

Protéines : Structure fonction et évolution.

Alexey Aleksandrov

▶ **To cite this version:**

Alexey Aleksandrov. Protéines : Structure fonction et évolution.. Biologie moléculaire. Ecole Polytechnique X, 2008. Français. NNT : . pastel-00004205

HAL Id: pastel-00004205
https://pastel.hal.science/pastel-00004205

Submitted on 22 Jul 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THESE

pour obtenir le titre de

DOCTEUR DE l'ECOLE POLYTECHNIQUE

Discipline: Biologie

Présentée et soutenue publiquement par

Alexey Aleksandrov

le 19 juin 2008

Tetracycline antibiotics and their targets: computational studies

Jury composé de

Prof. Johan Åqvist	Examinateur
Prof. Georgios Archontis	Membre invité
Dr. Martin Field	Rapporteur
Prof. Winfried Hinrichs	Membre invité
Dr. Richard Lavery	Rapporteur
Dr. Gilles Ohanessian	Examinateur
Prof. Thomas Simonson	Directeur de thèse

Abstract

Tetracyclines (Tc) are a family of important antibiotics, which bind specifically to the ribosome and several proteins. The most important mechanism of resistance to tetracycline is regulated by its binding as a Tc:Mg^{2+} complex to the Tet Repressor protein (TetR). It is thus of interest to increase our understanding of both Tc:TetR and Tc:ribosome binding. Crystal structures of several tetracyclines in complexes with proteins and ribosome have provided essential information. A complementary approach is to develop computer simulation models, which can be used to investigate the structure, dynamics, and thermodynamics of Tc:protein or Tc:ribosome complexes. Despite its importance, tetracycline has rarely been subjected to computer modeling, partly because of the need to first develop a molecular mechanics model for Tc.

Here, we developed such a model, so as to be compatible with the CHARMM27 force field for proteins and nucleic acids, for 12 important tetracycline analogs, including plain tetracycline. Intermolecular force field parameters were derived from a standard supermolecule approach. The model reproduces the ab initio geometry and flexibility of each Tc. As tests, we did simulations of a Tc crystal, of Tc:Mg^{2+} and Tc:Ca^{2+} complexes in aqueous solution, and of a solvated complex between Tc:Mg2+ and the TetR. The model compares well with a wide body of experimental data.

We first used our model to study Tc:TetR recognition which is a complex problem. We used free energy simulations to investigate the electrostatic interactions between protein and ligand and the possible role of induced fit in Tc binding. We found that tetracycline prefers an extended, zwitterionic state both in solution and in complex with the protein. Tc is thus preorganized for binding.

In the absence of Tc, TetR is tightly bound to its operator DNA; upon binding of Tc it dissociates from the DNA, allowing expression of the repressed genes.

Its tight control by Tc makes TetR broadly useful in genetic engineering. The Tc binding site is over 20 $\mathring{A}$ from the DNA, so the binding signal must propagate a long distance. We use molecular dynamics simulations and continuum electrostatic calculations to elucidate the allosteric mechanism. When $[Tc:Mg]^+$ binds, the Mg^{2+} ion makes interactions with helix 8 of one TetR monomer and helix 6 of the other monomer, and helix 6 is pulled in towards the central core of the structure. Hydrophobic interactions with helix 6 then pull helix 4 in a pendulum motion, with a maximal displacement at its N-terminus: the DNA interface. The N-terminal residue of helix 4, Lys48, is highly conserved in DNA-binding regulatory proteins of the TetR class and makes the largest contribution of any amino acid to the TetR:DNA binding free energy. Thus, the conformational changes lead to a drastic reduction in the TetR:DNA binding affinity, allowing TetR to detach itself from the DNA.

We next used the model to study Tc binding to the ribosome and to elongation factor Tu (Ef-Tu). Crystal structures of Tc bound to the *Thermus thermophilus* 30S subunit show the same primary Tc binding site (called TET1), with the strongest Tc electron density, close to the A-site, consistent with an inhibitory role. A secondary Tc-binding site, called TET5 was also observed in two structures. We have done molecular dynamics (MD) simulations of 30S ribosomal subunits to characterize Tc binding and help resolve the ambiguity regarding the number and strength of Tc binding sites. We have presented evidence for predominant binding to TET1, showing that other reported binding sites are weaker and not highly occupied at physiological Tc concentrations.

Recently, the crystal structure of a complex between elongation factor Tu (Ef-Tu) and Tc was solved, raising the question whether Tc's binding to Ef-Tu has a role in its inhibition of protein synthesis. We show that the direct contribution of Ef-Tu to the free energy of Tc binding to the Ef-Tu:GDP:Mg complex is negligible;

rather, the binding can be solely attributed to Tc interactions with the Mg ion and the GDP phosphate groups. We also show that Ef-Tu does not exhibit any binding preference for Tc over the non-antibiotic, 4-dedimethyl-Tc, and Ef-Tu does not bind the Tc analogue tigecycline, which is a potent antibiotic. Overall, our results support the idea that Ef-Tu is not the primary target of tetracycline.

The articles presented below include both computational and experimental results. All the experimental work was done by Winfried Hinrichs and his collaborators. All the computational work was done by myself. The insights obtained in this work and the modeling techniques employed should be of interest for engineering improved Tc antibiotics and improved TetR proteins for gene regulation.

Contents

Abstract **2**

1 Introduction to tetracycline and its function **7**

1.1 Biochemical background 7

1.1.1 Discovery and development of tetracyclines 7

1.1.2 Structure of tetracycline (Tc) and its relation to function . . . 8

1.1.3 Tetracycline uptake 12

1.2 Mode of action 13

1.3 Resistance to tetracyclines 15

1.3.1 Ribosome protection 15

1.3.2 Pumping tetracyclines out of the cell 17

1.3.3 Efflux genes. Regulation of resistance gene expression 18

2 Molecular mechanics models for tetracycline analogs **25**

Article 1: The Tetracycline:Mg^{2+} Complex: A Molecular Mechanics Force Field. A. Aleksandrov and T. Simonson. Journal of Computational Chemistry (2006) 27:1517-1533 26

Article 2: Molecular mechanics models for tetracycline analogs. A. Aleksandrov and T. Simonson. Journal of Computational Chemistry (2008) in press 44

3 Tetracycline:Tet Repressor Recognition: Simulations and Experiments 79

Article 3: Protonation Patterns in Tetracycline:Tet Repressor Recognition: Simulations and Experiments. A. Aleksandrov, J. Proft, W. Hinrichs, and T. Simonson. ChemBioChem (2007) 8:675-685 79

Article 4: Tet Repressor Induction by Tetracycline: A Molecular Dynamics, Continuum Electrostatics, and Crystallographic Study. A. Aleksandrov, L. Schuldt, W. Hinrichs and T. Simonson. Journal of Molecular Biology (2008) 378:896-910. 81

4 Tetracycline binding to the ribosome and Elongation Factor Tu 111

Article 5: Molecular Dynamics Simulations of the 30S Ribosomal Subunit Reveal a Preferred Tetracycline Binding Site. 113

Article 6: Binding of tetracycline to elongation factor Tu, the Tet Repressor, and the ribosome: a molecular dynamics and continuum electrostatics study. A. Aleksandrov and T. Simonson. submitted 130

Conclusions 162

Bibliography 166

Appendix: Computational sidechain placement and protein mutagenesis with implicit solvent models. A. Lopes, A. Alexandrov, C. Bathelt, G. Archontis, and T. Simonson. Proteins (2007) 67:853-867. 171

Acknowledgments 187

Chapter 1

Introduction to tetracycline and its function

1.1 Biochemical background

> ... those sciences are vain and full of errors which are not born from experiment, the mother of certainty ...
>
> Leonardo da Vinci, 1452-1519

1.1.1 Discovery and development of tetracyclines

Chlortetracycline and oxytetracycline, were the first members of the tetracycline group, discovered at the end of 1940s as natural products of *Streptomyces aureofaciens* and *S. rimosus*, respectively. Other tetracyclines were identified later, either as naturally occurring molecules (e.g. tetracycline) or as semisynthetic products (e.g., doxycycline, minocycline). The most recently discovered tetracyclines are the semisynthetic group, known as glycylcyclines. One of the glycylcyclines, tigecycline

(Tygacyl), was selected for further development and was approved in the USA as a drug in 2005 [31].

1.1.2 Structure of tetracycline (Tc) and its relation to function

Tetracycline molecules comprise a linear fused tetracyclic nucleus (designated as rings A, B, C and D) to which a variety of functional groups are attached. The tetracycline scaffold is shown in Figure 1.1. The simplest tetracycline which displays antibacterial activity is 6-deoxy-6-demethyl-Tc, which is regarded as a minimum pharmacophore.

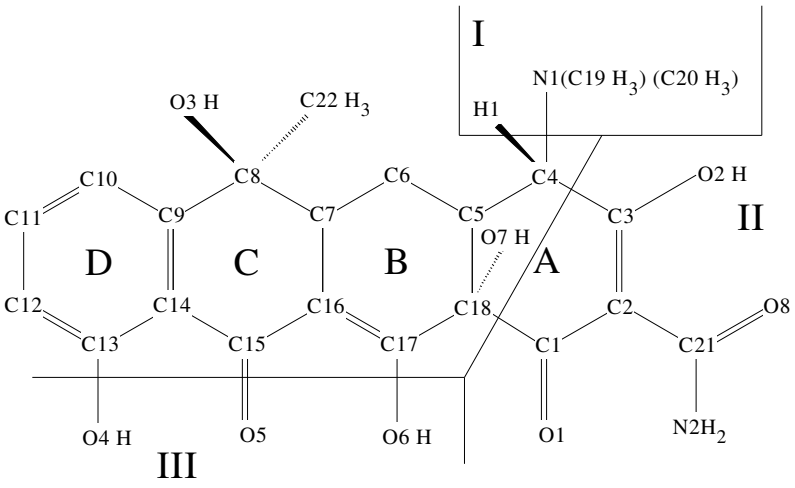


Figure 1.1: The neutral tautomer, Tc_N. The molecule is divided into regions I, II, III, to which specific acid/base behavior has been assigned (see text). The individual rings of the main scaffold are A–D.

The tetracycline features which are important for antibiotic activity are well established: maintenance of the linear, fused–ring conformation; a dimethylamino group at position 4b and the ketol-enol system (positions 11, 12). Under physio-

logical conditions, tetracycline binds its targets as a metal complex, by chelating a metal ion. The most important chelation site corresponds to the β-diketone system (positions 11 and 12). Thiatetracyclines, anhydrotetracycline and a number of other tetracyclines, referred to collectively as atypical tetracyclines, exhibit a different structure-activity relationship from the majority of tetracyclines. These molecules directly perturb the bacterial cytoplasmic membrane, leading to a bactericidal response. This contrasts with the typical tetracyclines which inhibit protein biosynthesis. These molecules are not the main interest of the current work since they interact nonspecifically with eukaryotic as well as prokaryotic cell membranes.

21 crystal structures of close Tc analogs are available in the Cambridge Crystallographic Data Bank (CCDB, 2006). These structures can be grouped into two conformational families, called the extended and twisted conformations, respectively (shown in Figure 1.2).

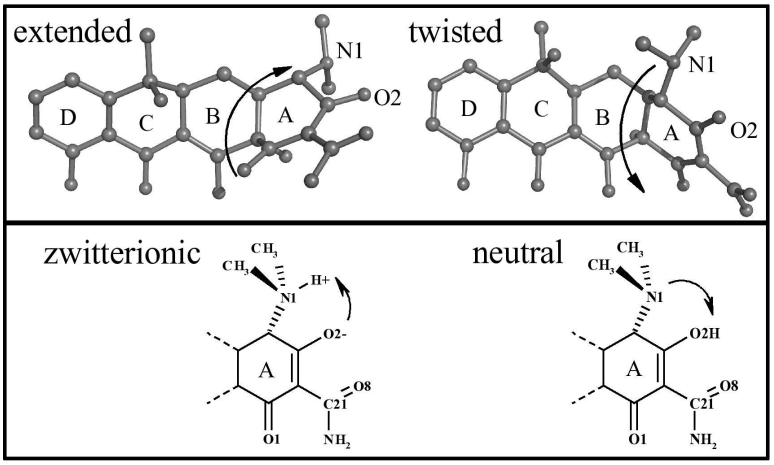


Figure 1.2: Upper panel: extended and twisted conformations of tetracycline; Bottom panel: Ring A in the zwitterionic tautomer, Tc_Z: O2 is deprotonated and N1 is protonated; The neutral tautomer, Tc_N: O2 is protonated and N1 is deprotonated.

The protonation of Tc can vary through binding/release of a proton in the region III and by proton exchange between O2 (neutral Tc) and N1 (zwitterionic Tc). The experimental proton dissociation constants, or pK_a's, for Tc in aqueous solution were reported as 3.3, 7.7, and 9.7 [30]. For a pH between 4.4 and 7.8, the N1/O2 region is globally neutral, with a proton on either N1 or O2.

In the CCDB [6, 7] structures, the oxygen protonation states can be inferred from the corresponding CO bond lengths, as summarized in Figure 1.3. We found that O5 is always deprotonated, while O4 and O6 are protonated. For the O2/N1 region, we find a mixture of the neutral and zwitterionic tautomers in the CCDB. From the observed bond lengths, we inferred that 13 structures are zwitterionic, and all of these are in the extended conformation. Six structures are neutral, and all of them are in the twisted conformation. The remaining two structures lack the atom N1.

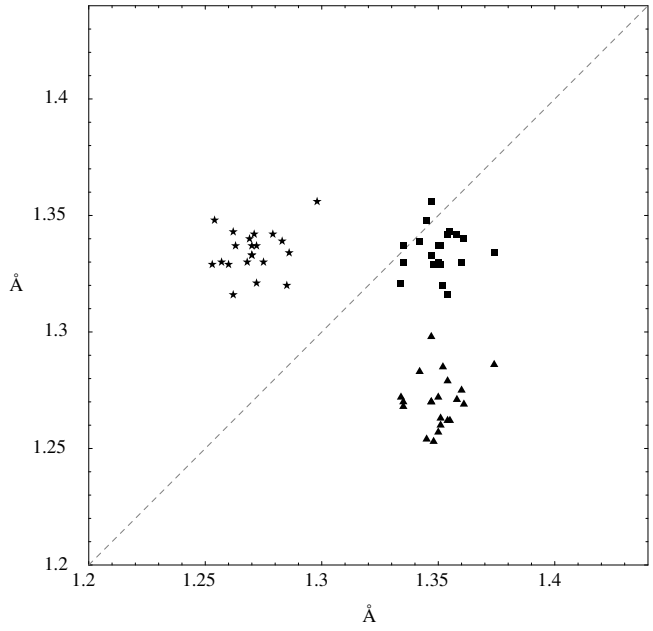


Figure 1.3: Selected CO bond lengths in the experimental Tc structures. Triangles – C13–O4 (X) and C15–O5 (Y) bond lengths for different experimental structures in the CCDB; boxes – C13–O4 (X) and C17–O6 (Y); stars – C15–O5 (X) and C17–O6 (Y). The short C15–O5 length indicates a deprotonated O5.

All structures of tetracycline complexes available in the PDB contain only tetracycline bound in the extended conformation. In the PDB, the limited experimental resolution does not allow the protonation state to be inferred. Thus the question of which Tc tautomer (neutral or zwitterionic) binds to TetR, Ef-Tu or the ribosome cannot be directly resolved from the PDB structures. In order to answer this question we need to know the free energy difference between the two tautomers in solvent; see Chapter 3. The energies of the most stable structures are shown in Table 2 for several Tc variants (see Chapter 2). It is known that substitutions

at the 4-dimethylamino position, for example in 4-dedimethylamino-tetracycline, invariably reduce antibacterial activity [12]. For all three Tc analogs that have a 4-dimethyl group, the most stable conformation in vacuum is neutral–twisted, while in solution it is zwitterionic–extended. In solution, all neutral Tcs are in the twisted conformation and all zwitterionic Tcs are in the extended one, consistent with the structures found in the CCDB (see above).

1.1.3 Tetracycline uptake

To interact with their targets, tetracyclines need to traverse one or more membrane systems depending on whether the susceptible organism is gram positive or gram negative [12]. Tcs traverse the outer membrane of gram-negative bacteria through the OmpF and OmpC porin chanels [23], as positively charged metal (probably magnesium) coordination complexes ($[Tc:Mg]^+$). The $[Tc:Mg]^+$ is attracted by the Donnan potential across the outer membrane, leading to accumulation in the periplasm, where tetracyclines probably dissociate from ions and become uncharged. This weakly lipophilic molecule is then able to diffuse through lipid bilayers of the inner (cytoplasmic) membrane and does not depend on protein channels. Similarly, the electroneutral, lipophilic form of tetracycline is assumed to be the species transferred across the cytoplasmic membrane of gram-positive bacteria. Uptake of tetracyclines across the cytoplasmic membrane is energy-dependent and driven by the δpH component of the proton motive force [23]. In the cytoplasm, Tc can be converted to an ionic compound again since the internal pH and the Mg^{2+} concentration are higher than in the periplasm. In accord with this, the uphill accumulation of tetracycline is only driven by pH and not by the transmembrane electrical potential and is strongly temperature-dependent and also influenced by pH and Mg2+ concentration in the extracellular medium. The active drug species which binds to the ribosome is a metal-tetracycline complex (see Chapter 4), thus Mg ion is needed

for the Tc proper function.

1.2 Mode of action

interaction between the ribosome and antibiotic

It was established that the tetracycline mode of action is associated with its inhibition of protein synthesis [12]. Thus tetracycline belongs to the class of bacteriostatic antibiotics, which only inhibit growth and reproduction of bacteria without killing them.

It was proposed that tetracyclines inhibit bacterial protein synthesis by preventing the binding of aminoacyl-tRNA to the A-site of the bacterial ribosome [11]. Extensive biochemical and genetic studies have given results that are somewhat difficult to interpret. Photoaffinity labeling and chemical footprinting studies indicate that protein S7 and several 16S rRNA bases (G693,A892,U1052,C1054,G1300, and G1338) contribute to the binding pocket [18, 23]. Interestingly, in the available X-ray structures of 30S ribosomal subunit in complex with tetracycline, tetracycline does not interact with any proteins directly. It was pointed out [8, 23] that these apparent sites for the drug in the ribosome may not reflect the actual binding sites, because tetracycline photoproducts may react further with the ribosome.

Crystal structures of Tc bound to the *Thermus thermophilus* 30S subunit were determined by two groups [9, 20]. Both groups found the same high affinity, primary, site (called TET1), with the strongest Tc electron density. Additionally, both groups reported a secondary Tc-binding site, called TET5. Apart from these two Tc binding sites, one of the groups reported four other low–occupancy sites [20]. TET1 is close to A-site of the ribosome which is consistent with an inhibitory role; the less occupied site is at the interface between three RNA domains in the body of the subunit. In the primary site, Tc interacts exclusively with the 3' major domain of

16S RNA, right above the binding site for aa-tRNA. The binding pocket is formed be an irregular minor grove of H34 and residues 964–967 of H31. Tc interacts primarily with the exposed sugar-phosphate backbone of H34. A putative Mg^{2+} ion of the Tc:Mg chelate coordinates the phosphate oxygens of C1054, G1197 and G1198. This ion is present practically in the same position in the X-ray structure determined in the absence of Tc (PDB refcode 2V46 [33]). There is no overall conformational change in the 16S RNA upon Tc binding to the ribosome. The second binding site is located in the body of the subunit near H44 and sandwiched between H27 and H11. Again, all interactions are mediated by the rRNA; there are no interactions between Tc's in the primary or secondary sites and ribosomal proteins. No Mg^{2+} ion was observed, but this is not conclusive since electron density was weaker in this region. One of the groups reported four additional Tc sites. Both groups discussed that the primary binding site alone could explain Tc's bacteriostatic effect alone.

Molecular dynamics simulations of the ribosome reveal a preferred tetracycline binding site

Chapter 3 describes molecular dynamics (MD) simulations of the 30S ribosomal subunits to characterize Tc binding and help resolve the ambiguity regarding the number and strength of Tc binding sites. We compared the binding free energies of Tc to TET1 and TET5; other secondary sites were not considered, because they were unseen by one group and thus expected to have affinities lower than TET5. The free energy calculations then showed that TET1 binding is indeed stronger than TET5 binding, as suggested by the weaker TET5 electron density. With roughly-physiological concentrations, [Ribosome]=[Tc]=1microM, TET5 is largely unoccupied. This suggests that allosteric coupling between binding sites is not important for the function of tetracycline.

Tetracycline binding to EF-Tu

The bacteriostatic mode of action of tetracycline has been extensively studied over the last 50 years and it is well established that tetracycline blocks protein biosynthesis [11, 23]. As discussed above, the primary molecular target of tetracycline binding is believed to be the ribosome. Recently, however, the possibility was raised that elongation factor Tu (Ef-Tu) might be one of several tetracycline targets, since tetracycline was reported to crystallize as a 1:1 complex with the trypsin-modified form of *E.coli* Ef-Tu-Mg:GDP (tm-Ef-Tu-Mg:GDP) [17]. The structure of the complex was recently published [15]. The results demonstrate that tetracycline not only binds to tm-Ef-Tu-Mg:GDP, but does so by interacting with an important structural motif that is shared by all GTPases and many ATPases [15]. We address this question in Chapter 3 by modern molecular dynamics and free energy simulation methods. Our results show that the contribution of Ef-Tu to the binding free energy of tetracycline to Ef-Tu:GDP:Mg complex is negligible, while the binding is solely attributed by tetracycline interactions with the Mg ion and GDP phosphate groups. Further free energy calculations showed that Ef-Tu does not discriminate binding of tetracycline and of the non-antibiotic 4-dedimethyl-tetracycline, while the ribosome and also Tet Repressor do show specifity for these compounds. Also, tigecycline which is one of the most potent tetracycline antibiotics, does not bind to Ef-Tu. Overall, our results suggest that Ef-Tu is not the primary target of tetracycline.

1.3 Resistance to tetracyclines

1.3.1 Ribosome protection

Ribosomal protection represents one of the important tactics for tetracycline resistance in both gram-positive and gram-negative bacteria [13]. There are only two ribosomal protection proteins (RPP), TetO and TetM, that have been studied in de-

tail. Other members of the RPP class (TetS, TetT, TetQ, TetBP, TetW and OtrA) are assumed to function through similar mechanisms [13].

TetO and TetM are soluble cytoplasmic proteins with 75% sequence similarity. It is believed that RPPs have originated in the natural producer of tetracyclines and then spread throughout the eubacteria via lateral gene transfer events [13]. The RPPs display sequence similarity to the ribosomal elongation factors EF-G and EF-TU and are in the group of the translation factor superfamily of GTPases.

RPPs can dislodge tetracycline from the ribosome, and thus increase the apparent dissociation constant. The process is strictly dependent on the presence of GTP [10], although there is some discrepancy concerning the role of GTP, since a nonhydrolyzable GTP analogue was active with TetO and partially active with TetM. In accord with the ability of RPPs to remove tetracycline, it was shown that tRNA binding to the A site, which is normally inhibited by tetracycline, is protected in the presence of TetM [10]. Thus it seems that RPPs confer tetracycline resistance by releasing tetracycline from the ribosome and thereby freeing the ribosome from the inhibitory effects of the drug.

It was shown that the RPP binds to same binding site as elongation factor G and suggested that a component of this common site is the L11 region on the 50S subunit. Studies by cryo-EM showed that the overall shape of the Rib:TetO complex is similar to the Rib:EF-G complex, in agreement with the sequence similarity noted above [29]. A comparison of the EF-G and TetO ribosomal contacts indicates that they differ primarily in the vicinity of the domain IV, where EF-G contacts H69 of the 23S rRNA and TetO interacts with H18/H34 of the 16S rRNA [29]. It was postulated that differences in domain IV may serve to distinguish TetO and EF-G with respect to their activities.

A detailed mechanism for TetO-mediated tetracycline resistance can thus be proposed. In the presence of tetracyclines, the ribosome is blocked in the postranslo-

cational state because subsequent A-site occupation is inhibited by tetracycline. This blockage is due to direct steric clashes between tetracycline and the accommodating aa-tRNA; no gross conformational changes were observed in the crystal structures of tetracycline bound to the 30S subunit[9, 20]. In the presence of TetO, TetO would bind to the tetracycline-blocked ribosome, probably to form the ternary complex from which tetracycline is released. It was shown by cryo-EM studies [29] that when when TetO binds to the ribosome, it does not overlap with the primary tetracycline binding site. This suggests that TetO triggers the release of tetracycline through an allosteric mechanism [29]. The proposed conformational changes resulting in tetracycline release probably involve H34, as cryo-EM reconstructions show that domain IV of TetO contacts bases of H34, and H34 forms an integral part of the primary site.

In the presence of tetracycline, the ribosome is blocked with an open A site, which could provide a kinetic window for TetO to act, thus distinguishing the tetracycln-blocked ribosome from a translating ribosome [13].

1.3.2 Pumping tetracyclines out of the cell

Antibiotic activity of the drug depends on its presence in the bacterium. Lowering the amount of tetracyclines in the bacterium can effectively decrease suceptibility of the cell. A reduced tetracycline concentration is achieved by pumping out of the bacterial cell via special efflux proteins. The efflux proteins are the best studied Tet proteins. The genes encoding efflux proteins are found in both gram-positive and gram-negative species. Most of these efflux proteins confer resistance to tetracycline but not to glycylcyclines and minocycline [12] (only in gram-positive bacteria).

Over a dozen classes of tetracycline resistance determinates, encoding drug-specific efflux proteins, are known to date, and designated as TetA–E, G–L, P, Z, and OtrB. TetA–E, G and H are found in gram-negative bacteria; TetK, L, P,

and OtrB occur in gram-positive bacteria. Each efflux gene codes for a membrane bound protein. The tetracycline resistance proteins share high sequence similarity, 40–80% amino acid identity [12]. The proteins in gram-negative species have 12–14 predicted transmembrane alpha–helices and thus appear to reside in the lipid bilayer. The efflux proteins exchange a proton for a tetracycline–cation complex transferred out of the cell against a concentration gradient.

1.3.3 Efflux genes. Regulation of resistance gene expression

Efflux determinants in gram-negative bacteria share a common genetic organization, which is different in gram–positive species: tetR, which encodes a tetracycline-responsive repressor, is located next to tetA in a divergent orientation. In the absence of tetracyclines, TetR is bound to two palindromic operator sites (tetO1,2), thereby repressing the expression of tetR and tetA. Binding of [MgTc]+ to TetR abolishes DNA binding to the operator sites and thus allows transcription of tetA and tetR.

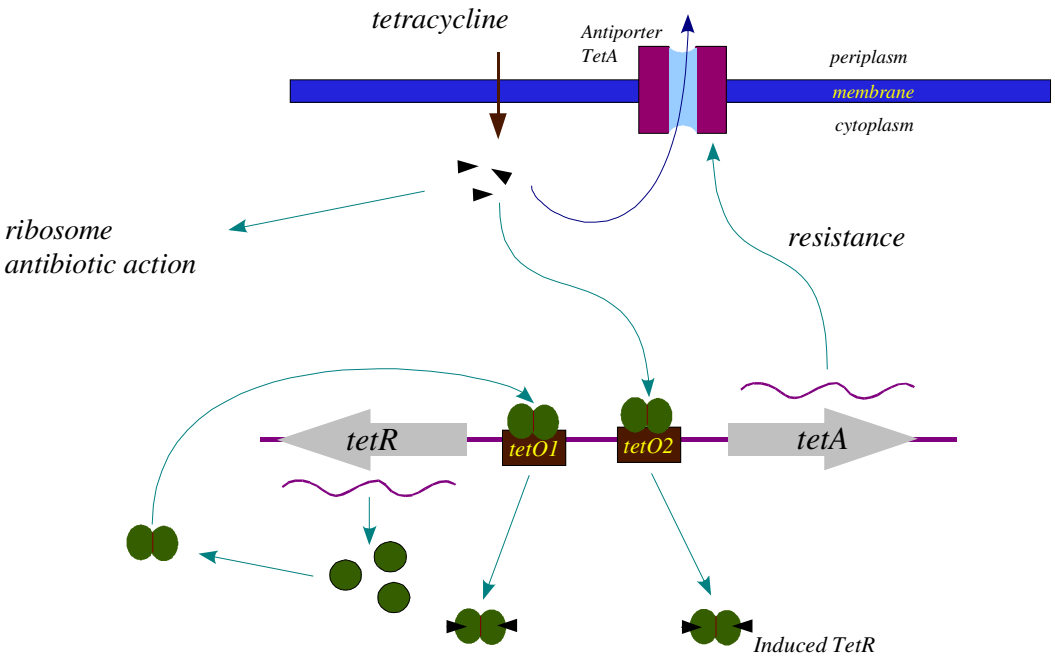


Figure 1.4: Schematic description of tetracycline efflux resistance reproduced from [22]. Tet repressor (green spheres) is bound as a homodimer to the operators tetO1 and tetO2 in the absence of tetracyclines. Tc (black triangles) diffuses into the cell and binds to TetR:tetO to enforce conformational changes so that the induced TetR releases the operators and transcription of genes tetR (coding TetR) and tetA (coding the resistance protein TetA) can now proceed (wavy lines represent mRNA). After translation, TetA binds into the membrane to export invading tetracyclines out of the cell to prevent Tc attack on the ribosome.

High level expression of TetA is lethal for the bacterial cell because nonspecific

cation transport causes the transmembrane potential to collapse. This demands a well-defined repressed status of the tetA gene by strong binding of TetR to tetO, with an association constant Ka 10^{11} M^{-1}). Tc binding to the ribosome is five orders of magnitude lower (Ka = 10^6 M^{-1}), thus TetR is required to be a particularly sensitive switch. Binding of [MgTc]+ to TetR (Ka = 10^9 M^{-1}) reduces the affinity of TetR for tetO by nine orders of magnitude [22]. This guarantees that the silent state (TetR bound to tetO) and the active state (TetR bound to [MgTc]+) of tetO are efficiently discriminated and subinhibitory Tc concentrations are sufficient to allow transcription to proceed. This makes TetR a unique biological switch, which is commonly used for the regulation of gene expression in transgenic organisms.

Structure of the Tet repressor

X-Ray structure analysis showed that the tetracycline repressor (TetR) occurs as a homodimer $(TetR)_2$ see Figure 5, and in the absence of $[Mg:Tc]^+$ it binds via two identical helix-turn-helix (HTH) motifs to two adjacent major grooves of the DNA.

The two monomers of TetR are identical, consisting of around 200 amino acids, each folded into ten α-helices. In the final TetR complex, the two monomers are related by a twofold symmetry [22]. The DNA binding domain (DBD) consists of the N-terminal three-helix-bundles, which is linked to the regulatory core domain by α-helix 4. The central part of the core domain consists of the four-helix-bundle (antiparallel helices 8 and 10 which cross the related helices 8' and 10' of the other monomer). It is possible to see that this four-helix-bundle is the most rigid part of TetR when X-Ray structures of TetR in complex with DNA or Tc are superimposed. There are two clearly-seen, tunnellike [Mg:Tc]+ binding cavities formed by helices 5, 6, 7, 8' (a prime is used to designate the second TetR monomer) and 9'. In the TetR:Tc complex, tetracyclines are completely buried inside their binding pockets with distance to the DNA interface of over 20 $\AA$.

Tetracycline:Tet Repressor recognition

Crystal structures of TetR have provided essential information on interactions between tetracycline and TetR. Tetracycline makes important hydrogen bonds between its ring A and His64, Asn82 and Gln116 of Tet repressor. These amino acids are found in comparable positions in the crystal structures of apo-TetR, TetR bound to the DNA, and in the induced form, TetR:Tc. Tetracycline also makes van der Waals contacts with Val113', Leu131',Ile134' Leu170', Leu174', Met177' and Pro105. Tc-chelated Mg^{2+} coordinates three water molecules, which mediate interactions with Glu147' and Thr103. Interactions with the Tc-chelated Mg^{2+} ion displace His100 and Thr103 of helix 6, which in turn shifts helix 6 and peels off of its C-terminal turn to form a type II β–turn in the induced TetR:Tc structure.

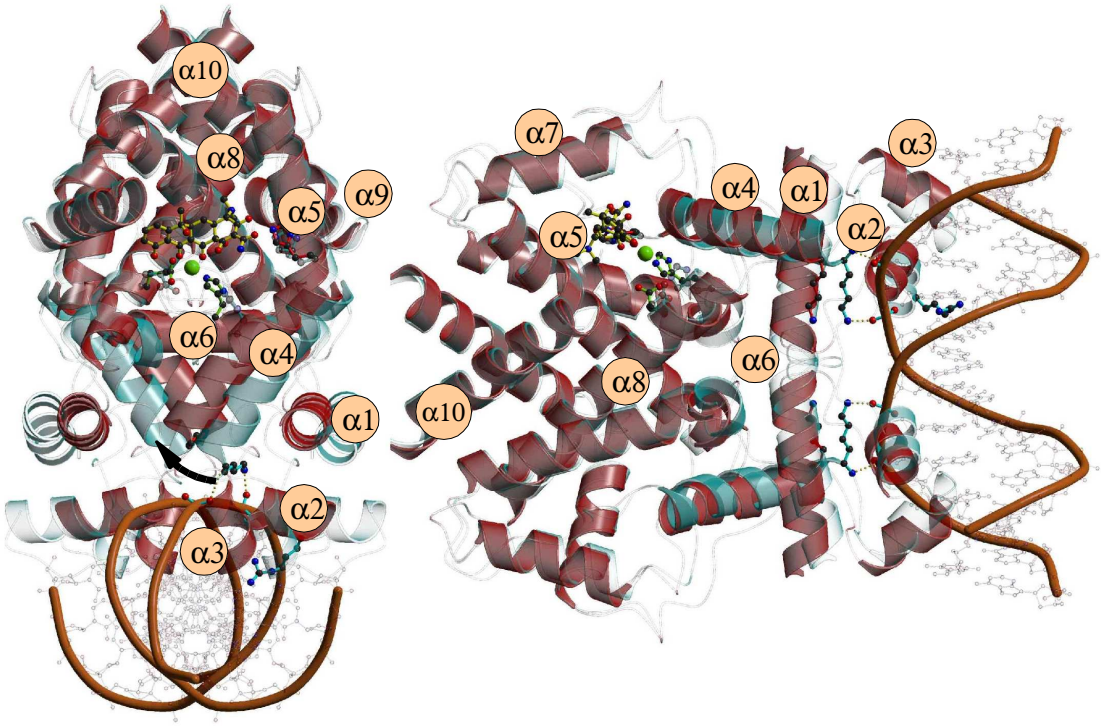


Figure 1.5: An overview of the TetR structure in its induced and non-induced states. Two orthogonal views are shown. The TetR:DNA complex is red; the DNA backbone is shown as a tube; bases are shown as sticks (light grey). Superimposed (light blue) is the induced, TetR:Tc structure. Tc is shown with a ball-and-sticks representation. The Mg^{2+} ion is a yellow sphere. Helices are labelled. Only one Tc is shown (monomer 1). Lys48 is shown (ball-and-sticks) making hydrogen bonds to the DNA. The helix 4 shift upon induction is highlighted by an arrow.

Many details of the Tc:TetR binding process are not evident from the available experimental data. In particular, tetracycline has two main conformations and two main protonation states at neutral pH. When Tc binds to TetR its conformation/protonation may change. Similarly, protein groups that interact with Tc have several possible orientations and protonation states, which could change when

Tc binds. The medium-resolution X-Ray structure does not reveal the protonation states of Tc or the protein sidechains, nor the orientation of several key recognition residues. In particular, two orientations are possible for Asn82, Gln116 and His64. To investigate the protonation patterns and conformational states in the TetR:Tc complex, we have performed molecular dynamics free energy simulations, which were subsequently confirmed by crystallographic calculations. We computed the Tc binding free energy as a function of its protonation state and the protonation and conformation state of the binding pocket. In order to validate the results, two different free energy methods were employed, which use either an explicit treatment of aqueous solvent or a less expensive, continuum dielectric treatment. We predict the most stable form of the TetR:Tc complex and the the extent of the protein and ligand preorganization. This work is described in Chapter 3.

Tet repressor induction by tetracycline

Binding of tetracyclines to TetR triggers a sequence of conformational changes, which reduce the TetR binding affinity for the DNA by 6–10 orders of magnitude. TetR then dissociates from tetO and expression of tetA and tetR can proceed. The amino acids His64, Asn82 and Gln116 which make hydrogen bonds with the tetracycline ring A are found practically in the same positions in the TetR:DNA structure and in the induced TetR:Tc X–Ray structure, which suggests that they do not transfer a binding signal directly, but rather provide binding specificity for Tc, as well as rigidifying the induced structure. In the absence of Tc, the binding tunnel is presumably filled with disordered waters, which are not seen in the electron density maps. The small alpha helix 6 is two turns long, with residues conserved or type-conserved among different classes of TetR. Upon Tc binding, helix six is displaced to form direct and water-mediated interactions with the Mg^{2+} ion chelating Tc. The C-terminal end of helix 6 unwinds and forms a type II beta-turn (residues 100–103).

As a result, the loop connecting alpha 6 and 7 is shifted by 2.5 Å, so Arg 104 can form a salt bridge with Asp178', which closes the binding tunnel. The shift of helix 6 is accompanied by a pendulum motion of two other alpha helices, helices 4 and 4'. This displaces the DBDs, so that they no longer fit into adjacent major grooves of the DNA. According to the model inferred from the above structural data, the shifts near Tc:Mg ligand trigger the helix displacements in a manner that has not yet been determined. In this model, the main role of Tc is to act as a Mg^{2+} carrier, binding specifically to TetR, rigidifying the surrounding groups, and bringing Mg^{2+} into its proper position to trigger the allosteric motions.

The model discussed above is mainly based on crystal structures of the end-point states, TetR in complex with DNA and induced TetR. Additional information has been provided by biochemical and biophysical measurements. In Chapter 3, we reexamine the allosteric mechanism of TetR induction using molecular dynamics and continuum electrostatics calculations.

Chapter 2

Molecular mechanics models for tetracycline analogs

In order to model Tc interactions with protein and RNA, we have first developed a molecular mechanics force field model for 12 tetracyclines, including doxycycline, anhydrotetracycline, minocycline and tigecycline, which is consistent with the CHARMM27 force field for proteins and nucleic acids. We used structures from the Cambridge Crystallographic Data Base to identify the conformations that are likely to be present in solution and in biomolecular complexes. A conformational search was also undertaken, where simulated annealing was used to generate structures of tetracycline, anhydrotetracycline, doxycycline and tigecycline. Several resulting, low-energy structures were optimized with an ab initio model. We found that Tc and its analogs all adopt an extended conformation in the zwitterionic tautomer and a twisted one in the neutral tautomer, and the zwitterionic-extended state is the most stable in solution, consistent with experimental data. We considered a zwitterionic tautomeric form of each Tc variant, in the extended conformation, with and without bound Mg^{2+}. Intermolecular parameters were derived from a supermolecule ab initio approach; The final, rms deviation between the ab initio and force

field energies, averaged over all forms, was 0.35 kcal/mol. The model reproduces the ab initio geometry and flexibility of each Tc. As tests, we did simulations of a solvated complex between three Tcs and the Tet repressor protein (TetR).

Article 1: The Tetracycline:Mg^{2+} Complex: A Molecular Mechanics Force Field. Alexey Aleksandrov and Thomas Simonson. *J. Comput. Chem.,* 27:1517-1533, 2006

Article 2: Molecular mechanics models for tetracycline analogs Alexey Aleksandrov and Thomas Simonson. *J. Comput. Chem.,* 2008, in press.

WILEY InterScience® DISCOVER SOMETHING GREAT

The Tetracycline:Mg^{2+} Complex: A Molecular Mechanics Force Field

ALEXEY ALEKSANDROV, THOMAS SIMONSON*
Laboratoire de Biochimie (CNRS UMR7654), Department of Biology, Ecole Polytechnique, 91128 Palaiseau, France

Received 10 December 2005; Revision 23 January 2006; Accepted 6 February 2006
DOI 10.1002/jcc.20453
Published online in Wiley InterScience (www.interscience.wiley.com).

Abstract: Tetracycline (Tc) is an important antibiotic, which binds specifically to the ribosome and several proteins, in the form of a Tc^{-}:Mg^{2+} complex. To model Tc:protein and Tc:RNA interactions, we have developed a molecular mechanics force field model of Tc, which is consistent with the CHARMM force field for proteins and nucleic acids. We used structures from the Cambridge Crystallographic Data Base to identify the main Tc conformations that are likely to be present in solution and in biomolecular complexes. A conformational search was also done, using the MM3 force field to perform simulated annealing of Tc. Several resulting, low-energy structures were optimized with an *ab initio* model and used in developing the new Tc force field. Atomic charges and Lennard-Jones parameters were derived from a supermolecule *ab initio* approach. We considered the *ab initio* energies and geometries of a probe water molecule interacting with Tc at 36 different positions. We considered both a neutral and a zwitterionic Tc form, with and without bound Mg^{2+}. The final rms deviation between the *ab initio* and force field energies, averaged over all forms, was just 0.35 kcal/mol. The model also reproduces the *ab initio* geometry and flexibility of Tc. As further tests, we did simulations of a Tc crystal, of Tc:Mg^{2+} and Tc:Ca^{2+} complexes in aqueous solution, and of a solvated complex between Tc:Mg^{2+} and the Tet repressor protein (TetR). With slight, *ad hoc* adjustments, the model can reproduce the experimental, relative, Tc binding affinities of Mg^{2+} and Ca^{2+}. It performs well for the structure and fluctuations of the Tc:Mg^{2+}:TetR complex. The model should therefore be suitable to investigate the interactions of Tc with proteins and RNA. It provides a starting point to parameterize other compounds in the large Tc family.

© 2006 Wiley Periodicals, Inc. J Comput Chem 27: 1517–1533, 2006

Key words: force field; antibiotic; molecular dynamics

Introduction

Tetracycline (Tc) is one of the most important antibiotics in use today.[1,2] Several thousand varieties have been used as broad spectrum antibiotics for both humans and livestock. The most common bacteriostatic action of Tc is by inactivation of the bacterial ribosome, so that protein biosynthesis is interrupted and the bacteria die. The most common mechanism of resistance in gram-negative bacteria is associated with the membrane protein TetA, which exports Tc out of the bacterial cell before it can attack its target, the ribosome. The expression of TetA is tightly regulated by the homodimeric Tet repressor (TetR), which binds specifically to operator DNA, upstream of the tetA gene. When Tc diffuses into the cell, it is negatively charged, having released a proton, and chelated to form a Tc^{-}:Mg^{2+} complex. It is the Tc^{-}:Mg^{2+} complex that binds to TetR, inducing conformational changes that sharply reduce the affinity of TetR for the DNA.[3] TetR is then released from the DNA and expression of TetA can proceed, conferring resistance on the bacterial cells.

It is thus of interest to increase our understanding of both Tc:TetR and Tc:ribosome binding. Crystal structures of TetR and TetR:DNA complexes have provided essential information,[3] and the structure of the entire ribosome was recently determined.[4] A complementary approach is to develop computer simulation models, which can be used to investigate the structure, dynamics, and thermodynamics of Tc:protein or Tc:ribosome complexes. Despite its importance, TetR has rarely been subjected to computer modeling,[5,6] partly because of the need to first develop a molecular mechanics model for Tc. Here, we

***Correspondence to:** T. Simonson; e-mail: thomas.simonson@polytechnique.fr

Contract/grant sponsors: Commmissariat pour l'Énergie Atomique; Centre National pour la Recherche Scientifique

© 2006 Wiley Periodicals, Inc.

report such a model, developed so as to be compatible with the CHARMM22 force field for proteins and nucleic acids,[7,8] and with the TIP3P water model.[9]

Thirty structures of molecules of the Tc family are available in the Cambridge Crystallographic Data Base (CCDB) and there are six TetR:Tc complexes in the Protein Data Bank (PDB), involving three different Tc variants. Tcs are fairly rigid, and they exist in two main conformations: the so-called twisted and extended conformations, characterized by the orientation of the main rings, A–D, of the four-ring Tc scaffold (Fig. 1). All the PDB complexes involve the extended Tc conformation. Tc has several probable protonation states at neutral pH, which must be taken into account.[10] Proton binding constants corresponding to different titratable groups of Tc were measured (Fig. 1), but the assignment of these values to specific proton acceptors is still unclear. The "neutral" state, Tc_N, and the "zwitterionic" state, Tc_Z, are shown in Figure 1. They differ by the transfer of a proton from the O2 hydroxyl to N1 of the nearby dimethylammonium group. Binding constants between Tc and several divalent metal ions (Mg, Ca, Ba, Mn, Fe, Co, Ni, Zn, Cd) are known, but the specific binding sites are not all known (W. Hinrichs, unpublished results). Binding constants of most of the corresponding (metal:Tc)$^{+}$ complexes to TetR are also known (W. Hinrichs, unpublished). From the PDB structures, Mg^{2+} is known to bind to Tc between O5 and O6. When it binds, the H6 proton is released, giving a Tc^{-}:Mg^{2+} complex. The question of which Tc protonation state, zwitterionic or neutral, binds to TetR is still open.

Only limited computational work has been done on Tc, partly due to the complexity of this molecule. Several studies explored its conformation space.[5,6] Clark and coworkers used quantum mechanical, Density Functional Theory (DFT) to investigate the conformations of neutral tetracycline (Tc_N) and anhydrotetracycline (aTc) in aqueous solution. They found that the extended Tc_N conformation is about 3 kcal/mol more stable than the twisted one, and the energetic preference for the extended conformation is a solvent effect. For aTc, solvent contributions make the neutral form (in its extended conformation) more favorable than the zwitterionic aTc. For non-ionized aTc, the twisted conformation is the most stable in water at 298 K. Thus, aTc is predicted to adopt the twisted conformation in its neutral form in solution, and the extended conformation in its zwitterionic form.

Here, we report force field parameters for the most important Tc protonation states at neutral pH: Tc_N, Tc_Z, Tc_N^{-}:Mg^{2+}, and Tc_Z^{-}:Mg^{2+}. We only considered the extended conformation, which is found in all six PDB structures. We used experimental data from the CCDB to identify the main Tc conformations that are likely to be present in solution and in biomolecular complexes. A conformational search was also undertaken, using the MM3 force field to perform molecular dynamics and simulated annealing of Tc. Clustering of the resulting conformations led to several low-energy structures that were further optimized with an *ab initio*, quantum mechanical model. These structures were used in developing the four CHARMM force field models. The atomic charges and Lennard-Jones parameters were derived from a supermolecule *ab initio* approach. We considered the *ab initio* energies and geometries of a water molecule interacting with Tc at 36 different positions. This supermolecule approach is known

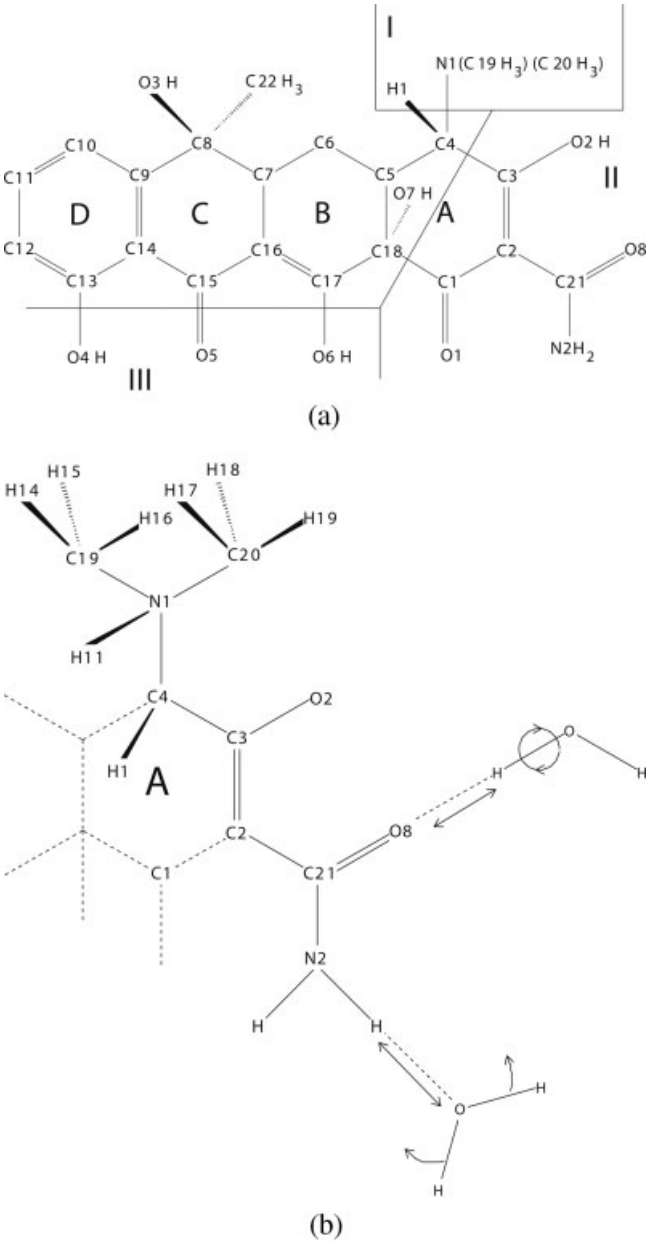


Figure 1. (A) The neutral tautomer, Tc_N. The molecule is divided into regions I–III, to which specific acid/base behavior has been assigned (see text). The rings of the main scaffold are labeled A–D. (B) Ring A in the zwitterionic tautomer, Tc_Z. O2 is deprotonated and N1 is protonated. Two probe water positions are shown; for each one a single distance and angle may be optimized.

to give a good balance between solute-water and water-water interactions in the force field.[7,8] Good agreement was obtained between the *ab initio* and force field data (rms deviation for the energies of 0.35 kcal/mol). The model also reproduces well the *ab initio* geometry and flexibility of Tc, and the experimental crystal structures. To test the parameters further, we performed simulations of a Tc_Z crystal, of Tc_Z^{-}:Mg^{2+} and Tc_Z^{-}:Ca^{2+} com-

plexes in aqueous solution, and of solvated Tc_N^-:Mg^{2+}:TetR and Tc_Z^-:Mg^{2+}:TetR complexes. With slight, *ad hoc* adjustments, the model can be made to reproduce the experimental, relative, Tc binding affinities of Mg^{2+} and Ca^{2+}. It performs well for the structure and fluctuations of the Tc:Mg^{2+}:TetR complexes. The model should therefore be suitable to investigate the interactions of Tc with proteins and RNA, and particularly the Tc:TetR and Tc:ribosome complexes. It provides a starting point to parameterize other compounds in the large Tc family.

In the following sections, we present our Computational Methods, Results, and Conclusions.

Computational Methods

Experimental Tc Structures

We searched the 2005 release of the Cambridge Crystallographic Data Bank (CCDB)[11,12] and retrieved 21 crystal structures of Tc, out of 30; the other nine were chemically too different from the Tc variant of interest here (Fig. 1). We also retrieved six crystal structures from the Protein Data Bank, corresponding to complexes between a Tc and the Tet Repressor. Based on these structures, we considered two main protonation states of Tc: a "neutral" state, Tc_N, and a "zwitterionic" state, Tc_Z, schematized in Figure 1. The structures can also be grouped into two conformational families, called the "extended" and "twisted" conformations, respectively.

Conformational Searching with Simulated Annealing

The MM3 force field[13] and the TINKER program[14] were used to do molecular dynamics and simulated annealing of the Tc molecule. The temperature was varied from 1200 to 0 K over a period of 50 ns. Conformations were drawn from the trajectory every 5 ps and quenched by energy minimization. Clustering was performed with the CHARMM program,[15] using the ART-2 algorithm.[16] The distance between structures was measured by the dihedral angles of the main Tc rings. With a 10° threshold, we obtained about 50 clusters. The structures corresponding to the cluster centers with the lowest energies were further optimized quantum mechanically at the B3LYP/6-31G(d) level.

Optimization of the Intermolecular Force Field Parameters

We adopt a force field of the CHARMM22 form,[7] which will later be used to simulate complexes between Tc and proteins or RNA in aqueous solution. The intermolecular energy terms are Lennard-Jones and Coulomb terms. In accord with the development of the CHARMM22 force field, we rely on supermolecule, quantum mechanical calculations on complexes between Tc and a single water molecule. The position of the water is varied around the Tc, primarily at sites involving hydrogen-bonding interactions with polar Tc atoms. The geometry of Tc was taken from the conformational search, above. For the force field calculations, the water geometry was taken from the TIP3P model.[9]

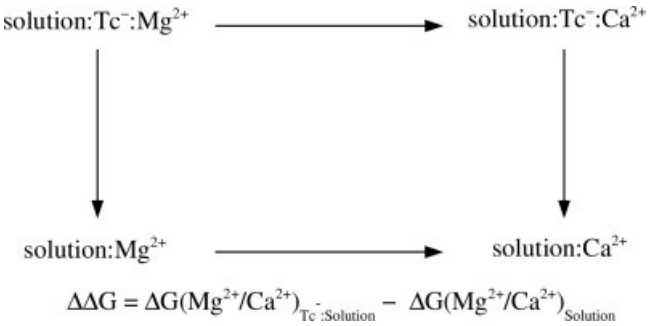


Figure 2. Thermodynamic cycle for Tc:cation binding. Vertical legs represent cation binding/dissociation; horizontal legs represent the alchemical transformation of Mg^{2+} into Ca^{2+} in solution (below) or in the solvated cation:Tc complex (above).

Each supermolecule structure was optimized at the HF/6-31G(d) level by varying the interaction distance and a single angle (Fig. 1B), to find the local minimum for the water position. From the resulting optimal structure, the interaction energy was calculated. No correction for basis set superposition error was made. The *ab initio* interaction energies were scaled by a factor of 1.16. This factor leads to good agreement between the HF/6-31G(d) Hamiltonian and the TIP3P model for the energy of a water dimer.[17] Applying the same factor to the Tc-water energies should thus lead to a force field with a correct balance between the Tc-water and water-water interactions. Also, the *ab initio* interaction distances were reduced by 0.2 Å,[7] as in previous CHARMM22 force field work. This is done partly to compensate for overestimated interaction distances with the Hartree–Fock model (due to neglected electron correlation) and partly to compensate for the expected volume decrease in going from the gas phase to an aqueous condensed phase.

The force field parameters were then adjusted to reproduce these "corrected" *ab initio* interaction energies and water positions. Initial partial charges were obtained from related fragments in the CHARMM22 force field, where available, and otherwise from a Mulliken population analysis of the HF/6-31G(d) wavefunction. Initial Lennard-Jones parameters were taken from similar molecules for which CHARMM parameters are known.

For the neutral form of Tc, we considered 26 water positions. For 22 of them, we optimized the interaction distance and one angle, corresponding to a rocking of the water around the water-Tc axis. For eight positions, including four new ones, only the interaction distance was optimized. The (corrected) *ab initio* data were then fitted by varying manually the Tc charges and Lennard-Jones parameters. This involved reoptimizing the Tc-water distance and angle after each parameter change.

The parameters optimized for the neutral form of Tc then served as initial guesses for the zwitterionic form. For this form, only eight water positions were considered, with two degrees of freedom each. Indeed, the differences between the zwitterionic and neutral Tc forms are fairly local, so we could concentrate on one portion of the Tc molecule. The parameters were refined by manual adjustment.

Table 1. Selected Bond Lengths (Å) in Tc Crystal Structures.

CCDB Structure	C13-O4	C15-O5	C17-O6
AOTETC	1.350	1.272	1.337
BINNAN	1.348	1.253	1.329
CMTCYH10	1.350	1.257	1.330
DEMXTC10	1.347	1.270	1.333
DEMXTC20	1.347	1.270	1.333
DXTETC	1.360	1.275	1.330
DXXTEC	1.352	1.285	1.320
EPCTCY10	1.335	1.268	1.330
EPMXTC10	1.351	1.260	1.329
EPOXTC10	1.355	1.262	1.343
JAVCIS10	1.351	1.263	1.337
$JUNLUZ_1$	1.334	1.272	1.321
OTETCB	1.347	1.298	1.356
OXTETD	1.374	1.286	1.334
OXTETH1	1.345	1.254	1.348
OXYTET	1.335	1.270	1.337
OXYTET01	1.361	1.269	1.340
TBDMXT01	1.354	1.279	1.342
TCYMBP10	1.358	1.271	1.342
TCYURT10	1.354	1.262	1.316
TETCYH01	1.342	1.283	1.339

The first column gives the structure name in the Cambridge Crystallographic Data Bank. Atom names are explained in Fig. 1.

Ionized Tc with Bound Mg^{2+}

Based on the available crystal structures, we assume Tc binds Mg^{2+} in an ionized form, Tc^-, with the O6 oxygen deprotonated. We consider the same two tautomeric forms as before, which we continue to refer to as neutral and zwitterionic. To obtain parameters for the two corresponding Tc^-:Mg^{2+} complexes, we take the neutral Tc parameters, above, as an initial guess. For Tc_N^-:Mg^{2+}, we consider nine water positions with two degrees of freedom each (distance and rocking angle). Because the manual fitting procedure was slow, an automatic fitting procedure that varies parameters in a random fashion was implemented and applied. At each step of the procedure, one atom and one associated parameter (charge or van der Waals parameter) were chosen. A combined rms deviation of the interaction energies, distances and angles was optimized by a linear search with respect to this parameter. After 15,000 such steps, the procedure converged (the rms deviation over ten subsequent steps did not change).

For Tc_Z^-:Mg^{2+}, we do not derive parameters specifically, but make use of a fragment transfer approach, using the molecular regions shown in Figure 5. For region II, we adopt the force field parameters of Tc_N^-:Mg^{2+}; for region I, we adopt those of Tc_Z. For parameters that involve both regions, such as dihedrals straddling the I/II boundary, the parameters are identical in Tc_Z and Tc_N^-:Mg^{2+}, so that there is no ambiguity.

Optimization of the Intramolecular Force Field Parameters

In a force field treatment, the intramolecular geometry is mainly determined by the minimum energy values of the bond length and bond angle terms, and by the phase and multiplicity of the dihedrals. These parameters were optimized by fitting to the structures from the conformation space search and *ab initio* optimizations, in the case of Tc_N, and from the X-ray crystal structure TETCYH01 in the CCDB in the case of Tc_Z. At each parameter optimization step, the structure was minimized with the force field model, using a Powell conjugate gradient algorithm, stopping when the rms energy gradient reached 10^{-6} kcal/mol/Å. The quality of the parameters was measured by the sum of the rms deviations from the X-ray structure and the *ab initio* structures of both the neutral and zwitterionic forms. The most difficult part was the adjustment of the dihedral force constants and phases. Starting guesses were taken from either the CHARMM[7] or the CHARMm Version 22[18] force fields for related, small molecule fragments. A few of the phases were then varied manually in an *ad hoc* way, to minimize the total rms deviation. The final changes in phases were less than 20°. Similarly, some of the dihedral force constants were increased, to force the structure into the desired geometry.

The normal modes were then computed with the force field and with the B3LYP/6-31G* Hamiltonian. The B3LYP frequen-

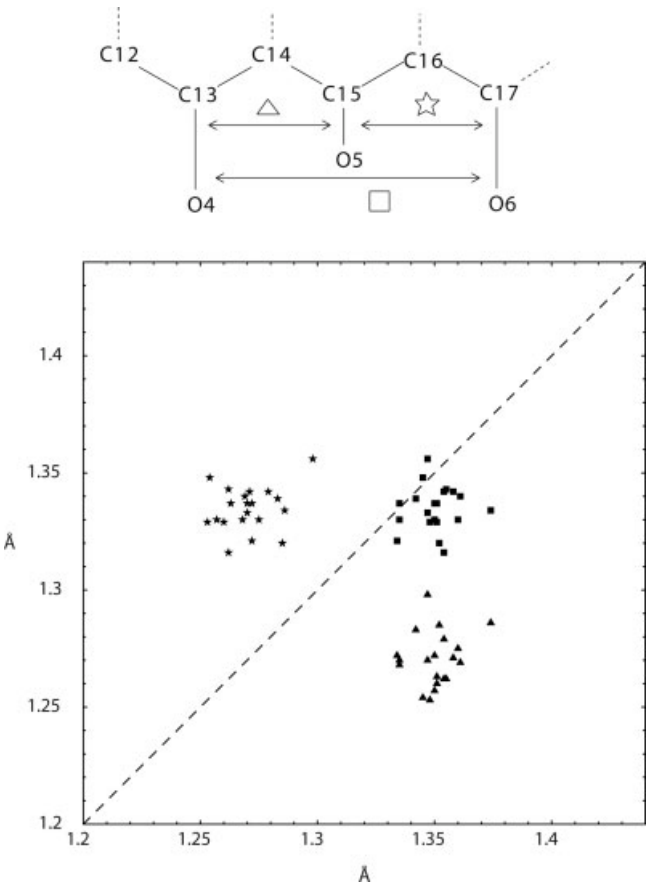


Figure 3. Selected CO bond lengths in the experimental Tc structures. Triangles: C13-O4 (X) vs. C15-O5 (Y) bond lengths for different experimental structures in the CCDB; boxes: C13-O4 (X) vs. C17-O6 (Y); stars: C15-O5 (X) vs. C17-O6 (Y). The short C15-O5 length indicates a deprotonated O5.

Table 2. Rms Deviations (Å) Between Experimental and *Ab Initio* Tc Structures.

	N_1	N_2	N_3	N_4	Z_1	Z_2	Z_3	Z_4
N_2	0.838							
N_3	0.804	**0.071**						
N_4	0.912	**0.192**	0.225					
Z_1	0.837	**0.117**	**0.107**	0.216				
Z_2	0.860	**0.076**	**0.127**	**0.179**	**0.106**			
Z_3	0.223	0.857	0.827	0.913	0.873	0.882		
Z_4	0.253	0.828	0.799	0.880	0.847	0.855	**0.063**	
AOTETC	0.267	0.839	0.807	0.900	0.861	0.870	**0.145**	**0.149**
BINNAN	0.349	0.936	0.904	0.998	0.951	0.967	0.297	0.307
DHCTET10	0.286	0.864	0.843	0.924	0.876	0.881	0.260	0.280
EPCTCY10	0.354	0.990	0.957	1.045	0.996	1.015	0.302	0.324
OXYTET	0.225	0.946	0.912	1.024	0.929	0.962	0.422	0.457
TBDMXT01	**0.178**	0.824	0.795	0.906	0.816	0.841	0.372	0.388
ATETCY10	0.906	0.305	0.308	0.287	0.231	0.273	0.952	0.927
CMTCYH10	0.923	0.239	0.270	**0.069**	0.254	0.223	0.923	0.890
$DXTETC_1$	0.863	0.245	0.263	**0.118**	0.228	0.221	0.875	0.845
$DXTETC_2$	0.919	0.257	0.281	**0.090**	0.260	0.242	0.919	0.887
DXXTEC	0.923	0.247	0.276	**0.081**	0.273	0.236	0.914	0.881
EPOXT10	0.926	**0.152**	**0.174**	**0.154**	**0.135**	**0.136**	0.946	0.917
JAVCIS10	0.907	0.280	0.309	**0.119**	0.285	0.254	0.904	0.874
$JUNLUZ_1$	0.894	0.234	0.241	**0.198**	**0.161**	**0.199**	0.929	0.903
$JUNLUZ_2$	0.910	**0.151**	**0.181**	**0.133**	**0.132**	**0.121**	0.933	0.904
OXTETD	0.921	0.303	0.322	**0.175**	0.267	0.270	0.941	0.912
TCYMBP10	0.917	0.277	0.303	**0.108**	0.297	0.265	0.907	0.875
TCYURT10	0.990	0.380	0.413	0.272	0.451	0.393	0.946	0.909
TETCYH01	0.921	0.267	0.298	**0.101**	0.293	0.252	0.911	0.878
TETCYH10	0.924	0.265	0.297	**0.096**	0.292	0.250	0.914	0.881
WIBBOY	0.896	0.286	0.310	**0.139**	0.275	0.254	0.900	0.871
XAYCAB	0.953	0.282	0.317	**0.129**	0.327	0.278	0.935	0.901

Deviations are computed for the non-hydrogen atoms of the main Tc rings, A–D. Values lower than 0.2 Å are in bold face. Experimental structures are from the Cambridge Crystallographic Data Bank (CCDB).

cies were scaled by 0.96, since this factor has been found to improve the agreement with experiment for organic molecules.[19]

Simulation of Tc Crystals

Crystal minimizations and molecular dynamics simulations were performed with the CHARMM program.[15] The starting structure was TETCYH01, from the CCDB.[20] This structure corresponds to the $P2_12_12_1$ space group; its asymmetric unit contains four zwitterionic Tc molecules and 24 water molecules. The lattice parameters were free to vary during both minimization and dynamics. Minimization was terminated when the rms energy gradient reached 10^{-6} kcal/mol/Å. Truncation of the nonbonded interactions was done using an atom-based shifting function for the Coulomb term and an atom-based switching function for the van der Waals term, with a minimum image convention and a 14 Å cutoff. Molecular dynamics were done for the same crystal, with constant, room temperature and pressure, using the temperature and pressure coupling scheme of Berendsen et al.[21] as implemented in CHARMM, in conjunction with the leapfrog integrator. A time step of 2 fs was used, with a temperature coupling constant of 5.0 ps and a pressure coupling constant of 5.0 ps. SHAKE was used to constrain the length of all covalent bonds involving hydrogens.[22] The trajectory lasted 0.5 ns.

Ca^{2+} Versus Mg^{2+} Binding to Tc: Free Energy Simulations

We used the thermodynamic cycle in Figure 2. We used the horizontal legs, which correspond to alchemical molecular dynamics free energy simulations (MDFE). The Mg^{2+} ion is reversibly transformed into Ca^{2+} during a series of MD simulations; the corresponding work is derived from a thermodynamic integration formula, eq. (2).[23] To simulate the two legs of the thermodynamic cycle, we considered two systems: the Mg^{2+}/Ca^{2+} ion in the center of a water sphere with a 24 Å radius, and the same sphere containing the ion bound to Tc. In each system, the energy function can be expressed as a linear combination of Mg^{2+} and Ca^{2+} terms:

$$U(\lambda) = U_0 + (1 - \lambda)U(\mathrm{Mg}^{2+}) + \lambda U(\mathrm{Ca}^{2+}) \quad (1)$$

where λ is a coupling parameter and U_0 represents interactions between parts of the system other than the mutated ion. The free energy derivative with respect to λ has the form:

$$\frac{\partial G}{\partial \lambda}(\lambda) = \langle U(\mathrm{Ca}^{2+}) - U(\mathrm{Mg}^{2+}) \rangle_\lambda \quad (2)$$

where the brackets indicate an average over an MD trajectory with the energy function $U(\lambda)$.[23] We gradually mutated Mg^{2+}

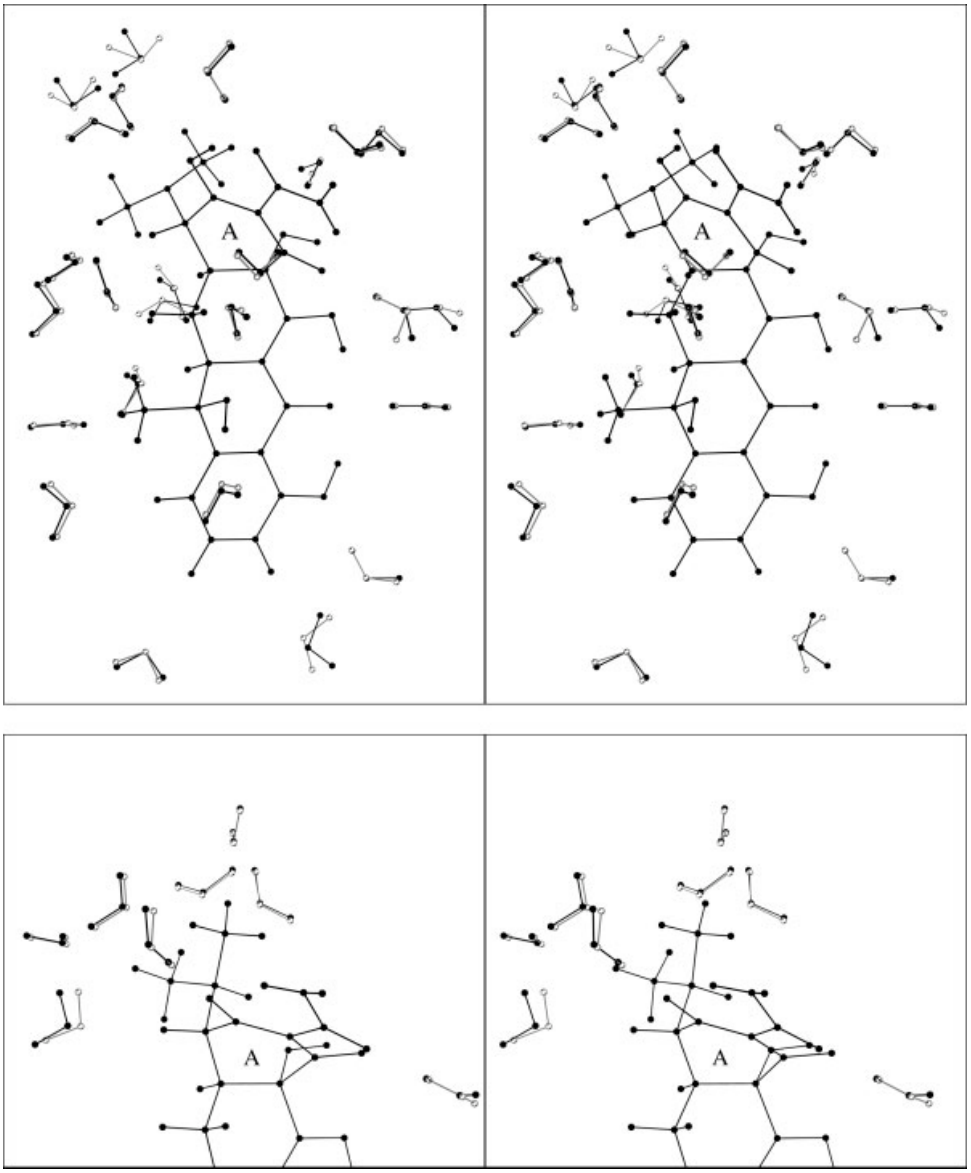


Figure 4. Probe water positions around Tc_N (above) and Tc_Z (below, closeup). *Ab initio* (white) and force field (black) positions. Stereo views.

into Ca^{2+} by changing λ from 1 to 0. The successive values of the coupling parameter were: 0.999, 0.99, 0.95, 0.9, 0.8, 0.6, 0.4, 0.2, 0.1, 0.05, 0.01, and 0.001. The free energy derivatives were computed at each λ value from a 100 ps MD simulation; the last 80 ps were used for averaging. The total mutation run thus corresponded to 1.2 ns. We performed two runs: forward (from $\lambda = 0$ to 1) and backward (from $\lambda = 1$ to 0). The ion position was harmonically restrained at the center of the water sphere with a force constant of 2 kcal/mol/Å^2. This restraint does not affect the free energy, since its contribution cancels between the two legs. We used the "Spherical Solvent Boundary Potential", or SSBP solvent model, which treats the region outside the 24 Å sphere as a uniform dielectric continuum with a dielectric constant of 80,[24] thus simulating a bulk solution. The position of the dielectric boundary between the inner sphere and the outer continuum was determined "on the fly," by the position of the outermost water molecule. A multipolar expansion with 20 terms was used to approximate the reaction field due to the surrounding continuum. Electrostatic interactions between atoms within the sphere were computed without any cutoff, using a multipole approximation for distant groups.[25]

Protein Simulations

The crystal structure of the class D Tet Repressor dimer (TetR), complexed with Tc, was taken from the Protein Data Bank

Table 3. Interactions Between a Probe Water and Selected Tc_N Sites.

	Ab Initio/Force Field Results		
Probe Site	Energy (kcal/mol)	Distance (Å)	Angle (°)
C6–H3	3.3/3.1	2.5/2.6	134/131
C22-H23	2.2/1.9	2.5/2.7	134/107
C7-H4	2.2/2.2	2.4/2.6	116/103
C12-H9	2.6/2.3	2.2/2.6	170/205
C22-H24	3.7/3.6	2.8/2.8	162/167
O3-H6	6.6/6.6	1.8/2.0	133/133
C10-H7	2.6/2.3	2.7/2.9	118/114
N2-H20	5.8/5.6	1.8/1.9	104/98
C11-H8	2.6/2.6	2.5/2.5	137/145
C20-H18	2.3/2.7	2.6/2.6	134/112
C5-H2	1.9/2.2	2.9/3.0	133/157
C20-H19	1.4/1.4	2.5/2.6	105/80
C19-H14	0.9/0.8	2.6/2.6	136/76
C4-N1	6.3/6.2	2.0/2.0	77/79
C6-H5	5.0/5.3	2.2/2.5	67/53
C1-O1	4.6/4.8	1.9/1.8	97/174
C13-O4	5.2/5.5	1.9/1.9	59/46
C18-O7	1.6/1.9	2.1/2.0	23/44
C17-O6	1.4/1.5	2.2/2.1	5/32
C21-N2A	1.6/1.4	2.4/2.2	120/123
C21-N2B	1.2/1.5	2.4/2.2	59/51
C21-O8	4.0/4.2	1.9/1.8	151/148
C15-O5	2.2/2.5	1.9/1.8	–
C1-O1	4.5/3.6	1.9/1.8	–
C4-H1	2.9/3.0	2.4/2.6	–
C19-H14	0.9/0.7	2.6/2.7	–
C3-O2	4.5/4.7	2.1/2.0	–
C13-O4	5.0/4.6	1.9/2.3	–
C17-O6	1.4/1.5	2.2/2.1	–
C19-N1	2.9/5.3	2.1/2.1	–

(PDB), entry 2TRT.[26] The simulations included protein residues within a 24 Å sphere, centered on the Mg^{2+} ion of one of the two Tc:Mg^{2+} complexes bound to the protein. Hydrogens were constructed with ideal stereochemistry. In addition to crystal waters, a 24 Å sphere of water was overlaid and waters that overlapped protein, crystal waters, Tc, or Mg^{2+} were removed. Protein atoms between 20 and 24 Å from the initial sphere's center were harmonically restrained to their experimental positions, while protein residues beyond 23.5 Å had their charges switched off. Simulations of the protein system were performed with the SSBP solvent model, which treats the region outside the 24 Å sphere as a uniform dielectric continuum (see above), with a dielectric constant of 80.[24] This is reasonable, since most of the outer region is water. Newtonian dynamics were used for the innermost region, within 20 Å of the sphere's center; Langevin dynamics were used for the outer part of the sphere, referred to as the buffer region. We used a 2 fs timestep and a bath temperature of 292 K. The CHARMM22 force field was used for the protein.[7] The TIP3P model was used for water.[9]

Results

Protonation States and Conformations of Tc

Protonation States and Conformations in the CCDB and the PDB

The protonation state of Tc can vary through binding/release of a proton in the region of O4, O5, O6 (Fig. 1), and by proton exchange between O2 (neutral Tc) and N1 (zwitterionic Tc). The experimental proton dissociation constants, or pK_a's, for Tc hydrochloride in aqueous solution were reported as 3.3, 7.7, and 9.7.[27] These values were assigned, respectively, to the regions II, I, and III in Figure 1. The pK_a values reported from an ^1H NMR study in 50/50% methanol:water were 4.4, 7.8, and 9.4, which were assigned to the C3 hydroxyl proton, the dimethylammonium proton, and the C17 hydroxyl proton, respectively.[10] For a pH between 4.4 and 7.8, the N1/O2 region is globally neutral, with a proton on either N1 (zwitterionic Tc) or O2 (neutral Tc).

In the 21 CCDB structures, the protonation state can be inferred from the OH bond lengths, summarized in Table 1 and Figure 3. We find that O5 is always deprotonated, while O4 and O6 are always protonated. None of these structures include a metal cation bound in this region. We assume that when Mg^{2+} binds between O5 and O6, as seen in the PDB structures, O6 will become deprotonated, and Tc will be negatively charged.

For the O2/N1 region, we find a mixture of the neutral and zwitterionic tautomers in the CCDB. From the observed bond lengths, we infer that 13 structures are zwitterionic, and all of these are in the extended conformation. Six structures are neutral, and all of them are in the twisted conformation. The remaining two structures lack N1.

In the experimental PDB structures, the limited experimental resolution does not allow the protonation state to be inferred. Only the extended Tc conformation is found. Although this correlates with the zwitterionic form in the CCDB structures, we cannot assume the same correlation holds in protein complexes. Indeed, protein sidechains could stabilize the neutral form, even when Tc is extended. The question of which Tc tautomer binds to TetR remains open.

Simulated Annealing Conformational Search and Ab Initio Optimizations

For both the neutral and zwitterionic forms of Tc, simulated annealing with the MM3 force field was followed by conformational clustering. Using a 10° threshold distance (in dihedral angle space) between cluster centers, we obtained about 50 clusters each for Tc_N and Tc_Z. Four low energy Tc_N structures were optimized quantum mechanically, at the B3LYP/6-31G(d) level; i.e., the lowest twisted and the three lowest extended structures. For Tc_Z, two extended and two twisted structures were optimized quantum mechanically.

The quantum mechanical structures agree closely with the CCDB structures, even though the CCDB structures have slightly different ring substituents and different crystal environments. The rms deviations between experimental and quantum mechanical structures are given in Table 2. The rms deviation between each quantum mechanical structure and the nearest

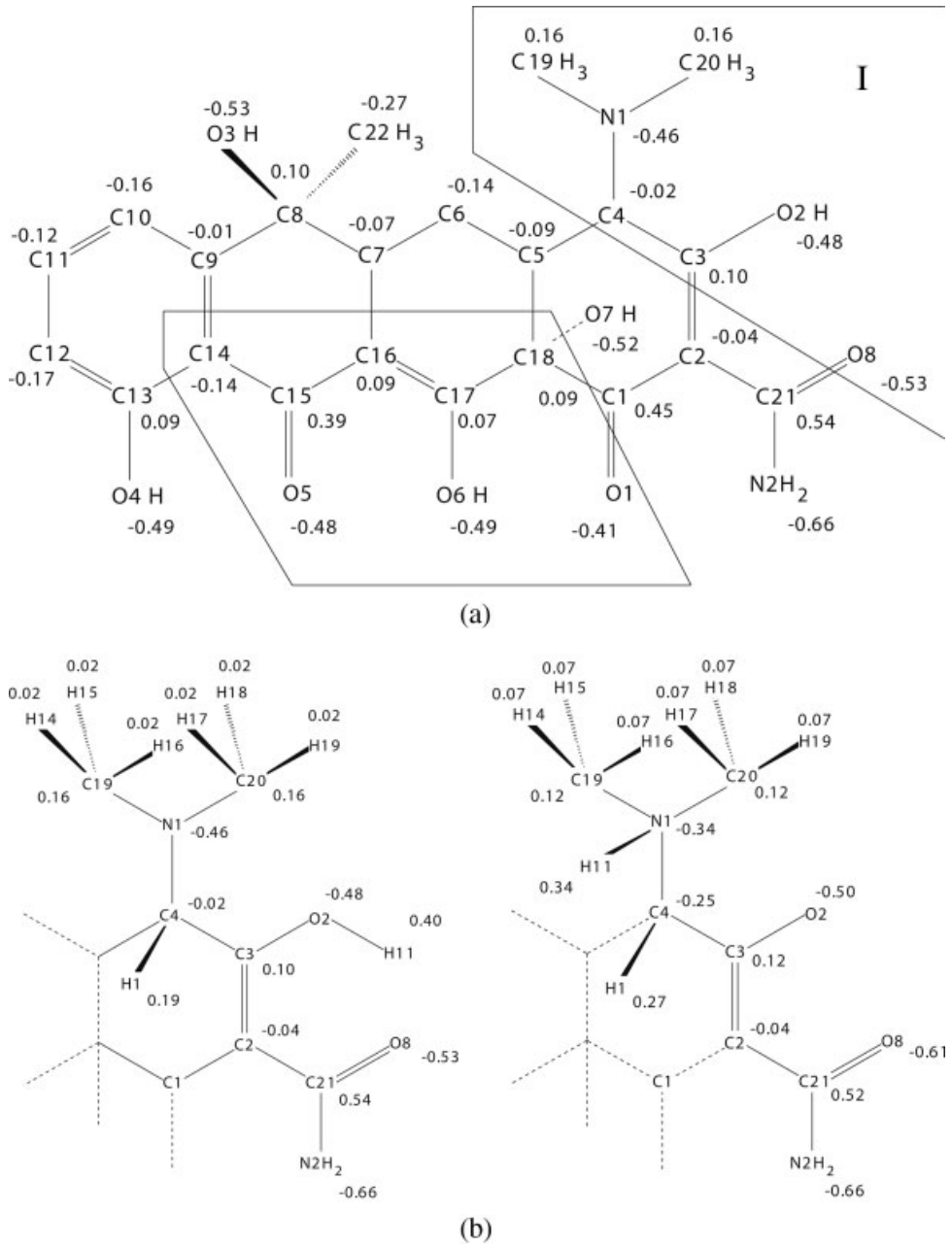


Figure 5. (A) Atomic charges in Tc_N. (B) Ring IV in Tc_N (left) and Tc_Z (right).

CCDB structure varies between 0.08 Å (neutral, extended) and 0.18 Å (neutral, twisted). This good agreement shows that the quantum mechanical structures are a valid starting point for molecular mechanics force field parameterization.

Force Field Determination

We limit ourselves to the extended Tc conformation, which is the only one observed in protein complexes. Since the CCDB extended conformations are all in the zwitterionic tautomeric form, we might have limited ourselves to this tautomer. However, the mixture seen in the CCDB suggests that the energy difference between the neutral and zwitterionic forms is not very large, and we must allow for the possibility that the protein environment could shift the equilibrium in favor of an extended neutral form. In the complexes with TetR, O2, and N1 are both involved in hydrogen bonds with protein residues. Also, the presence of a metal cation could change the balance between the neutral and zwitterionic forms. In all, we consider four Tc states:

Table 4. Interactions Between a Probe Water and Selected Tc_Z Sites.

	Ab Initio/Force Field Results		
Probe Site	Energy (kcal/mol)	Energy (kcal/mol)	Angle (°)
C4-H1	4.7/4.5	2.2/2.5	154/145
N1-H11	7.3/7.0	3.0/3.0	166/163
C19-H14	5.7/5.5	2.2/2.3	136/137
C19-H15	6.2/5.9	2.1/2.2	124/121
C20-H18	2.0/2.2	2.3/2.4	131/131
C20-H19	5.5/5.5	2.2/2.3	50/149
C3-O2	5.3/5.5	1.8/1.9	55/65
C21-O8	9.2/9.4	1.8/1.7	−69/−56

neutral and zwitterionic, with and without bound Mg^{2+}. When Mg^{2+} is bound, O6 is deprotonated. The four states are denoted: Tc_N, Tc_Z, $Tc_{\bar{N}}$:Mg^{2+}, and $Tc_{\bar{Z}}$:Mg^{2+}.

Lennard-Jones Parameters and Atomic Partial Charges: Tc_N and Tc_Z

For the neutral Tc form, the water positions employed are shown in Figure 4. The *ab initio* optimal positions are compared to the positions obtained at the end of the force field fitting. Table 3 summarizes the agreement for the interaction energies and geometries. The *ab initio* data are "corrected" by energy scaling and distance shifting, as described in Methods. The rms deviation between the *ab initio* and force field data for interactions with two degrees of freedom are 0.22 kcal/mol for the energies, 0.16 Å for the Tc-water distances, and 26° for the rocking angle. The largest errors are 0.39 kcal/mol, 0.32 Å, and 77°. The corresponding force field parameters are listed in the Appendix (Table A1). The charge distribution is represented in Figure 5A. The initial guess was based on the CHARMM22 charges, in cases where a very similar group existed, or the HF/6-31G(d) Mulliken charges, in all other cases. The largest change between the initial and final charges was 0.2*e*; the average unsigned change was 0.12*e*.

In several cases, we optimized the supermolecular interaction using just one degree of freedom (distance) for the water molecule. Agreement with the quantum results was rather poor (Table 3), whereas the same interaction gave good agreement when two degrees of freedom were optimized (distance and rocking angle; see Fig. 1B). In one case, for example, we obtained a molecular mechanics/*ab initio* difference of just 0.2 kcal/mol by changing the rocking angle by only 2°, whereas the nonoptimized angle gave an energy difference of 2.4 kcal/mol. Only the results optimized with two degrees of freedom were used for parameter fitting.

Several chemical groups in Tc can be compared to groups already present in the CHARMM22 force field. Thus, Tc contains three methyl groups. CHARMM charges for methyl groups in uncharged model compounds vary from −0.30*e* to +0.16*e* for the carbon and from 0.05*e* to 0.09*e* for the hydrogens. In the dimethylamine group in neutral Tc, the charges are 0.16*e* for the carbons and 0.02*e* for the hydrogens. For the methyl attached to C8, the charges are −0.27*e* for the carbon and 0.09*e* for the hydrogens, close to a typical CHARMM methyl. For hydroxyl groups, CHARMM22 charges for the oxygens are −0.54*e* in phenol and −0.66*e* in methanol, ethanol and 2-propanol. In Tc_N, we have 5 hydroxyls, whose oxygen charges vary from −0.475*e* to −0.530*e*, slightly less negative than in the CHARMM22 force field. The CHARMM22 hydroxyl hydrogens have charges of 0.43*e* in methanol, ethanol, phenol, and 2-propanol. In Tc_N, the hydroxyl hydrogen charges vary from 0.37*e* to 0.43*e*. Interestingly, the hydroxyls in the CHARMM force field all have the same van der Waals types. In our case, we changed the van der Waals types of some hydroxyl hydrogens from the CHARMM22 values of $\varepsilon = -0.046$ kcal/mol and $R/2 = 0.224$ Å to −0.050 kcal/mol and 0.76 Å (H6 and H13). The hydroxyl oxygen van der Waals type was left the same as in CHARMM. Another small molecule of interest is acetamide. In the CHARMM glutamine sidechain, the hydrogens *cis* and *trans* to the oxygen have different charges: 0.32*e* and 0.30*e*, respectively, while in the CHARMM formamide, they have the same charge: 0.35*e*. In the Tc acetamide group, we found that the best fit is obtained when the *cis* hydrogen charge is 0.327*e* and the *trans* hydrogen charge is 0.317*e*. In general, the charges of the Tc acetamide group are in good agreement with the CHARMM22 glutamine.

Finally, we consider the relation between phenol and the D ring of Tc_N. The charges of C9 and C14 in Tc are equal to −0.07*e* and −0.14*e*, respectively, compared to zero for the corresponding atoms in CHARMM22 phenol. The H10 charge is 0.37*e*, less than the CHARMM phenol hydrogen charge of 0.43*e*. In the CHARMM22 phenol, all the carbons (except the one bonded to oxygen) have the same charge, and each carbon/hydrogen pair is neutral. In our study, we did not aim to create neutral groups within the Tc molecule, and the D ring carbon charges are all slightly different, ranging from −0.12*e* for C11 (close to the CHARMM value of −0.115*e*) to −0.165*e* for C12.

For the zwitterionic Tc form, the water positions are shown in Figure 4. Table 4 summarizes the agreement for the interaction energies and geometries. The rms deviation between the *ab initio* and force field data are 0.23 kcal/mol for the energies, 0.16 Å for the Tc-water distances, and 7° for the rocking angle. The largest errors are 0.31 kcal/mol, 0.34 Å, and 13°. The corresponding force field parameters are listed in the Appendix (Table A2). The charge distribution is represented in Figure 5B. We see that the transfer of a proton from O2 to N1 has made the N1 methyl groups more polar. The largest change between the Tc_N charges and the final Tc_Z charges is 0.23*e*.

Intramolecular Parameters: Tc_N and Tc_Z

The intramolecular parameters are reported in Supplementary Material, for both Tc_N and Tc_Z. The minimum energy bond lengths, bond angles, and torsion angles were taken directly from the

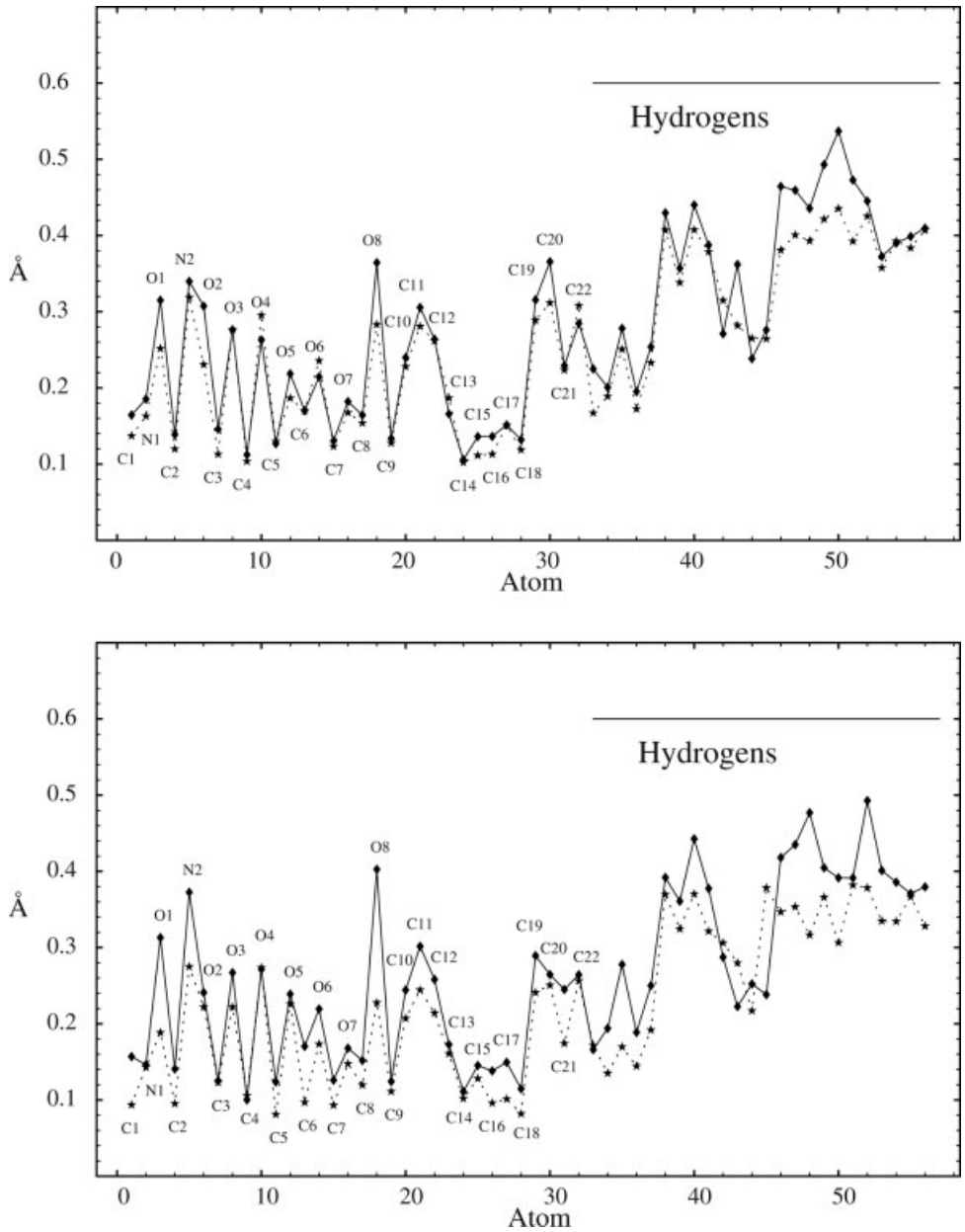


Figure 6. Atomic rms fluctuations (Å) for Tc_N (above) and Tc_Z (below) from *ab initio* (solid line) and force field (dashed line) models.

ab initio structures, optimized using the B3LYP/6-31G(d), hybrid, Hartree–Fock/density functional theory. To derive the corresponding force constants, the normal modes of vibration were computed at the B3LYP/6-31G* level. Normal modes were also computed with the force field model. Vibrational frequencies, rms atomic fluctuations, and interatom pairwise correlations were derived. The interatom pairwise correlation, for atoms i, j, is defined by $\langle u_i \cdot u_j \rangle$, where u_i represents the instantaneous displacement of atom i from its mean position and the brackets represent a thermal average (an average over the normal mode spectrum, in this case). Comparisons between the quantum mechanical and force field results are given in Figures 6 and 7. Results are shown for both Tc_N and Tc_Z. Agreement is generally quite good, especially for the atoms of the main, four-ring Tc scaffold. The long-range atomic pairwise correlations are also well reproduced, as are the lowest vibrational frequencies. This shows that the overall plasticity of Tc is reproduced, which is especially important for future studies of protein binding.

The rms deviation between the *ab initio* minimized structure and the force field minimized structure, after superposition of

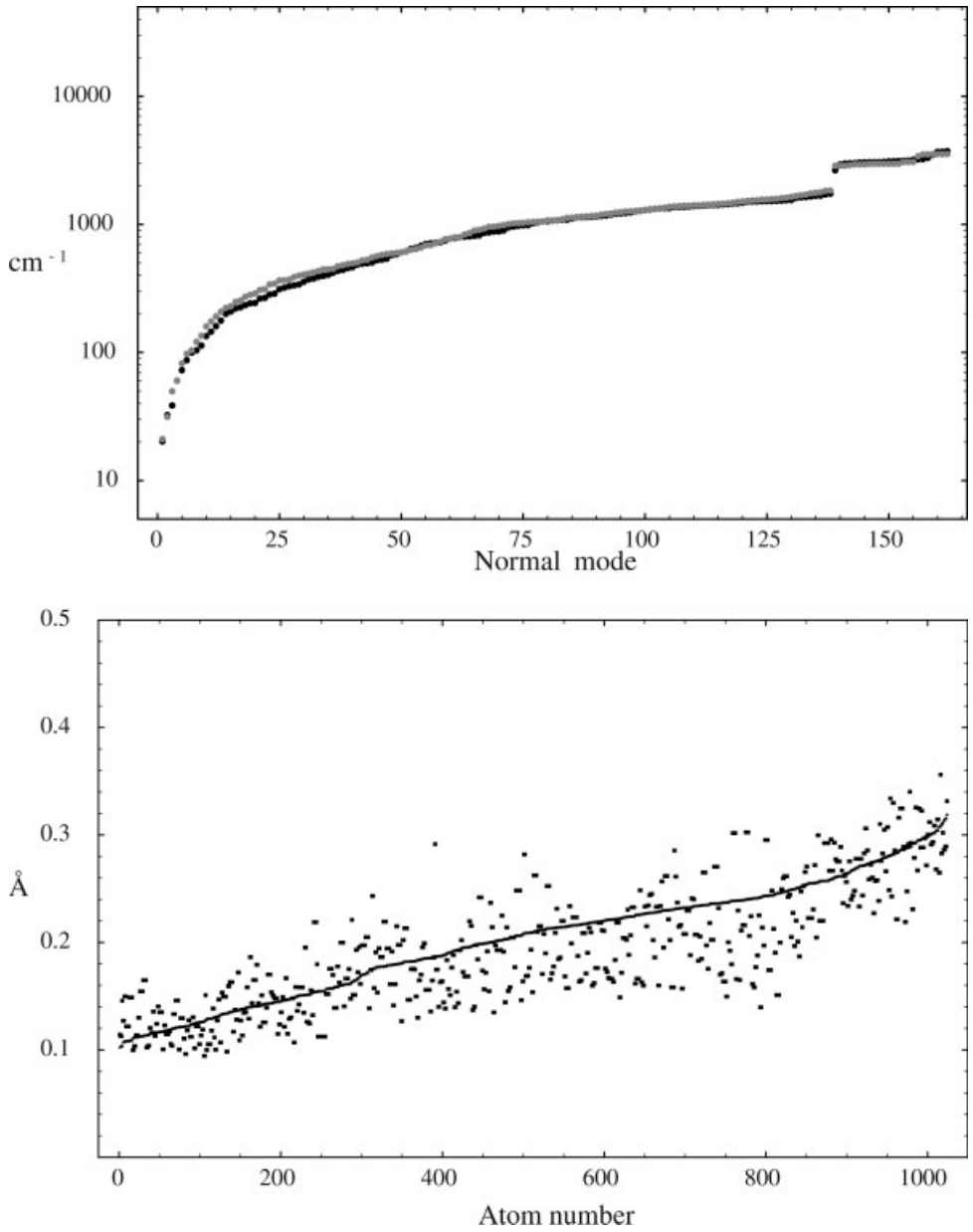


Figure 7. Upper panel: *Ab initio* (grey) and force field (black) vibrational frequencies (cm^{-1}). Lower panel: *Ab initio* (solid line) and force field (dots) interatom correlations; arbitrary atom numbers.

the whole molecule, is 0.20/0.23 Å for the neutral/zwitterionic forms of Tc, respectively. The rms deviation for individual groups ranges from 0.13/0.14 Å for the D ring system (including hydrogens) to 0.40/0.23 Å for the acetamide groups of Tc_N/Tc_Z.

To test the intramolecular energies further, we also computed the energy variations when the Tc_N molecule is distorted along the 10 lowest-frequency, *ab initio* normal modes. 150 structures were produced (15 per normal mode); the corresponding *ab initio* energies were computed from the *ab initio* frequencies. The B3LYP/6-31G* modes and frequencies were used (the frequencies being scaled by 0.96; see Methods). The molecular mechanics energies were computed for the same structures. The two sets are in good agreement, as shown in Figure 8.

Tc^- with Bound Mg^{2+}

The $Tc_N^-:Mg^{2+}$ complex was modeled with four water molecules coordinating Mg^{2+}, in addition to the water that is used to probe Tc in the supermolecule calculations. The Tc oxygens O5 and O6 also coordinate the cation, giving a standard, octahedral coordination. Nine probe positions were considered, shown in Figure 9. The *ab initio* and force field energies and geometries

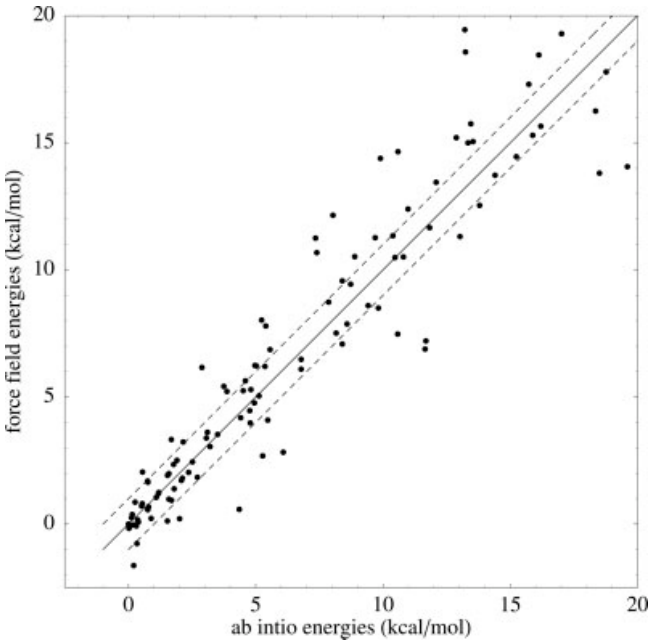


Figure 8. *Ab initio* and force field energies (kcal/mol) for 150 Tc_N structures, corresponding to displacements along the 10 lowest *ab initio* (B3LYP/6-31G*) normal modes. The dashed lines are 1 kcal/mol above and below the diagonal.

are given in Table 5. The atomic charges are shown in Figure 10. The rms deviations from the *ab initio* data are 0.65 kcal/mol for the energies, 0.20 Å for the Tc-water distances, and 73° for the angles. For two sites, the energy error is as large as 1 kcal/mol. Thus, the interactions of a water with $Tc^-:Mg^{2+}$ cannot be captured quite as accurately as the interactions with Tc alone.

The force field parameters are listed in the Appendix (Table A3). Deprotonating O6 and introducing Mg^{2+} changes the Tc charge distribution locally. O5, O6, and C16 are more negatively charged, compared to Tc_N. The C16 charge has decreased from $0.09e$ in Tc_N to $-0.50e$ in $Tc_N^-:Mg^{2+}$. C15 and C17 are more positively charged in $Tc_N^-:Mg^{2+}$; the C17 charge has increased by $0.46e$. The charges of C1, O4, C12 and C13 are unchanged.

$Tc_Z^-:Mg^{2+}$ was not modeled explicitly. Since the Mg^{2+} site is distant from the O2-N1 region, we assume the $Tc_Z^-:Mg^{2+}$ complex can be modeled by combining the Tc_Z^- charges for region I and the $Tc_N^-:Mg^{2+}$ charges for region II (the regions are defined in Fig. 5).

Testing the Force Field

Tc Crystal Simulations

The CCDB crystal structures correspond to several Tc variants. The most similar to the Tc of interest here (shown in Fig. 1), is TETCYH01.[20] This is an extended Tc in its zwitterionic form. We therefore considered the Tc_Z molecule in the TETCYH01 crystal lattice. The unit cell contains four Tc molecules and 24 waters, allowing us to directly test the balance between Tc-water and water-water interactions. There are no ions in the structure. We performed both energy minimizations and MD simulations. The only symmetry constraint was the orthogonal symmetry of the $P2_12_12_1$ crystal space group. The lattice parameters and the internal organization of the unit cell were otherwise free to vary.

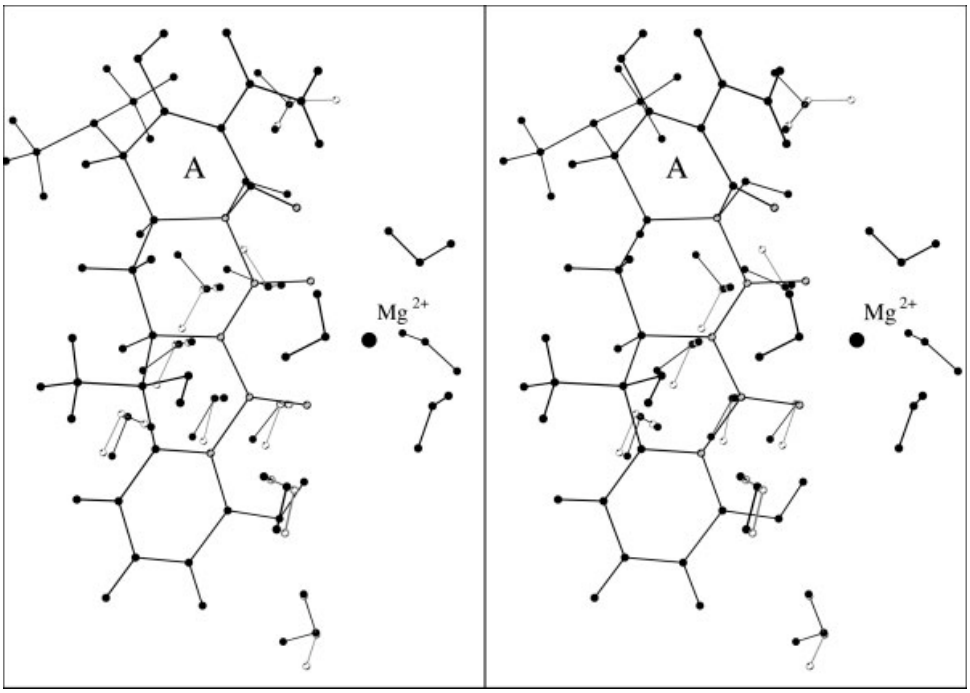


Figure 9. Probe water positions around $Tc:Mg^{2+}$ (stereo view). *Ab initio* (white) and force field (black) positions. Tc atoms whose charges change upon Mg^{2+} binding are light-colored. Mg^{2+} is coordinated by Tc O5 and O6 and by four waters.

Table 5. Interactions Between a Probe Water and Selected Tc_Z^-:Mg^{2+} Sites.

	Ab Initio/Force Field Results		
Probe Site	Energy (kcal/mol)	Distance (Å)	Angle (°)
C15	1.4/1.5	2.7/2.6	156/173
C16	1.8/2.2	2.4/2.2	229/259
C14(B)	2.4/2.2	2.7/2.5	−142/−146
C14(F)	7.0/7.5	2.5/2.2	138/137
O5	3.6/2.8	2.1/2.2	101/118
O6	3.8/2.5	2.2/2.1	−117/−153
C17	1.2/1.5	2.7/2.4	173/−71
O4	3.4/4.4	1.9/1.9	140/−118
O7	2.2/2.2	2.1/2.0	64/−81

The C14 sites are in the front (F) and back (B) of the stereo Figure 5, and above and below the plane of Fig. 1, respectively.

To compare the force field and the experimental structures, we superimposed either an individual Tc molecule or the entire unit cell on the experimental structure. By superimposing a single Tc molecule, we measure the intramolecular deformations, which are sensitive to the intramolecular Tc force field parameters. By superimposing the whole unit cell, we measure intermolecular rearrangements, which are sensitive to the Lennard-Jones parameters and the atomic charges. The rms deviations between the force field data and the crystal data are given in Table 6. The intramolecular rms deviations are between 0.17 Å and 0.27 Å, demonstrating the quality of the geometrical force field parameters. The rms deviation for the four Tc's in the unit cell is 0.39 Å after energy minimization and 0.76 Å after 500 ps of molecular dynamics (rms averaged over the last 50 ps of dynamics). Including the 24 waters and the four Tc's, the rms deviation after MD is 2.7 Å. The MD lattice parameters are 12.00, 22.36, and 9.86 Å, in good agreement with those in TETCYH01: 12.08, 21.58, 9.56 Å.

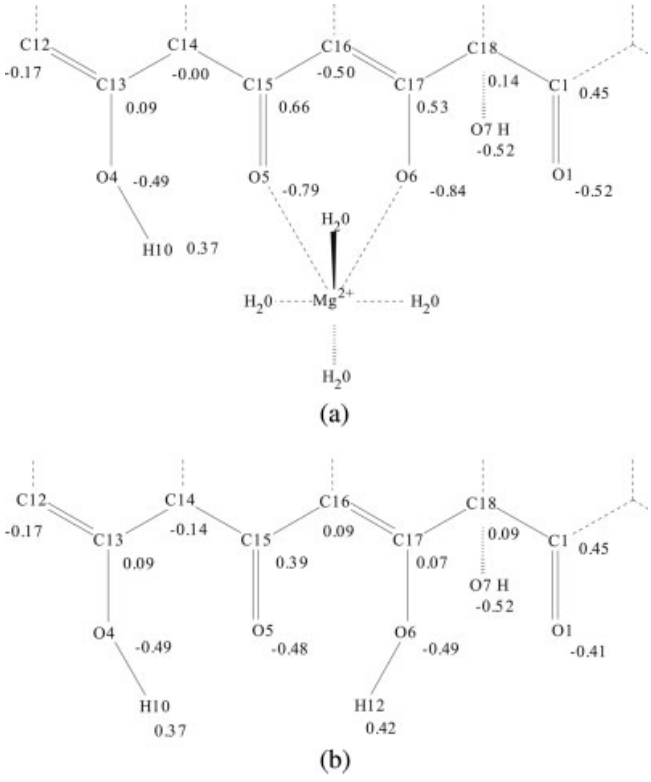


Figure 10. (A) Atomic charges of the Tc^-:Mg^{2+} complex. O6 is deprotonated. The first hydration sphere of the cation is explicitly represented. (B) Atomic charges of Tc_N, for comparison.

Table 6. Rms Deviations (Å) of Tc Structure in Different Environments.

Environment	Molecules[a]	Minimized	MD
Vacuum	Tc_N	0.19	–
	Tc_Z	0.17	–
Crystal	Tc_Z	0.25	0.27
	Four Tc_Z in unit cell	0.39	0.76
	Tc and water in unit cell	2.87	2.71
TetR:Tc_N/Tc_Z:Mg^{2+}	Tc_N/Tc_Z[a]	0.24/0.26	0.27/0.28
	Tc_N/Tc_Z[b]	–	0.53/0.68
	Protein backbone	–	0.68/0.81
	Protein	–	0.96/1.03

Deviations between the structure from the force field calculation and a reference structure: the ab initio structure for the vacuum case; the TETCYH01 CCDB structure for the crystal case, and the PDB structure 2TRT for the TetR complex. For the TetR complex, the data correspond to two of the five variants studied, which include either Tc_N or Tc_Z, and to atoms in the inner 10 Å of the spherical simulation model. The results are typical of all five model variants.

[a]After superimposing the Tc molecule on the crystal structure.

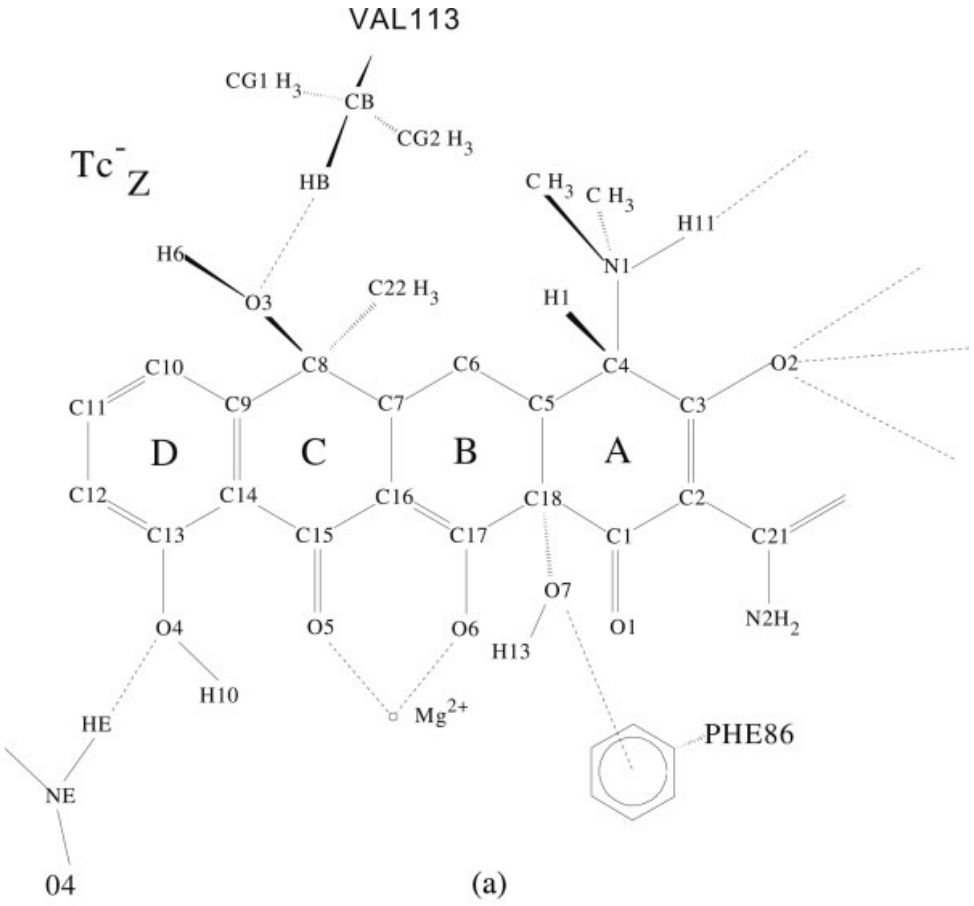
[b]After superimposing the protein backbone on the crystal structure.

Table 7. Free Energy Simulations Comparing Tc:Mg^{2+} and Tc:Ca^{2+} Binding.

Medium	Direction of Alchemical Mutation Run	ΔG	Average	ΔΔG
Solution:Mg^{2+}	Forward	51.00	51.42	
	Backward	49.83		
Solution:Tc:Mg^{2+}	Forward	50.41	50.76	
	Backward	51.10		0.34
Experiment[a]				0.16

Energies are in kcal/mol. ΔG corresponds to the alchemical mutation of Mg^{2+} into Ca^{2+} "Forward" mutation runs transform Mg^{2+} into Ca^{2+}; "backward" runs transform Ca^{2+} into Mg^{2+}. A positive ΔΔG corresponds to preferential Mg^{2+} binding to Tc.

[a]Winfried Hinrichs, personal communication.



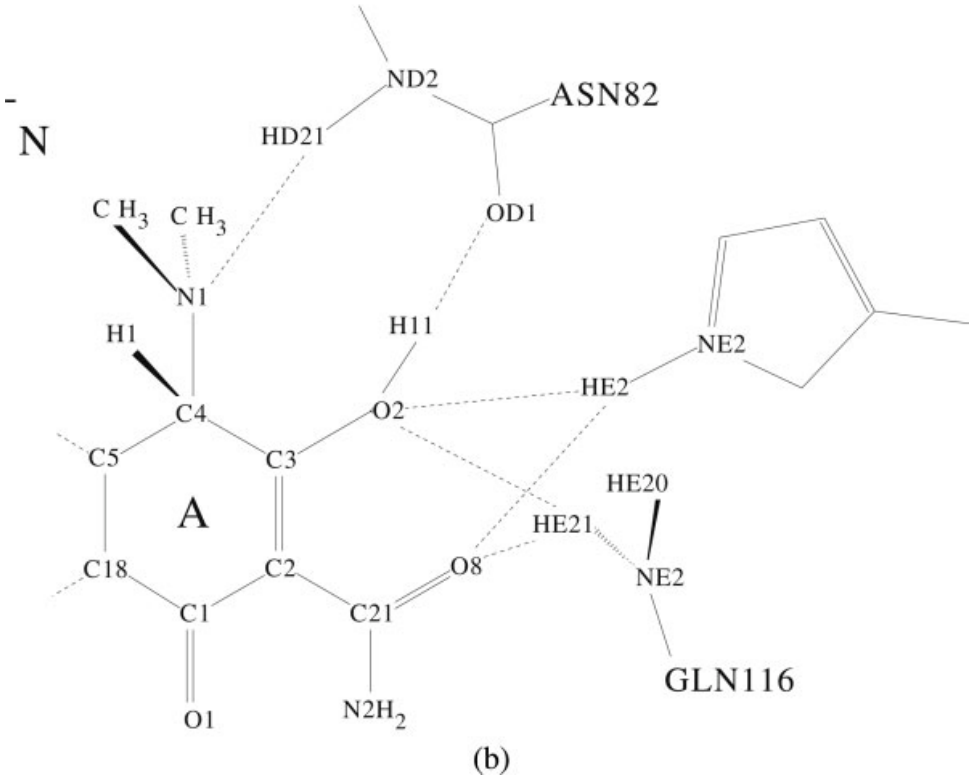


Figure 11. (A) Tc_Z:TetR interactions in the best Tc_Z MD model. (B) Tc_N:TetR interactions in the best Tc_N MD model.

Ca^{2+} versus Mg^{2+} Binding: Free Energy Simulations

To compute the relative binding affinities of Ca^{2+} and Mg^{2+} for Tc, we gradually mutated Mg^{2+} into Ca^{2+}, both in solution and in complex with Tc (horizontal legs in Fig. 2). We used the CHARMM22 force field parameters for Ca^{2+} and Mg^{2+}, the TIP3P water model, and the Tc parameters derived here. The CHARMM22 parameters reproduce the experimental solvation free energies of both Ca^{2+} and Mg^{2+}.[28,29]

The experimental binding free energy difference between Ca^{2+} and Mg^{2+} is only 0.2 kcal/mol, favoring Mg^{2+}. This value results from a competition between several effects. Since Ca^{2+} is larger than Mg^{2+}, it interacts more weakly with both Tc and water. In addition, it polarizes Tc to a lesser extent than Mg^{2+} does. A force field with fixed atomic charges, such as the one developed here, cannot reproduce the polarization effects without specific parameter adjustment. Thus, we used a trial-and-error process to modify the Lennard-Jones interactions between Tc and the two cations, and obtained the free energies reported in Table 7. The final, computed, binding free energy difference is 0.3 kcal/mol. The only parameter change that was needed to obtain this fit was a decrease of the van der Waals radius of the oxygens that bind to Mg^{2+} from 1.750 Å to 1.665 Å. The small magnitude of the parameter adjustment that is needed provides evidence that the Tc force field is reasonable.

Tc:Mg^{2+}:TetR Interactions: Molecular Dynamics Simulations

We also performed simulations of the Tc:TetR complex, using five different model variants, for 5 ns each. We considered both Tc_N and Tc_Z, along with two protonation states for His64 and two starting orientations for Asn82. These two amino acids interact directly with the Tc ligand, as schematized in Figure 11. The protein was solvated by a 24 Å water sphere centered on the Tc ligand, and more distant regions were modeled as a dielectric continuum (see Methods).

We computed rms deviations from the crystal structure, after energy minimization and after molecular dynamics, averaging over the last 50 ps of 5 ns simulations (Table 6). We consider here the best model variant with Tc_N and the best variant with Tc_Z. By superimposing the Tc molecule on the crystal structure, we obtain intramolecular Tc deformations of 0.27–0.28 Å, similar to the Tc crystal case above. Superimposing the protein backbone on the crystal structure, we obtain Tc rms deviations of 0.53/0.68 Å for Tc_N/Tc_Z, which correspond partly to an overall shift of the Tc molecule. The protein backbone rms deviations are 0.68/0.81 Å for Tc_N/Tc_Z. The interactions between the Tc molecule and the protein are well reproduced, indicating that the Tc shift is accompanied by a corresponding shift of the protein environment. Indeed, the main Tc-protein interactions are schematized in Figure 11; they involve His64, Asn82, Phe86, Arg104, Val113, and Gln116. In the best model variant, eight out of nine contacts with these groups are reproduced to within less than 0.5 Å (usually less than 0.3 Å), after 5 ns of MD. Thus, the MD model is similar to other good quality, current, protein simulations.

Conclusions

Despite its importance, Tc has only recently been the object of theoretical studies.[5,6] In this work, we developed a molecular mechanics force field for Tc that is compatible with the CHARMM22 force field for proteins and nucleic acids. Our method for developing the force field is consistent with the CHARMM22 methodology, but presents also some specific aspects. First, we did not simplify the parameterization by breaking Tc down into smaller fragments, which would then be parameterized separately. Rather, the whole Tc molecule was treated together. This was due to the fused ring structure of Tc, which makes it hard to split it into smaller pieces. Second, we considered a probe water molecule with two degrees of freedom, instead of one in some of the CHARMM22 development.[7,8] This led to a good balance between the quality of interaction energies and geometries. Third, we proceeded in two steps, parameterizing Tc alone and then introducing the Mg^{2+} cation. To obtain a reasonable model using fixed atomic charges, we explicitly included the first hydration sphere of the cation. Our model makes no attempt to describe hydration/dehydration of the Mg^{2+} cation.

To check the validity of the model, we performed molecular dynamics simulations in several environments, including simulations of Tc complexed with the TetR protein. Because protein crystal structures do not ordinarily reveal the positions of hydrogen atoms, there are uncertainties about the Tc tautomeric form and the orientation of nearby histidine and asparagine sidechains that make direct interactions with Tc. Therefore, we considered five different variants of the system. All five simulations led to reasonable agreement with experiment for the structure of the Tc:TetR complex. Overall, the force field model shows good agreement with a wide range of experimental and *ab initio* data, showing that it is suitable for the investigation of Tc:protein and Tc:RNA complexes. It also provides a starting point to parameterize other compounds from the large and important Tc family.

Acknowledgments

An Egide Ph.D. fellowship was awarded to Alexey Aleksandrov. We thank Georgios Archontis and Winfried Hinrichs for helpful discussions and Martin Karplus for the CHARMM program. We thank Tim Clark for communicating unpublished data.

Appendix: Force Field Parameters

The CHARMM22 magnesium parameters are: $Q = +2$; $\varepsilon = -0.0150$ kcal/mol; $R/2 = 1.18500$ Å. The other force field parameters are listed in Tables A1–A3 and in Supplementary Material.

Table A1. Charges Q and van der Waals Parameters ε, R for Tc_N.

N	Atom	Q	ε	$R/2$
1	C1	0.450	−0.070	2.000
2	N1	−0.462	−0.200	1.850
3	O1	−0.412	−0.120	1.700
4	C2	−0.040	−0.040	1.960
5	N2	−0.660	−0.200	1.850
6	O2	−0.475	−0.152	1.770
7	C3	0.100	−0.040	1.960
8	O3	−0.530	−0.152	1.770
9	C4	−0.020	−0.080	2.060
10	O4	−0.490	−0.152	1.770
11	C5	−0.090	−0.080	2.060
12	O5	−0.480	−0.120	1.700
13	C6	−0.140	−0.080	2.060
14	O6	−0.485	−0.152	1.770
15	C7	−0.065	−0.080	2.060
16	O7	−0.520	−0.152	1.770
17	C8	0.100	−0.080	2.060
18	O8	−0.530	−0.120	1.700
19	C9	−0.007	−0.070	1.993
20	C10	−0.155	−0.070	1.993
21	C11	−0.120	−0.070	1.993
22	C12	−0.165	−0.070	1.993
23	C13	0.090	−0.070	1.993
24	C14	−0.140	−0.070	1.993
25	C15	0.390	−0.070	2.000
26	C16	0.090	−0.040	1.960
27	C17	0.065	−0.040	1.960
28	C18	0.090	−0.080	2.060
29	C19	0.159	−0.080	2.060
30	C20	0.159	−0.080	2.060
31	C21	0.540	−0.070	2.000
32	C22	−0.270	−0.080	2.060
33	H1	0.192	−0.042	1.330
34	H2	0.090	−0.042	1.330
35	H3	0.075	−0.042	1.330
36	H4	0.065	−0.042	1.330
37	H5	0.065	−0.042	1.330
38	H6	0.430	−0.050	0.760
39	H7	0.125	−0.030	1.278
40	H8	0.195	−0.030	1.278
41	H9	0.130	−0.030	1.278
42	H10	0.370	−0.046	0.225
43	H11	0.395	−0.046	0.225
44	H12	0.420	−0.046	0.225
45	H13	0.430	−0.050	0.760
46	H14	0.020	−0.039	1.210
47	H15	0.020	−0.039	1.210
48	H16	0.020	−0.039	1.210
49	H17	0.020	−0.039	1.210
50	H18	0.020	−0.039	1.210
51	H19	0.020	−0.039	1.210
52	H20	0.320	−0.046	0.225
53	H21	0.330	−0.046	0.225
54	H22	0.090	−0.042	1.330
55	H23	0.090	−0.042	1.330
56	H24	0.090	−0.042	1.330

Units of e, kcal/mol and Å.

Table A2. Charges Q and van der Waals Parameters ε, R for Tc_Z.

N	Atom	Q	ε	$R/2$
1	C1	0.450	−0.070	2.000
2	N1	−0.338	−0.200	1.850
3	O1	−0.412	−0.120	1.700
4	C2	−0.040	−0.040	1.960
5	N2	−0.660	−0.200	1.850
6	O2	−0.500	−0.179	1.850
7	C3	0.119	−0.040	1.960
8	O3	−0.530	−0.152	1.770
9	C4	−0.247	−0.080	2.060
10	O4	−0.490	−0.152	1.770
11	C5	−0.090	−0.080	2.060
12	O5	−0.480	−0.120	1.700
13	C6	−0.140	−0.080	2.060
14	O6	−0.485	−0.152	1.770
15	C7	−0.065	−0.080	2.060
16	O7	−0.520	−0.152	1.770
17	C8	0.100	−0.080	2.060
18	O8	−0.610	−0.120	1.700
19	C9	−0.007	−0.070	1.993
20	C10	−0.155	−0.070	1.993
21	C11	−0.120	−0.070	1.993
22	C12	−0.165	−0.070	1.993
23	C13	0.090	−0.070	1.993
24	C14	−0.140	−0.070	1.993
25	C15	0.390	−0.070	2.000
26	C16	0.090	−0.040	1.960
27	C17	0.065	−0.040	1.960
28	C18	0.090	−0.080	2.060
29	C19	0.118	−0.080	2.060
30	C20	0.118	−0.080	2.060
31	C21	0.520	−0.070	2.000
32	C22	−0.270	−0.080	2.060
33	H1	0.272	−0.042	1.330
34	H2	0.090	−0.042	1.330
35	H3	0.075	−0.042	1.330
36	H4	0.065	−0.042	1.330
37	H5	0.065	−0.042	1.330
38	H6	0.430	−0.050	0.760
39	H7	0.125	−0.030	1.278
40	H8	0.195	−0.030	1.278
41	H9	0.130	−0.030	1.278
42	H10	0.370	−0.046	0.225
43	H11	0.334	−0.046	0.225
44	H12	0.420	−0.046	0.225
45	H13	0.430	−0.050	0.760
46	H14	0.066	−0.023	0.990
47	H15	0.066	−0.023	0.990
48	H16	0.066	−0.023	0.990
49	H17	0.066	−0.023	0.990
50	H18	0.066	−0.023	0.990
51	H19	0.066	−0.023	0.990
52	H20	0.320	−0.046	0.225
53	H21	0.330	−0.046	0.225
54	H22	0.090	−0.042	1.330
55	H23	0.090	−0.042	1.330
56	H24	0.090	−0.042	1.330

Units of e, kcal/mol and Å.

Table A3. Charges Q and van der Waals Parameters ε, R for Tc^{-}:Mg^{2+}.

N	Atom	Q	ε	$R/2$
1	C1	0.447	−0.070	2.000
2	N1	−0.465	−0.200	1.850
3	O1	−0.515	−0.120	1.700
4	C2	−0.043	−0.040	1.960
5	N2	−0.663	−0.200	1.850
6	O2	−0.478	−0.152	1.770
7	C3	0.097	−0.040	1.960
8	O3	−0.533	−0.152	1.770
9	C4	−0.023	−0.080	2.060
10	O4	−0.493	−0.152	1.770
11	C5	−0.093	−0.080	2.060
12	O5	−0.795	−0.136	1.665
13	C6	−0.143	−0.080	2.060
14	O6	−0.837	−0.136	1.665
15	C7	−0.068	−0.080	2.060
16	O7	−0.523	−0.152	1.770
17	C8	0.097	−0.080	2.060
18	O8	−0.533	−0.120	1.700
19	C9	−0.010	−0.070	1.982
20	C10	−0.158	−0.070	1.982
21	C11	−0.123	−0.070	1.982
22	C12	−0.168	−0.070	1.982
23	C13	0.087	−0.070	1.982
24	C14	−0.001	−0.070	1.982
25	C15	0.655	−0.070	2.000
26	C16	−0.502	−0.040	1.960
27	C17	0.534	−0.040	1.960
28	C18	0.139	−0.080	2.060
29	C19	0.155	−0.080	2.060
30	C20	0.155	−0.080	2.060
31	C21	0.537	−0.070	2.000
32	C22	−0.273	−0.080	2.060
33	H1	0.189	−0.042	1.330
34	H2	0.087	−0.042	1.330
35	H3	0.072	−0.042	1.330
36	H4	0.062	−0.042	1.330
37	H5	0.062	−0.042	1.330
38	H6	0.427	−0.050	0.760
39	H7	0.122	−0.030	1.278
40	H8	0.192	−0.030	1.278
41	H9	0.127	−0.030	1.278
42	H10	0.367	−0.046	0.225
43	H11	0.392	−0.046	0.225
44	H13	0.427	−0.050	0.760
45	H14	0.017	−0.039	1.210
46	H15	0.017	−0.039	1.210
47	H16	0.017	−0.039	1.210
48	H17	0.017	−0.039	1.210
49	H18	0.017	−0.039	1.210
50	H19	0.017	−0.039	1.210
51	H20	0.317	−0.046	0.225
52	H21	0.327	−0.046	0.225
53	H22	0.087	−0.042	1.330
54	H23	0.087	−0.042	1.330
55	H24	0.087	−0.042	1.330

Units of e, kcal/mol and Å.

References

1. Lederer, T.; Kintrup, M.; Takahashi, M.; Sum, P.; Ellestad, G.; Hillen, W. Biochemistry 1996, 35, 7439.
2. Chopra, I.; Roberts, M. Microbiol Mol Biol Rev 2001, 65, 232.
3. Saenger, W.; Orth, P.; Kisker, C.; Hillen, W.; Hinrichs, W. Angew Chem Int Ed Engl 2000, 39, 2042.
4. Yusupov, M.; Yusupova, G.; Baucom, A.; Lieberman, K.; Earnest, T.; Cate, J.; Noller, H. Science 2001, 292, 883.
5. Othersen, O.; Beierlein, F.; Lanig, H.; Clark, T. J Phys Chem B 2003, 107, 13743.
6. Meindl, K.; Clark, T. J Phys Chem B 2005, 109, 4279.
7. Mackerell, A.; Bashford, D.; Bellott, M.; Dunbrack, R.; Evanseck, J.; Field, M.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph, D.; Kuchnir, L.; Kuczera, K.; Lau, F.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D.; Prodhom, B.; Reiher, W.; Roux, B.; Smith, J.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. J Phys Chem B 1998, 102, 3586.
8. Mackerell, A.; Wiorkiewicz-Kuczera, J.; Karplus, M. J Am Chem Soc 1995, 117, 11946.
9. Jorgensen, W.; Chandrasekar, J.; Madura, J.; Impey, R.; Klein, M. J Chem Phys 1983, 79, 926.
10. Rigler, N.; Bag, S.; Leyden, D.; Sudmeler, T.; Reilley, C. Anal Chem 1965, 37, 872.
11. Allen, F.; Kennard, O. Chem Des Autom News 1993, 8, 1.
12. Allen, F.; Kennard, O. Chem Des Autom News 1993, 8, 31.
13. Allinger, N.; Yuh, Y.; Lii, J. J Am Chem Soc 1989, 111, 8551.
14. Ren, P.; Ponder, J. J Phys Chem B 2003, 107, 5933.
15. Brooks, B.; Bruccoleri, R.; Olafson, B.; States, D.; Swaminathan, S.; Karplus, M. Charmm: a program for macromolecular energy, minimization, and molecular dynamics calculations. J Comput Chem 1983, 4, 187.
16. Carpenter, G.; Grossberg, S. Appl Opt 1987, 26, 4919.
17. MacKerell, A.; Karplus, M. J Phys Chem 1991, 95, 10559.
18. Accelerys, Inc.CHARMM, Version 22. San Diego, CA, 1992.
19. Rauhut, G.; Pulay, P. J Phys Chem 1995, 99, 3093.
20. Caira, M.; Nassimbeni, L.; Russell, J. Acta Cryst B 1977, 33, 1171.
21. Berendsen, H.; Postma, J.; van Gunsteren, W.; DiNola, A.; Haak, J. J Chem Phys 1984, 811, 3684.
22. Ryckaert, J.; Ciccotti, G.; Berendsen, H. J Comput Phys 1977, 23, 327.
23. Simonson, T. InComputational Biochemistry and Biophysics; Becker, O.; Mackerell, A., Jr.; Roux, B.; Watanabe, M., Eds.; Marcel Dekker: New York, 2001;Ch. 9.
24. Beglov, D.; Roux, B. J Chem Phys 1994, 100, 9050.
25. Stote, R.; States, D.; Karplus, M. J Chim Phys 1991, 88, 2419.
26. Hinrichs, W.; Kisker, C.; Duvel, M.; Muller, A.; Tovar, K.; Hillen, W.; Saenger, W. Science 1994, 264, 418.
27. Stephens, C.; Murai, K.; Brunings, K.; Woodward, R. J Am Chem Soc 1956, 78, 4155.
28. Burgess, M. Metal Ions in Solution; Ellis Horwood: Chichester, UK, 1978.
29. Marchand, S.; Roux, B. Proteins 1998, 33, 265.

Molecular mechanics models for tetracycline analogs

Alexey Aleksandrov and Thomas Simonson*

Laboratoire de Biochimie (CNRS UMR7654), Department of Biology,
Ecole Polytechnique, 91128 Palaiseau, France.

*Email: thomas.simonson@polytechnique.fr

Abstract

Tetracyclines (Tcs) are an important family of antibiotics that bind to the ribosome and several proteins. To model Tc interactions with protein and RNA, we have developed a molecular mechanics force field for 12 tetracyclines, consistent with the CHARMM force field. We considered each Tc variant in its zwitterionic tautomer, with and without a bound Mg^{2+}. We used structures from the Cambridge Crystallographic Data Base to identify the conformations likely to be present in solution and in biomolecular complexes. A conformational search by simulated annealing was undertaken, using the MM3 force field, for tetracycline, anhydrotetracycline, doxycycline and tigecycline. Resulting, low-energy structures were optimized with an ab initio method. We found that Tc and its analogs all adopt an extended conformation in the zwitterionic tautomer and a twisted one in the neutral tautomer, and the zwitterionic-extended state is the most stable in solution. Intermolecular force field parameters were derived from a standard supermolecule: we considered the ab initio energies and geometries of a water molecule interacting with each Tc analog at several different positions. The final, rms deviation between the ab initio and force field energies, averaged over all forms, was 0.35 kcal/mol. Intramolecular parameters were adopted from either the standard CHARMM force field, the ab initio structure, or the earlier, plain Tc force field. The model reproduces the ab initio geometry and flexibility of each Tc. As tests, we describe MD and free energy simulations of a solvated complex between three Tcs and the Tet repressor protein.

1 Introduction

Tetracycline (Tc) is one of the most important antibiotics in use today [13]. Several thousand varieties have been synthesized; just a few can be used as broad spectrum antibiotics for both humans and livestock. The most common bacteriostatic action of Tc is to inactivate the bacterial ribosome, so that protein biosynthesis is interrupted and the bacteria die. The most common mechanism of resistance in gram negative bacteria is mediated by the membrane protein TetA, which exports Tc out of the bacterial cell before it can attack its target, the ribosome. The expression of TetA is tightly regulated by the homodimeric Tet repressor (TetR), which binds specifically to operator DNA, upstream of the *tetA* gene. When Tc diffuses into the cell, it is negatively charged, having released a proton, and chelated to form a Tc^{-}:Mg^{2+} complex. It is the Tc^{-}:Mg^{2+} complex that binds to TetR, inducing conformational changes that sharply reduce the affinity of TetR for the DNA [34]. TetR is then released from the DNA and expression of TetA can proceed, conferring resistance on the bacterial cells. Resistance can also arise through mutations in the bacterial ribosome [33, 30].

It is thus of interest to increase our understanding of both Tc:TetR and Tc:ribosome binding. Crystal structures of TetR and TetR:DNA complexes have provided essential information [34], and structures of the 30S ribosome with and without Tc were recently determined [9]. A complementary approach is to develop computer simulation models, which can be used to investigate the structure, dynamics, and thermodynamics of Tc:protein or Tc:ribosome complexes. Recently, we have developed a molecular mechanics model for Tc ("plain" tetracycline) [3] and used it to study Tc:TetR binding [2, 1], as well as Tc:ribosome binding [4].

Tcs are fairly rigid, and they exist in two main conformations: the so-called twisted and extended conformations, characterized by the orientation of the main rings, A–D, of the four-ring Tc scaffold (Fig. 1). All the PDB complexes involve extended Tc. Tc also has several probable protonational states at neutral pH. The "neutral" state, Tc_N, and the "zwitterionic" state, Tc_Z, are shown in Fig. 1. They differ by the transfer of a proton from the O2H2 hydroxyl to N1 of the nearby dimethyl-ammonium group. From the PDB structures, Mg^{2+} is known to bind to Tc between O5 and O6. When it binds, the H6 proton is released, giving a Tc^{-}:Mg^{2+} complex. We found that tetracycline prefers an extended, zwitterionic state both in solution and in complex with the TetR protein [2], the elongation factor EF-Tu (AA & TS, unpublished results) and the ribosome [4]. Tc

is thus preorganized for binding.

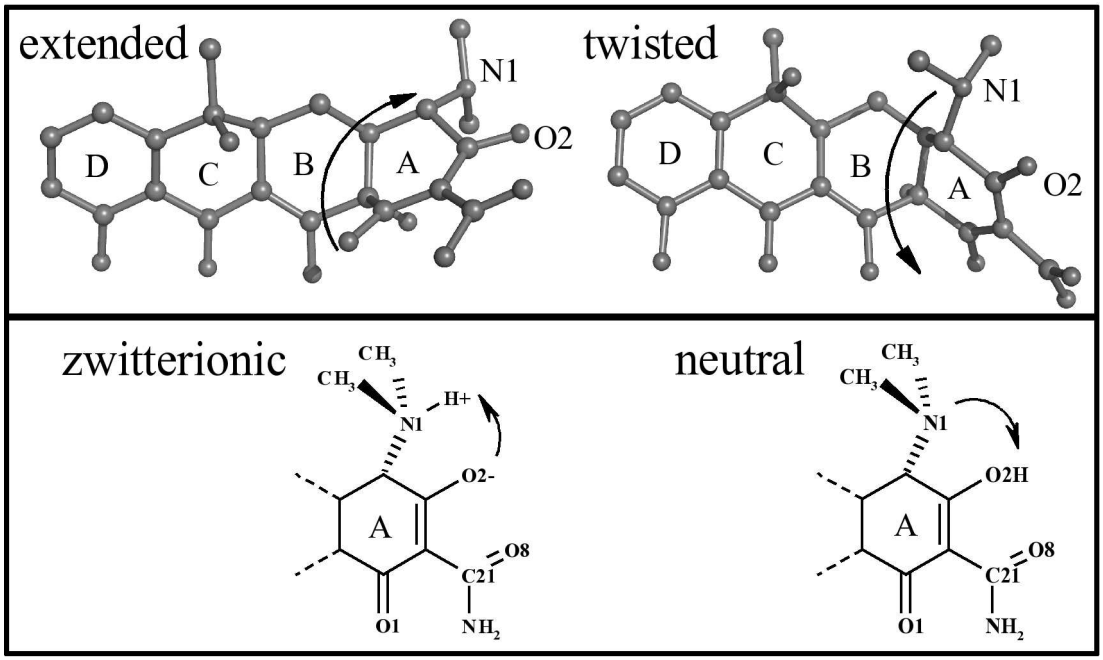


Figure 1: Upper left: extended conformation of tetracycline; upper right: twisted conformation. Lower left: zwitterionic tautomer, Tc_Z; O2 is deprotonated and N1 is protonated. Lower right: neutral tautomer, Tc_N; O2 is protonated and N1 is deprotonated.

Several studies have explored the Tc conformational space [29, 27]. Clark and coworkers used quantum mechanical, Density Functional Theory (DFT) to investigate the conformations of neutral tetracycline (Tc_N) and anhydrotetracycline (aTc) in aqueous solution. They found that the extended Tc_N conformation is about 3 kcal/mol more stable than the twisted one, and the energetic preference for the extended conformation is a solvent effect. For aTc, solvent contributions make the neutral form (in its extended conformation) more favorable than the zwitterionic aTc. For non-ionized aTc, the twisted conformation is the most stable in water at 298 K. Thus, aTc is predicted to adopt the

twisted conformation in its neutral form in solution, and the extended conformation in its zwitterionic form.

Here, we perform a similar conformational analysis for anhydrotetracycline, doxycycline, 4-dedimethylamino-tetracycline and tigecycline. The conformational searches were done by molecular dynamics and simulated annealing, using the MM3 force field. Clustering of the resulting conformations led to several low energy structures that were further optimized with an ab initio, quantum mechanical model. We find that in solution, all four tetracyclines adopt the extended conformation and the three that have a 4-dimethylamino group are in the zwitterionic state (all but 4-ddma-Tc; see abbreviations in Fig. 2). In vacuum, the most stable conformation is twisted, and the tetracyclines having a 4-dimethylamino group are in the neutral tautomer. Thus the presence of the 4-dimethylamino does not change the conformation of the Tcs in vacuum or aqueous solvent, but it changes the protonation state of the 4-hydroxy group in solvent from protonated in 4-ddma-Tc to deprotonated, as in doxycycline.

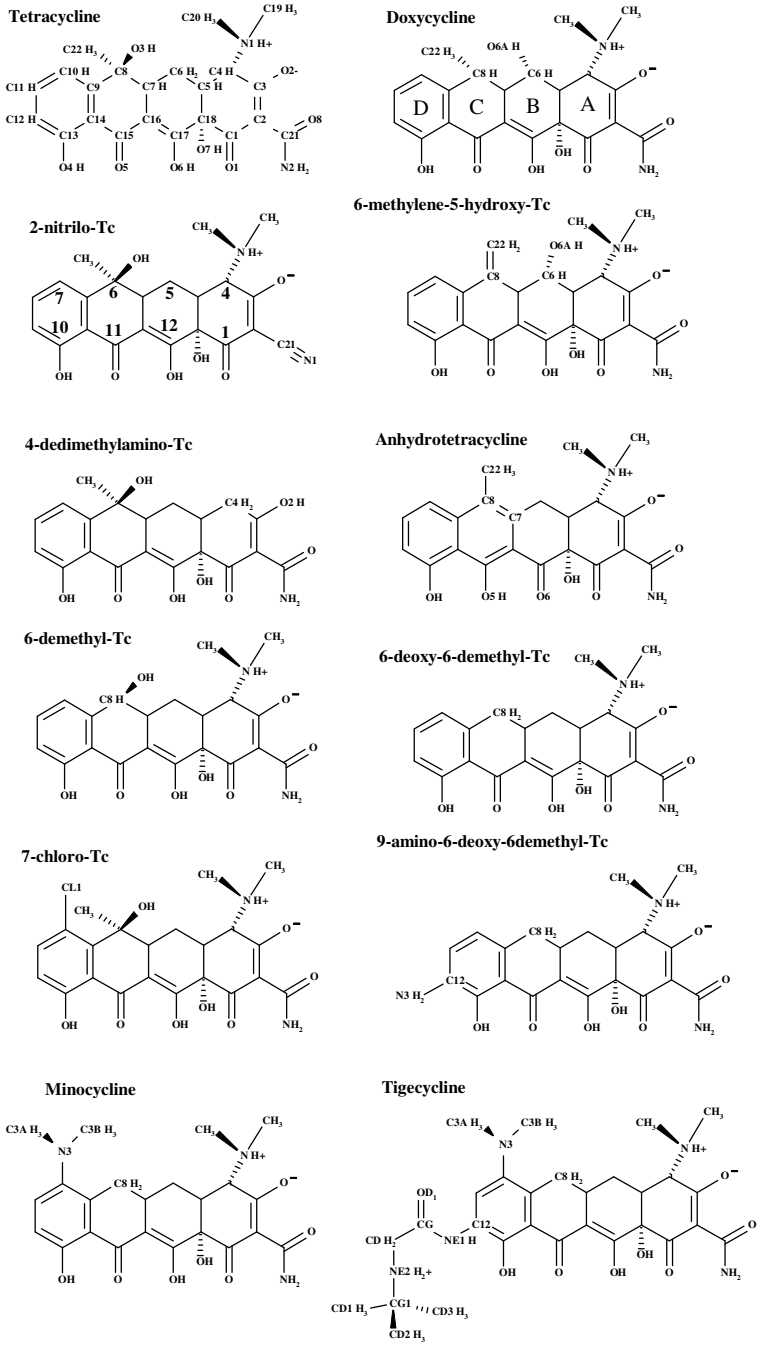


Figure 2: Structures of the 12 tetracycline variants studied here.

Next, we report a force field model for 12 Tcs (including plain Tc, whose force field was described earlier [3]); see Fig. 2. The force field is developed so as to be compatible with the CHARMM27 force field for proteins and nucleic acids [25, 26], and with the TIP3P water model [20]. The 11 tetracycline analogs considered (in addition to plain Tc) are: 4-dedimethylamino-Tc, doxycycline (6-deoxy-5-hydroxy-Tc), 6-demethyl-Tc, 6-deoxy-6-demethyl-Tc, 6-methylene-5-hydroxy-Tc, 7-chloro-Tc, 9-amino-6-deoxy-6-demethyl-Tc, anhydrotetracycline, 2-nitrilo-Tc, minocycline and tigecycline. Chlorotetracycline and oxytetracycline were the first members of the Tc group to be discovered, in the late 1940s. Other tetracyclines were identified later, either as naturally occurring molecules, like tetracycline, or as products of semisynthetic approaches, like doxycycline and minocycline. Tigecycline belongs to the most recently discovered semisynthetic group, referred to as glycylcyclines. Anhydrotetracycline belongs to the group of atypical tetracyclines, which exhibit a different structure-activity relationship from the majority of tetracyclines. They appear to directly perturb the bacterial cytoplasmic membrane, leading to a bactericidal response. 4-dedimethylamino-Tc is exceptional among all considered Tc variants: it is ineffective against Gram-negative bacteria and is devoid of activity *in vivo* against Gram-positive strains [16].

For each Tc variant, force field parameters were developed for the most important protonation and complexation states at neutral pH: Tc_Z (zwitterionic Tc) and the Tc_Z^-:Mg^{2+} complex. We only considered the extended conformation, which is found in all 13 PDB structures involving Tcs. Structures were taken from the conformational analysis described above. Atomic charges were derived from a supermolecule, ab initio approach; Lennard-Jones parameters were adopted from the CHARMM27 force field [25, 26, 14]. Charges developed previously for plain Tc [3] were used as an initial guess for the other Tc variants. We considered the ab initio energies and geometries of a water molecule interacting with each Tc at a few, selected positions. This supermolecule approach is known to give a good balance between solute–water and water–water interactions in the force field [25, 26]. Good agreement was obtained between the ab initio and force field data (rms deviation for the energies of 0.35 kcal/mol). The model also reproduces well the ab initio geometry and flexibility of each Tc. With this model, we could reproduce experimental binding free energy differences between tetracycline and several analogs binding to the Tet repressor (see below) and to the ribosome [4]. The model should thus be suitable to investigate the interactions of tetracycline analogs with proteins and RNA, and particularly the Tc:TetR and Tc:ribosome complexes that are important for

drug design.

In the following sections, we present our Computational Methods, Results, and Conclusions.

2 Computational Methods

2.1 Experimental tetracycline structures

We searched the 2006 release of the Cambridge Crystallographic Data Bank (CCDB) and retrieved 21 crystal structures of tetracyclines (Tc). We also retrieved 6 crystal structures from the Protein Data Bank, corresponding to complexes between a tetracycline and a Tet repressor protein. Based on these structures, we considered two main protonation states of Tc and its analogs: a "neutral" tautomer, Tc_N, and a "zwitterionic" tautomer, Tc_Z, schematized in Fig. 1. The structures can also be grouped into two conformational families, called the "extended" and "twisted" conformations, respectively (Fig. 1).

2.2 Conformational searching with simulated annealing

The MM3 force field [6] and the TINKER program [32] were used to do molecular dynamics and simulated annealing of (plain) tetracycline, doxycycline, anhydrotetracycline and tigecycline molecules. The temperature was varied from 1200 to 0 K over a period of 50 ns. Conformations were drawn from the trajectory every 5 ps and quenched by energy minimization. Clustering was performed with the CHARMM program [10], using the ART-2 algorithm [12] and a 10° threshold for the distances between clusters, measured in the torsion angle space corresponding to the nonhydrogen ring atoms. This led to a total of 50 clusters. 10 structures were retained, corresponding to the cluster centers with the lowest energies. These structures were further optimized quantum mechanically at the B3LYP/6-31G(d) level.

2.3 Solution free energies: ab initio calculations

We used quantum mechanical density functional methods to compute the free energy difference between Tc_N and Tc_Z in solution. We used the B3LYP hybrid functional [7, 23] and the 6-31G* basis set [18]. Geometries were fully optimized with the Gaussian03 program [15]. Single-point energy calculations on the optimized structures were

performed using a dielectric continuum representation of aqueous solvent, through the polarizable continuum model (PCM) [43], and using the MP2 method to account for electron correlation [42].

2.4 Optimization of the intermolecular force field parameters

We adopt a force field of the CHARMM22 form [25], which will later be used to simulate complexes between Tc analogs and proteins or RNA in aqueous solution. The intermolecular energy terms are Lennard-Jones and Coulomb terms. In accord with the development of the CHARMM22 force field, we rely on supermolecule, quantum mechanical calculations on complexes between a Tc variant and a single water molecule. The position of the water is varied around the Tc, primarily at sites involving hydrogen-bonding interactions with polar Tc atoms. The geometry of the Tc variant was either taken from the conformational search, above, or constructed from plain Tc and optimized at the B3LYP/6-31G* level. For the force field calculations, the water geometry was taken from the TIP3P model [20].

Each supermolecule structure was optimized at the HF/6-31G(d) level by varying the interaction distance and a single angle, to find the local minimum for the water position. From the resulting optimal structure, the interaction energy was calculated. No correction for basis set superposition error was made. The ab initio interaction energies were scaled by a factor of 1.16. This factor leads to good agreement between the HF/6-31G(d) Hamiltonian and the TIP3P model for the energy of a water dimer [24]. Applying the same factor to the Tc-water energies should thus lead to a force field with a correct balance between the Tc-water and water-water interactions. Also, to compensate for overestimated interaction distances with the Hartree-Fock model (due to neglected electron correlation), the ab initio interaction distances were reduced by 0.2 Å [25].

The force field parameters were then adjusted to reproduce these "corrected" ab initio interaction energies and water positions. Initial partial charges were obtained from a Mulliken population analysis of the HF/6-31G(d) wavefunction and were compared with charges of similar molecules for which CHARMM parameters are known. Lennard-Jones parameters were adapted from the CHARMM27 force field for similar chemical groups. The parameters optimized for the zwitterionic form of Tc [3] served as an additional guide for the Tc analogs. A few water positions were considered in regions of the Tc variant that differ from plain Tc. Indeed, the differences between Tc and its analogs

are fairly local, so we could usually concentrate on one portion of each Tc variant. For tigecycline and anhydrotetracycline, the molecular differences are greater, so that more water positions were considered: 12 and 17 positions, respectively. We optimized the interaction distance and one angle, corresponding to a rocking of the water around the water–Tc axis. The (corrected) ab initio data were then fitted by varying manually the Tc charges. This involved reoptimizing the Tc-water distance and angle after each parameter change.

2.5 Testing intramolecular force field parameters

In a force field treatment, the intramolecular geometry is mainly determined by the minimum energy values of the bond length and bond angle terms, and by the phase and multiplicity of the dihedrals. These parameters were optimized for plain Tc previously [3]. The Tc analogs considered here can be obtained from plain Tc by fairly local modifications; for example, we get 6-deoxy-6-demethyl-Tc by changing two hydroxyl and methyl groups of Tc into two hydrogens. Since 6-hydrogens are analogous to hydrogens at position 5 in Tc, parameters of 5-hydrogens in plain Tc can be transferred to the new atoms. Thus, intramolecular parameters were adopted either from the standard CHARMM force field for proteins and nucleic acids or from the plain Tc force field. As a further test, ab initio normal modes were computed and compared to the molecular mechanics modes. The B3LYP/6-31G(d) ab initio frequencies were scaled by 0.96, since this factor has been found to give good results in other studies of organic molecules [31].

2.6 Protein simulations

MD simulations were performed for Tc and two variants bound to the class D Tet Repressor (TetR). The crystal structure of TetR, complexed with two tetracyclines, was taken from the PDB [19]. A Tc analog was superimposed on each Tc, based on carbon atoms of the main ring system, and Tc was removed. The simulations included protein residues within a 24 Å sphere, centered on the Mg^{2+} ion of one of the two Tc:Mg^{2+} complexes bound to the protein. Hydrogens were constructed with ideal stereochemistry. In addition to crystal waters, a 24 Å sphere of water was overlaid and waters that overlapped protein, crystal waters, Tc, or Mg^{2+} were removed. Protein atoms between 20 Å and 24 Å from the sphere's center were harmonically restrained to their experimental positions, while protein residues beyond 23.5 Å had their charges switched off. Simu-

lations of the protein system were performed with the SSBP solvent model [8], which treats the region outside the 24 Å sphere as a uniform dielectric continuum (see above), with a dielectric constant of eighty. This is reasonable, since most of the outer region is water. Newtonian dynamics were used for the innermost region, within 20 Å of the sphere's center; Langevin dynamics were used for the outer part of the sphere, referred to as the buffer region. We used a two femtosecond timestep and a bath temperature of 292 K. The CHARMM22 forcefield was used for the protein [25]. The TIP3P model was used for water [20].

2.7 Tc vs. dTc and 7ClTc binding to TetR: free energy simulations

To compare TetR binding by Tc and an analog (dTc or 7ClTc), we use the MD free energy, or MDFE method [11, 36]. It relies on the thermodynamic cycle in Figure 3.

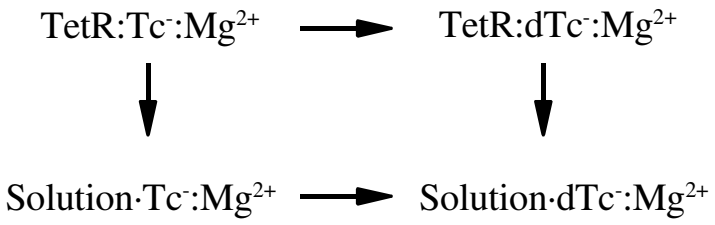


Figure 3: Thermodynamic cycle for TetR:Tc binding. Vertical legs represent Tc and dTc binding/dissociation; horizontal legs represent the alchemical transformation of Tc into dTc in solution (below) or in the solvated TetR:Tc complex (above).

We perform MD simulations of the horizontal legs, which correspond to the reversible, alchemical transformation of one ligand into the other, either in solution, or in complex

with TetR. The corresponding work is derived from a thermodynamic integration formula, below (Eq. 2) [35]. To simulate the two legs of the thermodynamic cycle, we considered two systems: the Tc^-:Mg^{2+} complex in the center of a water sphere with a 24 Å radius, and the same sphere containing Tc^-:Mg^{2+} bound to TetR(D). Protein groups outside this sphere were deleted and replaced by dielectric continuum solvent (see below). In each system, the energy function can be expressed as a linear combination of Tc and dTc (or 7ClTc) terms:

$$U(\lambda) = U_0 + (1 - \lambda)\ U(\mathrm{Tc}) + \lambda\ U(\mathrm{dTc})) \tag{1}$$

where λ is a "coupling parameter" and U_0 represents interactions between parts of the system other than the ligand. The free energy derivative with respect to λ has the form:

$$\frac{\partial G}{\partial \lambda}(\lambda) = \langle U(\mathrm{dTc}) - U(\mathrm{Tc}) \rangle_\lambda \tag{2}$$

where the brackets $\langle\rangle_\lambda$ indicate an average over an MD trajectory with the energy function $U(\lambda)$ [35]. We gradually mutated Tc into dTc by changing λ from zero to one. The successive values used were: 0.001, 0.01, 0.05, 0.1, 0.2, 0.4, 0.6, 0.8, 0.9, 0.95, 0.99, and 0.999. The free energy derivatives were computed at each λ value from a 100 ps MD simulation; the last 80 ps were used for averaging. The total mutation run thus corresponded to 1.2 ns. We performed two runs: forward (from $\lambda = 0$ to 1) and backward (from $\lambda = 1$ to 0). The ion position was harmonically restrained at the center of the water sphere with a force constant of 2 kcal/mol/Å^2. This restraint does not affect the free energy, since its contribution cancels between the two legs. We used the "Spherical Solvent Boundary Potential", or SSBP solvent model, which treats the region outside the 24 Å sphere as a uniform dielectric continuum with a dielectric constant of 80 [8], thus neglecting distant protein groups and simulating a bulk solution. The position of the dielectric boundary between the inner sphere and the outer continuum was determined "on the fly", by the position of the outermost water molecule. A multipolar expansion with 20 terms was used to approximate the reaction field due to the surrounding continuum. Electrostatic interactions between atoms within the sphere were computed without any cutoff, using a multipole approximation for distant groups [39].

3 Results

3.1 Protonation states and conformations of tetracycline variants

3.1.1 Protonation states and conformations in the CCDB and the PDB

Based on the CCDB structures, we previously showed that in plain Tc, O5 is always deprotonated, while O4 and O6 are always protonated [3]. None of the CCDB structures include a metal cation bound in this region. We assume that when Mg^{2+} binds between O5 and O6, as seen in the PDB structures, O6 will become deprotonated and Tc will be negatively charged. For the O2/N1 region, we find a mixture of neutral and zwitterionic tautomers in the CCDB. From the observed bond lengths, we can infer that 13 structures are zwitterionic, and all of these are in the twisted conformation. Six structures are neutral, and all of them are in the extended conformation.

In the experimental PDB structures, the limited experimental resolution does not allow the protonation state to be inferred. Only the extended Tc conformation is found. In our previous study [2] we showed by MD free energy methods that Tc bound to TetR is in its zwitterionic state. In the ribosome [4] and elongation factor Tu (AA & TS, unpublished results), tetracycline was also found in its zwitterionic form, suggesting that only this, zwitterionic–extended form of tetracycline is relevant in biological complexes.

3.1.2 Simulated annealing conformational search and ab initio optimizations

In this study, the conformation search performed earlier for plain Tc was extended to three Tc analogs in two tautomeric states: neutral and zwitterionic. Simulated annealing with the MM3 force field was followed by conformational clustering. Using a 10° threshold distance between cluster centers, computed from the dihedral angles of the main rings, in combination with all possible orientation of hydroxyl hydrogens, we obtained around 50 clusters. The ten lowest-energy structures in the twisted and extended conformation (each) were optimized quantum mechanically, at the B3LYP/6-31G(d) level.

It is known that substitutions at the 4-dimethylamino position invariably reduce antibacterial activity [13], thus it is of general interest to investigate the structure of this group in solvent. We find that the 4-dimethylamino moiety adopts preferentially one of two conformations: one where it makes a hydrogen bond to oxygen O7, and one where it makes a hydrogen bond to the hydroxyl group at position 3. The rms deviations

(RMSD) between the lowest energy structures corresponding to the two orientations of the 4-dimethylamino group and the experimental structures are summarized in Table 1. In the Tet Repressor, the ribosome, and Elongation Factor Tu, the 4-dimethylamino group points towards the hydroxyl group at position 3, while in the TtgR protein, which belongs to the TetR family [5], the 4-dimethylamino group interacts with oxygen O7.

Table 1: Rms deviations (Å) between experimental and ab initio Tc structures

structure	TetR[a]	TTGR[b]	Ribosome[c]	Ef-Tu[d]
S_{O7}[e]	0.8	**0.4**	0.9/1.0	0.6
S_{O2}[e]	**0.3**	0.6	**0.4/0.6**	**0.5**
TetR	–	0.7	0.4/0.5	0.5
TTGR	0.7	–	0.7/0.8	0.5
$Ribosome_1$	0.4	0.7	–/0.4	0.6
$Ribosome_2$	0.5	0.8	0.4/–	0.8
Ef-Tu	0.5	0.5	0.6/0.8	–

Deviations are computed for the non-hydrogen atoms of Tc. The lowest values for each column are in bold face. [a]PDB code 2TRT [19]. [b]PDB code 2UXO [5]. [c]PDB code 1HNW [9] (there are two bound Tc molecules in the structure). [d]PDB code 2HCJ [17]. [e]The two lowest energy structures of tetracycline in its zwitterionic-extended conformation found in this work. In S_{O7}, O7 makes a hydrogen bond with the 4-amino hydrogen. In S_{O2}, the 4-amino hydrogen interacts with the oxygen O2.

The energies of the most stable structures are shown in Table 2. For all three Tc analogs that have a 4-dimethyl group, the most stable conformation in vacuum is neutral–twisted, while in solution it is zwitterionic–extended. In solution, all the neutral Tcs are in the twisted conformation and all the zwitterionic Tcs are in the extended one, consistent with the structures found in the CCDB (see above).

Table 2: Results of the conformation analysis for tetracycline and its analogs

Tc analog	tautomer	conformation[a]	Solute energy[b]	Solvent contribution[c]	Relative energy
4-dedimethylamino-Tc	neutral	extended	0.0	**0.0**	**0.0**
		twisted	**-0.2**	2.3	2.1
plain Tc	neutral	extended	-13.5	19.3	5.8
		twisted	**-18.3**	22.4	4.1
	zwitterionic	extended	0.0	**0.0**	**0.0**
		twisted	0.4	2.1	2.5
anhydro-Tc	neutral	extended	-11.3	18.3	6.9
		twisted	**-18.4**	19.6	1.2
	zwitterionic	extended	0.0	**0.0**	**0.0**
		twisted	-6.4	11.7	5.3
doxycycline	neutral	extended	-11.0	18.7	7.7
		twisted	**-21.6**	25.2	3.6
	zwitterionic	extended	0.0	**0.0**	**0.0**
		twisted	-0.4	3.6	3.2

Energies (kcal/mol) are given for the lowest energy structures for each tautomer. [a]Refers to the two main classes of Tc conformations. [b]Calculated by the MP2/6-31G* method in vacuum. [c]Contribution from dielectric continuum solvent.

For 4-dedimethylamino-Tc, the twisted conformation is the most favorable in vacuum, while in solution, the extended conformation is more favorable. Thus, in solution, the 4-dimethyl amino group does not change the Tc conformation (extended) but it changes the protonation state of O4 to deprotonated. This can have a dramatic affect on the binding of Tc to TetR, since Tc makes important interactions with His64, Gln116 and Asn82 of TetR via the O4 hydroxy group. Indeed, it was found experimentally that 4-ddma-Tc does not bind to TetR of class B [22]. It probably does not bind to TetRs of other classes, since His64, Gln116 and Asn82 are strictly conserved among all TetR classes [1].

3.1.3 Protonation states of 9-amino-6-deoxy-6-demethyl-Tc

The experimental pK_a's of 2-aminophenol are 9.99 and 4.86 for the 1-hydroxyl and 2-amino groups, respectively [37]. The pK_a of the hydroxyl group is thus close to the pK_a of the hydroxyl group in phenol, 9.95 pH units. The low pK_a of the amino group of 2-aminophenol suggests that the 9-amino group in 9-amino-6-deoxy-6-demethyl-Tc is deprotonated. In its complex with Mg^{2+}, electrostatic interactions between the 9-amino group and the Mg^{2+} ion can only decrease its pK_a further.

3.1.4 Protonation states of minocycline

The experimental pK_a's of 4-dimethylaminophenol are 6.15 and 10.1 for the amino and hydroxy groups, respectively [28]. The 4-dimethylamino substitute thus does not have a strong influence on the pK_a of the hydroxyl group, whose pK_a in plain phenol is 9.95. The pK_a of the amino group is slightly lower than the physiological pH. Thus, the 7-dimethylamino group in minocycline is probably in its neutral state.

3.1.5 Protonation states of tigecycline

Tigecycline can be obtained from minocycline by attaching a t-butylglycylamido group at position 9. The carbonyl oxygen is deprotonated because of the lower pK_a of the acetamide oxygen [40]. The pK_a of dimethylamine is 10.7 for the second hydrogen of the amine nitrogen [40], suggesting that the long 9-substitute has a positive charge on the nitrogen. The distance between the positively-charged amine group of tigecycline and the Mg^{2+} ion is more than 8 Å when this compound is simulated in solvent (not shown), suggesting that it has a small influence on Mg^{2+} binding. The unique feature of glycylcyclines (including tigecycline) is their ability to overcome both the ribosomal protection and the efflux resistance mechanisms associated with other tetracyclines [41]. This may be explained by the positive charge on the 9-(t-butylglycylamido) group of tigecycline, which could influence its export by the TetA protein, and also provide favorable interactions with negatively-charged RNA in the ribosome.

We also performed a conformational analysis of tigecycline in its zwitterionic form. In the corresponding MD simulation, tigecycline adopted both extended and twisted conformations. Its long, 9-tert-butyl-glycylamido tail adopted two distinct conformations, *cis* and *trans* relative to the ring D, as shown in Fig. 4. Representative structures in the *cis* and *trans* conformations were optimized quantum mechanically at the HF/6-31G(d)

level. Table 3 summarizes the results. The free energy difference between the two conformations was within the ≈2 kcal/mol uncertainty of the method, suggesting that in complexes with proteins and/or RNAs, tigecycline can adopt either conformation.

Table 3: Conformational analysis of tigecycline

tail conformation[a]	Solute energy[b]	Solvent contribution[c]	Relative energy
towards front	**0.0**	**0.0**	**0.0**
towards back	0.2	1.9	2.1

Energies (kcal/mol) are given for the lowest structures of each tautomer. [a]Refers to the cis and trans positions of the 9-tert-butylglycylamido sidechain relative to the front of the molecule (orientation as for Tc in Figure 1). [b]Calculated by the B3LYP/6-31G* method in vacuum. [c]Dielectric continuum solvent contribution.

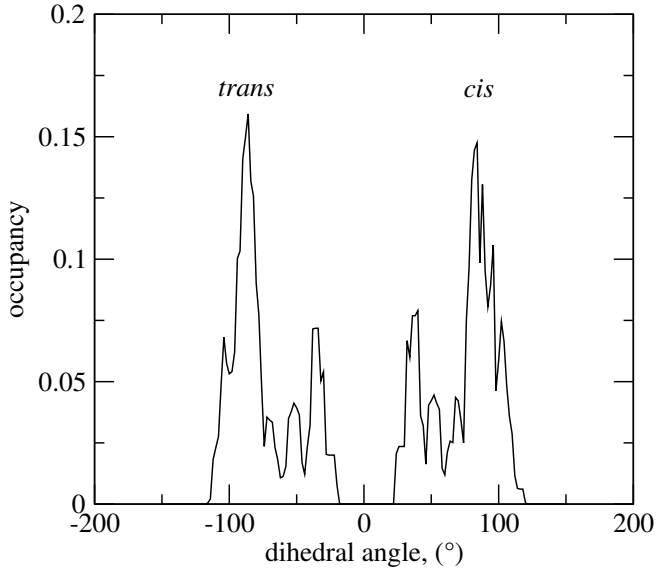


Figure 4: Distribution of the 9-tert-butyl-glycylamido sidechain dihedral angle of tigecycline computed from simulated annealing. The values sampled during the MD simulation were binned in 5° intervals and the corresponding histogram was normalized. The plot thus reveals the occupancy of the cis (towards the front of tigecycline) and trans (towards the back) conformations.

3.1.6 Protonation states of 4-ddma-6-deoxy-6-demethyl-Tc and Mg^{2+} binding

An experimental pK_a of 9.7 was reported for tetracycline hydrochloride [38] and assigned to O4, O5, O6 and O7. We computed the pK_a's for proton binding to O4, O5, O6, O7, and O2 in 4-ddma-6-deoxy-6-demethyl-Tc. The computed free energy differences between protonation states allowed us to deduce the corresponding pK_a's. The results are summarized in the Table 4

Table 4: Analysis of protonation states of 4-dedimethylamino-6-deoxy-6-demethyl-Tc

	Solute energy[b]			Solvent	Relative	
Tc[a]	HF/6-31G*	MP2/6-31G*	MP2/6-31++G*	contribution[c]	energy	pK_a
O4	15.3	18.0	16.4	**-5.5**	10.9	17.6
O6	**0.0**	**0.0**	**0.0**	0.0	**0.0**	9.7
O7	22.7	25.5	20.6	-2.7	17.9	22.7
O2	1.7	4.7	2.2	-2.3	**-0.1**	9.6

[a]Deprotonation site. [b]In vacuum. [c]Contribution from dielectric continuum solvent.

The lowest pK_a's are 9.7 and 9.6, and correspond to O6 and O2. O6 is deprotonated in the Tc^-:Mg^{2+} complex. The pK_a's of O4 and O7 are 17.6 and 22.7, respectively. The high pK_a of O4, compared to O6, suggests that binding of a metal ion between O4 and O5 is unlikely. Thus, free 4-ddma-6-deoxy-6-demethyl-Tc is expected to be neutral in solution. The relatively low pK_a of O2 suggests that Mg^{2+} could chelate between O2 and O8. To test this, we compared the two putative Mg^{2+} binding sites by ab initio analysis. An Mg^{2+} ion with four waters was positioned to chelate 4dma-Tc between O2 and O8, and the structure was optimized at the HF/6-31G(d) level. The energies corresponding to the two Mg^{2+} binding sites are shown in Table 5. The energy difference was 3.7 kcal/mol, which is within the uncertainty of the method, suggesting that metal binding between O2 and O8 may be possible.

Table 5: Possible Mg^{2+} binding sites in 4-dedimethylamino-6-deoxy-6-demethyl-Tc

Mg^{2+}	Solute energy[a]			Solvent	Relative
site	HF/6-31G*	HF/6-31++G*	MP/6-31G*	contribution[b]	energy
O5, O6	**0.0**	**0.0**	**0.0**	0.0	**0.0**
O2, O8	3.5	4.6	5.6	**-1.9**	3.7

Energies in kcal/mol. [a]In vacuum. [b]Dielectric continuum solvent contribution.

3.2 Force field determination

We limit ourselves to the extended conformation of each Tc variant, since this is the only conformation observed in crystal structures of protein and RNA complexes. In the CCDB, extended conformations are all in the zwitterionic tautomer; the conformational analysis showed that the extended-zwitterionic state is predominant in solvent. In our previous study [2], we showed that TetR strongly prefers the zwitterionic tautomer of Tc. The same result was observed for the ribosome [4]. In all, we consider 11 Tc analogs (Fig. 2): 4-dedimethylamino-Tc, doxycycline (6-deoxy-5-hydroxy-Tc), 6-demethyl-Tc, 6-deoxy-6-demethyl-Tc, 6-methylene-5-hydroxy-Tc, 7-chloro-Tc, 9-amino-6-deoxy-6-demethyl-Tc, anhydro-Tc, 2-nitrilo-Tc, minocycline and tigecycline.

3.2.1 Atomic partial charges of Tc analogs in the absence of Mg^{2+}

For each Tc:water complex, the ab initio optimized water position is compared to the position obtained at the end of the force field fitting. Interaction energies are reported in Fig. 5. More details are given in Supplementary Material. The ab initio data are "corrected" by energy scaling and distance shifting, as described in Methods. The rms deviation between the ab initio and force field data for interactions with two degrees of freedom (Tc:water distance and water rocking angle) are 0.35 kcal/mol for the energies, 0.16 Å for the Tc–water distances, and 26° for the rocking angle. The largest errors are 1.3 kcal/mol, 0.32 Å, and 77°. The resulting force field parameters (charges and Lennard-Jones parameters) are given in Supplementary Material. The initial guess for the charges was based either on the tetracycline charges developed previously [3], or the CHARMM27 charges in cases where a very similar group existed in the CHARMM27 force field, or the HF/6-31G(d) Mulliken charges in all other cases. The largest change between the initial and final charges was $0.2e$; the average change was $0.1e$. Normally, only a limited number of water molecules around sites of Tc variation were considered, except for anhydrotetracycline, which has a different electronic structure of the main ring atoms (see Fig. 2), and tigecycline, which has a large substituting group. For these two Tc variants, 17 and 12 water molecules were considered, respectively.

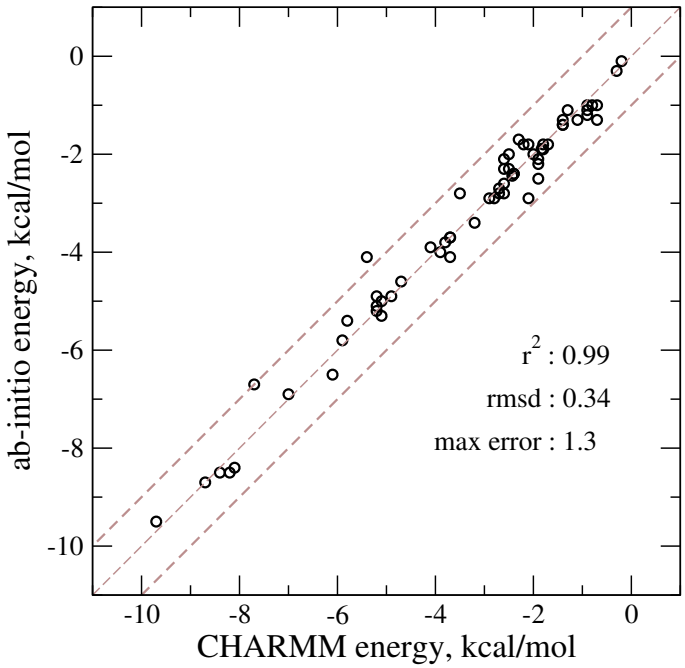


Figure 5: Comparison between ab initio and empirical interaction energies between the Tc analogs and the probe water molecules.

For 4-dedimethylamino-tetracycline, we started from the charges of tetracycline in its neutral state developed previously [3]. Two water molecules interacting with the C4 hydrogens were considered. Only the charges of carbon C4 and its bonded hydrogens were varied during the optimization. In 7-chloro-tetracycline, the interaction between the chloride and a water molecule was weak, just -0.3 kcal/mol (Supplementary Material). Thus, the chloride is more hydrophobic than the hydrogen at this position in Tc, which has an interaction energy with water of -2.6 kcal/mol [3]. Indeed, in the Tc complex with TetR, the D ring is in a hydrophobic pocket [2], and 7-chloro-Tc has a higher association constant with TetR than Tc; the binding free energy difference between the two is -0.7 kcal/mol [22].

In the CHARMM force field for small molecular compounds, the charge of hydroxyl oxygens is -0.54e in phenol, -0.61e in acetic acid, and -0.66e in methanol, ethanol and 2-propanol. In the Tc analogs, we have 6 hydroxyl groups; the charges of the oxygen varies

from -0.475e to -0.595e; thus, it is usually less negative than in the CHARMM force field. We see for example that removing the 6-methyl group increases the negative charge on the 6-hydroxyl oxygen from -0.53e in tetracycline to -0.595e in 6-demethyl-tetracycline. The same is true for the charge of the 5-hydroxyl oxygen in doxycycline and 6-methylene-5-hydroxy-Tc, which is -0.573e (see Supplementary Material). Thus, the 6-methyl group has a hydrophobic effect and can substantially change the interactions of Tc with its targets. In the ribosome and EF-Tu, this group is solvent-exposed [9, 17] while in the Tet Repressor, which is a part of the resistance mechanism, it is in a hydrophobic pocket [2].

For 6-deoxy-6-demethyl-Tc, the charges of 6-demethyl-Tc were used as an initial guess; only two waters were used. The charges of 6-deoxy-6-demethyl-Tc were used as an initial guess for 9-amino-6-deoxy-6-demethyl-Tc, minocycline and tigecycline. The charge of the 9-amino nitrogen of 9-amino-6-deoxy-6-demethyl-Tc is -0.798e. There are no anilines in the standard CHARMM force field and the closest analog is cytosine, which has a -0.75e charge on the amino nitrogen. The charge of the amino hydrogen *cis* to oxygen O4 is 0.47e; it is 0.37e for the *trans* hydrogen. These values are fairly close to cytosine's hydrogen charges: 0.37e when *cis* to the hydroxyl group and 0.33e when trans.

Doxycycline charges were used as a starting point to develop the 6-methylene-5-hydroxy-Tc partial charges. One hydrogen was removed from the 6-methyl group and one from the 6-carbon bonded to the C ring. The structure was optimized at the B3LYP/6-31G* level. Four water molecules were considered around the 6-methylene group, rings B and D. The interaction energies are listed in Supplementary Material. During charge optimization, only the charges of the 6-methylene group and the main ring carbon, C8, were varied. The deviation between the empirical and ab initio energy was maximal for 6-methylene-5-hydroxy-Tc among all Tc analogs: 1.3 kcal/mol for a water molecule interacting with the hydrogen of C5 in the main ring. Though the absolute difference is large, the relative error is just 24%. For the original CHARMM27 force field, the maximum error was 0.79 kcal/mol, for a guanine-water interaction [14]. We note that it is more difficult to optimize a force field for a larger compound such as Tc. The Tc analogs have an average of more than 50 atoms each; tigecycline has over 80 atoms. In the present work, we could not simplify the parametrization by breaking Tc down into smaller fragments, which would then be parameterized separately. Rather, the whole Tc molecule was treated together, due to its fused ring structure.

The 2-dimethylamino group of tetracycline in its neutral state served as an initial guess for the charges of the minocycline 7-dimethyl-amino group. For the rest of minocycline, charges from 6-deoxy-6-demethyl-Tc were used. The charges of the 7-dimethylamino-group and the carbon of ring D to which it is bonded, C10, were varied during the optimization. The final charges of the 7-methyl-groups were smaller than those of the 2-dimethyl-amino group of plain Tc, while the nitrogen is more negatively charged, -0.517e, versus -0.462e in plain Tc. The 7-amino nitrogen withdraws more electron density from the carbon of the D ring and thus borrows less electron density from its dimethyl group, compared to the 2-dimethylamino group of plain Tc.

Though anhydrotetracycline differs from Tc by just one hydroxyl group and a hydrogen, the electronic structure of its main ring system is substantially different. In anhydrotetracycline, the carbons C8 and C7 are double-bonded, which leads to a rearrangement in the 11, 12 enolate groups: O5 is protonated and O6 is deprotonated. Ring C is planar and the 6-methyl carbon is coplanar with it. We employ 17 water molecules for optimization of the charges. The charges and interaction energies are listed in Supplementary Material. The 6-methyl carbon has a charge of -0.232e, less negative than in tetracycline and in the standard CHARMM methyls (-0.27e), but more negative than the methyl group of thymine (-0.11e). The minimum and maximum errors were 0.0 and 0.7 kcal/mol, respectively. Notice that the hydroxyl group that gives the maximum error is difficult to parametrize, since it is just 3 Å from the 4-amino-dimethyl group and the probe water molecule strongly interacts with this latter group. The rms deviation from the ab initio interaction energies was just 0.37 kcal/mol.

The minocycline charges served as an initial guess for tigecycline. Charges at positions that correspond to minocycline were frozen during the optimization, except for the carbon C12, which is bonded to the 9-substitute. The optimized charges of the 9-amide group can be compared to the standard CHARMM charges for an N-methylamide C-terminus. The charges of the carbonyl group were 0.52e and -0.52e for carbon and oxygen, respectively, in good agreement with the usual CHARMM charges of 0.51e and -0.51e. The charge of the amide nitrogen, -0.74e, is closer to the cytosine nitrogen charge of -0.75e. The methyl carbons at the end of the 9-substitute were substantially more charged than a "typical" CHARMM methyl carbon: -0.53e instead of the standard -0.27e.

3.2.2 Intramolecular parameters

All the Tc analogs considered here can be obtained from plain Tc by small modifications, such as the change of a 6-methyl into a 6-H for 6-demethyl-Tc (Fig. 2). Thus, for the bonded force field parameters, we used either the Tc parameters developed previously [3] or parameters from the standard CHARMM model compounds [25, 26]. In some cases, the minimum energy bond lengths, bond angles and torsion angles were taken directly from the ab initio structures, optimized using the B3LYP/6-31G(d), hybrid, Hartree-Fock/density functional theory. To help validate the intramolecular and intermolecular parameters, the normal modes of vibration were computed at the B3LYP/6-31G* level for some of the Tc variants. Normal modes were also computed with the force field model. Vibrational frequencies and rms atomic fluctuations were derived. Comparisons between the quantum mechanical and force field results are given in Figs. 6, 7. Results are shown for doxycycline and tigecycline. Agreement is generally quite good, especially for the atoms of the main, four-ring Tc scaffold. The long-range, atomic, pairwise correlations are also well-reproduced (not shown), as are the lowest vibrational frequencies. This shows that the overall plasticity of doxycycline is reproduced, which is especially important for future studies of protein binding. The rms deviation computed between the ab initio minimized structure and the force field minimized structure (after superposition of the whole molecule) is 0.3 Å for doxycycline. The rms deviation for particular groups varies from 0.15 Å for the D ring system (including hydrogens) to 0.40 Å for the acetamide groups of doxycycline. Similar results were obtained for the other Tc analogs.

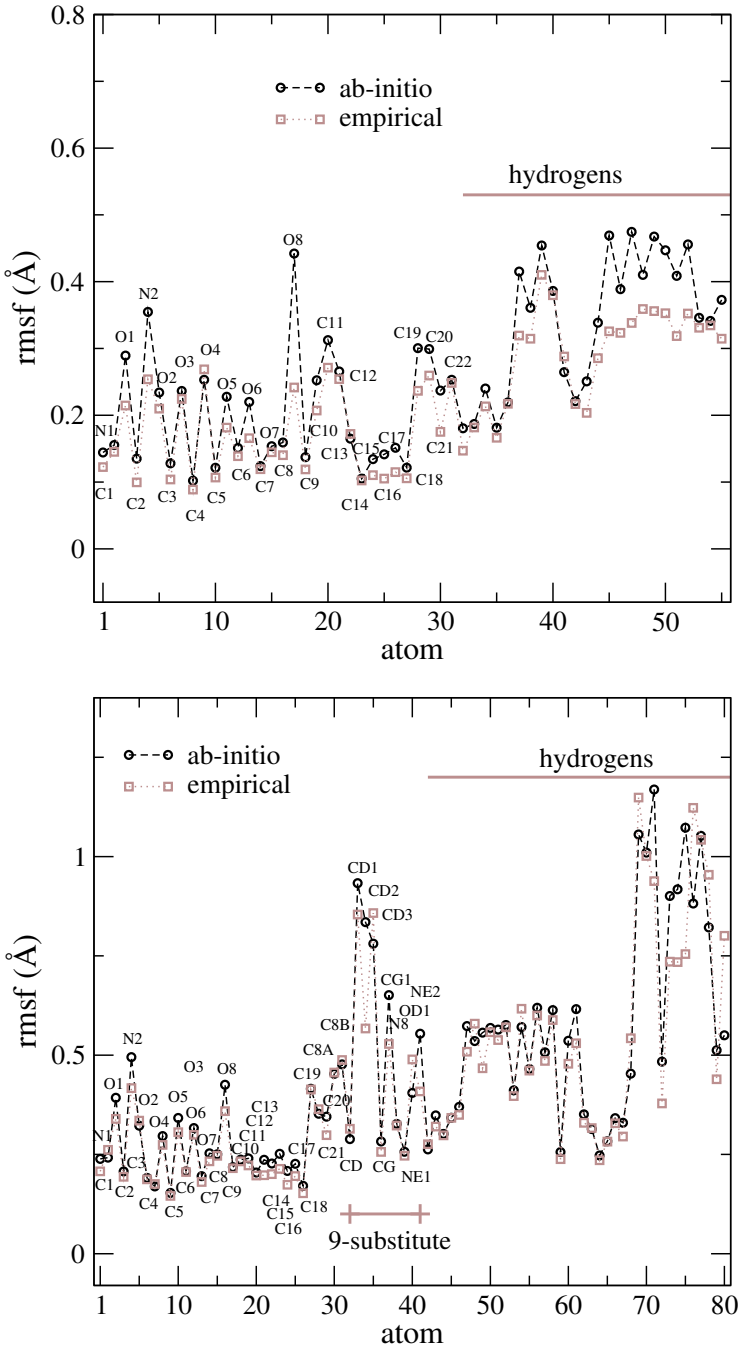


Figure 6: Rms fluctuations (Å) of doxycycline (above) and tigecycline (below), computed in the normal mode approximation. Ab initio result: solid line; force field result: dashed line. Atom numbers are arbitrary.

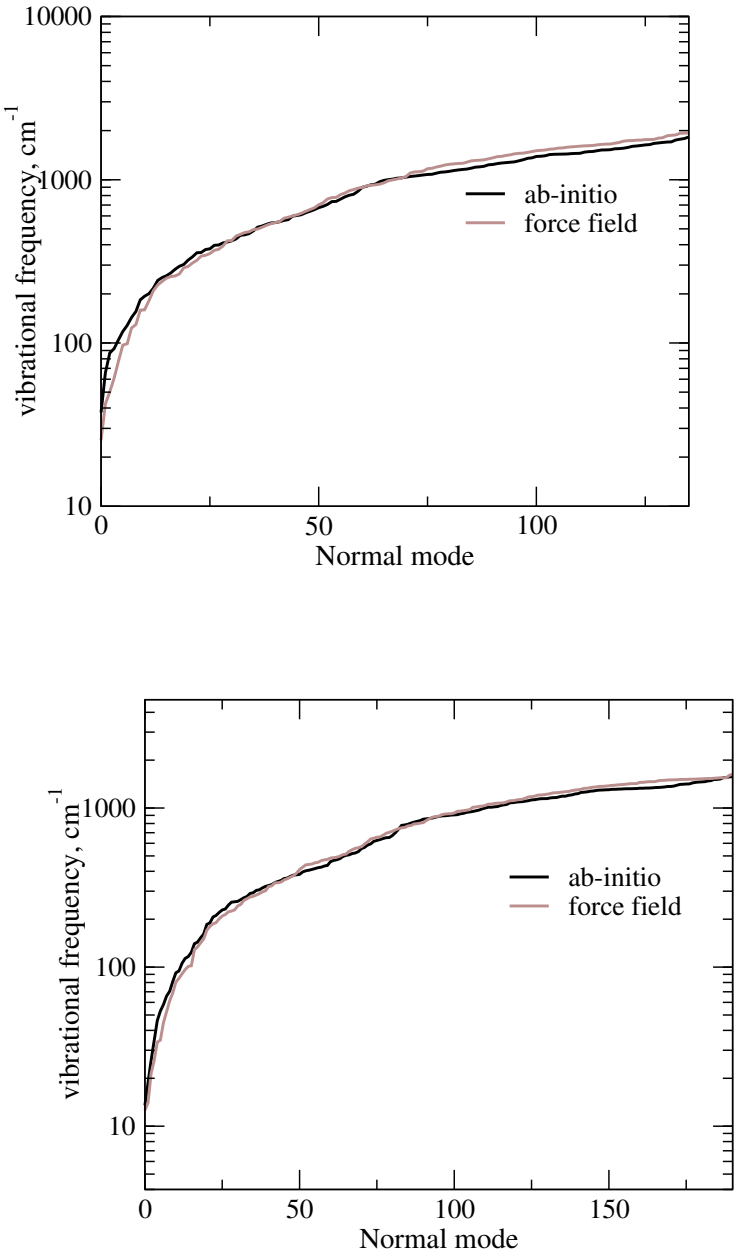


Figure 7: Upper panel: ab initio (grey) and force field (black) vibrational frequencies (cm^{-1}) of doxycycline; lower panel: vibrational frequencies of tigecycline.

3.2.3 Tc^- variants with bound Mg^{2+}

Consider a particular Tc variant, say xTc. To construct a model of the xTc^-:Mg^{2+} complex, we used the following strategy. Mg^{2+} introduces a fairly local polarization within the xTc moiety, and also changes its protonation state. Therefore, we first removed the hydrogen from the oxygen chelating the Mg^{2+}. Then, xTc^-:Mg^{2+} charges were assigned by combining the xTc charges optimized above with the charges of plain Tc and Tc^-:Mg^{2+}, optimized earlier [3]. For an atom i present in both xTc and plain Tc, the new charge Q_i is obtained by the following rule:

$$\begin{aligned} Q_i[\mathrm{xTc^-}\colon \mathrm{Mg^{2+}}] &= Q_i[\mathrm{xTc}] + \Big(Q_i[\mathrm{Tc^-}\colon \mathrm{Mg^{2+}}] - Q_i[\mathrm{Tc}]\Big) \\ &= Q_i[\mathrm{Tc^-}\colon \mathrm{Mg^{2+}}] + (Q_i[\mathrm{xTc}] - Q_i[\mathrm{Tc}]) \end{aligned} \quad (3)$$

Since the total number of atoms is different for some Tc analogs, there was an extra charge which was then distributed on all the xTc atoms. The extra charge on every atom was almost always less than $0.001e$, which cannot substantially change the interactions of xTc with its surroundings.

3.2.4 Tc^-:Mg^{2+}:TetR interactions: molecular dynamics simulations

We also performed MD simulations of the doxycycline:TetR and 7ClTc:TetR complexes for 5 nanoseconds each. We modelled each Tc variant in its zwitterionic form. His64 in TetR was fully protonated, following our previous work [2]. The protein:Tc complex was solvated by a 24 Å water sphere centered on the Tc-bound Mg^{2+}; more distant regions were modeled as a dielectric continuum (see Methods). The Tc analog was positioned by superposition on the Tc present in the crystal structure.

We computed rms deviations from the crystal structure, averaging over the last 50 ps of the 5 ns simulations (Table 6). By superimposing the Tc analog on the crystal structure, we obtain intramolecular Tc deformations of less than 0.3 Å. Superimposing the protein backbone on the crystal structure, we obtain Tc rms deviations of 0.44/0.53 Å for doxycycline/7-chloro-Tc, which correspond partly to an overall shift of the Tc molecule. The protein backbone rms deviations are 0.73/0.67 Å for doxycycline/7-chloro-Tc. Figure 8 shows the good structural agreement between the MD simulation of 7-chloro-Tc:TetR and the experimental X-ray structure [21] (PDB entry code 2TCT). The interactions between the Tc variant and the protein are reasonably well-reproduced (Tables 7, 8), indicating that the Tc shift is accompanied by a corresponding shift of the

protein environment. Overall, the MD model is similar to other good quality, current simulations.

Table 6: Rms deviations (Å) for Tc:TetR complexes

Molecule	ligand[a]	ligand[b]/backbone	backbone	10 Å sphere[c]
tetracycline	0.28	0.53	0.68	1.03
doxycycline	0.29	0.44	0.73	1.00
7-chloro-Tc	0.15	0.53	0.67	1.01

Deviations between the MD and crystal structures. [a]After superimposing the MD ligand on the ligand in the crystal structure. [b]After superimposing the protein backbone on the crystal structure. [c]Atoms in the inner 10 Å of the spherical simulation model. Nonhydrogen atoms only in each case.

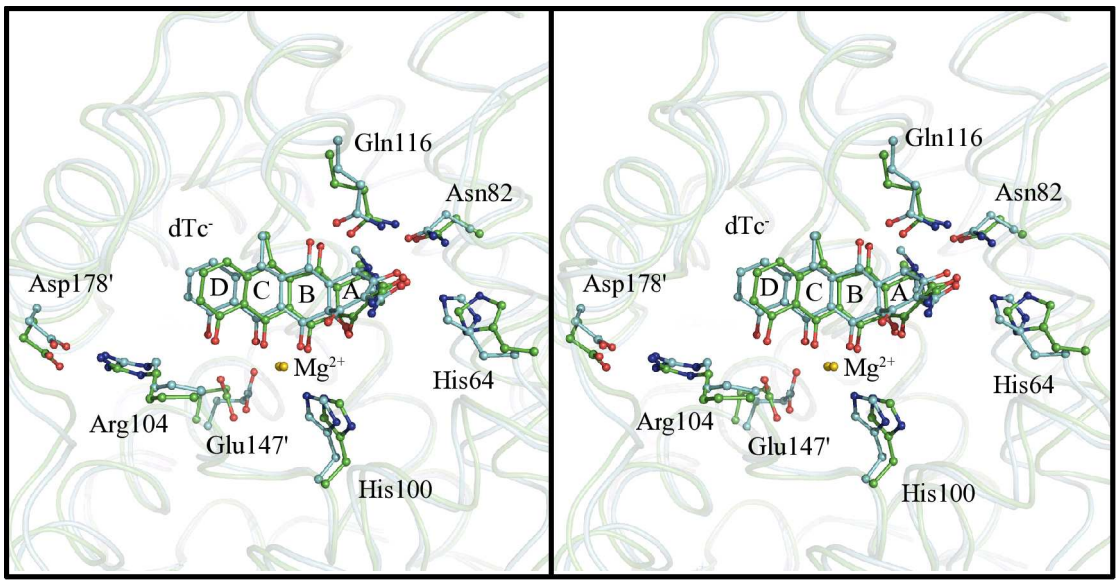


Figure 8: Wall-eyed stereo view of the dTc^{-}:Mg^{2+} binding site from the final nanosecond of the MD trajectory and the experimental X-ray structure (PDB entry 2O7O). The Mg^{2+} cation is shown as a gold sphere; doxycycline as well as important residues are in ball-and-stick representation; the experimental structure is green, and the MD structure is cyan.

Table 7: Selected distances (Å) between atoms of 7-chloro-Tc and TetR(D)

atom pair	MD simulations	Xray structures 2TCT	2TRT
$CD2_{L174}$ $C11_{7ClTc}$	3.5	3.6	3.6
CB_{V113} $O3_{7ClTc}$	3.3	3.5	3.1
CG_{L170} Cl_{7ClTc}	4.6	5.1	5.2
$CD2_{L131}$ Cl_{7ClTc}	5.2	4.0	4.2
CG_{L117} Cl_{7ClTc}	5.1	4.8	4.7

TetR atoms (left) are labelled by the amino acid to which they belong.

Table 8: Selected distances (Å) between atoms of doxycycline and TetR(D)

atom pair	MD simulations	Xray structures 2O7O [2]	2TRT [19]
$OE1_{Q116}$ $O3X_{dTc}$	2.9	2.6	–
$CG2_{I134}$ $O3X_{dTc}$	3.8	3.3	–
CB_{V113} $C22_{dTc}$	4.4	3.9	4.3
CG_{L117} $C22_{dTc}$	5.6	4.0	4.3
$CD1_{L170}$ $C22_{dTc}$	6.6	5.1	5.5

TetR atoms (left) are labelled by the amino acid to which they belong.

3.2.5 Tc vs. dTc and Tc vs. 7ClTc binding: free energy simulations

To compute the relative binding affinites of Tc, dTc, and 7ClTc for TetR(D), we gradually mutated Tc into either dTc or 7ClTc, both in solution and in complex with TetR (horizontal legs in the thermodynamic cycle in Fig. 3). We used the CHARMM22 force field parameters for the protein and Mg^{2+}, the TIP3P water model, the parameters developed previously for Tc [3], and the dTc and 7ClTc parameters derived here. The experimental binding free energy difference between Tc and 7ClTc is only -0.5 kcal/mol,

favoring 7ClTc (Winfried Hinrichs, personal communication). This value presumably results from the solvent penalty to transfer the hydrophobic Tc chloride from solvent into the protein. The computed binding free energy difference is -0.8 (0.5) kcal/mol. The experimental binding free energy difference between Tc and dTc is -1.5 kcal/mol, favoring dTc. This value is explained by the hydrogen bonding between Gln116' and the 5-hydroxyl group of dTc, and also by the fact that the 6-hydroxyl group of Tc is in the hydrophobic pocket formed by the Val113, Pro105, Leu170, Leu117 residues of TetR(D). The computed binding free energy difference, -2.1 (0.9) kcal/mol, is in reasonable agreement with the experimental one.

4 Conclusions

Recently, we parametrized a tetracycline force field [3] and used it to study interactions between tetracycline and the Tet Repressor [2] and the ribosome [4]. We found that Tc prefers an extended, zwitterionic state both in solution and in complex with the protein [2]. Tc is thus preorganized for binding. Here, we extended the earlier conformation analysis to anhydrotetracycline, doxycycline, 4-dedimethylamino-tetracycline and tigecycline. The conformational search was done using the MM3 force field to perform molecular dynamics and simulated annealing of the Tcs. Clustering of the resulting conformations led to several low energy structures which were further optimized with an ab initio, quantum mechanical model. We find that in solution, all the tetracyclines studied adopt the extended conformation and all tetracyclines that have a 4-dimethylamino group are in the zwitterionic state. In vacuum, the most stable conformation is twisted, and tetracyclines that have a 4-dimethylamino group are in the neutral state. Thus, the presence of the 4-dimethylamino does not change the conformation of Tc in vacuum or solvent, but changes the protonation state of the 4-hydroxy group in solvent from protonated in 4-ddma-Tc to deprotonated in variants like tetracycline or doxycycline.

In this work, we also developed a molecular mechanics force field for several important Tc analogs, compatible with the CHARMM27 force field for proteins and nucleic acids. Our method for developing the force field is consistent with the CHARMM22 methodology, but presents also some specific aspects. We started from tetracycline parameters developed previously [3] and extensively tested in recent work [2, 4, 1]. We did not simplify the parametrization by breaking the Tcs down into smaller fragments, which would then be parameterized separately. Rather, the whole Tc was treated at

once. This was due to the fused ring structure of Tc, which makes it hard to split the molecule into smaller pieces. Second, we considered a probe water molecule with two degrees of freedom. This led to a good balance between the quality of interaction energies and geometries. Last, we proceeded in two steps, parameterizing each Tc alone and then introducing the Mg^{2+} cation, making use of the similarity between plain Tc and each variant.

With this model, we were able to reproduce experimental binding free energy differences between tetracycline and two analogs binding to TetR. Good results have also been obtained for additional analogs (unpublished data). Overall, the force field model shows good agreement with both ab initio and experimental data, showing that it is suitable for investigating Tc:protein and Tc:RNA complexes.

5 Acknowledgements

Support was provided by the Nuclear Toxicology program of the Commmissariat pour l'Énergie Atomique and the Centre National pour la Recherche Scientifique. An Egide PhD fellowship was awarded to AA. We thank M. Karplus for the CHARMM program.

References

[1] A. Aleksandrov, L. Schuldt, W. Hinrichs, and T. Simonson. Tet repressor induction by tetracycline: a molecular dynamics, continuum electrostatics, and crystallographic study. *J. Mol. Biol.*, in press:0000, 2008.

[2] Alexey Aleksandrov, Juliane Proft, Winfried Hinrichs, and Thomas Simonson. Protonation patterns in tetracycline:Tet repressor recognition: Simulations and experiments. *ChemBioChem*, 8:675–685, 2007.

[3] Alexey Aleksandrov and Thomas Simonson. The tetracycline:Mg^{2+} complex: a molecular mechanics force field. *J. Comp. Chem.*, 13:1517–1533, 2006.

[4] Alexey Aleksandrov and Thomas Simonson. Molecular dynamics simulations of the 30S ribosomal subunit reveal a preferred tetracycline binding site. *J. Am. Chem. Soc.*, 130:1114–1115, 2008.

[5] Y. Alguel, C. Meng, W. Teran, T. Krell, J. L. Ramos, M. Gallegos, and X. Zhang. Crystal structures of multidrug binding protein TtgR in complex with antibiotics and plant antimicrobials. *J. Mol. Biol.*, 2007.

[6] N.L. Allinger, Y.H. Yuh, and J.H. Lii. Molecular mechanics. the MM3 force field for hydrocarbons. *J. Am. Chem. Soc.*, 111:8551–8566, 1989.

[7] A.D. Becke. Density functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.*, 98:5648, 1993.

[8] D. Beglov and B. Roux. Finite representation of an infinite bulk system: solvent boundary potential for computer simulations. *J. Chem. Phys.*, 100:9050–9063, 1994.

[9] D. E. Brodersen, W. M. Clemons Jr, A. P. Carter, R. J. Morgan-Warren, B. T. Wimberly, and V. Ramakrishnan. The structural basis for the action of the antibiotics tetracycline, pactamycin, and hygromycin B on the 30S ribosomal subunit. *Cell*, 103:1143–1154, 2000.

[10] B. Brooks, R. Bruccoleri, B. Olafson, D. States, S. Swaminathan, and M. Karplus. Charmm: a program for macromolecular energy, minimization, and molecular dynamics calculations. *J. Comp. Chem.*, 4:187–217, 1983.

[11] C.L. Brooks, M. Karplus, and M. Pettitt. Proteins: a theoretical perspective of dynamics, structure and thermodynamics. *Adv. Chem. Phys.*, 71:1–259, 1987.

[12] G.A. Carpenter and S. Grossberg. ART 2: Self-organization of stable category recognition codes for analog input patterns. *Appl. Optics*, 26:4919–4930, 1987.

[13] I. Chopra and M. Roberts. Tetracycline antibiotics: Mode of action, applications, molecular biology, and epidemiology of bacterial resistance. *Microbiology and Molecular Biology Reviews*, 65:232–260, 2001.

[14] Nicolas Foloppe and Alexander D. MacKerell Jr. All-atom empirical force field for nucleic acids: I. Parameter optimization based on small molecule and condensed phase macromolecular target data. *J. Comp. Chem.*, 21:86–104, 2000.

[15] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. A. Montgomery, Jr., T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev,

A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, C. Gonzalez, and J. A. Pople. Gaussian 03, Revision C.02. Gaussian, Inc., Wallingford, CT, 2004.

[16] R. A. Greenwald, W. Hillen, and M. L. Nelson. *Tetracyclines in Biology, Chemistry and Medicine*. Birkhäuser Verlag/Switzerland, 2001.

[17] Susan E. Heffron, Suet Mui, Annette Aorora, Kenton Abel, Ernst Bergmann, and Frances Jurnak. Molecular complementarity between tetracycline and the GTPase active site of elongation factor Tu. *Acta Cryst. D*, 62:1392–1400, 2006.

[18] W.J. Hehre, L. Radom, P.v.R. Schleyer, and J.A. Pople. *Ab initio molecular orbital theory*. John Wiley and Sons, New York, 1987.

[19] W. Hinrichs, C. Kisker, M. Duvel, A. Muller, K. Tovar, W. Hillen, and W. Saenger. Structure of the Tet repressor-tetracycline complex and regulation of antibiotic resistance. *Science*, 264:418–420, 1994.

[20] W. Jorgensen, J. Chandrasekar, J. Madura, R. Impey, and M. Klein. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.*, 79:926–935, 1983.

[21] C. Kisker, W. Hinrichs, K. Tovar, W. Hillen, and W. Saenger. The complex formed between tetracycline repressor and tetracycline-Mg^{2+} reveals mechanism of antibiotic resistance. *J. Mol. Biol.*, 247:260–280, 1995.

[22] T. Lederer, M. Kintrup, M. Takahashi, P.E. Sum, G.A. Ellestad, and W. Hillen. Tetracycline analogs affecting binding to Tn10-Encoded Tet repressor trigger the same mechanism of induction. *Biochemistry*, 35:7439–7446, 1996.

[23] C. Lee, W. Yang, and R.G. Parr. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B*, 37:785–789, 1988.

[24] A. D. MacKerell and M. Karplus. Importance of attractive van der Waals contribution in empirical energy function models for the heat of vaporization of polar liquids. *J. Phys. Chem.*, 95:10559–10560, 1991.

[25] A.D. Mackerell, D. Bashford, M. Bellott, R.L. Dunbrack, J. Evanseck, M.J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph, L. Kuchnir, K. Kuczera, F.T.K. Lau, C. Mattos, S. Michnick, T. Ngo, D.T. Nguyen, B. Prodhom, W.E. Reiher, B. Roux, J. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, D. Yin, and M. Karplus. An all-atom empirical potential for molecular modelling and dynamics study of proteins. *J. Phys. Chem. B*, 102:3586–3616, 1998.

[26] A.D. Mackerell, J. Wiorkiewicz-Kuczera, and M. Karplus. An all-atom empirical energy force-field for the study of nucleic acids. *J. Am. Chem. Soc.*, 117:11946–11975, 1995.

[27] K. Meindl and T. Clark. Conformations and tautomers of 5a,6-anhydrotetracycline. *J. Phys. Chem. B*, 109:4279–4284, 2005.

[28] T. J. Meredith. *Antidotes for poisoning by cyanide*. Cambridge University Press, 1993.

[29] O.G. Othersen, F. Beierlein, H. Lanig, and T. Clark. Conformations and tautomers of tetracycline. *J. Phys. Chem. B*, 107:13743–13749, 2003.

[30] M. Pringle, C. Fellstrom, and K. E. Johansson. Decreased susceptibility to doxycycline associated with a 16S rRNA gene mutation in *brachyspira hyodysenteriae*. *Veterinary Microbiology*, 123(1–3):245–248, 2007.

[31] G. Rauhut and P. Pulay. Transferable scaling factors for density functional derived vibrational force fields. *J. Phys. Chem.*, 99:3093–3100, 1995.

[32] P. Ren and J. Ponder. Polarizable atomic multipole water model for molecular mechanics simulation. *J. Phys. Chem. B*, 107:5933–5947, 2003.

[33] J. Ross, E. Eady, J. Cove, and W. Cunliffe. 16S rRNA mutation associated with tetracycline resistance in a Gram-positive bacterium. *Antimicrob. Agents Chemoth.*, 42:1702–1705, 1998.

[34] W. Saenger, P. Orth, C. Kisker, W. Hillen, and W. Hinrichs. The tetracycline repressor: A paradigm for a biological switch. *Angew. Chem. Int. Ed.*, 39:2042–2052, 2000.

[35] T. Simonson. Free energy calculations. In O. Becker, A.D. Mackerell Jr., B. Roux, and M. Watanabe, editors, *Computational Biochemistry & Biophysics*, chapter 9. Marcel Dekker, N.Y., 2001.

[36] Thomas Simonson, Georgios Archontis, and Martin Karplus. Free energy simulations come of age: the protein–ligand recognition problem. *Acc. Chem. Res.*, 35:430–437, 2002.

[37] P. Sims. The stability constants of some metal chelates of ortho-aminophenols. *J. Chem. Soc.*, NOV:3648–3649, 1959.

[38] C.R. Stephens, K. Murai, K.J. Brunings, and R.B. Woodward. Acidity constants of the tetracycline antibiotics. *J. Am. Chem. Soc.*, 78:4155–4158, 1956.

[39] Roland Stote, David States, and Martin Karplus. On the treatment of electrostatic interactions in biomolecular simulation. *J. Chim. Phys.*, 88:2419–2433, 1991.

[40] A. Streitweiser Jr. and C.H. Heathcock. *Introduction to organic chemistry.* McMillan, New York, 1981.

[41] P. Sum. Case studies in current drug development: glycylcyclines. *Curr. Opin. Chem. Biol.*, 10:374–379, 2006.

[42] Attila Szabo and Neil S. Ostlund. *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory.* Dover Publ., New York, 1996.

[43] J. Tomasi, R. Cammi, B. Mennucci, C. Cappelli, and S. Corni. Molecular properties in solution described with a continuum solvation model. *Phys. Chem. Chem. Phys.*, 4:5697–5712, 2002.

Chapter 3

Tetracycline:Tet Repressor Recognition: Simulations and Experiments

Tc:TetR recognition is a complex problem, with the protein and ligand each having several possible conformations and protonation states, which are difficult to elucidate by experiment alone. We used a combination of free-energy simulations and crystallographic analysis to investigate the electrostatic interactions between protein and ligand and the possible role of induced fit in Tc binding.

Upon Tc binding, TetR switches to its induced conformation, so that The high TetR:DNA affinity is sharply reduced upon Tc binding. Here, we use molecular dynamics simulations and continuum electrostatic calculations to help elucidate the detailed molecular mechanism of TetR induction by Tc.

Article 3: Protonation Patterns in Tetracycline:Tet Repressor Recognition: Simulations and Experiments Alexey Aleksandrov, Juliane Proft, Winfried Hinrichs, and Thomas Simonson. *ChemBioChem,* 2007, 8, 675 685.

Article 4: Tet Repressor Induction by Tetracycline: A Molecular Dynamics, Continuum Electrostatics, and Crystallographic Study Alexey Aleksandrov, Linda Schuldt, Winfried Hinrichs and Thomas Simonson. *J. Mol. Biol.,* 2008, 378, 896910.

DOI: 10.1002/cbic.200600535

Protonation Patterns in Tetracycline:Tet Repressor Recognition: Simulations and Experiments

Alexey Aleksandrov,[a] Juliane Proft,[b] Winfried Hinrichs,*[b] and Thomas Simonson*[a]

Resistance to the antibiotic tetracycline (Tc) is regulated by its binding as a Tc:Mg^{2+} complex to the Tet Repressor protein (TetR). Tc:TetR recognition is a complex problem, with the protein and ligand each having several possible conformations and protonation states, which are difficult to elucidate by experiment alone. We used a combination of free-energy simulations and crystallographic analysis to investigate the electrostatic interactions between protein and ligand and the possible role of induced fit in Tc binding. Tc in solution was described quantum mechanically, while Tc:TetR interactions were described by a recent, high-quality molecular-mechanics model. The orientations of the amide and imidazole groups were determined experimentally by a careful analysis of Debye–Waller factors in alternate crystallographic models. The agreement with experiment for these orientations suggested that the simulations and their more detailed, thermodynamic predictions were reliable. We found that the ligand prefers an extended, zwitterionic state both in solution and in complexation with the protein. Tc is thus preorganized for binding, while the protein combines lock-and-key behavior for regions close to the ligand's amide, enolate, and ammonium groups, with an induced fit for regions close to the Mg^{2+} ion. These insights and the modeling techniques employed should be of interest for engineering improved TetR ligands and improved TetR proteins for gene regulation, as well as for drug design.

Introduction

Tetracycline (Tc) is one of the most important antibiotics in use today,[1–3] its bacteriostatic function being based on inhibition of polypeptide elongation by the bacterial ribosome. Over one thousand Tc varieties have been synthesized, but only about ten have proved useful, as broad-spectrum antibiotics for humans and livestock.

Tc resistance in Gram-negative bacteria is regulated by the Tet Repressor protein (TetR). TetR binds specifically to DNA, repressing both its own expression and that of the cytoplasmic membrane protein TetA, which exports Tc out of the cell. When Tc enters the cell, it binds to TetR, inducing conformational changes that sharply reduce the TetR:DNA affinity.[4] TetR is then released from the DNA and expression of TetA can proceed, conferring resistance. Today, the TetR/Tc system is of general interest because of widespread application in molecular and cell biology as a sensitive switch for target gene regulation.[5,6]

It is thus of great interest to increase our understanding of Tc:TetR binding, but Tc:TetR recognition is a complex problem and several important questions remain open. We focus here on protonation states and side chain conformations in the binding pocket and their possible changes in the protein:ligand binding reaction, a question with general implications for molecular recognition and drug design.

Tcs are large, complex molecules, with two main conformations and two main protonation states at neutral pH; these conformations and/or protonation states may change when the Tc binds to TetR. Similarly, protein groups in the Tc binding pocket have several possible orientations and protonation states, which could also change when Tc binds.[7] While crystal structures of TetR have provided essential information,[4,8,9] and biochemical and thermodynamic data are also available,[10–14] these data do not reveal the protonation states of Tc or the protein side chains, nor the orientations of several key recognition residues.[12] In particular, two orientations are possible for Asn, Gln, and His side chains close to the ligand. Furthermore, the predominant Tc conformation and protonation state in solution are not known, so that its preorganization for binding cannot be assessed. All these unknown features have to do with the electrostatic properties of the ligand and the binding pocket; indeed, protonation of a histidine side chain or transfer of a proton from one Tc group to another do not change the shapes of the two partners, but they have a large impact on the electrostatic field and potential in the binding pocket. A

[a] *A. Aleksandrov, Prof. T. Simonson*
Laboratoire de Biochimie (CNRS UMR7654)
Department of Biology, Ecole Polytechnique
91128 Palaiseau (France)
Fax: (+33) 169-333-013
E-mail: thomas.simonson@polytechnique.fr

[b] *J. Proft, Prof. W. Hinrichs*
Institute for Biochemistry
Department of Molecular Structural Biology, University of Greifswald
Felix-Hausdorff-Strasse 4, 17489 Greifswald (Germany)
E-mail: winfried.hinrichs@uni-greifswald.de

 © 2007 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

change in any of these properties induced by the Tc:TetR binding reaction can thus be viewed as an electrostatic induced fit. In contrast, if the electrostatic state of the partners is maintained, then the partners effectively function as an electrostatic lock and key.

Computer simulations can be used to investigate electrostatic interactions in complex systems such as these.[15,16] Several computational studies of Tc have been reported,[17–21] but work on Tc in interaction with its biological partners has been limited to two very recent studies of the allosteric properties of TetR.[22,23] This is due to the considerable difficulty involved, with a molecule as complex as Tc, in the development of a molecular-mechanics forcefield of sufficient quality for quantitative molecular recognition studies. We recently reported such a force field[24] that is compatible with and of the same quality as current protein forcefields. It has been tested for molecular recognition between Tc and the ribosome, giving good agreement with experiment (A.A. and T.S., unpublished results).

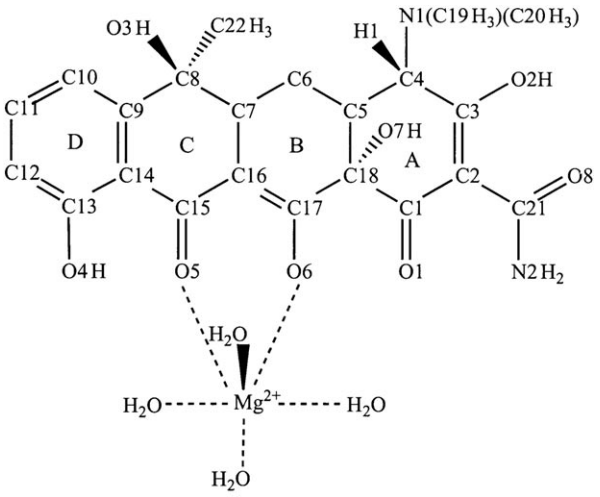
To investigate the protonation patterns and conformational states in the Tc:TetR complex, we have performed both molecular dynamics free-energy simulations and crystallographic calculations. Free-energy simulations have matured significantly in recent years and represent a valuable tool, complementary to experimentation,[25–29] that allows the Tc binding free energy to be computed as a function of its protonation state and the protonation and conformational states of the binding pocket. Two different free-energy methods, using either an explicit treatment of aqueous solvent or a continuum dielectric treatment, have been employed. We have also computed the structure and protonation state of Tc in solution, using a quantum mechanical treatment of Tc and a continuum dielectric treatment of the solvent.[30] By combining these data, we predict the most stable form of the TetR:Tc complex and the extent of protein and ligand preorganization.

To test the simulation results, crystallographic refinements have also been performed, through the use of X-ray diffraction data to a resolution of 1.89 Å. We have constructed two models of the binding pocket, both of which are compatible with the experimentally determined electron density, but which differ in the orientations of His, Asn, and Gln side chains in the binding pocket. Indeed, these side chains can be flipped by 180° and still occupy the same space within the electron density map. Nevertheless, through a systematic refinement of the two models and a careful analysis of the resulting thermal disorder parameters, the correct orientations of these groups can be deduced. If, for example, an Asn side chain is flipped with respect to the correct structure, so that the terminal OD1 and ND2 atoms are interchanged, the refined Debye–Waller factors will reveal the mismatch, with the heavier OD1 having an artificially enlarged Debye–Waller factor. Simulations and experiment have been found to be in good agreement. The structure, fluctuations, and thermodynamics that emerge from the data have then been analyzed to identify the interactions that govern the binding affinity of Tc for TetR. We have found that the protonation states and conformations both of the ligand and of a large portion of the binding pocket do not change upon binding, but are preorganized. In contrast, side chains close to the Mg^{2+} ion are already known, from the available X-ray structures, to reorganize upon binding, as confirmed by the simulations, so the system has features both of an induced fit and of a lock-and-key. These insights and the modeling techniques employed should be of interest for engineering of improved TetR ligands, improved TetR proteins, and for drug design in general.

Results

TetR:Tc binding: free-energy calculations

We considered several modes of Tc binding, involving two Tc protonation states (Tc_N, Tc_Z), three protonation states for nearby His64, and two orientations for the nearby Asn82. All these degrees of freedom are difficult to deduce from the available medium-resolution TetR:Tc crystal structures, because they involve the positions of hydrogens. Views of Tc and the binding pocket are shown in Figure 1, the different states are



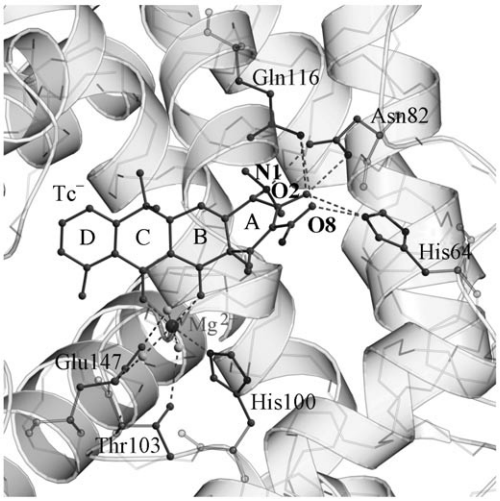


Figure 1. A) The neutral tautomer of Tc (protonated O2, deprotonated N1) in complexation with Mg^{2+} and four water molecules. B) Tc in the TetR binding pocket, with Mg^{2+} coordination and selected groups and hydrogen bonds highlighted.

 © 2007 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

schematized in Figure 2 (the states that were directly compared are connected by arrows), while free-energy results are summarized in Figure 3. We used the rigorous, but expensive, MDFE technique to compare one Tc_N state (model N_3; Figure 2) and two Tc_Z states (models Z_2 and Z_3), and a faster but more approximate method—PBLRA—to compare all three Tc_N and Tc_Z states systematically. Where comparable, the two methods agree very well. We first consider the three TetR complexes with Tc_N, and then the complexes with Tc_Z. Finally, we compare the Tc_N complexes to the Tc_Z complexes, to obtain a complete ranking of all six complexes. Comparison with the crystallographic data is done in the next section.

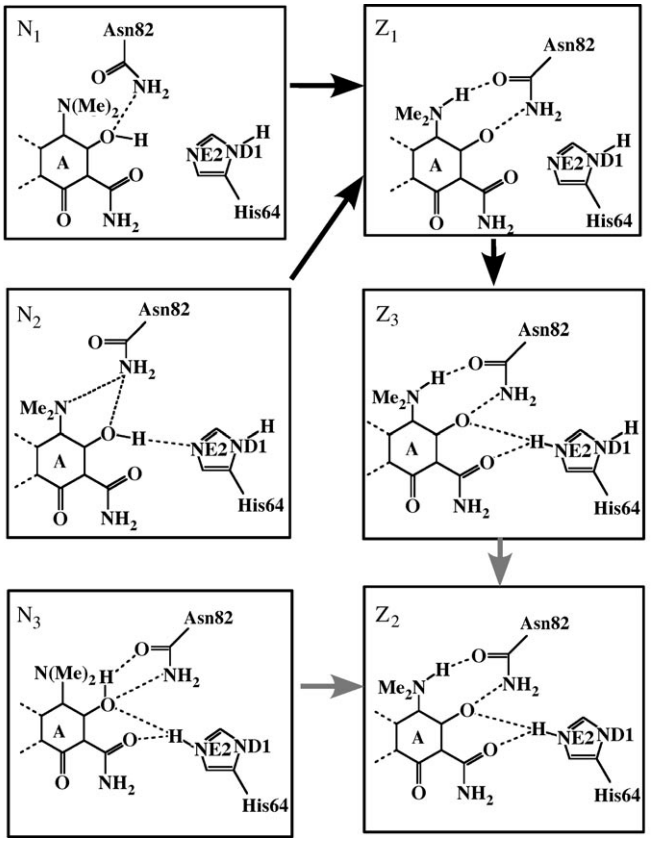


Figure 2. Six different Tc:TetR models. The models involve neutral and zwitterionic forms of Tc, and different protonation and rotation states of His64 and Asn82. Arrows indicate direct connections between states by PBLRA (and MDFE in some cases; gray arrows).

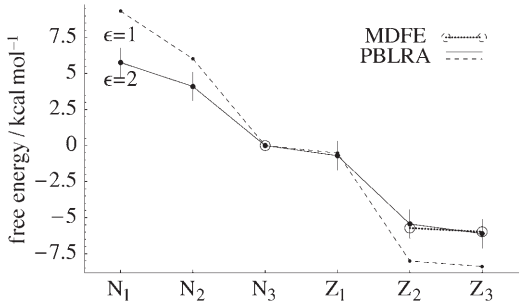


Figure 3. Relative free energies of all six states. PBLRA is performed with a solute dielectric ϵ of one or two, as indicated. States connected by MDFE are circled: $N_3 \rightarrow Z_2$ and $Z_2 \rightarrow Z_3$. Connections between the N and Z states also make use of the quantum mechanical solution data (see text, Table 3). PBLRA error bars are shown for the $\epsilon = 2$ data.

Comparison among TetR:Tc_N complexes

We considered TetR:Tc_N complexes with a neutral His64 in two protonation states, termed HSD and HSE (proton on ND1 and NE2, respectively), and with two Asn82 orientations (Figure 2). Complexes with charged His64 were not considered, because double protonation of His64 was found not to increase significantly the stabilities of either apo-TetR or TetR:Tc_Z (see below). The three TetR:Tc_N complexes considered were compared by the approximate (PBLRA) method. For technical reasons, the Tc_N states were compared indirectly, by using Tc_Z states as intermediates. For example, N_1 and N_2—differing in the orientation of Asn82—were both compared directly to Z_1 (see connections in Figure 2). N_3 is the most stable, and N_1 the least stable Tc_N state; indeed, flipping the side chain so that it donates a hydrogen bond to the Tc dimethylammonium lowers the free energy by about 2–3 kcal mol^{-1} ($\Delta G(N_1 \rightarrow N_2)$; Table 1). The span of values corresponds to a range of 1–2 for the dielectric constant of the protein; this is the physically reasonable range with the PBLRA method.[29,31,32] Statistical error bars, estimated by pooling the MD data into two batches, are about 1 kcal mol^{-1} (Table 2). This moderate statistical uncertainty is due partly to the long simulations and partly to the tight organization and moderate fluctuations of the TetR binding pocket. Note also that the PBLRA method makes use of the interaction energy between the ligand and its environment, which has intrinsically smaller fluctuations than the total energy (as used in, for example, the well known MM/PBSA method[29,33]).

Table 1. Free-energy simulations comparing Tc_N^-:Mg^{2+} and Tc_Z^-:Mg^{2+} binding to TetR.

Initial state	Final state	Method	ΔG_{prot}[a]	ΔG_{sol}[a]	$\Delta\Delta G$
N_3:Tc*N*	Z_2:Tc*Z*	MDFE	−32.78 (0.28)	−30.06 (0.20)	−2.73 (0.34)
N_3:Tc*N*	Z_2:Tc*Z*	PBLRA (ϵ=1/ϵ=2)	22.6/11.4	27.6/13.8	−5.0/−2.5

Energies are in kcal mol^{-1}. [a] Columns 4–5 are the free energies to go reversibly from the initial ligand (Tc_N^-:Mg^{2+}) to the final one (Tc_Z^-:Mg^{2+}), either in the protein (ΔG_{prot}) or in solution (ΔG_{sol}). $\Delta\Delta G = \Delta G_{prot} - \Delta G_{sol}$. A negative $\Delta\Delta G$ means the zwitterionic form binds more strongly to TetR. For the MDFE data, error bars are given in parentheses (twice the standard deviation over two runs, one in each direction). For PBLRA, results with a solute dielectric of one and two are given before and after the back-slash.

Comparison among TetR:Tc_Z complexes

We considered TetR:Tc_Z complexes with three His64 protonation states: HSD (state Z_1), HSE (state Z_2), and the doubly protonated form HSP (state Z_3). States Z_2 and Z_3 were compared

 © 2007 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

Table 2. Free-energy simulations comparing different protonation states of His64.

Initial state	Final state	method	dielectric	Protein ΔG_{prot}[a]	ΔG_{sol}[a]	$\Delta\Delta G$
Z1:HSD	Z_3:HSP	PBLRA	1.0	64.4	72.6	−8.2
Z1:HSD	Z_3:HSP	PBLRA	2.0	30.6	36.4	−5.8
Z_2:HSE	Z_3:HSP	PBLRA	1.0	66.1	66.9	−0.8
Z_2:HSE	Z_3:HSP	PBLRA	2.0	32.5	33.5	−1.0
Z_2:HSE	Z_3:HSP	MDFE		−1.10	−0.16	−0.94
apoTetR:HSD	apoTetR:HSP	PBLRA	1.0	62.2	72.2	−10.0
apoTetR:HSD	apoTetR:HSP	PBLRA	2.0	30.5	35.5	−5.0
apoTetR:HSE	apoTetR:HSP	PBLRA	1.0	64.4	64.7	−0.3
apoTetR:HSE	apoTetR:HSP	PBLRA	2.0	31.5	32.5	−1.0

Energies are in kcal mol^{-1}. [a] Columns 5–6 are the free energies to go reversibly from the initial protonation state (HSD or HSE) to the final one (HSP), either in the protein (ΔG_{prot}) or in the reference methylimidazole in solution (ΔG_{sol}). $\Delta\Delta G = \Delta G_{prot} - \Delta G_{sol}$. A negative $\Delta\Delta G$ means the doubly protonated His64 is preferred in the protein, relative to the reference: the His64 pK_a is up-shifted relative to methylimidazole. apoTetR is the TetR protein without Tc.

by use both of PBLRA and of the more rigorous MDFE method, both of which follow the horizontal legs of the thermodynamic cycle shown in Figure 4 A. The protonation state is reversibly

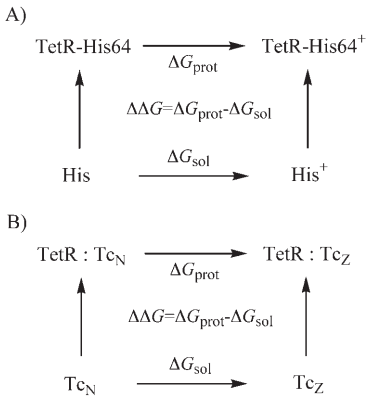


Figure 4. Thermodynamic cycles. A) His64 protonation. B) Tc^-:Mg^{2+} binding. Vertical legs represent Tc^-:Mg^{2+} binding/dissociation to TetR; horizontal legs represent the alchemical transformation of Tc_N^- into Tc_Z^- in solution (below) or in TetR:Tc^-:Mg^{2+} complex (above).

changed by gradually changing the atomic charges on the His64 side chain. The upper leg makes the change in the protein; the lower leg corresponds to a side chain analogue (methylimidazole in solution) that serves as a reference molecule. Subtraction yields the double free-energy difference $\Delta\Delta G$(HSE→HSP). The proton binding constant of His64 is related to $\Delta\Delta G$ by the relationship:

$$pK_{a,prot} = pK_{a,model} + \frac{1}{2.303\,kT}\Delta\Delta G(\mathrm{HSE} \rightarrow \mathrm{HSP}) \qquad (1)$$

Here, p$K_{a,prot}$ and p$K_{a,model}$ are the pK_a values of the titrating group in the protein and in the reference molecule, respectively. Results are summarized in Table 2. Combining the experimentally measured value for methylimidazole (p$K_{a,model}$ = 6.5) with the MDFE result yields $\Delta G(Z_2 \rightarrow Z_3) = -0.3$ kcal mol^{-1}. This number arises from two contributions: the increased stability of the imidazole charge in the protein context, relative to in solution ($\Delta\Delta G = -0.94$ kcal mol^{-1}, Table 2), and the additional cost required to protonate imidazole at a pH of 7 (the pH of interest here), slightly above its solution pK_a.

With a protein dielectric of 2, PBLRA yields −0.3 kcal mol^{-1}, while with a dielectric of 1 it yields −0.1 kcal mol^{-1}. PBLRA thus agrees very well with MDFE, and the doubly protonated His64 state appears to be very slightly favored, if it is borne in mind that the $\Delta G(Z_2 \rightarrow Z_3)$ values are within the uncertainty of both free-energy methods. The Z_1 state has the single His64 proton on the less favorable HSD side, losing the hydrogen bond to Tc_Z. This raises the free energy considerably, by 5–8 kcal mol^{-1} (Table 2).

We also studied the His64 protonation state in the absence of Tc (Table 2). Results are very similar, with the doubly protonated state almost equistable with HSE, and with HSD 5–10 kcal mol^{-1} higher. Tc binding thus does not affect the protonation state: His64 is preorganized for Tc binding.

Comparison between Tc_N and Tc_Z in solution

To compare the stabilities of Tc_N and Tc_Z complexes (below), we must first compare the two tautomers in solution. Tc has been modeled in solution by use of a quantum mechanical description of the solute and a continuum dielectric description of the solvent.[30] We considered two Tc forms: Tc_N^-:Mg^{2+} and Tc_Z^-:Mg^{2+}. Four water molecules that chelate the cation are treated as part of the solute; these and the Mg^{2+} cation are included in the quantum mechanical treatment. For each form, we considered several extended and twisted conformations, and the energies corresponding to the most stable structures are shown in Table 3. The extended, Tc_Z^-:Mg^{2+} complex is seen to be the more stable in solution: by about 1.5 kcal mol^{-1}, which is comparable to the uncertainty of the computational method. The preference for a tautomer that is both extended and zwitterionic is consistent with the structures found in the Cambridge Crystallographic Databank (CCDB): all the zwitterionic Tc structures are extended, whereas all the neutral structures are twisted.[24] This free-energy difference arises from two competing effects: Tc_Z is solvated much more strongly than Tc_N (as would be expected from its greater charge separation), with a solvent contribution 9 kcal mol^{-1} larger than for Tc_N (fourth column in Table 3); this is partly offset by a 6 kcal mol^{-1} penalty in the solute intramolecular energy. The lowest extended Tc_N has a relative free energy of about 3 kcal mol^{-1}, 1.5 kcal mol^{-1} higher than the twisted conformer.

 © 2007 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

Table 3. Energies of Tc^{-}:Mg^{2+} complexes in solution (kcal mol^{-1}).

Tc tautomer	Conformation[a]	Solute energy[b]	Solvent contribution[c]	Relative energy
zwitterionic	extended	0.00	−84.26	**0.00**[d]
zwitterionic	extended	8.66	−92.03	0.89
neutral	twisted	−8.28	−74.46	1.53
zwitterionic	extended	6.15	−87.63	2.82
neutral	extended	−5.90	−75.39	**2.99**[d]
zwitterionic	twisted	−6.46	−73.39	4.40
neutral	extended	−1.19	−77.14	5.94
neutral	twisted	3.08	−79.85	7.50

Energies are given for the four lowest structures for each tautomer. [a] Refers to the two main classes of Tc conformations. [b] Calculated by the B3LYP/6-31G* density functional method. [c] Contribution from dielectric continuum solvent. [d] The most stable extended structures used to define the solution free energies of Tc_N and Tc_Z in the thermodynamic cycle (Figure 2B).

Comparison between TetR:Tc_N and TetR:Tc_Z complexes

To compare the Tc_N and Tc_Z complexes, we linked the N_3 and Z_2 states by both MDFE and PBLRA; results are summarized in Table 1. Both methods follow the horizontal legs of the thermodynamic cycle in Figure 4B: the Tc_N^-:Mg^{2+} moiety is alchemically transformed into Tc_Z^-:Mg^{2+}, both in solution and in complexation with the protein. By comparing the two legs, we obtain a binding free-energy difference, $\Delta\Delta G_{bind}(Tc_N \rightarrow Tc_Z)$. The overall difference in stability between the N_3 and Z_2 states can be written:

$$\Delta G(N_3 \rightarrow Z_2) = \Delta\Delta G_{bind}(Tc_N \rightarrow Tc_Z) + \Delta G_{sol}(Tc_N \rightarrow Tc_Z) \qquad (2)$$

The rightmost quantity, $\Delta G_{sol}(Tc_N \rightarrow Tc_Z)$, corresponds to the stability difference in solution, which was estimated quantum mechanically (above) to be -3 kcal mol^{-1} (we consider here the extended N and Z tautomers). MDFE gives $\Delta\Delta G_{bind}(Tc_N \rightarrow Tc_Z) = -2.7 \pm 0.3$ kcal mol^{-1}, with a very good consistency between the two MDFE runs performed. The PBLRA results with protein dielectric values of one (-5.0 kcal mol^{-1}) and two (-2.5 kcal mol^{-1}) bracket the MDFE result and show that a protein dielectric of two is more realistic. A dielectric of two can be interpreted as an effect of electronic polarizability of the protein. Including the quantum mechanical data for Tc in solution, $\Delta G_{sol}(Tc_N \rightarrow Tc_Z) = -3.0$ kcal mol^{-1}, we find that Z_2 is 5.7 kcal mol^{-1} lower in free energy than N_3.

We are now in a position to rank all six states considered (see Figure 3). Error bars in the Figure (± 1 kcal mol^{-1}) correspond to the average statistical uncertainty for all the PBLRA runs. The solution quantum calculations add an additional uncertainty for the positions of all the N states relative to all the Z states. This uncertainty is harder to quantify, but is unlikely to be more than 2–3 kcal mol^{-1}, so that it should not affect the overall ranking of Z_2 and Z_3. Z_2 and Z_3 are the lowest states and almost equistable, with a very small preference for the doubly protonated His64 (Z_3). The increased stabilities of these states are the result of three main factors: a strong preference for the Tc_Z tautomer in solution, improved interactions of Asn82 with the Tc dimethylammonium group (compare N_2 to N_3), and improved interactions of His64 with the Tc amide carbonyl and Tc-O2 (compare Z_1 to Z_2 and Z_3). The dominant Tc state, the extended Tc_Z, is the same in the protein and in solution: the Tc ligand is preorganized for TetR binding.

TetR:Tc:Mg^{2+} binding—crystallographic refinement

The initially refined crystallographic model (model 1) was manually modified to create a second, alternative model (model 2). The amide side chains of Asn82, Gln116, the amide group of doxyTc, and the imidazole ring of His64 were all flipped by 180°. The use of the doxyTc complex instead of a Tc complex is not a concern, because the known atomic positions agree very closely for these two complexes. Asn5 and His100 were also flipped, as controls. Indeed, for these two side chains, the correct orientation is obvious from the hydrogen bonding patterns in the structure (e.g., His100 coordinates the Mg^{2+} cation). These two alternative models were refined to convergence, and the correct conformations of each group were then identified by comparison of the individual, isotropic, Debye–Waller factors of the corresponding amide and imidazole atoms; see Table 4. The Debye–Waller factor measures the thermal disorder of a given atom; a large value indicates a mobile atom that is less effective at scattering X-rays. For a correctly oriented Asn or Gln side chain, the mobilities of the ND1 and OD2 atoms would be expected to be similar, due to their similar environments. Wrong conformations are therefore identified unambiguously by unusual discrepancies of the Debye–Waller factors between alternative O/N- or C/N-positions. Consider, for example, the Asn5 side chain in model 2, in which the discrepancy is large. Evidently, the terminal group is

Table 4. Debye–Waller factors for selected atoms [Å^2].

His64		Asn82		His100		Gln116		doxyTc		Asn5	
CA	29.90/29.89	CA	25.37/25.46	CA	27.00/27.08	CA	31.74/31.89	C2	25.18/24.87	CA	39.43/39.60
CB	27.89/27.91	CB	24.04/24.16	CB	26.71/26.71	CB	30.28/30.51	C21	27.22/26.87	CB	39.47/39.64
CG	28.48/28.30	CG	30.57/30.91	CG	27.02/26.65	CG	29.35/29.59	O21	26.61/31.21	CG	40.14/40.41
ND1	28.31/27.97	OD1	28.49/30.99	ND1	26.08/30.61	CD	24.33/24.77	N21	27.35/21.89	OD1	37.35/45.31
CE1	25.38/24.89	ND2	26.39/24.12	CE1	28.26/22.53	OE1	24.40/28.89			ND2	41.72/33.73
NE2	25.45/29.41			NE2	23.92/28.63	NE2	24.21/20.03				
CD2	25.89/22.22			CD2	24.95/20.07						

Values before and after the oblique are for models 1 and 2, respectively.

 © 2007 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

incorrectly flipped by 180°, exchanging the O and N positions, and in the structure refinement there is too much scattering power in the oxygen position and too little in the nitrogen position. To compensate, the refinement gives final Debye–Waller factors that are too high for the (incorrectly placed) oxygen and too low for the (incorrectly placed) nitrogen. The mismatch between the two Debye–Waller factors (Table 4, model 2) identifies this Asn5 orientation as incorrect. In contrast, model 1 has a much closer match between the N and O Debye–Waller factors, indicating that the Asn5 orientation is correct. Refinement statistics for other parameters such as coordinates of individual atoms are very similar in the two refined models (not shown). For all the groups considered—Asn5, His64, Asn82, His100, Gln116, and the doxyTc amide—there is a clear Debye–Waller factor mismatch in model 2, so that the correct orientation can be clearly identified as that of model 1. In all cases, a reasonable hydrogen bonding network is maintained in model 1.

Structure and dynamics of TetR and the TetR:Tc:Mg^{2+} complex

The MD and X-ray structures are compared in Figure 5A and Tables 5 and 6. All the MD models give low rms coordinate deviations from experiment for atoms within the Tc binding pocket, between 0.7 and 1.1 Å (Table 5). Model Z_2 gives the second lowest deviation (0.8 Å), while Z_3 gives the highest (1.1 Å). The protein–ligand interatomic distances (Table 6) are reproduced well both by Z_2 and by Z_3, with Z_3 giving a slightly better agreement. Figure 5A shows that the Z_2 and Z_3 structures superimpose almost perfectly with the X-ray structure, with the six groups considered in the X-ray refinement all having the correct orientation in both Z_2 and Z_3.

In models Z_1 and Z_2, His64 is neutral, singly protonated either on ND1 (Z_1) or on NE2 (Z_2), while His64 in Z_3 is doubly protonated. When a single proton is on ND1 (Z_1) the interaction with Tc is lost; the side chain shifts, and two water molecules move from the bulk into the initial His64 position to replace its interactions with Tc. One interacts with Tc-O8 and the other with Tc-O2. This disagrees with the experimentally determined structures, in which the Tc:His64 interaction is maintained. In models Z_2 and Z_3, the positions of Asn82, Gln116, and Tc are practically the same, while the Tc:His64 interactions differ slightly. In model Z_2, His64 interacts mainly with Tc-O2 (equilibrium distance of 2.9 Å). In Z_3, it shares its interaction between Tc-O2 and Tc-O8, in better agreement with the present X-ray structure, and with five out of six others listed in Table 6 (all but 1BJ0). In Z_3, the other His64 proton (on ND1) interacts with the Ser85 carbonyl oxygen, behind the binding pocket.

In all six MD models the Mg^{2+} ion coordinates three water molecules and interacts with His100. Two of these waters also interact with residue Glu147 from the other protomer (within the TetR dimer), while the third water interacts with Thr103. The same arrangement is found in the experimentally ascertained structure. Other than the waters coordinating Mg^{2+}, Tc interacts directly with just one other water molecule: through Tc-O1 and Tc-O6. This water is also seen in the X-ray structure, in the same position, suggesting that it is important for the structure.

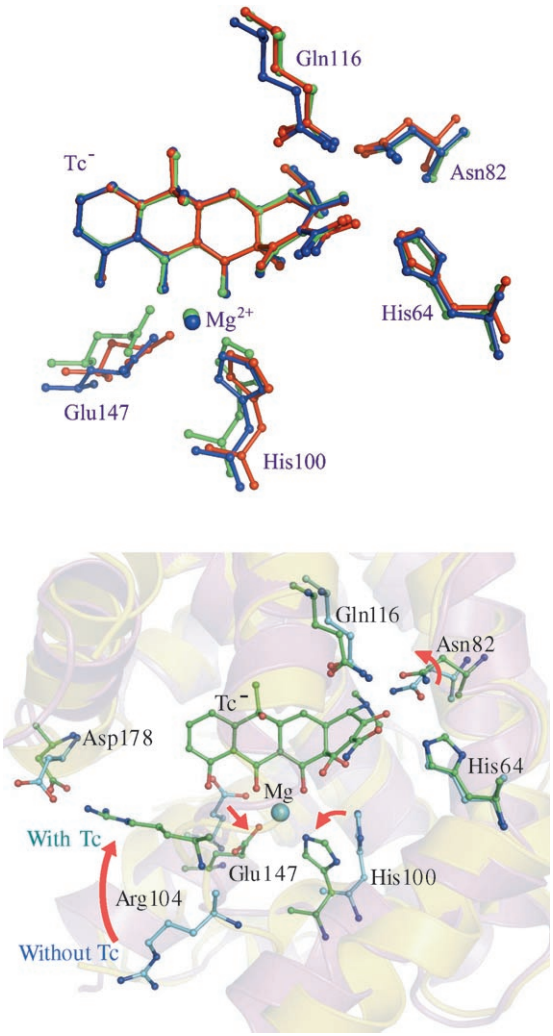


Figure 5. A) MD models vs. X-ray structure. X-ray: red; Z_2: green; Z_3: blue; except for the acetamide groups of Tc, Gln116, and Asn82, where N is blue and O is red (to compare their orientations). B) TetR with and without Tc. MD models in the presence (green) and absence (blue) of Tc. Conformational transitions shown as red arrows. Asp178 and Glu147 are from the other protomer in the TetR dimer.

Table 5. Rms deviations from X-ray structure [Å] for MD models.

Model	Backbone	10 Å sphere
N_1	0.76	1.01
N_2	0.53	1.02
N_3	0.95	1.13
Z_1	0.68	0.89
Z_2	0.68	1.03
Z_3	0.75	1.05

Deviation with respect to the crystal structure for protein backbone and for all non-hydrogen atoms within 10 Å of Tc (including Tc itself).

Finally, Figure 5B compares the MD structures in the presence and absence of Tc. There are few notable differences upon ligand binding, which agree with the known crystal

 © 2007 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

Table 6. Selected distances between atoms in MD and X-ray models.

	MD models			Crystal structures					
	Z_2	Z_3	this work	2TRT	1BJ0	1BJY[b]	2TCT	1ORK	1DU7
CB_{Val113} $O3_{Tc}$	3.5	3.5	–	3.0	3.8	3.1/3.2	3.5	–	4.1
NE_{Arg104} $O4_{Tc}$	5.1	5.1	–	3.2	7.4	4.5/3.5	3.3	5.2	3.9
Phe86[a] $O7_{Tc}$	3.7	3.6	–	3.8	3.9	3.6/3.7	3.8	3.8	3.3
$NE2_{Gln116}$ $O2_{Tc}$	3.2	3.3	3.3	3.4	3.0	3.3/3.4	3.3	3.4	3.4
$NE2_{Gln116}$ $O8_{Tc}$	3.1	3.0	3.1	3.4	3.2	3.0/3.2	3.2	3.3	3.4
$NE2_{His64}$ $O2_{Tc}$	2.9	3.1	2.8	2.7	3.3	2.7/2.7	2.9	2.8	3.7
$NE2_{Hsd64}$ $O8_{Tc}$	3.4	2.8	2.9	2.6	3.6	2.9/2.8	2.7	2.7	2.9
$ND2_{Asn82}$ $O2_{Tc}$	2.9	2.9	2.8	2.9	3.1	2.8/2.9	2.8	2.8	2.9
$OD1_{Asn82}$ $N1_{Tc}$	2.8	2.8	2.6	2.7	2.8	2.8/2.7	2.7	2.7	3.0

[a] Distance was computed between oxygen Tc-O7 and the center of the Phe86 ring. [b] There are two Tcs in the structure.

structures.[34] His100 and Glu147 shift to coordinate Mg^{2+} when Tc binds, and Arg104 moves closer to the ligand and forms a salt bridge with Asp178 from the other protomer. Its side chain surface is two-thirds exposed without Tc and two-thirds buried with Tc. The protonation state of His64 does not change upon Tc binding, from the PBLRA results above. From the simulations, Asn82 flips its terminal $(CO)NH_2$ by 180° to interact with His64 in the absence of Tc, but occupies approximately the same space. The His64 positions with and without Tc superimpose almost exactly. The electrostatic complementarity between the binding pocket and the Tc ligand is illustrated in Figure 6.

Free-energy components and Tc:TetR complementarity

From the structural and thermodynamic data, we can quantify the contributions of several individual groups to Tc binding, in the spirit of a hydrogen bond inventory model or an alanine-scanning experiment. In particular, the Poisson model has a useful property: the TetR:Tc binding free energy can be broken down into several distinct contributions,[29,35] including contributions from individual amino acids in the Tc binding pocket. We refer to them as free-energy components. We emphasize that the Poisson components, unlike free components calculated with MDFE, do not depend on a "pathway" connecting the bound and the unbound states, but only on the endpoints of the binding reaction.[29,35] The components are estimated from the electrostatic interactions between a particular amino acid and the Tc ligand (with its co-bound Mg^{2+}). The estimation is made by use of the native TetR:Tc structure. With the Poisson continuum model, structural changes associated with deletion of an amino acid side chain are modeled implicitly, through the solute dielectric constant ($\in = 4$; see Experimental Section). This should be reasonably accurate only for mutations that do not produce large structural changes. Figure 7 shows the contributions of individual amino acids to the Poisson binding free energy of Tc in the Z_3 state. In a few cases, experimentally measured binding free energies for TetR point mutants are

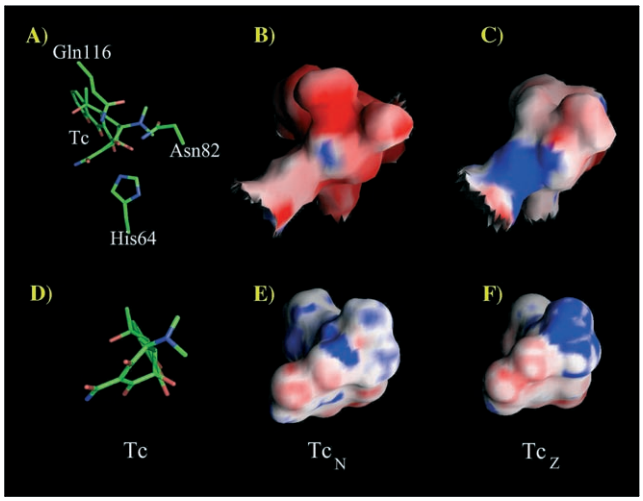


Figure 6. Electrostatic potential on the surfaces of Tc and the binding pocket. A) Tc with selected recognition residues, viewed from the back of the pocket. B)–C) The binding pocket surface, viewed from the back. Red areas have negative potential; blue areas are positive. His64 is neutral (B) or positive (C). D) Tc alone, same viewpoint. E) Tc_N surface. F) Tc_Z surface. The best complementarity is between (F) and (B). Produced with GRASP[64] and Pymol.[65]

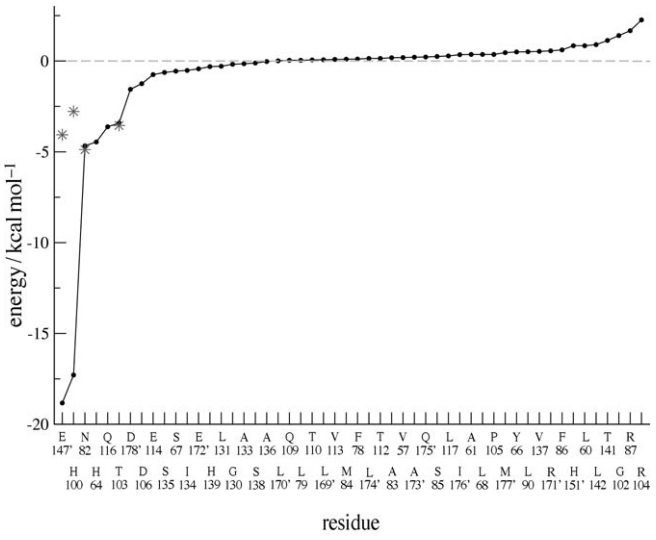


Figure 7. Binding free-energy components. Line/dots: contributions of individual amino acids to the Tc binding free energy, computed with the Poisson model for state Z_3. Stars: values from alanine-scanning experiments.[39]

 © 2007 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

given, where a side chain has been replaced by alanine.[36] The free-energy change upon alanine replacement can be interpreted as an experimental measure of the original side chain's contribution to the Tc binding free energy, so it is of interest to compare the Poisson free-energy components to these values.

The largest Poisson contributions come from Glu147 and His100, which coordinate the Mg^{2+} ion. We saw that these groups undergo a large conformational change when Tc binds (Figure 5 B). In fact, when these amino acids are mutated, Tc binds without any co-bound Mg^{2+},[36] so it is not surprising that the Poisson components for these amino acids are very different from the experimentally determined values (Figure 7). In contrast, the Poisson values for Asn82 (-4.7 kcal mol^{-1}) and Thr103 (-3.4 kcal mol^{-1}) match the experimental results perfectly, which gives confidence that the Poisson model yields good approximate measures of the (electrostatic) contributions of most individual amino acids to binding. Other groups representing significant components are His64 (-4.5 kcal mol^{-1}) and Gln116 (-3.6 kcal mol^{-1}). Arg104 also represents a sizeable positive component (2.3 kcal mol^{-1} opposing binding), but this estimate is not very reliable because Arg104, like Glu147 and His100, undergoes a large conformational change upon Tc binding.

Insights into the roles of His64 and Asn82 can also be obtained from the free-energy differences between the states N_3, Z_1, Z_2, and Z_3. Consider first His64. States Z_2 and Z_3 are almost equistable; this indicates that His64 does not stabilize Tc through its positive charge, but rather through its hydrogen bonds to Tc-O2 and Tc-O8. These provide about -5 kcal mol^{-1} (the free-energy difference between Z_2 and Z_1 and between Z_3 and Z_1). This is close to the His64 free-energy component of -4.5 kcal mol^{-1} obtained from the Poisson analysis above. A similar estimate is obtained by comparison of states N_2 and N_3 (Figure 2). Consider next Asn82. States N_3 and Z_2 differ by -2.7 kcal mol^{-1} (MDFE result, Table 1). This is over half the Asn82 Poisson free-energy component (-4.7 kcal mol^{-1}), indicating that when Tc is switched to its Z state, strengthened interactions with Asn82 account for most of the affinity gain. This is expected from the hydrogen bonds seen in the X-ray structure, but the simulation data give a more quantitative picture. We note that Asn82 is flipped in the apoprotein, suggesting that its interaction with His64 is slightly repulsive in the Z_2 and Z_3 states.

Conclusions

A great deal is known about Tc:TetR binding from X-ray crystallography and biochemistry. It is noteworthy that for the six groups whose orientations have been determined here by experimental Debye–Waller factor analysis, the original crystallographic model (model 1) turned out to be correct. Nevertheless, neither these orientations nor details of the binding-pocket protonation state could be inferred with complete confidence before this work. A program used to position hydrogens automatically, WHATCHECK,[37] protonated His64 incorrectly, probably because it is not able to take into account His64's interaction with the Tc ligand. The range of free energies spanned by the six states shown in Figure 3 illustrates the difficulty. By performing extensive simulations and X-ray structure refinements, a comparison between experiment and theory could be made for the orientations of the six active-site groups. The excellent agreement observed gives confidence that other properties are correctly predicted by the simulations. In particular, we find that the zwitterionic Tc tautomer is strongly preferred, and that the nearby His64 occupies a mixture of singly and doubly protonated states. Our prediction of the His64 protonation state could be tested by pK_a measurements in the presence and absence of Tc. Our prediction of a zwitterionic Tc is consistent with the tetracycline structures in the Cambridge Crystallographic Databank. Indeed, Tc in TetR complexes always has an extended conformation, and all the extended Tcs in the CCDB are zwitterionic, whereas all the neutral Tcs in the CCDB have twisted conformations.

The question arises of whether the Debye–Waller analysis used here can be generally applied to orient groups in protein structure refinements, and whether complementary simulations are generally useful. We cannot make a definite statement, but the agreement found between the X-ray refinement and simulations is encouraging and does suggest that, for crystallographic datasets with a comparable resolution, a careful Debye–Waller analysis should at least be explored.

Overall, TetR:Tc binding is a complex process, involving competition between several effects, and with electrostatics playing an important role. These effects include superior solvation of Tc_Z relative to Tc_N, a strengthened interaction with Asn82 for the *Z* tautomer, and the balance of interactions of His64 with Tc, Asn82, and solvent. Interestingly, while His64 interactions contribute several kcal mol^{-1} to Tc binding, its extra positive charge in the Z_3 state has almost no effect. Overall, the TetR binding pocket combines electrostatic preorganization in the region of the Tc amide and ammonium groups with an induced fit in the region that coordinates the Mg^{2+} ion.

Ligand binding and acid/base equilibria are important for most aspects of protein function, so the simulation techniques used here are of general interest. Free-energy simulations have been extensively applied to these problems in the past, but, after an initial burst of popularity in the early 1990s,[38] technical and fundamental difficulties were brought to light that led to a drop in their use. In recent years, substantial progress has been made,[26–28] as illustrated by the good agreement with experiment seen here. Two substantially different free-energy methods have been compared with each other and with experimental data, thus providing support both for the methodology and for the newly developed tetracycline forcefield.[24] More recently, we have also obtained very good agreement with experimental results[39] for binding affinities of a series of ten tetracycline analogues for TetR (A.A. and T.S., unpublished results). We expect that with the current level of simulation maturity and available computer power, joint experimental and theoretical studies will increasingly help to reveal the complexity of protein–ligand recognition.

 © 2007 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

Experimental Section

Tc in solution—computed structures and free energies

Structures: To generate models of Tc in solution and in complexation with the protein, we proceeded as described previously.[24] Briefly, we retrieved 20 crystal structures of tetracyclines from the Cambridge Crystallographic Data Bank (CCDB).[40] The structures can be grouped into two families, corresponding to "extended" and "twisted" conformations. Starting from selected CCDB structures, we performed molecular dynamics simulated annealing in vacuum with use of the MM3 forcefield.[41] and the Tinker program,[42] in order to generate a wide variety of Tc conformations. After clustering by use of the ART-2 algorithm[43] and the CHARMM program, a few representative structures were optimized quantum mechanically at the B3LYP/6–31G(d) level.[44–46] With this procedure, we generated models for both Tc_N^- and Tc_Z^- bound to a Mg^{2+} ion (Figure 1).

Solution free energies—ab initio calculations: We used quantum mechanical density functional methods to compute the free-energy difference between Tc_N^-:Mg^{2+} and Tc_Z^-:Mg^{2+} in solution. We used the B3LYP hybrid functional[44,45] and the 6–31G* basis set.[46] Geometries were fully optimized by use of the Gaussian03 program.[47] Single-point energy calculations on the optimized structures were performed with use of a dielectric continuum representation of aqueous solvent, with the aid of the polarizable continuum model (PCM).[30]

TetR:Tc binding—computed structures and free energies

Molecular dynamics simulations: The crystal structure of the class D Tet Repressor dimer (TetR) was taken from the Protein Data Bank (PDB), entries 2TRT (with bound Tc) and 1A6I (without bound Tc).[8,34] The simulations included protein residues within a 24 Å sphere, centered on the Tc binding site. In addition to crystal waters, a 24 Å sphere of water was overlaid and waters that overlapped protein, crystal waters, Tc (when present), or Mg^{2+} were removed. Protein atoms between 20 Å and 24 Å from the sphere's center were harmonically restrained to their experimentally determined positions. Simulations were performed with the aid of the SSBP solvent model,[48,49] which treats the region outside the 24 Å sphere as a uniform dielectric continuum, with a dielectric constant of 80. This is reasonable, since most of the outer region is water. Newtonian dynamics were used for the innermost region, within 20 Å of the sphere's center; Langevin dynamics were used for the outer part of the sphere, with a 292 K bath. The CHARMM22 forcefield was used for the protein[50] and the TIP3P model for water.[51] Tetracycline was described with the forcefield developed previously.[24] Electrostatic interactions were computed without any cutoff, with use of a multipole approximation for distant groups.[52] Calculations were done with the CHARMM program.[53]

Alchemical MD free-energy simulations (MDFE): To compare Tc_N^- and Tc_Z^- binding, we used the thermodynamic cycle shown in Figure 4. The MDFE method follows the horizontal legs of the cycle. Tc_N^- is reversibly transformed into Tc_Z^- during a series of MD simulations. For the lower leg, we simulated Tc:Mg^{2+} in the center of a 24 Å radius water sphere, surrounded by a dielectric continuum representing bulk solvent.[48] For the upper leg, we simulated a portion of the Tc:Mg^{2+}:TetR complex, solvated by the same 24 Å water sphere and embedded in the same dielectric continuum. Portions of TetR outside the 24 Å sphere would be expected to contribute very little to the binding free-energy difference.[49] In each simulation system, the energy function can be expressed as a linear combination of Tc_N^-:Mg^{2+} and Tc_Z^-:Mg^{2+} terms [Eq. (3)]:

$$U(\lambda) = U_0 + (1-\lambda)U(\mathrm{Tc_N}) + \lambda U(\mathrm{Tc_Z})) \quad (3)$$

where λ is a "coupling parameter" and U_0 represents interactions between parts of the system other than Tc. The free-energy derivative with respect to λ has the form [Eq. (4)]:

$$\frac{\partial G}{\partial \lambda} = \langle U(\mathrm{Tc_Z}) - U(\mathrm{Tc_N}) \rangle_\lambda \quad (4)$$

Here the brackets indicate an average over an MD trajectory with the energy function $U(\lambda)$.[26,54] We gradually mutated Tc_N into Tc_Z by changing λ from one to zero. In practice, the transformation between Tc_N^-:Mg^{2+} and Tc_Z^-:Mg^{2+} takes the form of a rearrangement of atomic charges on a few Tc atoms, because the displaced hydrogen does not enter into van der Waals interactions in this forcefield. The successive values of λ were: 0.999, 0.99, 0.95, 0.9, 0.8, 0.6, 0.4, 0.2, 0.1, 0.05, 0.01, and 0.001. The free-energy derivatives were computed at each λ value from a 100 ps MD simulation; the last 80 ps were used for averaging. A complete mutation run thus corresponded to 1.2 ns. One run was performed in each direction (Tc_N to Tc_Z and the reverse). The free-energy change was computed by numerical integration of the derivatives. We also used the MDFE method to estimate the pK_a values of His64 of TetR, which is part of the Tc binding site. In this case, the solution leg transforms methylimidazole into its doubly protonated state. Both the protein and the solution transformations were modeled as rearrangements of atomic charges on the imidazole group; see refs. [31,55] for details. A single run was performed for each leg, lasting for 4.8 ns (protein leg) or 2.4 ns (solution leg).

Poisson–Boltzmann Linear Response Approximation (PBLRA): A recently developed method that combines MD simulations with continuum electrostatics is the Poisson–Boltzmann Linear Response Approximation (PBLRA) method.[31,56,57] This method is well suited to treatment of rearrangements of atomic charges, as in the Tc_N to Tc_Z transformation or His64 protonation. The free-energy change ($\Delta G_{pro} - \Delta G_{sol}$ in Figure 4) is approximated by the continuum electrostatic free energy, averaged over the equilibrium states before and after the rearrangement. We applied the method with MD structures of the Tc_N^-:Mg^{2+} and Tc_Z^-:Mg^{2+} states (see above), and it was also used to compare protonation states of His64 in the Tc binding pocket. The equilibrium states were simulated for 3 ns each. Calculations are performed here at zero ionic strength, so that we solve the Poisson, and not the Poisson–Boltzmann, equation. Nevertheless, we refer to the method as PBLRA. For similar systems, we have found that PBLRA results are weakly sensitive to the ionic strength.[58] The Poisson equation was solved by use of a finite difference method implemented in the Charmm program, with a cubic grid of length 144 Å, with 0.4 Å between grid planes. 250 structures were used for each state. The solute dielectric constant was set to 1 or 2, and the solvent dielectric was set to 80, the experimentally determined value for bulk water.

Poisson calculations and free-energy component analysis: The Poisson model can also be used to compute the electrostatic contribution to the Tc binding free energy.[29,33,35] We simply compute the electrostatic free energy of the TetR complex and the separate partners and take the difference. This Poisson binding free energy then has a useful property: it can be broken down into several distinct contributions,[29,35] including contributions from individual amino acids in the Tc binding pocket. We refer to these as free-energy components. A higher solute dielectric, relative to the PBLRA situation, might be appropriate for this analysis. Indeed, with PBLRA we averaged over structures corresponding to the system before and after a charge modification (e.g., His protona-

 © 2007 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

tion). Dielectric relaxation of the protein is explicitly modeled by the structural differences between the two states. Here, in contrast, dielectric relaxation when the TetR:Tc complex is formed is implicitly modeled by the solute dielectric constant. Below, we have chosen the solute dielectric constant so that the binding free-energy difference between Tc_N and Tc_Z approximately reproduces that obtained by MDFE (with use of N_3 and Z_2). This gives a solute dielectric of four, similarly to previous work on similar problems.[33,35]

Crystallization, X-ray diffraction, and structure refinement: TetR of class D—TetR(D)—was co-crystallized with doxycycline and Mg^{2+} by previously described methods for other tetracycline complexes of TetR.[8,59,60] The doxycycline:Mg^{2+} complex of TetR(D) crystallized in space group $I4_122$, isomorphous with other TetR(D) complexes with tetracycline:Mg^{2+}. X-ray diffraction data were collected at the synchrotron beam line X13 of the EMBL outstation (DESY, Hamburg). Flash-frozen crystals were measured in a gaseous N_2 stream at 100 K. Data were reduced by use of the HKL package[61] and further processed with the aid of programs of the CCP4 suite.[62] Initial phasing and model building were based on the TetR(D) complex with 7-chlorotetracycline:Mg^{2+} (PDB ID: 2TCT). Refinement was performed with the program REFMAC.[63] Details and statistics for data collection and structure refinement are given in Table 7.

Table 7. Statistics for the X-ray data collection and structure refinement.

Data collection	
detector	MAR CCD 165 mm
radiation source	beam line X13, EMBL Hamburg c/o DESY
wavelength	0.8011 Å
resolution range	19.34—1.89 Å
tetragonal space group/*a*-axis/*c*-axis	$I4_122$/65.83 Å/179.5 Å
number of independent reflections	15374
completeness	99.75%
redundancy (highest shell)	15.8 (10.3)
$I/\sigma(I)$ (highest shell)	29.6 (7.3)
R_{sym} (%) (highest shell)	7.1 (36.2)
Wilson B-factor	29.9 Å^2
Refinement	
R factor (%)	20.2
R_{free} (test dataset with 5% data, 817 refl.)	24.6
number of atoms	1801
TetR/[Mg-doxyTc]/solvent	1661/33/107
average overall isotropic B-factors	35.9 Å^2
protein main chain/side chain	34.4/37.1 Å^2
Mg-doxyTc/solvent	25.6/40.6 Å^2
RMS deviations from ideal geometry	
bond lengths	0.026 Å
bond angles	2.308°
torsion angles	5.742°
PDB ID	2O7O

In a second step, the initially refined crystallographic model (model 1) was manually modified to create a second, alternative model (model 2). The amide side chains of Asn5, Asn82, and Gln116, the amide group of doxyTc, and the imidazole rings of His64 and His100 were all flipped by 180°. This alternative model was also refined to convergence. During the refinements, standard restraints were used for the Debye–Waller factors. Differences between the Debye–Waller factors of atoms separated by one or two covalent bonds were thus penalized, as is physically reasonable. The correct conformations of each group were then identified by comparison of the individual, isotropic, Debye–Waller factors of the corresponding amide and imidazole atoms in models 1 and 2. The coordinates of the best model (with the correct orientations of each individual group) have been submitted to the Protein Data Bank.

Acknowledgements

Support was provided by the Nuclear Toxicology program of the Commmissariat pour l'Énergie Atomique and the Centre National pour la Recherche Scientifique (to T.S.). An Egide PhD fellowship was awarded to A.A. We thank Martin Karplus for the CHARMM program. Some of the simulations were performed at the CINES supercomputing center of the French Ministry of Education.

Keywords: antibiotics · molecular dynamics · molecular recognition · protein structures · tetracycline binding

[1] I. Chopra, M. Roberts, *Microbiol. Mol. Biol. Rev.* **2001**, *65*, 232–260.
[2] I. Chopra, *Drug Resist. Updates* **2002**, *5*, 119–125.
[3] M. W. Olson, A. Ruzin, E. Feyfant, T. S. Rush, J. O'Connell, P. A. Bradford, *Antimicrob. Agents Chemother.* **2006**, *50*, 2156–2166.
[4] W. Saenger, P. Orth, C. Kisker, W. Hillen, W. Hinrichs, *Angew. Chem.* **2000**, *112*, 2122–2133; *Angew. Chem. Int. Ed.* **2000**, *39*, 2042–2052.
[5] M. Fussenegger, *Biotechnol. Prog.* **2001**, *17*, 1–51.
[6] C. Berens, S. Lochner, S. Lober, I. Usai, A. Schmidt, L. Drueppel, W. Hillen, P. Gmeiner, *ChemBioChem* **2006**, *7*, 1320–1324.
[7] J. Trylska, J. Antosiewicz, M. Geller, C. Hodge, R. Klabe, M. Head, M. Gilson, *Protein Sci.* **1999**, *8*, 180–195.
[8] W. Hinrichs, C. Kisker, M. Duvel, A. Muller, K. Tovar, W. Hillen, W. Saenger, *Science* **1994**, *264*, 418–420.
[9] P. Orth, D. Schnappinger, W. Hillen, W. Saenger, W. Hinrichs, *Nat. Struct. Biol.* **2000**, *7*, 215–219.
[10] O. Scholz, M. Köstner, M. Reich, S. Gastiger, W. Hillen, *J. Mol. Biol.* **2003**, *329*, 217–227.
[11] S. Kedracka-Krok, Z. Wasylewski, *Eur. J. Biochem.* **2003**, *270*, 4564–4573.
[12] S. Kedracka-Krok, A. Gorecki, P. Bonarek, Z. Wasylewski, *Biochemistry* **2005**, *44*, 1037–1046.
[13] E. M. Henssler, R. Bertram, S. Wisshak, W. Hillen, *FEBS J.* **2005**, *272*, 4487–4496.
[14] S. E. Reichheld, A. R. Davidson, *J. Mol. Biol.* **2006**, *361*, 382–389.
[15] A. Warshel, *Computer Modelling of Chemical Reactions in Enzymes and Solutions*, Wiley, New York, **1991**.
[16] T. Simonson, *Rep. Prog. Phys.* **2003**, *66*, 737–787.
[17] M. O. Schmitt, S. Schneider, *PhysChemComm* **2000**, *9*, 42–55.
[18] H. F. Dos Santos, C. S. Nascimento, P. Belletato, W. B. De Imeida, *J. Mol. Struct. THEOCHEM* **2003**, *626*, 305–319.
[19] C. F. Leypold, M. Reiher, G. Brehm, M. O. Schmitt, S. Schneider, P. Matousek, M. Towrie, *Phys. Chem. Chem. Phys.* **2003**, *5*, 1149–1157.
[20] O. Othersen, F. Beierlein, H. Lanig, T. Clark, *J. Phys. Chem. B* **2003**, *107*, 13743–13749.
[21] K. Meindl, T. Clark, *J. Phys. Chem. B* **2005**, *109*, 4279–4284.
[22] H. Lanig, O. G. Othersen, U. Seidel, F. Beierlein, T. E. Exner, T. Clark, *J. Med. Chem.* **2006**, *49*, 3444–3447.
[23] H. Lanig, O. G. Othersen, F. Beierlein, U. Seidel, T. Clark, *J. Mol. Biol.* **2006**, *359*, 1125–1136.
[24] A. Aleksandrov, T. Simonson, *J. Comput. Chem.* **2006**, *13*, 1517–1533.
[25] D. Pearlman in *The Encyclopedia of Computational Chemistry* (Eds.: P. v. R. Schleyer, N. L. Allinger, T. Clark, J. Gasteiger, P. A. Kollman, H. F. Schaefer III, P. R. Schreiner), Wiley, Chichester, **1998**.
[26] T. Simonson, G. Archontis, M. Karplus, *Acc. Chem. Res.* **2002**, *35*, 430–437.
[27] W. Jorgensen, *Science* **2004**, *303*, 1813–1818.

 © 2007 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

[28] C. F. Wong, J. A. McCammon, *Adv. Protein Chem.* **2003**, *66*, 87 – 121.
[29] T. Simonson in *Free Energy Calculations: Theory and Applications in Chemistry and Biology* (Eds.: C. Chipot, A. Pohorille), Springer, New York, **2006**, Chapter 12.
[30] J. Tomasi, R. Cammi, B. Mennucci, C. Cappelli, S. Corni, *Phys. Chem. Chem. Phys.* **2002**, *4*, 5697 – 5712.
[31] G. Archontis, T. Simonson, *Biophys. J.* **2005**, *88*, 3888 – 3904.
[32] C. Schutz, A. Warshel, *Proteins* **2001**, *8*, 211 – 217.
[33] P. Kollman, I. Massova, C. Reyes, B. Kuhn, S. Huo, L. Chong, M. Lee, T. Lee, Y. Duan, W. Wang, O. Donini, P. Cieplak, J. Srinivasan, D. Case, T. Cheatham, *Acc. Chem. Res.* **2000**, *33*, 889 – 897.
[34] P. Orth, F. Cordes, D. Schnappinger, W. Hillen, W. Saenger, W. Hinrichs, *J. Mol. Biol.* **1998**, *279*, 439 – 447.
[35] G. Archontis, T. Simonson, M. Karplus, *J. Mol. Biol.* **2001**, *306*, 307 – 327.
[36] O. Scholz, P. Schubert, M. Kintrup, W. Hillen, *Biochemistry* **2000**, *39*, 10914 – 10920.
[37] R. Hooft, C. Sander, G. Vriend, *Proteins* **1996**, *26*, 363 – 376.
[38] P. Kollman, *Chem. Rev.* **1993**, *93*, 2395.
[39] T. Lederer, M. Kintrup, M. Takahashi, P. Sum, G. Ellestad, W. Hillen, *Biochemistry* **1996**, *35*, 7439 – 7446.
[40] F. Allen, O. Kennard, *Chem. Des. Autom. News* **1993**, *8*, 31 – 37.
[41] N. Allinger, Y. Yuh, J. Lii, *J. Am. Chem. Soc.* **1989**, *111*, 8551 – 8566.
[42] P. Ren, J. Ponder, *J. Phys. Chem. B* **2003**, *107*, 5933 – 5947.
[43] G. Carpenter, S. Grossberg, *Appl. Opt.* **1987**, *26*, 4919 – 4930.
[44] A. Becke, *J. Chem. Phys.* **1993**, *98*, 5648.
[45] C. Lee, W. Yang, R. Parr, *Phys. Rev. B* **1988**, *37*, 785 – 789.
[46] W. Hehre, L. Radom, P. Schleyer, J. Pople, *Ab initio Molecular Orbital Theory*, Wiley, New York, **1987**.
[47] Gaussian 03, Revision C.02, M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. A. Montgomery, Jr., T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, C. Gonzalez, J. A. Pople, Gaussian, Inc., Wallingford, CT, **2004**.
[48] D. Beglov, B. Roux, *J. Chem. Phys.* **1994**, *100*, 9050 – 9063.
[49] T. Simonson, *J. Phys. Chem. B* **2000**, *104*, 6509 – 6513.
[50] A. Mackerell, D. Bashford, M. Bellott, R. Dunbrack, J. Evanseck, M. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph, L. Kuchnir, K. Kuczera, F. Lau, C. Mattos, S. Michnick, T. Ngo, D. Nguyen, B. Prodhom, W. Reiher, B. Roux, J. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, D. Yin, M. Karplus, *J. Phys. Chem. B* **1998**, *102*, 3586 – 3616.
[51] W. Jorgensen, J. Chandrasekar, J. Madura, R. Impey, M. Klein, *J. Chem. Phys.* **1983**, *79*, 926 – 935.
[52] R. Stote, D. States, M. Karplus, *J. Chim. Phys.* **1991**, *88*, 2419 – 2433.
[53] B. Brooks, R. Bruccoleri, B. Olafson, D. States, S. Swaminathan, M. Karplus, *J. Comput. Chem.* **1983**, *4*, 187 – 217.
[54] T. Simonson in *Computational Biochemistry & Biophysics* (Eds.: O. Becker, A. Mackerell, Jr., B. Roux, M. Watanabe), Marcel Dekker, New York, **2001**, Chapter 9.
[55] T. Simonson, J. Carlsson, D. A. Case, *J. Am. Chem. Soc.* **2004**, *126*, 4167 – 4180.
[56] Y. Sham, Z. Chu, A. Warshel, *J. Phys. Chem. B* **1997**, *101*, 4458 – 4472.
[57] I. Eberini, A. Baptista, E. Gianazza, F. Fraternali, T. Beringhelli, *Proteins* **2004**, *54*, 744 – 758.
[58] D. Thompson, P. Plateau, T. Simonson, *ChemBioChem* **2006**, *7*, 337 – 344.
[59] C. Kisker, W. Hinrichs, K. Tovar, W. Hillen, W. Saenger, *J. Mol. Biol.* **1995**, *247*, 260 – 280.
[60] P. Orth, D. Schnappinger, P. Sum, G. Ellestad, W. Hillen, W. Saenger, W. Hinrichs, *J. Mol. Biol.* **1999**, *285*, 455 – 461.
[61] Z. Otwinowski, W. Minor, *Methods Enzymol.* **1997**, *276*, 307 – 326.
[62] Collaborative Computational Project Number 4, *Acta Crystallogr. Sect. D Biol. Crystallogr.* **1994**, *50*, 760 – 776.
[63] G. Murshudov, A. Vagin, E. Dodson, *Acta Crystallogr. Sect. D Biol. Crystallogr.* **1997**, *53*, 240 – 255.
[64] A. Nicholls, *Grasp: Graphical Representation and Analysis of Structural Properties*, Columbia University, New York, **1991**.
[65] W. L. DeLano, *The PyMOL molecular graphics system*, DeLano, San Carlos, USA, **2002**.

Received: February 6, 2007
Published online on March 16, 2007

 © 2007 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

doi:10.1016/j.jmb.2008.03.022 J. Mol. Biol. (2008) 378, 896–910

JMB

Available online at www.sciencedirect.com

ScienceDirect

ELSEVIER

Tet Repressor Induction by Tetracycline: A Molecular Dynamics, Continuum Electrostatics, and Crystallographic Study

Alexey Aleksandrov[1], Linda Schuldt[2], Winfried Hinrichs[2]* and Thomas Simonson[1]*

[1]*Laboratoire de Biochimie (CNRS UMR7654), Department of Biology, Ecole Polytechnique, 91128 Palaiseau, France*

[2]*Institute for Biochemistry, Department of Molecular Structural Biology, University of Greifswald, Felix-Hausdorff-Strasse 4, D-17489 Greifswald, Germany*

Received 21 December 2007; received in revised form 27 February 2008; accepted 3 March 2008
Available online 19 March 2008

The Tet repressor (TetR) mediates the most important mechanism of bacterial resistance against tetracycline (Tc) antibiotics. In the absence of Tc, TetR is tightly bound to its operator DNA; upon binding of Tc with an associated Mg^{2+} ion, it dissociates from the DNA, allowing expression of the repressed genes. Its tight control by Tc makes TetR broadly useful in genetic engineering. The Tc binding site is over 20 Å from the DNA, so the binding signal must propagate a long distance. We use molecular dynamics simulations and continuum electrostatic calculations to test two models of the allosteric mechanism. We simulate the TetR:DNA complex, the Tc-bound, "induced" TetR, and the transition pathway between them. The simulations support the model inferred previously from the crystal structures and reveal new details. When $[Tc:Mg]^+$ binds, the Mg^{2+} ion makes direct and water-mediated interactions with helix 8 of one TetR monomer and helix 6 of the other monomer, and helix 6 is pulled in towards the central core of the structure. Hydrophobic interactions with helix 6 then pull helix 4 in a pendulum motion, with a maximal displacement at its N-terminus: the DNA interface. The crystal structure of an additional TetR reported here corroborates this motion. The N-terminal residue of helix 4, Lys48, is highly conserved in DNA-binding regulatory proteins of the TetR class and makes the largest contribution of any amino acid to the TetR:DNA binding free energy. Thus, the conformational changes lead to a drastic reduction in the TetR:DNA binding affinity, allowing TetR to detach itself from the DNA. Tc plays the role of a specific Mg^{2+} carrier, whereas the Mg^{2+} ion itself makes key interactions that trigger the allosteric transition in the TetR:Tc complex.

© 2008 Elsevier Ltd. All rights reserved.

Keywords: protein; allostery; computer simulation; binding free energy; gene regulation

Edited by D. Case

Introduction

Tetracyclines (Tcs) are widely used broad-spectrum antibiotics that affect both Gram-positive and Gram-negative bacteria, binding to the ribosome and inhibiting protein synthesis.[1] The most common resistance mechanism in Gram-negative bacteria is the active efflux of Tc out of the cell by the membrane transport protein TetA. This mechanism is tightly regulated at the transcription level by the Tet repressor (TetR) protein. TetR is a homodimer; in the absence of Tc, it binds tightly to two palindromic operators (binding constant $K_a = 10^{11}\ M^{-1}$),[2] blocking

*Corresponding authors. E-mail addresses: winfried.hinrichs@uni-greifswald.de; thomas.simonson@polytechnique.fr.

Present address: L. Schuldt, EMBL Hamburg, c/o DESY, Building 25A, Notkestrasse 85, D-22603 Hamburg, Germany.

Abbreviations used: TetR, Tet repressor; Tc, tetracycline; DBD, DNA-binding domain; MD, molecular dynamics; anTc, anhydrotetracycline.

0022-2836/$ - see front matter © 2008 Elsevier Ltd. All rights reserved.

the expression of the *tetA* and *tetR* genes. Tc binds to TetR as a $[Tc:Mg]^+$ complex with a nanomolar dissociation constant, inducing conformational changes that reduce the binding constant of TetR and DNA by 6–10 orders of magnitude.[3] After dissociation of TetR from DNA, the expression of *tetR* and *tetA* can proceed. Today, the TetR/Tc system is of general interest because of widespread application in molecular and cell biology as a sensitive switch for target gene regulation.[4–6]

Crystal structures of TetR are available in the "non-induced" state, complexed with DNA; in the "induced" state, complexed with various Tcs; and in the "apo" state (neither ligand). TetR consists of two identical monomers; the homodimer is divided into two N-terminal DNA-binding domains (DBDs) and a single, central, core domain (Fig. 1). The core domain is built around a four-helix bundle and includes two deep $[Tc:Mg]^+$ binding cavities, located over 20 Å from the DNA-binding interface. Upon induction, allosteric changes propagate from the Tc binding sites to the DNA-binding interface. Based on the experimental structures, a model for the allosteric changes was proposed.[7] In this model, conformational changes are triggered by the Mg^{2+} ion co-bound to Tc and by the amino acids that coordinate it. Indeed, when $[Tc:Mg]^+$ binds, the last turn of helix 6 unwinds (residues 100–103) and shifts to interact with the metal ion. His100 coordinates it directly, while Thr103 coordinates it indirectly, through bridging water molecules. The shift is accompanied by a pendulum motion of two alpha helices, helices 4 and 4′ (we use a prime to designate groups or residues belonging to the second TetR monomer). This displaces the DBDs, so that they no longer fit into adjacent major grooves of the DNA. According to the model, the structural shifts near the $[Tc:Mg]^+$ ligand trigger the helix displacements in a manner that has not yet been determined. A consequence of the model is that the main role of Tc is to act as a Mg^{2+} carrier, binding specifically to TetR, rigidifying the surrounding groups, and bringing Mg^{2+} into its proper position to trigger the allosteric motions.

The above model is mainly based on crystal structures of the endpoint states: induced and non-induced. Crystal structures for two intermediate states also support the model: (i) When Tc binds in the absence of Mg^{2+}, TetR remains in the non-induced state and can bind to DNA[8]; this supports the idea that Mg^{2+} triggers the motion.[8] (ii) In one "half-induced" crystal structure, $[Tc:Mg]^+$ is present in just one monomer; the dimer symmetry is broken, and the DNA binding constant is reduced by about 10^3.[8] Additional information has been provided by biochemical and biophysical measurements.[1,9,10] Notice that certain atypical ligands such as anhydrotetracycline (a type II Tc) can induce TetR without Mg^{2+}[1,11,12]; see below.

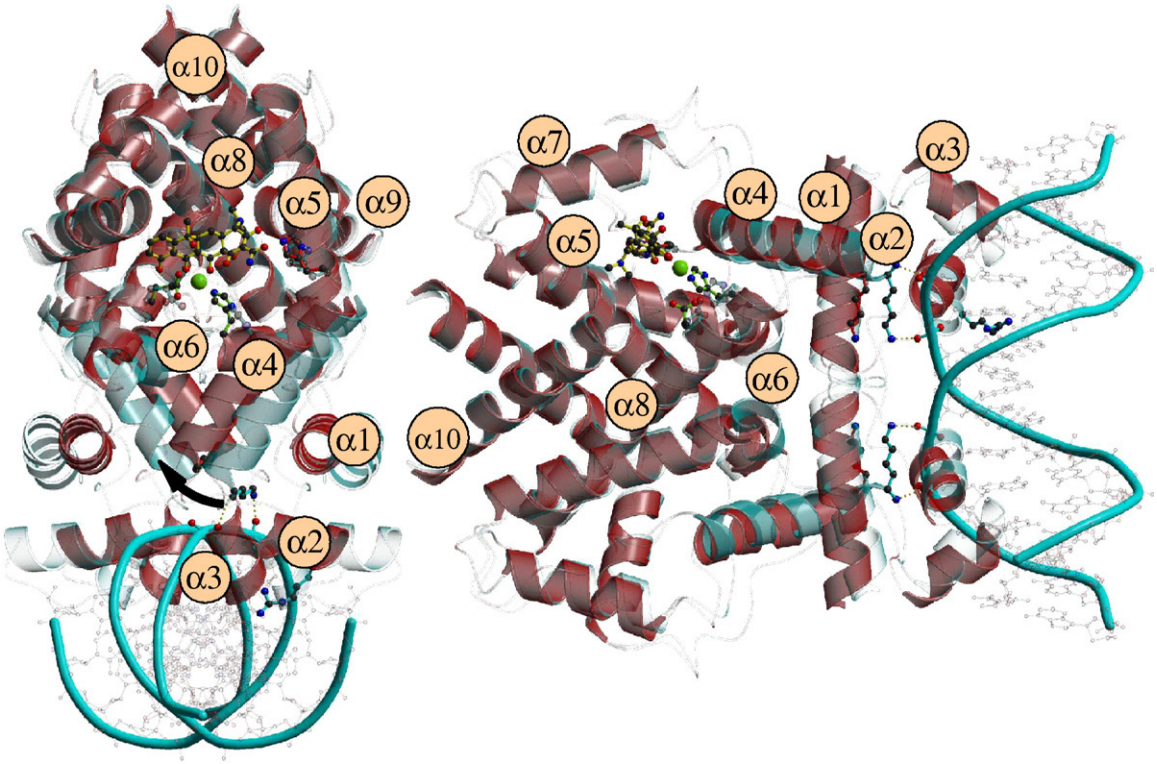


Fig. 1. An overview of the TetR structure in its induced (red) and non-induced (blue) states. Two orthogonal views are shown. The TetR:DNA complex is blue; the DNA backbone is shown as a tube; bases are shown as sticks (light grey). Superimposed (red) is the induced, TetR:Tc structure. Tc is shown with a ball-and-stick representation. The Mg^{2+} ion is a green sphere. Helices are labelled in one monomer. Only one Tc is shown (monomer 1). Lys48 is shown (ball and sticks) making hydrogen bonds to the DNA. The helix 4 shift upon induction is highlighted by an arrow.

A different model of the allosteric mechanism, based on molecular dynamics (MD) simulations, was proposed recently by Lanig *et al.*[13,14] MD simulations can in principle provide the most detailed picture, including a dynamical view of the transition between endpoint states. In this model, Mg^{2+} is assumed to be present in the non-induced state (i.e., in the absence of Tc) and coordinated by Asp157 (referred to as Asp156 by those authors). Upon Tc binding, Asp157 is displaced, salt bridges in the Tc binding site rearrange, and the structure gains in overall flexibility. This in turn is predicted to decrease the TetR affinity for DNA.

Several new findings suggest that the problem of TetR allostery should be revisited. First, the flexible loop 156–164, to which Asp157 belongs, was disordered and invisible in all but one X-ray structure[8] published before 2007. This was an unusual structure, with one induced and one non-induced monomer. A very recent structure of the doxycycline complex[15] also featured an ordered loop, with a conformation similar to the earlier one. Another very recent structure of an unusual TetR (a fusion of class B and D repressors) also had an ordered loop.[16] In these structures, the position of Asp157 is inconsistent with specific Mg^{2+} coordination; rather, coordination by Asp157 would appear similar to non-specific Mg^{2+} coordination by any of the other carboxylate groups at the TetR surface. Second, a key residue at the DNA-binding interface, Lys48, was slightly mispositioned in the original, medium-resolution, X-ray structure[7] so that its side chain did not appear to interact closely with the DNA, and it has not been considered in discussions of the allosteric mechanism or mutagenesis studies. Third, recent mutagenesis work showed that mutant TetRs can be constructed where Asp157 is mutated to Ala, and are still inducible.[17] Consistent with this, the Asp157 position shows no sequence conservation whatsoever, with a wide variety of polar and hydrophobic amino acids occupying this position in different TetRs. Finally, the most recent structural data strongly suggest that Mg^{2+} is not present in the DNA complex, contrary to the earlier MD model.[13]

For these reasons, we have reexamined the allosteric mechanism of TetR induction using MD simulations and continuum electrostatic calculations. The calculations are performed for the D class repressor, but the main conclusions are expected to apply to other TetRs. We make use of the more recent experimental information. We also report a new crystal structure of a homologous Tet repressor of the H class, which sheds additional light on the mechanical flexibility of TetR. The sequence identity between TetR(D) and TetR(H) is 57%. We benefit from a very recent, high-quality, force field model of Tc,[18] which has been tested in simulations of both TetR and the 30S ribosome subunit[15,19] and validated by comparisons to experimental measurements of Tc binding to both systems. Our earlier TetR simulations with this force field helped reveal some structural details not visible in available crystal structures, such as active-site protonation patterns, side-chain orientations, and the Tc tautomeric state,[15] all of which are important for the present study. We use MD simulations to study the endpoint, induced and non-induced states. We use an established continuum electrostatic method to obtain thermodynamic information.[20–24] Finally, we use a "targeted molecular dynamics" method to study the pathway between the endpoint states and to test the model described above. $[Tc:Mg]^+$ is added to the TetR:DNA complex to form a non-equilibrium, intermediate state. A weak, biasing force is applied to groups very close to the Tc, pulling them gently towards their induced conformation. With this method, structural rearrangements near the Tc occur very early, and we can examine how the changes propagate to the rest of the structure, including the DNA-binding interface. Thus, a detailed hypothesis can be formulated regarding the order and timescale of individual events, and the proposed models for the allosteric transition can be further tested.

The simulations support the model inferred previously from the crystal structures.[7] When $[Tc:Mg]^+$ binds, the Mg^{2+} ion coordinates directly to helix 6 (imidazole of His100) and makes water-mediated interactions with helix 6 (Thr103) of one TetR monomer and helix 8 (Glu147′) of the other monomer. Helix 6 is pulled in towards the central core of the structure; hydrophobic interactions with helix 6 then pull helix 4, in a pendulum motion, with a maximal displacement at its N-terminus—the DNA-binding interface. The N-terminal residue of helix 4, Lys48, is highly conserved in transcriptional regulatory domains of the TetR family.[25] It makes a strong interaction with a DNA phosphate group, producing the largest contribution of any TetR amino acid to the TetR:DNA binding free energy. Thus, the conformational changes upon $[Tc:Mg]^+$ binding lead to a drastic reduction in the TetR:DNA binding affinity. Tc plays the role of a specific Mg^{2+} carrier, whereas the Mg^{2+} ion itself makes the key interactions that trigger the allosteric transition.

Results

Analysis of the experimental and simulation structures

We first compare the X-ray structures of the two endpoint states, TetR:DNA and TetR:Tc, and recall some known features of the allosteric transition. Next, we compare the MD structures to the experiment structures, mainly to establish the quality of the simulations and also to identify possible artefacts in the X-ray structures.

X-ray structures of the endpoint states

The two monomers of TetR each include 10 alpha helices (sometimes denoted α1–10), and are related by 2-fold symmetry in both the non-induced and

induced states (Fig. 1). The two DBDs include helices 1–3; helices 2 and 3 form a classic helix–turn–helix motif. The center of the core domain is a four-helix bundle involving helices 8, 8′, 10, and 10′. We use a prime to designate groups or residues from monomer 2. The $[Tc:Mg]^+$ binding sites are formed by α5–α9, α8′ and by α5′–α9′, α8 and are located 24 Å from the DNA (distance from the Mg^{2+} to the DNA surface). The DNA fragment in the X-ray structure (and the simulations) contains 15 base pairs, including a central pair that does not interact directly with TetR, and 14 palindromic pairs. Sequence-specific contacts between TetR and the DNA bases are made by helices 2 and 3, especially by Arg28 and Gln38, which hydrogen bond to Gua2 and Ade3, respectively. The Arg28 positive charge also interacts with the DNA backbone through long-range electrostatics. Arg28 is highly conserved in TetRs (Fig. 2). Interactions with the DNA sugar–phosphate backbone are made by Thr26, Thr27, Thr40, Tyr42, and Lys48. Lys48 belongs to helix 4 and is not part of the DBD. It is highly conserved in transcriptional regulation domains of the TetR class.

In the structure with bound $[Tc:Mg]^+$, His100 and Thr103 have shifted to coordinate the ion, peeling off the C-terminus of α6 to form a type II beta-turn. Helices 4 and 4′ have swung in a pendulum motion, remaining anchored at their C-termini, where His64 hydrogen bonds to the A ring of Tc (Fig. 1). The motion drives helices 3 and 3′ apart, so that they no longer fit optimally into adjacent major grooves of the operator DNA. Mutations at position 103 that perturb the Thr103···Mg^{2+} interaction (e.g., T103A, T103I, T103K) lead to non-inducible TetRs: $[Tc:Mg]^+$ binds but TetR does not dissociate from the DNA.[28] Similarly, Tc variants that bind to TetR without Mg^{2+} have been obtained and shown to be non-inducing in most cases: the protein remains in its DNA-binding conformation. A probable exception is anhydrotetracycline (anTc), which appears able to induce TetR even though it binds without an associated Mg^{2+}.[11] Anhydrotetracycline has a chemical structure rather different from that of the classical Tcs.[1,12] Furthermore, it appears to bind TetR with an associated Na^+ in one TetR crystal structure (W.H., unpublished data), which raises the possibility that the observed induction[11] may have been assisted by sodium present in the buffer solution. Note that while TetR structures containing Tc but no Mg^{2+} have been observed, the reverse is not true: TetR crystal structures containing Mg^{2+} but no Tc have never been observed.

Motion of the DBDs upon induction is rather complex. They shift approximately perpendicular to the DNA in a scissor-like manner (Fig. 1). We analyzed the distances between symmetry-related C^α pairs in the TetR homodimers. Significant changes are observed for both the DBD (α1, α2, α3) and the adjacent α4. Helices 3 and 3′ increase their separation upon induction, whereas α1/α1′ and α2/α2′ come closer together. Thus, movement of the DBDs has to be described as a combined rotation and translation. The angle between α4 and α4′ increases from 54° to 92° on Tc binding, consistent with the pendulum motion mentioned above, while the angle between α3 and α3′ decreases. All the main TetR:DNA interface residues shift as a result of the conformational changes; for example, Thr27, Arg28, Gln38, Thr40, and Lys48 all shift by 4.3–6.5 Å.

In the TetR(H)–[Mg:anTc] complex, the movement of the DBDs and α4 is more distinct than in the TetR (D) structure. This can be explained by weaker crystal packing constraints, since the TetR(H) complex (PDB entry 2VPR) crystallized in a different space group with a higher solvent content. TetR(D) [Protein Data Bank (PDB) entries 2TCT/2TRT] has a tetragonal space group, $I4_122$; the homodimers are arranged in a head-to-head orientation with their DBDs facing each other and a solvent content of just 40%. For TetR(H), the space group is orthorhombic ($C222_1$), with a head-to-tail orientation of the homodimers and a solvent content of 46%. Comparing the different lattice contacts using the PISA program,[29,30] we found that in TetR(D):Tc, the homodimer contacts are mainly DBD contacts, whereas the TetR(H) homodimers are positioned mainly by contacts between the core

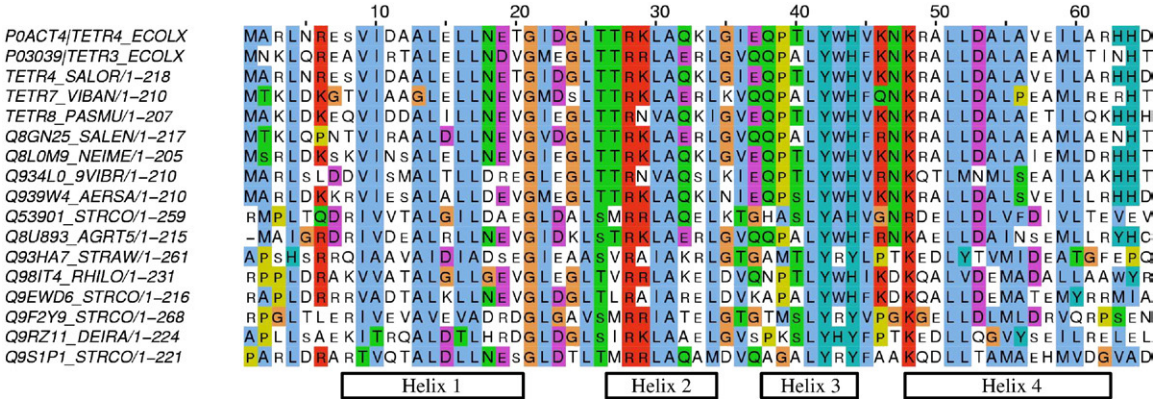


Fig. 2. Sequence alignment of the TetR DBD and helix 4. The top sequence is the *E. coli* D-class TetR studied here; others correspond to the diverse and non-redundant set of TetRs in the Pfam database.[26] The sequences were aligned with ClustalW.[27]

domains. This explains why the DBDs in TetR(H) have a higher mobility and the effects of induction are (slightly) more pronounced.

Simulation structures of the endpoint states

The MD simulations employ a large, cubic box containing the TetR(D) protein, 20,959 water molecules, and counterions, for a total of 70,288 atoms. The box is replicated periodically in all directions to avoid boundary effects. The simulations lasted 10 ns for TetR:Tc and 7 ns for TetR:DNA. These durations should give a reasonable picture of the equilibrium structure and dynamics. The MD structures deviate moderately from the crystal structures. Averaging over the last 2 ns of MD, the mean rmsd of the backbone atoms from the starting crystal positions is 1.9 Å for TetR:Tc and 1.7 Å for TetR:DNA. The largest deviations are for the DBD, which makes extensive crystal contacts in the experimental TetR (D):Tc structure (as discussed above). The interhelix distances α2–α2′, α3–α3′, and α4–α4′ are all somewhat longer in the TetR:Tc MD structure due to the lack of crystal contacts.

Root mean square fluctuations were computed for the C^{α} atoms in both states and compared to the crystallographic Debye–Waller factors, which measure positional disorder. The Debye–Waller factors include a contribution from overall rotation/translation of the protein (or protein:DNA complex), which can be described by four libration parameters (two directional parameters and two amplitudes).[31,32] This contribution must be subtracted to obtain an estimate of the intramolecular vibrations, which can then be compared to the MD predictions. The four libration parameters are adjusted to optimize the fit between the MD and X-ray estimates, which are compared in Fig. 3. Agreement between experiment and simulations is very good. If the experimental Debye–Waller factors are adjusted using a single scaling factor instead of the four librational parameters, the level of agreement is comparable (not shown). For most of the protein, the MD fluctuations match the X-ray data very well. The overall correlation between the computed and experimental data is 87% and 84% with and without DNA, respectively. In the TetR:Tc complex, for regions where there is a discrepancy, there is always a crystal contact in the experimental system, so that the experimental fluctuations are smaller than the MD ones. For the DBDs and helix 4 (residues 2–62), for example, the experimental motions are clearly limited by crystal contacts. Indeed, in the TetR(D) crystal structure, most of the crystal packing interactions are made by the DBD. Furthermore, we show below that in TetR structures with different crystal packing interactions (different space groups), the DBD has a well-conserved internal structure but different orientations with respect to the core domain. In both MD simulations, the loop between helices 8 and 9 has the largest fluctuations. This agrees with the X-ray structures, since this loop was either invisible or fragmented in the experimental electron density maps (residues 156–163 for TetR:DNA and 156–164 for TetR:Tc). Only recently was a structure obtained where this loop is fully ordered in both monomers.[15] For regions close to the loop, agreement between the MD and X-ray data is very good. The overall fluctuations, both from the X-ray data and the simulations, are slightly greater for the TetR:DNA complex than for the TetR:Tc complex (in contrast to another recent simulation study[13]).

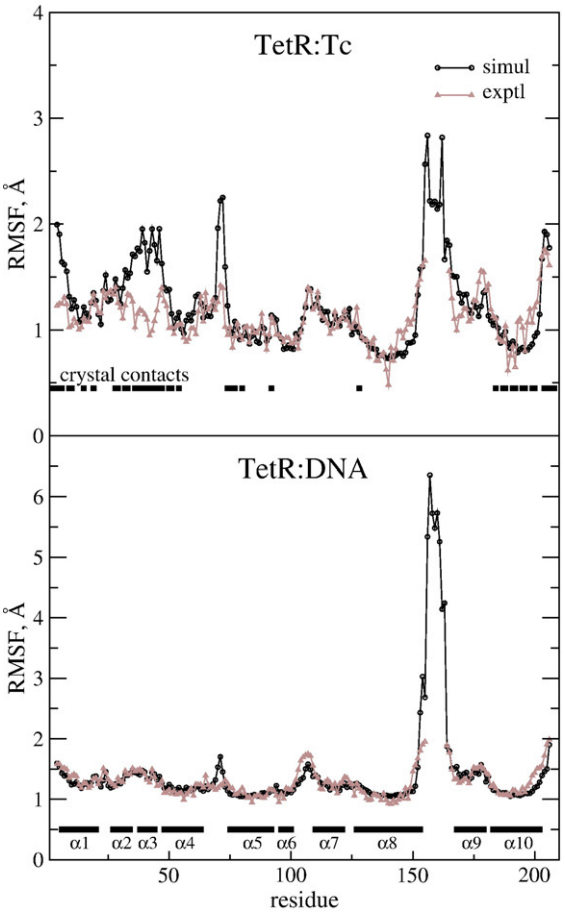


Fig. 3. RMS atomic fluctuations estimated from the MD simulations and the crystallographic Debye–Waller factors. The crystallographic data are processed to remove contributions from overall rotation/translation of the protein or the protein:DNA complex (see the text). Secondary-structure elements and crystal contacts (space group $I4_122$) are indicated.

In the TetR:DNA X-ray structure,[7] the Lys48 side chain is positioned to interact with the backbone carbonyl of Leu25. This position is probably incorrect. The Lys48-N^{ζ}–Leu25-O distance is 2.7 Å, while the nearest DNA phosphate group is Gua10-O1P, 3.8 Å away. After a brief period of molecular dynamics, Lys48 rotates to make a salt bridge to Gua10-O1P. The Lys48-N^{ζ}–Leu25-O and Lys48-N^{ζ}–Gua10-O1P distances change to 3.0 and 2.6 Å, respectively. Both the initial and final positions are compatible with the available electron density maps (not shown). Note that the simulations include an explicit and accurate description of the surrounding solvent and its shielding of the TetR:DNA interactions. In the X-ray

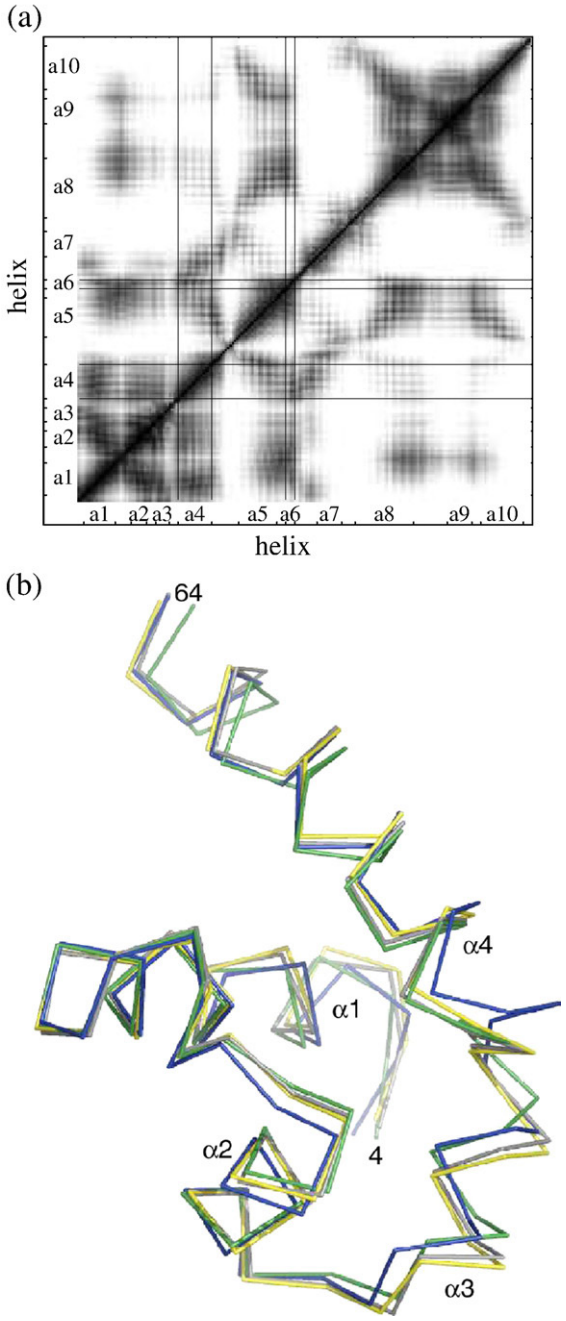


Fig. 4. (a) Covariance matrix for the atomic displacements of the backbone C^{α}s. Dark areas correspond to large values of the covariance (strong correlations between atoms). White areas correspond to uncorrelated motions. Only correlations within one monomer are shown. Correlations between monomers mainly reveal the presence of the core domain, which spans the two monomers. (b) Structures of the DBD and helix 4 from different TetRs in different states. Blue, TetR(D):DNA (PDB entry 1QPI). Yellow, TetR(D):Tc (PDB entry 2TRT). Grey, apoTetR(D) (PDB entry 1BJZ). Green, TetR(H):anTc (PDB entry 2VPR). The DBD and helix 4 are seen to behave practically as a rigid unit when TetR switches states, suggesting that they could be viewed as a single domain.

structure and the simulations, the Lys48 backbone hydrogen bonds to the adjacent phosphate 11; the lysine side-chain length matches exactly the interphosphate distance in B-DNA. The temperature factors of Lys48 atoms in the X-ray structure are lower than those of neighboring atoms (27 $Å^2$ for Lys48-N^{ζ}; 50–80 $Å^2$ for Arg49; around 40 $Å^2$ for Asn47 and Arg28). For the other interface residues, especially Arg28, Gln38, Thr26, and Thr27, the interactions seen in the X-ray structure are maintained throughout the simulations.

Dynamical and mechanical properties of the protein

We saw above that the simulations accurately reproduce the mean amplitude of the protein fluctuations (Fig. 3). Here, we characterize the dynamical properties of TetR in more detail by analyzing the correlations between atomic motions in different parts of the structure. The interatom correlations averaged over the TetR:Tc simulation are represented as a covariance matrix in Fig. 4a. The individual domains within the protein dimer, the rigid core and the two DBDs, are easily recognizable. Surprisingly, however, helix 4 is correlated with the DBD so highly that it appears to be a part of the same domain. Protein domains have sometimes been defined based on the mechanical properties of the structure[33]; with this definition, helices 1–4 could be viewed as a single domain. This contrasts markedly with the usual definition of the DBD.

Table 1. Statistics of the X-ray data collection and structure refinement

Data collection	
Radiation source	Beam line X13, EMBL Hamburg c/o DESY
Wavelength (Å)	0.8084
Detector (mm)	MAR CCD 165
Resolution range (Å)	54.39–2.49
Orthorhombic space group	$C222_1$
Cell dimensions $a/b/c$ (Å)	56.36/108.74/68.98
No. of independent reflections	7669
Completeness (highest shell) (%)	98.4 (97.7)
Redundancy (highest shell)	4.2 (3.6)
$I/\sigma(I)$ (highest shell)	14.6 (2.9)
R_{sym} (%) (highest shell)	9.6 (50.9)
Wilson ^{B}-factor ($Å^2$)	53.2
Refinement, 20–2.5 Å	
R−factor (%)	20.1
R_{free} (test data set with 10% data, 741 reflections)	28.4
No. of atoms	
TetR(H)	1547
[Mg-anTc]	32
Solvent	23
B-factor ($Å^2$)	
Average overall isotropic	42.2
Protein main chain/side chain	42.2/42.8
Mg-anTc/solvent	26.1/41.3
rmsds from ideal geometry	
Bond lengths (Å)	0.016
Bond angles (°)	1.576
PDB entry	2VPR

That the DBD and helix 4 behave as a single unit is illustrated even more strikingly in Fig. 4b. The DBD and helix 4 from four structures are superimposed: TetR(D):DNA, TetR(D):Tc, apo-TetR(D), and the TetR(H):anTc structure reported here (Table 1).

The DBDs and helices 4 superimpose almost perfectly (rmsd of ~1 Å between structures) despite the different environments, which lead to different positions of the DBD with respect to the core domain (see, e.g., Fig. 1). The present superimposition shows that the DBD and helix 4 move practically as a rigid unit upon induction. This property could facilitate rapid switching between the induced and non-induced conformations and possibly provide a selective advantage.

Protein:DNA binding free energy: amino acid contributions

To identify the most important TetR···DNA interactions, we analyze the MD structures using an established, continuum electrostatic model.[20–24] We compute the electrostatic contribution to the TetR:DNA binding free energy by solving the Poisson–Boltzmann equation numerically. An attractive feature of the continuum electrostatic model is that the free energy can be decomposed into a series of contributions, including contributions that depend explicitly on the atomic charges of a single amino acid side chain. Such a contribution is referred to as a free-energy component; it measures qualitatively the direct interaction between the amino acid charges and the DNA charges (partly shielded by solvent and by other protein groups). We treat the protein:DNA complex as one homogeneous dielectric medium, and the surrounding solvent as another. The solvent dielectric constant is set to the experimental value of 80. The solute value is set to 4, based on our earlier work on TetR.[15] Results with a dielectric constant of 6 are also computed. The range 4–6 is similar to values used for other, similar problems.[21,22]

With this analysis, the largest contributions to the binding free energy (Fig. 5) come from Arg28 and Arg28′ (13 kcal/mol each; 30% of the total protein contribution) and especially Lys48 and Lys48′ (15 kcal/mol each; 34% of the total). Lys48 has been neglected until now because of its presumably incorrect positioning in the TetR:DNA X-ray structure, as described above. Although Gln38 is at the N-terminus of helix 3, adjacent to the DNA, it makes only a 2 kcal/mol contribution. Nearby Glu37 makes a 5 kcal/mol contribution, opposing binding through its negative charge. Both Arg28 and Lys48 are positively charged and thus have strong favorable interactions with the negatively charged DNA. Their flanking residues, particularly Thr27, Lys29, and Arg49, make smaller contributions of about 3 kcal/mol each.

Free-energy components must be interpreted with caution, as discussed recently for TetR–Tc binding.[15] Not only is the continuum model approximate, but the components are obtained by considering only the TetR:DNA structure, not the unbound TetR and DNA structures. A comparison to mutagenesis data provides a rough test of their validity. Figure 5 shows the experimental effect of selected TetR mutations on the effectiveness of TetR as a repressor. Measured levels of repression of a reporter gene by TetR[9] have

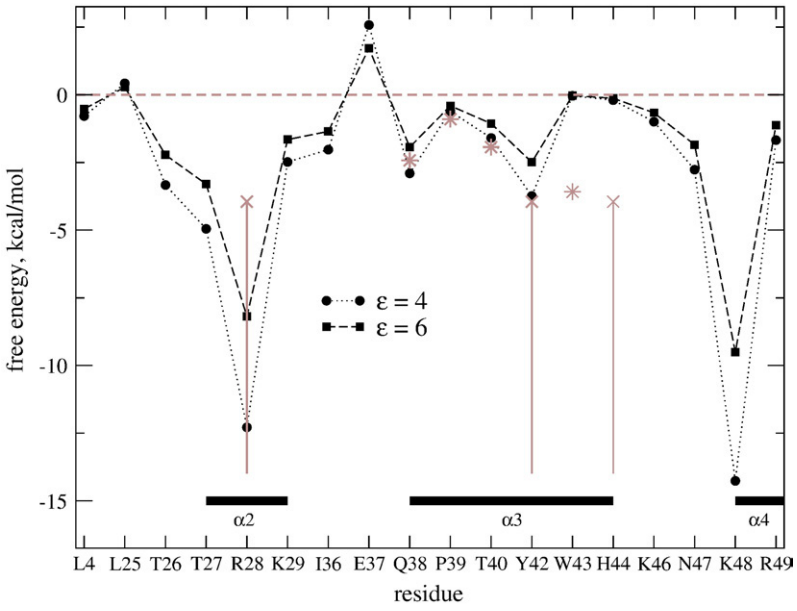


Fig. 5. Free-energy components from the continuum electrostatic model. Each component corresponds to the contribution of an amino acid to the TetR:DNA binding free energy. Results with two values for the protein and DNA dielectric constant, 4 and 6, respectively, are shown. Grey stars show experimental estimates obtained from alanine scanning mutagenesis. In some cases, the experiment only gives a lower bound, so that a grey line extends below the experimental point, which is shown as a cross for these cases. In both the experiments and the simulations, the changes were introduced into both monomers simultaneously. The contributions shown are scaled by one-half to correspond to a single amino acid and a single monomer.

been converted to DNA-binding free-energy changes, using a simple kinetic model (see Supplementary Material). The TetR:DNA binding equilibrium is considered, and the experimental level of expression of the reporter gene is assumed to be proportional to the amount of DNA not bound to TetR. Using the experimental TetR:DNA binding constant and the law of mass action, one obtains a relation between the binding free-energy change and the observed expression levels.[9] Details are given in the Supplementary Material. When the change in binding affinity is large, only a lower bound can be inferred because of experimental limitations.[9] In effect, once the DNA binding falls below a certain level, the reporter gene is fully expressed and further losses in binding cannot be detected. Figure 5 shows that mutations to alanine at positions 38, 39, and 40 lead to moderate decreases in DNA binding and repression, consistent with the computed free-energy components. Agreement with experiment is best with a solute dielectric of 4. For positions 28 and 42, the free-energy changes inferred from experiment are at the limit of detection and should be considered a lower bound. The computed Tyr42 component is about equal to the experimental lower bound. The Arg28 component is consistent with this lower bound but is much larger. Presumably, when Arg28 is changed to a hydrophobic group, the loss in binding affinity is larger than the minimum required to give essentially full expression of the reporter gene. For Trp43 and His44, the computed components are smaller than the measured changes. In the MD model, these amino acids are close to the end of the DNA fragment (Fig. 1); the experiments use a much longer DNA fragment.

In addition to the direct TetR:DNA interactions computed above, there are indirect contributions that arise from intraprotein interactions and are harder to quantify. For example, residues Asp23 and Asp23′, near the N-terminus of helix 2, are close together in the TetR:DNA complex, but move apart when Tc binds and the pendulum motion of helix 4 occurs; this presumably helps to stabilize the induced state. Similarly, Lys98 and Glu150 make a strong, favorable interaction in the induced state and a weaker one in the TetR:DNA complex. Substitution of either residue reduces, but does not abolish, inducibility and has little effect on the affinity of TetR for Tc.[34] As a last example, Arg104 and Asp178′ form a salt bridge in TetR:Tc, but are separated in the TetR:DNA complex; the Asp178Ala mutation reduces but does not abolish TetR inducibility.[28,34]

The allosteric transition: simulations of TetR:Tc:DNA

Structural changes

To characterize the first stages of the allosteric transition, we added $[Tc:Mg]^+$ to each monomer of the TetR:DNA complex, forming a non-equilibrium TetR:Tc:DNA state. We simulated the resulting structural adjustments over 2 ns of MD. The actual allosteric transition is likely to occur on a much longer timescale too expensive to simulate. Therefore, to induce a more rapid transition while preserving as many features as possible of the natural transition, we performed a "targeted molecular dynamics", biasing the simulation in the following way. A harmonic force was applied to residues within 3 Å of either Tc, pulling them towards the positions they occupy in the induced TetR:Tc complex. The pulling force was weak, with a force constant of 0.01 kcal/mol/Å^2. The force affects 29 amino acids and 474 atoms per monomer. For these 474 atoms, the rmsd between the starting structure and the target structure is about 2 Å. Therefore, the energy initially stored in the biasing restraint is about 0.04 kcal/mol per atom, 7% of the mean kinetic energy of an atom at room temperature. Within 2–3 ps, the total restraint energy has decreased to about 8 kcal/mol, or less than 0.01 kcal/mol per restrained atom. Only for nine amino acids that undergo significant conformational shifts upon induction is the restraint energy larger than 0.01 kcal/mol per atom. These include His64, His100, Thr103, Arg104, Pro105, and Glu147′. The other 5628 protein atoms, including the DNA-binding interface and the regions linking it to the Tc binding sites, were not subjected to any biasing force.

With this approach, we impose structural changes close to the Mg^{2+} ion, then observe how the changes propagate to the DNA-binding interface and affect the protein:DNA binding affinity. Due to the biasing forces, structural changes around the $[Tc:Mg]^+$ occur very early (Fig. 6). In monomer 1, the peptide group connecting Leu101 and Gly102 flips by 180° within about 20 ps, and a hydrogen bond forms between the CO of His100 and the NH of Thr103 (Fig. 6). Thus, the last turn of helix 6 shifts into the expected, type II beta-turn conformation within about 30 ps. Glu147′ from helix 8 of the other monomer also shifts to coordinate the Mg^{2+} ion. In monomer 2, the His100′–Thr103′ hydrogen bond also forms, after about 30–40 ps. The Leu101′–Gly102′ peptide group makes several transient and incomplete flips between about 10 and 30 ps, then settles back into its original orientation. Thus, a type I beta-turn forms in monomer 2 instead of the type II turn seen in monomer 1 and in the induced state. Presumably, monomer 2 would eventually convert to a type II turn if the simulation were long enough.

With His100, Thr103, and Glu147′ all moving towards the Mg^{2+} ion, α6 and α8 are pulled closer together. Since α8 belongs to the rigid core, α6 and the 6/7 loop are the ones to shift. After this move, the structure is further stabilized by a salt bridge between residues Asp178 (from α9) and Arg104′ (from the 6/7 loop).

The shift of α6 disrupts its contacts with the adjacent α4 and eventually pulls α4 along with it. Indeed, these helices interact through a cluster of hydrophobic side chains, illustrated (for monomer 1 only) in Fig. 7. When α6 shifts, at $t = 15$ ps, the cluster is seen to partly separate (Fig. 7, middle panel). After another 80 ps, α4 shifts to reform its close contact with helix 6. The hydrophobic contacts are reflected

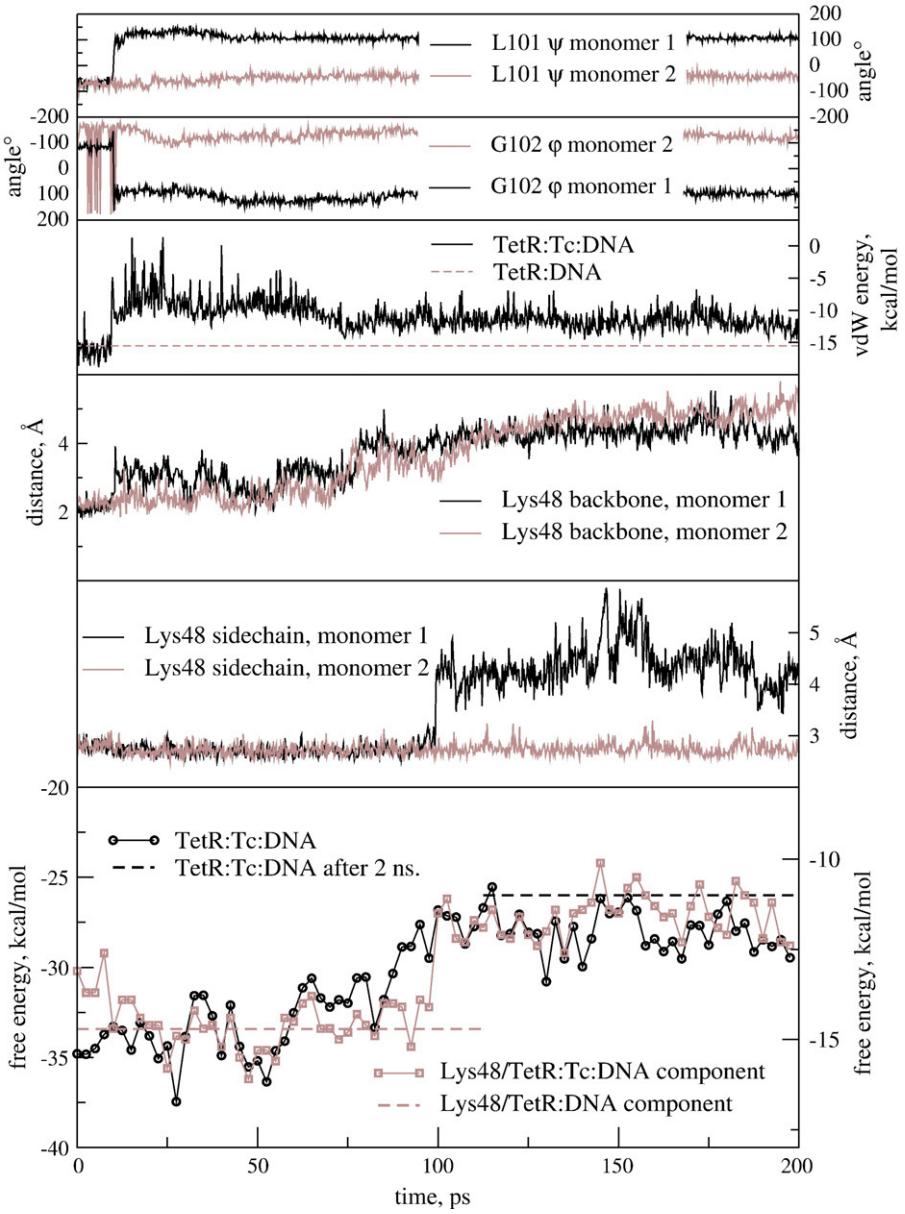


Fig. 6. Targeted MD trajectory. Several descriptors are plotted as a function of time. Top two panels: backbone angles for residues 101 and 102, illustrating the formation of the 100–103 type II beta turn. In monomer 1, the turn forms at $t = 15$ ps. In monomer 2, the backbone fluctuates, then a type I beta turn forms. Third panel: van der Waals interaction energy between helices 4 and 6. The energy increases abruptly at about $t = 15$ ps, then relaxes gradually back to its initial value as helix 4 shifts back towards helix 6. Fourth and fifth panels: the distance between Lys48 and the DNA in each monomer. The backbone breaks away from the DNA at $t = 15$ ps in monomer 1 and later in monomer 2. The side chain in monomer 1 breaks away at about $t = 100$ ps. Bottom panel: TetR:DNA binding free energy, estimated from the dielectric continuum model (see the text). Horizontal lines to guide the eye represent (left) the Lys48 contribution at the beginning of the simulation and (right) the total TetR contribution at the end of the simulation ($t = 2000$ ps).

in the van der Waals interaction energy between α4 and α6, shown in Fig. 6 as a function of time. The motion of α6 away from helix α4 appears as an abrupt, 10 kcal/mol loss of favorable interaction energy. This energy is regained abruptly after 100 ps; it then fluctuates strongly for a few picoseconds, then relaxes gradually back to its original value over a period of about 500 ps (Fig. 6).

Helix 4 transfers the structural changes all the way to the DNA-binding interface, 24 Å away from the

Fig. 7. Schematic view of the allosteric transition. Top: initial structure. Helices and selected groups are labelled, including Lys48 and His100. Lys48 hydrogen bonds to the DNA are shown. A hydrophobic cluster involving helix 4 and helix 6 side chains is shown (white spheres). Middle: structure after 70 ps. Important helix motions are highlighted. A Lys48 backbone hydrogen bond has been lost, and the hydrophobic contacts between helices 4 and 6 have been broken (schematized by a thick dashed line). Bottom: structure after 500 ps. Helix 4 has swung in a pendulum motion to restore its contacts with helix 6. Lys48 in monomer 1 has lost its hydrogen bonds to the DNA.

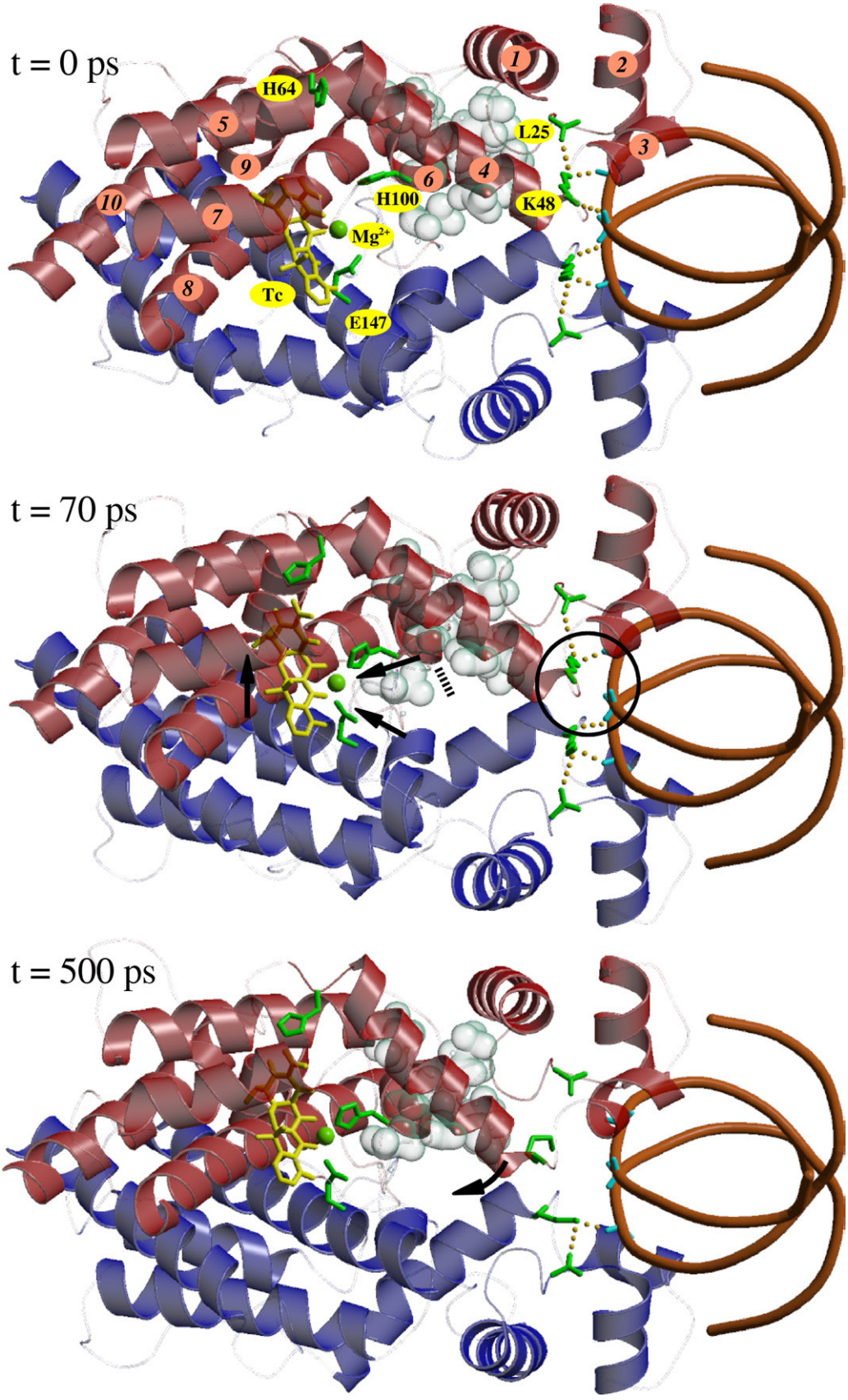


Fig. 7 (*legend on previous page*)

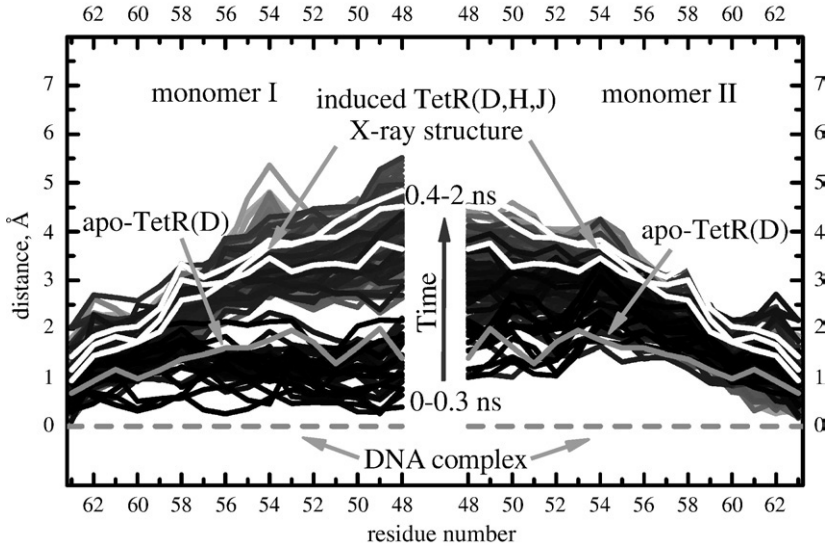


Fig. 8. Mobility of helix 4 during the targeted MD, characterized by the distance of each amino acid to its position in the TetR:DNA complex (zero distance; dashed line). Results for both TetR monomers are shown. The apo-TetR crystal structure (PDB code 1BJZ) is shown as a continuous grey line, with distances in the range 1–2 Å (deviation of each atom from its position in the TetR:DNA structure). Three induced TetR:Tc crystal structures are shown as white lines: TetR(D), TetR(H), and TetR(J). Here, the pendulum-like rotation of helix 4 is clearly seen. The C-terminus stays close to the TetR:DNA geometry, while the N-terminus is 3–4 Å from the TetR:DNA geometry. The amplitude is smaller in the TetR(D) structure (lowest white curve), because its DBD is involved in extensive crystal contacts. In all the crystal structures, both monomers are equivalent. MD conformations are shown with a grey scale that goes from black (beginning of the simulation; non-induced geometry) to light grey (end of the simulation; induced geometry). Agreement with experiment is very good. However, the simulations reveal a degree of mobility of helix 4 that is not evident from the endpoint, X-ray structures. In the simulations, the amplitude of the pendulum motion is somewhat smaller in monomer 2. This is a result of the finite simulation length (see the text).

Mg^{2+} ion. Indeed, the C-terminus of helix 4 is held in place by interactions in the core domain, in front of Tc, involving His64, Asn82, and Gln116. These interactions anchor the C-terminus, so that the shift of helix 4 takes the form of a pendulum motion. The largest displacement is for its N-terminal residue, Lys48.

Thus, at the precise moment when the Leu101-Gly102 peptide group flips in monomer 1, the DNA interface begins to change. At $t = 15$ ps, the backbone of Lys48 moves away from the DNA by about 1 Å (Figs. 6 and 7). It continues to adjust over a period of 100 ps, increasing its distance from the Gua10 ribose ring from 2 Å up to about 4 Å (distance measured from the Lys48 backbone N). The same adjustment occurs in monomer 2, with a slight delay: monomer 2 lags behind monomer 1 by about 20 ps (Fig. 6). Once the Lys48 backbone has adjusted in both monomers, at around $t = 100$ ps, the Lys48 side chain in monomer 1 breaks abruptly away from the DNA, losing its salt bridge to the Gua10 phosphate group. This is the moment when the α4–α6 interaction energy is regained in monomer 1 (Fig. 6). In monomer 2, the Lys48′ side chain maintains its salt bridge to the DNA throughout the simulation, probably due to the limited simulation length (2 ns).

The evidence provided by the present non-equilibrium simulation is in a sense anecdotal. Only one simulation was performed, and the initial changes, around the Mg^{2+} ion, were imposed at the outset. Nevertheless, it is notable that with weak restraints acting in a very limited region (essentially, the nine amino acids closest to the $[Tc:Mg]^{+}$), structural changes propagate within a short time to the DNA-binding interface, over 20 Å away, in both monomers, through the pulling of α4 towards α6. No part of α4 is subjected to any restraints. The C-terminal anchoring and the mechanical rigidity of α4 are enough to transmit the motions to Lys48, which is almost 12 Å from any of the amino acids subjected to biasing restraints (distance of 11.7 Å between its backbone and the Thr103 side chain).

The helix motions observed here can be compared to several available crystal structures. Figure 8 shows the shift that occurs between the non-induced and induced conformations for both monomers. The non-induced conformation is represented by the TetR:DNA complex and by the apo-TetR structure. The induced state is represented by the TetR(D) structure, as well as the H-class structure reported here and an unpublished J-class structure. The crystal structures show a simple, pendulum motion of the helix, as discussed above. The motion has a larger amplitude in TetR(H) and TetR(J) than in TetR(D). This is because the DBD is constrained by extensive crystal contacts in the TetR(D) X-ray structure. The MD conformations are in good agreement: the conformations sampled early in the trajectory resemble, on average, the non-induced conformation; those sampled late in the trajectory resemble, on average, the induced conformations, especially those in TetR(H) and TetR(J). The pendulum motion in the simulation is slightly less

pronounced in monomer 2 than in monomer 1. The agreement with experiment is a strong indication that the MD trajectory, despite its limited duration (2 ns), has sampled the key steps of the allosteric transition. At the same time, the simulations reveal a complexity that is not apparent in the X-ray structures: the internal structure of α4 fluctuates and bends considerably during the transition, deviating from the more linear geometry seen in the endpoint states.

Free-energy changes

The importance of Lys48 in the allosteric transition is further supported by an analysis of the TetR:DNA binding free energy and its evolution during the simulation. Indeed, the MD structures were analyzed with the continuum electrostatic model, as above. Figure 6f shows the estimated TetR:DNA binding free energy as a function of simulation time. The binding free energy starts to increase at $t = 60$ ps, then increases abruptly at $t = 90$ ps, immediately before the Lys48 side chain breaks away from the DNA. The direct contribution of Lys48 increases at the moment when it loses its salt bridge to Gua10, at $t = 100$ ps. The overall loss of TetR:DNA binding free energy upon adding $[Tc:Mg]^+$ is about 12 kcal/mol after 500 ps. The Lys48 contribution is about 5 kcal/mol, almost half of the total effect. The flanking residues, Lys46, Asn47, and Arg49, make up the remainder of the affinity loss (not shown). Thus, the breaking away of Lys48 can be considered a key step along the allosteric pathway.

Concluding Discussion

The TetR allosteric mechanism is of general interest. TetR mediates the main mechanism of resistance against the clinically important tetracycline antibiotics. Due to its tight control by Tc, it is also widely used in molecular and cell biology as a sensitive switch for target gene regulation.[4–6] Finally, TetR belongs to the large family of regulatory proteins of the helix–turn–helix class[25] and is a paradigm for the allosteric control of DNA binding.

A detailed model for the allosteric mechanism has been inferred from the TetR:Tc and TetR:DNA crystal structures. In this model, $[Tc:Mg]^+$ binds to the TetR:DNA complex and induces the necessary structural changes. The Tc-coordinated Mg^{2+} ion pulls helix 6 inwards, which in turn pulls helix 4. The pendulum motion of α4 and α4′ pulls the helices 3 and 3′ apart and leads to TetR:DNA dissociation.

A second model was proposed recently, based on an MD simulation study.[13,14] It assumes that Mg^{2+} is present in the absence of Tc, coordinated by Asp157 from the 8/9 loop. Tc binding is proposed to displace Asp157, leading to a cascade of structural changes and DNA dissociation. Experimental data call this model into question, as discussed. The flexible loop 156–163 is not conserved in TetRs, and mutation of Asp157 to Ala does not affect TetR function.[17] The loop is disordered in all but two TetR crystal structures,[8,15] where its conformation is not compatible with Mg^{2+} coordination. Most importantly, TetR structures with bound Mg^{2+} but no Tc have never been observed; Mg^{2+} is always seen to bind in complex with Tc.

A third model can probably be ruled out, where $[Tc:Mg]^+$ binds to the apo-Tet repressor, not the TetR:DNA complex. This would imply that TetR first dissociates transiently from the DNA—a rare event, given the small TetR:DNA dissociation constant. A rapid bacterial response for resistance to Tc probably requires that Tc bind initially to TetR:DNA.

The present simulations and crystal structure provide further details on TetR:DNA binding and allow us to test the induction model inferred from the earlier X-ray structures.[7] The targeted MD simulation, in particular, provides direct support for the main features of the model. $[Tc:Mg]^+$ is added to the TetR:DNA complex and a biasing force is applied, pulling nearby side chains towards their induced conformations.

His100 and Thr103 shift to coordinate the Mg^{2+} ion, and α6 is pulled inwards, towards the rigid core. Remarkably, the pendulum motion of α4 inferred from the X-ray structures then occurs spontaneously in the simulation (Figs. 7 and 8), so that Lys48 at its N-terminus breaks away from the DNA. Similar motions occur in monomer 2, but the sequence of events is less complete, probably because of the limited simulation length.

The simulations directly support the idea that the Mg^{2+} ion itself triggers the allosteric transition, with Tc acting mainly as a specific Mg^{2+} carrier. Consistent with this, modified Tcs that bind without an associated ion do not induce TetR.[8] Similarly, TetR mutations that disrupt the Mg^{2+} coordination, such as Thr103Ala, lead to non-inducible TetRs. Anhydrotetracycline appears to be an exception to the first rule, since it was found to induce TetR without the need for an associated Mg^{2+}. However, its chemical structure is rather different from that of classical Tcs; furthermore, it may be that a bound sodium ion participates in the observed induction, as discussed above.

Although the simulations broadly agree with the original model, they reveal several new insights. First, the main consequence of the α4 pendulum motion appears to be the dissociation of Lys48 from the DNA. Lys48 breaking away reduces the TetR:DNA binding free energy drastically (Fig. 6), so that TetR can then detach itself spontaneously. The van der Waals interactions between helices 4 and 6, transiently lost at $t = 15$ ps, amount to about 5 kcal/mol, very close to the DNA binding free energy lost by each Lys48 side chain during the transition (at around $t = 100$ ps). Interestingly, Lys48 is not part of the DBD. However, it is highly conserved in regulatory proteins of the TetRs class (Fig. 2; strict conservation in 77% of all sequences; homologous residues Arg, His in over 10% of all sequences[25]). Note that if Lys48 can trigger TetR dissociation, then helices 4 and 4′ do not need to push the DBDs strongly out of the DNA grooves. Rather, the DBDs

can relax fully into their induced orientations after the TetR has dissociated from the DNA. This picture differs from the initial, X-ray-based model.

A second insight concerns the mechanical rigidity of the DBD/α4 unit. The internal cohesion of this unit is remarkable: helices 1–4 appear to move practically as a rigid body when induced and non-induced crystal structures are compared (Fig. 4b). This cohesion leads to the dynamical correlations represented in Fig. 4a. The rigidity of this unit could facilitate rapid transitions between the induced and non-induced orientations, possibly providing a selective advantage. It suggests that helix 4 could be considered part of the same domain as helices 1–3, a striking departure from the usual view of TetR quaternary structure.

A third insight concerns the motion of α4, which is more complex than a simple pendulum. The conformations sampled early in the trajectory resemble, on average, the non-induced X-ray structure; those sampled late in the trajectory resemble, on average, the induced X-ray structures [especially those of TetR(H) and TetR(J); Fig. 8]. However, the individual MD conformations vary widely, revealing significant fluctuations and bending of α4 during the course of the allosteric transition. Thus, although a certain mechanical rigidity is needed to transmit the binding signal from one end of the helix to the other, α4 is also elastic.

Fourth, in our simulations, the allosteric changes are more complete in one monomer (monomer 1). In monomer 2, despite the biasing forces, residues 100′–103′ remain in a type I beta-turn conformation, instead of the expected type II turn. The pendulum motion of α4′ also has a slightly smaller amplitude than in monomer 1, and Lys48′ loses just one of its hydrogen bonds to the DNA. These differences between monomers are presumably due to the limited simulation length. The full induction seen in monomer 1 is consistent with the idea that Mg^{2+} and its binding pocket are the initiators of the allosteric transition. The differences between monomers are compatible with the existence of a half-induced state seen crystallographically, where $[Tc:Mg]^+$ is bound to just one monomer[8] and the other monomer has the non-induced conformation. Differences between monomers are presumably facilitated by the known absence of cooperativity between Tc binding sites.[11]

Overall, the simulations are consistent with a broad body of experimental data. They generally support the model of induction inferred earlier from the X-ray data. However, they provide important new details, some of which may be relevant for other regulatory proteins.

Materials and Methods

Molecular dynamics simulations

The starting structure for the TetR:Tc simulation was a 2.5 Å resolution crystal structure of the *Escherichia coli* class D Tet repressor homodimer with two bound $[Tc:Mg]^+$ molecules (PDB entry 2TRT). The starting structure for TetR:DNA was a 2.5 Å resolution crystal structure of the *E. coli* class D TetR homodimer bound to its operator DNA segment (PDB entry 1QPI). For both complexes, the disordered loop, residues 156–164, was taken from a 2.7 Å resolution crystal structure of the same Tet repressor with bound 7-chlorotetracycline:Mg^{2+} (PDB entry 1BJY). The loop was positioned by superimposing the backbone atoms of residues 155 and 165. Hydrogen atoms were constructed with ideal stereochemistry. Protonation states of histidines were assigned by visual inspection, except for His64 in the Tc binding site, which was assigned earlier through extensive simulations.[15] Orientations of His, Asn, and Gln side chains in the Tc binding site as well as the tautomeric state of Tc were assigned based on the previous simulations.[15] In addition to crystal water molecules, an 80 Å cubic box of water was overlaid, and water molecules overlapping the protein or ligand were removed. The final TetR:Tc model contained two bound $[Tc:Mg]^+$ complexes, 414 protein residues, and 14,158 water molecules, including 170 crystal water molecules. Periodic boundary conditions were assumed; that is, the entire 80 Å box was replicated periodically in all directions. All long-range electrostatic interactions were computed efficiently by the particle mesh Ewald method.[35] Six sodium counterions were included to reduce the formal charge of the system. The final TetR:DNA model contained 406 protein residues, 30 nucleotides (15 per strand), and 20,959 water molecules, including 82 crystal water molecules. Forty-two sodium counterions were included to reduce the formal charge. MD simulations were performed at constant room temperature and pressure with a Nosé–Hoover algorithm[36,37] following 5 ps of thermalization. During the first 5 ps of the simulation, a 5 kcal/mol/Å^2 harmonic restraint was applied to heavy atoms, maintaining them in their starting, crystal positions. The CHARMM22 force field was used for the protein[38] and DNA[39] and the TIP3P model for water.[40] Tc was described with the force field developed previously.[18] Calculations were done with the CHARMM program.[41]

We also performed MD simulations for a non-equilibrium state: the ternary complex between the protein, Tc, and DNA. We started from the 1QPI experimental structure used above. To position two $[Tc:Mg]^+$ complexes and their six coordinated water molecules within the TetR:DNA structure, we used an equilibrated TetR:Tc structure from our previous simulations.[15] Residues within 7 Å of each Tc were used as a reference and superimposed onto the TetR:DNA complex, thus defining the Tc position within the ternary complex. For residues within 3 Å of either Tc, weak geometric restraints were then applied, with a force constant of 0.01 kcal/mol/Å^2, pulling each atom towards its expected position in the equilibrium, binary, TetR:Tc complex.

Poisson–Boltzmann free energy method

The electrostatic contribution to the TetR:DNA binding free energy was obtained from a dielectric continuum model by solving the Poisson–Boltzmann equation numerically. The electrostatic free energy was computed for the TetR:DNA complex and for the separate DNA and protein. The Poisson–Boltzmann equation was solved using a cubic grid and a finite difference algorithm, implemented in CHARMM (grid size, 96 Å; grid spacing, 0.4 Å).[42] A physiological ionic strength was used (0.1 M monovalent ion concentration). The structure of the protein:DNA complex was taken from the MD simulations

(above). Calculations were performed for multiple structures, sampled every 16 ps along the equilibrated 4.0-ns trajectory, for a total of 250 structures. The separate DNA and protein structures were obtained by simply discarding the unwanted partner. Thus, structural relaxation on separating protein and DNA was not explicitly included (although it is implicit in the dielectric constant). The solvent dielectric constant was set to 80. We used either 4 or 6 for the solute dielectric constant. Pairwise interaction energies between groups were also computed, as follows. For a particular group A, such as a TetR side chain, and another group B (such as the DNA), the interaction energy has the form $G_{AB}=1/2\sum_{j\in B}q_jV_j$, where the sum is over the atoms j belonging to B, q_j is an atomic charge, and V_j is the potential produced at j by the partial charges of A.

Crystallization, X-ray diffraction, and structure refinement

TetR of class H, TetR(H), was co-crystallized with 5a,6-anTc, and Mg^{2+} according to protocols previously described for other tetracycline complexes of TetR.[8,43,44] Crystals were obtained at room temperature by vapor diffusion against precipitant solutions of 20% PEG 1500 and 0.1 M Na Hepes, pH 7.5, with protein concentrations of 3.3 mg/ml. The complex of TetR(H)/[Mg:anTc] crystallized in space group $C222_1$, which differs from structures of TetRs from other classes reported recently. The X-ray diffraction data were collected at the synchrotron beam line X13 of the EMBL outstation (DESY, Hamburg). A flash-frozen crystal was measured in a gaseous N_2 stream at 100 K. Data were reduced using the HKL package[45] and further processed with the programs of the CCP4 suite.[46] Initial phasing by molecular replacement (AMORE program[47]) and model building were based on an earlier TetR(D) structure, where the bound tetracycline was removed (PDB entry 2TCT). Refinement was performed with the program REFMAC5.[48] The inspection of geometric parameters was performed using the program PROCHECK.[49] Details and statistics of data collection and structure refinement are given in Table 1. In addition, an isomorphously crystallized complex of [Mg:anTc] was found with TetR of class J. Diffraction data with 2.1 Å resolution were collected and refinement is in progress.

Protein Data Bank accession numbers

Coordinates and structure factors have been deposited in the PDB with the accession numbers 2vpr (coordinates) and r2vprsf (structure factors).

Acknowledgements

Support was provided by the Nuclear Toxicology program of the Commmissariat pour l'Énergie Atomique and the Centre National pour la Recherche Scientifique (to T.S.). An Egide PhD fellowship was awarded to A.A. We thank Martin Karplus for the CHARMM program. Some of the simulations were performed at the CINES supercomputing center of the French Ministry of Education.

Supplementary Data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jmb.2008.03.022

References

1. Chopra, I. & Roberts, M. (2001). Tetracycline antibiotics: mode of action, applications, molecular biology, and epidemiology of bacterial resistance. *Microbiol. Mol. Biol. Rev.* **65**, 232–260.
2. Lederer, T., Takahashi, M. & Hillen, W. (1995). Thermodynamic analysis of tetracycline-mediated induction of Tet repressor by quantitative methylation protection assay. *Anal. Biochem.* **232**, 190–196.
3. Takahashi, M., Altschmied, L. & Hillen, W. (1986). Kinetic and equilibrium characterization of the Tet repressor–tetracycline complex by fluorescence measurements. *J. Mol. Biol.* **187**, 341–348.
4. Fussenegger, M. (2001). The impact of mammalian gene regulation concepts on functional genomic research, metabolic engineering, and advanced gene therapies. *Biotech. Prog.* **17**, 1–51.
5. Berens, C. & Hillen, W. (2003). Gene regulation by tetracyclines: constraints of resistance regulation in bacteria shape TetR for application in eukaryotes. *Eur. J. Biochem.* **270**, 3109–3121.
6. Berens, C., Lochner, S., Lober, S., Usai, I., Schmidt, A., Drueppel, L. *et al.* (2006). Subtype selective tetracycline agonists and their application for a two-stage regulatory system. *ChemBioChem*, **7**, 1320–1324.
7. Orth, P., Schnappinger, D., Hillen, W., Saenger, W. & Hinrichs, W. (2000). Structural basis of gene regulation by the tetracycline inducible Tet repressor–operator system. *Nat. Struct. Biol.* **7**, 215–219.
8. Orth, P., Saenger, W. & Hinrichs, W. (1999). Tetracycline-chelated Mg^{2+} ion initiates helix unwinding in Tet repressor induction. *Biochemistry*, **38**, 191–198.
9. Baumeister, R., Helbl, V. & Hillen, W. (1992). Contacts between Tet repressor and *tet* operator revealed by new recognition specificities of single amino acid replacement mutants. *J. Mol. Biol.* **226**, 1257–1270.
10. Lederer, T., Kintrup, M., Takahashi, M., Sum, P., Ellestad, G. & Hillen, W. (1996). Tetracycline analogs affecting binding to Tn10-Encoded Tet repressor trigger the same mechanism of induction. *Biochemistry*, **35**, 7439–7446.
11. Scholz, O., Schubert, P., Kintrup, M. & Hillen, W. (2000). Tet repressor induction without Mg^{2+}. *Biochemistry*, **39**, 10914–10920.
12. Chopra, I. (1994). Tetracycline analogs whose primary target is not the bacterial ribosome. *Antimicrob. Agents Chemother.* **38**, 637–640.
13. Lanig, H., Othersen, O. G., Beierlein, F., Seidel, U. & Clark, T. (2006). Molecular dynamics simulations of the tetracycline–repressor protein: the mechanism of induction. *J. Mol. Biol.* **359**, 1125–1136.
14. Lanig, H., Othersen, O. G., Beierlein, F., Exner, T. E. & Clark, T. (2006). Structural changes and binding characteristics of the tetracycline–repressor binding site on induction. *J. Med. Chem.* **49**, 3444–3447.
15. Aleksandrov, A., Proft, J., Hinrichs, W. & Simonson, T. (2007). Protonation patterns in tetracycline:Tet repressor recognition: simulations and experiments. *ChemBioChem*, **8**, 675–685.

16. Klotzsche, M., Goeke, D., Berens, C., Hillen, W. & Muller, Y. A. (2007). Efficient and exclusive induction of Tet repressor by the oligopeptide Tip results from co-variation of their interaction site. *Nucleic Acids Res.* **35**, 3945–3952.
17. Scholz, O., Kintrup, M., Reich, M. & Hillen, W. (2001). Mechanism of Tet repressor . by tetracyclines: length compensates for sequence in the α8–α9 loop. *J. Mol. Biol.* **310**, 979–986.
18. Aleksandrov, A. & Simonson, T. (2006). The tetracycline: Mg^{2+} complex: a molecular mechanics force field. *J. Comput. Chem.* **13**, 1517–1533.
19. Aleksandrov, A. & Simonson, T. (2008). Molecular dynamics simulations of the 30S ribosomal subunit reveal a preferred tetracycline binding site. *J. Am. Chem. Soc.* **130**, 1114–1115.
20. Hendsch, Z. & Tidor, B. (1999). Electrostatic interactions in the GCN4 leucine zipper: substantial contributions arise from intramolecular interactions enhanced on binding. *Protein Sci.* **8**, 1381–1392.
21. Kollman, P., Massova, I., Reyes, C., Kuhn, B., Huo, S., Chong, L. *et al.* (2000). Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. *Acc. Chem. Res.* **33**, 889–897.
22. Simonson, T., Archontis, G. & Karplus, M. (2002). Free energy simulations come of age: the protein–ligand recognition problem. *Acc. Chem. Res.* **35**, 430–437.
23. Simonson, T. (2001). Free energy calculations. In *Computational Biochemistry & Biophysics* (Becker, O., Mackerell, A., Jr, Roux, B. & Watanabe, M., eds), Marcel Dekker, New York; chapt. 9.
24. McCammon, J. (2005). Computation of noncovalent binding affinities. In *Theory and Applications of Computational Chemistry* (Dykstra, C., Frenking, G., Kim, K. & Scuseria, G., eds), pp. 41–46, Elsevier, Amsterdam.
25. Ramos, J. L., Martnez-Bueno, M., Molina-Henares, A. J., Teran, W., Watanabe, K., Zhang, X. *et al.* (2005). The TetR family of transcriptional repressors. *Microbiol. Mol. Biol. Rev.* **69**, 326–356.
26. Finn, R. D., Mistry, J., Schuster-Bckler, B., Griffiths-Jones, S., Hollich, V., Lassmann, T. *et al.* (2006). Pfam: clans, web tools and services. *Nucleic Acids Res.* **34**, D247–D251.
27. Higgins, J. D. & Gibson, T. J. (1995). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, 4673–4680.
28. Müller, G., Hecht, B., Helbl, V., Hinrichs, W., Saenger, W. & Hillen, W. (1995). Characterization of non-inducible Tet repressor mutants suggests conformational changes necessary for induction. *Nat. Struct. Biol.* **2**, 693–703.
29. European Bioinformatics Institute (2005). Pisa Server: protein interfaces, surfaces and assemblies. http://www.ebi.ac.uk/msd-srv/prot_int/pistart.html.
30. Krissinel, E. & Henrick, K. (2005). Detection of protein assemblies in crystals. pp. 163–174, Springer-Verlag, Berlin.
31. Diamond, R. (1990). On the use of normal modes in thermal parameter refinement: theory and application to the bovine pancreatic trypsin inhibitor. *Acta Crystallogr. A*, **46**, 425–435.
32. Howlin, B., Moss, D. & Harris, G. (1990). Segmented anisotropic refinement of bovine ribonuclease A by the application of the rigid-body TLS model. *Acta Crystallogr. A*, **45**, 851–861.
33. Navizet, I., Cailliez, F. & Lavery, R. (2004). Probing protein mechanics: residue-level properties and their use in defining domains. *Biophys. J.* **87**, 1426–1435.
34. Saenger, W., Orth, P., Kisker, C., Hillen, W. & Hinrichs, W. (2000). The tetracycline repressor: a paradigm for a biological switch. *Angew. Chem. Int. Ed.* **39**, 2042–2052.
35. Darden, T. (2001). Treatment of long-range forces and potential. In (Becker, O., Mackerell, A., Jr, Roux, B. & Watanabe, M., eds), Marcel Dekker, New York; chapt. 4.
36. Nose, S. (1984). A unified formulation of the constant temperature molecular dynamics method. *J. Chem. Phys.* **81**, 511–519.
37. Hoover, W. (1985). Canonical dynamics: equilibrium phase-space distributions. *Phys. Rev. A*, **31**, 1695–1697.
38. Mackerell, A., Bashford, D., Bellott, M., Dunbrack, R., Evanseck, J., Field, M. *et al.* (1998). An all-atom empirical potential for molecular modelling and dynamics study of proteins. *J. Phys. Chem. B*, **102**, 3586–3616.
39. Mackerell, A., Wiorkiewicz-Kuczera, J. & Karplus, M. (1995). An all-atom empirical energy force-field for the study of nucleic acids. *J. Am. Chem. Soc.* **117**, 11946–11975.
40. Jorgensen, W., Chandrasekar, J., Madura, J., Impey, R. & Klein, M. (1983). Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926–935.
41. Brooks, B., Bruccoleri, R., Olafson, B., States, D., Swaminathan, S. & Karplus, M. (1983). Charmm: a program for macromolecular energy, minimization, and molecular dynamics calculations. *J. Comput. Chem.* **4**, 187–217.
42. Im, W., Beglov, D. & Roux, B. (1998). Continuum solvation model: computation of electrostatic forces from numerical solutions to the Poisson–Boltzmann equation. *Comput. Phys. Commun.* **111**, 59–75.
43. Hinrichs, W., Kisker, C., Duvel, M., Muller, A., Tovar, K., Hillen, W. & Saenger, W. (1994). Structure of the Tet repressor–tetracycline complex and regulation of antibiotic resistance. *Science*, **264**, 418–420.
44. Kisker, C., Hinrichs, W., Tovar, K., Hillen, W. & Saenger, W. (1995). The complex formed between tetracycline repressor and tetracycline-Mg^{2+} reveals mechanism of antibiotic resistance. *J. Mol. Biol.* **247**, 260–280.
45. Otwinowski, Z. & Minor, W. (1997). Processing of x-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326.
46. Collaborative Computational Project Number 4 (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **50**, 760–776.
47. Navaza, J. (2001). Implementation of molecular replacement in AMoRe. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **57**, 1367–1372.
48. Murshudov, G., Vagin, A. & Dodson, E. (1997). Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **53**, 240–255.
49. Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thorton, J. (1993). PROCHECK: a program to check stereochemical quality of protein structures. *J. Appl. Crystallogr.* **26**, 283.

Tet repressor induction by tetracycline: a molecular dynamics, continuum electrostatics, and crystallographic study

Alexey Aleksandrov[1], Linda Schuldt[2,3], Winfried Hinrichs[2,*] and Thomas Simonson[1,*]

[1]Laboratoire de Biochimie (CNRS UMR7654), Department of Biology, Ecole Polytechnique, 91128 Palaiseau, France.
[2]Institute for Biochemistry, Department of Molecular Structural Biology, University of Greifswald, Felix-Hausdorff-Str. 4, D-17489 Greifswald, Germany.

[3]Present address: EMBL Hamburg, c/o DESY, Building 25A, Notkestrasse 85, D-22603 Hamburg, Germany.

Supplementary material

Relation between reporter gene activity and binding free energies

Baumeister et al report β-galactosidase activity and its repression by TetR, for the native and mutant TetRs (Baumeister, R., Helbl, V., and Hillen, W., 1992, J. Mol. Biol., 226:1257–1270). We seek a relation between the measured activities and the TetR:DNA binding free energy changes due to the point mutations. At thermodynamic equilibrium, the law of mass action can be written:

$$e^{-\beta \Delta G_0} = \frac{[P:DNA]_0}{[P]_0\ [DNA]_0}, \tag{1}$$

where ΔG_0 is the standard protein:DNA binding free energy, $[P:DNA]_0$ is the equilibrium concentration of protein:DNA complexes and $[DNA]_0$ and $[P]_0$ are the concentrations of free DNA and protein. We have defined $\beta = 1/kT$, where k is Boltzmann's

constant and T the temperature. The total concentration of protein, C_P, and of DNA, C_{DNA}, obey the relations:

$$[P:DNA]_0 + [P]_0 = C_P \tag{2}$$

$$[P:DNA]_0 + [DNA]_0 = C_{DNA} \tag{3}$$

Analogous expressions can be written for a mutant protein P':

$$e^{-\beta\ \Delta G_1} = \frac{[P':DNA]_1}{[P']_1\ [DNA]_1} \tag{4}$$

$$[P':DNA]_1 + [P']_1 = C_P \tag{5}$$

$$[P':DNA]_1 + [DNA]_1 = C_{DNA} \tag{6}$$

The total concentration of the mutant and native proteins and of DNA in both experiments are assumed to be the same.

The β-galactosidase activity A is assumed to be proportional to the free DNA concentration: $A \propto [DNA]$. Thus, for the ratio of native and mutant TetR activities, we have:

$$\frac{A_1}{A_0} = \frac{[DNA]_1}{[DNA]_0} \tag{7}$$

Equations (1–7) form a closed set.

We are mainly interested in the binding free energy difference $\Delta\Delta G = \Delta G_1 - \Delta G_0$. From Eqs. (1–7), it takes the form:

$$\Delta\Delta G = kT\ Log\left(\frac{A_{rel}\ ((2-A_{rel})\ K_0\ (C_P - C_{DNA}) + A_{rel}\ (B-1))}{1-2\ A_{rel} + K_0\ (C_P - C_{DNA}) + B}\right) \tag{8}$$

where

$$B = \sqrt{4K_0C_{DNA} + (1 + K_0(C_P - C_{DNA}))^2}$$

$K_0 = e^{-\beta\ \Delta G_0}$ is the association constant for the native protein; $A_{rel} = A_1/A_0$ is the β-galactosidase activity relative to the native protein. The equations can be simplified when $C = C_{DNA} = C_P$:

$$\Delta\Delta G = kT\ Log\left(\frac{A_{rel}^2\ (\sqrt{4K_0C+1} - 1)}{\sqrt{4K_0C+1} - 2A_{rel} + 1}\right) \tag{9}$$

We use the experimental association constant $K_0 = 10^{11}\ \mathrm{M}^{-1}$ for the native TetR protein. The concentrations of the DNA and protein in the β-galactosidase assay experiments, C_P and C_{DNA}, can be deduced from additional relations. Indeed, in the

absence of TetR, the activity was 100%. For the native protein the activity was 0.5%. Thus,

$$\frac{C_{DNA}}{[DNA]_0} = \frac{100\%}{0.5\%} \tag{10}$$

A 14% relative activity corresponds to a 10^3 ratio for the association constants or a $\Delta\Delta G$ of 4.1 kcal/mol. These relations, together with Eqs. (1–7) allowed us to deduce the experimental C_P and C_{DNA} to be the same, 0.44 μM. The figure below shows the relation between the binding free energy difference $\Delta\Delta G$ and the β-galactosidase activity. The activity in the absence of TetR protein was defined with 5% precision. When a mutant was present, the average uncertainty was 2%. Thus, the highest experimentally measured activity that could be distinguished from 100% was 93%, which corresponds to $\Delta\Delta G = 7.9$ kcal/mol (shown in the figure by dashed lines).

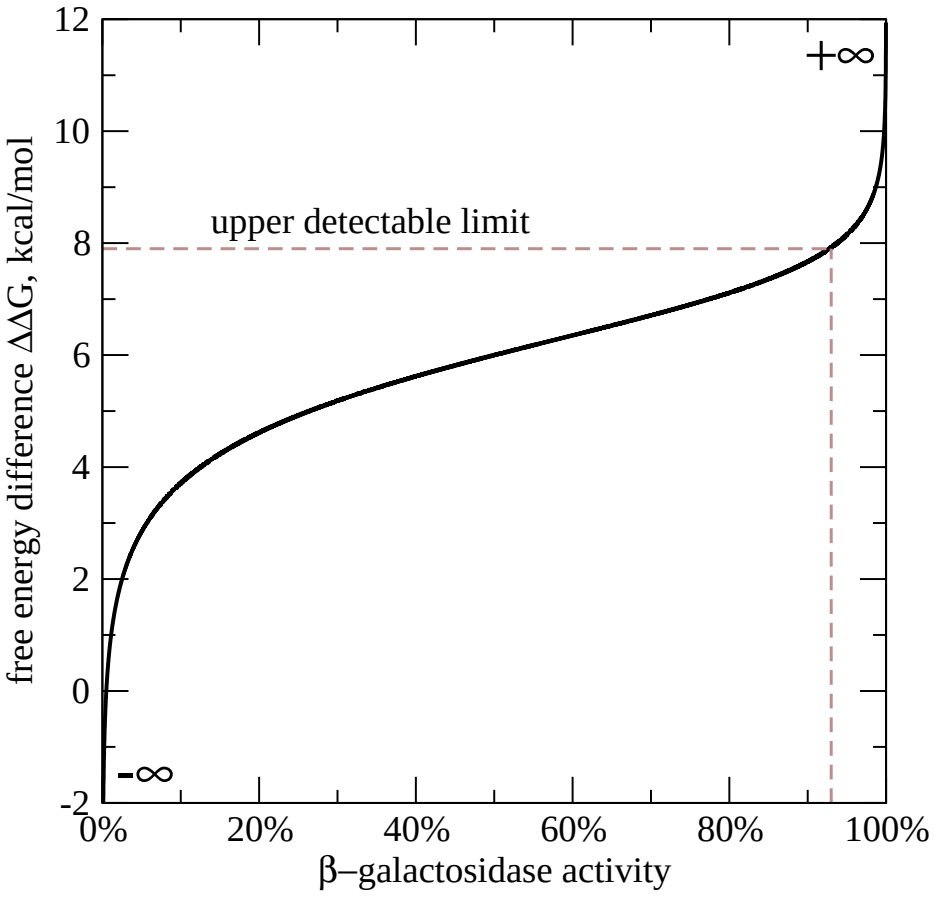

12
10
8
6
4
2
0
-2
+∞
-∞
upper detectable limit
free energy difference ΔΔG, kcal/mol
0%
20%
40%
60%
80%
100%
β–galactosidase activity

Figure 1:

Chapter 4

Tetracycline binding to the ribosome and Elongation Factor Tu

We use the developed Tc model to study Tc binding to the ribosome and to elongation factor Tu (Ef-Tu). Crystal structures of Tc bound to the *Thermus thermophilus* 30S subunit show the same primary Tc binding site (called TET1), with the strongest Tc electron density, close to the A-site, consistent with an inhibitory role. A secondary Tc-binding site, called TET5 was also observed in two structures. We have done molecular dynamics (MD) simulations of 30S ribosomal subunits to characterize Tc binding and help resolve the ambiguity regarding the number and strength of Tc binding sites. We have presented evidence for predominant binding to TET1, showing that other reported binding sites are weaker and not highly occupied at physiological Tc concentrations.

Recently, the crystal structure of a complex between elongation factor Tu (Ef-Tu) and Tc was solved, raising the question whether Tc's binding to Ef-Tu has a role in its inhibition of protein synthesis. We show that the direct contribution of

Ef-Tu to the free energy of Tc binding to the Ef-Tu:GDP:Mg complex is negligible; rather, the binding can be solely attributed to Tc interactions with the Mg ion and the GDP phosphate groups. We also show that Ef-Tu does not exhibit any binding preference for Tc over the non-antibiotic, 4-dedimethyl-Tc, and Ef-Tu does not bind the Tc analogue tigecycline, which is a potent antibiotic. Overall, our results support the idea that Ef-Tu is not the primary target of tetracycline.

Article 5: Molecular Dynamics Simulations of the 30S Ribosomal Subunit Reveal a Preferred Tetracycline Binding Site. Alexey Aleksandrov and Thomas Simonson. *J. Am. Chem. Soc.,* 2008, 130, 1114-1115

Article 6: Binding of tetracycline to elongation factor Tu, the Tet Repressor, and the ribosome: a molecular dynamics and continuum electrostatics study Alexey Aleksandrov and Thomas Simonson. submitted.

J|A|C|S
COMMUNICATIONS

Published on Web 01/10/2008

Molecular Dynamics Simulations of the 30S Ribosomal Subunit Reveal a Preferred Tetracycline Binding Site

Alexey Aleksandrov and Thomas Simonson*

Laboratoire de Biochimie (CNRS UMR 7654), Department of Biology, Ecole Polytechnique, Palaiseau, France

Received June 8, 2007; E-mail: thomas.simonson@polytechnique.fr

Tetracycline (Tc) and its analogues are important antibiotics affecting both gram-positive and gram-negative bacteria.[1,2] Tc inhibits protein synthesis by binding to the 30S subunit of the bacterial ribosome, preventing tRNA binding to the A-site.[2,3] Clinical use has led to several resistance mechanisms, including export of Tc from the cell and mutations in the bacterial rRNA. Much effort is being made to develop new Tcs that can evade resistance while effectively inhibiting protein synthesis.

Extensive biochemical and genetic studies[4−7] have given results that are somewhat difficult to interpret. Crystal structures of Tc bound to the *Thermus thermophilus* 30S subunit were determined by two groups.[8,9] Though nonpathogenic, *Thermus* is considered a valid model organism to understand ribosomal antibiotics.[10−12] Both groups found the same primary site (called TET1), with the strongest Tc electron density, close to the A-site, consistent with an inhibitory role. A Tc-bound Mg^{2+} ion was inferred from the electron density. Both groups reported a secondary Tc-binding site, called TET5. No Mg^{2+} ion was observed, but this is not conclusive, since electron density[8] was weaker in this region. One of the groups reported four additional Tc sites.[9] Both groups discussed whether multiple Tc binding sites can work cooperatively, and suggested that the primary binding site alone could explain Tc's bacteriostatic action.

We have done molecular dynamics (MD) simulations of 30S ribosomal subunits to characterize Tc binding and help resolve the ambiguity regarding the number and strength of Tc binding sites. We compared the binding free energies of Tc to the TET1, primary site and the TET5, secondary site. Other secondary sites were not considered, because they were unseen by one group[8] and are thus expected to have affinities lower than TET5. We benefit from a recent, high quality Tc force field[13] and from the maturity of current free energy methods.[14] We use structures obtained by MD, along with a well-established, continuum electrostatic (CE) method for the free energies.[14,15] The method is first parametrized and tested by computing the TET1 binding free energy differences between Tc and its analogues minocycline (Mc) and doxycycline (Dc), giving good agreement with experiment.[18,19] We also computed the free energy to introduce an Mg^{2+} cation in the TET1 and TET5 sites, using a more rigorous method, where the Mg^{2+} is gradually introduced during an MD simulation.[13,14e] CE gave good agreement with the more rigorous method; both methods predict that Mg^{2+} is pre-bound in each site in the absence of Tc. The free energy calculations then show that TET1 binding is indeed stronger than TET5 binding, as suggested (but not proven) by the weaker TET5 electron density,[8] resolving the previous ambiguity.

The simulations were done as follows. The starting point was the crystal structure of the 30S subunit with Tc bound to the TET1 and TET5 sites.[8] Simulations were done separately for both sites, including RNA and protein within a 26 Å-radius sphere centered on the Tc site, along with 1650−1850 water molecules and 80−100 sodium and choride counterions. Building models from the entire 70S ribosome,[3] instead of the 30S subunit, would presumably not affect our results significantly, since the closest atoms of the 50S subunit are over 24 Å from the centers of the TET1 and TET5 spheres used here. The medium outside each 26 Å sphere was modeled as a dielectric continuum with the dielectric constant of bulk water, $\epsilon = 80$. Protein, RNA, water, and ions were described by standard force fields;[16] the ligands were described by a force field developed recently.[13] For each site, we considered eight structural models, including neutral and zwitterionic Tc tautomers, with two possible orientations of the Tc acetamide group. For each site and model, MD was run for 5 nanoseconds. We also computed the free energy to introduce an Mg^{2+} cation in each site, using the rigorous, MD free-energy simulation method[13,14e] (see Supporting Information for more details). The MD structures were then used to compute the contribution of electrostatic interactions to the ligand/ribosome binding free energy. Nonelectrostatic contributions are expected to cancel approximately when the different models and binding sites are compared[14] and are neglected. Protein, RNA, and Tc (or Mc or Dc) were treated as one homogeneous dielectric medium, with a dielectric ϵ, and solvent as another, with a dielectric of 80. The free energy of the bound and unbound states was obtained by solving numerically the Poisson−Boltzmann equation of continuum electrostatics.[17] The unbound structures were obtained by removing Tc (or Mc or Dc or Mg^{2+}) from the bound structures. The dielectric ϵ was set to 8, to reproduce the experimental binding free energy differences between Tc, Mc, and Dc (see above and Supporting Information).

The free energy calculations indicate that Tc binds to the primary site TET1 in its zwitterionic tautomer. The co-bound Mg^{2+} interacts directly with three ribosomal phosphate groups (Figure 1A). The MD structure agrees with the crystal structure, with an rms deviation for atomic positions of just 1.5 Å for Tc and its immediate neighbors. Experimental, Tc-ribosome, intergroup distances are well-reproduced. In the secondary site TET5, the zwitterionic tautomer is also preferred, and Mg^{2+} is predicted to be co-bound. This is not inconsistent with the experimental electron density data, which is moderately precise.[8] The Tc is more shifted (by 2.3 Å) with respect to the crystal structure.[8] In this site, the co-bound Mg^{2+} interacts directly with just one ribosomal phosphate group (Figure 1B), completing its coordination sphere with two water molecules. Binding to the TET5 secondary site is thus weaker, with a binding free energy difference $\delta\delta G$ of 3 ± 2 kcal/mol relative to the TET1, primary site. This is consistent (Figure 1C) with the weaker, observed, electron density in TET5.[8] The continuum electrostatic treatment of the ribosome and the surrounding solution[17] includes one adjustable parameter: the dielectric constant ϵ used for the protein and RNA. Our choice of $\epsilon = 8$ yields good agreement with the experimental binding free energy differences between Tc, Mc, and Dc[18,19] (see Supporting Information for details). It also yields agreement with the more rigorous, MD free energy simulations for the Mg^{2+} binding free energies to TET1 and TET5.

10.1021/ja0741933 CCC: $40.75 © 2008 American Chemical Society

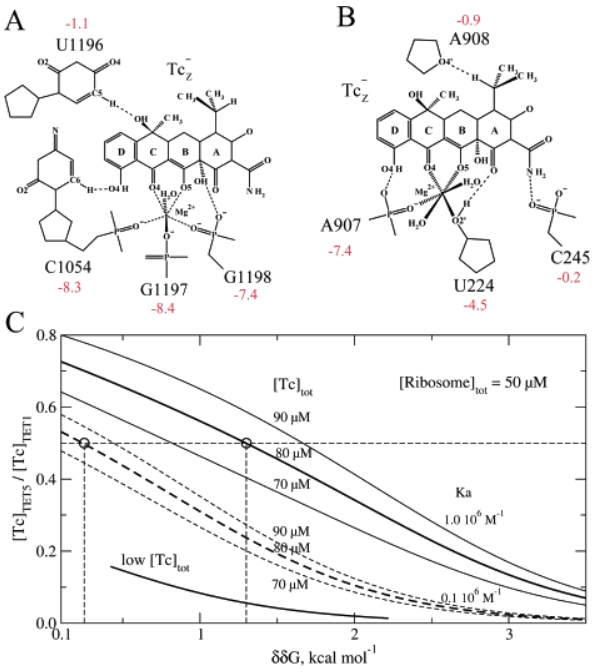


Figure 1. The Tc binding sites TET1 (A) and TET5 (B); selected group contributions to the binding free energies in kcal/mol. (C) Relative occupancy $[Tc]_{TET5}/[Tc]_{TET1}$ of TET5 and TET1 vs the computed binding free energy difference $\delta\delta G$ between sites. In the crystal cell,[8] total ribosome and Tc concentrations are 50 and 80 μM. With an association constant $K_a = 10^6$ M^{-1}, TET1 is fully occupied, while TET5 half-occupancy (qualitatively consistent with the X-ray data) occurs for $\delta\delta G \approx 1.3$ kcal/mol. The "low [Tc]" curve crudely mimics physiological conditions: [ribosome] = [Tc] = 1 μM (□1000 ribosomes per bacterial cell, □1 g of Tc in the bloodstream); TET5 is then largely unoccupied.

Varying ϵ between 4 and 16 changes $\delta\delta G$ by just ± 1 kcal/mol. The remaining uncertainty in $\delta\delta G$ is due to the finite MD simulation length.

Considering the experimental association constant K_a for a single Tc binding to the 30S subunit,[18,20] the experimental Tc and ribosome concentrations in the crystal cell,[8] and the TET1/TET5 binding free energy difference $\delta\delta G$, we can compute the occupancies of the TET1 and TET5 sites (see Supporting Information and Figure 1C). With $K_a = 10^6$ M^{-1} and $\delta\delta G = 1$ kcal/mol, we find a TET1 occupancy of ≈100% and a TET5 occupancy of about 57%, which appears compatible with the experimental electron density maps.[8] Taking into account the K_a uncertainty, the moderate X-ray resolution, and differences between the crystalline and solution conditions, the computed range for $\delta\delta G$ appears consistent with the X-ray data; the true $\delta\delta G$ value is likely to be in the lower part of the computed range, closer to 1−2 than to 4−5 kcal/mol. With more physiological concentrations, [Ribosome] = [Tc] = 1 μM, TET5 is largely unoccupied (Figure 1C).

In summary, we have presented evidence for the existence of a single, predominant binding site for Tc on the ribosome, suggesting that other reported binding sites are weaker and not highly occupied at physiological Tc concentrations. This lends support to the view that allosteric coupling between binding sites is not important or at least not crucial for the function of tetracycline. Our simulations complement the X-ray data, giving a more detailed picture of Tc binding, including thermodynamic information such as individual residue free energy contributions. The continuum electrostatic model has many caveats[14,15] and is only semiquantitative, so that additional, experimental tests are needed. Nevertheless, protein, RNA, water, and ions were described by well-tested force fields;[16] the Tc/Mc/Dc force field was developed in the same painstaking way.[13] The simulations were long, and 16 different structural models were compared. The results are robust with respect to the details of the Poisson−Boltzmann treatment, with a large range of parameter choices all leading to the same qualitative picture: the primary site is computed to have a larger Tc affinity than the putative secondary site TET5.

Acknowledgment. A.A. was supported by an Egide doctoral fellowship. We thank Marat Yusupov, Winfried Hinrichs, and Yves Méchulam for discussions, and Satoru Shimizu for crystallographic calculations.

Supporting Information Available: Full ref 16a; details of simulation methods and results. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Hlavka, J. J., Boothe, J. H., Eds. *The Tetracyclines;* Springer-Vertag: Heidelberg, Germany, 1985.
(2) (a) Chopra, L.; Roberts, M. *Microbiol. Mol. Biol. Rev.* **2001**, *65*, 232−260. (b) Polacek, N.; Mankin, A. S. *Crit. Rev. Biochem. Mol. Biol.* **2005**, *40*, 285−311.
(3) (a) Yusupov, M. M.; Yusupova, G. Z.; Baucom, A.; Lieberman, K.; Earnest, T. N.; Cate, J. H.; Noller, H. F. *Science* **2001**, *292*, 883−896. (b) Selmer, M.; Dunham, C. M.; Murphy, F. V.; Weixlbaumer, A.; Petry, S.; Kelley, A. C.; Weir, J. R.; Ramakrishnan, V. *Science* **2006**, *313*, 1935−1942.
(4) (a) Cundliffe, E. In *The Molecular Basis of Antibiotic Action*; Gale, E. F., Cundliffe, E., Reynolds, P. E., Richmond, M. H., Waring, M. J. Eds.; **1981**, 402−547. (b) Weisblum, B. *Antimicrob. Agents Chemother.* **1995**, *39*, 577−585. (c) Spahn, C. M.; Prescott, C. D. *J. Mol. Med.* **1996**, *74*, 423−439.
(5) Connell, S. R.; Tracz, D. M.; Nierhaus, K. H.; Taylor, D. E. *Antimicrob. Agents Chemother.* **2004**, *47*, 3675−3681.
(6) Auerbach, T.; Bashan, A.; Harms, J.; Schluenzen, F.; Zarivach, R.; Bartels, H.; Agmon, I.; Kessler, M.; Pioletti, M.; Franceschi, F.; Yonath, A. *Curr. Drug Targets Infect. Disord.* **2002**, *2*, 169−186.
(7) Anokhina, M. M.; Barta, A.; Nierhaus, K. H.; Spiridonova, V. A.; Kopylov, A. M. *Nucleic Acids Res.* **2004**, *32*, 2594−2597.
(8) Brodersen, D. E.; Clemons, W. M., Jr.; Carter, A. P.; Morgan-Warren, R. J.; Wimberly, B. T.; Ramakrishnan, V. *Cell* **2000**, *103*, 1143−1154.
(9) Pioletti, M.; Schlunzen, F.; Harms, J.; Zarivach, R.; Gluhmann, M.; Avila, H.; Bashan, A.; Bartels, H.; Auerbach, T.; Jacobi, C.; Hartsch, T.; Yonath, A.; Franceschi, F. *EMBO J.* **2000**, *20*, 1829−39.
(10) Gregory, S. T.; Carr, J. F.; Rodriguez-Correa, D.; Dalberg, A. E. *J. Bacteriol.* **2005**, *187*, 4804−4812.
(11) Thompson, J.; Dalberg, A. E. *Nucleic Acids Res.* **2004**, *32*, 5954−5961.
(12) Kaper, J. B.; Nataro, J. P.; Mobley, H. L. *Nat. Rev. Microbiol.* **2004**, *2*, 123−140.
(13) (a) Aleksandrov, A.; Simonson, T. *J. Comp. Chem.* **2006**, *13*, 1517−1533. (b) Aleksandrov, A.; Proft, J.; Hinrichs, W.; Simonson, T. *ChemBioChem* **2007**, *8*, 675−685.
(14) (a) Warshel, A. *Computer Modelling Of Chemical Reactions in Enzymes and Solutions*; Wiley: New York, 1991. (b) Simonson, T.; Archontis, G.; Karplus, M. *Accts. Chem. Res.* **2002**, *35*, 430−437. (c) Simonson, T. In *Free Energy Calculations*; Chipot, C., Pohorille, A., Eds.; Springer: New York, 2007; Chapter 13. (d) Jorgensen, W. L. **2003**, *303*, 1813−1818. (e) Thompson, D.; Simonson, T. *J. Biol. Chem.* **2006**, *33*, 23792−23803.
(15) (a) McDowell, S. E.; Spackova, N.; Sponer, J.; Walter, N. G. *Biopolymers 85*, 169−184. (b) Spackova, N.; Cheatham, T. E.; Ryjacek, F.; Lankas, F.; Van Meervelt, L.; Sponer, J. *J. Am. Chem. Soc.* **2003**, *125*, 1759−1769. (c) Yang, G.; Trylska, J.; Tor, Y.; McCammon, J. A. *J. Med. Chem.* **2007**, *49*, 5478−5490.
(16) (a) Mackerell, A.; et al. *J. Phys. Chem. B* **1998**, *102*, 3586−3616. (b) Mackerell, A. D.; Wiorkiewicz-Kuczera, J.; Karplus, M. *J. Am. Chem. Soc.* **1995**, *117*, 11946−11975. (c) Jorgensen, W.; Chandrasekar, J.; Madura, J.; Impey, R.; Klein, M. *J. Chem. Phys.* **1983**, *79*, 926−935.
(17) Im, W.; Beglov, D.; Roux, B. *Comp. Phys. Commun.* **1998**, *111*, 59−75.
(18) Olson, M. W.; Ruzin, A.; Feyfant, E.; Rush, T. S., III; O'Connell, J.; Bradford, P. A. *Antimicrob. Agents Chemother.* **2006**, *50*, 2156−2166.
(19) Ross, J.; Eady, E.; Cove, J.; Cunliffe, W. *Antimicrob. Agents Chemother.* **1998**, *42* (7), 1702−1705.
(20) Epe, B.; Wolley, P. *EMBO J.* **1984**, *3*, 121−126.

JA0741933

Supplementary Material: Molecular dynamics simulations of the 30S ribosomal subunit reveal a preferred tetracycline binding site

Alexey Aleksandrov and Thomas Simonson
Laboratoire de Biochimie (CNRS UMR7654), Department of Biology, Ecole Polytechnique, 91128 Palaiseau, France.

Full reference 16a

Mackerell, A., Bashford, D., Bellott, M., Dunbrack, R., Evanseck, J., Field, M., Fischer, S., Gao, J., Guo, H., Ha, S., Joseph, D., Kuchnir, L., Kuczera, K., Lau, F., Mattos, C., Michnick, S., Ngo, T., Nguyen, D., Prodhom, B., Reiher, W., Roux, B., Smith, J., Stote, R., Straub, J., Watanabe, M., Wiorkiewicz-Kuczera, J., Yin, D., and Karplus, M. An all-atom empirical potential for molecular modelling and dynamics study of proteins. *J. Phys. Chem. B* 102 (1998), 3586–3616.

Computational methods: Molecular dynamics (MD) simulations

The crystal structure of the Thermus thermophilus 30S ribosomal subunit, complexed with tetracycline (Tc), was taken from the Protein Data Bank (PDB), entry 1HNW [1]. The complexes with minocycline (Mc) and doxycycline (Dc) were built by superimposing Mc or Dc on Tc in the 1HNW structure. MD simulations were performed for the primary site (TET1) and the secondary site (TET5; only for Tc) and included RNA and protein residues within a 26 Å sphere, centered on the C22 atom of Tc (or Mc or Dc) in the primary site, or on the C6 atom of Tc in the secondary site. Notice that the sphere center does not play any special role in the model, so the precise choice is not important. Sensitivity to the position of the model boundary, on the other hand, is discussed below. Hydrogens were constructed with ideal stereochemistry. A 29 Å sphere of water was overlaid and waters that overlapped protein/RNA, crystal ions, Tc, or Mg^{2+} were removed. To zero the total charge and simulate ionic strength explicitly, in addition to crystal ions, 24 chloride ions and 59 sodium ions were added to the models of the primary site, and 20 chloride ions and 76 sodium ions were added to the models of the secondary site. The ions were added one by one, choosing the water

molecule that experienced the lowest (respectively, the highest) electrostatic potential and replacing that water by a sodium (respectively, a chloride) [2]. RNA/protein atoms between 20 Å and 26 Å from the initial sphere's center were harmonically restrained to their Xray positions. Simulations were performed with the SSBP solvent model [2, 3], which treats the region outside the 26 Å sphere as a uniform dielectric continuum, with a dielectric constant of 80. This is reasonable, since most of the outer region is water. Indeed, if a parallelepiped is adjusted to fit the 30S particle, its inner volume is about 91% water. Newtonian dynamics were used for the innermost region, within 20 Å of the sphere's center; Langevin dynamics were used for the outer part of the sphere, referred to as the buffer region. We used a two femtosecond timestep and a bath temperature of 292 K. The CHARMM22 force field was used for the protein [4]. The TIP3P model was used for water [5]. Electrostatic interactions between atoms within the 26 Å sphere were computed without any cutoff, using a multipole approximation for distant groups [6]. Tetracycline was described with the force field developed previously [7]. The Mc and Dc force fields were developed in the same way and will be described elsewhere. The Mc and Dc force fields yield good agreement for the Tc/Mc and Tc/Dc binding free energy differences, computed with alchemical, MD free energy simulations (not to be confused with Poisson-Boltzmann free energy calculations; see below) [8, 9]. Calculations were performed with the CHARMM program [10]. Including the two binding sites, the different ligands and Tc forms, and the alchemical MD free energy simulations (see below), we performed a total of more than 150 ns of dynamics.

Computational methods: Mg^{2+} binding free energies

To establish that Mg^{2+} is bound in the TET1 and TET5 sites in the absence of Tc or Mc, we employed a more rigorous, well-established free energy simulation method; see [11, 12] for reviews. MD simulations were performed with the setup described above, without Tc. Water (within the 26 Å-radius sphere) was described explicitly, in contrast to the dielectric continuum treatment above. During the simulations, the Mg^{2+} cation is gradually deleted. The same procedure is performed for Mg^{2+} in a sphere of pure water. The free energy change is obtained from a "thermodynamic integration" formula [8, 12]. Taking the difference between the free energies computed in the ribosome complex and in pure water yields the binding free energy [13]; Table 1. For more details, see a recent application of the same method to an enzyme:ATP:Mg^{2+} complex [14]. This explicit solvent, MD free energy approach does not involve any adjustable parameters. As a further test, the same explicit solvent approach was used

to compute the binding free energy difference between Tc and Mc or doxycycline (Dc) in the primary site, yielding good agreement with experiment (not shown; manuscript in preparation).

Computational methods: Poisson-Boltzmann (PB) binding free energies

As explained in the main text, several of the binding processes are studied with a simplified, semi-empirical model, which is reasonably well-established, although a number of variants have been used. The processes studied in this way include the competition between Tc binding to TET1 and TET5, as well as the relative stabilities of different Tc tautomers binding to either TET1 or TET5. The electrostatic contribution to the binding free energy is obtained by solving the Poisson-Boltzmann (PB) equation for the bound and unbound forms. This approach has been extensively described in the literature and textbooks; see eg [8, 9, 15]. Importantly, the method does not attempt to compute absolute binding free energies, but only binding free energy differences between two similar ligands (eg, Tc and minocycline), or two putative sites (eg, TET1 and TET5). Because an Mg^{2+} is already in place in both TET1 and TET5 before Tc:Mg^{2+} binds, we consider Mg^{2+} dissociation from the binding site (TET1 or TET5), followed by Tc:Mg^{2+} binding to the site. Thus, to compute the (electrostatic contribution to the) binding free energy in the TET1 site, for example, we consider two systems: the ribosome:Tc:Mg^{2+} complex along with a free Mg^{2+} ion; and the ribosome:Mg^{2+} complex along with a free Tc:Mg^{2+}. Subtracting their electrostatic free energies yields the electrostatic contribution to the binding free energy in the TET1 site. Subtracting the analogous free energy for the TET5 site, we obtain the binding free energy difference between sites.

Previous work with this class of models has sometimes involved the explicit treatment of other, "non-electrostatic" contributions. This is well-illustrated by a recent study of aminoglycosidic antibiotics binding to the 30S ribosome particle by McCammon and coworkers [16]; see their Eq. 1. Additional terms treated there include rotation/translation entropy of the ligand and loss of conformational entropy due to torsional degrees of freedom of the ligands that become ordered upon binding. These terms are not relevant here. The first will largely cancel when TET1 and TET5 are compared, or Tc and Mc. The rigidity of Tc and its fused-ring structure makes the second unimportant. Two other terms considered by McCammon and several other groups are a solvent-accessible surface area term (ΔG_{np} in Eq. 1 of [16]) and changes

in the van der Waals interactions between the ligand and its surroundings upon binding (ΔG_{strain} in Eq. 1 of [16]; ΔG_{vdW} in [17, 18]). These terms are neglected in the present work, for reasons discussed in the following section and in several review articles [8, 15, 16, 19].

The details of the PB calculations are as follows. The electrostatic free energy is obtained by numerically solving the Poisson-Boltzmann equation, using a cubic grid and a finite-difference algorithm implemented in CHARMM [20]. The ionic strength is set to the physiological value of 100 mM (monovalent salt concentration). Doubling this value changes the results only slightly (by 0.3 kcal/mol for the TET1/TET5 Tc binding free energy difference). Repeating the calculations with a 32 Å radius sphere changes the results negligibly (Tc binding to TET1 changes by 0.1 kcal/mol), even though the number of ribosome atoms almost doubles with this larger model. This strongly suggests that the 26 Å model size is reasonable, both for the continuum electrostatic and the MD calculations. The structure of the ribosome:Tc, :Mc, and :Dc complexes are taken from the MD simulations. Calculations are performed for multiple structures, taken every 4 ps along the equilibrated 2.0 ns trajectory, for a total of 250 structures per ligand. The separate ligand and ribosome structures are obtained by simply discarding the unwanted partner. Thus, structural relaxation on separating protein and ligand is not explicitly included (but it is implicit in the dielectric constant). The solvent dielectric constant was set to 80. The solute dielectric constant was varied from 4 to 16. The data in Table 2 show that Tc binds to both the primary and secondary sites in its zwitterionic form with an associated Mg. The data in Table 3 and 4 show that a dielectric of 6-8 is optimal and that Tc binding to the secondary site is about 3 kcal/mol weaker than the primary site.

Computational methods: neglecting non-electrostatic terms in the Poisson-Boltzmann binding calculations

Poisson-Boltzmann binding calculations have been used to study a wide variety of protein:ligand complexes [8, 9, 15, 16, 19, 21–26]. One either compares two ligands binding to the same site, or (more rarely) one ligand binding to two sites. In many cases, the so-called MMPB-SA method has been used [23], where van der Waals contributions are explicitly calculated, as well as surface area terms and even vibrational entropy contributions. This poses two problems: (a) these contributions are often noisy, being obtained as differences between two large numbers (eg, the total van der Waals energies of the bound and unbound complexes); (b) the corresponding model

cannot be derived rigorously, either from thermodynamic perturbation theory or from continuum electrostatics; rather it has a heuristic character; see [15] for a discussion. In other cases [16, 21, 22, 24, 25], the focus has been on problems where electrostatic interactions are thought to play a dominant role, so that the continuum electrostatics model is used and other contributions are assumed to cancel. For the present system, where Tc^- binding to TET1 and TET5 are compared, four arguments suggest that the PB, electrostatic contribution provides a reasonable approximation to the overall effect. First, the ligand is an anion, binding directly to a Mg^{2+} cation, close to several ribosome phosphate groups (Fig. 1 in main text). Thus, electrostatic effects should be predominant. Note that this is true not only for the relative binding to TET1 and TET5, but for comparing the neutral and zwitterionic Tc tautomers binding to either site. Second, by adjusting the solute dielectric constant to reproduce binding free energy differences between Tc, Dc, Mc, as well as the Mg^{2+} binding free energy difference between TET1 and TET5 (see below), we implicitly include other energy terms in the model. In continuum electrostatics, the dielectric constant takes into account the work needed to polarize the medium around a new charge [27]; this work arises partly from elastic deformation of the medium and atomic packing effects mediated by van der Waals and stereochemical energy terms. Notice that both TET1 and TET5 are preorganized for Tc binding: when the MD structures with and without Tc are compared, the rms deviation between the two is just 1.0 Å for both sites and for groups within 7 Å of the ligand. Thus, elastic deformation of either site in response to the ligand is small. Third, we can make a rough estimate of the contribution of van der Waals and accessible surface area effects to the relative binding free energy of Tc to TET1 and TET5, using the Extended Linear Response method of Aqvist, Jorgensen and coworkers; see [17, 18, 28] for details. The mean van der Waals interaction energy of Tc with its surroundings is computed for Tc bound to TET1 or TET5. These energies are denoted E_{vdW}^{TET1} and E_{vdW}^{TET5}. We emphasize that they represent interactions between Tc and its surroundings, not total van der Waals energies. With the Linear Response method, the van der Waals contribution to the TET1/TET5 binding free energy difference has the form $\Delta G_{vdW}^{TET1/TET5} = \alpha(E_{vdW}^{TET1} - E_{vdW}^{TET5})$ (see, eg, Eq. 4 of [17]). Values of α of between 0.10 and 0.40 have been commonly used [17, 18, 28]. For the present system, we obtain $E_{vdW}^{TET1} - E_{vdW}^{TET5}$ = -5±2 kcal/mol. Thus, the Extended Linear Response estimate for $\Delta G_{vdW}^{TET1/TET5}$ would be around -1.0±0.5 kcal/mol; the negative sign indicates that this contribution favors the TET1 site. Since this contribution is at least partly included in the PB term, the value obtained is likely to be an overestimate. Note that the mean van der Waals interactions between the empty

TET1 site and its environment, and between the empty TET5 site and its environment, are practically the same. Another term often included in simple binding models is a surface area term, which attempts to capture contributions from the hydrophobic effect. The solvent-accessible surface areas lost by the Tc ligand when it binds to TET1 and TET5 are 715 and 915 Å^2, respectively. McCammon and coworkers used a surface term in combination with PB calculations to study aminoglycosidic antibiotics binding to the 30S ribosome particle [16]. Using their surface tension of about 5 cal/mol/Å^2, this leads to a 1 kcal/mol contribution to the TET1/TET5 binding free energy, favoring TET5. This contribution almost exactly cancels the van der Waals estimate above. Both terms are compatible with the estimated uncertainty of our TET1/TET5 comparison (see main text). Both terms are comparable to values obtained for other ligands binding to the 30S ribosome particle (see ΔG_{strain} and ΔG_{np} terms in [16]; see ΔG_{vdW} term in [29]), both in absolute value and in comparison to the electrostatic term. Finally, the fourth argument suggesting that non-electrostatic contributions are small, is the good agreement between our computed TET1/TET5 binding free energy difference and the observed electron density maps.

Computed occupancies of TET1 and TET5

30S complexes with 0, 1, or 2 bound Tcs are compared through the thermodynamic cycle in Fig. 1. State 0 (top) corresponds to the unbound partners; states 1 and 1' correspond to Tc bound to either the primary or the secondary site; in state 2, both binding sites are occupied. The partition functions of these states can be written:

$$\begin{aligned} Z_0 &= e^{-\beta G_0} \\ Z_1 &= e^{-\beta G_1} + e^{-\beta G_1'} \\ Z_2 &= e^{-\beta G_2} \end{aligned} \tag{1}$$

where G_0, G_1, G_1', G_2 represent free energies of the corresponding states. Then,

$$\begin{aligned} \Delta G_a &= -kT \ln(Z_1/Z_0) \\ \Delta G_b &= -kT \ln(Z_2/Z_1) \end{aligned} \tag{2}$$

We define $\beta = 1/kT$, where k is Boltzmann's constant and T the temperature; $\alpha = exp(-\beta\delta\delta G)$; $\delta\delta G = \delta G_1' - \delta G_1$ is the binding free energy difference between Tc in the secondary and primary sites. Equations (1) and (2) give:

$$\Delta G_b = \Delta G_a + kT(1+\alpha)(1+\alpha^{-1}) \tag{3}$$

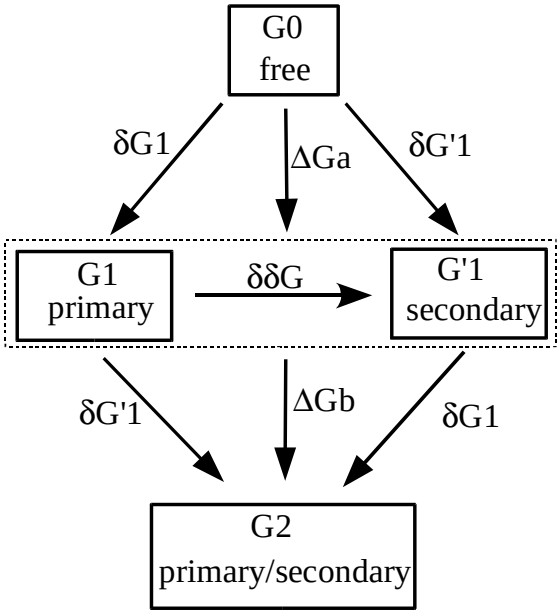


Figure 1: Thermodynamic cycle for Tc:ribosome binding.

At equilibrium:

$$\begin{aligned} \Delta G_a &= -kT \ln[R:T]_1/([R][T]) \\ \Delta G_b &= -kT \ln[R:T]_2/([R:T]_1[T]) \end{aligned} \tag{4}$$

where $[R:T]_1$ is the concentration of ribosomes with a single Tc bound (either in the primary or secondary site); $[R:T]_2$ is the concentration of ribosomes with Tc bound in both sites; [R], [T] are the concentrations of free ribosomes and tetracyclines. The total concentrations of ribosomes and tetracyclines are:

$$[R]_{tot} = [R:T]_1 + [R:T]_2 + [R] \tag{5}$$

$$[T]_{tot} = [R:T]_1 + 2[R:T]_2 + [T] \tag{6}$$

The association constant is:

$$K_a = [R:T]_1/([R][T]) \tag{7}$$

Together, (3–7) form a closed system of equations. For the concentration of complexes with Tc bound to the primary site, we obtain:

$$[R:T]_1^{\mathrm{TET1}} = [R:T]_1/(1+\alpha) \tag{8}$$

For the ratio of complexes with Tc in TET5 and TET1, we obtain:

$$\frac{[Tc]_{\mathrm{TET5}}}{[Tc]_{\mathrm{TET1}}} = \frac{[R:T]_2}{[R:T]_1^{\mathrm{TET1}} + [R:T]_2} \tag{9}$$

The experimental total concentration of ribosomes in the crystal [1] was estimated from the volume of the unit cell as 50 10^{-6} M; the total concentration of tetracyclines was 80 10^{-6} M. The association constant K_a is thought to be in the range 0.1 10^6 M^{-1} to 1.0 10^6 M^{-1} [30–32].

Results: Mg^{2+} is pre-bound in TET1 and TET5

By reversibly deleting Mg^{2+} from either TET1, TET5, or from bulk water, we can show that Mg^{2+} is strongly bound in both sites in the absence of any ligand. MD simulations are done with explicit solvent and the free energy change is obtained by thermodynamic integration (see above). The data are given in Table 1.

Table 1: Free energy to delete Mg^{2+} cations from the primary and secondary ribosomal binding sites using MD with explicit solvent

Medium	$\Delta G(Mg^{2+} \rightarrow nothing)$	$\Delta\Delta G$
solution	440[a]	-
TET1 site	465.9/469.0	27.5
TET5 site	454.5/450.6	12.5

Energies are in kcal/mol. [a]Solution value is for Mg^{2+} in pure water, from a previous study [33]. For TET1 and TET5, two simulations were done, giving two free energy results. $\Delta\Delta G$ is the difference between Mg^{2+} deletion in the ribosome and in pure water. A positive value indicates a favorable binding to the ribosome site.

Results: Tc binds in the zwitterionic form to TET1 and TET5

The Poisson-Boltzmann free energy results are in Table 2. The zwitterionic state is preferred in solution [21]. Its binding free energies are lower than those of the neutral form, so that it is the predominant form in both sites.

Table 2: Poisson-Boltzmann free energy simulations comparing the binding of Tc^{-}:Mg^{2+} to the primary and secondary site. Energies are in kcal/mol.

binding site	Tc form	Mg^{2+}	ϵ, dielectric constant			
			4	6	8	12
primary	neutral	yes	-54.0	-39.4	-31.6	-23.5
primary	zwitterionic	yes	-78.2	-55.8	-44.0	-31.8
secondary	neutral	yes	-25.1	-19.5	-16.4	-13.3
secondary	zwitterionic	yes	-40.5	-30.2	-24.7	-19.0

Results: Calibrating the dielectric constant to match Dc and Mc binding

We may assume the ligands all bind mainly to the primary site in their zwitterionic forms with Mg present. For Mc, the experimental binding free energy difference is -2.6 kcal/mol, favoring Mc over Tc. Since Dc has the same biological activity as Tc, we assume their binding free energies are equal. These experimental values are best reproduced with a dielectric constant of 6 or 8 (Table 3).

Table 3: Poisson-Boltzmann free energy simulations comparing binding of zwitterionic tetracycline (Tc), minocycline (Mc) and doxycycline (Dc) to the TET1 site.

ligand	ϵ, dielectric constant				$\Delta\Delta G_{exptl}$
	4	6	8	12	
Tc^-:Mg^{2+}	-78.2	-55.8	-44.0	-31.8	
Mc^-:Mg^{2+}	-81.1	-57.8	-45.1	-32.9	
$\Delta\Delta G(Mc \rightarrow Tc)$	-2.9 (2.0)	-2.0 (1.3)	-1.5 (1.0)	-1.1 (0.8)	-2.6
Dc^-:Mg^{2+}	-76.6	-54.7	-43.2	-31.4	
$\Delta\Delta G(Dc \rightarrow Tc)$	1.6 (2.0)	1.1 (1.3)	0.8 (1.0)	0.4 (0.8)	0.0

Poisson-Boltzmann free energies, in kcal/mol. Values correspond to the binding of Tc^-:Mg^{2+}, Dc^-:Mg^{2+}, and Mc^-:Mg^{2+} to the primary site. Negative $\Delta\Delta G(Mc \rightarrow Tc)$ values correspond to stronger Mc binding. The experimental result for Mc is -2.6 kcal/mol [30] and 0.0 kcal/mol for Dc [34].

Results: Tc binds preferentially to the primary, TET1 site

We use Poisson-Boltzmann free energy calculations. We follow a protocol that compares the binding of Mg^{2+} alone, and the binding of Tc^-:Mg^{2+}. This has the advantage that Tc is always associated with Mg^{2+}, and the large energy associated with Tc^-:Mg^{2+} separation never appears explicitly. When the two sites and ligands are compared, the contributions due to Mg^{2+} cancel, as they should. Thus, the line "secondary - primary" in Table 4 gives the Tc binding free energy difference $\delta\delta G$ between sites. For all values of the solute dielectric constant, $\delta\delta G$ is negative, ie, the primary site is favored. Individul group contributions to TET1 binding are shown in Table 5. The phosphate groups of Cyt1054, Gua1197, and Gua1198 make the largest contributions by far.

Table 4: Tc^{-}:Mg^{2+} and Mg^{2+} binding to the primary and secondary sites.

binding site	ligand	ϵ, dielectric constant			
		4	6	8	12
primary	Mg^{2+}	-95.5	-71.2	-58.3	-44.7
secondary	Mg^{2+}	-59.9	-48.5	-42.1	-34.8
difference	Mg^{2+}	-35.6	-22.7	-16.2	-9.9
primary	Tc^{-}:Mg^{2+}	-78.2	-55.8	-44.0	-31.8
secondary	Tc^{-}:Mg^{2+}	-40.5	-30.2	-24.8	-19.0
difference	Tc^{-}:Mg^{2+}	-37.7	-25.6	-19.2	-12.8
secondary-primary		2.1 (2.0)	2.9 (2.0)	3.0 (2.0)	2.9 (2.0)

Poisson-Boltzmann free energies, in kcal/mol. Values correspond to the binding of Mg^{2+} or Tc^{-}:Mg^{2+} to the primary or secondary site. Tc was deleted from the pocket when Mg2+ binding was computed. Positive secondary-primary values correspond to preferential Tc binding to the primary site.

Table 5: Contributions (kcal/mol) of individual groups to the Poisson-Boltzmann free energy of binding Tc^{-}:Mg^{2+} to the TET1 site; dielectric constant $\epsilon = 8$

nucleotide	backbone	base	nucleotide	backbone	base
Gua1197	-0.0	-5.6	Gua963	0.0	-0.1
Cyt1054	-0.9	-4.6	Ura1052	0.0	-0.1
Gua1198	0.0	-5.0	Ade1055	0.0	0.0
Gua1053	-0.0	-1.1	Cyt967	-0.2	0.2
Ura1199	0.1	-1.0	Ura1194	0.0	0.0
Ura1196	-0.1	-0.6	Cyt1195	-0.7	0.8
Ade964	0.0	-0.3	Ade968	0.0	0.1
Ade965	0.0	-0.2	Gua966	0.0	1.0

References

[1] Brodersen, D. E., Clemons Jr., W. M., Carter, A. P., Morgan-Warren, R. J., Wimberly, B. T., and Ramakrishnan, V. The structural

basis for the action of the antibiotics tetracycline, pactamycin, and hygromycin B on the 30S ribosomal subunit. *Cell 103* (2000), 1143–154.

[2] Nina, M., and Simonson, T. Molecular dynamics of the tRNA^{Ala} acceptor stem: comparison between continuum reaction field and particle-mesh Ewald electrostatic treatments. *J. Phys. Chem. B 106* (2002), 3696–3705.

[3] Beglov, D., and Roux, B. Finite representation of an infinite bulk system: solvent boundary potential for computer simulations. *J. Chem. Phys. 100* (1994), 9050–9063.

[4] Mackerell, A., Bashford, D., Bellott, M., Dunbrack, R., Evanseck, J., Field, M., Fischer, S., Gao, J., Guo, H., Ha, S., Joseph, D., Kuchnir, L., Kuczera, K., Lau, F., Mattos, C., Michnick, S., Ngo, T., Nguyen, D., Prodhom, B., Reiher, W., Roux, B., Smith, J., Stote, R., Straub, J., Watanabe, M., Wiorkiewicz-Kuczera, J., Yin, D., and Karplus, M. An all-atom empirical potential for molecular modelling and dynamics study of proteins. *J. Phys. Chem. B 102* (1998), 3586–3616.

[5] Jorgensen, W., Chandrasekar, J., Madura, J., Impey, R., and Klein, M. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys. 79* (1983), 926–935.

[6] Stote, R., States, D., and Karplus, M. On the treatment of electrostatic interactions in biomolecular simulation. *J. Chim. Phys. 88* (1991), 2419–2433.

[7] Aleksandrov, A., and Simonson, T. The tetracycline:Mg^{2+} complex: a molecular mechanics force field. *J. Comp. Chem. 13* (2006), 1517–1533.

[8] Simonson, T., Archontis, G., and Karplus, M. Free energy simulations come of age: the protein–ligand recognition problem. *Acc. Chem. Res. 35* (2002), 430–437.

[9] McCammon, J. Computation of noncovalent binding affinities. In *Theory and Applications of Computational Chemistry*, C. Dykstra, G. Frenking, K. Kim, and G. Scuseria, Eds. Elsevier, Amsterdam, 2005, pp. 41–46.

[10] Brooks, B., Bruccoleri, R., Olafson, B., States, D., Swaminathan, S., and Karplus, M. Charmm: a program for macromolecular energy, minimization, and molecular dynamics calculations. *J. Comp. Chem. 4* (1983), 187–217.

[11] Kollman, P. Free energy calculations: applications to chemical and biochemical phenomena. *Chem. Rev. 93* (1993), 2395.

[12] Simonson, T. Free energy calculations. In *Computational Biochemistry & Biophysics*, O. Becker, A. Mackerell Jr., B. Roux, and M. Watanabe, Eds. Marcel Dekker, N.Y., 2001, ch. 9.

[13] Jorgensen, W., Buckner, K., Boudon, S., and Tirado-Rives, J. Efficient computation of absolute free energies of binding by computer simulations. Application to the methane dimer in water. *J. Chem. Phys. 89* (1988), 3742–3746.

[14] Thompson, D., and Simonson, T. Molecular dynamics simulations show that bound mg^{2+} contributes to amino acid and aminoacyl adenylate binding specificity in aspartyl-trna synthetase through long range electrostatic interactions. *J. Biol. Chem. 281* (2006), 23792–23803.

[15] Simonson, T. Free energy calculations: approximate methods for biological macromolecules. In *Free energy calculations: theory and applications in chemistry and biology*, C. Chipot and A. Pohorille, Eds. Springer Verlag, N.Y., 2006, ch. 12.

[16] Yang, G., Trylska, J., Tor, Y., and McCammon, J. A. Binding of aminoglycosidic antibiotics to the oligonucleotide A-site model and 30S ribosomal subunit: Poisson-boltzmann model, thermal denaturation, and fluorescence studies. *J. Med. Chem. 49* (2006), 5478–5490.

[17] Ostrovsky, D., Udier-Blagovic, M., and Jorgensen, W. Analyses of activity for factor Xα inhibitors based on Monte Carlo simulations. *J. Med. Chem. 46* (2003), 5691–5699.

[18] Kroeger-Smith, M., Hose, B., Hawkins, A., Lipchock, J., Farnsworth, D., Rizzo, R., Tirado-Rives, J., Arnold, E., Zhang, W., Hughes, S., Jorgensen, W., Michedja, C., and Smith, R. Molecular modeling calculations of HIV-1 reverse transcriptase nonnucleoside inhibitors: correlation of binding energy with biological activity for novel 2-aryl-substituted benzimidazole analogues. *J. Med. Chem. 46* (2003), 1940–1947.

[19] McDowell, S. E., Spackova, N., Sponer, J., and Walter, N. G. Molecular dynamics simulations of RNA: An in silico single molecule approach. *Biopolymers 85* (2007), 169–184.

[20] Im, W., Beglov, D., and Roux, B. Continuum solvation model: computation of electrostatic forces from numerical solutions to the Poisson-Boltzmann equation. *Comp. Phys. Comm. 111* (1998), 59–75.

[21] Aleksandrov, A., Proft, J., Hinrichs, W., and Simonson, T. Protonation patterns in tetracycline:tet repressor recognition: Simulations and experiments. *ChemBioChem 8* (2007), 675–685.

[22] Hendsch, Z., and Tidor, B. Electrostatic interactions in the GCN4 leucine zipper: substantial contributions arise from intramolecular interactions enhanced on binding. *Prot. Sci. 8* (1999), 1381–1392.

[23] Kollman, P., Massova, I., Reyes, C., Kuhn, B., Huo, S., Chong, L., Lee, M., Lee, T., Duan, Y., Wang, W., Donini, O., Cieplak, P., Srinivasan, J., Case, D., and Cheatham, T. Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. *Acc. Chem. Res. 33* (2000), 889–897.

[24] Archontis, G., Simonson, T., and Karplus, M. Binding free energies and free energy components from molecular dynamics and Poisson-Boltzmann calculations. Application to amino acid recognition by aspartyl-tRNA synthetase. *J. Mol. Biol. 306* (2001), 307–327.

[25] Thompson, D., Lazennec, C., Plateau, P., and Simonson, T. Ammonium scanning in an enzyme active site: the chiral specificity of aspartyl-tRNA synthetase. *J. Biol. Chem. in press* (2007), 0000.

[26] Wong, C., Hunenberger, P., Akamine, P., Narayana, N., Diller, T., McCammon, J., Taylor, S., and Xuong, N. Computational analysis of PKA–balanol interactions. *J. Med. Chem. 44* (2001), 1530–1539.

[27] Fröhlich, H. *Theory of Dielectrics*. Clarendon Press, Oxford, 1949.

[28] Aqvist, J., Luzhkov, V., and Brandsal, B. Ligand binding affinities from MD simulations. *Acc. Chem. Res. 35* (2002), 358–365.

[29] Almlöf, M., Ander, M., and Aqvist, J. Energies of codon-anticodon recognition on the small ribosomal subunit. *Biochemistry 46* (2007), 200–209.

[30] Olson, M. W., Ruzin, A., Feyfant, E., Rush, T. S., O'Connell, J., and Bradford, P. A. Functional, biophysical, and structural bases for antibacterial activity of tigecycline. *Antimicrob. Agents Chemoth. 50* (2006), 2156–2166.

[31] Wissmann, A., and Hillen, W. *Forum Mikrobiol.* (1989), 292–299.

[32] Epe, B., and Wolley, P. *EMBO J. 3* (1984), 121–126.

[33] Thompson, D., Plateau, P., and Simonson, T. Free energy simulations reveal long-range electrostatic interactions and substrate-assisted specificity in an aminoacyl-tRNA synthetase. *ChemBioChem 7* (2006), 337–344.

[34] Ross, J., Eady, E., Cove, J., and Cunliffe, W. *Antimicrob. Agents Chemoth. 42* (1998), 1702–1705.

Binding of tetracyclines to elongation factor Tu, the Tet Repressor, and the ribosome: a molecular dynamics simulation study

Alexey Aleksandrov and Thomas Simonson*
Laboratoire de Biochimie (CNRS UMR7654), Department of Biology, Ecole Polytechnique, 91128 Palaiseau, France.
*Email: thomas.simonson@polytechnique.fr

Abstract

Tetracycline (Tc) is a broad-spectrum antibiotic that kills bacteria by interrupting protein biosynthesis. It is thought that the bacteriostatic action of Tc is associated with its binding to the acceptor site (or A site) in the bacterial ribosome, interfering with the attachment of aminoacyl-tRNA. Recently, however, the crystal structure of a complex between Tc and elongation factor Tu (EF-Tu) was solved, raising the question whether Tc's binding to EF-Tu has a role in its inhibition of protein synthesis. We address this question using computer simulations. As controls, we first compute relative ribosome binding free energies for seven Tc variants for which experimental data are available, obtaining good agreement. We then consider the binding of Tc to the EF-Tu:GDP complex. We show that the direct contribution of EF-Tu to the binding free energy is negligible; rather, the binding can be solely attributed to interactions of Tc with a bridging Mg ion and the GDP phosphate groups. Further, our calculations show that EF-Tu does not exhibit any binding preference for Tc over the non-antibiotic, 4-dedimethyl-Tc, and EF-Tu does not bind the Tc analogue tigecycline, which is a potent antibiotic. In contrast, both the ribosome and the Tet Repressor protein (involved in Tc resistance) do show a binding preference for Tc over 4-dedimethyl-Tc, and the ribosome prefers to bind tigecycline over Tc. Overall, our results provide insights into the binding properties of tetracyclines and support the idea that EF-Tu is not their primary target.

1 Introduction

Tetracycline (Tc) is a broad-spectrum antibiotic, with activity against a wide range of gram-positive and gram-negative bacteria [1–3]. Tetracyclines also have diverse applications in molecular biology and biotechnology, being used in conjunction with the Tet Repressor protein to artificially control the expression of target genes [4–6]. In addition, they have a range of interesting, non-antibiotic properties; for example, they bind to and inhibit several human enzymes [7, 8]. The bacteriostatic action of tetracycline has been extensively studied over the last 50 years, and it is well-established that Tc blocks protein biosynthesis [9, 10]. Tc is thought to prevent the binding of aminoacylated tRNA to the A-site of messenger RNA-programmed ribosomes [3]. Thus, the primary molecular target of Tc is thought to be the ribosome. Two structural studies in which Tc was diffused into pregrown crystals of the 30S ribosomal subunit identified, respectively, two and six binding sites in which Tc makes direct contacts with ribosomal RNA [11, 12]. Recently, we reported a computational study of Tc binding to the ribosome [13]. Molecular dynamics and continuum electrostatic methods were employed, providing evidence that one site is predominant, consistent with its higher crystallographic occupancy in the X-ray structures. In this binding site, Tc primarily interacts with the H34 region of the 16S rRNA [11, 12], which is involved in the binding of aminoacyl-tRNA. This suggests further that this site is the primary target of Tc action. However, another possibility was raised when Tc was shown to crystallize as a 1:1 complex with the trypsin-modified form of *Escherichia coli* EF-Tu:Mg:GDP [14]. Thus, EF-Tu might be one of several Tc targets. Several classes of antibiotics exist that bind to EF-Tu and block protein synthesis [15–17]. The structure of the EF-Tu:Mg:GDP:Tc complex was recently reported [18], showing that Tc not only binds to EF-Tu:Mg:GDP, but does so by interacting with an important structural motif that is shared by all GTPases and many ATPases [18].

To clarify the possible biological relevance of Tc:EF-Tu binding, we use molecular dynamics and free energy simulations [19–24]. Such calculations have become possible because we recently developed the first high quality force field model for tetracycline and eleven of its analogues [25, 26]. As a first step, we perform a series of control calculations, considering ribosome binding by Tc and several analogs for which experimental data are available. We use alchemical, molecular dynamics free energy simulations (MDFE) to compute the relative binding free energies, obtaining good agreement with experiment. Next, we consider EF-Tu binding by Tc and two of its analogues: the potent antibiotic, tigecycline, and the non-antibiotic, 4-dedimethyl-Tc

(Fig. 1). We compare their binding to EF-Tu, the ribosome, and the Tet Repressor (TetR), which mediates the main mechanism of Tc resistance. We show that when Tc binds to the EF-Tu:GDP:Mg complex, the contribution of EF-Tu to the binding free energy is negligible; rather, the favorable binding free energy can be solely attributed to Tc interactions with the Mg ion and the GDP phosphate groups. Further calculations show that EF-Tu does not exhibit any preference for Tc binding over the non-antibiotic analog, 4-dedimethyl-Tc (4-ddma-Tc). In contrast, both the ribosome and TetR do show a marked preference for Tc binding. We also performed free energy simulations to compare EF-Tu binding of Tc and its analogue tigecycline. Tigecycline is found to bind much more weakly to EF-Tu than Tc. Finally, we examined the G1058C mutation in the ribosome, which is known to produce Tc resistance [27]. This mutation is shown to increase the Tc binding free energy. Along with recent results for Tc:ribosome binding and TetR induction by Tc [13, 26, 28], these data illustrate the modular structure of tetracyclines. Indeed, these molecules have one edge that primarily confirms receptor binding specificity, and another edge that controls metal binding and positioning, and can be used to act, through the metal ion, on a given environment [26]. Overall, our results strongly support the idea that the ribosome, not EF-Tu, is the primary target of tetracycline.

2 Results

2.1 Control calculations: ribosome binding by Tc variants

We considered ribosome binding by eight Tc variants, shown in Fig. 1. Experimental data are available for plain Tc and six of the variants (Table 1). The relative binding free energies were computed through alchemical, molecular dynamics free energy simulations (referred to as "MDFE") [20, 23, 29–31]. Earlier computational work strongly suggests that Tc binds to the ribosome in its zwitterionic form, predominantly in the primary, TET1 site in the 30S subunit, stabilized by contacts between its co-bound Mg and three phosphate groups [11, 13]. We assume the other seven ligands bind in the same manner. The simulations follow the horizontal legs of the thermodynamic cycle in Fig. 2A, alchemically transforming the ligand from one Tc variant into another. The corresponding work is obtained from a thermodynamic integration formula (see Methods). Results are summarized in Table 1. More details on convergence and uncertainty are given in Supplementary Material. For two ligands (chlortetracycline, oxytetracycline), the experimental data have the form of *in vivo* Minimum Inhibitory Concentra-

tions [32], which only provide a qualitative test. For one other ligand (6-demethyl-Tc), no experimental data are available. For 4-ddma-Tc, there is no observable binding [33], which is consistent with the large, positive, computed $\Delta\Delta G$. Overall, the calculations agree very well with experiment. This indicates that the force field is accurate and the MDFE protocol is reasonable. Both the force field and the MDFE methodology have been further tested in the past [23, 26, 28].

2.2 Analysis of the experimental structures of EF-Tu with and without Tc

A recent X-ray structure [18] revealed that Tc binds on the surface of domain 1 of EF-Tu:GDP, as shown in panel A of Fig. 3. Tc's phenol-diketone moiety interacts directly with the Mg ion in the GTPase active site of EF-Tu. The O12a oxygen of Tc interacts with the phosphate group of GDP. Thr25 coordinates the Mg ion through its hydroxyl group; Asp80 also interacts with the Thr25 hydroxyl group. Tc does not interact directly with either sidechain; the distance between the Asp80 oxygen and the Thr25 oxygen is 2.57 Å, whereas the Tc oxygen O10 is 3.27 Å away. To interact with Asp80, which is out of the plane of the D ring of Tc, the Tc hydrogen of the O12 hydroxyl group would have to be rotated by 90° out-of-plane, which is very unfavorable (for the hydroxyl group of phenol, the rotation barrier is about 3 kcal/mol [34]). The second oxygen of Asp80 makes a hydrogen bond to Thr16 and to a water molecule that coordinates the Mg ion. No hydrogen bonds are apparent between Tc and any residues of EF-Tu. Tc replaces two well-ordered water molecules found in the Mg^{2+} coordination sphere in all other EF-Tu structures [18]. Overall, the Mg ion is coordinated by two water molecules, the Thr25 oxygen, two oxygens of Tc, and one oxygen of the terminal phosphate group of GDP. Superposition of EF-Tu:GDP with and without Tc (PDB entry code 1DG1) shows that there are no significant conformational changes upon Tc binding (see panel A of Fig. 3). The rms deviation between the two structures for the backbone atoms within a 10 Å sphere around GDP is just 0.31 Å; for the GDP atoms, the rms deviation is 0.30 Å.

2.3 Analysis of the structures from the simulations

The complex between Tc in its zwitterionic tautomer and EF-Tu:GDP was simulated by molecular dynamics (MD) for 5 nanoseconds (ns). The structures sampled in the simulation are in good agreement with the X-ray structure. The rms deviation, aver-

aged over the last 1 ns of dynamics and over backbone atoms is 0.62 Å; the average for nonhydrogen atoms within 10 Å of Tc is 0.91 Å. The EF-Tu ligands, Mg^{2+}, GDP, and Tc all maintain their positions relative to the X-ray structure. Mg^{2+} coordinates the OG1 oxygen of Thr25, a phosphate oxygen of GDP, two oxygens of Tc, and makes water-mediated interactions with Asp80 throughout the MD simulation. Tetracycline does not make direct hydrogen bonds to EF-Tu; the 4-dimethylamino and 3-enolate groups of the Tc ring A, which are important for its antibiotic function (discussed below) are completely solvent-exposed.

2.4 Mg^{2+} stabilizes the EF-Tu:GDP complex

In the EF-Tu:GDP crystal structure without Tc, an Mg ion is present [18]. As a first step, we examine here its role in GDP binding to EF-Tu. This step is needed for the analysis of Tc binding below. We use MDFE simulations where the Mg ion is gradually introduced, either in complex with GDP, or in the protein, according to the thermodynamic cycle in Fig. 2C. In principle, these simulations should allow us to compute the relative free energies for GDP:EF-Tu binding, with and without a co-bound Mg. In practice, we employ a so-called "fixed-charge" force field: atomic charges have fixed values, and the electronic polarization of water, GDP, or protein by the divalent Mg^{2+} is not explicitly modelled (see Methods). A consequence is that the force field cannot provide very accurate binding thermodynamics for reactions involving Mg. However, it can still give useful information on the incremental contribution of Mg to GDP:EF-Tu binding. Indeed, in the protein complex, the Mg^{2+} charge is shielded and neutralized by the GDP phosphate groups. Furthermore, the Mg ion does not interact closely with highly-polarizable protein groups. Its octahedral coordination sphere only includes one EF-Tu atom: the OG1 oxygen of Thr25 (Fig. 3). Thus, any force field artefacts should be fairly small when we consider the incremental effect of Mg on EF-Tu:GDP binding.

We therefore consider the thermodynamic cycle in Fig. 2C. In the vertical legs, GDP binds to EF-Tu. The horizontal legs introduce Mg, to form a complex either with GDP (upper leg) or with EF-Tu:GDP (lower leg). We assume that the allosteric, structural changes associated with EF-Tu:GDP binding are the same, whether or not Mg is co-bound to GDP. With this assumption, we can limit our analysis to the GDP-bound conformation of EF-Tu. Indeed, we may assume that to bind GDP, apo-EF-Tu is first brought into this conformation. The free energy to do this cancels, when the left and righthand legs of the thermodynamic cycle are compared. Therefore, the

allosteric changes do not play a role when analyzing the incremental effect of Mg on GDP binding.

The free energies of the horizontal legs are calculated by MDFE, giving ΔG_{EfTu} = -492.5±2.0 and ΔG_{GDP} = -478.3±6.0 kcal/mol (notations are defined in Fig. 2C). The difference between Mg introduction in the protein and in the GDP complex is $\Delta\Delta G$ = ΔG_{EfTu} - ΔG_{GDP} = -14.2 kcal/mol. Thus, Mg binding to GDP:EF-Tu is much stronger than Mg binding to GDP alone. Most importantly, $\Delta\Delta G$ can be interpreted as the difference in GDP binding free energies in the presence and absence of the Mg ion. We see that in the absence of Mg^{2+}, GDP binding to EF-Tu is much less favorable. Indeed, it was found experimentally that removing Mg^{2+} from EF-Tu:GDP accelerates the dissociation of GDP by a factor of 150–300 [35]. Therefore, we may conclude that GDP always binds to EF-Tu with a co-bound Mg ion, consistent with the GDP:EF-Tu X-ray structure [18]. In what follows, we will therefore assume that in EF-Tu, when GDP is present, Mg is always co-bound.

2.5 Tc binds to EF-Tu in its zwitterionic form

As a second step, we now establish which Tc tautomer is preferentially bound by the EF-Tu:GDP complex. We showed earlier that the preferred Tc tautomer in solution is the zwitterionic one, Tc_Z [28]. Furthermore, both TetR and the ribosome bind more strongly to Tc_Z than to the neutral form, Tc_N [13, 28]. Here, we compare Tc_N and Tc_Z binding to EF-Tu, using a Poisson-Boltzmann/Linear Response Approximation method, or PB/LRA [24, 28, 36–38]. This method follows the horizontal legs of the thermodynamic cycle in Fig. 2B: the Tc_N moiety is transformed into Tc_Z both in solution and in complex with the protein (see Methods). Notice that to bind Tc to EF-Tu:GDP, an Mg ion must first be removed from the protein (see above), since Tc brings its own Mg with it. This step is the same for Tc_N and Tc_Z, so it does not play any role in the Tc_N/Tc_Z comparison. Notice also that an Mg ion is co-bound to Tc both in solution and in the complex with EF-Tu:GDP; therefore, the reaction considered here does not require us to accurately describe the thermodynamics of Mg binding/release (see discussion in the previous section).

Structural ensembles for each state (Tc_Z or Tc_N) are obtained from MD simulations. The electrostatic potential on the ligand atoms is then computed using continuum electrostatics, by solving the Poisson-Boltzmann equation (see Methods). Finally, the potentials are combined with a Linear Response Approximation to obtain a free energy change [24, 28, 36–38]. Subtracting the two legs of the cycle yields the binding

free energy difference, $\Delta\Delta G$. The Poisson-Boltzmann calculations are done with a low protein dielectric constant of either one or two. This is the appropriate range for the present PB/LRA method [24, 28, 36–38], because most of the protein reorganization is explicitly modelled by performing MD simulations of the two endpoints (Tc_N and Tc_Z complexes). Results are summarized in Table 2. Both dielectric values lead to a $\Delta\Delta G$ of 0.1 kcal/mol, with an estimated uncertainty of about $\pm$0.5 kcal/mol. This means that EF-Tu:Mg:GDP does not exhibit any preference between the neutral and zwitterionic states of Tc. This contrasts with the Tet Repressor and the ribosome, both of which have a strong preference for Tc_Z binding. The lack of preference can be explained by the fact that the 4-dimethylamino and enolate moieties of Tc are solvent exposed in the EF-Tu complex (see panel B of Fig. 3), and do not interact directly with any protein residues or with GDP. The Tc_N/Tc_Z stability difference in solution has been estimated quantum mechanically to be about -3 kcal/mol in favor of Tc_Z [25, 39]; this means that Tc binds to EF-Tu in its zwitterionic form, even though EF-Tu itself does not discriminate between Tc_Z and Tc_N.

2.6 A component analysis of the binding free energy between Tc and EF-Tu

We now turn to a group or "component" analysis of the Tc:EF-Tu binding free energy. This analysis will show that Tc binding is mainly promoted not by EF-Tu itself, but by the GDP-associated Mg ion. MDFE and PB/LRA methods were used above to compute binding free energy differences. A third method is used here: the Poisson-Boltzmann Free Energy method, or PBFE [19, 21, 23, 24]. This method gives an estimate of the contribution of electrostatic interactions to the binding free energy, using the vertical legs of the thermodynamic cycle in Fig. 2A; see Methods. The PPBFE method is less rigorous than MDFE or PB/LRA; its advantage is that it provides an intuitive decomposition of the binding free energy into contributions from individual protein groups [13, 23, 24, 40, 41]. PBFE involves a single adjustable parameter, the protein/ligand dielectric constant, ϵ_P. It is set here to four, based on extensive applications to other protein:ligand complexes [19, 21, 23, 24, 28]. This differs from the PB/LRA situation above, where protein reorganization was explicitly modelled through MD simulations, so that a lower dielectric (1–2) was appropriate. For comparison, we also did calculations with a value of $\epsilon_P = 8$. Although the absolute magnitude of the group contributions varies with ϵ_P, the relative contributions of different groups, which are of interest here, are robust with respect to the choice of ϵ_P. We use the PBFE

method to analyze Tc_Z binding to the EF-Tu:GDP:Mg complex.

Results are summarized in Table 3. The largest single-residue contribution to the binding free energy between Tc and EF-Tu comes from Lys24, which interacts directly with the terminal phosphate group of the GDP. This is explained by the opposite charges on the negative Tc and the positive lysine, which prefers to interact with $[Tc:Mg]^+$, rather than Mg^{2+} alone. In contrast, interactions with Asp80, Thr25, Asp21 and Asp86 all favor binding of Mg^{2+} over $[Tc:Mg]^+$. Strikingly, the total contribution of all EF-Tu residues to the binding free energy of Tc is small: 0.9 kcal/mol, within the $\sim$2 kcal/mol uncertainty of the PBFE method. As expected, the largest positive contribution comes from GDP, which is negatively charged, and prefers Mg^{2+} over $[Tc:Mg]^+$ binding. Overall, in the standard state conditions simulated here, it is the Mg ion itself that is solely responsible for the binding of $[Tc:Mg]^+$, rather than Mg^{2+} alone, to EF-Tu. In the crystallographic experiment [18], Tc without any bound Mg was diffused into the EF-Tu:GDP crystals. The Mg concentration was thus sub-stoechiometric with respect to Tc and GDP: there is not enough Mg to saturate both EF-Tu:GDP and any unbound Tc. Therefore, each Tc shares a Mg with EF-Tu:GDP, and the interactions with Mg drive Tc binding. As the Mg concentration becomes larger, unbound Tc becomes associated with Mg, and Tc will not bind to EF-Tu. Overall, the lack of any intrinsic preference for $[Tc:Mg]^+$ over Mg^{2+} on the part of EF-Tu suggests that EF-Tu does not play the role of a specific Tc target.

2.7 Tc vc. non-antibiotic 4-dedimethylamino-Tc binding to the ribosome, EF-Tu, and the Tet Repressor

The possible role of EF-Tu as a Tc target was analyzed further by comparing the binding of two Tc analogs, one of which is a potent antibiotic, while the other is a non-antibiotic. We first consider 4-dedimethylamino-Tc (4-ddma-Tc; Fig. 1), which is not an antibiotic [33]. Additionally, it has a much weaker binding affinity than Tc to the Tet Repressor (TetR), which is a part of the antibiotic resistance mechanism [2, 42]. We test the competition between Tc and 4-ddma-Tc binding to EF-Tu, TetR and the ribosome. We use MDFE simulations (see Methods), which do not involve any adjustable parameters [23, 30, 31]. The simulations predict a Tc/4-ddma-Tc binding free energy difference for TetR of 4.6$\pm$1.6 kcal/mol (Table 4). This is consistent with experiment, which indicates a binding free energy difference of 5.0 kcal/mol or more [2]. Thus, 4-ddma-Tc binding is much weaker than Tc binding, mainly because of lost hydrogen bonds with His62, Asn82, and Gln116. Details on the convergence and

uncertainty of the MDFE simulations are given in Supplementary Material.

The Tc/4-ddma-Tc free energy difference for ribosome binding was computed to be $\Delta\Delta G = 4.0 \pm 1.1$ kcal/mol, strongly favoring Tc binding (Table 4). We attribute this result to the loss of hydrogen bonds between the Tc 4-amino group and the phosphate of Gua966, and to water-mediated interactions between Tc and rRNA. Indeed, we can decompose the free energy changes into components associated with the Coulomb electrostatic and the van der Waals energy terms, respectively. Subtracting the results in solvent and the ribosome complex, we find that $\Delta\Delta G$ arises mostly from the Coulomb terms (+2.5 kcal/mol), with a smaller van der Waals component (+1.5 kcal/mol).

Finally, in contrast to TetR and the ribosome, the Tc/4-ddma-Tc free energy difference for EF-Tu binding was found to be very small: $\Delta\Delta G$ = -0.1±0.6 kcal/mol (Table 4). The estimated uncertainty is consistent with the value that would be inferred from the 2 kcal/mol forward/backward hysteresis (±1 kcal/mol; data not shown). The small $\Delta\Delta G$ value arises because the 4-dimethylamino and 3-enolate groups of Tc are solvent-exposed in the complex with EF-Tu; see Fig. 3. The portion of Tc that binds directly to EF-Tu:GDP:Mg is the same in Tc and 4-ddma-Tc. Thus, EF-Tu does not distinguish between the antibiotic Tc and the non-antibiotic 4-ddma-Tc. This appears inconsistent with a role for EF-Tu as a specific Tc target.

2.8 Tc vc. tigecycline binding to the ribosome and EF-Tu

Tigecycline, the 9-t-butylglycylamido derivative of minocycline (Fig. 1), is a broad-spectrum antibiotic that is not affected by classical Tc resistance mechanisms, including ribosomal protection and efflux by Tc-specific pumps [43, 44]. Tigecycline was also found to bind five times more strongly than Tc to the ribosome [45, 46]. Bauer et al. showed by Fe^{2+}-mediated cleavage that tigecycline binds to the same high-affinity site in the 16S rRNA as Tc [47]. Here, we test the competition between Tc/tigecycline binding to EF-Tu and the ribosome using MDFE (see Methods). Results are summarized in Table 5. For ribosome binding, the simulations predict a Tc/tigecycline binding free energy difference of -0.8±1.0 kcal/mol, in good agreement with the experimental value of -1.0 kcal/mol [45]. This value results from the attractive electrostatic interaction between the positively-charged 9-t-butylglycylamido group of tigecycline and the negative phosphate group of Ade1055 in the ribosome.

The same transformation was performed in the elongation factor EF-Tu. The computed binding free energy difference is extremely large, 25.1±2.0 kcal/mol, showing that tigecycline does not bind to EF-Tu at all. More precisely, it cannot bind in the

same manner as Tc; other binding sites and modes cannot be ruled out. The large $\Delta\Delta G$ value results from steric clashes between the 9-t-butylglycylamido substituent of tigecycline and residues 62–65 and 80–82 of EF-Tu (see panel A of Fig. 3). If tigecycline and Tc share the same type of bacteriostatic function, this result strongly suggests that EF-Tu cannot be the primary target of Tc.

2.9 The ribosome resistance mutation G1058C reduces Tc binding

As a final illustration of the binding properties of Tc to the ribosome, we consider the effect of a resistance mutation on Tc binding. The G1058C mutation in the 16S rRNA was identified in clinical isolates of Tc-resistant *Cutaneous Propionibacteria* [27]. This mutation was also found in isolates of *Brachyspira Hyodysenteriae* that have increased Minimal Inhibitory Concentrations (MICs) for doxycycline, another Tc analogue [48]. The experimental binding free energy difference was estimated from the MICs to be 2.2±1.0 kcal/mol [27, 48]. We study the effect of the G1058C mutation on Tc binding. MD simulations were performed for both the wild type ribosome:Tc^{-}:Mg^{2+} complex and the G1058C mutant (Fig. 4). The binding free energy difference was then estimated by PBFE (see above). This method is less rigorous than the MDFE and PB/LRA methods used above. However, it should give a qualitative estimate of the binding free energy change. The simulation and PBFE setup were described earlier [13] and in Methods. With a ribosome dielectric constant of eight, used earlier [13], the computed binding free energy difference is 4.5 (0.7) kcal/mol, in rough agreement with the experimental value of 2.2 kcal/mol. With a somewhat higher dielectric of twelve, we would obtain 2.4 kcal/mol, close to the experimental value. Thus, the Tc-resistance phenotype is reflected in the Tc binding properties of the ribosome. In the MD simulations, the G1058C mutation leads to rearrangements in helix 34 of the 16S RNA, which in turn affect the Mg^{2+} coordination (Fig. 4), and lead to a decreased binding affinity for Tc^{-}:Mg^{2+}. The Mg ion that coordinates Tc also interacts with O1P of Cyt1054 in the wildtype ribosome, while in the G1058C mutant simulations, it loses this interaction and chelates O4' of Cyt1054 instead. As a result, the Tc position is slightly rotated compared to its wild type position.

3 Concluding discussion

Tetracycline binding to protein and RNA is a complex process, involving a competition between several effects, with electrostatics playing an important role. This is illustrated by our earlier, detailed studies of ribosome and TetR binding. Protonation states of the protein and the ligand were determined, as well as TetR sidechain orientations [13, 28]. Similarly, it was necessary here to determine which Tc tautomer binds to EF-Tu and whether the Mg ion is always pre-bound. Free energy simulations are a powerful technique to study such complex processes, complementary to experiment. After an initial burst of popularity in the early 1990s [30], technical and fundamental difficulties were brought to light that led to a drop in their use. In recent years, substantial progress has been made [20, 23, 24, 49], as illustrated here by the good agreement with experiment for our control calculations. Another major technical difficulty for simulation studies is the force field parameterization of Tc and its analogues, which must be done at a high level of accuracy. Our Tc parameterization [25] followed the same rigorous procedure that was used for the rest of the CHARMM force field [50, 51]. The same procedure was used more recently to parameterize 11 other Tc analogues, including 4-ddma-Tc and tigecycline [39]. This force field approach allows us to study both protein and nucleic acid binding with a comparable accuracy. Thus, we obtained good agreement with experiment above for Tc analogues binding to the ribosome; free energy calculations for several Tc analogues binding to TetR have given similar agreement (AA & TS, unpublished data).

To test the hypothesis of a functional role for Tc:EF-Tu binding, we have compared the binding strength of Tc and two analogues: tigecycline, one of the most potent Tc antibiotics, and the non-antibiotic, 4-ddma-Tc. Removal of the 4-dimethylamino group of Tc (giving 4-ddma-Tc) abolishes all its antimicrobial properties, which underlines the functional importance of this group [33]. The behavior of EF-Tu was compared to that of the ribosome, usually thought to be the primary biological target, and to that of the Tet Repressor, which mediates the primary mechanism of Tc resistance. With this approach, we have obtained evidence suggesting that EF-Tu is not part of the Tc bacteriostatic mechanism. Indeed, both TetR and the ribosome have a binding preference for Tc over 4-ddma-Tc, and the ribosome prefers to bind tigecycline over Tc. In contrast, EF-Tu does not have any preference for Tc over 4-ddma-Tc, and does not bind tigecycline at all. The lack of specificity for Tc over 4-ddma-Tc is not surprising, given the X-ray structure, where Tc does not make any direct hydrogen bonds to EF-Tu. However, for a quantitative statement, the present simulations were

necessary. Furthermore, the simulations reveal that EF-Tu does not contribute directly to the Tc binding affinity; rather, binding of Tc to EF-Tu:GDP:Mg is mediated by the Mg ion. Overall, our results are consistent with a broad range of experimental data. They suggest that EF-Tu is not the primary target of Tc.

The present data illustrate further the modular structure of tetracyclines. Indeed, these molecules are seen to have a "specificity" edge that primarily confers receptor binding specificity (roughly, the upper edge in Fig. 1), and another, "metal-display" edge that controls metal binding and positioning. Recent crystallographic and simulation studies [13, 52, 53] have shown that in the Tet Repressor protein, for example, the co-bound, divalent magnesium ion is primarily responsible for driving the protein through its allosteric transition into the induced state; tetracycline itself mainly plays the role of an adaptor, or metal carrier. In the ribosome, we recently used a free energy decomposition analysis to show that Tc binding is mainly driven by magnesium:phosphate interactions; the "specificity" edge of Tc presumably serves to fine-tune the binding and make it specific. Here, we have an even more extreme example of the modular Tc structure: only the metal-display edge is used to bind EF-Tu; the metal itself contributes all of the affinity, and there is no specificity at all. In the future, these and similar insights could help to engineer new antibiotics and new systems for gene regulation.

4 Materials and Methods

4.1 EF-Tu:Tc binding: computed structures and free energies

Molecular dynamics simulations Crystal structures of the elongation factor Tu (EF-Tu) were taken from the Protein Data Bank (PDB), entries 1DG1 (without bound Tc) and 2HCJ (with bound Tc) [18, 54]. The simulations included protein residues within a 26 Å sphere, centered on the Tc binding site. In addition to crystal waters, a 26 Å sphere of water was overlaid and waters that overlapped protein, GDP, crystal waters, Tc (when present), or Mg^{2+} were removed. Protein atoms between 20 Å and 26 Å from the sphere's center were harmonically restrained to their experimental positions. Simulations were done with the SSBP solvent model [55–57], which treats the region outside the 26 Å sphere as a uniform dielectric continuum, with a dielectric constant of 80. This is reasonable, since most of the outer region is water. Newtonian dynamics were used for the innermost region, within 20 Å of the sphere's center; Langevin dynamics were used for the outer part of the sphere, with a 292 K bath. The

CHARMM22 force field was used for the protein [51] and the TIP3P model for water [58]. Tc was described with the force field developed previously [25]. The analogues tigecycline and 4-ddma-Tc were parameterized in a similar way [26]. Electrostatic interactions were computed without any cutoff, using a multipole approximation for distant groups [59]. Calculations were done with the CHARMM program [60]. The same MD protocols were used for the class D Tet Repressor (TetR) and the *Thermus thermophilus* 30S ribosomal particles; see [13, 28] for more details. Notice that there are a few non-standard nucleotides in the *Thermus thermophilus* 30S ribosomal particle [61]. Only two are within 12 Å of either Tc binding site: m^2Gua-966 has a methyl group about 11 Å from the nearest TET1 atom, pointing away from the ligand; m^5Cyt-967 is somewhat closer, with its additional methyl about 6 Å from the acetamide nitrogen of Tc in the TET1 site (lower-rightmost atom in the Fig. 1 orientation). In the absence of force field parameters for m^5Cyt, we replaced both nucleotides by standard nucleotides, assuming that the additional methyl groups do not significantly affect the relative binding affinities of the various Tc variants. This should be reasonable, since the Cyt-967 methyl is 6 Å away and the Tc variants all share the exact same chemistry in the vicinity of their acetamide group (Fig. 1).

Alchemical MD free energy simulations (MDFE) To compare Tc and a Tc analog (say, Tc') binding to a receptor (TetR, EF-Tu or the ribosome), we use the thermodynamic cycle in Fig. 2A. The MDFE method follows the horizontal legs of the cycle. Tc is reversibly transformed into its analog during a series of MD simulations. The corresponding work is derived from a thermodynamic integration formula [31].

Most of the simulations were performed with the spherical boundary conditions described above (see also [62, 63]): we considered receptor atoms within a sphere with a 25 Å radius, centered on an Mg ion that coordinates tetracycline in the TetR, EF-Tu, or the ribosome. However, the Tc/4-ddma-Tc transformations in the TetR protein and the ribosome were performed using a cubic box with a 74 Å edge. For these simulations, periodic boundary conditions were assumed and electrostatic interactions were approximated by the Particle Mesh Ewald (PME) method [64]; as before, portions of the receptor over 25 Å from the Mg were neglected, and receptor atoms between 20 and 25 Å were harmonically restrained.

For the lower leg of the thermodynamic cycle, we simulate Tc:Mg^{2+} in solution. For the upper leg, we simulate a portion of the Tc^-:Mg^{2+}:receptor complex (respectively, Tc^-:Mg^{2+}:EF-Tu or Tc^-:Mg^{2+}:Ribosome), solvated by the same 25 Å sphere or 74 Å water box. In each simulation system, the energy function can be expressed as a linear

combination of Tc^{-}:Mg^{2+} and Tc'^{-}:Mg^{2+} terms:

$$U(\lambda) = U_0 + (1 - \lambda)U(\mathrm{Tc}) + \lambda U(\mathrm{Tc}')) \tag{1}$$

where λ is a "coupling parameter" and U_0 represents interactions between parts of the system other than Tc. The free energy derivative with respect to λ has the form:

$$\frac{\partial G}{\partial \lambda}(\lambda) = \langle U(\mathrm{Tc}) - U(\mathrm{Tc}') \rangle_\lambda \tag{2}$$

where the brackets indicate an average over an MD trajectory with the energy function $U(\lambda)$ [23, 31]. We gradually mutated Tc into Tc' by changing λ from one to zero. The successive values of λ were: 0.999, 0.99, 0.95, 0.9, 0.8, 0.6, 0.4, 0.2, 0.1, 0.05, 0.01, 0.001. The free energy derivatives were computed at each λ value from a 100 ps MD simulation, or "window"; the last 80 ps of each window were used for averaging. A complete mutation run thus corresponded to 12 windows and 1.2 ns of simulation. Usually, one run was performed in each direction (Tc into Tc' and the reverse).

Accurate uncertainty estimation with MDFE is difficult and expensive [65–67]. A widely-used approach is to perform multiple runs and measure the dispersion between runs. While this seems plausible, it neglects some forms of systematic error, and can actually lead to an overestimated uncertainty. In particular, it is well known that runs performed in opposite directions ("forward" and "backward" transformations; λ increasing or decreasing) exhibit systematic hysteresis effects [68–70]. Thus, in the past, we have used error estimators that consider pairs of runs, one in each direction, forming a "forward/backward pair" [62, 63, 71]. The forward/backward averages were much more reproducible than the individual values. Here, we use a variant of this approach. For most of the transformations considered, we performed a single pair of runs, one in each direction. For a given run, we split each trajectory window (corresponding to a particular λ value) into two halves; each collection of "half windows" can be viewed as a "half run". A forward and a backward half run are viewed as a pair of runs, and the corresponding ΔG values are averaged. In this way, starting from one forward and one backward run, we get four "half runs", and two averaged ΔG values. The uncertainty is taken to be their difference (*i.e.*, two standard deviations). In practice, this method is roughly equivalent to estimating the uncertainty by block-averaging a single run [72, 73], and assuming systematic hysteresis effects mostly cancel whenever a forward and a backward run are averaged. To test this idea, we performed several pairs of forward/backward runs for two of the larger transformations (Tc vs. 4-ddma-Tc, binding to TetR and to the ribosome). In both cases, the dispersion of the forward/backward

averages was close to the uncertainty estimated by the previous method. The error estimated from the mean hysteresis (the difference between a forward and a backward run) was also not much greater; see Supplementary Material for details.

Poisson-Boltzmann Linear Response Approximation (PB/LRA) A recently-developed method that combines MD simulations with continuum electrostatics is the Poisson-Boltzmann/Linear Response Approximation, or PB/LRA method [24, 28, 36–38]. This method is well-suited to treat rearrangements of atomic charges, as in the Tc_N to Tc_Z transformation. The free energy change (ΔG_{prot}, ΔG_{solv} in Fig. 2B) is approximated by the continuum electrostatic free energy, averaged over the equilibrium states before and after the rearrangement. We applied the method using MD structures of the Tc_N^-:Mg^{2+} and Tc_Z^-:Mg^{2+} states (see above). We averaged free energies over 250 structures from the last 1 ns of the MD simulations.

Acknowledgements

Support was provided by the Nuclear Toxicology program of the Commmissariat pour l'Énergie Atomique and the Centre National pour la Recherche Scientifique (to TS). An Egide PhD fellowship was awarded to AA. We thank Winfried Hinrichs for discussions and Martin Karplus for the CHARMM program. Some of the simulations were performed at the CINES supercomputing center of the French Ministry of Education.

Supporting Information Available:

Full ref 32; details of free energy simulations and their convergence.

References

[1] Taubes, G. The bacteria fight back. *Science 321* (2008), 356–361.

[2] Lederer, T., Kintrup, M., Takahashi, M., Sum, P., Ellestad, G., and Hillen, W. Tetracycline analogs affecting binding to Tn10-Encoded Tet repressor trigger the same mechanism of induction. *Biochemistry 35* (1996), 7439–7446.

[3] Chopra, I., and Roberts, M. Tetracycline antibiotics: Mode of action, applications, molecular biology, and epidemiology of bacterial resistance. *Microbiology and Molecular Biology Reviews 65* (2001), 232–260.

[4] Fussenegger, M. The impact of mammalian gene regulation concepts on functional genomic research, metabolic engineering, and advanced gene therapies. *Biotech. Prog. 17* (2001), 1–51.

[5] Berens, C., and Hillen, W. Gene regulation by tetracyclines: Constraints of resistance regulation in bacteria shape TetR for application in eukaryotes. *Eur. J. Biochem. 270* (2003), 3109–3121.

[6] Berens, C., Lochner, S., Lober, S., Usai, I., Schmidt, A., Drueppel, L., Hillen, W., and Gmeiner, P. Subtype selective tetracycline agonists and their application for a two-stage regulatory system. *ChemBioChem 7* (2006), 1320–1324.

[7] Pruzanski, W., Stefanski, E., Vadas, P., McNamara, T. E., Ramamurthy, N., and Golub, L. M. Chemically modified non-antimicrobial tetracyclines inhibit activity of phospholipases A(2). *J. Rhumat. 25* (1998), 1807–1812.

[8] Hamdan, I. I., Afifi, F., and Taha, M. O. In vitro alpha amylase inhibitory effect of some clinically-used drugs. *Pharmazie 59* (2004), 799–801.

[9] Schnappinger, D., and Hillen, W. Tetracyclines: antibiotic action, uptake, and resistance mechanisms. *Arch. Microbiol. 165* (1996), 359–369.

[10] Chopra, I., Hawkey, P. M., and Hinton, M. Tetracyclines, molecular and clinical aspects. *J. Antimicrob. Chemother. 29* (1992), 245–277.

[11] Brodersen, D. E., Clemons Jr, W. M., Carter, A. P., Morgan-Warren, R. J., Wimberly, B. T., and Ramakrishnan, V. The structural basis for the action of the antibiotics tetracycline, pactamycin, and hygromycin B on the 30S ribosomal subunit. *Cell 103* (2000), 1143–1154.

[12] Pioletti, M., Schlünzen, F., Harms, J., Zarivach, R., Gluhmann, M., Avila, H., Bashan, A., Bartels, H., Auerbach, T., Jacobi, C., Hartsch, T., Yonath, A., and Franceschi, F. Crystal structures of complexes of the small ribosomal subunit with tetracycline, edeine and IF3. *EMBO J. 20* (2001), 1829–1839.

[13] Aleksandrov, A., and Simonson, T. Molecular dynamics simulations of the 30S ribosomal subunit reveal a preferred tetracycline binding site. *J. Am. Chem. Soc. 130* (2008), 1114–1115.

[14] Mui, S., Delaria, K., and Jurnak, F. Preliminary crystallographic analysis of a complex between tetracycline and the trypsin-modified form of *escherichia coli* elongation factor Tu. *J. Mol. Biol. 212* (1990), 445–447.

[15] Krab, I. M., and Parmeggiani, A. Mechanisms of EF-Tu, a pioneer GTPase. *Prog. Nucleic Acids Res. Mol. Biol. 71* (2002), 513–551.

[16] Parmeggiani, A., Krab—, I. M., Okamura, S., Nielsen, R. C., Nyborg, J., and Nissen, P. Structural basis of the action of pulvomycin and GE2270 A on Elongation factor Tu. *Biochemistry 45* (2006), 6846–6857.

[17] Zuurmond, A.-M., Olsthoorn-Tieleman, L. N., de Graaf, J. M., Parmeggiani, A., and Kraal, B. Mutant EF-Tu species reveal novel features of the enacyloxin IIa inhibition mechanism on the ribosome. *J. Mol. Biol. 294* (1999), 627–637.

[18] Heffron, S. E., Mui, S., Aorora, A., Abel, K., Bergmann, E., and Jurnak, F. Molecular complementarity between tetracycline and the GTPase active site of elongation factor Tu. *Acta Cryst. D 62* (2006), 1392–1400.

[19] Kollman, P., Massova, I., Reyes, C., Kuhn, B., Huo, S., Chong, L., Lee, M., Lee, T., Duan, Y., Wang, W., Donini, O., Cieplak, P., Srinivasan, J., Case, D., and Cheatham, T. Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. *Acc. Chem. Res. 33* (2000), 889–897.

[20] Jorgensen, W. The many roles of computation in drug discovery. *Science 303* (2003), 1813–1818.

[21] McCammon, J. Computation of noncovalent binding affinities. In *Theory and Applications of Computational Chemistry*, C. Dykstra, G. Frenking, K. Kim, and G. Scuseria, Eds. Elsevier, Amsterdam, 2005, pp. 41–46.

[22] Becker, O., Mackerell Jr., A., Roux, B., and Watanabe, M., Eds. *Computational Biochemistry & Biophysics*. Marcel Dekker, New York, 2001.

[23] Simonson, T., Archontis, G., and Karplus, M. Free energy simulations come of age: the protein–ligand recognition problem. *Acc. Chem. Res. 35* (2002), 430–437.

[24] Simonson, T. Free energy calculations: approximate methods for biological macromolecules. In *Free energy calculations: theory and applications in chemistry and biology*, C. Chipot and A. Pohorille, Eds. Springer Verlag, N.Y., 2006, ch. 12.

[25] Aleksandrov, A., and Simonson, T. The tetracycline:Mg^{2+} complex: a molecular mechanics force field. *J. Comp. Chem. 13* (2006), 1517–1533.

[26] Aleksandrov, A., Schuldt, L., Hinrichs, W., and Simonson, T. Tet repressor induction by tetracycline: a molecular dynamics, continuum electrostatics, and crystallographic study. *J. Mol. Biol. 378* (2008), 896–910.

[27] Ross, J., Eady, E., Cove, J., and Cunliffe, W. 16S rRNA mutation associated with tetracycline resistance in a Gram-positive bacterium. *Antimicrob. Agents Chemoth. 42* (1998), 1702–1705.

[28] Aleksandrov, A., Proft, J., Hinrichs, W., and Simonson, T. Protonation patterns in tetracycline:Tet repressor recognition: Simulations and experiments. *ChemBioChem 8* (2007), 675–685.

[29] Tembe, B., and McCammon, J. Ligand-receptor interactions. *Comp. Chem. 8* (1984), 281–283.

[30] Kollman, P. Free energy calculations: applications to chemical and biochemical phenomena. *Chem. Rev. 93* (1993), 2395.

[31] Simonson, T. Free energy calculations. In *Computational Biochemistry & Biophysics*, O. Becker, A. Mackerell Jr., B. Roux, and M. Watanabe, Eds. Marcel Dekker, N.Y., 2001, ch. 9.

[32] Trinh, H., Billington, S., Field, A., Songer, J., and Jost, B. Susceptibility of *arcanobacterium pyogenes* from different sources to tetracycline, macrolide and lincosamide antimicrobial agents. *Vet. Microbiol. 85* (2002), 353–359.

[33] Greenwald, R. A., Hillen, W., and Nelson, M. L. *Tetracyclines in Biology, Chemistry and Medicine*. Birkhäuser Verlag/Switzerland, 2001.

[34] Foloppe, N., and MacKerell Jr., A. D. All-atom empirical force field for nucleic acids: I. Parameter optimization based on small molecule and condensed phase macromolecular target data. *J. Comp. Chem. 21* (2000), 86–104.

[35] Gromadski, K., Wieden, H.-J., and Rodnina, M. Kinetic mechanism of Elongation Factor Ts-catalyzed nucleotide exchange in Elongation Factor Tu. *Biochemistry 41* (2002), 162–169.

[36] Sham, Y., Chu, Z., and Warshel, A. Consistent calculations of pK_a's of ionizable residues in proteins: semi-microscopic and microscopic approaches. *J. Phys. Chem. B 101* (1997), 4458–4472.

[37] Eberini, I., Baptista, A., Gianazza, E., Fraternali, F., and Beringhelli, T. Reorganization in apo- and holo-β-lactoglobulin upon protonation of Glu89: molecular dynamics and pK_a calculations. *Proteins 54* (2004), 744–758.

[38] Archontis, G., and Simonson, T. Proton binding to proteins: a free energy component analysis using a dielectric continuum model. *Biophys. J. 88* (2005), 3888–3904.

[39] Aleksandrov, A., and Simonson, T. Molecular mechanics models for tetracycline analogs. *J. Comp. Chem. in press* (2008), 0000.

[40] Archontis, G., Simonson, T., and Karplus, M. Binding free energies and free energy components from molecular dynamics and Poisson-Boltzmann calculations. Application to amino acid recognition by aspartyl-tRNA synthetase. *J. Mol. Biol. 306* (2001), 307–327.

[41] Hendsch, Z., and Tidor, B. Electrostatic interactions in the GCN4 leucine zipper: substantial contributions arise from intramolecular interactions enhanced on binding. *Prot. Sci. 8* (1999), 1381–1392.

[42] Scholz, O., Kostner, M., Reich, M., Gastiger, S., and Hillen, W. Teaching TetR to recognize a new inducer. *J. Mol. Biol. 329* (2003), 217–227.

[43] Boucher, H. W., Wennersten, C. B., and Eliopoulos, G. M. In vitro activities of the glycylcycline GAR-936 against gram-positive bacteria. *Antimicrob. Agents Chemother. 44* (2000), 2225–2229.

[44] Chopra, I. New developments in tetracycline antibiotics: glycylcyclines and tetracycline efflux pump inhibitors. *Drug Resist. Updates 5* (2002), 119–125.

[45] Bergeron, J., Ammirati, M., Danley, D., James, L., Norcia, M., Retsema, J., Strick, C., Su, W., Sutcliffe, J., and Wondrack, L. Glycylcyclines bind to the high-affinity tetracycline ribosomal binding site and evade Tet(M)- and Tet(O)-mediated ribosomal protection. *Antimicrob. Agents Chemother. 40* (1996), 2226–2228.

[46] Olson, M. W., Ruzin, A., Feyfant, E., Rush, T. S., O'Connell, J., and Bradford, P. A. Functional, biophysical, and structural bases for antibacterial activity of tigecycline. *Antimicrob. Agents Chemoth. 50* (2006), 2156–2166.

[47] Bauer, G., Berens, C., Projan, S. J., and Hillen, W. Comparison of tetracycline and tigecycline binding to ribosomes mapped by dimethylsulphate and drug-directed Fe^{2+} cleavage of 16S rRNA. *J Antimicrob Chemother. 53*, 4 (2004), 592–599.

[48] Pringle, M., Fellstrom, C., and Johansson, K. E. Decreased susceptibility to doxycycline associated with a 16S rRNA gene mutation in *brachyspira hyodysenteriae*. *Veterinary Microbiology 123*, 1–3 (2007), 245–248.

[49] Wong, C., and McCammon, J. Protein simulation and drug design. *Adv. Prot. Chem. 66* (2003), 87–121.

[50] Mackerell, A., Wiorkiewicz-Kuczera, J., and Karplus, M. An all-atom empirical energy force-field for the study of nucleic acids. *J. Am. Chem. Soc. 117* (1995), 11946–11975.

[51] Mackerell, A., Bashford, D., Bellott, M., Dunbrack, R., Evanseck, J., Field, M., Fischer, S., Gao, J., Guo, H., Ha, S., Joseph, D., Kuchnir, L., Kuczera, K., Lau, F., Mattos, C., Michnick, S., Ngo, T., Nguyen, D., Prodhom, B., Reiher, W., Roux, B., Smith, J., Stote, R., Straub, J., Watanabe, M., Wiorkiewicz-Kuczera, J., Yin, D., and Karplus, M. An all-atom empirical potential for molecular modelling and dynamics study of proteins. *J. Phys. Chem. B 102* (1998), 3586–3616.

[52] Hinrichs, W., Kisker, C., Duvel, M., Muller, A., Tovar, K., Hillen, W., and Saenger, W. Structure of the Tet repressor-tetracycline complex and regulation of antibiotic resistance. *Science 264* (1994), 418–420.

[53] Saenger, W., Orth, P., Kisker, C., Hillen, W., and Hinrichs, W. The tetracycline repressor: A paradigm for a biological switch. *Angew. Chem. Int. Ed. 39* (2000), 2042–2052.

[54] Abel, K., Yoder, M. D., Hilgenfeld, R., and Jurnak, F. An alpha to beta conformational switch in EF-Tu. *Structure 4* (1996), 1153–1159.

[55] Beglov, D., and Roux, B. Finite representation of an infinite bulk system: solvent boundary potential for computer simulations. *J. Chem. Phys. 100* (1994), 9050–9063.

[56] Simonson, T. Electrostatic free energy calculations for macromolecules: a hybrid molecular dynamics/continuum electrostatics approach. *J. Phys. Chem. B 104* (2000), 6509–6513.

[57] Nina, M., and Simonson, T. Molecular dynamics of the $tRNA^{Ala}$ acceptor stem: comparison between continuum reaction field and particle-mesh Ewald electrostatic treatments. *J. Phys. Chem. B 106* (2002), 3696–3705.

[58] Jorgensen, W., Chandrasekar, J., Madura, J., Impey, R., and Klein, M. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys. 79* (1983), 926–935.

[59] Stote, R., States, D., and Karplus, M. On the treatment of electrostatic interactions in biomolecular simulation. *J. Chim. Phys. 88* (1991), 2419–2433.

[60] Brooks, B., Bruccoleri, R., Olafson, B., States, D., Swaminathan, S., and Karplus, M. Charmm: a program for macromolecular energy, minimization, and molecular dynamics calculations. *J. Comp. Chem. 4* (1983), 187–217.

[61] Guymon, R., Pomerantz, S. C., Crain, P. F., and McCloskey, J. A. Influence of phylogeny on posttranscriptional modification of rRNA in thermophilic prokaryotes: The complete modification map of 16S rRNA of *thermus thermophilus*. *Biochemistry 45* (2006), 4888–4899.

[62] Thompson, D., Plateau, P., and Simonson, T. Free energy simulations reveal long-range electrostatic interactions and substrate-assisted specificity in an aminoacyl-tRNA synthetase. *ChemBioChem 7* (2006), 337–344.

[63] Thompson, D., and Simonson, T. Molecular dynamics simulations show that bound Mg^{2+} contributes to amino acid and aminoacyl adenylate binding specificity in aspartyl-tRNA synthetase through long range electrostatic interactions. *J. Biol. Chem. 281* (2006), 23792–23803.

[64] Sagui, C., and Darden, T. Molecular dynamics simulations of biomolecules: long-range electrostatic effects. *Ann. Rev. Biophys. Biomol. Struct. 28* (1999), 155–179.

[65] Hodel, A., Simonson, T., Fox, R. O., and Brünger, A. T. Conformational substates and uncertainty in macromolecular free energy calculations. *J. Phys. Chem. 97* (1993), 3409–3417.

[66] Reinhardt, W., Miller, M., and Amon, L. Why is it so difficult to simulate entropies, free energies, and their differences? *Acc. Chem. Res. 34* (2001), 607–614.

[67] Shirts, M. R., Pitera, J. W., Swope, W. C., and Pande, V. S. Extremely precise free energy calculations of amino acid side chain analogs: Comparison of common molecular mechanics force fields for proteins. *J. Chem. Phys. 119* (2003), 5740–5761.

[68] Hermans, J. Simple analysis of noise and hysteresis in (slow-growth) free energy simulations. *J. Phys. Chem. 95* (1991), 9029–9032.

[69] Wood, R. H. Estimation of errors in free energy calculations due to the lag between the Hamiltonian and the system configuration. *J. Phys. Chem. 95* (1991), 4838–4842.

[70] Hummer, G. Fast growth thermodynamic integration: error and efficiency analysis. *J. Chem. Phys. 114* (2001), 7330–7337.

[71] Thompson, D., Lazennec, C., Plateau, P., and Simonson, T. Ammonium scanning in an enzyme active site: the chiral specificity of aspartyl-tRNA synthetase. *J. Biol. Chem. 42* (2007), 30856–30868.

[72] Straatsma, T., Berendsen, H., and Stam, A. Estimation of the statistical error in molecular simulation calculations. *Mol. Phys. 57* (1986), 89–95.

[73] Allen, M., and Tildesley, D. *Computer Simulations of Liquids*. Clarendon Press, Oxford, 1991.

[74] Failing, K., Theis, P., and Kaleta, E. Determination of the inhibitory concentration 50% (IC50) of four selected drugs (chlortetracycline, doxycycline, enrofloxacin and difloxacin) that reduce in vitro the multiplication of *chlamydophila psittaci*. *Deutsch. Tierarztl. Wochenschr. 113* (2006), 412–417.

Table 1: Tc analogs binding to the ribosome: MD free energy simulations

Tc analog	$^a\Delta G_{rib}$	$^a\Delta G_{sol}$	$\Delta\Delta G$	$\Delta\Delta G_{expl}$
minocycline	-17.2(1.1)	-15.0(0.4)	-2.2(1.2)	-2.6 [46]
doxycycline	25.2(0.4)	25.0(0.3)	0.2(0.5)	0.0 [27], 0.2 [74]
tigecycline	-112.3(0.8)	-111.5(0.7)	-0.8 (1.0)	-1.0 [45], -1.1 [46]
oxytetracycline	7.4(0.2)	7.0(0.3)	0.4(0.4)	0.0^b [32]
6-demethyl-Tc	8.4(0.2)	8.9(0.2)	-0.5(0.3)	NA
chlortetracycline	5.0(0.2)	5.1(0.2)	0.1(0.3)	-0.4^b [32]
4-ddma-Tc	-25.9(1.0)	-28.8(0.6)	4.0(1.1)	No binding^c

Free energies in kcal/mol. ^{a}Columns 2-3 are the free energy to reversibly go from plain Tc to one of its analogs, either in the ribosome (ΔG_{rib}) or in solution (ΔG_{sol}). $\Delta\Delta G = \Delta G_{rib} - \Delta G_{sol}$ is the relative binding free energy. A negative $\Delta\Delta G$ means the Tc analog binds more strongly. MDFE error bars are given in parentheses. ^{b}Derived from Minimum Inhibitory Concentrations *vivo*. ^{c}No observable binding [33].

Table 2: PB/LRA free energy simulations comparing Tc_N^- and Tc_Z^- binding to EF-Tu

	Protein dielectric constant	
	$\epsilon_P = 1$	$\epsilon_P = 2$
$\Delta G_{\mathrm{protein}}$	55.9 (0.4)	28.0 (0.2)
$\Delta G_{\mathrm{solvent}}$	55.8 (0.5)	27.9 (0.3)
$\Delta\Delta G$	0.1 (0.6)	0.1 (0.4)

Energies in kcal/mol; uncertanties are in parentheses. $\Delta G_{\mathrm{protein}}$ and $\Delta G_{\mathrm{solvent}}$ correspond to the alchemical transformation of Tc from one form into the other. $\Delta\Delta G = \Delta G_{\mathrm{protein}} - \Delta G_{\mathrm{solvent}}$ is the binding free energy difference. $\Delta\Delta G$ is very small, indicating that EF-Tu does not discriminate between the zwitterionic and neutral forms of Tc.

Table 3: Contributions of selected EF-Tu residues to the Tc binding free energy

residue	free energy contribution	residue	free energy contribution
Lys24	-9.0 (0.7)	Asp86	1.5 (0.8)
Asp21	2.1 (0.6)	Thr25	2.7 (0.8)
Asp80	7.8 (0.6)	GDP	25.1 (0.9)
		GDP:Mg^{2+}	-67.0 (1.0)
	all protein residues together		0.9 (1.0)

PB free energies are in kcal/mol; uncertainties are in parentheses. Only amino acids contributing more than than 1 kcal/mol (in absolute magnitude) are shown. Calculations were done with a protein dielectric constant $\epsilon_P = 4$.

Table 4: Tc binding to EF-Tu, the ribosome, and TetR: binding free energy differences $\Delta\Delta G$ between Tc and the non-antibiotic analog, 4-dedimethylamino-Tc

Target	$\Delta\Delta G_{comp}$	$\Delta\Delta G_{exptl}$
Tet Repressor	4.6 (1.6)	$> 5.0^a$
EF-Tu:GDP	-0.1 (0.6)	NA^b
Ribosome	4.0 (1.1)	NA^b

Energies in kcal/mol. Uncertainties are in parentheses. A positive $\Delta\Delta G$ corresponds to preferential Tc binding. ^{a}Experiment only gives a lower bound [2]. ^{b}Experimental values are not known.

Table 5: Tc binding to EF-Tu and the ribosome: binding free energy differences $\Delta\Delta G$ between Tc and the potent antibiotic, tigecycline

Target	$\Delta\Delta G_{comp}$	$\Delta\Delta G_{exptl}$
EF-Tu:GDP	25.1 (2.0)	NA[a]
Ribosome	-0.8 (1.0)	-1.0[b], -1.1[b]

Energies in kcal/mol. Uncertainties are in parentheses. A positive $\Delta\Delta G$ corresponds to preferential Tc binding. [a]Experimental value unknown. [b]Experimental values from references [45, 46].

Figure Captions

1. Plain tetracycline (Tc) and the seven analogues considered in this work.

2. **Thermodynamic cycles for ligand:receptor binding. A)** Comparing Tc and 4-ddma-Tc. Vertical legs represent Tc^{-}:Mg^{2+} or 4-ddma-Tc^{-}:Mg^{2+} binding to a receptor (TetR, EF-Tu, or ribosome). Horizontal legs represent the alchemical transformation of Tc^{-} into 4-ddma-Tc^{-} in solution (below) or in the receptor:Tc^{-}:Mg^{2+} complex (above). **B)** Comparing Tc tautomeric states. **C)** GDP:EF-Tu binding with and without co-bound Mg.

3. **(A):** superposition of the X-ray structures of EF-Tu with and without bound Tc, showing that there are no conformational changes upon Tc binding. **(B):** close-up view of the Tc binding site, showing that there are no hydrogen bonds between Tc and EF-Tu.

4. **Native and mutant ribosome interacting with Tc.** The MD structure of Tc bound to the native ribosome is black; the G1058 mutant is grey. The Mg^{2+} ions are shown as spheres. Tetracycline and the ribosomal RNA residues 1058, 1199, 1054 are in stick representation.

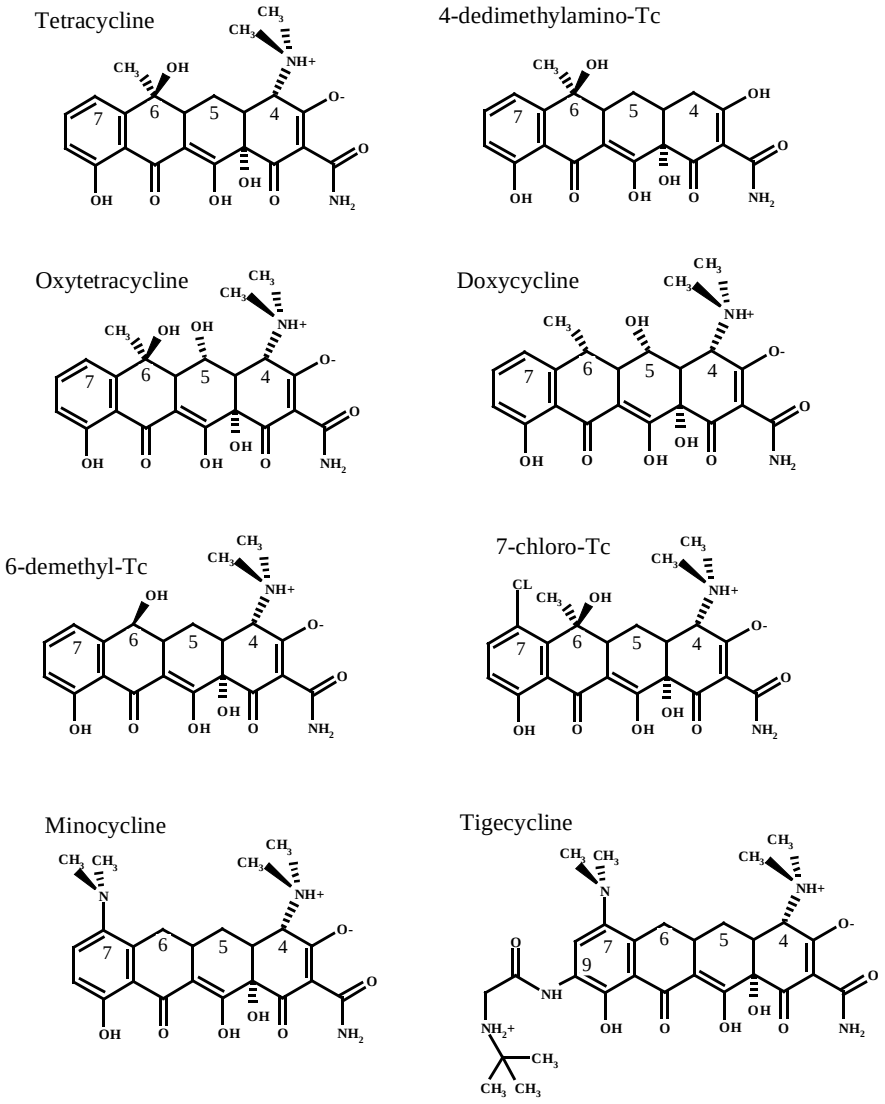


Figure 1: Plain tetracycline (Tc) and the seven analogues considered in this work.

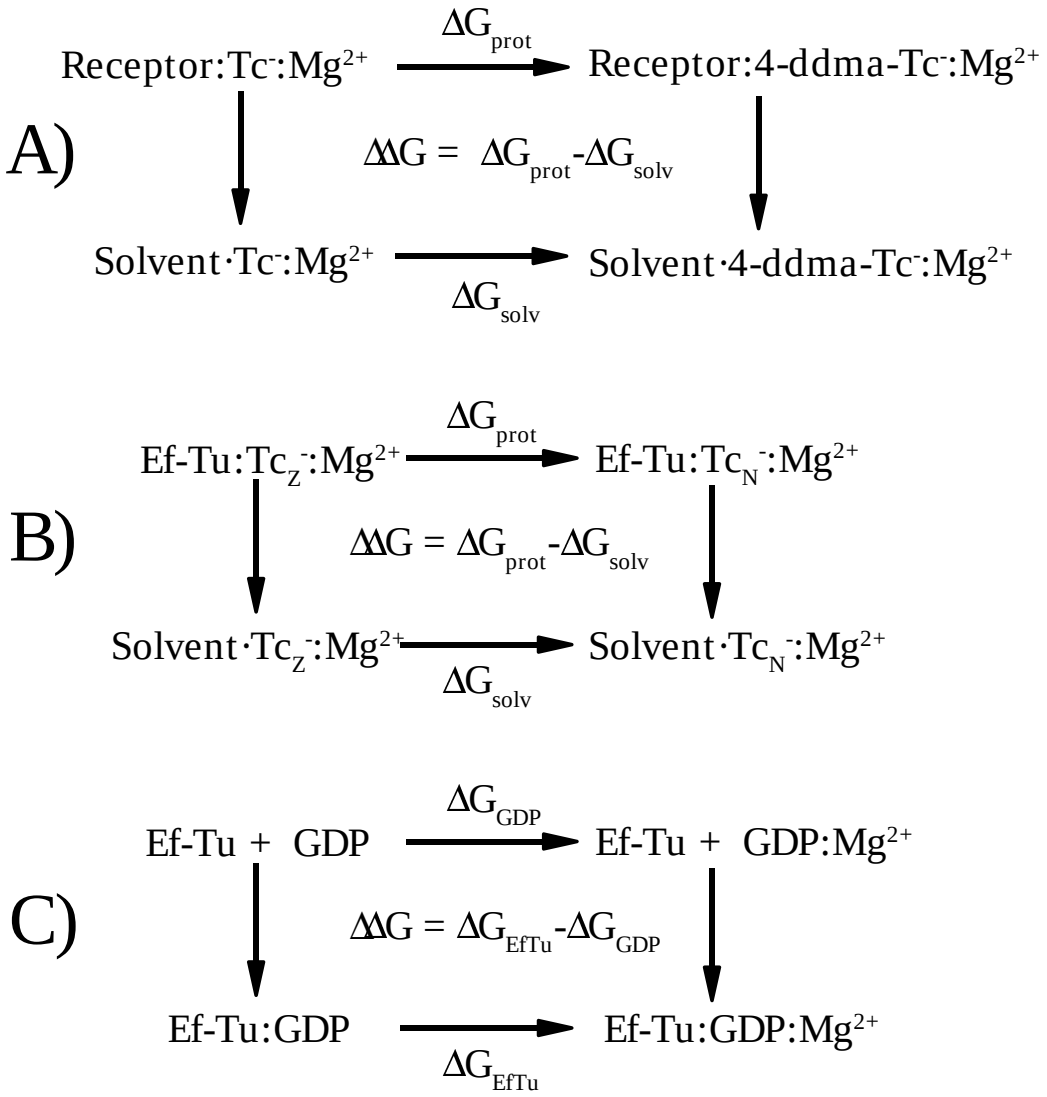


Figure 2: **Thermodynamic cycles for ligand:receptor binding. A)** Comparing Tc and 4-ddma-Tc. Vertical legs represent Tc^{-}:Mg^{2+} or 4-ddma-Tc^{-}:Mg^{2+} binding to a receptor (TetR, EF-Tu, or ribosome). Horizontal legs represent the alchemical transformation of Tc^{-} into 4-ddma-Tc^{-} in solution (below) or in the receptor:Tc^{-}:Mg^{2+} complex (above). **B)** Comparing Tc tautomeric states. **C)** GDP:EF-Tu binding with and without co-bound Mg.

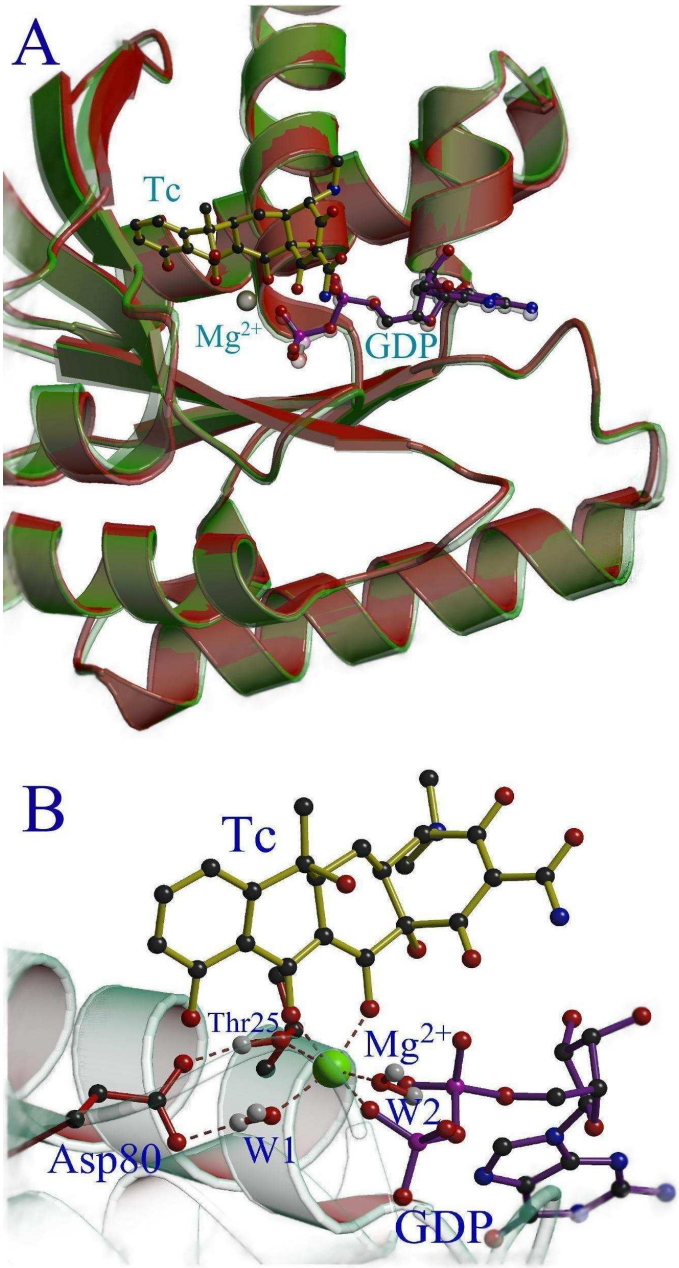


Figure 3: **(A):** superposition of the X-ray structures of EF-Tu with and without bound Tc, showing that there are no conformational changes upon Tc binding. **(B):** close-up view of the Tc binding site, showing that there are no hydrogen bonds between Tc and EF-Tu.

Supplementary Material: Binding of tetracyclines to elongation factor Tu, the Tet Repressor, and the ribosome: a molecular dynamics simulation study

Alexey Aleksandrov and Thomas Simonson
[1]Laboratoire de Biochimie (CNRS UMR7654), Department of Biology, Ecole Polytechnique, 91128 Palaiseau, France.

Full reference 32

Mackerell, A., Bashford, D., Bellott, M., Dunbrack, R., Evanseck, J., Field, M., Fischer, S., Gao, J., Guo, H., Ha, S., Joseph, D., Kuchnir, L., Kuczera, K., Lau, F., Mattos, C., Michnick, S., Ngo, T., Nguyen, D., Prodhom, B., Reiher, W., Roux, B., Smith, J., Stote, R., Straub, J., Watanabe, M., Wiorkiewicz-Kuczera, J., Yin, D., and Karplus, M. An all-atom empirical potential for molecular modelling and dynamics study of proteins. *J. Phys. Chem. B* 102 (1998), 3586–3616.

Free energy difference between Tc and 4-ddma-Tc binding to the ribosome and TetR: convergence and uncertainty

To illustrate the convergence in the MDFE calculations, we consider here the results for Tc and 4-ddma-Tc binding to the ribosome and the Tet Repressor (Table 1 and Figure 1). Four and six runs were performed for the ribosome and TetR, respectively (Table 1). Tc binding is favored over 4-ddma binding in both the ribosome and TetR. The behavior of the free energies as a function of the coupling parameter λ in typical MDFE runs is shown in Figure 1. To characterize the uncertainty in each ΔG value, the data for two runs, or four "half runs" are shown. A "half run" is obtained by splitting each trajectory segment (corresponding to a particular λ value) into two halves. The collection of half-segments defines a "half run". Each "half run" is plotted separately in Figure 1, giving four curves each for the ribosome and TetR. In Table 1, three error estimators are compared: (1) "Error" (column 5) is the difference between half runs, averaged over the unsigned forward and backward differences. For example, for the first TetR pair of runs, the unsigned differences between half runs are 0.4 and 0.7 kcal/mol, respectively, for the forward and backward half runs, for an average of 0.6

kcal/mol. (2) In column 6, we first average the $\Delta\Delta G$ values obtained from a forward and a backward run ("pairwise mean"). A standard error is computed from the three (TetR) or two (ribosome) resulting values. For example, for TetR, the three pairwise means are 3.6, 6.5, and 3.8 kcal/mol. The corresponding average is 4.0 kcal/mol; the standard error is 1.6 kcal/mol. (3) Finally, in column 7, we report one half of the unsigned difference between a forward and a backward run (the "hysteresis"). When the results are averaged (numbers in bold face), we see that the TetR error estimates range from 0.6 kcal/mol to 1.2 kcal/mol. For both TetR and the ribosome, the three "error" estimates are in rough agreement. In the body of the text, the first, somewhat more optimistic, "error" estimate is used.

Table 1: Alchemical MDFE runs transforming Tc into 4-ddma-Tc

	ΔG (kcal/mol)		pairwise mean		pairwise mean	$\frac{1}{2}\times$ unsigned
medium	forward runs[a]	backward runs[a]	ΔG[b]	"error"[c]	$\Delta\Delta G$[d]	hysteresis[e]
solvent	-28.5 (28.5/28.4)	-29.1 (29.1/29.2)	-28.8	0.1	0.0	0.3
TetR	-22.7 (22.5/22.9)	-27.6 (27.3/28.0)	-25.2	0.6	3.6	2.5
	-22.4 (23.0/21.8)	-22.1 (21.9/22.3)	-22.3	0.8	6.5	0.2
	-24.0 (24.4/23.7)	-26.0 (26.0/26.1)	-25.0	0.4	3.8	1.0
	TetR averages:		-24.1	**0.6**	4.6±**1.6**	**1.2**
Ribosome	-24.9 (25.9/23.9)	-25.8 (25.9/25.7)	-25.4	1.1	3.4	0.5
	-22.9 (22.2/23.6)	-25.8 (26.0/25.7)	-24.3	0.8	4.5	1.5
	Ribosome averages:		-24.8	**1.0**	4.0±**1.1**	**1.0**

All energies in kcal/mol. Forward runs transform Tc into 4-ddma-Tc; backward runs do the reverse. [a]Half runs in parentheses (unsigned values, for simplicity). [b]Average over the forward and backward runs. [c]We compute the unsigned difference between the two forward half runs, giving δ_f, and between the two backward half runs, giving δ_b; "error" is the average, $(\delta_f + \delta_b)/2$. [d]$\Delta\Delta G$ is the difference between ΔG in the current medium and in solvent; it is equal to the Tc/4-ddma-Tc binding free energy difference. [e]Hysteresis is the difference between a forward and a backward run.

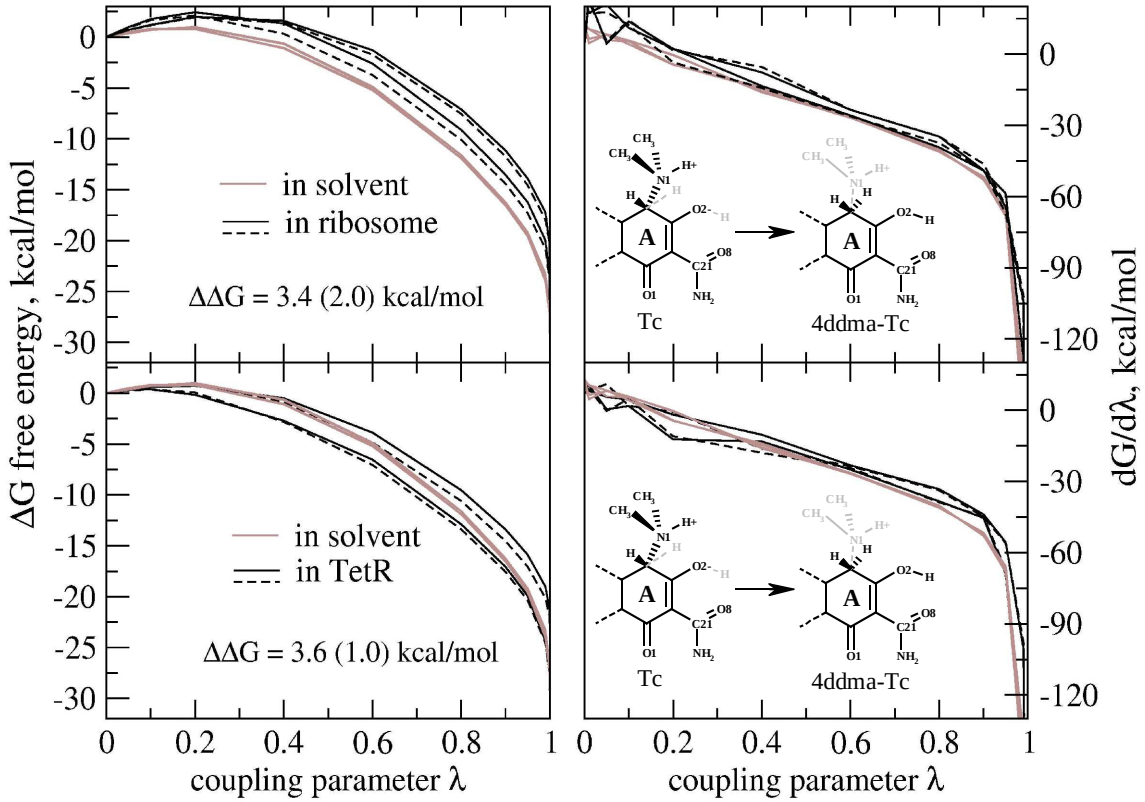


Figure 1: The free energy, ΔG, and its derivative, $dG/d\lambda$, as a function of the coupling parameter, λ, in solvent, the ribosome, and in the protein TetR. Two runs are shown in each case, corresponding to the "forward" and "backward" tranformation of the ligand. The "forward" transformation, converting plain Tc into 4-ddma Tc is shown as an inset, below. A so-called "dual topology" approach is used: atoms shown in light grey are present at each endpoint, but have a zero weight in the energy function, due to the dependency of the energy on λ. For the ribosome and TetR, each run is split here into two "half runs", by dividing each simulation window into two. Each "half run" is represented in the Figure. Solid black lines correspond to one forward and one backward "half run"; dashed black lines correspond to the other forward/backward pair of "half runs". With our procedure, the uncertainty in ΔG corresponds closely to the difference between two solid lines or two dashed lines, rather than to the difference between the solid and back lines.

Conclusions

Antibiotics are normally used for the treatment of deadly infectious diseases caused by pathogenic bacteria. Tetracycline (Tc), which was discovered in the late 40s, is one of the most important antibiotics in use today[12, 14]. Although the tetracyclines retain important roles in both human and veterinary medicine, the emergence of microbial resistance has limited their effectiveness[14]. This is also true for all currently used antibiotics and the drug pipeline is verging on empty[32]. On the other hand, virtually all new antibiotics are second generation drugs, that is, modifications of previously existing drugs[32]. This suggests that the only route to new antibiotics will be through more fundamental research on the basic biology of the bacteria.

It is thus of interest to increase our understanding of tetracycline and its biological partners, such as the ribosome and TetR. Extensive experimental research has established the main features. For example, crystal structures of TetR:Tc and TetR:DNA complexes are known[16, 19], and the structure of tetracycline in complex with the 30S ribosome was recently determined[9, 20].

However, the experimental approaches have limitations. For example crystal structures of Tc bound to the Ribosome do not reveal the number of Tc binding sites, their relative strength, probable allosteric coupling, and do not show which site(s) is functional, nor give thermodynamic information such as individual residue free energy contributions. Crystal structures of Tet repressor do not assess the magnitude, timescale, order/coupling of different events during induction of TetR nor even answer directly the question why TetR:Tc unbinds from DNA. A complementary approach is to develop computer simulation models, which can be used to investigate the structure, dynamics, and thermodynamics of Tc:protein or Tc:RNA complexes. Despite its importance, TetR has rarely been subjected to computer modeling, partly because of the need to first develop a molecular mechanics model

for Tc.

I developed such a model for the twelve most important tetracyclines and for their most important protonation and complexation states at neutral pH[3, 5]. The model was first tested before its application to real biological problems. The force field model shows good agreement with large body of both ab initio and experimental data, showing that it is suitable for investigating Tc:protein and Tc:RNA complexes.

I used this model to examine several aspects of the mechanism of Tc action and bacterial resistance. Recently, the crystal structure of a complex between elongation factor Tu (Ef-Tu) and Tc was solved, raising the question whether Tc's binding to Ef-Tu has a role in its inhibition of protein synthesis. My free energy simulations strongly support the idea that the ribosome, not Ef-Tu, is the primary target of tetracycline.

Crystal structures of Tc bound to the 30S subunit were also determined recently. I compared the binding free energies of Tc to the primary site and a secondary site[4]; these free energy calculations provided evidence for the existence of a single, predominant binding site for Tc on the ribosome.

Finally, I studied Tc interactions with TetR[1]. Crystal structures of TetR do not reveal the protonation states of Tc or the protein sidechains, nor the orientation of several key recognition residues. I focused on the extent of protein preorganization and the possible role of induced fit in the protein:ligand binding reaction. I found that the protonation state and the conformation of both the ligand and the binding pocket do not change significantly upon binding: the system functions as an electrostatic lock and key near important recognition sidechains.

The "static" X-Ray structures also do not reveal the transition pathway of induction nor answer directly the question why TetR:Tc unbinds from DNA. I used molecular dynamics simulations and continuum electrostatic calculations to study the allosteric mechanism and simulations of the TetR:DNA complex, the Tc-bound,

"induced" TetR, and the transition pathway between them[2]. The simulations are consistent with a broad body of experimental data. They generally support the model of induction inferred previously from the crystal structures and reveal new important details, some of which may be relevant for other regulatory proteins.

My work has thus filled important gaps in knowledge which are experimentally difficult to access. Only recently such work became possible, mainly because of several reasons: advances in computer hardware, maturity of computation methods[21, 24–27], recent success of computational biophysics[28] and availability of new crystal structures[9, 19, 20].

In the future, the tetracycline forcefield can be used to develop forcefield models for other tetracycline analogs. In this work I considered the most important twelve Tc variants. In the future this list can be easily expanded for new potential tetracycline–like drugs.

For example glycotetracyclines, which belong to a new class of antibiotics, induce Tet Repressor (TetA is produced), but are not exported by TetA. New drugs which do not interact with TetR are also a good alternative, so that information about tetracycline binding to the Tet repressor, will be important in the rational–drug design of new tetracyclines.

Nowadays the TetR/Tc system is of general interest because of widespread application in molecular and cell biology as a sensitive switch for target gene regulation. Experimental studies on TetR:Tc are continuing, this includes recent discoveries how a single mutation allows reversal of the allosteric response of Tet repressor and how a small peptide efficiently induces the tetracycline repressor TetR. This suggests that the use of TetR:Tc system will only increase in the future.

Understanding the mechanism of induction of TetR is thus important for the future rational design of new Tet Repressors with new properties for gene regulation. For example, the mechanism of induction of reversal TetR is not known so far, though

two models were proposed recently (one based on molecular dynamics simulations and the other one was based on the recent crystal structure of the reverse TetR). Here a new computational study could very helpful and would possibly involve mutations of TetR in MD simulations (for example Leu17Gly) and calculation of the binding free energy of TetR to the DNA, similarly to the Tc insertion into the TetR:DNA complex in our published study. More generally, understanding TetR can be helpful in understanding the machinery of other repressors in the cell (not only in prokaryotes)

Considering how much experimental work was done, it is surprising how little we know about Tc function. Obviously studies on Tc and TetR function will continue. Here a joined experimental and computation approach likely would play a bigger role than today. In summary, I hope my work will be useful and serve as a seed for future studies and new interesting findings.

Bibliography

[1] A. Aleksandrov, J. Proft, W. Hinrichs, and T. Simonson. Protonation patterns in tetracycline:Tet repressor recognition: Simulations and experiments. *ChemBioChem*, 8:675–685, 2007.

[2] A. Aleksandrov, L. Schuldt, W. Hinrichs, and T. Simonson. Tet repressor induction by tetracycline: a molecular dynamics, continuum electrostatics, and crystallographic study. *JMB*, 378:896–910, 2008.

[3] A. Aleksandrov and T. Simonson. The tetracycline:Mg^{2+} complex: a molecular mechanics force field. *J. Comp. Chem.*, 13:1517–1533, 2006.

[4] A. Aleksandrov and T. Simonson. Molecular dynamics simulations of the 30S ribosomal subunit reveal a preferred tetracycline binding site. *J. Am. Chem. Soc.*, 130:1114–1115, 2008.

[5] A. Aleksandrov and T. Simonson. Molecular mechanics models for tetracycline analogs. *J. Comp. Chem.*, "in press", 2008.

[6] F.H. Allen and O. Kennard. *Chem. Des. Autom. News*, 8:1, 1993.

[7] F.H. Allen and O. Kennard. 3D search and research using the Cambridge Cristallogaphic Database. *Chem. Des. Autom. News*, 8:31–37, 1993.

[8] G. Bauer, C. Berens, SJ. Projan, and W. Hillen. Comparison of tetracycline and tigecycline binding to ribosomes mapped by dimethylsulphate and drug-

directed fe2+ cleavage of 16s rrna. *J Antimicrob Chemother.*, 53(4):592–599, 2004.

[9] D. E. Brodersen, W. M. Clemons Jr, A. P. Carter, R. J. Morgan-Warren, B. T. Wimberly, and V. Ramakrishnan. The structural basis for the action of the antibiotics tetracycline, pactamycin, and hygromycin B on the 30S ribosomal subunit. *Cell*, 103:1143–1154, 2000.

[10] V. Burdett. Tet(m)-promoted release of tetracycline from ribosomes is gtp dependent. *J. Bacteriol.*, 178:3246–3251, 1996.

[11] I. Chopra, P. M. Hawkey, and M. Hinton. Tetracyclines, molecular and clinical aspects. *Antimicrob. Agents Chemother.*, 29(3):245–277, 1992.

[12] I. Chopra and M. Roberts. Tetracycline antibiotics: Mode of action, applications, molecular biology, and epidemiology of bacterial resistance. *Microbiology and Molecular Biology Reviews*, 65:232–260, 2001.

[13] SR. Connell, DM. Tracz, KH. Nierhaus, and DE. Taylor. Ribosomal protection proteins and their mechanism of tetracycline resistance. *Antimicrob. Agents Chemother.*, 47(12):3675–3681, 2003.

[14] R. A. Greenwald, W. Hillen, and M. L. Nelson. *Tetracyclines in Biology, Chemistry and Medicine*. Birkhäuser Verlag/Switzerland, 2001.

[15] Susan E. Heffron, Suet Mui, Annette Aorora, Kenton Abel, Ernst Bergmann, and Frances Jurnak. Molecular complementarity between tetracycline and the GTPase active site of elongation factor Tu. *Acta Cryst. D*, 62:1392–1400, 2006.

[16] C. Kisker, W. Hinrichs, K. Tovar, W. Hillen, and W. Saenger. The complex formed between tetracycline repressor and tetracycline-Mg^{2+} reveals mechanism of antibiotic resistance. *J. Mol. Biol.*, 247:260–280, 1995.

[17] S. Mui, K. Delaria, and F. Jurnak. Preliminary crystallographic analysis of a complex between tetracycline and the trypsin-modified form of escherichia coli elongation factor tu. *JMB*, 212(3):445–447, 1990.

[18] R. Oehler, N. Polacek, G. Steiner, and Barta A. Interaction of tetracycline with rna: photoincorporation into ribosomal rna of escherichia coli. *NAR*, 25(6):1219–1224, 1997.

[19] P. Orth, D. Schnappinger, W. Hillen, W. Saenger, and W. Hinrichs. Structural basis of gene regulation by the tetracycline inducible Tet repressor-operator system. *Nat. Struct. Biol.*, 7:215–219, 2000.

[20] M. Pioletti, F. Schlnzen, J. Harms, R. Zarivach, M. Gluhmann, H. Avila, A. Bashan, H. Bartels, T. Auerbach, C. Jacobi, T. Hartsch, A. Yonath, and F. Franceschi. Crystal structures of complexes of the small ribosomal subunit with tetracycline, edeine and if3. *EMBO*, 20(8):1829–1839, 2001.

[21] B. Roux and T. Simonson. Implicit solvent models. *Biophys. Chem.*, 78:1–20, 1999.

[22] W. Saenger, P. Orth, C. Kisker, W. Hillen, and W. Hinrichs. The tetracycline repressor: A paradigm for a biological switch. *Angew. Chem. Int. Ed.*, 39:2042–2052, 2000.

[23] D. Schnappinger and W. Hillen. Tetracyclines: antibiotic action, uptake, and resistance mechanisms. *Arch Microbiol.*, 165(6):359–369, 1996.

[24] T. Simonson. Free energy calculations. In O. Becker, A.D. Mackerell Jr., B. Roux, and M. Watanabe, editors, *Computational Biochemistry & Biophysics*, chapter 9. Marcel Dekker, N.Y., 2001.

[25] Thomas Simonson. Dielectric relaxation in proteins: macroscopic and microscopic models. *Int. J. Quant. Chem.*, 73:45–57, 1999.

[26] Thomas Simonson. Electrostatic free energy calculations for macromolecules: a hybrid molecular dynamics/continuum electrostatics approach. *J. Phys. Chem. B*, 104:6509–6513, 2000.

[27] Thomas Simonson. Macromolecular electrostatics: continuum models and their growing pains. *Curr. Opin. Struct. Biol.*, 11:243–252, 2001.

[28] Thomas Simonson. Electrostatics and dynamics of proteins. *Rep. Prog. Phys.*, 66:737–787, 2003.

[29] CM. Spahn, G. Blaha, RK. Agrawal, P. Penczek, RA. Grassucci, CA. Trieber, SR. Connell, DE. Taylor, KH. Nierhaus, and J. Frank. Localization of the ribosomal protection protein tet(o) on the ribosome and the mechanism of tetracycline resistance. *Molecular Cell*, 5(7):1037–1045, 2001.

[30] C.R. Stephens, K. Murai, K.J. Brunings, and R.B. Woodward. Acidity constants of the tetracycline antibiotics. *J. Am. Chem. Soc.*, 78:4155–4158, 1956.

[31] P. Sum. Case studies in current drug development: glycylcyclines. *Curr. Opin. Chem. Biol.*, 10:374–379, 2006.

[32] G. Taubes. The bacteria fight back. *Science*, 321(5887):355–361, 2008.

[33] A. Weixlbaumer, S. Petry, CM. Dunham, M. Selmer, AC. Kelley, and V. Ramakrishnan. Crystal structure of the ribosome recycling factor bound to the ribosome. *NSB*, 14(8):733–737, 2007.

Appendix

Appendix: Computational sidechain placement and protein mutagenesis with implicit solvent models. Anne Lopes, Alexey Alexandrov, Christine Bathelt, Georgios Archontis, and Thomas Simonson *Proteins,* 2007, 67, 853-867.

PROTEINS: Structure, Function, and Bioinformatics 67:853–867 (2007)

Computational Sidechain Placement and Protein Mutagenesis With Implicit Solvent Models

Anne Lopes,[1†] Alexey Alexandrov,[1†] Christine Bathelt,[1] Georgios Archontis,[2*] and Thomas Simonson[1*]
[1]*Laboratoire de Biochimie (UMR CNRS 7654), Department of Biology, Ecole Polytechnique, 91128, Palaiseau, France*
[2]*Department of Physics, University of Cyprus, Nicosia, Cyprus*

ABSTRACT **Structure prediction and computational protein design should benefit from accurate solvent models. We have applied implicit solvent models to two problems that are central to this area. First, we performed sidechain placement for 29 proteins, using a solvent model that combines a screened Coulomb term with an Accessible Surface Area term (CASA model). With optimized parameters, the prediction quality is comparable with earlier work that omitted electrostatics and solvation altogether. Second, we computed the stability changes associated with point mutations involving ionized sidechains. For over 1000 mutations, including many fully or partly buried positions, we compared CASA and two generalized Born models (GB) with a more accurate model, which solves the Poisson equation of continuum electrostatics numerically. CASA predicts the correct sign and order of magnitude of the stability change for 81% of the mutations, compared to 97% with the best GB. We also considered 140 mutations for which experimental data are available. Comparing to experiment requires additional assumptions about the unfolded protein structure, protein relaxation in response to the mutations, and contributions from the hydrophobic effect. With a simple, commonly-used unfolded state model, the mean unsigned error is 2.1 kcal/mol with both CASA and the best GB. Overall, the electrostatic model is not important for sidechain placement; CASA and GB are equivalent for surface mutations, while GB is far superior for fully or partly buried positions. Thus, for problems like protein design that involve all these aspects, the most recent GB models represent an important step forward. Along with the recent discovery of efficient, pairwise implementations of GB, this will open new possibilities for the computational engineering of proteins. Proteins 2007;67:853–867.** **© 2007 Wiley-Liss, Inc.**

Key words: structure prediction; solvation; mean field; Generalized Born; Poisson equation; protein design

INTRODUCTION

Homology-based structure prediction and computational protein design are areas whose importance is increasing with the development of structural genomics.[1–4] Both techniques usually rely on a simplified description of protein conformational space, taking into account one or a few fixed backbone conformations and a discrete set of sidechain rotamers.[5–8] The most stable structure within this discrete space can be found by exact or approximate search methods.[6,9–15] This paper focusses on another important ingredient: the energy or scoring function and, in particular, the treatment of aqueous solvent. We consider the performance of several implicit solvent models for two key problems that occur in computational protein design: sidechain placement and the calculation of stability changes due to point mutations. Both problems have been extensively studied, using a range of models.[13–20] However, they must be considered together if one is to parameterize and test solvent models for protein design; i.e., for searching sequence and conformational space simultaneously. For example, previous solutions of the sidechain placement problem that omit electrostatics[14,16,21] are not acceptable in this context. Such combined analyses are much less common.[15,19,20] Furthermore, with the rapid progress of implicit solvent models, especially generalized Born models, it is important to reconsider this problem.[22,23]

Indeed, aqueous solvent plays an important role in the structure and stability of proteins.[24] Structure prediction and protein design are done almost exclusively with "implicit" solvent models, for efficiency. The solvent degrees of freedom are not explicitly represented; rather, they are taken into account through their effect on the intraprotein interactions.[25] The energy function for the protein is referred to as a Potential of Mean Force (PMF), or "effective" energy function. To obtain correct energetics for the protein, the PMF for any given protein conformation should coincide with the Boltzmann

The Supplementary Material referred to in this article can be found at http://www.interscience.wiley.com/fpages/0887-3585/suppmat/

[†]These authors contributed equally.

*Correspondence to: Thomas Simonson, Laboratoire de Biochimie (UMR CNRS 7654), Department of Biology, Ecole Polytechnique, 91128, Palaiseau, France. E-mail: thomas.simonson@polytechnique.fr or Georgios Archontis, Department of Physics, University of Cyprus, Nicosia, Cyprus. E-mail: archonti@ucy.ac.cy

Received 4 August 2006; Revised 26 October 2006; Accepted 16 December 2006

Published online 8 March 2007 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/prot.21379

© 2007 WILEY-LISS, INC.

average of the energy over the solvent configurations.[25] In practice, only approximate PMFs can be constructed.

An implicit solvent model that has a clear physical basis is the Poisson model, which treats the solvent as a dielectric continuum,[26–29] by numerically solving the Poisson equation (PE). The essential physical ingredients are (1) the strong, attractive interactions between charged protein groups and the surrounding, high-dielectric solvent, and (2) the large shielding of protein–protein electrostatic interactions by solvent. The solvent contribution to the PMF is obtained as the electrostatic free energy of a collection of point charges in a dielectric cavity.[25,29] The PE model provides good accuracy for many applications,[30] including small molecule solvation,[31,32] acid/base equilibria,[28,33] ligand binding,[34] protein–protein binding,[35,36] and protein dynamics.[37] Unfortunately, PE methods cannot be used routinely for computational protein design. Indeed, in continuum electrostatics, the effective interaction between two protein residues depends on the entire protein's shape and the complementary volume occupied by high-dielectric solvent. Therefore, continuum electrostatic energies are many-body quantities that cannot ordinarily be expressed as a sum over residue or atom pairs.[19,20,23,29,38] This is a prohibitive limitation for protein design.

A more efficient alternative is the generalized Born (GB) model.[22,30,39,40] GB is based on the same physical picture as PE, with a dielectric continuum solvent surrounding a protein cavity. But it makes additional approximations that allow an analytical expression of the PMF. It has become feasible to use GB in a protein design context, because residue–pairwise implementations were recently discovered.[19,23] Several GB variants and parameterizations exist.[41–46] GB has been used for many applications, including small molecule solvation,[42–44,47] protein solvation,[45,48] acid/base equilibria,[49,50] protein dynamics,[51,52] ligand binding,[53] protein folding,[54,55] loop structure prediction,[56] and scoring native folds vs. decoys.[57] The best variants have an accuracy that is not much inferior to explicit solvent models,[22,46,49] and further improvements can be expected.

A third, even simpler class of implicit solvent models are the so-called Accessible Surface Area models.[58–60] These models characterize different atom types by "atomic solvation parameters," which reflect their hydrophobicity or hydrophilicity, and consider the fraction of each atom's surface area that is accessible to solvent. Each atom contributes to the PMF through the product of its solvation parameter and its solvent accessible surface area. Usually, this accessible surface area contribution is supplemented by another electrostatic term, which attempts to capture the shielding of protein–protein electrostatic interactions by the high-dielectric solvent. The simplest approach is to add to the PMF a screened Coulomb energy, so that protein–protein electrostatic interactions are reduced by a constant factor ε. We refer to this as the Coulomb/Accessible Surface Area (CASA) model. This class of models has been used for protein molecular dynamics,[60–62] structure prediction,[14] and protein–ligand binding.[63] It is routinely used for computational protein design.[4,64–67]

With the ongoing development of GB models, and with the discovery of methods to implement them efficiently in computational protein design,[19,23] it is important to systematically compare the behavior of the PE, GB, and CASA models for protein structure prediction and design. In the context of protein design, structure prediction is usually limited to sidechain placement with a fixed protein backbone. Many authors have analyzed this problem and shown that good results are obtained with very simple models, without any electrostatic or solvent treatment.[14,16,21] In computational protein design, however, sidechain placement must be done repeatedly, following rounds of random mutagenesis. The mutations will often modify the protein charge, frequently introducing charged sidechains into buried positions. For these effects, an accurate electrostatic treatment is desirable. Therefore, an integrated treatment of sidechain placement and ionized mutations is needed. Previous solutions of the placement problem that omit electrostatics are not acceptable. Many authors have studied the accuracy of PE and GB models for protein electrostatics.[18,29,30,68] However, very few studies have considered sidechain placement and ionized mutations simultaneously. Even fewer (none that we are aware of) have compared surface area models (CASA) to GB in this context. While CASA is the most common solvent model in computational protein design, GB models have greatly progressed in the last few years, so that a reevaluation and comparison of models is needed. For example, one recent study considered a suboptimal GB/ACE parameterization.[20]

In this paper, we consider both sidechain placement and the effect of amino acid mutations on protein stability. For sidechain placement, we use the CASA model. In a related test, we use CASA, GB, and PE to estimate the stability of large libraries of protein conformations, where the sidechain positions have been randomized. We then consider over 1000 mutations involving ionized sidechains, and use CASA, GB, and PE to compute the corresponding stability changes. The data set includes 140 mutations in 12 proteins for which experimental measurements are available. The experimental data span only a limited free energy range, and they do not include any mutations that introduce or delete ionized sidechains in the protein core, since these are usually not experimentally tractable.[69] The other mutations are mostly in buried positions and correspond to much larger stability changes, but no experimental measurements are available. For computational reasons, most of these mutations are not "biochemically exact": they consist in charge modifications, as opposed to real mutations. For example, negative charges are introduced onto valine sidechain methyls, roughly mimicking a valine → aspartate mutation. Nevertheless, they contain the relevant physical effects and represent a valid test of the solvent models. We refer to these two data sets as the "experimental" and "artificial" mutations, respectively.

An important aspect of this work is the parameterization of the CASA and GB models. With CASA, for example, we consider three sets of atomic surface parameters, a wide range of dielectric constants, and a range of scaling factors for various model terms. We consider two GB models: GB/ACE,[42] in combination with the Charmm19 force field,[70] and GB/HCT,[71] in combination with the AMBER force field.[72] A limited parameter optimization is done for GB/ACE and a more extensive one for GB/HCT. An essential point is that the parameterization must be useful for both structure prediction and stability changes. The goal of the GB/HCT parameter optimization is to take into account very recent improvements in the set of atomic radii used for continuum electrostatics calculations with the AMBER atomic charges.[73]

The "experimental" mutations are not used in the parameterization, for several reasons. First, they correspond to very small stability changes. More importantly, they cannot be computed without assuming a specific model for the unfolded protein's structure. Thus, for a rigorous optimization, the unfolded state model and the energy parameters would have to be tested and varied simultaneously. We prefer to optimize the energy function separately; therefore, we use only the artifical mutations for parameter optimization. Since experimental measurements are not available, we take as reference values the PE results. Indeed, PE has been extensively used to study protein electrostatics.[29,68] Many recent studies have shown that it is a valid reference for the parameterization and testing of implicit solvent models.[17–19] With this strategy, the unfolded state treatment does not influence the parameterization, since the same treatment is used for CASA, GB, and PE.

Once the energy parameters are chosen, we can use the experimental mutations for a blind test of the CASA and GB models. We use a very simple, tripeptide representation of the unfolded state, with noninteracting amino acids, which is commonly used in protein design. With CASA, the mean unsigned deviation from experiment is 2.1 kcal/mol. This appears comparable to the accuracy reported by Serrano et al.[74] for a comparable set of mutations, although they did not describe their subset of charged mutations separately. Remarkably, it is similar to the accuracy of molecular dynamics free energy simulations using explicit solvent models,[49,75] which are much more difficult and expensive (but which give additional structural and dynamical information). With GB/ACE and GB/HCT, the mean unsigned deviations from experiment are 2.9 and 3.4 kcal/mol, respectively. If a surface term is included in the GB/HCT model, to help represent dispersion and hydrophobic contributions,[32] the GB/HCT mean unsigned error drops to 2.1 kcal/mol, the same level as CASA.

In summary, we find that (a) the choice of electrostatic model is not very important for the sidechain placement calculations (confirming several earlier studies[14,16,21]); (b) GB/HCT and CASA give the same accuracy for the experimental mutations; (c) GB/HCT yields an enormous improvement for the total solvation free energies and for the artificial mutations. Thus, for problems like protein design that involve all these aspects, the most recent GB models represent an important step forward. Along with the recent discovery of efficient implementations of GB in protein design,[19,23] this will open new possibilities for the computational engineering of new proteins.

MATERIALS AND METHODS

Effective Energy Function

We tested several effective energy functions, or PMFs, which have the form:

$$E = E_{\text{bonds}} + E_{\text{angles}} + E_{\text{dihe}} + E_{\text{impr}} + E_{\text{vdw}} + E_{\text{coul}} + E_{\text{solv}}. \quad (1)$$

The first six terms in Eq. (1) represent the protein internal energy. They are taken from either the CHARMM19 or the AMBER empirical energy function.[70,72] They represent a covalent bond energy, a bond angle energy, a torsion energy associated with sidechain dihedrals, a term maintaining the chirality or planarity of selected atomic centers, a van der Waals energy, and a Coulomb electrostatic energy,

$$E_{\text{coul}} = \sum_{i<j} \frac{q_i q_j}{r_{ij}}, \quad (2)$$

where the sum is over all pairs of protein atoms i, j, of charges q_i, q_j, and r_{ij} is the distance between a pair. The last term on the right of Eq. (1), E_{solv}, represents the contribution of solvent.[25] In this work, we compare three different solvent treatments, described below.

CASA Solvent Treatment

Our first solvent treatment is an accessible surface area treatment: the CASA model. E_{solv} includes two energy terms that describe protein–solvent electrostatic and hydrophobic interactions:[25]

$$E_{\text{solv}} \equiv E_{\text{CASA}} = E_{\text{screen}} + E_{\text{surf}} = \left(\frac{1}{\varepsilon} - 1\right) E_{\text{coul}} + \alpha \sum_i A_i \sigma_i. \quad (3)$$

E_{screen} is a screened Coulomb energy; ε is the dielectric constant of the medium (relative to vacuum). Notice that $E_{\text{coul}} + E_{\text{screen}} = E_{\text{coul}}/\varepsilon$. E_{surf} is related to the atomic solvent-accessible surface areas. The sum is over all protein atoms i; A_i is the solvent-accessible area of atom i; σ_i is an atomic solvation parameter (measured in kcal/mol/Å^2); and α is an overall weight for the surface energy term. The coefficients σ_i reflect the preference of particular atom types to be exposed or hidden from solvent, and incorporate both electrostatic and nonelectrostatic effects.[25] Surface areas were computed by the Lee and Richards algorithm,[76] implemented in XPLOR,[77] using a

1.4 Å probe radius. Three different sets of atomic solvation parameters were tested,[59–61] along with different values of the dielectric constant ε and the weight α.

GB Solvent Treatment

Our second solvent treatment is a GB model[39,41,42]:

$$E_{\text{solv}} \equiv E_{\text{GB}} = \sum_i \Delta G_i^{\text{self}} + \sum_{i<j} \Delta G_{ij}^{\text{screen}} = \tau \sum_i \frac{q_i^2}{2b_i} + \tau \sum_{i<j} \frac{q_i q_j}{\sqrt{r_{ij}^2 + b_i b_j exp[-r_{ij}^2/(4b_i b_j)]}} \quad (4)$$

where $\tau = 1/\varepsilon_w - 1/\varepsilon_p$, r_{ij} is the distance between charges q_i and q_j, b_i is the effective Born radius of atom i, ε_w is the dielectric constant of water (set to 80), and ε_p is the protein dielectric constant (set to 1 unless otherwise mentioned). The first term is a sum of atomic self-energies $\Delta G_i^{\text{self}} = \tau {q_i}^2/2b_i$, corresponding to the interaction of each atomic charge q_i with its own reaction field in the environment of the solvated biomolecule. The second term models the interaction of a charge q_i with the reaction field produced by a different charge q_j, and accounts for the screening of electrostatic interactions by the high-dielectric solvent.[39]

The self-energy term in Eq. (4) requires the calculation of an electrostatic energy density integral over the solute volume. We use both the GB/ACE and the GB/HCT models, which assume a Coulomb functional form for the electric field inside the solute and partition the volume into atomic contributions.[42,71] The GB/HCT parameters are described further on. The GB/ACE atomic volumes corresponded to the Voronoi database V01,[78] scaled by a factor of 0.8 as described in Ref. 52. The smoothing parameter that controls the width of atomic volumes[42] in GB/ACE was set to 1.3. The partial charges corresponded to the CHARMM19 energy function.[70] The calculations were performed with the XPLOR program.[77,79]

Poisson Equation Solvent Treatment

Our third solvent treatment is a continuum electrostatic model with numerical solution of the Poisson equation. We refer to it as the PE model. The solvation energy has the form

$$E_{\text{solv}} \equiv E_{\text{PE}} = \frac{1}{2} \sum_i q_i V_i^{\text{reac}} \quad (5)$$

where q_i is the charge of atom i and V_i^{reac} is the electrostatic potential on atom i due to the polarization charge at the protein–solvent interface, induced by every atomic charge. For a given protein structure, we computed V_i^{reac} by solving the Poisson equation numerically for the protein in solution and in the gas phase and taking the difference. The finite-difference program UHBD was used.[80] The protein–solvent dielectric boundary was defined by the molecular surface of the protein. The solution used a two-step focussing procedure and a cubic grid with spacings of 0.8 and 0.4 Å. The molecular surface was constructed with 2000 points per atom, using a probe sphere of radius 2 Å and the boundary smoothing method in UHBD.[80] Small voids in the interior of the rotameric structures were filled by dummy atoms, to prevent the occurence of artificial, high-dielectric internal cavities (sometimes overlooked in PE applications). Two PE parameterizations are used. The first uses Charmm19 atomic radii and charges,[70] except for hydrogen radii, which were set to 1 Å. The second uses atomic charges from the AMBER, all-atom force field,[72] along with atomic radii specifically and carefully optimized for PE with AMBER charges.[73]

Protein Set and Rotamer Library

Twenty-nine proteins were used for sidechain placement calculations; they are listed in Table III. Their sizes varied from 36 to 212 amino acids, with a mean of 110. Their structures were all determined using X-ray crystallography with a resolution of 1.8 Å or better. The rotamer construction and energy calculations were performed with the XPLOR program.[77] The backbone and C_β atoms of the X-ray structure were fixed during the construction. The sidechain atoms were geometrically constructed from the position of the N, C_α, C, and C_β atoms, using standard bond lengths and bond angles from the CHARMM19 force field. Sidechain dihedral angles were taken from the Tuffery rotamer library.[6]

Additional, solvation energy calculations were performed for the proteins BPTI (4PTI), lysozyme (1LZ1), thioredoxin (2TRX), and ubiquitin (1UBQ). We created a set of 750–1000 random structures for each one by holding the backbone in its native conformation and randomizing the orientation of all sidechains except Pro, Ala, and Cys.

Mean Field Optimization

We did sidechain placement using a mean field approximation.[14] This method calculates iteratively the Boltzmann probability $P(i,k)$ of each rotamer k of each residue i, which is related to the mean energy $E(i,k)$ of sidechain i:

$$E(i,k) = -RT \ln P(i,k). \quad (6)$$

$E(i,k)$ is the Boltzmann average of the interaction energy between sidechain i and its environment; R is the ideal gas constant, and T is the temperature. Since the protein backbone is fixed, we can write

$$E(i,k) = E_{BB}(i,k) + \sum_{j \neq i} \sum_l E(ik, jl) P(j,l) \quad (7)$$

where E_{BB} is the interaction energy with the backbone, the first sum is over protein sidechains j, the second sum is over the rotamers l of sidechain j, and $E(ik,jl)$ is

the interaction energy between sidechains i and j when they occupy rotamers k and l. We assume the optimal sidechain positions correspond to the most probable rotamers.

The probabilities were computed iteratively. The current estimate $P(i,k)^{(n)}$ was updated according to

$$P(i,k)^{(n+1)} = \lambda P(i,k)^{(n)} + (1-\lambda)P(i,k)^{(n-1)}. \quad (8)$$

Different λ values were tested. The best results (reported below) were obtained with $\lambda = 0.35$ and uniform starting rotamer probabilities. About 20 cycles were typically needed for convergence.

An Approximate, Pairwise Surface Area Calculation

In the sidechain placement, the solvent accessible surfaces are calculated by an approximate, but very efficient procedure. The buried surface of a sidechain is computed by summing over the neighboring sidechain and backbone groups. For each neighboring group, the contact area with the sidechain of interest is computed, independently of other surrounding groups. The contact areas are then summed. This approach assumes the contact areas between sidechains are independent, and the total area of a sidechain is a pairwise sum over its neighbors. In fact, surface area buried by one sidechain may also be buried by another. Mayo et al.[81] showed that this approach overestimated the surface areas of buried sidechains, but had little effect on solvent exposed sidechains. Therefore, we performed most of the calculations with a scaling factor applied to the contact areas involving buried sidechains. To determine the optimal scaling factor, following Mayo et al.,[81] we considered the total surface area that is buried within the protein, A_{buried}. A_{buried} is defined as the difference between the total area of all residues, taken separately, and the total surface area of the protein structure; see Ref. 81 for details. With the pairwise approximation, it is computed as

$$A_{\text{buried}} \approx A^{\text{pairwise}}_{\text{buried}} = \sum_i A^0_{i_r t3} - \left(\sum_i A_{i_r t} + \frac{1}{2} \sum_{ij} s_i (A_{i_r j_s t} - A_{i_r t} - A_{j_s t}) \right). \quad (9)$$

The first sum on the right is over all residues i and represents the total area of all residues, taken separately. $A^0_{i_r t3}$ is the exposed area of sidechain i when it occupies the rotamer r, in the presence of just the backbone of residues $i-1$, i, $i+1$ (a tripeptide indicated by the subscript $t3$). The sums in parentheses represent the approximate total protein surface area. $A_{i_r t}$ is the exposed area of the sidechain and its backbone, in the presence of the entire protein backbone. $A_{i_r j_s t}$ is the exposed area of the sidechain pair ij in the presence of the entire protein backbone. Finally, s_i is the scaling parameter, which is set to 1 if residue i is exposed to solvent and to a smaller value $s < 1$ if it is buried. To validate this approximation and determine the optimal s, we performed surface area calculations on the 29 test proteins above, and compared the exact and approximate buried surface areas. Using a scaling factor of $s_i = 0.5$ for buried amino acids [Eq. (9)], we observed a correlation of 0.999 between the approximate and the true surfaces, with a slope of 0.993 and an offset of 73.3 Å^2. The RMSD between the approximate and true surfaces was just 199 Å^2, 1% of the mean surface. Therefore, a value of $s_i = 0.5$ was used for all the sidechain placements, unless otherwise mentioned.

Validation Methods

Two criteria were used to assess the accuracy of the sidechain placements. First, we calculated the percentage of sidechain dihedral angles within 40° of the value in the crystal structure. Second, we computed the RMSD between the sidechain atom positions in the model and the X-ray structure, excluding the C_β. The RMSD calculation took into account the rotational symmetry axis of Asp, Phe, Glu, and Tyr residues.

Later in this article, we compare protocols that use several different effective energy functions, corresponding to different solvent treatments. To determine the significance of the differences between protocols, statistical tests were done. Analysis of variance (ANOVA) was performed to evaluate groups of protocols. The null hypothesis is that for an observable of interest (e.g., RMSD), all the means observed for the various protocols are identical. Rejection of the null hypothesis means that there is at least one pair of means that are different from each other. To identify this pair, we then performed Student's t-tests on all pairs of means.

Charge Mutations

To test the effective energy functions in a situation that mimics protein design calculations, we performed two sets of charge mutations. In the first set, the atomic charges of selected sidechains were modified artificially. These mutations can be classified into three types: (1) charge deletions, which remove the net charge on Arg, Lys, Asp, Glu, and doubly-protonated His; (2) charge insertions, which add a ±1 charge to Ala, Ile, Leu, Val, Met, Pro, Thr, or Tyr; (3) polarity changes, which either make Asn, Gln, or singly-protonated His apolar, or introduce a dipole onto the Cys sidechain. The details of the charge modifications are given in Supplementary Material. While these charge transformations do not correspond to real amino acid mutations, they pose the same physical problems and represent a realistic test of the methodology. They were performed with the CASA, GB, and PE solvent models. Experimental data are not available for this set, so we take PE as a reference, since it is commonly judged to be the most realistic of the three solvent models.[25,29,40] The PE model uses a protein dielectric constant of one, because we wish to model

solvent relaxation, not relaxation of the protein. The CASA model employed the Fraternali atomic solvation parameters:[61] a positive value of 0.0119 kcal/mol/Å^2 for carbon atoms and a negative value of −0.0598 kcal/mol/Å^2 for nitrogen and oxygen atoms. To optimize further the model for charge mutations, we explored several values of the CASA dielectric constant, and we introduced a new solvation parameter for charged atoms, by the following scheme. In charge deletions, the atoms involved have a large charge in their native state and a small one in the mutated state. These atoms were assigned a native-state surface coefficient σ_C (to be determined), and a mutant-state coefficient of 0.0119 kcal/mol/Å^2, reflecting their hydrophobic character after the mutation. The optimum σ_C value was determined by comparison with free energy differences evaluated by Poisson calculations (PE solvent model). In charge insertions, most of the atoms involved are carbons, with a zero initial charge and a large (positive or negative) final charge. For these mutations, the coefficients of all the relevant atoms were set to their chemical-type value at the native state, and to a value of σ_C in the mutant state. In polar-to-neutral mutations, the atoms involved were assigned a native-state coefficient according to their chemical type, and a mutant-state coefficient of −0.0598 kcal/mol/Å^2 (Cys) or 0.0119 kcal/mol/Å^2 (Asn, Gln, His). In all cases, hydrogen atoms were assigned a zero atomic surface coefficient in both states. A list of atom types associated with the new coefficient σ_C is given in Supplementary Material.

The second data set included 140 mutations in 12 proteins for which the experimental stability changes are known (listed in Supplementary Material). In each case, an ionized sidechain is introduced or removed. Each native protein structure is first energy minimized with its backbone fixed. The new sidechain is then positioned successively in each of its rotamers, minimized with the surrounding protein structure fixed, and the best rotamer is retained. To obtain the difference in stability between the mutant and the native protein, the charge modification must also be introduced in the unfolded protein. By subtracting the wildtype/mutant energy differences in the folded and unfolded states, the stability of the mutant relative to the native protein is obtained (see Fig. 1). In our calculations, the "unfolded" state corresponded to the mutated residue, isolated from the rest of the sequence (and with the same coordinates as in the native structure). This simple unfolded model is commonly used in protein design.[65–67] Mutations were performed using the CASA, GB/ACE, GB/HCT, and PE solvent treaments.

Optimization of the GB/HCT Parameters

With the GB/HCT model, each atom is assigned a volume and a scaling factor.[48] For the proteins considered here, there are 29 different atom types in the AMBER protein force field,[72] giving 58 atomic parameters. These were adjusted by an iterative, least-squares procedure to reproduce the stability changes computed with PE for artificial mutations, similar to those described above. 1020 of the mutations were chosen randomly to be optimized; the other 7130 were used for a cross-validation test of the optimization. The PE model employed the AMBER atomic charges and atomic radii optimized very recently by McCammon and coworkers[73] to reproduce a large body of experimental and simulation data. The optimized parameters are given in Supplementary Material.

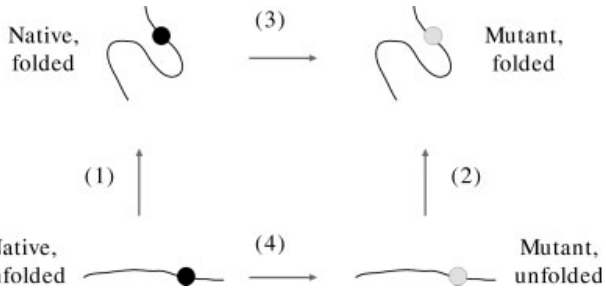


Fig. 1. Thermodynamic cycle used to calculate the change in protein stability due to charge mutations.

RESULTS

Sidechain Placement

We first describe sidechain placement with the CASA solvent model [Eq. (3)]. Twenty-nine proteins were used as a test set (Table III). Several parameterizations were compared in order to optimize the model.

Choice of the dielectric constant

We first considered the effect of the dielectric constant ε and the Coulomb energy term [Eq. (3)], using the CHARMM19 force field and three different sets of atomic solvation parameters, referred to as the Wesson, Ooi, and Fraternali sets, respectively.[59–61] The quality of the predicted structures was assessed by the agreement of sidechain dihedrals with the crystal structure and by the comparison of RMSD of the sidechain atoms with the crystal structure. Results for $\varepsilon = 4$ and $\varepsilon = 20$ are shown in Table I.

The dihedral prediction and sidechain RMSD values are comparable to those of Koehl and Delarue and Yang et al.[14,82] We compared the models with the two ε values and the three solvation parameter sets using the ANOVA statistical test. Differences of χ_1 predictions among the six models are statistically significant at the 1% significance level. The same is true for the predictions of both χ_1 and χ_2 (χ_{1+2}) and for the sidechain RMSD.

Student's test was then done for pairs of models with different ε values. For the Fraternali parameters, differences of the χ_1, χ_{1+2}, and RMSD predictions with the two ε values are not significant at the 1% significance level. The same is true for the Ooi and Wesson parameter sets. Thus, the difference detected by ANOVA is not

TABLE I. CASA Sidechain Predictions with Different Dielectric Constants ε

	Atomic parameters	χ_1 Mean (sd)	χ_{1+2} Mean (sd)	RMSD (Å) Mean (sd)	RMSD, no C_β (Å) Mean (sd)
$\varepsilon = 20$	Fraternali	0.76 (7)	0.61 (6)	1.85 (32)	1.62 (28)
	Wesson	0.71 (7)	0.56 (6)	2.05 (31)	1.77 (26)
	Ooi	0.70 (7)	0.54 (7)	2.12 (37)	1.85 (33)
$\varepsilon = 4$	Fraternali	0.77 (6)	0.63 (6)	1.85 (31)	1.61 (28)
	Wesson	0.72 (6)	0.58 (7)	1.97 (32)	1.73 (29)
	Ooi	0.69 (6)	0.55 (7)	2.13 (32)	1.86 (29)

Fraction of successful dihedral predictions and sidechain RMSD relative to crystal structure, averaged over 29 test proteins. Standard deviation in parentheses (in units of last digit). Rightmost column omits the C_β from RMSD.

TABLE II. CASA Sidechain Predictions with Different Surface Weights α

	Atomic parameters	χ_1 Mean (sd)	χ_{1+2} Mean (sd)	RMSD (Å) Mean (sd)	RMSD, no C_β (Å) Mean (sd)
$\alpha = 1/2$	Fraternali	0.79 (5)	0.66 (6)	1.79 (30)	1.57 (26)
	Wesson	0.76 (6)	0.62 (7)	1.86 (30)	1.63 (27)
	Ooi	0.74 (6)	0.62 (8)	1.94 (34)	1.70 (29)
$\alpha = 1/3$	Fraternali	0.80 (6)	0.67 (6)	1.75 (33)	1.53 (30)
	Wesson	0.78 (6)	0.65 (6)	1.82 (30)	1.59 (27)
	Ooi	0.76 (6)	0.64 (8)	1.88 (30)	1.65 (27)
	Fraternali[a]	0.80 (6)	0.66 (6)	1.74 (33)	1.52 (30)
$\alpha = 0$	none	0.84 (5)	0.68 (6)	1.68 (29)	1.46 (24)

Fraction of successful dihedral predictions and sidechain RMSD relative to crystal structure, averaged over 29 test proteins. Standard deviation in parentheses (in units of last digit). Rightmost column omits the C_β from RMSD. Dielectric constant is $\varepsilon = 4$.
[a] The surface areas of buried sidechains are not downscaled [$s_i = 1$ in Eq. (9)].

related to the choice of ε, which does not affect sidechain prediction. This is consistent with several earlier studies, in which good results were obtained without any electrostatic term.[14,16,21] As pointed out above, however, a model without electrostatics is not appropriate here, since our goal is to do protein design. All the following calculations include the Coulomb energy term with $\varepsilon = 4$.

Choice of the atomic solvation parameters and the weight α

Next, statistical tests were done to determine whether the three solvation parameter sets were significantly different (using $\varepsilon = 4$). Results are summarized in Table II. ANOVA indicates that the null hypothesis is rejected at the 1% significance level for the three quality criteria considered (χ_1, χ_{1+2}, RMSD). Thus, there is at least one prediction protocol that differs. Results with the Ooi and Wesson parameters are equivalent according to the Student test at the 5% significance level for all three quality criteria. However, differences in the χ_1 predictions and the RMSD calculations between Fraternali and the two other parameter sets are statistically significant at the 5% significance level, and differences for χ_{1+2} are significant at the 1% level. Thus, the Fraternali parameters perform better than the other two sets. The Fraternali protocol gave an average success rate for χ_1 and χ_{1+2} predictions of 77% and 63%, respectively.

We next considered a model with $\varepsilon = 4$ but no surface area term [$\alpha = 0$ in Eq. (3)]. The results are actually improved (see Table II). Indeed, the average of the χ_1 and χ_{1+2} predictions are about 84% and 68%, somewhat better than those of Koehl and Delarue and Yang et al.[14,82] It appears that the fully-weighted surface area term is too large, compared to the other terms in the energy function. Nevertheless, further calculations were done with a nonzero α. Indeed, we show below that $\alpha = 1/2$ gives much better results for experimental mutations and their associated stability changes. For protein design, with sidechain reconstruction and mutagenesis occurring simultaneously, we need a consensus model that performs well at both tasks.

Further calculations were therefore done with $\alpha = 1/2$ and $\alpha = 1/3$ (Table II). Results improve with decreasing α. Successful χ_1 prediction with the Fraternali parameters increases to 80% with $\alpha = 1/3$, while χ_{1+2} increases to 67%. These percentages approach those obtained with $\alpha = 0$. The Ooi and Wesson parameters give somewhat poorer results.

We performed ANOVA for each parameter set to compare the different α values (1/3, 1/2, 1). For the Fraternali set, for example, the three protocols were not equivalent according to the χ_1 and χ_{1+2} criteria at the 1% significance level. The $\alpha = 1/2$ and $\alpha = 1/3$ Fraternali models are not significantly different according to the Student's t-test; both are superior to the $\alpha = 1$ model.

Overall, the different Student's tests show that Fraternali with $\alpha = 0$ to 1/2 and Wesson with $\alpha = 1/3$ are equivalent and provide the best results. Table III compares Fraternali, $\alpha = 1/3$ with the work of Koehl and Delarue and Yang et al.[14,82]

TABLE III. CASA Sidechain Predictions Compared to Selected Earlier Work

		CASA, this work (Fraternali, $\varepsilon = 4$, $\alpha = 1/3$)				Koehl and Delarue			Yang et al.		
PDB code	Chain length	χ_1	χ_{1+2}	RMSD[a]	RMSD[b]	χ_1	χ_{1+2}	RMSD[a]	χ_1	χ_{1+2}	RMSD[b]
1PPT	36	0.73	0.64	1.59	1.40	0.73	0.67	1.57	–	–	–
1CRN	46	0.89	0.76	0.88	0.75	0.78	0.68	1.42	0.95	0.84	0.84
2OVO	56	0.90	0.79	1.30	1.12	0.69	0.60	2.35	–	–	–
4PTI	58	0.78	0.70	2.05	1.81	0.87	0.74	1.85	0.80	0.65	1.26
1IGD	61	0.84	0.72	1.22	1.06	–	–	–	0.76	0.74	1.28
1ISU	62	0.95	0.79	1.36	1.19	–	–	–	0.84	0.74	1.12
2SN3	65	0.81	0.68	2.31	2.03	0.59	0.45	2.76	–	–	–
1PTX	68	0.82	0.74	2.16	1.92	–	–	–	0.74	0.61	1.73
1UBQ	76	0.75	0.57	2.19	1.90	0.74	0.57	2.06	–	–	–
1PLC	99	0.77	0.66	1.40	1.21	0.76	0.67	1.55	0.82	0.70	1.24
1LN4	104	0.73	0.62	1.90	1.64	–	–	–	–	–	–
1AAC	105	0.81	0.68	1.81	1.58	–	–	–	0.92	0.69	1.05
256B	106	0.80	0.67	2.01	1.75	–	–	–	0.73	0.53	1.53
2TRX	108	0.82	0.61	1.88	1.65	–	–	–	–	–	–
5CPV	109	0.78	0.61	1.69	1.48	0.68	0.54	1.99	–	–	–
1CCR	112	0.75	0.64	1.86	1.62	–	–	–	0.84	0.60	1.21
1THX	115	0.76	0.65	1.53	1.32	–	–	–	–	–	–
1WHI	122	0.77	0.67	1.92	1.66	–	–	–	0.73	0.65	1.52
1PMY	123	0.80	0.63	1.95	1.70	–	–	–	0.83	0.63	1.23
3RN3	124	0.68	0.60	2.03	1.76	0.66	0.58	2.24	–	–	–
1LZ1	130	0.85	0.70	1.66	1.47	0.80	0.73	1.55	0.82	0.66	1.16
2END	137	0.79	0.63	1.89	1.66	–	–	–	0.79	0.64	1.37
2FOX	138	0.79	0.69	1.46	1.27	–	–	–	0.74	0.56	1.35
2HBG	147	0.77	0.61	1.79	1.57	–	–	–	0.76	0.61	1.34
2RN2	155	0.81	0.67	2.01	1.77	–	–	–	–	–	–
2CPL	165	0.86	0.71	1.49	1.31	–	–	–	–	–	–
1KOE	172	0.73	0.63	1.85	1.61	–	–	–	–	–	–
1XNB	185	0.81	0.70	1.58	1.39	–	–	–	0.78	0.66	1.89
1ES9	212	0.78	0.61	1.92	1.68	–	–	–	–	–	–
Mean	110	0.80	0.67	1.74	1.53	0.72	0.62	1.89	0.80	0.66	1.36

Fraction of successful dihedral predictions and sidechain RMSD (Å) relative to crystal structure.
[a] C_β excluded.
[b] C_β included.

To test for a possible force-field dependency of these results, calculations with the AMBER force field[72] were performed for six of the proteins. The accuracy of sidechain prediction is equivalent to that observed with the CHARMM19 force field, with the $\varepsilon = 4$, $\alpha = 1/3$, Fraternali protocol (data not shown). For example, the average χ_1 prediction rate is 75% with both AMBER and CHARMM19 for these six proteins.

The rate of successful prediction was analyzed as a function of amino acid type. In agreement with earlier work, it is easier to predict the rotamers of hydrophobic residues. Their rate of successful χ_1 prediction is 92% with the best parameterization, compared to 71% for hydrophilic residues. This can be explained because hydrophobic residues are mainly located in the protein core and are constrained by van der Waals packing interactions. There is little difference in accuracy between large and small residues. A low χ_{1+2} accuracy is observed for Asn, Gln, and His, since these residues' χ_2 angle can often be flipped by 180° with only a small change in energy.

Solvation Energies: Comparing Solvent Treatments

To directly compare the three solvent models, CASA, GB, and PE, we computed Coulomb and solvation energies for four proteins: trypsin inhibitor (4PTI), thioredoxin (2TRX), lysozyme (1LZ1), and ubiquitin (1UBQ). For each protein, we generated 750–1000 structures by randomizing sidechain rotamers. In the CASA energies, the Coulomb term is divided by a dielectric constant ε [Eq. (3)]. We initially searched a range of values, $\varepsilon = 1$–20, to find the optimum for each protein.

The RMSDs between PE and CASA are listed in Table IV. The optimal dielectric constants are fairly uniform among proteins: $\varepsilon = 1.5$–2.5. With these ε values, the PE/CASA RMSD ranges from 33 kcal/mol (4PTI) to 58 kcal/mol (2TRX). Ultimately, we require a consensus CASA model that performs well for sidechain placement, solvation energies, and charge mutations, and for all proteins. Sidechain placement (above) is insensitive to ε, while charge mutations (below) work best with a large value, $\varepsilon = 16$–20. Table IV also reports the CASA solva-

TABLE IV. Protein Solvation: RMS Deviation Between the CASA and GB Energies and the PE Reference Values

	CASA dielectric constant ε								
Protein	1	1.5	2.0	2.5	3.0	4.0	20.0	GB/ACE	GB/HCT
4PTI	46.9	**32.5**	32.8	35.7	38.5	42.9	55.6	71.7	20.7
2TRX	87.7	64.8	59.1	**58.3**	59.0	61.1	71.0	99.6	30.8
1LZ1	80.6	59.1	**56.2**	57.7	60.0	64.1	78.3	163.5	44.7
1UBQ	70.1	52.0	**49.8**	51.3	53.4	57.1	69.6	102.1	26.4

All values in kcal/mol. For each protein, the deviations are evaluated over 750–1000 rotameric structures. Best CASA results in boldface.

TABLE V. Comparison Between the CASA and Poisson Models for Charge Mutations

Protein	No of mutations	Optimal σ_C (kcal/mol/Å^2)	Optimal ε	PE-CASA RMSD[a] (kcal/mol)	Success[b] rate (%)
AspRS	518	−0.15 (−0.20)	16	17.5 (17.9)	80.1 (80.1)
3RN3	106	−0.20	16	13.1	79.0
4PTI	51	−0.20	16	11.0	74.4
5CYT	84	−0.20	16	13.5	79.8
2TRX	94	−0.20	16	12.3	80.8
1LZ1	108	−0.20	16	12.2	83.4
1UBQ	67	−0.20	16	10.7	86.5

Proteins are indicated by their PDB code, except AspRS, which is *Escherichia coli* aspartyl-tRNA synthetase. The surface coefficient σ_C is associated with atoms on charged sidechains (see text).
[a]RMS deviation between the surface area (CASA) and Poisson (PE) energy differences.
[b]Percentage of mutations that are predicted to have a positive or negative stability change by both CASA and PE.

tion results with $\varepsilon = 20$. The RMS deviations from PE increase to 56—71 kcal/mol. Thus, the loss in performance is significant when one takes the charge mutation dielectric as a consensus value for all proteins.

The CASA, GB/ACE, and GB/HCT total solution energies are compared with PE in Figure 2. Note that CASA and GB/ACE use the Charmm19 atomic charges, and so they are compared with PE performed with these charges. GB/HCT uses the AMBER charges, and is compared with PE performed with the AMBER charges. A CASA ε of 20 is used. There is a considerable energy dependency on the rotameric structure, with values between about −300 and +300 kcal/mol. The trend of the PE energies is reproduced approximately by both GB/ACE and CASA. The CASA surface term is almost constant for all structures (not shown), so that the CASA behavior is governed by the Coulomb term [Eq. (3)]. The CASA/PE correlation indicates, therefore, that the PE solvation energies correlate quite well with the Coulomb energy. The best correlation by far is obtained with GB/HCT. The RMS deviation between GB/HCT and PE is also quite small (21–45 kcal/mol; Table IV), considerably lower than that of CASA with the consensus dielectric constant.

Charge Modifications: Comparing Solvent Treatments

The third and most important test of the solvent models is the calculation of stability changes due to mutations that introduce or remove charged groups. In protein design, mutations are introduced randomly, and the largest stability changes will usually be associated with changes in the net protein charge. This is especially true for mutations that affect charges in fully or partly buried positions.[69] We performed two sets of mutations. The first included over 1000 mutations in seven proteins for which experimental data are not available, described in this section. The second included 140 mutations in 12 proteins for which experimental data are available, described in the next section.

Artificial mutations: CASA vs. PE

We used the Fraternali atomic surface parameters, along with a new atomic parameter σ_C for certain atoms belonging to ionized groups. The new parameter allows us to optimize the accuracy of the CASA model for charge deletions and insertions. The atom types associated with this new coefficient are listed in Supplementary Material. Values of the atomic parameter σ_C between −0.20 kcal/mol/Å^2 and the original value, −0.0598 kcal/mol/Å^2, were tried, in combination with dielectric constants in the range $\varepsilon = 2$–20. The CASA and PE energy changes for each mutation were compared. The combination of σ_C and ε that maximizes the agreement between CASA and PE is listed in Table V for the seven proteins. The optimum coefficient for atoms on charged sidechains (in their charged state) is between −0.15 and −0.20 kcal/mol/Å^2, and the optimum dielectric constant is 16–20. With these values, the RMSD between the CASA and PE stability changes ranges from 17.5 kcal/mol for the largest protein (AspRS) to 10.0–11.0 kcal/mol for the smallest proteins (ubiquitin and BPTI).

As explained above, we would like to obtain a consensus parameterization that works well for all proteins. The best σ_C and ε values are the same for all but one protein: a slightly smaller σ_C of −0.15 kcal/mol/Å^2 is

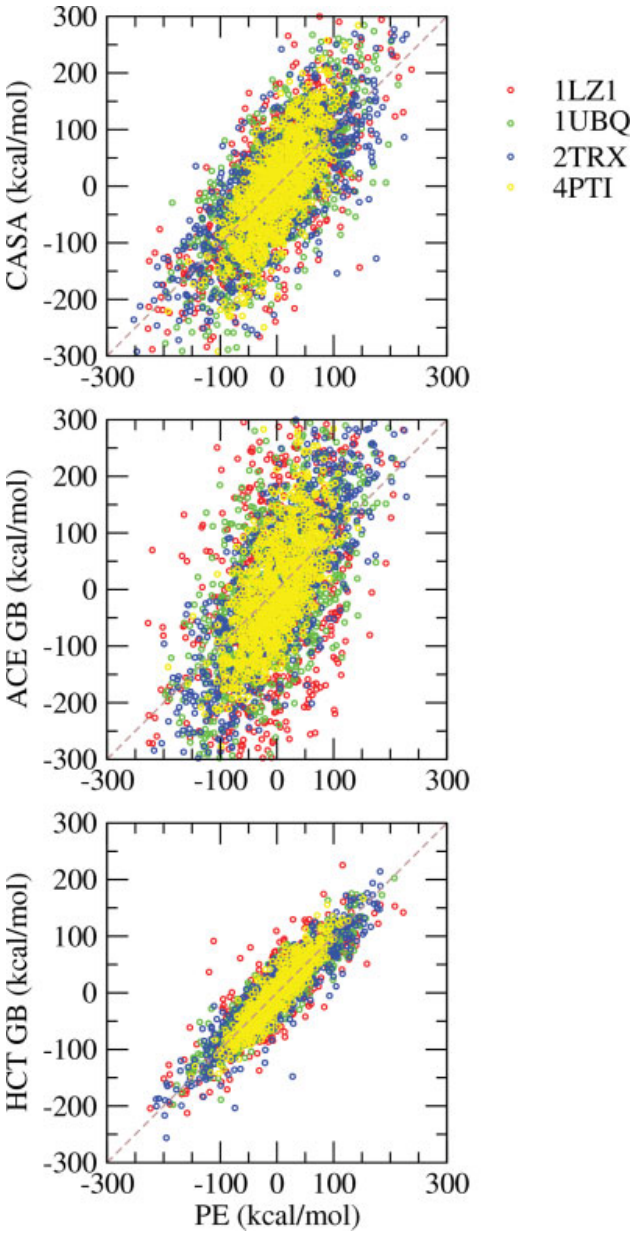


Fig. 2. Total solution energies of protein rotamer structures with the GB/ACE, surface area (CASA), and GB/HCT models, versus the corresponding Poisson values (PE). Points are colored according to the protein, listed by their PDB codes: 1LZ1, lysozyme; 1UBQ, ubiquitin; 2TRX, thioredoxin; 4PTI, bovine pancreatic trypsin inhibitor.

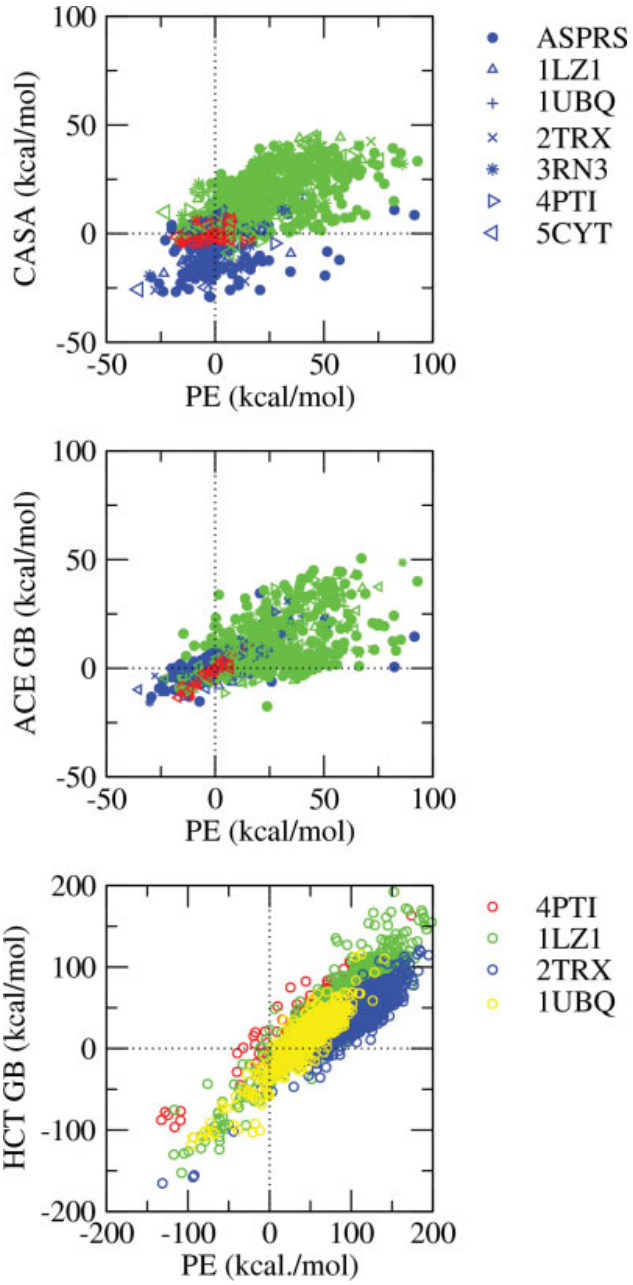


Fig. 3. Stability changes due to charge mutations with CASA, GB/ACE, and GB/HCT, vs. the corresponding changes in the PE model. For CASA and GB/ACE, green, blue and red points correspond, respectively, to charge insertions, charge deletions and polar-to-neutral conversions; different symbols correspond to different proteins, listed by their PDB codes. For GB/HCT, points are colored by protein.

best for AspRS. Using the consensus σ_C value for AspRS gives almost the same result (Table V). The same σ_C and ε values work well for the experimental mutations, explained later. We conclude that they are likely to work well for most mutations in most proteins.

The CASA stability changes are plotted against the PE ones in Figure 3. Most charge insertions decrease the protein stability, both in the CASA and PE models (i.e., they give a positive double free energy difference, $\Delta G_{mut} - \Delta G_{nat}$, for the thermodynamic cycle in Fig. 1). On the other hand, most charge deletions increase the protein stability. Most charged residues are at the protein surface and their interactions with the rest of the protein are screened by the nearby solvent. Eliminating the charge on such residues destabilizes the folded state, mainly due to the loss of interactions between the charge and the solvent. In the unfolded state, the same charges

TABLE VI. Comparison Between the GB and Poisson Models for Charge Mutations

		GB/ACE, Approx 1		GB/ACE, Approx 2		GB/HCT	
Protein	No of mutations	RMSD[a]	% success[b]	RMSD[a]	% success[b]	RMSD[a]	% success[b]
AspRS	518	40.7	47.9	18.2	82.2		
3RN3	106	21.6	66.9	14.0	84.0		
4PTI	51	13.7	74.5	7.0	70.6	8.3	99.2
5CYT	84	15.2	65.5	12.6	73.8		
2TRX	94	26.3	64.9	17.4	91.5	12.1	98.5
1LZ1	108	22.8	59.2	12.3	80.6	16.5	98.2
1UBQ	67	17.1	56.8	11.6	76.1	10.8	92.3

[a]RMS deviation between the GB and Poisson (PE) total energy differences (kcal/mol).
[b]Percentage of mutations predicted to have a positive or negative stability change by both GB and PE. Approximations 1 and 2 are explained in the text. Proteins are indicated by their PDB code, except AspRS, which is *Escherichia coli* aspartyl-tRNA synthetase.

are even more exposed to solvent. Thus, the charge elimination destabilizes the unfolded state even more, yielding a negative (favorable) double free energy difference. CASA and PE produce the same stability order (i.e., sign of the stability change) for 81% of mutations in all proteins. The CASA energy differences vary in a range of ≈80 kcal/mol, whereas the corresponding PE range is 150 kcal/mol; nevertheless, the two distributions are reasonably well-correlated.

Calculations were also done with a "distance-dependent" dielectric constant, $\varepsilon(r) = \varepsilon_0 r$, where r is the separation between a pair of charges and ε_0 is a constant. The success rate increased to about 86% if one uses $\varepsilon_0 = 2$. However, the performance of this model variant is much poorer for the experimental mutations, below. Increasing ε_0 improves the experimental mutations but deteriorates the artificial ones (not shown).

Artificial mutations: GB/ACE and GB/HCT vs. PE

The mutations are of three types: (1) charge insertions (usually, introduction of charge onto hydrophobic atom types); (2) charge deletions (elimination of charge on hydrophilic atom types), and (3) polar-to-neutral conversions. Thus, the mutated atoms have a combination of charge and van der Waals radius that is different from the optimum value for GB calculations. For this reason, it was necessary to modify some of the GB/ACE parameters. Table VI contains the results of two approximations. The first approximation corresponds to the usual GB/ACE parameterization (i.e., V01 volumes scaled by 0.8).[52] In the second approximation, the hydrogen types HC are assigned a Voronoi volume of 4.0, the V01 volumes of atom types CH1E, CH2E, CH3E (carbons of hydrophobic sidechains, carrying a significant partial charge in the charge-insertion mutations) were scaled by a factor 0.4, and nitrogen types NH3, NC2 were set to the original V01 volumes; for all other atom types, the V01 volumes were scaled by 0.8. The same parameterization is used for all proteins. As seen from Table VI, approximation 2 yields an RMSD between GB/ACE and PE of 7–18 kcal/mol. The fraction of mutations predicted with a positive or negative stability change by models is 80%. Thus, with this set of parameters, the GB/ACE model is comparable to CASA. Another recent study reported poorer results with GB/ACE for a similar test.[20] This was presumably due to the use of an early and nonoptimal parameterization of GB/ACE; see Calimet et al.[52] for a critical discussion of the early parameterization. The GB/ACE stability changes with approximation 2 are plotted against the corresponding PE results in Figure 3. Most charge insertions destabilize the protein, as with CASA. On the other hand, most charge deletions increase the stability.

The GB/HCT results are also given in Table VI and Figure 3. The GB/HCT parameters were extensively fitted to the PE model in this work (see Methods section). However, all the GB/HCT—PE comparisons correspond to cross validation tests, using PE data not employed during the fits. The set of proteins and mutations is slightly different from GB/ACE, with fewer (four) proteins, but a larger number of mutations per protein. The RMSD between GB/HCT and PE is 8—16 kcal/mol, smaller than for CASA and GB/ACE. Correlation between GB/HCT and PE is excellent, far superior to CASA and GB/ACE, with 97% of the stability changes predicted to have the correct sign.

Charge Modifications: Comparing to Experiment

We considered 140 mutations in 12 proteins (Table VII), with experimental stability changes taken either from the ProTherm database[83] or Ref. 74. All mutations introduced or removed an ionized residue. The stability changes computed with CASA, GB/ACE, GB/HCT, and PE are summarized in Table VII and detailed in Supplementary Material. We report the mean unsigned error (MUE) for each protein, with selected outliers left out. The outliers were identified by a large van der Waals contribution to the stability change (10 kcal/mol or more). Such a large contribution reflects steric conflict in the mutant structure that results from our simple sidechain construction method. Overall, we identify 7 outliers with GB/HCT, 8 with CASA, and 21 with GB/ACE. Calculations with GB are slower than with CASA; e.g., with GB/HCT, a typical

TABLE VII. Comparison Between Models and Experiment for Charge Mutations

Protein	PDB code	Total number of mutants	Mean unsigned error (kcal/mol)			
			CASA $\varepsilon = 16$	GB/ACE $\varepsilon_P = 4$	GB/HCT $\varepsilon_P = 4$	PE $\varepsilon_P = 4$
Protein G, B_1 domain	1EM7	2	0.31 (2)	3.5 (2)	2.7 (2)	1.6 (2)
Chymotrypsin inhibitor	1YPC	15	3.3 (15)	4.3 (13)	4.4 (15)	4.6 (15)
Lysozyme	2LZM	14	2.6 (13)	2.6 (12)	4.2 (14)	4.4 (14)
Ribonuclease	2RN2	25	2.5 (24)	3.4 (23)	3.6 (24)	4.8 (24)
Src SH3 domain	1SHG	5	1.8 (5)	4.6 (5)	3.6 (5)	3.2 (5)
Staphylococcal nuclease	1STN	45	1.9 (40)	2.2 (33)	3.8 (42)	3.9 (42)
Thioredoxin	2TRX	3	5.1 (3)	6.8 (2)	1.4 (2)	1.8 (2)
Trypsine	1BPI	11	1.4 (10)	2.4 (10)	1.7 (9)	6.7 (9)
Ubiquitin	1UBQ	8	2.1 (8)	1.5 (8)	1.6 (8)	1.4 (8)
pepT1[c]	–	4	1.6 (4)	2.4 (4)	2.8 (4)	1.9 (4)
K2AE2[c]	–	4	1.9 (4)	2.3 (4)	1.8 (4)	1.4 (4)
KEAKE[c]	–	4	1.8 (4)	2.0 (4)	1.6 (4)	1.2 (4)
Total		140	2.1 (132)[a]	2.9 (120)	3.4 (133)	3.9 (133)
					2.1[b] ($\varepsilon_P = 8$)	2.6[b] ($\varepsilon_P = 8$)

[a]In parentheses: the number of mutations after discarding those with a van der Waals contribution of 10 kcal/mol or more.
[b]GB/HCT or PE supplemented by a surface area term and using a protein dielectric of 8 (see text).
[c]Experimental data from Ref. 88. Natives structures are not known experimentally; they were therefore modelled using the Swiss PDB Viewer, which constructs an ideal α-helix and places sidechains in favorable rotamers.[89]

mutation takes 2–4 times more CPU time than with CASA.

Excluding the outliers, the MUE is 2.1 kcal/mol with CASA. Poorer results are obtained if the constant dielectric in the Coulomb electrostatic term is replaced by a distance-dependent dielectric, $\varepsilon(r) = \varepsilon_0 r$: with $\varepsilon_0 = 2$, for example, the error increases to 7.9 kcal/mol. Similarly, while sidechain reconstruction (above) works well without any surface term [$\alpha = 0$ in Eq. (3)], results here are much poorer: the MUE increases to 4.6 kcal/mol if the surface term is left out. These results illustrate further that the best model for protein design should be a consensus model, not necessarily optimal for each prediction task but reasonably competent for all of them.

In the case of GB/ACE, GB/HCT, and PE, the protein dielectric constant ε_P is adjusted empirically to minimize the MUE. Indeed, our simple sidechain modeling does not allow dielectric relaxation of the protein structure when an ionized sidechain is introduced or removed. For all three models, GB/ACE, GB/HCT, and PE, the best results are obtained with a protein dielectric of 3–5. Note that the use of GB models with a protein dielectric greater than 1 [Eq. (4)] is not very common, but is straightforward.[42,84] A consensus value of four for the protein dielectric works well with all three models, giving MUEs of 3.4 kcal/mol with GB/HCT and 3.9 kcal/mol with PE. Results with a dielectric of one are very poor (e.g., the MUE is 8.7 kcal/mol with PE). With GB/ACE, mutations involving Asp and Glu sidechains were found to be poorly described. Therefore, another adjustable parameter was introduced, which empirically increases by 6.3 kcal/mol the contribution of Asp or Glu sidechains to the stability of the unfolded state. This leads to a MUE of 2.9 kcal/mol with GB/ACE (using $\varepsilon_P = 4$, and with 21 outliers omitted). With GB/HCT and PE, the prediction quality did not depend noticeably on the amino acid type.

Mutation of a charged sidechain could, in some cases, alter the protonation state of surrounding residues. It is not practical to model this effect in detail, because there are too many possible protonation states to consider. It is implicitly incorporated into the model through the choice of a protein dielectric constant greater than 1. Furthermore, for each protein, we systematically did CASA calculations with the histidine sidechains either all ionized or all neutral. Results for the mutations were reasonably similar (not shown).

The performance of the three models, GB/ACE, GB/HCT, and PE, could be improved by adding an additional, surface area term to describe hydrophobic solvation and dispersion interactions with the surrounding solvent, as is commonly done in protein modeling, see e.g. Refs. 32,85. Using a surface coefficient of $\sigma = -0.04$ kcal/mol/Å^2 for all atoms and a somewhat larger dielectric of $\varepsilon = 8$, the mean unsigned error for GB/HCT drops to 2.1 kcal/mol, identical to the CASA value. For PE, the mue drops to 2.6 kcal/mol, using a surface coefficient of $\sigma = -0.05$ kcal/mol/Å^2 for all atoms and a dielectric of $\varepsilon = 8$. We did not try this procedure with GB/ACE, because of its poorer overall performance.

The experimental agreement is slightly worse with PE than with CASA and GB/HCT. Nevertheless, we took PE to be the reference model for the artifical mutations, above. Indeed, it is the model with the clearest physical basis and the best performance in general for protein electrostatics.[29,68] CASA, in contrast, is not expected to give quantitative accuracy for mutations of buried sidechains, since the surface energy cannot distinguish between positions that are deeply buried and positions that are closer to the protein surface. GB/HCT has a MUE close to PE, largely because its parameters have been optimized to reproduce PE. Thus, even though the PE mue is slightly larger, our results do not contradict

many earlier studies of protein electrostatics showing that PE is a valid reference model.

CONCLUSIONS

We have examined three problems that are important ingredients of computational protein design: sidechain placement, protein solvation, and mutagenesis involving charged sidechains. We have tested the behavior of four implicit solvent models: the Poisson model, considered to be the standard of accuracy, the GB/ACE and GB/HCT generalized Born models, and the CASA model. The CASA model is commonly used for protein design; GB/ACE and GB/HCT are more recent and sophisticated solvent models. Several methods have been proposed that allow GB models to be used for protein design.[19,23,38] The most recent one achieves a residue–pairwise GB implementation without any loss in accuracy.[23]

Using a standard mean field method for sidechain placement along with the CASA solvent model, we obtain results of similar quality to earlier workers.[14,82] The results are weakly sensitive to the details of the CASA model: solvent dielectric, overall weight of the surface term, force field. This is consistent with the good results obtained by some earlier workers without any electrostatic term or solvent model.[14,16,21]

In contrast, the mutagenesis results are much more sensitive to the solvent treatment. This is expected, since almost all the mutations involved insertions or deletions of a net charge. Many of the insertions were on buried sidechains. All four solvent models were compared. Different kinds of parameter optimization were performed. In the CASA model, we introduced a new atomic surface parameter for oxygen and nitrogen atoms in charged sidechains and we adjusted the dielectric constant. In the GB/ACE model, we adjusted the atomic volumes of selected atom types belonging to charged sidechains. In the GB/HCT model, we optimized the atomic volumes and scaling factors (58 parameters in all). All these optimizations used a large data set of artifical mutations, and took the Poisson model as the reference. The Poisson model employed a dielectric constant of one for the protein (and 80 for solvent), because we are optimizing the implicit solvent treatments, which try to reproduce the relaxation of solvent, not protein, in response to the mutations. Similarly, the GB models employed a protein dielectric of one and a solvent dielectric of 80. CASA uses a single dielectric constant, which tries to capture the average effect that solvent relaxation exerts within the protein interior. A value of about 20 worked well—intermediate between the solvent value of 80 and the value of one that is appropriate for the protein interior in these calculations. With these parameter adjustments and dielectric constants, the quality of CASA and GB/ACE were very similar, as compared to the Poisson reference. In 80–81% of the charged mutations, they correctly captured the sign and order of magnitude of the protein stability change. With GB/HCT, the sign was correct for 97% of the mutations, a dramatic improvement with this more recent GB variant.[48] Protein solvation energies were also strikingly better with GB/HCT than with CASA or GB/ACE.

Finally, in a separate test, we compared the models to experiment for a set of 140 point mutations. Comparison with experiment introduces several major difficulties. First, the stability changes are very small, so that a simple null model will usually give better agreement than any current implicit solvent model. Second, we need a model for the unfolded protein, whose structure is not known. Third, we must describe the structural relaxation *of the protein*, and not just that of the solvent. Fourth, we must describe hydrophobic contributions to stability, which are notoriously hard to capture with simple models.[25,86] To describe the unfolded state, we adopted the simplest possible, commonly-used, tripeptide model. To describe the protein relaxation, we increased the protein dielectric in the GB and PE models, exploring values between 2 and 32. With CASA, we kept a dielectric of 20, and verified that increasing it had only a small effect on the results. With GB/ACE, we also added an empirical correction for Asp/Glu sidechains in the unfolded state. This correction presumably reflects a problem with the GB/ACE parameterization of carboxylate groups. Finally, we incorporated hydrophobic contributions, along with solvent dispersion interactions into the GB/HCT and PE models by adding a surface term. All the solvent models gave fair agreement with experiment, with mean unsigned errors of 2.1 kcal/mol for CASA, 2.9 kcal/mol for GB/ACE, and 2.1 kcal/mol for GB/HCT supplemented by the surface term. GB/HCT also gave excellent results for the artifical mutations and the solvation energies.

In summary, (a) we confirm earlier observations that the choice of electrostatic model is not very important for sidechain placement; (b) GB/HCT and CASA give the same accuracy for surface mutations; (c) GB/HCT yields an enormous improvement for total solvation free energies and for mutations in fully or partly buried positions. Thus, for problems like protein design that involve all these aspects, the most recent GB models, and their best pairwise implementations,[87] represent an important step forward.

ACKNOWLEDGMENTS

We thank A. Jaramillo and M. Schmidt am Busch for useful discussions. Support was provided by the Egide program ZENON between Cyprus and France (to GA and TS), by the ACI IMPBio Program of the French Ministry of Research (to TS), by the HPC-EUROPA project RII3-CT-2003-506079 (to GA and TS), and by a Cyprus "Research Reinforcement of Young Cypriot Researchers" (PENEK) grant (to GA).

REFERENCES

1. Hellinga H. Metalloprotein design. Curr Opin Biotech 1996;7: 437–441.
2. Al-Lazikani B, Jung J, Xiang Z, Honig B. Protein structure prediction. Curr Opin Chem Biol 2001;5:51–56.
3. Marti-Renom M, Stuart A, Fiser A, Sanchez R, Melo F, Sali A. Comparative protein structure modeling of genes and genomes. Annu Rev Biophys Biomol Struct 2000;29:291–325.

4. Bolon D, Mayo S. Enzyme-like proteins by computational design. Proc Natl Acad Sci USA 2001;98:14274–14279.
5. Ponder J, Richards FM. Tertiary templates for proteins: use of packing criteria in the enumeration of allowed sequences for different structural classes. J Mol Biol 1988;193:775–791.
6. Tuffery P, Etchebest C, Hazout S, Lavery R. A new approach to the rapid determination of protein side chain conformations. J Biomol Struct Dyn 1991;8:1267.
7. Dunbrack R, Karplus M. Backbone-dependent rotamer library for proteins. Application to sidechain prediction. J Mol Biol 1993; 230:543–574.
8. Dunbrack R, Cohen F. Bayesian statistical analysis of protein sidechain rotamer preferences. Prot Sci 1997;6:1661–1681.
9. Desmet J, De Mayaer M, Hazes B, Lasters I. The dead-end elimination theorem and its use in protein sidechain positioning. Nature 1992;356:539–542.
10. Goldstein R. Efficient rotamer elimination applied to protein sidechains and related spin glasses. Biophys J 1994;66:1335–1340.
11. Looger L, Hellinga H. Generalized dead-end elimination algorithms make large-scale protein sidechain structure prediction tractable: implications for protein design and structural genomics. J Mol Biol 2001;307:429–445.
12. Holm L, Sander C. Database algorithm for generating protein backbone and sidechain coordinate from a C_α trace: application to model building and detection of coordinates errors. J Mol Biol 1991;218:183–194.
13. Lee C, Subbiah S. Prediction of protein sidechain conformation by packing optimization. J Mol Biol 1991;217:373–388.
14. Koehl P, Delarue M. Application of a self-consistent mean field theory to predict protein sidechain conformations and estimate their conformational entropy. J Mol Biol 1994;239:249–275.
15. Mendes J, Baptista A, Carrondo M, Soares C. Implicit solvation in the self-consistent mean field theory method: sidechain modelling and prediction of folding free energies of protein mutants. J Computer Aided Mol Des 2001;15:721–740.
16. Liang S, Grishin N. Sidechain modeling with an optimized scoring function. Prot Sci 2002;11:322–331.
17. Onufriev A, Bashford D, Case D. Exploring protein native states and large-scale conformational changes with a modified generalized Born model. Proteins 2004;55:383–394.
18. Feig M, Onufriev A, Lee M, Im W, Case D, Brooks CL, III. Performance comparison of generalized Born and Poisson methods in the calculation of electrostatic solvation energies for protein structures. J Comput Chem 2004;25:265–284.
19. Pokola N, Handel T. Energy functions for protein design I: efficient and accurate continuum electrostatics and solvation. Prot Sci 2004;13:925–936.
20. Jaramillo A, Wodak S. Computational protein design is a challenge for implicit solvation models. Biophys J 2005;88:156–171.
21. Canutescu AA, Shelenkov AA, Dunbrack R. A graph-theory algorithm for rapid protein side-chain prediction. Prot Sci 2003; 12:2001–2014.
22. Feig M, Brooks CL, III. Recent advances in the development and application of implicit solvent models in biomolecule simulations. Curr Opin Struct Biol 2004;14:217–224.
23. Archontis G, Simonson T. A residue-pairwise Generalized Born scheme suitable for protein design calculations. J Phys Chem B 2005;109:22667–22673.
24. Tanford C. The hydrophobic effect. New York: John Wiley; 1980.
25. Roux B, Simonson T. Implicit solvent models. Biophys Chem 1999;78:1–20.
26. Warwicker J, Watson H. Calculation of the electrostatic potential in the active site cleft due to α helix dipoles. J Mol Biol 1982;157:671–679.
27. Honig B, Nicholls A. Classical electrostatics in biology and chemistry. Science 1995;268:1144–1149.
28. Schaefer M, Vlijmen Hv, Karplus M. Electrostatic contributions to molecular free energies in solution. Adv Prot Chem 1998;51:1–57.
29. Simonson T. Electrostatics and dynamics of proteins. Rep Prog Phys 2003;66:737–787.
30. Simonson T. Macromolecular electrostatics: continuum models and their growing pains. Curr Opin Struct Biol 2001;11:243–252.
31. Sitkoff D, Sharp K, Honig B. Accurate calculation of hydration free energies using macroscopic solvent models. J Phys Chem 1994;98:1978–1988.
32. Simonson T, Brünger AT. Solvation free energies estimated from macroscopic continuum theory: an accuracy assessment. J Phys Chem 1994;98:4683–4694.
33. Bashford D, Karplus M. The pK_a's of ionizable groups in proteins: atomic detail from a continuum electrostatic model. Biochemistry 1990;29:10219–10225.
34. Archontis G, Simonson T, Karplus M. Binding free energies and free energy components from molecular dynamics and Poisson-Boltzmann calculations. Application to amino acid recognition by aspartyl-tRNA synthetase. J Mol Biol 2001;306:307–327.
35. Hendsch Z, Tidor B. Electrostatic interactions in the GCN4 leucine zipper: substantial contributions arise from intramolecular interactions enhanced on binding. Prot Sci 1999;8:1381–1392.
36. Gohlke H, Kiel C, Case D. Insight into protein-protein binding by binding free energy calculation and free energy decomposition for the Ras-Raf and Ras-RalGDS complexes. J Mol Biol 2003;330:891–913.
37. David L, Luo R, Gilson M. Comparison of Generalized Born and Poisson models: energetics and dynamics of HIV protease. J Comput Chem 2000;21:295–309.
38. Wisz M, Hellinga H. An empirical model for electrostatic interactions in proteins incorporating multiple geometry-dependent dielectric constants. Proteins 2003;51:360–377.
39. Still WC, Tempczyk A, Hawley R, Hendrickson T. Semianalytical treatment of solvation for molecular mechanics and dynamics. J Am Chem Soc 1990;112:6127–6129.
40. Bashford D, Case D. Generalized Born models of macromolecular solvation effects. Ann Rev Phys Chem 2000;51:129–152.
41. Hawkins G, Cramer C, Truhlar D. Pairwise descreening of solute charges from a dielectric medium. Chem Phys Lett 1995;246:122–129.
42. Schaefer M, Karplus M. A comprehensive analytical treatment of continuum electrostatics. J Phys Chem 1996;100:1578–1599.
43. Qiu D, Shenkin P, Hollinger F, Still W. The GB/SA continuum model for solvation. A fast analytical method for the calculation of approximate Born radii. J Phys Chem A 1997;101:3005–3014.
44. Ghosh A, Rapp C, Friesner RA. Generalized Born model based on a surface area formulation. J Phys Chem B 1998;102:10983–10990.
45. Dominy B, Brooks CL, III. Development of a Generalized Born model parameterization for proteins and nucleic acids. J Phys Chem B 1999;103:3765–3773.
46. Lee M, Salsbury Jr, F, Brooks CL, III. Novel generalized Born methods. J Chem Phys 2002;116:10606–10614.
47. Wagner F, Simonson T. Implicit solvent models: combining an analytical formulation of continuum electrostatics with simple models of the hydrophobic effect. J Comput Chem 1999;20:322–335.
48. Onufriev A, Bashford D, Case D. Modification of the generalized Born model suitable for macromolecules. J Phys Chem B 2000; 104:3712–3720.
49. Simonson T, Carlsson J, Case DA. Proton binding to proteins: pK_a calculations with explicit and implicit solvent models. J Am Chem Soc 2004;126:4167–4180.
50. Lee M, Salsbury F, Jr., Brooks C, III. Constant pH molecular dynamics using continuous titration coordinates. Proteins 2004;56: 738–752.
51. Cornell W, Abseher R, Nilges M, Case D. Continuum solvent molecular dynamics study of flexibility in interleukin-8. J Mol Graph Model 2001;19:136–145.
52. Calimet N, Schaefer M, Simonson T. Protein molecular dynamics with the generalized Born/ACE solvent model. Proteins 2001; 45:144–158.
53. Majeux N, Scarsi M, Apostolakis J, Ehrhardt C, Caflisch A. Exhaustive docking of molecular fragments with electrostatic solvation. Proteins 1999;37:88–105.
54. Bursulaya B, Brooks C, III. Comparative study of the folding free energy landscape of a three-stranded β-sheet protein with explicit and implicit solvent models. J Phys Chem B 2001;104: 12378–12383.
55. Simmerling C, Strockbine B, Roitberg A. All-atom structure prediction and folding simulations of a stable protein. J Am Chem Soc 2002;124:11258–11259.
56. Rapp C, Friesner R. Prediction of loop geometry using a generalized Born model of solvation effects. Proteins 1999;35:173–183.
57. Felts A, Gallicchio E, Wallqvist A, Levy R. Distinguishing native conformations of proteins from decoys with an effective free energy estimator based on the OPLS all-atom force field and the surface generalized Born solvent model. Proteins 2002;48:404–422.

58. Eisenberg D, McClachlan A. Solvation energy in protein folding and binding. Nature 1986;319:199–203.
59. Ooi T, Oobatake M, Nemethy G, Scheraga H. Accessible surface areas as a measure of the thermodynamic hydration parameters of peptides. Proc Natl Acad Sci USA 1987;84:3086–3090.
60. Wesson L, Eisenberg D. Atomic solvation parameters applied to molecular dynamics of proteins in solution. Prot Sci 1992;1:227–235.
61. Fraternali F, van Gunsteren W. An efficient mean solvation force model for use in molecular dynamics simulations of proteins in aqueous solution. J Mol Biol 1996;256:939–948.
62. Ferrara P, Apostolakis J, Caflisch A. Evaluation of a fast implicit solvent model for molecular dynamics simulations. Proteins 2002;46:24–33.
63. Pei J, Wang Q, Zhou J, Lai L. Estimating protein-ligand binding free energy: atomic solvation parameters for partition coefficient and solvation free energy calculation. Proteins 2004; 57:661–664.
64. Dahiyat B, Mayo S. Protein design automation. Prot Sci 1996;5: 895–903.
65. Koehl P, Levitt M. Protein topology and stability define the space of allowed sequences. Proc Natl Acad Sci USA 2002;99: 1280–1285.
66. Ogata K, Jaramillo A, Cohen W, Briand J, Conan F, Wodak S. Automatic sequence design of MHC class-I binding peptides impairing CD8+ T cell recognition. J Biol Chem 2003;278:1281.
67. Liang S, Grishin N. Effective scoring function for protein sequence design. Proteins 2004;54:271–281.
68. Baker N. Poisson-Boltzmann methods for biomolecular electrostatics. Methods Enzym. 2004;383:94.
69. Dwyer J, Gittis A, Karp D, Lattman E, Spencer D, Stites W, Garcia-Moreno B. High apparent dielectric constants in the interior of a protein reflect water penetration. Biophys J 2000;79: 1610–1620.
70. Brooks B, Bruccoleri R, Olafson B, States D, Swaminathan S, Karplus M. Charmm: a program for macromolecular energy, minimization, and molecular dynamics calculations. J Comput Chem 1983;4:187–217.
71. Hawkins G, Cramer C, Truhlar D. J Phys Chem 1996;100:19824.
72. Cornell W, Cieplak P, Bayly C, Gould I, Merz K, Ferguson D, Spellmeyer D, Fox T, Caldwell J, Kollman P. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. J Am Chem Soc 1995;117:5179–5197.
73. Swanson J, Adcock S, McCammon J. Optimized radii for Poisson-Boltzmann calculations with the AMBER force field. J Chem Theory Comput 2005;1:484–493.
74. Guérois R, Nielsen J, Serrano L. Predicting changes in the stability of proteins and protein complexes: a study of more than 1000 mutations. J Mol Biol 2002;320:369–387.
75. Simonson T, Archontis G, Karplus M. Free energy simulations come of age: the protein–ligand recognition problem. Acc Chem Res 2002;35:430–437.
76. Lee B, Richards F. The interpretation of protein structures: estimation of static accessibility. J Mol Biol 1971;55:379–400.
77. Brünger AT. X-plor version 3.1, A system for X-ray crystallography and NMR. New Haven: Yale University Press; 1992.
78. Schaefer M, Bartels C, Leclerc F, Karplus M. Effective atom volumes for implicit solvent models: comparison between Voronoi volumes and minimum fluctuation volumes. J Comp Chem 2001; 22:1857–1879.
79. Moulinier L, Case D, Simonson T. X-ray structure refinement of proteins with the generalized Born solvent model. Acta Cryst D 2003;59:2094–2103.
80. Madura J, Briggs J, Wade R, Davis M, Luty B, Ilin A, Antosiewicz J, Gilson M, Baheri B, Scott L, McCammon J. Electrostatics and diffusion of molecules in solution: simulations with the University of Houston Brownian Dynamics program. Comput Phys Commun 1995;91:57–95.
81. Street A, Mayo S. Pairwise calculation of protein solvent-accessible surface areas. Folding Des 1998;3:253–258.
82. Yang J, Tsai C, Hwang M, Tsai H, Hwang J, Kao C. GEM: a Gaussian evolutionary method for predicting protein sidechain conformations. Prot Sci 2002;11:1897–1907.
83. Kumar M, Bava K, Gromiha M, Parabakaran P, Kitajima K, Uedaira H, Sarai A. ProTherm and ProNIT: thermodynamic databases for proteins and protein–nucleic acid interactions. Nucl Acids Res 2006;34:D204–206.
84. Sigalov G, Scheffel P, Onufriev A. Incorporating variable dielectric environments into the generalized born model. J Chem Phys 2005; 122:094511.
85. Simonson T. Free energy calculations: approximate methods for biological macromolecules. In Chipot C, Pohorille A, editors. Free energy calculations: theory and applications in chemistry and biology. New York: Springer Verlag; 2006, Ch. 12.
86. Gallicchio E, Kubo M, Levy R. Enthalpy-entropy and cavity decomposition of alkane hydration free energies: numerical results and implications for theories of hydrophobic hydration. J Phys Chem B 2000;104:6271–6285.
87. Archontis G, Simonson T. Proton binding to proteins: a free energy component analysis using a dielectric continuum model. Biophys J 2005;88:3888–3904.
88. Pace C, Scholtz J. A helix propensity scale based on experimental studies of peptides and proteins. Biophys J 1998;75:422–427.
89. Guex N, Peitsch M. SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. Electrophoresis 1997;18:2714–2723.

Acknowledgments

I'd like to express my gratitude to my teacher, professor Thomas Simonson, for the great comprehension and patience, during all these years.

I wish to thank our secretary, Catherine, for the necessary help and advice. I'd like to acknowledge all colleagues for creating a very good atmosphere in the lab, especially, Marcel, Guillaume, Anne, David and Najette, and to thank all our colleagues in the biochemistry department.

I wish to acknowledge financing. This work was initially funded by EGIDE for three years. The last year was supported by the the French National Research Agency (L'Agence nationale de la recherche, the ANR). Some of the simulations were performed at the CINES supercomputing center of the French Ministry of Education, though most of calculations were accomplished on the local cluster.

Exclusively, I would like to acknowledge our colloborators. Working in community is very interesting (especially with professionals and experts). I would like to acknowledge Georgios Archontis for his very useful advice, and ingenious ideas. I appreciate our experimental collaborator, professor Winfried Hinrichs, who supervised my work as well. It is very hard to find an expert, like Winfried, who deeply understands the topic in every detail.

Finally I would like to thank all members of jury, especially those who had to travel a long distance to come to Paris for my defense. Thank you for your interesting questions and discussions, and of course, for your very warm compliments.