonstrate that DgrA and its homologs containing a PilZ domain are members of a class of small diguanylate receptor proteins, which bind c-di-GMP, but not other guanine nucleotides, with high

Table 1. Binding constants of diguanylate receptor proteins determined by UV cross-linking with c-[³³P]di-GMP

Organism	Protein	<i>K</i> _d , nM	Δ <i>K</i> _d *
<i>C. crescentus</i>	DgrB	132	36
	DgrA		
	WT	<50	14
	R11AR12A	ND	—
	D38A	761	149
<i>S. typhimurium</i>	W75A	6,200	496
	YcgR	182	29
<i>P. aeruginosa</i>	PA4608	<50	27

The protein concentration used for binding assay was 50 nM. ND, not detectable.

*S.E.

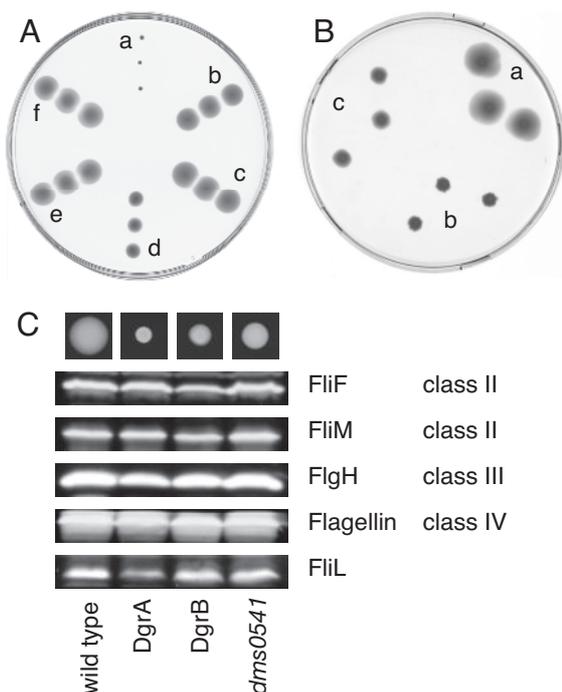


Fig. 3. DgrA and DgrB are involved in motility control by c-di-GMP. Motility behavior of *C. crescentus* wild-type strain CB15 and mutants is shown on semisolid agar plates. Three different colonies from independent conjugation experiments are shown. (A) The following strains containing plasmid pUJ142::*dgcA* or control plasmid pUJ142 were analyzed: CB15/pUJ142::*dgcA* (a), CB15 Δ *dgrA*/pUJ142::*dgcA* (b), CB15*dgrAW75A*/pUJ142::*dgcA* (c), CB15 Δ *dgrB*/pUJ142::*dgcA* (d), CB15 Δ *dgrA* Δ *dgrB*/pUJ142::*dgcA* (e), and CB15/pUJ142 (f). (B) Overexpression of *dgrA* or *dgrB* from the lactose promoter (Plac) repressed *C. crescentus* motility. (a) CB15/pBBR (vector control). (b) CB15/pBBR::*dgrA*. (c) CB15/pBBR::*dgrB*. (C) Levels of class II, class III, and class IV structural components of the *C. crescentus* flagellum were determined by immunoblot analysis for the following strains: CB15/pBBR (wild-type), CB15/pBBR::*dgrA* (DgrA), CB15/pBBR::*dgrB* (DgrB), and the extragenic diguanylate receptor motility suppressors CB15*dms0541* pBBR::*dgrA* (*dms0541*). The motility behavior of each strain is shown on top of the graph.

affinity. Thus, these proteins represent bona fide diguanylate receptor proteins and may be involved in the response of specific cell functions to fluctuating concentrations of c-di-GMP (2).

DgrA and DgrB Mediate c-di-GMP-Dependent Motility Control in *C. crescentus*. Low concentrations of c-di-GMP are generally associated with flagella- or pili-based motility of single planktonic cells, whereas increased concentrations of c-di-GMP promote multicellular traits and efficiently block cell motility (2). In agreement with this finding, *C. crescentus* cells are nonmotile in the presence of a plasmid-borne copy of *dgcA*, which encodes a highly active, soluble diguanylate cyclase (15) (Fig. 3A). Electron micrographs and immunoblot experiments showed that these cells were flagellated and expressed normal levels of flagellins (data not shown), arguing that increased c-di-GMP concentrations interfere with flagellar function rather than with the expression or assembly of flagellar components. To test whether motility control by c-di-GMP involves *dgrA* or *dgrB*, single and double in-frame deletion mutants were generated by using a two-step homologous recombination procedure (see *SI Materials and Methods*). In contrast to *C. crescentus* wild-type, Δ *dgrA* and Δ *dgrB* mutants were motile even in the presence of the *dgcA* plasmid (Fig. 3A). Motility was not the result of a reduction of the c-di-GMP concentration because cellular levels of c-di-GMP in these mutants were indistinguishably high (data not shown). At low cellular concentrations of c-di-GMP, motility phenotypes

were not significantly altered in the deletion mutants (data not shown), indicating that DgrA and DgrB affect cell motility primarily at conditions where the level of c-di-GMP is elevated. Together, these data suggested that the c-di-GMP-binding proteins DgrA and DgrB are part of a signal transduction pathway that interferes with flagellar function in response to high concentrations of c-di-GMP. In agreement with this suggestion, overexpression of *dgrA* or *dgrB* from a plasmid efficiently blocked motility on swarmer plates (Fig. 3B) and in liquid media as observed microscopically (data not shown).

Analysis of c-di-GMP Binding to the Diguanylate Receptor by NMR Spectroscopy. The available NMR structure and resonance assignments of the DgrA homolog PA4608 from *P. aeruginosa* (25) [Protein Data Bank (PDB) ID code 1YWU; Biological Magnetic Resonance Bank (BMRB) ID code 6514][‡] provided an opportunity to characterize the ligand-binding site on a molecular level and to investigate the structural consequences of ligand binding. PA4608 carrying an N-terminal His₆ tag was produced in uniformly ¹⁵N- and ¹³C-labeled form for NMR spectroscopy. The ¹H and ¹⁵N chemical shifts observed for pure PA4608 were in good agreement with those reported in BMRB entry 6514. When c-di-GMP was added to the protein, ¹H-¹⁵N heteronuclear sequential quantum correlation spectra (HSQC) changed dramatically (SI Fig. 8). Free and ligand-bound PA4608 were in slow exchange on the NMR chemical shift time scale, and titration curves were in agreement with a K_D in the sub- μ M range (data not shown). To assign resonances of the PA4608*c-di-GMP complex, exchange (EXSY) spectra were recorded on a roughly 3:1 mixture of free and c-di-GMP-bound PA4608 at 313 K. Standard triple-resonance NMR spectra recorded on PA4608 saturated with c-di-GMP were used to complete the backbone resonance assignments. No resonances were observed for residues Met³-His¹² (His₆ tag), His²², Phe³³-Ile³⁶, Gly⁷³, Ile⁹¹, Glu¹²⁵, Leu¹²⁸, and Asp¹³⁰-Leu¹³⁸. These residues are probably flexible on a microsecond to millisecond time scale, and peaks are broadened beyond detection because of intermediate chemical exchange. Secondary ¹³C α and ¹³C β shifts (26) showed that the secondary structure of PA4608 remained essentially unchanged after ligand binding (SI Fig. 9).

To localize the ligand-binding site on the protein surface, backbone amide ¹H and ¹⁵N chemical shifts of the PA4608*c-di-GMP complex were compared with those of the free protein, and the differences were mapped on the structure of the free protein (see Fig. 5 and SI Fig. 8). Large shift differences were found on one face of the β -barrel (around Val⁵⁸, Ile⁶³), in the C terminus (Val¹⁴², Ala¹⁴⁴), and in the N terminus (Arg³⁰-Asp³⁹). We conclude that c-di-GMP binds to the outside of the β -barrel close to Val⁵⁸ and that the termini, which are partially flexible in the apo form, fold around the bound ligand. Presumably the side chain N-H group of Trp⁹⁹ forms a hydrogen bond with the ligand because the ¹⁵N^{e1} and ¹H^{e1} resonances strongly shift toward higher chemical shifts by 8.24 and 1.66 ppm, respectively.

Because of their distinct chemical shifts (>10.7 ppm), the H1 imino hydrogens of guanine in c-di-GMP could be identified once the assignment of protein backbone ¹H^N and tryptophan ¹H^{e1} resonances had been completed. Because four separate H1 resonances of about equal intensity are observed for c-di-GMP in complex with PA4608 and each molecule of c-di-GMP contains two guanine bases, c-di-GMP binds to PA4608 as a dimer. Consistent with the ligand-binding site outlined above, two of these H1 imino resonances show intermolecular NOEs to Leu⁶⁴ and Trp⁹⁹ (SI Fig. 10).

[‡]Residue numbering for PA4608 as in Biological Magnetic Resonance Bank ID code 6514, which differs from that in Protein Data Bank structure 1YWU by +22, is used throughout this work.

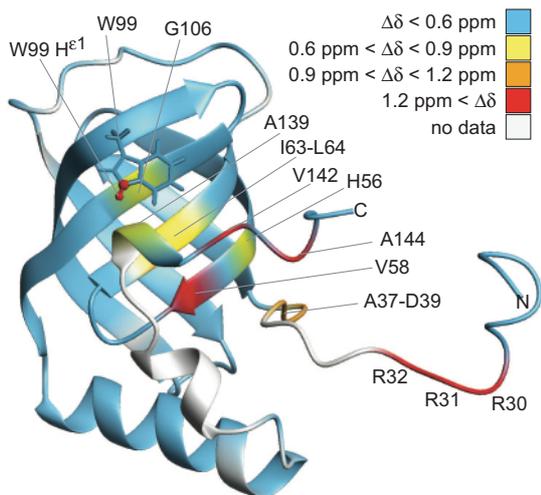


Fig. 4. Combined amide ^1H and ^{15}N shift differences ($\Delta\delta$) between PA4608 in its free and ligand-bound form. Shift differences are color-coded on the structure of free PA4608 (PDB 1YVU, model 12). Combined chemical shift differences were calculated as $\Delta\delta = \sqrt{((\Delta\delta_{\text{H}})^2 + (\Delta\delta_{\text{N}})^2)/2}$. These data are also shown in **SI Fig. 8**. Residue Trp⁹⁹ is shown as sticks, and N $^{\epsilon 1}$ and H $^{\epsilon 1}$ are shown in red to highlight the large $\Delta\delta$ value (1.67 ppm) for these atoms.

Amide ^{15}N T_1 and T_2 relaxation times and heteronuclear $\{^1\text{H}\}$ - ^{15}N NOEs were measured at 293 K for free and c-di-GMP-bound PA4608 (data not shown). Isotropic rotational correlation times (τ_c) were determined from these data with the program TENSOR (27) as 11.3 and 12.3 ns for free and ligand-bound protein, respectively. These τ_c are in reasonable agreement with values expected for monomeric apo-PA4608 (16.7 kDa, 9.8 ns) and c-di-GMP-bound PA4608 (18.1 kDa, 10.6 ns). Thus, PA4608 is a monomer before and after ligand binding.

C-di-GMP-Binding Mutants of DgrA Are Unable to Control Motility. Alignments of the amino acid sequences of PA4608, DgrA, DgrB, and YcgR revealed that the key residues that were postulated based on NMR data to be involved in c-di-GMP binding to PA4608 are conserved among other diguanylate receptor proteins (see Fig. 6). To probe the c-di-GMP-binding site of DgrA and to define the minimal requirements for c-di-GMP binding, residues Arg¹¹, Arg¹², Asp³⁸, and Trp⁷⁵ were replaced with Ala, and the mutant proteins were analyzed for c-di-GMP binding. Mutants R11A/R12A and W75A strongly reduce c-di-GMP binding, whereas mutant D38A is still able to bind c-di-GMP (see Fig. 5A). In agreement with this result, the binding constant for the D38A mutant was marginally increased to 740 nM, whereas the K_d for the W75A mutant (6.4 μM) was increased 100- to 1,000-fold compared with wild type (Table 1). Binding of c-di-GMP was completely abolished in the R11A/R12A mutant. To analyze the importance of c-di-GMP binding for DgrA-mediated signaling, the *dgrA* mutant alleles were tested for functionality *in vivo*. As indicated above, overexpression of wild-type *dgrA* renders cells nonmotile (Fig. 4B; see also Fig. 6B). In contrast, overexpression of *dgrAD38A*, *dgrAR11AR12A*, or *dgrAW75A* only partially affected motility (Fig. 5B). In particular, changing Trp⁷⁵ to Ala almost completely abolished the ability of DgrA to block motility under these conditions (Fig. 5B). Also, when the *dgrAW75A* mutant allele was expressed in single copy from its original chromosomal locus, cells were fully motile even in the presence of the *dgcA* plasmid, arguing that DgrAW75A can no longer control motility in response to increased c-di-GMP levels (Fig. 3). We isolated suppressors that alleviated the *dgrA*-mediated motility block (see *Materials and Methods*). One of the intragenic *dms* (diguanylate

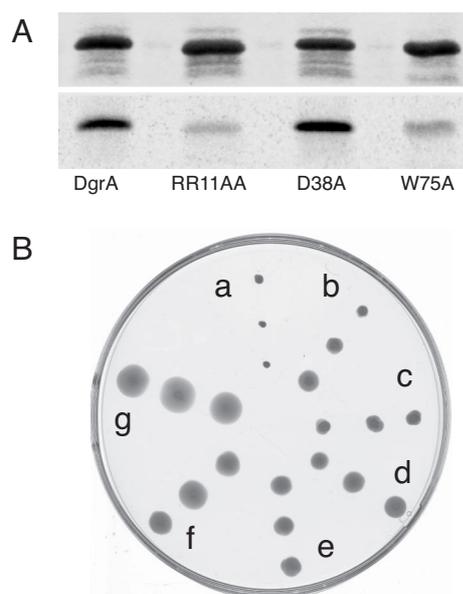


Fig. 5. C-di-GMP binding and motility control of DgrA mutants. (A) UV cross-linking of different DgrA mutant proteins with c- ^{33}P di-GMP. Coomassie blue-stained SDS/PAGE (Upper) and autoradiograph (Lower) with purified wild-type and mutant DgrA proteins (10 μM) are shown. (B) Motility behavior of *C. crescentus* wild-type CB15 overexpressing different *dgrA* alleles: CB15/pBBR::*dgrA* (a), CB15/pBBR::*dgrAR11AR12A* (b), CB15/pBBR::*dgrAD38A* (c), CB15/pBBR::*dgrAV74A* (d), CB15/pBBR::*dgrAR11AR12AV74A* (e), CB15/pBBR::*dgrAW75A* (f), and CB15/pBBR (vector control) (g). Three different colonies from independent conjugation experiments are shown.

receptor motility suppressors) mutations mapped to Val⁷⁴, in the immediate vicinity of the Trp residue critical for c-di-GMP binding (Figs. 5B and 6). Other intragenic *dms* mutations (Asp⁶², Gly⁸²) mapped to conserved residues of DgrA, emphasizing the functional importance of these residues (Fig. 6). In conclusion, these results support the view that ligand binding is essential for the regulatory function of the diguanylate receptor and suggest that DgrA blocks motility in its c-di-GMP-bound state.

Motility Control by DgrA Correlates with Cellular Levels of the FliL Motor Protein. Immunoblot analysis revealed that overexpression of *dgrA* or *dgrB* blocks motility without interfering with the expression of known class II, class III, or class IV components of the flagellar hierarchy (Fig. 3C). Because the expression of each class of genes depends on the successful expression and assembly of components of the preceding class of the hierarchy (28), this result suggested that flagella are assembled normally in cells overexpressing *dgrA* or *dgrB*. In agreement with this suggestion, flagella were readily observed by electron microscopy in these nonmotile cells (data not shown). The only flagellar protein whose concentration was severely affected in cells overexpressing *dgrA* was FliL (Fig. 3C). The *C. crescentus* *fliL* gene is not part of the flagellar hierarchy, and its product is not assembled into the flagellar structure (29). However, *fliL* is required for flagellar rotation (29). To examine whether reduced FliL levels are linked to motility, we screened the pool of *dms* mutants (see above) for extragenic suppressors (see *Materials and Methods*). From a total of 120 independently isolated motile suppressors, only one mapped to the chromosome. This suppressor mutation (*dms0541*), which mapped to gene CC3587 coding for the ribosomal protein S1, not only restored motility but also reestablished normal levels of FliL (Fig. 3C).

Discussion

Motility control by c-di-GMP is implemented through gene expression, organelle assembly, or motor function (2). In *C.*

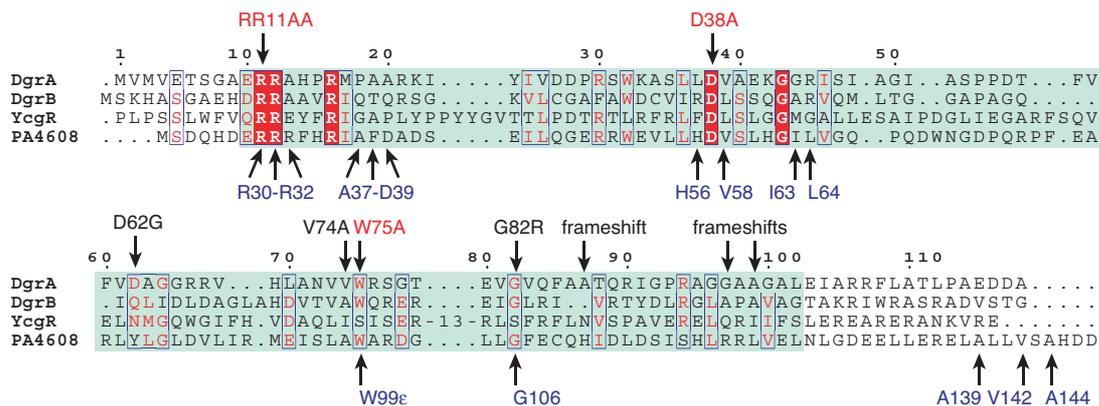


Fig. 6. Sequence alignment of the c-di-GMP-binding proteins DgrA, DgrB, YcgR, and PA4608 according to the PilZ Pfam entry PF07238. The PilZ domain is highlighted in green. DgrA residues shown to be important for c-di-GMP binding and *in vivo* function (red) and the positions of intragenic *dms* suppressor mutations (black) are highlighted above the alignment. Residues of PA4608 with large chemical shift differences upon c-di-GMP binding (blue) are indicated below the alignment.

crenscentus, increased cellular concentrations of c-di-GMP block motility by interfering with motor function rather than by altering expression or assembly of structural components of the flagellum (13). How are increased levels of c-di-GMP sensed and how is this information transmitted to the flagellar motor? The data presented here show that DgrA and DgrB are high affinity receptors for c-di-GMP that, in a ligand-bound form, interfere with the flagellar motor either directly or indirectly. Motor control by DgrA-like proteins is not unique to *Caulobacter*. *Escherichia coli* H-NS mutants lack flagella because the expression of the flagellar master control operon *flhCD* is reduced. Ectopic expression of *flhCD* restores flagellation but leaves the motors partially paralyzed (24). Under these conditions, flagellar function can be restored either by a mutation in *ycgR*, coding for the *E. coli* DgrA homolog, or by providing multiple copies of *yjhH*, which encodes a presumable c-di-GMP-specific phosphodiesterase (24). Together with our data demonstrating that the *Salmonella* YcgR protein specifically binds c-di-GMP, this finding suggests that in *C. crescentus* and in enteric bacteria flagellar motor function might be controlled by c-di-GMP through similar mechanisms.

But how would DgrA or YcgR interfere with the function of the flagellum? Our data propose the FliL protein as a candidate for such a role. FliL was the only flagellar protein that showed significantly reduced levels in nonmotile cells overexpressing *dgrA*. In *C. crescentus* the FliL protein is not part of the flagellar structure but is required for flagellar rotation (29). Intriguingly, *fliL* mutant strains exhibit a motility phenotype identical to that of cells that have high levels of c-di-GMP or overexpress *dgrA* (29). Because the expression of *fliM*, the gene located immediately downstream from *fliL* in the same operon (30), was not affected by DgrA, FliL changes must be the result of altered translation or protein stability. An extragenic suppressor mutation that restored motility under these conditions also reestablished normal FliL concentrations, indicating that the two phenotypes are linked. The simplest model that is in agreement with these results predicts that DgrA, upon binding of c-di-GMP, represses FliL by a so far unknown mechanism, and through this mechanism blocks motor function. The extragenic suppressor mutation restoring FliL levels was mapped to the coding region of *rpsA* (*ribosomal protein S1*). RpsA enhances translation initiation by binding to mRNA regions upstream from the Shine-Dalgarno sequence and by tethering the mRNAs on the 30S subunit of the ribosome (31–33). How DgrA and its ligand c-di-GMP modulate FliL levels is a subject for future investigation. Recently, FliL was reported to be involved in surface

sensing and virulence gene expression in the urinary tract pathogen *Proteus mirabilis* (34). Thus, it is possible that FliL has a more general role in controlling the switch between a planktonic and a surface-associated lifestyle.

A bioinformatics study originally proposed that the PilZ domain is a specific c-di-GMP-binding module (22). This proposal was recently substantiated by the demonstration that YcgR, a PilZ protein from *E. coli*, is able to bind c-di-GMP (23). Here we presented genetic, biochemical, and structural evidence that further validates this hypothesis, and we propose a model for ligand binding and activation of proteins containing a PilZ domain. NMR studies with the DgrA homolog PA4608 showed that a dimer of c-di-GMP binds to a well defined binding site on the surface of the β -barrel (Fig. 4). Large chemical shift differences between free and ligand-bound PA4608, which indicate changes in the local environment, were also observed in both termini of the protein, with the largest differences observed for residues Arg³⁰–Arg³², Val¹⁴², and Ala¹⁴⁴. These regions are structurally ill defined in the absence of ligand (25) and are probably flexible. The observed chemical shift differences indicate that these regions come in direct contact with the ligand after complex formation. The N-terminal part of PA4608 contains three consecutive Arg residues, which are conserved in most PilZ domains (22) (Fig. 6). Arg side chains are likely to be involved in hydrogen bonds or in electrostatic or π stacking interactions with c-di-GMP, as in the allosteric binding site of the diguanylate cyclases PleD and DgcA (15, 35). Furthermore, it is conceivable that the positively charged head groups of Arg are sufficient for transient binding to the phosphate groups of c-di-GMP and that their position on the flexible N terminus increases the ligand capture radius of the protein, as in the “fly-casting mechanism” proposed in ref. 36. Alternatively, the observed folding of previously flexible parts of the protein may be responsible for communication of the c-di-GMP signal to downstream elements, either by forming new interaction surfaces or by determining the relative position of neighboring domains. Similarly, the chemical shift differences of the C-terminal part of PA4608 could be explained by a specific role in ligand binding. However, the fact that residues Val¹⁴² and Ala¹⁴⁴, which showed the largest chemical shift differences, are not conserved, argues against this possibility. Several of the motile *dgrA* loss-of-function suppressors that were isolated had frameshift mutations in the very C terminus of DgrA (Fig. 6), suggesting that this part of the protein is critical for its *in vivo* function. One possibility is that the C terminus contributes to the specific readout mechanism of this protein family. Upon c-di-GMP binding to the β -barrel surface, the C terminus could be untied to interact with downstream components. In accordance

with such a view, the very C terminus of the *P. aeruginosa* PilZ protein has recently been proposed to interact with the PilF protein required for type 4 pilus assembly (37). To complement our picture of the c-di-GMP circuitry, future studies will have to focus on interaction partners of DgrA and related PilZ domain proteins.

It is intriguing that genetic and biochemical studies of the *C. crescentus* DgrA protein and structural analysis of PA4608 from *P. aeruginosa* identified the same set of key amino acids involved in c-di-GMP binding (Fig. 6). This finding is a strong indication that these proteins bind c-di-GMP in a similar way and suggests that they may share a common signaling mechanism. Based on these results, we postulate that most or all PilZ domain proteins function as diguanylate receptor proteins.

Materials and Methods

Strains, Plasmids, and Media. *E. coli* strains were grown in Luria broth. *C. crescentus* strains were grown in complex peptone yeast extract (38) supplemented with antibiotics, where necessary. For the exact procedure of strain and plasmid construction, see *SI Materials and Methods*.

UV Cross-Linking with C-[³³P]di-GMP and Isolation of DgrA. Procedures for enzymatic production of c-[³³P]di-GMP and UV cross-linking with c-[³³P]di-GMP were published earlier (6, 15). For a detailed protocol used for the isolation of DgrA, see *SI Materials and Methods*.

Preparation of Isotope-Labeled Protein, NMR Samples, and NMR Spectroscopy. The detailed procedures for overexpression and ¹³C, ¹⁵N labeling of PA4608 are described in *SI Materials and*

Methods. NMR samples (Shigemi microtubes) were prepared as 0.8 mM U-¹³C, ¹⁵N-labeled protein in 300 μl of 95% H₂O/5% D₂O/250 mM NaCl/10 mM DTT/1 mM Na₂S₂O₃/10 mM Tris at pH 7.1. C-di-GMP was added at suitable molar ratios from a 7.7 mM stock solution. NMR spectra were recorded on Bruker (Billerica, MA) DRX 600 and 800 MHz spectrometers at 293 K with the exception of EXSY spectra that were recorded at 313 K for faster exchange. Standard one-, two-, and three-dimensional spectra were recorded and processed as described elsewhere (39).

Isolation and Mapping of Motile *dgrA* Suppressors. A plasmid carrying *dgrA* (pBBR::*dgrA*) was conjugated into a *C. crescentus* *recA* mutant strain, and 150 individual transconjugants were patched onto PYE swarmer plates. Motile *dms* mutants were isolated and analyzed by immunoblot with an α-DgrA antibody. Mutants with reduced DgrA levels were discarded. The rest was analyzed by retransforming plasmids into the *recA* mutant strain to distinguish between intra- and extragenic suppressors. Intragenic mutations were identified by sequencing. The extragenic suppressor (*dms0541*) was mapped by *Tn5* linkage (40) and cotransduction with phage ΦCR30, and it was identified by sequencing.

We thank Suzette Moes for handling of MS samples, Martha Gerber and Flora Mauch for help with the suppressor screen, Jacob Malone (University of Basel) for a gift of *P. aeruginosa* DNA, and Colin Manoil (University of Washington, Seattle, WA) for providing plasmid pIT2. This work was supported by Swiss National Science Foundation Fellowships 3100A0-108186 (to U.J.) and 31-109712 (to S.G.).

1. Kolter R, Greenberg EP (2006) *Nature* 441:300–302.
2. Jenal U, Malone J (2006) *Annu Rev Genet* 40:385–407.
3. Paul R, Weiser S, Amiot NC, Chan C, Schirmer T, Giese B, Jenal U (2004) *Genes Dev* 18:715–727.
4. Ryjenkov DA, Tarutina M, Moskvina OV, Gomelsky M (2005) *J Bacteriol* 187:1792–1798.
5. Schmidt AJ, Ryjenkov DA, Gomelsky M (2005) *J Bacteriol* 187:4774–4781.
6. Christen B, Christen M, Folcher M, Schauer A, Jenal U (2005) *J Biol Chem* 280:30829–30837.
7. Bobrov AG, Kirillina O, Perry RD (2005) *FEMS Microbiol Lett* 247:123–130.
8. Tamayo R, Tischler AD, Camilli A (2005) *J Biol Chem* 280:33324–33330.
9. Galperin MY, Nikolskaya AN, Koonin EV (2001) *FEMS Microbiol Lett* 203:11–21.
10. Ulrich LE, Koonin EV, Zhulin IB (2005) *Trends Microbiol* 13:52–56.
11. Huitema E, Pritchard S, Matteson D, Radhakrishnan SK, Viollier PH (2006) *Cell* 124:1025–1037.
12. Aldridge P, Jenal U (1999) *Mol Microbiol* 32:379–391.
13. Aldridge P, Paul R, Goymer P, Rainey P, Jenal U (2003) *Mol Microbiol* 47:1695–1708.
14. Levi A, Jenal U (2006) *J Bacteriol* 188:5315–5318.
15. Christen B, Christen M, Paul R, Schmid F, Folcher M, Jenoe P, Meuwly M, Jenal U (2006) *J Biol Chem* 281:32015–32024.
16. Choy WK, Zhou L, Syn CK, Zhang LH, Swarup S (2004) *J Bacteriol* 186:7221–7228.
17. Boles BR, McCarter LL (2002) *J Bacteriol* 184:5946–5954.
18. Kazmierczak BI, Lebron MB, Murray TS (2006) *Mol Microbiol* 60:1026–1043.
19. Kader A, Simm R, Gerstel U, Morr M, Romling U (2006) *Mol Microbiol* 60:602–616.
20. Ross P, Mayer R, Weinhouse H, Amikam D, Huggiratt Y, Benziman M, de Vroom E, Fidler A, de Paus P, Sliedregt LA, et al. (1990) *J Biol Chem* 265:18933–18943.
21. Weinhouse H, Sapir S, Amikam D, Shilo Y, Volman G, Ohana P, Benziman M (1997) *FEBS Lett* 416:207–211.
22. Amikam D, Galperin MY (2006) *Bioinformatics* 22:3–6.
23. Ryjenkov DA, Simm R, Romling U, Gomelsky M (2006) *J Biol Chem* 281:30310–30314.
24. Ko M, Park C (2000) *J Mol Biol* 303:371–382.
25. Ramelot TA, Yee A, Cort JR, Semesi A, Arrowsmith CH, Kennedy MA (2006) *Proteins* 66:266–271.
26. Spera S, Bax A (1991) *J Am Chem Soc* 113:5490–5492.
27. Dosset P, Hus J-C, Blackledge M, Marion D (2000) *J Biomol NMR* 16:23–28.
28. Guber JW, E JC (2000) in *Prokaryotic Development*, eds Shimkets LJ, Brun YVB (Am Soc Microbiol, Washington, DC), pp 319–339.
29. Jenal U, White J, Shapiro L (1994) *J Mol Biol* 243:227–244.
30. Yu J, Shapiro L (1992) *J Bacteriol* 174:3327–3338.
31. Sengupta J, Agrawal RK, Frank J (2001) *Proc Natl Acad Sci USA* 98:11991–11996.
32. Komarova AV, Tchufistova LS, Supina EV, Boni IV (2002) *RNA* 8:1137–1147.
33. Sorensen MA, Fricke J, Pedersen S (1998) *J Mol Biol* 280:561–569.
34. Belas R, Suvanasuthi R (2005) *J Bacteriol* 187:6789–6803.
35. Chan C, Paul R, Samoray D, Amiot NC, Giese B, Jenal U, Schirmer T (2004) *Proc Natl Acad Sci USA* 101:17084–17089.
36. Shoemaker BA, Portman JJ, Wolyne PG (2000) *Proc Natl Acad Sci USA* 97:8868–8873.
37. Kim K, Oh J, Han D, Kim EE, Lee B, Kim Y (2006) *Biochem Biophys Res Commun* 340:1028–1038.
38. Ely B (1991) *Methods Enzymol* 204:372–384.
39. Grzesiek S, Bax A, Hu JS, Kaufman J, Palmer I, Stahl SJ, Tjandra N, Wingfield PT (1997) *Protein Sci* 6:1248–1263.
40. Jacobs MA, Alwood A, Thaipisuttikul I, Spencer D, Haugen E, Ernst S, Will O, Kaul R, Raymond C, Levy R, et al. (2003) *Proc Natl Acad Sci USA* 100:14339–14344.

Corrections

DEVELOPMENTAL BIOLOGY. For the article “*DMY* gene induces male development in genetically female (XX) medaka fish,” by Masaru Matsuda, Ai Shinomiya, Masato Kinoshita, Aya Suzuki, Tohru Kobayashi, Bindhu Paul-Prasanth, En-lieng Lau, Satoshi Hamaguchi, Mitsuru Sakaizumi, and Yoshitaka Nagahama, which appeared in issue 10, March 6, 2007, of *Proc Natl Acad Sci*

USA (104:3865–3870; first published February 28, 2007; 10.1073/pnas.0611707104), due to a printer’s error, in Table 4, column 3, the number of *DMRT1* embryos hatched was given as “30”; it should instead be “80.” Also, in Table 4, column 6, the number of *DMY* transgenic XX females was given as “–4”; it should instead be “4.” The corrected table appears below.

Table 4. CMV promoter-directed overexpression of *DMY* and *DMRT1*

cDNA	Embryos injected	Embryos hatched	Hatching rate, %	XY male	XX female	XX male
<i>DMY</i>	195	86	44.1	20	4	4
<i>DMRT1</i>	139	80	57.6	30	13	0

www.pnas.org/cgi/doi/10.1073/pnas.0702707104

MEDICAL SCIENCES. For the article “The omega-3 fatty acid docosahexaenoate attenuates endothelial cyclooxygenase-2 induction through both NADP(H) oxidase and PKC ϵ inhibition,” by Marika Massaro, Aida Habib, Laura Lubrano, Serena Del Turco, Guido Lazzerini, Todd Bourcier, Babette B. Weksler, and Raffaele De Caterina, which appeared in issue 41, October 10, 2006, of *Proc Natl Acad Sci USA* (103:15184–15189; first published October 3, 2006; 10.1073/pnas.0510086103), the authors note that the following should be added as an affiliation for Marika Massaro: Department of Biological and Environmental Sciences and Technologies, University of Lecce, Ecotekne, 73100 Lecce, Italy. The corrected author and affiliation lines, and related footnotes, appear below.

Marika Massaro^{*}, Aida Habib^{§5}, Laura Lubrano[¶], Serena Del Turco[¶], Guido Lazzerini[¶], Todd Bourcier[¶], Babette B. Weksler^{**}, and Raffaele De Caterina^{†††}**

*Institute of Clinical Physiology, Consiglio Nazionale delle Ricerche, 73100 Lecce, Italy; †Department of Biological and Environmental Sciences and Technologies, University of Lecce, Ecotekne, 73100 Lecce, Italy;

§Departments of Biochemistry and Internal Medicine, American University of Beirut, Beirut 1107 2020, Lebanon; ¶Institute of Clinical Physiology, Consiglio Nazionale delle Ricerche, 56124 Pisa, Italy; ¶Department of Anesthesia, Brigham and Women’s Hospital–Harvard Medical School, Boston, MA 02115; **Weill Medical College of Cornell University, New York, NY 10021; and ††Institute of Cardiology and Center of Excellence on Aging, “Gabriele d’Annunzio” University, 66100 Chieti, Italy

[†]M.M. and A.H. contributed equally to this work.

^{††}To whom correspondence should be addressed. E-mail: rdecater@unich.it.

www.pnas.org/cgi/doi/10.1073/pnas.0702704104

MICROBIOLOGY. For the article “Specific polysaccharide conjugate vaccine-induced IgG antibodies prevent invasion of *Shigella* into Caco-2 cells and may be curative,” by Yehuda Chowers, Joachim Kirschner, Nathan Keller, Iris Barshack, Simon Bar-Meir, Shai Ashkenazi, Rachel Schneerson, John Robbins, and Justen H. Passwell, which appeared in issue 7, February 13, 2007, of *Proc Natl Acad Sci USA* (104:2396–2401; first published February 7, 2007; 10.1073/pnas.0610833104), the authors note an error in the title. The title should read: “O-specific polysaccharide conjugate vaccine-induced antibodies prevent invasion of *Shigella* into Caco-2 cells and may be curative.” The online version has been corrected. Additionally, on page 2396, in the first line of the Abstract, in the first line of the second paragraph of the main text, and in the Abbreviations footnote, “O-specific polysaccharide” should instead appear as “O-specific polysaccharide.”

www.pnas.org/cgi/doi/10.1073/pnas.0702299104

MICROBIOLOGY. For the article “DgrA is a member of a new family of cyclic diguanosine monophosphate receptors and controls flagellar motor function in *Caulobacter crescentus*,” by Matthias Christen, Beat Christen, Martin G. Allan, Marc Folcher, Paul Jenö, Stephan Grzesiek, and Urs Jenal, which appeared in issue 10, March 6, 2007, of *Proc Natl Acad Sci USA* (104:4112–4117; first published February 27, 2007; 10.1073/pnas.0607738104), on page 4112, column 1, the key term “c-di-Gmp” should instead appear as “c-di-GMP.” The online version has been corrected.

www.pnas.org/cgi/doi/10.1073/pnas.0702705104

Christen *et al.* 10.1073/pnas.0607738104.

Supporting Information

Files in this Data Supplement:

[SI Figure 7](#)

[SI Table 2](#)

[SI Materials and Methods](#)

[SI Figure 8](#)

[SI Figure 9](#)

[SI Figure 10](#)

[SI Figure 7](#)

Fig. 7. Schematic of c-di-GMP signaling. (A) Chemical structure of the bacterial second messenger c-di-GMP. (B) Synthesis, hydrolysis, and effector mechanism of c-di-GMP. The GGDEF domain harboring the diguanylate cyclase activity (red), the EAL domain responsible for degradation of the second messenger into the linear dinucleotide pGpG (blue), and the newly identified diguanylate receptor DGR (green) are indicated. Dashed lines symbolize product inhibition of diguanylate cyclases by c-di-GMP (3) and allosteric activation of c-di-GMP-specific phosphodiesterases by GTP (2).

[SI Figure 8](#)

Fig. 8. Changes in ^1H - ^{15}N heteronuclear sequential quantum correlation

spectra of PA4606 upon addition of c-di-GMP. Combined ^1H and ^{15}N shift differences ($\Delta\delta$) between PA4608 in its free and ligand-bound form, plotted against the amino acid sequence, are shown.

This Article

▶ [Abstract](#)

▶ [Full Text](#)

Services

▶ [Email this article to a colleague](#)

▶ [Alert me to new issues of the journal](#)

▶ [Request Copyright Permission](#)

Combined chemical shift differences were calculated as $= \sqrt{[(\Delta\delta_H)^2 + (\Delta\delta_N/5)^2]/2}$. These data are also shown in Fig. 4 with the color codes matching the horizontal lines indicated here.

SI Figure 9

Fig. 9. Secondary $^{13}\text{C}\alpha$ and $^{13}\text{C}\beta$ shifts for PA4680 with and without bound c-di-GMP. Secondary shifts are defined as the difference between observed chemical shifts and chemical shifts of the same atom in the context of a random coil. Secondary shifts for $^{13}\text{C}\alpha$ and $^{13}\text{C}\beta$ atoms in proteins ($\Delta\text{C}\alpha$, $\Delta\text{C}\beta$) are mainly determined by secondary structure. In α -helices, $\Delta\text{C}\alpha = 3.09 \pm 1.00$ ppm, and $\Delta\text{C}\beta = -0.38 \pm 0.85$ ppm; in β -sheets, $\Delta\text{C}\alpha = -1.48 \pm 1.23$ ppm, and $\Delta\text{C}\beta = 2.16 \pm 1.91$ ppm (5). In coils, $\Delta\text{C}\alpha = \Delta\text{C}\beta = 0$ ppm by definition. The amino acid sequence of the PA4608 construct used in this work is shown with an N-terminal His tag at the top. $\Delta\text{C}\alpha$ and $\Delta\text{C}\beta$ values and secondary structures are indicated for the free protein (1YWU) and for the ligand bound form. Differences between the two states of the protein are minimal.

SI Figure 10

Fig. 10. Excerpt from a two-dimensional NOESY spectrum of PA4608 with bound c-di-GMP. The spectral region shown contains ^1H atoms with extraordinarily high chemical shifts on the horizontal axis. Peaks are labeled where assignments are known; ^1H belonging to PA4608 are labeled in black, whereas ^1H belonging to c-di-GMP are labeled in red.

Table 2. Purification of c-di-GMP binding proteins from *C. crescentus* CB15 crude extracts

Sample	Total yield of protein		Total yield of binding activity, %	Purification, fold
	Weight, μg	Percent		
100,000 μg supernatant	680,000	100	100	1
60 % $(\text{NH}_4)_2\text{SO}_4$ precipitation	323,000	47.5	95	2
Blue Sepharose 0.4-0.7 M	10,800	1.59	40	26

NaCl	10,000	1.57	40	20
Blue Sepharose 0.7-0.9 M NaCl	7,800	1.15	41	35
GTP Sepharose 125 mM NaCl	24	0.0035	11	3,200

SI Materials and Methods

Construction of Strains and Plasmids. *In-frame deletions of dgrA and dgrB in Caulobacter crescentus.* Seven hundred base pairs flanking the 3' and 5' DNA regions of both *dgrA* and *dgrB* were PCR-amplified from *C. crescentus* CB15 genomic DNA using primers 1464, 1465, 1466, 1467, 1472, 1473, 1474, and 1475 (see list below). The PCR products were digested with SpeI and HindIII or with HindIII and NheI and were subcloned in a triple ligation reaction into the SpeI and NheI sites of vector pNPTS138 to give pNPTS::*dgrA* 3-115 and pNPTS::*dgrB* 3-113. Strains with in-frame deletions in *dgrA* or *dgrB* were constructed by using the suicide plasmids pNPTS::*dgrA* 3-115 or pNPTS::*dgrB* 3-113 and a double-recombination procedure as described (1). The deletion mutants were verified by PCR.

Chromosomal replacement of wild-type *dgrA* with *dgrAW75A*. The *dgrA* wild-type gene flanked by a 700-bp 3' and 5' DNA region was PCR amplified from *C. crescentus* CB15 genomic DNA using primers 1465 and 1466 and cloned into pGEM-T-easy. The point mutations changing Trp⁷⁵ to Ala were introduced by a SOE PCR using primers 528, 529, 1585, and 1586. The SpeI- and NheI-digested SOE PCR product was cloned into the SpeI and NheI sites of vector pNPTS138 to give pNPTS::*dgrAW75A*. The suicide plasmid pNPTS::*dgrAW75A* was then conjugated into a CB15 *dgrA* strain, and double-recombinants were selected as described (1). Strains expressing a single copy of *dgrAW75A* under the native promoter were screened by PCR and the W75A mutation was confirmed by genomic sequencing.

Plasmids for the overexpression of *dgrA* and *dgrB* in *C. crescentus*. The *dgrA* and *dgrB* genes were PCR-amplified from *C. crescentus* CB15 genomic DNA by using primer pairs 1460, 1461 and 1468, 1469. The PCR products were digested with KpnI and BamHI and subcloned into the vector pBBRMCS2 to give pBBR::*dgrA* or pBBR::*dgrB*, respectively. Plasmids for the overexpression of *dgrA*, *dgrB*, PA4608, and *ycgR* in *Escherichia coli* *dgrA* and *dgrB* were PCR-amplified from *C. crescentus* CB15 genomic DNA using primer pairs 1462, 1463 and 1470, 1471, digested with NdeI and XhoI, and cloned into vector pET42b (Novagen, Madison, WI) to give pET42B::*dgrA* and pET42b::*dgrB*. PA4806 was PCR-amplified from *Pseudomonas aeruginosa* PAO1 genomic DNA using primers 1476 and 1477, and it was digested with NdeI and BamHI and cloned into vector pET15b to produce pET15::PA4608. *ycgR* was amplified from *Salmonella enterica* serovar Typhimurium genomic DNA with primers 1525 and 1526. The PCR product was digested with NdeI and XhoI and cloned into vector pET42b to generate pET42b::*ycgR*.

Plasmid for the overexpression of *dgcA* (CC3285) under *Pxyl* control in *C. crescentus*. *dgcA* was PCR-amplified from *C. crescentus* CB15 genomic DNA by using primers 1081 and 1082. The PCR product was digested with EcoRI and KpnI and subcloned into vector pUJ142 behind the xylose-inducible promoter *Pxyl*, resulting in construct pUJ142::*dgcA*.

Site-directed mutagenesis of *dgrA*. Primers 1461, 1463, and 1527 were used to replace the *dgrA* wild-

type allele with the *dgrAD38A* allele in plasmids pBBR::*dgrA* or pET42b::*dgrA*. An StuI-BamHI fragment or an StuI-XhoI fragment of *dgrA* was replaced to give pBBR::*dgrAD38A* and pET42b::*dgrAD38A*. Primers 1461, 1463, and 1528 were used to PCR-amplify the *dgrA* RR11AA allele, which was introduced into the vectors pBBRMCS2 and pET42b via KpnI and BamHI or NdeI and XhoI sites to generate pBBR::*dgrAR11AR12A* and pET42b::*dgrAR11AR12A*. Primer pairs 1585, 1463 and 1586, 1462 were used to introduce the W75A point mutation in *dgrA* by SOE PCR. The PCR product was digested with NdeI and XhoI and subcloned into vector pET42b to produce pET42b::*dgrAW75A*. The *dgrAW75A* allele was PCR-amplified using primers 1460 and 1461, digested with KpnI and BamHI, and subcloned into pBBRMCS2 to give pBBR::*dgrAW75A*.

Primer List.

528 GAC GTT GTA AAA CGA CGG CCA GT.

529 TTT CAC ACA GGA AAC ACG TAT GAC.

1081 AAA GGT ACC TCA AGC GCT CCT GC.

1082 AAA GAA TTC ATG AAA ATC TCA GGC G.

1460 AAG GTA CCG ATG GTC ATG GTC GAG ACC T.

1461 TTG GAT CCT TAG TGA TGA TGA TGA TGA TGG GCG TCG TCT TCG GCG GGC AAA
GTA GCG AGA AAC CGA.

1462 AAC ATA TGG TCA TGG TCG AGA CCT.

1463 TTG GAT CCT TAC TCG AGG GCG TCG TCT TCG GCG GGC AAA GTA GCG AGA AAC
CGA.

1464 AGT AAG CTT GAC GAC GCC TAG AGC GCC TTT CGA TCA GA.

1465 GGC TAG CAG ACC TTC ATC AAC GAA CAC GAC CTC G.

1466 AAC TAG TGT GGA CAG CGC CAG CGT GTG GCT CTG.

1467 CTA AGC TTG ACC ATG ACC ATC CGC TTC TCA AAC C.

1468 AAG GTA CCG ATG TCG AAG CAC GCG TCT GGC GCG GAA CAC G.

1469 TTG GAT CCT TAG TGA TGA TGA TGA TGA TGG CCT GTG GAG ACG TCC GCC CGT
GAC GC.

1470 TTC ATA TGT CGA AGC ACG CGT CTG GCG CGG AAC ACG.

1471 TTG GAT CCT TAC TCG AGG CCT GTG GAG ACG TCC GCC CGT GAC GC.

1472 AGT AAG CTT TCC ACA GGC TGA CGC CCG AGA.

1473 GGC TAG CGC ACT CGC GCC TTG CGG GCG TCG AT.

1474 TAC TAG TTG CCC ACC CTC GCT GAC TAC G.

1475 CTA AGC TTG TGC TTC GAC ATG GGT CGA TAA ACG CAC.

1476 TCC ATA TGA GTG ACC AGC ACG ACG AAC.

1477 TTG GAT CCT TAG TCG TCG TGG GCG CTG ACC AGC.

1525 CCATATGAGTGGTTACAATGAGCAGTTCC.

1526 CCT CGA GTT CTC GCA CTT TAT TCG CTC TTT C.

1527 AGG CCT CTC TGC TGG CGG TCG CCG AGA AGG GCG GGC GGA TCT CCA T.

1528 CAT ATG GTC ATG GTC GAG ACT AGT GGC GCT GAG GCA GCC GCG CAT CCC C.

1585 ATC TCG CCA ACG TCG TCG CAC GCT CCG GGA CCG AGG TC.

1586 ACC TCG GTC CCG GAG CGT GCG ACG ACG TTG GCG AGA T.

Lysis and Ammonium Sulfate Precipitation. A total of 12.44 g of *C. crescentus* CB15 cell pellet was resuspended in 110 ml of ice-cold 1' PBS buffer A containing 10% (vol/vol) glycerol, 1 mM DTT, 1 tablet COMPETE protease inhibitor (Roche, Indianapolis, IN), and sonicated for 5 min on ice. The lysate was centrifuged at 10,000 $\times g$ for 20 min at 4°C, and the supernatant was harvested and centrifuged at 100,000 $\times g$ for 60 min. The soluble protein fraction was saturated to 30% (NH₄)₂SO₄ at 4°C, and the 30% precipitate was removed by an additional centrifugation step at 100,000 $\times g$ for 1 h at 4°C. The pellet was discarded, and the remaining supernatant was saturated to 60% (NH₄)₂SO₄. The precipitate was collected by centrifugation at 100,000 $\times g$ for 1 h at 4°C, resuspended in 50 ml of buffer A, and dialyzed overnight using a SpectraPor dialysis membrane cutoff of 8 kDa (Spectrum, Rancho Dominguez, CA).

Synthesis of GTP-Sepharose 4B Column. A total of 550 mg of epoxy-activated Sepharose 4B (Sigma-Aldrich, St. Louis, MO) was washed with 100 ml of H₂O for 30 min. The matrix was resuspended in 2.5 ml of 0.5 mM NaOH, and 36.4 mg of GTP (Amersham Pharmacia, Piscataway, NJ) was added. After the reaction mixture was shaken for 72 h at 25°C, the Sepharose was collected by centrifugation. UV absorption measurements at 254 nm before and after the reaction revealed that 9.92 mg of GTP was bound to the matrix. The matrix was blocked with 20 ml of 1 M ethanolamine (pH 8.0) for 1 h at 25°C. Unbound GTP and ethanolamine was removed by washing with three cycles of 10 ml of H₂O/10 ml of acetate buffer, pH 4.0/500 mM NaCl/10 ml of Tris buffer, pH 8.0/500 mM NaCl.

Blue Sepharose CL-6B Column. Fifty milliliters of resuspended and dialyzed 60% (NH₄)₂SO₄ precipitate containing 323 mg of protein (6.63 mg/ml) was loaded on a Blue Sepharose CL-6B column (Amersham Pharmacia) (flow, 2 ml/min; 30 ml resin) with buffer A as loading buffer [1' PBS buffer A containing 10% (vol/vol) glycerol/1 mM DTT]. Protein (151 mg) was collected in the flow-through, and an additional 160 mg was eluted during washing with 200 ml of buffer A. Bound proteins were eluted with a gradient of 50-1,000 mM NaCl at a flow rate of 3 ml/min. The following fractions were pooled, dialyzed overnight in buffer A, and concentrated to a final protein concentration of 1 mg/ml by using a 50-ml Amicon ultrafiltration cell and regenerated cellulose ultrafiltration membrane NMWL 10,000 (Millipore Corporation, Billerica, MA): weak binders (3.3 mg of protein, fractions 7-11), medium binders (10.8 mg of protein, fraction 12-20), and tight binders (7.8 mg of protein, fractions 21-29).

GTP-Sepharose 4B Column. Three milliliters of the concentrated tight-binder fraction from the Blue Sepharose column was loaded on a 500-ml GTP-Sepharose 4B affinity resin in a 15-ml spin column and incubated for 5 min at 4°C. The loading was removed by centrifugation at 800 rpm in a tabletop

centrifuge for 2 min, and the flow-through was collected and stored on ice. The column was washed four times with 2 ml of buffer A, and the bound proteins were eluted by sequentially adding 2 ml of buffer A with 125, 250, 500, and 1,000 mM NaCl. Fractions eluted from the GTP-Sepharose 4B column were washed with buffer A in 1-mm centrifugal filter devices (Millipore Corporation) and concentrated to a final volume of 250 ml.

³³P-Labeled c-di-GMP. ³³P-labeled c-di-GMP was produced enzymatically using [³³P]GTP (3,000 Ci/mmol) and purified as described earlier (2, 3). Nickel-nitrilotriacetic acid-purified protein samples or *C. crescentus* CB15 protein preparations were incubated for 10 min on ice in PBS buffer supplemented with 1 mM DTT/1 mM c-di-GMP/³³P- radiolabeled c-di-GMP (0.75 mCi, 6,000 Ci/mmol). Samples were irradiated at 254 nm for 20 min in an ice-cooled, Parafilm-wrapped, 96-well aluminum block in an RPR-100 photochemical reactor with a UV lamp RPR-3500 (Southern New England Ultraviolet Company, Branford, CT). After irradiation, samples were mixed with 2' SDS/polyacrylamide sample buffer [250 mM Tris·HCl, pH 6.8/40% (vol/vol) glycerol/8% SDS/b2-mercaptoethanol/0.06% bromophenol blue/40 mM EDTA] and heated for 5 min at 95°C. Labeled proteins were separated by SDS/PAGE using 12.5% acrylamide gels or Tris/Tricine gradient gels and quantified by autoradiography. Bands corresponding to c-di-GMP-binding proteins were cut from the gels and were identified by MS/MS using a LTQ Orbitrap mass spectrometer (Finnigan, San José, CA). The nucleotide composition of the sample was analyzed by PEI-cellulose chromatography to exclude enzymatic cleavage of c-di-GMP during the UV cross-linking procedure.

Determination of the c-di-GMP-Binding Constant for Diguanylate Receptor Proteins. Binding constants were determined by performing UV cross-linking experiments with 50 ml of 50 nM His tag-purified diguanylate receptor protein and 0.02 mg/ml BSA in the presence of 0, 50, 100, 200, 375, 500, 750, and 1,000 nM ³³P-labeled c-di-GMP. Samples were separated by SDS/PAGE and stained with colloidal Coomassie blue. The gel was dried and exposed on a storage PhosphorImaging screen (Molecular Dynamics, Eugene, OR). Incorporation was measured by quantifying the intensities of the relevant spots using ImageJ software version 1.33 (National Institutes of Health, Bethesda, MD). Incorporation was plotted against the c-di-GMP concentration, and K_D values were determined with the Software Pro Fit 5.6.7.

Preparation of Isotope-Labeled Protein. *E. coli* strain BL21 (Novagen) carrying the pET15::PA4608 expression plasmid was grown in 3 liters of minimal medium containing 1.5 g of ¹⁵NH₄Cl and 4.8 g of unlabeled or ¹³C-labeled glucose as sole nitrogen and carbon sources. Protein expression was induced with 0.4 mM IPTG for 3 h. PA4608 was purified by nickel-nitrilotriacetic acid affinity chromatography (Qiagen, Valencia, CA) followed by size-exclusion chromatography using an Amersham Pharmacia HiLoad 26/60 Superdex 75 column.

NMR Samples and NMR Spectroscopy. NMR samples (Shigemi microtubes) were prepared as 0.8 mM U-¹³C/¹⁵N-labeled protein in 300 μl of 95% H₂O/5% D₂O with 250 mM NaCl/10 mM DTT/1 mM sodium azide/10 mM Tris at pH 7.1. C-di-GMP was added at suitable molar ratios from a 7.7 mM stock solution. NMR spectra were recorded on Bruker DRX 600 and 800 MHz spectrometers at 293 K (20°C) with the exception of EXSY spectra that were recorded at 313 K for faster exchange. Standard one-, two-, and three-dimensional spectra were recorded and processed as described elsewhere (4).

1. Stephens C, Reisenauer A, Wright R, Shapiro L (1996) *Proc Natl Acad Sci USA* 93:1210-1214.
2. Christen M, Christen B, Folcher M, Schauerte A, Jenal U (2005) *J Biol Chem* 280:30829-30837.

3. Christen B, Christen M, Paul R, Schmid F, Folcher M, Jenoe P, Meuwly M, **Jenal** U (2006) *J Biol Chem* 281:32015-32024.
4. Grzesiek S, Bax A, Hu JS, Kaufman J, Palmer I, Stahl SJ, Tjandra N, Wingfield PT (1997) *Protein Sci* 6:1248-1263.
5. Spera S, Bax, A (1991) *J Am Chem Soc* 113:5490-5492.

This Article

- ▶ [Abstract](#)
- ▶ [Full Text](#)

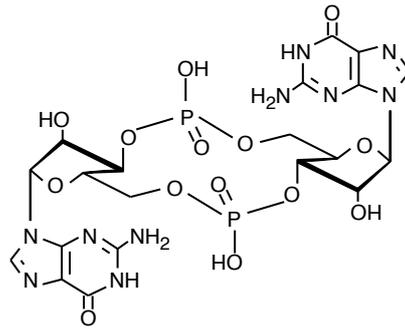
Services

- ▶ [Email this article to a colleague](#)
- ▶ [Alert me to new issues of the journal](#)
- ▶ [Request Copyright Permission](#)

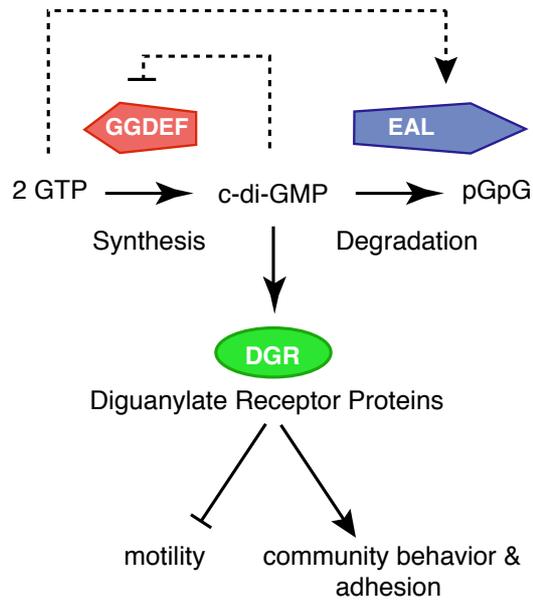
[Current Issue](#) | [Archives](#) | [Online Submission](#) | [Info for Authors](#) | [Editorial Board](#) | [About](#)
[Subscribe](#) | [Advertise](#) | [Contact](#) | [Site Map](#)

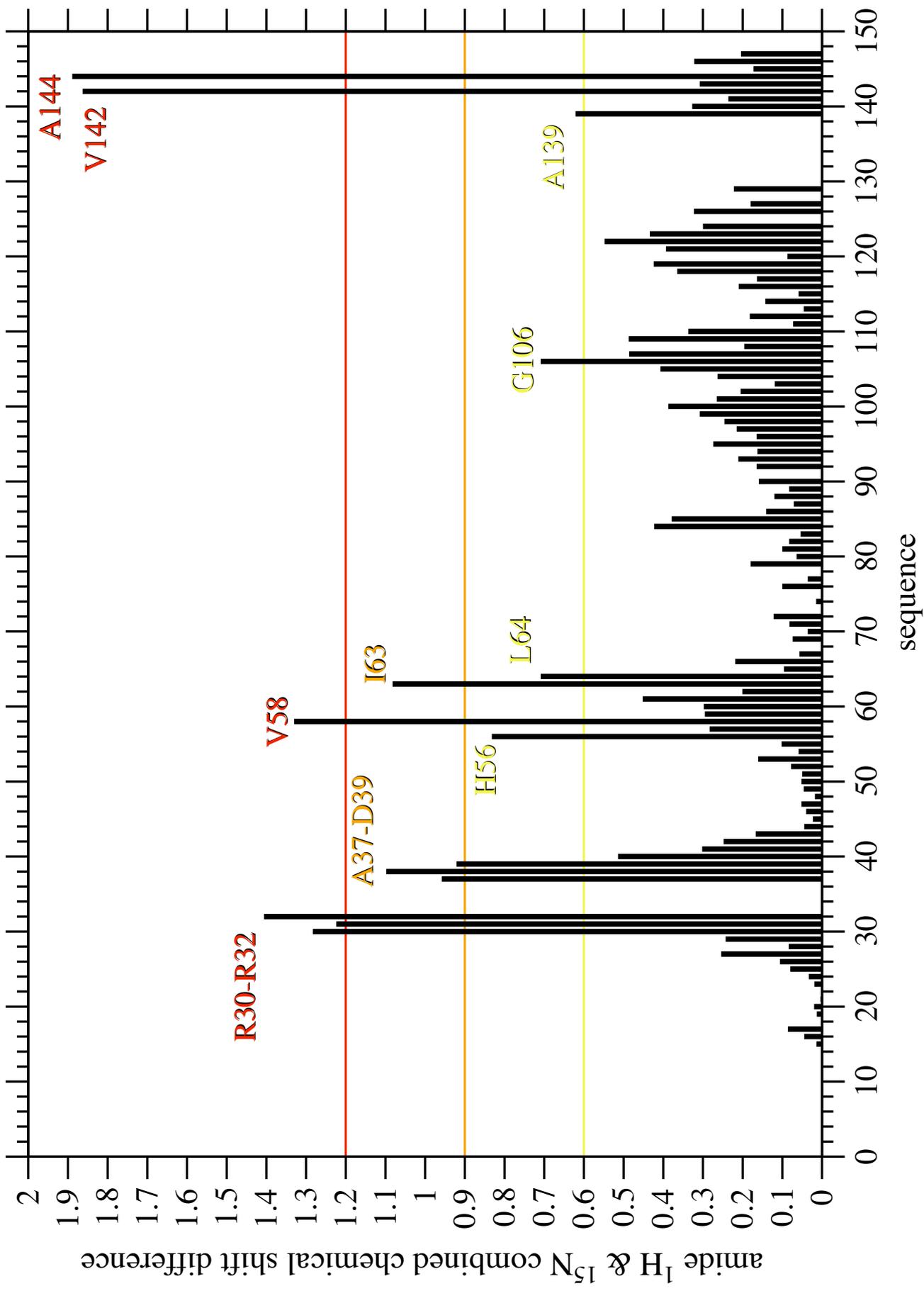
[Copyright © 2007 by the National Academy of Sciences](#)

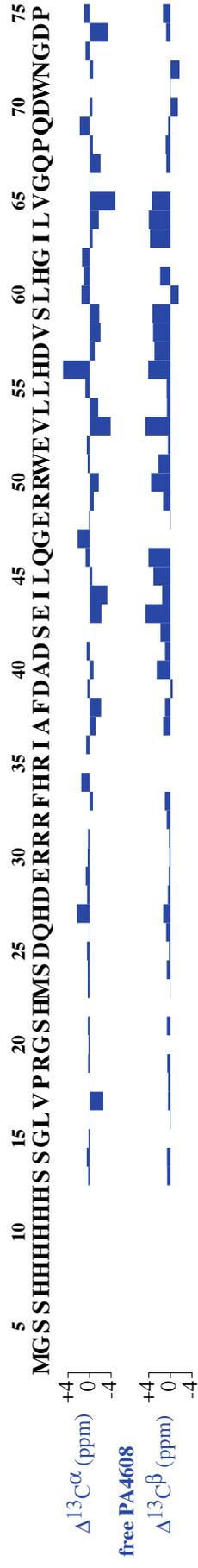
A



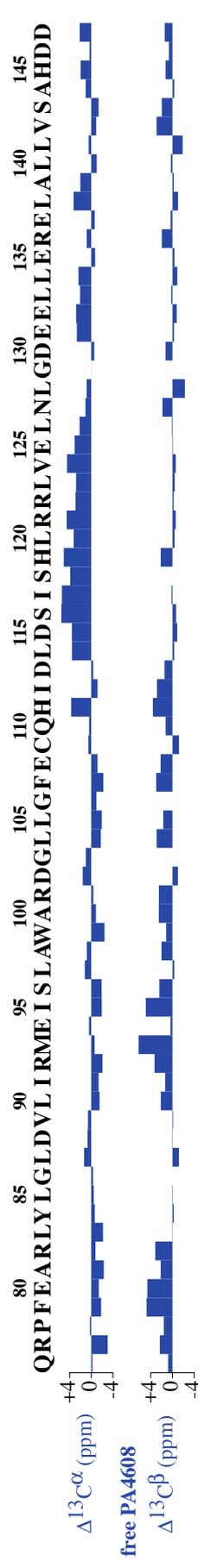
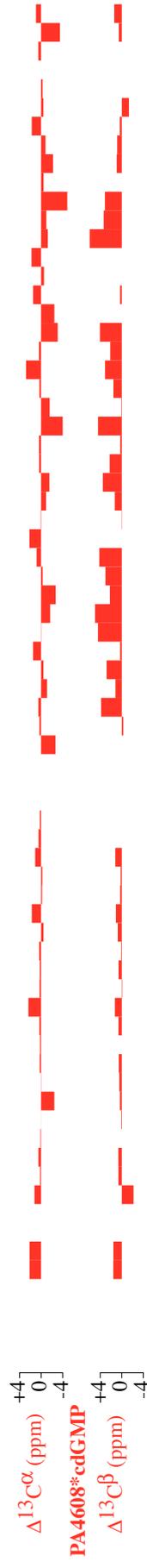
B



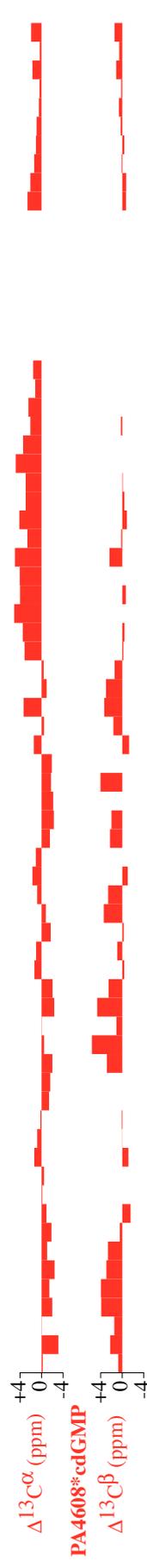




secondary structure from 1YWU



secondary structure from 1YWU



Solution structure of the diguanylate receptor PA4608 with bound c-di-GMP*

Martin G. Allan, Matthias Christen, Beat Christen, Regula Aregger, Urs Jenal and
Stephan Grzesiek

Biozentrum, Universität Basel, Switzerland

Abstract

Cyclic di-guanosine monophosphate (c-di-GMP) is a bacterial second messenger molecule that controls the switch from a planktonic, single-cell lifestyle to surface-attached, multicellular communities such as biofilms. The natural receptors of c-di-GMP have recently been found to be members of a domain family named PilZ. We have solved the solution structure of the PilZ domain protein PA4608 in complex with its ligand c-di-GMP by NMR spectroscopy. C-di-GMP binds to the side of the PA4608 β -barrel as an intercalated dimer. The previously flexible, N-terminal RxxxR motif of the protein contacts the bound ligand, which in turn displaces two C-terminal 3_{10} helices. We propose that this movement of the termini of the protein forms the basis of the molecular mechanism by which ligand binding elicits downstream responses.

* *Author contributions:* MGA prepared NMR samples, recorded, processed and evaluated NMR spectra, calculated the structure, and wrote the report. MC, BC, RA and UJ produced and purified isotope-labeled c-di-GMP and PA4608. SG supervised the structure determination.

Introduction

Cyclic di-guanosine monophosphate (c-di-GMP) is a second messenger molecule that appears to be ubiquitous in, and also unique to, the bacterial kingdom. Elevated intracellular levels of c-di-GMP generally cause bacteria to change from single, motile cells to surface-attached multicellular communities. In *Gluconacetobacter xylinum* (formerly *Acetobacter xylinum*), c-di-GMP activates cellulose synthase;¹ in *Caulobacter crescentus*, it is involved in the transition from flagellated swarmer cells to surface-attached stalked cells;² in *Vibrio cholerae*, it regulates biofilm formation and the expression of virulence genes;³ and in *Pseudomonas aeruginosa*, it is required for alginate biosynthesis.⁴

C-di-GMP consists of two guanosine moieties that are symmetrically connected by 5'-3' phosphodiester bonds, forming a 12-membered ring (Figure 1a). C-di-GMP has been shown to be extraordinarily polymorphic: five different solution structures ranging from monomer to octamer have been described.⁵ An X-ray crystal structure of c-di-GMP⁶ reveals a compact, self-intercalated dimer structure that is stabilized by base stacking and two hydrogen bonds between a H1 atom of one monomer, and a phosphate oxygen of the other (Figure 1b). A very similar self-intercalated structure was also observed in the crystal structure of PleD, which contained two allosterically bound dimers of c-di-GMP besides the extended, monomeric c-di-GMP molecule bound to the active site⁷ (PDB 1W25).

C-di-GMP is produced from GTP by diguanylate cyclase (DGC) enzymes containing GGDEF domains, and degraded to pGpG by phosphodiesterases (PDE) containing EAL domains. One or several of each of these domains are found in virtually all bacteria.

The long elusive natural receptor proteins of c-di-GMP were recently identified *in silico*. Based on its context in multidomain proteins including proteins known to be regulated by c-di-GMP, and on its occurrence in genomes that also contain GGDEF and EAL domain proteins, the PilZ domain was proposed as a c-di-GMP receptor.⁸ Biochemical studies on overexpressed *Escherichia coli* proteins confirmed that PilZ proteins indeed bind c-di-GMP with high affinity.⁹ Independently, c-di-GMP receptors were isolated from *C. crescentus* using biochemical methods, by probing for c-di-GMP binding, and identified by mass spectrometry.¹⁰ C-di-GMP binding to PilZ domains has since been shown by several groups.^{3,11}

The structures of two PilZ proteins had been solved by structural genomics initiatives before their function became known. The crystal structure of VCA0042 from *Vibrio cholerae*¹² (PDB 1YLN) reveals two six-stranded β -barrel domains, the C-terminal one of which shows good sequence homology to other PilZ domains. The solution

structure of PA4608 from *P. aeruginosa*¹¹ (PDB 1YWU) contains a single PilZ domain. Its N-terminal part comprising residues M23-R35, which contain the signature RxxxR motif found in the vast majority of PilZ domains, is dynamic and disordered.

We here present the solution structure of PA4608 with bound c-di-GMP.

Materials and methods

Production and purification of PA4608

Cloning and transformation. The *PA4608* gene was amplified from *P. aeruginosa* strain PAO1 genomic DNA by PCR, digested with the restriction enzymes NdeI and XhoI, and cloned into vector pET15b. A hexahistidine tag followed by a short linker including a thrombin cleavage site (MGSSHHHHHSSGLVPRGSH) precede the PA4608 sequence. The pET15::PA4608 plasmid was transformed into *E. coli* strain BL21 by electroporation, and the resulting strain was named BC1015. Bacteria were grown in 1.15 L of a modified M9 medium containing 0.5 g ¹⁵NH₃ and 2.0 g ¹³C₆-glucose as sole sources of nitrogen and glucose, respectively; the composition was 100 mL 10 × phosphate buffer (130 g KH₂PO₄, 100 g K₂HPO₄, 227 g Na₂HPO₄ · 12 H₂O, ad 1 L H₂O, pH adjusted to 7.2); 0.45 g ¹⁵NH₄Cl; 1.8 g ¹³C₆-glucose; 4.75 mL 1M MgCl₂; 4.75 mL trace elements mixture [41 mM CaCl₂, 22 mM FeSO₄, 5.8 mM MnCl₂, 5.8 mM CoCl₂; 2.4 mM ZnSO₄, 1.8 mM CuCl₂, 0.32 mM H₃BO₃, 0.20 mM (NH₄)₂Mo₂O₇, 12 mM EDTA]; 100 mg ampicillin sodium salt; ad 1 L H₂O. Cells were grown to an OD_{600nm} of 0.8 in shaking flasks at 37°C. PA4608 expression was induced by addition of IPTG to a final concentration of 0.4 mM, and maintained for 2 h at 37°C. Cells were harvested by centrifugation at 4000 rpm and 4°C for 20 min. The wet weight of the cell pellet was approximately 8 g. Cells were resuspended in about 30 mL Phosphate Buffered Saline (PBS; 80 g NaCl, 2.0 g KCl, 14.4 g Na₂HPO₄, 2.4 g KH₂PO₄, ad 1 L H₂O, pH about 7.4) and stored at -20°C until further processing.

Cell rupture & fractionation. The cell pellet was thawed, and cells were disrupted using a FA-031 FRENCH pressure cell (Spectronic Unicam) at approximately 1200 bar. This and all subsequent steps were carried out at 4°C. The suspension was centrifuged for 18 min at 6400 rpm, and the pellet was discarded. The supernatant was again centrifuged at 30000 rpm for 48 min, and the pellet discarded.

Nickel affinity chromatography. Cell lysate was incubated with 2 mL Ni-NTA Superflow resin (Qiagen) for 1 h. Two 15 mL batch columns (Promega) were loaded with the mixture, and packed by centrifugation at 3200 rpm. Each column was washed with 10 mL PBS, followed by 10 mL PBS containing 10 mM imidazol and 10

mL PBS containing 50 mM imidazol. The protein was then eluted with 8 mL PBS containing 250 mM imidazol per column. 2 mL fractions were collected, and their protein content was assessed by SDS-PAGE using the Amersham PhastGel system: small amounts of eluate were mixed with loading buffer (10% (v/v) glycerol, 5% (v/v) β -mercaptoethanol, 3% (w/v) SDS, 130 mM Tris-HCl at pH 6.8, 0.2 g/L bromophenol blue) at a ratio of 1:1 and separated on homogenous 20% polyacrylamide gels, which were stained with Coomassie Brilliant Blue. Fractions containing PA4608 protein were pooled and concentrated to a total volume of 8 mL using a 20 mL Vivaspin concentrator.

Size exclusion chromatography. The elution buffer used for size exclusion chromatography contained: 10 mM Tris-HCl at pH 7.0; 250 mM NaCl; 10 mM dithiothreitol (DTT). The protein solution was loaded on a HiLoad 26/60 Superdex 75 size exclusion column connected to an FPLC system (Amersham), and eluted at a flow rate of 2 mL/min. 5 mL fractions were collected, and UV absorption of the eluate at 280 nm was monitored. PA4608 eluted as a single, large peak at an elution volume of 190-210 mL.

Assessment of purity and yield. The homogeneity of the protein solution was tested by SDS-PAGE as described above. No contaminant protein bands were detected. A ^1H - ^{15}N -HSQC NMR spectrum of the preparation was recorded, and was found to be in good agreement with previously recorded reference spectra; no significant contaminant peaks were detected. The extinction coefficient of PA4608 was estimated as 17990 AU/M/cm, and the molecular mass was calculated as 16.737 kDa, with the ExPASy tool.¹³ UV absorption spectra of a 1:10 dilution of the protein solution in water were recorded, and the total protein yield was determined to be 23.4 mg.

Production and purification of c-di-GMP

C-di-GMP was produced enzymatically from GTP, using a not feedback-inhibited mutant of the diguanylate cyclase DgcA named DgcA0646, in which the inhibition site RESD was replaced with GRDC.¹⁴

Production and purification of DgcA0646. *E. coli* strain BL21 carrying the plasmid pET42b::dgcA0646 was grown at 37°C in 1.5 L LB medium containing 0.6 mM kanamycin in shaking flasks. Protein expression was induced at OD_{600nm} 0.9 by addition of IPTG to a final concentration of 0.4 mM, and maintained for 2 h. Cells were harvested by centrifugation and disrupted using a type FA-031 French press (Spectronic Unicam). Cell debris was removed by centrifugation, and the supernatant was incubated with 800 μL Ni-NTA Superflow resin (Qiagen) at 4°C for 1 h. 15 mL batch columns (Promega) were loaded with the mixture, and packed by centrifugation at 3200 rpm. Each column was washed with 10 mL PBS, followed by 10 mL PBS containing 10 mM imidazol and 10 mL PBS containing 50 mM imidazol. The protein

was eluted with 8 mL PBS containing 250 mM imidazol. 2 mL fractions were collected, and their protein content was assessed by SDS-PAGE. The protein yield was assessed using the Bio-Rad protein assay; 6 mg were obtained.

Production and purification of c-di-GMP. 6 mL 1 mM DgcA0646 was incubated with 0.6 mL 1 mM GTP, 50 mM MgCl₂, 20 mM Tris-HCl (pH 8) at 30°C for 1.5 h. The mixture was centrifuged at 13000 rpm for 5 min to remove precipitated enzyme, and 1 mL fresh 1 mM DgcA0646 was added to the supernatant. The mixture was incubated at room temperature overnight. Protein was removed by addition of 150 µL Ni-NTA Superflow resin (Qiagen) and incubation at 4°C for 20 min, followed by centrifugation. The supernatant was loaded on a C18 reverse phase HPLC column, and eluted with a 25 ml gradient from 5 to 30% methanol in 250 mM triethylammonium carbonate (TEAC), pH 7 in water. The flow rate was 1 mL/min. C-di-GMP eluted after 20.2 min as a single, sharp peak, and was collected based on UV absorption at 254 nm. The c-di-GMP solution was lyophilized at <1 mbar. In order to remove residual TEAC, c-di-GMP was redissolved in 600 µL H₂O and lyophilized again. The yield was measured by UV absorption at 252 and 260 nm and was found to be 1.3 µmol, or 920 µg; an extinction coefficient of 28.18 AU/mM/cm at 252 nm was assumed, based on guanosine monophosphate.¹⁵ C-di-GMP labeled with ¹³C and ¹⁵N was produced in the same manner, starting with ¹³C-, ¹⁵N-GTP (Spectra Stable Isotopes).

NMR sample preparation

Protein samples. PA4608 sample buffer contained 250 mM NaCl, 10 mM Tris-HCl at pH 6.7, 0.01% (w/v) NaN₃ as a preservative, and 5% (v/v) D₂O for locking. The protein concentration was 0.8 – 1.2 mM, and the sample volume was 200 – 400 µL. Samples were put in NMR microtubes with their magnetic susceptibility matched to D₂O, with an outer diameter of 5 mm (Shigemi, Inc.).

Protein-ligand complex samples. C-di-GMP was added to PA4608 samples in NMR tubes as an approximately 19 mM solution. ¹H-¹⁵N HSQC spectra were recorded to monitor the fraction of protein with bound ligand; two equivalents of c-di-GMP were necessary to saturate the protein. The sample pH, which was lowered after addition of acidic c-di-GMP, was readjusted by adding concentrated NaOH; the sample pH was followed using a pH meter and by observing histidine N^{δ1} and N^{ε2} chemical shifts in long-range ¹H-¹⁵N HSQC spectra.

D₂O samples. H₂O samples were transferred to 1.5 mL polypropylene tubes, shock frozen in liquid nitrogen, and lyophilized at <1 mbar. The lyophilisate was redissolved in D₂O. This process was repeated once. The sample was transferred to a Shigemi tube that had been rinsed with D₂O and dried.

Oriented sample for measurement of RDCs. The oriented sample had a final volume of 290 μ L and contained: 0.5 mM PA4608; 1.0 mM c-di-GMP; 56 mM NaCl; 20 mM Tris-HCl at pH 6.7; 0.01% (w/v) NaN_3 ; 5% (v/v) D_2O ; and 18 g/L Pfl phage (Asla Biotech).

The following samples were prepared: uniformly ^{13}C - and ^{15}N -labeled protein with unlabeled c-di-GMP in H_2O and D_2O ; uniformly ^{13}C - and ^{15}N -labeled protein with ^{13}C - and ^{15}N -labeled c-di-GMP in H_2O , D_2O , and H_2O plus Pfl phage for measurement of RDCs; and ^{15}N -labeled protein with less than one equivalent of unlabeled c-di-GMP for exchange spectroscopy (EXSY).

NMR spectroscopy

Standard 1-, 2- and 3D NMR spectra similar to those described in Grzesiek & *al.*, 1997¹⁶ were recorded on Bruker DRX 600 and DRX 800 NMR spectrometers with TXI and TCI probe heads, respectively. The sample temperature was 293 K except where indicated otherwise. A list of NMR spectra is given in Table 1. NMR data were processed using the NMRPipe suite of programs.¹⁷

Resonance assignments

Protein resonances were assigned using standard techniques.¹⁶

C-di-GMP ribose resonances were connected using ^{13}C -resolved 3D COSY and TOCSY spectra recorded on ^{13}C -, ^{15}N -labeled c-di-GMP bound to ^{13}C -, ^{15}N -labeled PA4608. Ribose moieties and guanine H1/N1 and H8/C8 resonances were assigned to c-di-GMP molecules based on NOE contacts.

Structure calculations

Conformational restraints. Cross peaks in 3D NMR spectra were assigned using the program STAPP,^{18,19} and assignments were iteratively refined based on preliminary structure calculations. Backbone torsion angle restraints were derived from $^{13}\text{C}^\alpha$ and $^{13}\text{C}^\beta$ chemical shifts using the program TALOS.²⁰ Side chain χ_1 torsion angles were obtained from H^N - H^β ROE distances and $^3J_{\text{H}^\alpha\text{H}^\beta}$, $^3J_{\text{N}^\text{C}^\gamma}$ and $^3J_{\text{C}^\text{C}^\gamma}$ coupling constants. Hydrogen bonds were included as distance restraints between H^N and O and between N and O atoms where hydrogen bond donors and acceptors could be unambiguously assigned after preliminary structure calculations. Analysis of $^3J_{\text{HH}}$ couplings in c-di-GMP ribose moieties revealed that all four ribose moieties were in a 3'-endo conformation; accordingly, five torsion angle restraints were introduced for each ribose ring. 29 protein-ligand NOE contacts were observed in 2D, isotope-filtered 2D and ^{13}C -edited 3D NOESY spectra. 11 ^{13}C - ^1H RDCs were measured for c-di-GMP

ribose moieties. Comprehensive statistics on conformational restraints are given in Table 2.

Structure calculations. Structure calculations were performed with the program XPLOR-NIH²¹ using a slightly modified version of the annealing protocol “sa.inp” included therein. C-di-GMP molecules were defined as two-residue, circular strands of RNA using parameters from the XPLOR-NIH file nucleic.par. The 20 residue His tag and linker was not included in structure calculations. Statistics for the complex structure are given in Table 2.

Kinetic measurements. The rates of disappearance of the free protein species and of appearance of protein with bound c-di-GMP were measured by continuously recording 1D ¹H spectra of PA4608 for 2.5 h after adding c-di-GMP and mixing. Each spectrum was recorded for 1.5 min. Steady-state association and dissociation rates were measured with a 3D ¹H-¹⁵N EXSY spectrum recorded at a temperature of 313 K; exchange was not detectable at 293 K.

Results and discussion

Resonance assignments

Resonance assignments were found for roughly 90% of the protein backbone in the PA4608 · c-di-GMP complex. Two stretches of the sequence, F33-I36 and D130-L138, were not visible in NMR spectra; presumably these residues are mobile, and NMR signals are broadened beyond detection by intermediate exchange. This situation did not improve after cooling to 278K or heating to 303 or 313 K. In protein-bound c-di-GMP, all imino H1 and carbon-bound hydrogens were assigned except for Gua1 H5'' and Gua4 H4', H5', and H5''. Assignment statistics are given in Table 2.

Secondary ¹³C^α and ¹³C^β shifts

Secondary ¹³C^α and ¹³C^β chemical shifts, *i.e.* observed chemical shifts minus chemical shifts expected in a random coil, are shown in Figure 2. Secondary ¹³C^α and ¹³C^β chemical shifts are primarily defined by secondary structure.²² Thus for residues I36-L128, *i.e.* the β barrel plus helix α1, the secondary structure of PA4608 with bound c-di-GMP is identical to that of the free protein. The N-terminus up to residue E29 is a random coil in both free and ligand-bound protein; resonance assignments are incomplete for R31-I36. The region constituting helix 3₁₀ 1 in free PA4608 and comprising residues G129-E137 could not be observed in the PA4608 · c-di-GMP complex, indicating that it becomes more flexible after ligand binding. The C-terminal residues A139-S143, which form helix 3₁₀ 2 in free PA4608, have different

secondary $^{13}\text{C}^\alpha$ and $^{13}\text{C}^\beta$ chemical shifts in the complex, implying that structural rearrangement takes place at the C-terminus.

Kinetics of association and dissociation

In order to follow the association reaction of c-di-GMP with PA4608, 1D ^1H NMR spectra of the reaction mixture were recorded continuously after mixing the two components. Peaks corresponding to ligand-bound protein appeared surprisingly slowly, on a time scale of about 100 min. Peaks corresponding to unbound protein disappeared at the same rate. Several peaks were attributed to free c-di-GMP. One set with chemical shifts of 11.08 and 11.14 ppm disappeared at a slightly faster rate than free protein, while another set with chemical shifts of 11.27 and 11.29 ppm increased in intensity for approximately 10 minutes before disappearing again, at a rate comparable to free protein resonances. Considering that upon mixing c-di-GMP stock solution with protein sample, c-di-GMP is transferred from a concentrated, acidic solution to a more dilute, buffered environment, we suggest that the observed slow processes represent a rearrangement of c-di-GMP to a form that is suitable for binding to PA4608. It has been reported previously that the transition between different states of c-di-GMP can be very slow and take up to several hours to complete.⁵

The PA4608 · c-di-GMP structure

The structure of the PA4608 · c-di-GMP complex was solved using a total of 1521 conformationally restricting constraints, including 1150 distance restraints as well as RDC, hydrogen bond, TALOS, and χ_1 angle constraints. The position and conformation of the ligand was defined by 11 ^{13}C - ^1H RDCs in ribose moieties, 29 NOE contacts between protein and ligand, and 56 NOE contacts within or between ligand molecules. The rmsd coordinate precision of the final model was 0.34 Å for backbone atoms of residues in the β barrel, 0.97 Å for all non-hydrogen atoms in the structured part of the protein (residues F38-N127), and 1.98 Å for non-hydrogen atoms in c-di-GMP. Figure 5 illustrates the model precision. Detailed statistics on conformational restraints and model quality are give in Table 2.

The PA4608 · c-di-GMP complex structure is shown in Figure 6. C-di-GMP binds to the side of the PA4608 β barrel. The N-terminal part comprising residues M23-R35 appears to be flexible and unstructured: $^{13}\text{C}^\alpha$ and $^{13}\text{C}^\beta$ chemical shifts are indicative of a random coil conformation (Figure 2), and no long-range protein-protein NOE contacts were found. Four NOE contacts were identified, however, between c-di-GMP residue Gua3 and protein residues R30-R32 (Figure 4). Thus the N terminus containing the highly conserved RxxxR motif contacts the bound ligand on its distal side. The position of these residues could only be determined with moderate precision

as residues F33-R35 were not observable in NMR spectra because of unfavorable dynamics.

Residues G129-E137 could not be observed in PA4608 · c-di-GMP, again because these residues were mobile. Residues L138-D147 contained a large number of short range NOEs indicating that residues L139-S143 form an α helix containing. However, no long-range contacts between this C-terminal helix and the β barrel were identified, and no structural information is available for residues D130-E137; therefore the position of the C-terminal α helix relative to the β barrel could not be determined.

C-di-GMP binds to PA4608 as an intercalated dimer

Four sets of guanosine resonances with similar intensity and relaxation properties were observed in NMR spectra of PA4608 · c-di-GMP, implying that the complex contains two molecules of c-di-GMP. Four sets of NOESY peaks originating from c-di-GMP imino H1 atoms are shown in Figure 3. Many NOE contacts were observed between c-di-GMP Gua4 H1 and the ribose moieties of Gua1 and Gua2, and between Gua4 H1 and the ribose moieties of Gua1 and Gua2 (Figure 4), consistent with the previously observed self-intercalated structure of free c-di-GMP.^{6,7} From this structure, contacts were also expected between Gua3 H1 and the ribose moiety of Gua1, and between Gua2 H1 and the ribose moiety of Gua4. However, the guanine moiety of Gua3 is dynamic as inferred from NMR peak shapes; consequently, NOESY diagonal and cross peaks were weak (Figure 3). The ribose moiety of Gua2 appears to be displaced from the c-di-GMP stack by the side chain of W99 and is too far away from the ribose moiety of Gua4 for NOE transfer to occur.

Dynamics

¹⁵N relaxation data for free and ligand-bound PA4608 are shown in Figure 8. Transverse (T_2) and longitudinal (T_1) relaxation time constants indicate that the overall tumbling rate increases slightly after ligand binding, presumably because the previously flexible N-terminus folds around the bound ligand, reducing hydrodynamic drag. This is supported by all relaxation parameters measured for residue R30, where the $\{^1\text{H}\}$ -¹⁵N NOE rate and ¹⁵N T_1 are drastically increased, whereas ¹⁵N T_2 is reduced such that all parameters are comparable to the core of the protein after ligand binding. Residue W99, which is in direct contact with bound c-di-GMP, is less flexible in the ligand-bound state. The C-terminal part of the protein comprising residues V142-D147 becomes more flexible after ligand binding.

Comparison of free and ligand-bound PA4608

An overlay of free and ligand-bound PA4608 is shown in Figure 7. Helix $\alpha 1$ appears to tilt towards the β barrel by approximately 12° ; however, considering the very similar $^{13}\text{C}^\alpha$ and $^{13}\text{C}^\beta$ chemical shifts in this region (Figure 2) we believe this to be an artifact of structure determination. More importantly, free PA4608 contains two C-terminal 3_{10} helices that contact the β barrel. In PA4608 \cdot c-di-GMP, residues D130-L138 were not observed in NMR spectra, and the only available structural information about this region is that it has a certain flexibility. Residues A139-A144 form an α helix whose position relative to the β barrel remains unknown; this helix cannot take the place of helix $3_{10} 2$ in free PA4608 since that position is occupied by the bound ligand.

Biological context

After recognition of c-di-GMP by PA4608, the signal must be further relayed to downstream effectors to elicit a response *in vivo*. The interaction partners of PA4608 remain unknown; we can only assume that they recognize interaction surfaces formed in PA4608 after ligand binding. Unlike PA4608, many PilZ domains are parts of larger multidomain proteins. In such a context, the formation of additional structure in the N-terminus and rearrangement of the C-terminus of the PilZ domain described here may change the interactions of output domains, and thus elicit downstream responses.

Bibliography

1. Ross P *et al.* Regulation of cellulose synthesis in *Acetobacter xylinum* by cyclic diguanylic acid. *Nature* **325**, 279-281 (1987).
2. Paul R, Weiser S, Amiot NC, Chan C, Schirmer T, Giese B & Jenal U. Cell cycle-dependent dynamic localization of a bacterial response regulator with a novel di-guanylate cyclase output domain. *Genes & Development* **18**, 715-727 (2004).
3. Pratt JT, Tamayo R, Tischler AD & Camilli A. PilZ domain proteins bind cyclic diguanylate and regulate diverse processes in *Vibrio cholerae*. *Journal of Biological Chemistry* **282**, 12860-12870 (2007).
4. Merighi M, Lee VT, Hyodo M, Hayakawa Y & Lory S. The second messenger bis-(3'-5')-cyclic-GMP and its PilZ domain-containing receptor Alg44 are required for alginate biosynthesis in *Pseudomonas aeruginosa*. *Molecular Microbiology* **65**, 876-895 (2007).
5. Zhang ZY, Kim S, Gaffney BL & Jones RA. Polymorphism of the signaling molecule c-di-GMP. *Journal of the American Chemical Society* **128**, 7015-7024 (2006).

6. Egli M, Gessner RV, Williams LD, Quigley GJ, Vandermarel GA, Vanboom JH, Rich A & Frederick CA. Atomic resolution structure of the cellulose synthase regulator cyclic diguanylic acid. *Proceedings of the National Academy of Sciences of the United States of America* **87**, 3235-3239 (1990).
7. Chan C, Paul R, Samoray D, Amiot NC, Giese B, Jenal U & Schirmer T. Structural basis of activity and allosteric control of diguanylate cyclase. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 17084-17089 (2004).
8. Amikam D & Galperin MY. PilZ domain is part of the bacterial c-di-GMP binding protein. *Bioinformatics* **22**, 3-6 (2006).
9. Ryjenkov DA, Simm R, Romling U & Gomelsky M. The PilZ domain is a receptor for the second messenger c-di-GMP - The PilZ domain protein YcgR controls motility in enterobacteria. *Journal of Biological Chemistry* **281**, 30310-30314 (2006).
10. Christen M, Christen B, Allan MG, Folcher M, Jenö P, Grzesiek S & Jenal U. DgrA is a member of a new family of cyclic diguanosine monophosphate receptors and controls flagellar motor function in *Caulobacter crescentus*. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 4112-4117 (2007).
11. Ramelot TA, Yee A, Cort JR, Semesi A, Arrowsmith CH & Kennedy MA. NMR structure and binding studies confirm that PA4608 from *Pseudomonas aeruginosa* is a PilZ domain and a c-di-GMP binding protein. *Proteins* **66**, 266-271 (2007).
12. Zhang R, Zhou M, Moy S, Collart F & Joachimiak A. The crystal structure of the hypothetical protein VCA0042 from *Vibrio cholerae* O1. *to be published* (2005).
13. Gasteiger E, Gattiker A, Hoogland C, Ivanyi I, Appel RD & Bairoch A. ExPASy: the proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Research* **31**, 3784-3788 (2003).
14. Christen B, Christen M, Paul R, Schmid F, Folcher M, Jenö P, Meuwly M & Jenal U. Allosteric control of cyclic di-GMP signaling. *Journal of Biological Chemistry* **281**, 32015-32024 (2006).
15. Cavaluzzi MJ & Borer PN. Revised UV extinction coefficients for nucleoside-5'-monophosphates and unpaired DNA and RNA. *Nucleic Acids Research* **32** (2004).
16. Grzesiek S, Bax A, Hu JS, Kaufman J, Palmer I, Stahl SJ, Tjandra N & Wingfield PT. Refined solution structure and backbone dynamics of HIV-1 Nef. *Protein Science* **6**, 1248-1263 (1997).
17. Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J & Bax A. NMRPipe - a multidimensional spectral processing system based on UNIX pipes. *Journal of Biomolecular NMR* **6**, 277-293 (1995).
18. Garrett DS, Gronenborn AM & Clore GM. Automated and interactive tools for assigning 3D and 4D NMR spectra of proteins – CAPP, STAPP and PIPP. *Journal of Cellular Biochemistry*, 71-71 (1995).
19. Garrett DS, Powers R, Gronenborn AM & Clore GM. A common sense approach to peak picking in 2-dimensional, 3-dimensional, and 4-dimensional spectra using automatic computer analysis of contour diagrams. *Journal of Magnetic Resonance* **95**, 214-220 (1991).

20. Cornilescu G, Delaglio F & Bax A. Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *Journal of Biomolecular NMR* **13**, 289-302 (1999).
21. Schwieters CD, Kuszewski JJ, Tjandra N & Clore GM. The XPLOR-NIH NMR molecular structure determination package. *Journal of Magnetic Resonance* **160**, 65-73 (2003).
22. Spera S & Bax A. Empirical correlation between protein backbone conformation and C^α and C^β ¹³C nuclear magnetic resonance chemical shifts. *Journal of the American Chemical Society* **113**, 5490-5492 (1991).
23. Grzesiek S & Bax A. Improved 3D triple resonance NMR techniques applied to a 31-kDa protein. *Journal of Magnetic Resonance* **96**, 432-440 (1992).
24. Piotto M, Saudek V & Sklenar V. Gradient-tailored excitation for single quantum NMR spectroscopy of aqueous solutions. *Journal of Biomolecular NMR* **2**, 661-665 (1992).
25. Kay LE, Ikura M, Tschudin R & Bax A. 3-dimensional triple resonance NMR spectroscopy of isotopically enriched proteins. *Journal of Magnetic Resonance* **89**, 496-514 (1990).
26. Ikura M, Kay LE & Bax A. A novel approach for sequential assignment of ¹H, ¹³C, and ¹⁵N spectra of larger proteins - heteronuclear triple resonance 3-dimensional NMR spectroscopy - application to calmodulin. *Biochemistry* **29**, 4659-4667 (1990).
27. Grzesiek S & Bax A. An efficient experiment for sequential backbone assignment of medium-sized isotopically enriched proteins. *Journal of Magnetic Resonance* **99**, 201-207 (1992).
28. Grzesiek S & Bax A. Correlating backbone amide and side chain resonances in larger proteins by multiple relayed triple resonance NMR. *Journal of the American Chemical Society* **114**, 6291-6293 (1992).
29. Grzesiek S & Bax A. Amino acid type determination in the sequential assignment procedure of uniformly ¹³C/¹⁵N-enriched proteins. *Journal of Biomolecular NMR* **3**, 185-204 (1993).
30. Bax A, Clore GM & Gronenborn AM. ¹H-¹H correlation via isotropic mixing of ¹³C magnetization, a new 3-dimensional approach for assigning ¹H and ¹³C spectra of ¹³C-enriched proteins. *Journal of Magnetic Resonance* **88**, 425-431 (1990).
31. Kay LE, Xu GY, Singer AU, Muhandiram DR & Formankay JD. A gradient-enhanced HCCH TOCSY experiment for recording side-chain ¹H and ¹³C correlations in H₂O samples of proteins. *Journal of Magnetic Resonance Series B* **101**, 333-337 (1993).
32. Grzesiek S, Kuboniwa H, Hinck AP & Bax A. Multiple-quantum line narrowing for measurement of H^α-H^β J-couplings in isotopically enriched proteins. *Journal of the American Chemical Society* **117**, 5312-5315 (1995).
33. Hu JS, Grzesiek S & Bax A. Two-dimensional NMR methods for determining χ₁ angles of aromatic residues in proteins from three-bond J_{C^αC^γ} and J_{N^{C^γ}} couplings. *Journal of the American Chemical Society* **119**, 1803-1804 (1997).
34. Grzesiek S & Bax A. The importance of not saturating H₂O in protein NMR – application to sensitivity enhancement and NOE measurements. *Journal of the American Chemical Society* **115**, 12593-12594 (1993).

35. Lippens G, Dhalluin C & Wieruszeski JM. Use of a water flip-back pulse in the homonuclear NOESY experiment. *Journal of Biomolecular NMR* **5**, 327-331 (1995).
36. Ikura M, Kay LE, Tschudin R & Bax A. 3-dimensional NOESY-HMQC spectroscopy of a ^{13}C -labeled protein. *Journal of Magnetic Resonance* **86**, 204-209 (1990).
37. Zuiderweg ERP, McIntosh LP, Dahlquist FW & Fesik SW. 3-dimensional ^{13}C -resolved proton NOE spectroscopy of uniformly ^{13}C -labeled proteins for the NMR assignment and structure determination of larger molecules. *Journal of Magnetic Resonance* **86**, 210-216 (1990).
38. Laskowski RA, Rullmann JAC, MacArthur MW, Kaptein R & Thornton JM. AQUA and PROCHECK-NMR: Programs for checking the quality of protein structures solved by NMR. *Journal of Biomolecular NMR* **8**, 477-486 (1996).
39. Schwieters CD & Clore GM. The VMD-XPLOR visualization package for NMR structure refinement. *Journal of Magnetic Resonance* **149**, 239-244 (2001).
40. DeLano WL. (DeLano Scientific LLC, Palo Alto, CA, USA, <http://www.pymol.org>, 2007).

Figures and Tables

Table 1: NMR spectra recorded on PA4608 · c-di-GMP.

Experiment	Sample ^a	time	Acquisition times in indirect dimensions(ms)	¹ H freq. (MHz)	Purpose
¹ H- ¹⁵ N HSQC	1		N: 60	600	Overview
Aliphatic constant-time ¹ H- ¹³ C HSQC	1		C: 26.0	800	Overview
Aromatic ¹ H- ¹³ C HSQC	1		C: 12.4	800	Overview
Constant-time HNCO ²³ with WATERGATE ²⁴	1	11 h	C': 39.2; N:27	600	Overview
HNCA ^{25,26}	1	37 h	C: 10.8; N: 23.4	600	Assignments
CBCANH ²⁷	1	36 h	C: 6.8; N: 22.9	600	Assignments
CBCA(CO)NH ²⁸ with WATERGATE ²⁴	1	35 h	C: 6.8; N: 25.6	600	Assignments
C(CO)NH TOCSY ²⁸	1	40 h	C: 7.0; N: 25.6	600	Assignments
HBHA(CO)NH ²⁹ with WATERGATE ²⁴	1	49 h	H: 11.0; N: 25.6	600	Assignments
Aliphatic HCCH TOCSY ^{30,31}	1	37 h	H: 15.6; C: 10.0	800	Assignments
Ribose HCCH TOCSY	3	37 h	H: 16; C: 5.2	600	Ribose assignments
Ribose HCCH COSY	3, 303 K ^b	20 h	H: 15.4; C: 6.1	600	Ribose assignments
HAHB ³²	2	32 h	H ^α H ^β : 9.9; C ^α : 23.6	800	³ J _{HαHβ} couplings
HAHB ³² optimized for ribose ³ J couplings	3 / 4	46 h / 47 h	H: 12.2; C: 14.5 H: 6.9; C: 14.5	800	Ribose ³ J _{HH} couplings & RDCs
¹³ C ^γ -{ ¹³ C ^γ } spin-echo difference ¹ H- ¹⁵ N HSQC ³³	5	20 h	N: 50	600	aromatic ³ J _{C^γC^γ} couplings
¹⁵ N-{ ¹³ C ^γ } spin-echo difference ¹ H- ¹⁵ N HSQC ³³	5	20 h	N: 59.8	800	aromatic ³ J _{C^γC^γ} couplings
Methyl C	2	16 h	C: 51.2	600	methyl ³ J _{C^γC^γ} couplings
Methyl N	2	28 h	C: 51.2	600	methyl ³ J _{N^γC^γ} couplings
¹⁵ N-edited 3D ROESY	5	47 h	H: 15.0; N: 23.9	800	ROE distances
¹⁵ N-edited 3D NOESY with water flip-back, ³⁴ radiation damping during mixing period ³⁵	1	33 h	H: 14.4; N: 20.3	800	NOE distances
Aliphatic ¹³ C-edited 3D NOESY ^{36,37}	1	53 h	H: 15.1; C _{aliph.} : 9.9	800	NOE distances
Aromatic ¹³ C-edited 3D NOESY	1	51 h	H: 15.3; C _{arom.} : 8.0	800	NOE distances
Ribose ¹³ C-edited 3D NOESY	5	61 h	H: 10.2; C _{ribose} : 6.0	800	NOE distances
4D ¹³ C- ¹³ C-NOESY-HMQC	2	94 h	C ₁ : 5.0; H ₁ : 8.0; C ₂ : 5.0	800	NOE distances
2D NOESY	1	19 h	H: 23.3	600	c-di-GMP H1 NOEs

2D NOESY filtered against ^1H bound to ^{13}C or ^{15}N	1	11 h	H: 23.3	600	c-di-GMP H8 NOEs
2D NOESY filtered against ^1H bound to ^{13}C or ^{15}N , without decoupling in indirect dimension	1	6 h	H: 23.3	600	Distinguish protein and c-di-GMP in indirect dimension
2D NOESY filtered against ^1H bound to ^{13}C or ^{15}N	2	9 h	H: 23.3	600	c-di-GMP H1' NOEs
IPAP ^1H - ^{15}N HSQC	1 / 4	6 h / 1.5 h	N: 76.8	800	N- H^{N} RDCs
J-resolved constant-time ^{13}C -HSQC	5 / 4	7h / 11 h	J: 22.5; C: 23.5 / J: 22.7; C: 23.5	800	Ribose C-H RDCs
2D ^1H -resolved ^{15}N EXSY	6, 293 K / 303 K / 313 K	16 h / 13 h / 15 h	H: 15.9	600	Exchange rates
3D ^1H -, ^{15}N -resolved ^{15}N EXSY	6, 313 K	47 h	H: 17.0; N: 43.2	800	Assignments, exchange rates

^aThe sample temperature was 293 K unless indicated otherwise. The following samples were used:

- 1: ^{13}C -, ^{15}N -protein + unlabeled ligand in H_2O ;
- 2: ^{13}C -, ^{15}N -protein + unlabeled ligand in D_2O ;
- 3: ^{13}C -, ^{15}N -protein + ^{13}C -, ^{15}N -ligand in D_2O ;
- 4: ^{13}C -, ^{15}N -protein + ^{13}C -, ^{15}N -ligand in H_2O + Pf1 phage;
- 5: ^{13}C -, ^{15}N -protein + ^{13}C -, ^{15}N -ligand in H_2O ;
- 6: ^{15}N -protein + approx. 0.5 equivalents unlabeled ligand in H_2O .

See Materials and Methods for details.

^b This spectrum was recorded at an elevated temperature for better relaxation properties. ^1H - ^{13}C HSQC spectra showed that ribose ^1H and ^{13}C chemical shifts remained largely unchanged after heating from 293 to 303 K.

Table 2: Structure statistics.

Completeness of resonance assignments:	
—Protein, backbone ^a	89.0% (445/500)
—Protein, all atoms ^b	83.5% (1117/1338)
—c-di-GMP ^c	83.3% (50/60)
Distance constraints ^d :	
—Protein, intraresidue ($i = j$)	106
—Protein, sequential ($ i - j = 1$)	367
—Protein, short range ($2 \leq i - j \leq 5$)	213
—Protein, long range ($ i - j \geq 6$)	379
—Protein to ligand	29
—Within ligand	56
—Total	1150
—NOE constraints per residue	9.20
RDCs, protein	154
RDCs, c-di-GMP	11
Hydrogen bonds	53
TALOS torsion angle constraints ²⁰	128
χ_1 angle constraints ^e	25
Total number of constraints	1521
Total constraints per residue	12.17
Number of structures calculated / used	60/10
Constraint violations:	
—Average number of distance violations >0.1 Å per structure	193
—RMSD of distance violations ^f	0.21 Å
—RMSD of dihedral angle violations	4.1°
—RMSD of RDC violations	1.44 Hz
Coordinate precision (RMSD to average):	
—Protein backbone atoms, ^g β barrel ^h	0.34 Å
—Protein backbone atoms, ^g residues F38-N127	0.53 Å
—All non-hydrogen atoms, residues F38-N127	0.97 Å
—c-di-GMP non-hydrogen atoms	1.98 Å
Ramachandran plot summary from PROCHECK-NMR ^{38,1}	
—Most favored regions	82.6%
—Additionally allowed regions	13.6%
—Generously allowed regions	3.8%
—Disallowed regions	0.0%

^a Considering $^1\text{H}^{\text{N}}$, ^{15}N (except Pro), $^{13}\text{C}^{\alpha}$, and $^1\text{H}^{\alpha}$ resonances of residues M23-D147.

^b Considering routinely assignable ^1H , ^{15}N , and ^{13}C resonances of residues M23-D147, excluding N-terminal and Lys amino groups, Arg guanidino groups, hydroxyl protons of Ser, Thr, Tyr, thiol protons of Cys, carboxyl resonances of Asp and Glu, nonprotonated aromatic carbons, and Pro ^{15}N . ^1H belonging to the same methyl group and Phe, Tyr $^1\text{H}^{\delta}$, $^1\text{H}^{\epsilon}$ are counted as one.

^c Considering H1', C1', H2', C2', H3', C3', H4', C4', H5', H5'', C5', H1, N1, H8, C8.

^d Conformationally restricting restraints, *i.e.* excluding distances between geminal hydrogens or protons of the same aromatic group.

^e Derived from $\text{H}^{\alpha}\text{-H}^{\beta}$ ^3J coupling constants and $\text{H}^{\text{N}}\text{-H}^{\beta}$ ROE distances.

^f NOEs and hydrogen bonds.

^g N^{H} , C^{α} , and C'.

^h Residues D41-Q46, R49-V58, G62-G66, P78-Y84, L90-R101, L104-I112.

ⁱ Residues F38-N127 and L138-L141, excluding Gly and Pro).

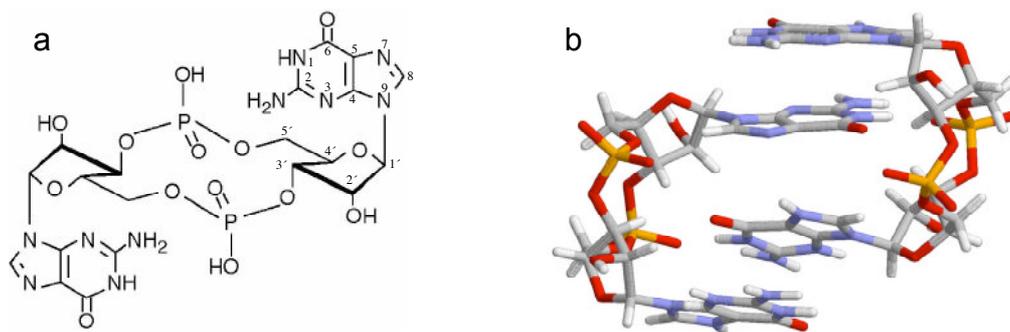


Figure 1: C-di-GMP. (a) Chemical structure. (b) Crystal structure.⁶

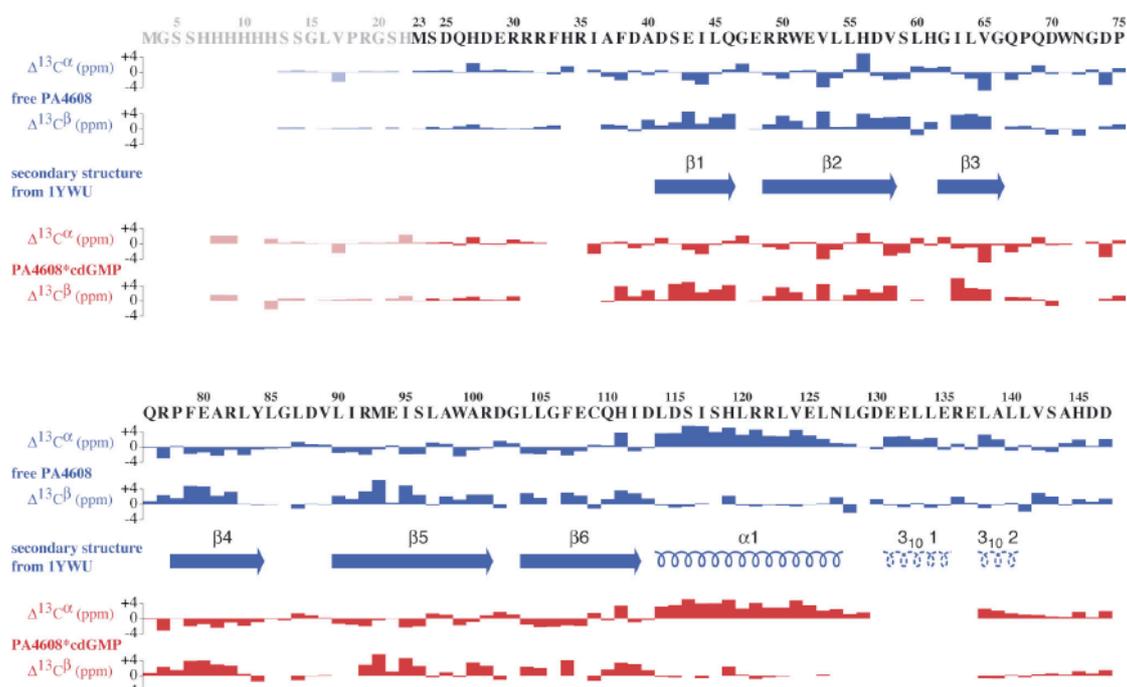


Figure 2: Secondary $^{13}\text{C}^\alpha$ and $^{13}\text{C}^\beta$ shifts for free PA4608 and PA4608 · c-di-GMP. The secondary structure found in free PA4608 is indicated, with arrows representing β sheet and spirals representing α and 3_{10} helices. Residues M3-H22 form the His tag and linker and are shown in gray / pale colors. See text for details.

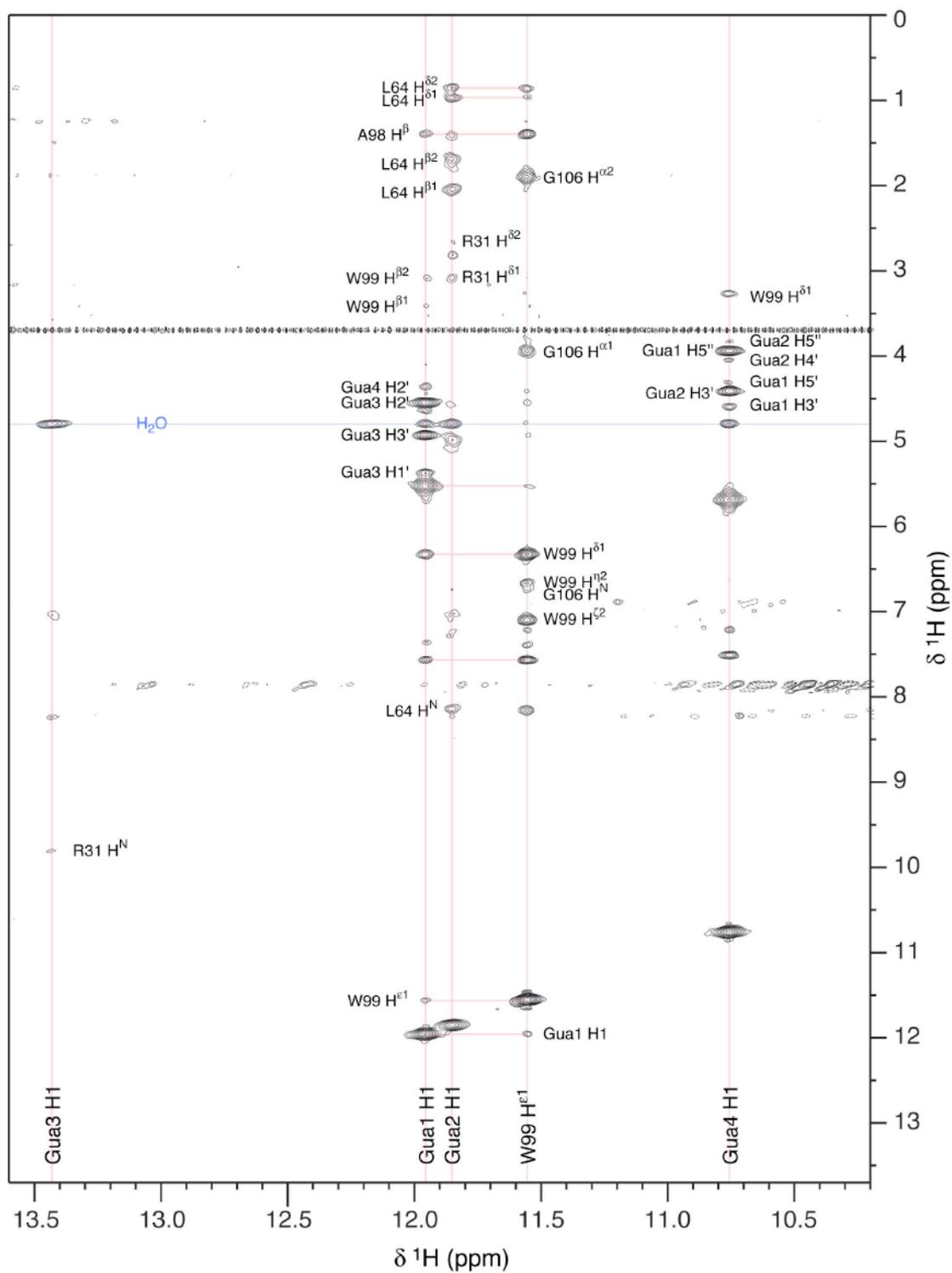


Figure 3: Excerpt from a 2D ^1H - ^1H NOESY spectrum of PA4608 with bound c-di-GMP. The spectral region shown contains 1H atoms with extraordinarily high chemical shifts on the horizontal axis. Peaks are labeled where assignments are known.

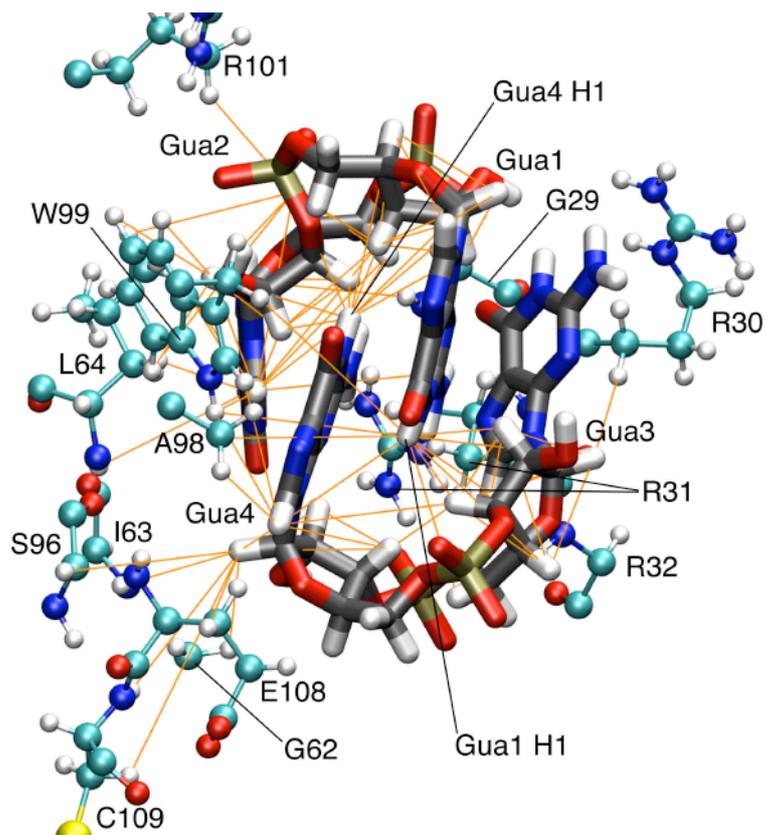


Figure 4: Intra-c-di-GMP and c-di-GMP to protein NOEs. C-di-GMP is shown as sticks; protein residues contacting the ligand are shown as balls and sticks. Protein side chains are only shown where necessary. NOE distance constraints are shown as orange lines. The figure was prepared with VMD-XPLOR.³⁹

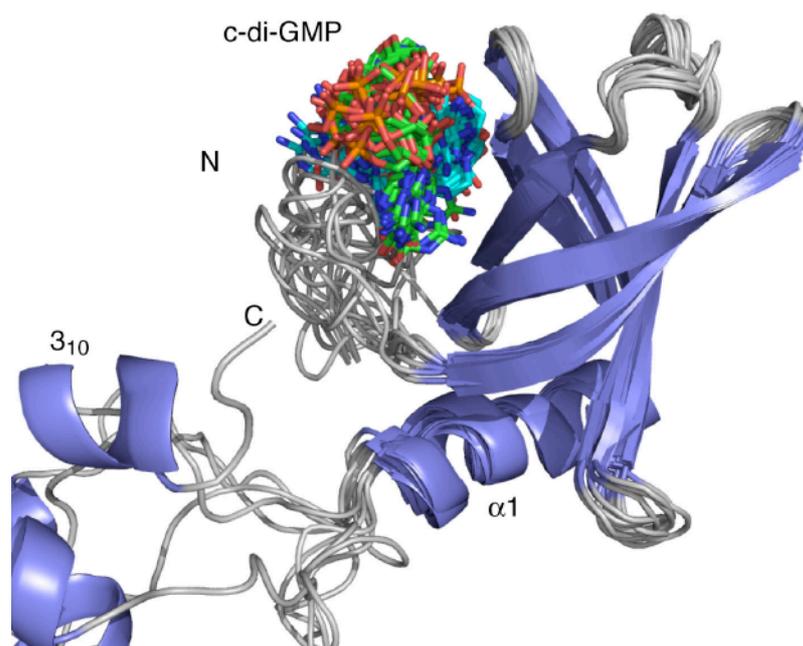


Figure 5: Overlay of the 10 lowest energy models out of 60 calculated, illustrating the precision of the structure (see also Table 2). Secondary structure elements and coils of PA4608 are shown in blue and gray, respectively. The two molecules of c-di-GMP are shown as green/CPK and cyan/CPK sticks. Hydrogen atoms were omitted for clarity. Parts of the protein C terminus are not shown.

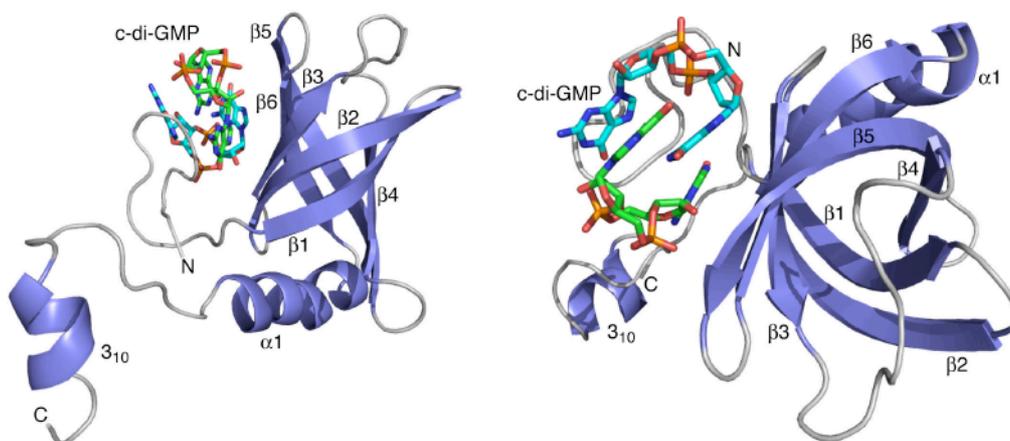


Figure 6: Structure of PA4608 with bound c-di-GMP. **Left:** view emphasizing the arrangement of secondary structure elements; **right:** view revealing the intercalated c-di-GMP dimer. Colors are as in Figure 5.




Figure 7: Overlay of free PA4608 (PDB 1YWU¹¹ model 12; yellow) with PA4608 · c-di-GMP (blue). Figures 5 and 6 were prepared with MacPyMol.⁴⁰




Figure 8: amide ¹⁵N relaxation data for PA4608 with (red) and without (black) bound c-di-GMP at 293 K.

Curriculum Vitae: **Martin G. Allan**

EDUCATION

- 2003 – 2007 PhD in biophysics with Prof S. Grzesiek, Biozentrum, University of Basel, Switzerland.
- 2000 – 2002 MSc in molecular biology, Biozentrum, University of Basel, Switzerland.
- 1998 – 2000 Undergraduate studies in biochemistry, University of Fribourg, Switzerland.
- 1994 – 1998 Baccalaureate: Kollegium Heilig Kreuz, Fribourg.
- 1996 – 1997 Exchange year at Northrop High School, Fort Wayne, Indiana, USA.

RESEARCH EXPERIENCE

- 2003 – 2007 PhD at Biozentrum, University of Basel, Switzerland under the guidance of Prof. S. Grzesiek and in collaboration with Prof. Tilman Schirmer, Prof. Urs Jenal, and Jörg Stetefeld.
- Established protocols for the production and purification of several isotope- and SeMet-labeled proteins, achieving yields and purities sufficient for structural biology.
 - Identified crystallization conditions, acquired and analyzed synchrotron diffraction data, and determined the structure of TipAN bound to target DNA.
 - Investigated the binding of various antibiotics to the bacterial protein TipAS, and proposed a mechanism for transcriptional regulation. Determined NMR solution structures of TipAS with bound antibiotics.
 - Determined the binding site of c-di-GMP on the surface of the recently discovered PA4608 / DgrA protein, and characterized the binding process by NMR.
- 2001 – 2003 MSc student / research assistant at Biozentrum, University of Basel, Switzerland. under the guidance of Prof. S. Grzesiek and Jan Kahmann
- Discovered that binding of a drug to TipAS protein induces drastic structural changes.
 - Produced, purified and characterized a protein involved in multidrug resistance of *M. tuberculosis*.
- 2000 Internship at Nestlé Research Center, Lausanne.
- Wrote the final report on a study concluding that amaranth grain does not lower cholesterol in hamsters.

PEER-REVIEWED PUBLICATIONS

- Allan M, Christen M, Christen B, Jenal U, Grzesiek S. Solution structure of the diguanylate receptor PA4608 with bound cyclic di-GMP. In preparation.
- Allan M, Newberry K, Schuman J, Brennan R, Thompson C, Stetefeld J, Schirmer T, Grzesiek S. Structure of the transcriptionally inactive MerR domain TipAN in complex with DNA. In preparation.
- Christen M, Christen B, Allan M, Folcher M, Jenö P, Grzesiek S, Jenal U. DgrA is a member of a new family of cyclic diguanosine monophosphate receptors and controls flagellar motor function in *Caulobacter crescentus*. *PNAS* **104**, 4112-4117 (2007).
- Kahmann J, Sass HJ, Allan M, Seto H, Thompson C, Grzesiek S: Structural basis for antibiotic recognition by the TipA class of multidrug-resistance transcriptional regulators. *EMBO J* **22**, 1824-1834 (2003).