95- 1798

LA-UR-

Title:

Three-Dimensional Model of a Selective
Theophylline-Binding RNA Molecule

Author(s):

Chang-Shung Tung, Tudor I. Oprea, Gerhard Hummer, and
Angel E. Garcia

Submitted to:

11th Int. Symp. on Affinity Chromatography and Biological
Recognition, Trinity University, San Antonio, Texas,
May 25-30, 1995.

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

Los Alamos
NATIONAL LABORATORY

Los Alamos National Laboratory, an affirmative action/equal opportunity employer, is operated by the University of California for the U.S. Department of Energy under contract W-7405-ENG-36. By acceptance of this article, the publisher recognizes that the U.S. Government retains a nonexclusive, royalty-free license to publish or reproduce the published form of this contribution, or to allow others to do so, for U.S. Government purposes. The Los Alamos National Laboratory requests that the publisher identify this article as work performed under the auspices of the U.S. Department of Energy.

Form No. 836 HS
ST 2629 10/91

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

MASTER

DISCLAIMER

Portions of this document may be illegible in electronic image products. Images are produced from the best available original document.

THREE-DIMENSIONAL MODEL OF A SELECTIVE THEOPHYLLINE-BINDING RNA MOLECULE

C.-S. Tung¹, T.I. Oprea¹, G. Hummer^{1,2}, and A.E. Garcia¹

¹Theoretical Biology and Biophysics (T-10), MS K710

²Center for Nonlinear Studies (CNLS)

Theoretical Division

Los Alamos National Laboratory

Los Alamos, NM 87545 U.S.A.

Abstract

We propose a three-dimensional (3D) model for an RNA molecule that selectively binds theophylline but not caffeine. This RNA, which was found using SELEX [Jenison, R.D., et al., Science (1994) 263:1425], is 10,000 times more specific for theophylline ($K_d = 320$ nM) than for caffeine ($K_d = 3.5$ mM), although the two ligands are identical except for a methyl group substituted at N7 (present only in caffeine). The binding affinity for ten xanthine-based ligands was used to derive a Comparative Molecular Field Analysis (CoMFA) model ($R^2 = 0.93$ for 3 components, with cross-validated R^2 of 0.73), using the SYBYL and GOLPE programs. A pharmacophoric map was generated to locate steric and electrostatic interactions between theophylline and the RNA binding site. This information was used to identify putative functional groups of the binding pocket and to generate distance constraints. Based on a model for the secondary structure (Jenison et al., idem), the 3D structure of this RNA was then generated using the following method: each helical region of the RNA molecule was treated as a rigid body; single-stranded loops with specific end-to-end distances were generated. The structures of RNA- xanthine complexes were studied using a modified Monte Carlo algorithm. The detailed structure of an RNA-ligand complex model, as well as possible explanations for the theophylline selectivity will be discussed.

Introduction

Large variability and flexibility of nucleic acid molecules make them ideal to design specific sequences for performing various tasks. The SELEX procedure (Tuerk and Gold, 1990) allows the selection of nucleic acid molecules that bind molecular targets with high affinity and specificity. For example, a selected RNA molecule that folds into a pseudoknot motif can serve as a ligand that inhibits the function of the human immunodeficiency virus type I reverse transcriptase (Tuerk et al., 1992). Another RNA molecule was selected as a ligand to basic fibroblast growth factor to inhibit receptor binding (Jellinek et al., 1993). Other RNA molecules were found that have high binding affinity to small molecular ligands such as ATP (Sassanfar and Szostak, 1993) and a variety of organic dyes (Ellington and

Szostak, 1990). Single-stranded DNA molecules that bind and inhibit human thrombin were also found using the SELEX procedure (Bock et al., 1992). SELEX was proposed as an alternative route for drug discovery, since this procedure is capable of selecting nucleic acids that recognize and bind to specific molecular targets (Jenison et al., 1994).

Recently, RNA molecules with high affinity for theophylline, but not caffeine, were obtained using the SELEX procedure. Theophylline, a naturally occurring methyl-xanthine, is currently used as bronchodilator in the treatment of airway-obstructive diseases. Chemical differentiation from other related xanthines, e.g. caffeine and theobromine, when monitoring blood levels of theophylline, is important for avoiding toxicity problems. All theophylline-binding RNA molecules share a similar secondary structure that includes a conserved CCU bulge and a 6 bases symmetric internal loop flanking a conserved 3 base pairs stem. Based on equilibrium filtration analysis with [^{14}C]theophylline, the binding affinities of these RNA molecules are similar to that observed for monoclonal antibodies raised against the particular antigen (Poncelet et al., 1990). Detailed physico-chemical and NMR studies were reported for one of these RNA molecules, labelled TCT8-4 (Jenison et al., 1994). Thermal denaturation studies showed that TCT8-4 forms a stable folded structure with a T_m of 72°C. The binding affinity of ten xanthine derivatives to TCT8-4 RNA was also established in competition studies using radiolabelled theophylline (see Table 1 in Jenison et al., 1994).

To understand the molecular basis of theophylline-RNA specificity, we examined the series of ten ligands binding to TCT8-4 and their physico-chemical properties using molecular modeling procedures. Based on pharmacophoric pattern matching, we developed a three-dimensional quantitative structure-activity relationship (3D-QSAR) of the ten xanthine derivatives using the Comparative Molecular Field Analysis (CoMFA) (Cramer III et al. 1988) method.

Based on a set of structural constraints generated using the pharmacophoric map and stereochemistry, we constructed a three-dimensional model in atomic detail for the RNA molecule (TCT8-4). A complete set of reduced coordinates developed in our laboratory was used for the modeling of nucleic acid structures (Soumpasis and Tung, 1988; Tung,

1993; Tung et al., 1994). This method has the advantage that every parameter in the set has intuitive physical meaning, therefore it is ideal for modeling nucleic acid molecules with limited tertiary structure information.

Methods

Structure-activity relationships

To determine the pharmacophoric pattern of theophylline we used the XED (Extended Electron Definition) procedure (Vinter 1994), which explicitly treats pi-electrons and lone pairs of polar groups in a molecular mechanics framework. After XED points are added to a given molecule, the local minima for the interaction with three probes (H⁺, HO⁻ and H₂O) are located, similar to the GRID (Goodford 1985) method. The geometric location of the H⁺ and HO⁻ minima was used to define a pharmacophoric map for theophylline, as presented in Figure 1.

Structural models of the xanthine derivatives were calculated using the AM1 Hamiltonian, as implemented in Mopac 6.0 (Stewart 1990), with full geometry optimization (keywords: AM1 EF). For these optimized molecules, the free energy of solvation was estimated using the AM1-SM2 method (Cramer and Truhlar 1992) (keywords AM1 SM2 1SCF). Since purine, the scaffold of xanthine derivatives, can exist in two tautomeric structures (proton can migrate from N7 to N9 - see Table 1), N7-N9 tautomers were theoretically investigated for these compounds (except 1,3 dimethyl-uric acid, 7-methylxanthine, theobromine and caffeine). Mopac AM1 and Amsol AM1-SM2 heats of formation (HoF) were used to compare the stability of these compounds, in conjunction with calculated and measured LogP, the octanol-water partition coefficient (Leo, 1993) values obtained from Daylight CIS (Daylight 1994).

The relationship between chemical structure and binding affinity of xanthine derivatives to TCT8-4 was studied using two 3D-QSAR programs, CoMFA and GOLPE. A brief description of these methods is included. The CoMFA (Cramer III, Patterson et al. 1988)

method postulates that observed variations in experimental activity among different compounds may be explained with calculated or measured physico-chemical properties, and are dependent on the compounds' intermolecular interactions with the binding site. In CoMFA, these non-covalent interactions are represented as the steric (Lennard-Jones) and electrostatic (Coulombic) energies of a ligand with a grid of regularly spaced probe atoms. Both steric and electrostatic energy are tabulated for each molecule at every grid point. The resulting matrix is statistically analyzed using Partial Least Squares (PLS) (Wold, Johansson et al. 1993). CoMFA fields are thus related to the binding affinity. This procedure emphasizes receptor characteristics that were probed by the initial data set.

The following settings were used for CoMFA: 1.5 Å spaced grid (18 x 15 x 10 Å); probe atom: Csp3; dielectric constant of 1.0; a +/-20 kcal/mol cutoff value with no electrostatic interactions at steric overlap points (i.e., electrostatic field values are not considered at grid points overlapping with ligand atoms). The regression analyses were performed using the Sybyl (SYBYLTM) implementation (Cramer III et al. 1988) of the PLS algorithm (Wold et al. 1993), with cross-validation (leave-one-out model), and three principal components. A 0.4 kcal/mol energy column filter was applied.

The predictive ability of this model was evaluated by calculating Q^2 :

$$Q^2 = 1 - \frac{PRESS}{SD}, \quad [1]$$

$$PRESS = \sum (Y_p - Y_a)^2, \quad [2]$$

$$SD = \sum (Y_a - Y_m)^2, \quad [3]$$

where *PRESS* is the sum of squared deviations between predicted (Y_p) and actual (Y_a) activities, and *SD* is the sum of the squared deviations between actual activities and the mean (Y_m) biological activity value for all molecules. The best predictive model yields a Q^2 of 1, but values can go to zero (or below) for models that predict values equal to, or worse than Y_m .

GOLPE (Baroni et al. 1993) (Generating Optimal Linear PLS Estimations) was used in conjunction with CoMFA to further refine the PLS model using variable selection procedure according to fractional factorial design (Cruciani et al. 1993). The internal predictivity of the xanthine model system was determined by leave-three-out cross-validation. In this procedure, the data set is randomly divided in three equal groups, one set being excluded from the model and predicted. The process was repeated 25 times, using recalculation of weights. Results were also compared with the leave-one-out cross-validation. The distribution of both steric and electrostatic fields was studied in GOLPE, and a 0.4 standard deviation (SDEV) filter was applied to each field. The steric/electrostatic weighting ratio, automatically suggested in CoMFA, was applied: 1.0 for the steric, and 0.55 for the electrostatic field.

RNA Folding

From the predicted secondary structure, the whole RNA molecule can be divided into two motifs (helices and loops). To simplify the problem, all helices in the molecule were modeled as rigid bodies assuming a canonical A-conformation (Arnott 1972). Six parameters (three rotational and three translational) are required to describe the position and orientation of each helix in space with respect to one another. By positioning and orienting in space all helices, the lengths of all the interconnecting loops are determined. To model the structures of these interconnecting loops, we have developed an algorithm that is capable of generating single-stranded loops with fixed end-to-end distances.

If $\vec{S}_n$ indicates the vector for C5'-O3' of the n^{th} sugar and $\vec{P}_n$ indicates the vector for O3'-C5' between the n^{th} and the $(n+1)^{th}$ sugars, the span of the three bases loop can be represented as seven connecting vectors, shown in Fig. 2. This seven vectors group can be further decomposed into two long vectors ($\vec{AD}$ and $\vec{EH}$) and one short vector ($\vec{DE}$). The end-to-end distance (d) of a three connecting bonds is related to the torsional angle (α) of the three bonds according to the following equation (Tung, 1993):

$$d^2 = l_1^2 \sin^2 \theta_1 + l_3^2 \sin^2 \theta_2 + (l_1 \cos \theta_1 + l_3 \cos \theta_2 - l_2)^2 - 2l_1 l_3 \sin \theta_1 \sin \theta_2 \cos \alpha. \quad [4]$$

where l_1, l_2, l_3 are the three bond lengths and θ_1, θ_2 are the two bond angles. For a specific end-to-end distance, a torsional angle (α) can be found that satisfies the condition as specified in equation [4]. For example, the structure of the ADEH segment can be completely defined using three lengths (l_{AD}, l_{DE}, l_{EH}), two angles ($\theta_{ADE}, \theta_{DEH}$) and one torsional angle (ϕ) corresponding to the rotation of the ADEH segment with respect to the end-to-end pseudobond (A-H). In a similar way, the two short segments (ABCD and EFGH) can each be described by a set of six parameters. A total of 18 parameters are required to completely describe the structure of the entire segment for a single-stranded three bases loop.

Once the structure of the ABCDEFG segment is defined, every base, sugar and phosphate group in the loop can be inserted with one rotational degree of freedom. In this representation, a set of 28 parameters completely describes the structure of a three-bases loop.

To address the complexity of the system, a modified Metropolis Monte Carlo simulation algorithm has been developed. The conventional Metropolis Monte Carlo simulation (Metropolis et al., 1953) constitutes the core of our method. The AMBER force field (Weiner et al., 1984) was used to calculate conformational energies (E_c) of the drug-RNA complex. A quadratic energy term (E_d) was added to the system to account for the set of structural constraints derived from the pharmacophoric map. The force constant (k_d) for these constraints was arbitrarily chosen to be $200 \text{ Kcal/mol} \cdot \text{\AA}^2$ (the same order of magnitude as that of a covalent bond). The structure of the complex is allowed to move in conformational space on the set of reduced coordinates for the complex as previously described. The average step sizes are $0.01 \text{ \AA}$ for translational parameters and 1° for rotational parameters, such that the acceptance rate is close to 50 per cent.

The first part of our modified MC simulation involves the generation of random structures of the drug-RNA complex with the set of reduced coordinates. Those structures with a total energy below the cutoff value, 10^5 Kcal/mol , were selected for 5,000 steps of regular MC simulation. This modified MC simulation provides a better sampling in the conformational space with less chance of being trapped in the subspace close to the

starting conformation. A typical run involves the selection of 100 random structures each equilibrated with 5,000 steps of regular MC simulation.

Force Field Parameters for Theophylline and Caffeine

The structures for theophylline and caffeine were optimized using *ab initio* quantum mechanical calculations. Bond lengths and bond angles were optimized using STO-3G basis functions. These were used as equilibrium values in the force field calculations. For additional quantum mechanical calculations we used 6-31G* basis functions on the STO-3G optimized structures. We obtained a set of atom-centered partial charges by doing a least square fitting of the electrostatic potential, as discussed in the literature (Besler et al. 1990). We evaluated the electrostatic potential (ESP) on four layers of points describing Connolly surfaces as implemented in Gaussian 92 (Frish et al. 1992). We performed least-square fitting of the ESP after imposing equal charges on hydrogens of the same methyl groups. Partial charges change within 10%, by varying the number of Connolly surface layers and the density of points. The fitted potential with only four surface layers give a better fitting of the ESP at close range. Off-atom-centered point charges were also considered, but the advantage of more complicated fitting was not obvious. Bond lengths, bond angles, Lennard Jones and 10-12 hydrogen bond force constants are similar to those of nucleic acids in the all atom force field of Weiner, Case and Kollman (Weiner et al, 1986). The force field parameters for theophylline and caffeine are listed in Table II.

Secondary structure

The 3-dimensional fold of any RNA model depends on the presumed secondary structure. The sequence specificity of RNA molecules that bind theophylline is influenced by two factors: (i) changes in the binding pocket and loss of favorable interactions and (ii) drastic changes in the secondary structure. To monitor these factors, we determined the effect of single base substitutions in the unpaired regions of TCT8-4 by studying changes in the secondary structure (Zuker, 1989).

Atomic Model Refinement

Selected structures for the drug-RNA complexes were subject to further refinement via energy minimization using AMBER (Pearlman et al., 1991). No cutoff was applied for the non-bonded interactions. The electrostatic interactions were screened with a dielectric constant of 80 corresponding to bulk water. The geometric parameters and charges of the theophylline were determined as described in the previous section. Further refinement of the intramolecular potentials of theophylline (e.g., fitting of vibrational spectra) were considered unnecessary for this study, where the intermolecular interactions of RNA and theophylline are dominant.

Results and Discussion

Structure-activity relationships

Prior to modeling the pharmacophoric pattern and the structure- affinity relationship, we have obtained local minima for each structure and examined the tautomers of susceptible compounds. CP- theophylline has two conformational minima (using AM1): at -91.68 and -91.57 kcal/mol, respectively. The carboxy-propyl (CP) linker is perpendicular to the purine scaffold, above and below the plane. The calculated free energy of solvation, ΔG_{SOL} (using AM1- SM2), is equal for both conformers, since they have similar energies in water: -115.85 and -115.72 kcal/mol, respectively. Thus, both conformers of the linker can be important in the binding process and should be considered in the modeling of TCT8-4 RNA.

The N9 tautomers of theophylline, CP-theophylline and 3- Methylxanthine are not energetically relevant, either in gas phase (AM1) or in water (AM1-SM2) (see table 3). The N9 tautomers of xanthine and 1-methylxanthine seem to gain importance in aqueous phase as opposed to gas phase: ΔH_oF of N7 compounds is 3 kcal/mol higher in gas phase, but only 2 kcal/mol higher in water. Since the N9 tautomer's ΔG_{SOL} is better than that of the N7 ones, it is possible that tautomeric structures become relevant for these two compounds, but with a smaller molar fraction compared to N7 tautomers. Calculated LogP values are identical for both N7 and N9 tautomers, and not significantly different

from the measured values (see Table 3). The N9 tautomer of hypoxanthine has a lower HoF than the N7 tautomer, both in gas phase (ΔHoF 1 kcal/mol) and in solvent (FHoF 2.45 kcal/mol). The latter suggests that N9 hypoxanthine has a higher molar fraction in aqueous solution. Both tautomers were included in the CoMFA study. However, the 3D-QSAR models were not able to distinguish between the two structures.

For hypoxanthine, the N3-N7 and N3-N9 tautomers were also examined (Table 3). These tautomers are unlikely to occur, since the calculated LogP values (-0.59 in both cases) are significantly different than the measured one (-1.11). For these tautomers, the ΔHoF in both gas and aqueous phase is higher than 1 *RT*.

The pharmacophoric map of theophylline (Fig. 1) was determined using the XED procedure. Three hydroxide minima were observed, at less than $3\text{\AA}$, in coplanarity with the bicyclic compound. One minimum, considered energetically relevant (39.5 kcal/mol), is situated between N7 and C8, somewhat closer to the former. This minimum is not located directly above the N7 hydrogen due to the slightly acidic character of the C8 hydrogen. Two other minima, located above the N1 and N3 methyl groups at more than $2.7\text{\AA}$, were not included in the constraint map due to the weak nature of their interaction with hydroxide (6.5-7.5 kcal/mol). These minima are a consequence of the increased polarity of the N-substituted CH₃ groups. Among the six proton minima observed, the strongest interaction is at N9 (-28.5 kcal/mol), which is the site for tautomeric shifts. Both carbonyl oxygens exhibit strong proton affinity: two minima at O2 are located above the oxygen's lone pairs (-22.5 kcal), whereas one is located above O6 (-19.5 kcal). The other O6 lone pair does not have a minimum due to the proximity of the N7 hydrogen. Two other minima (-23.9 kcal/mol) are located above and below the bicyclic plane, between C5 and C6, and are caused by the pi-electron cloud. These minima were included in the constraint map, after merging the positions of the O2 proton minima into a single location.

Prior to CoMFA studies, the binding affinity of xanthine derivatives to TCT8-4 was correlated with other physico-chemical properties, such as LogP, pK_a and ΔG_{sol} . There was no correlation between LogP and pK_D, excluding hydrophobicity as a driving force in the binding process. This important result is supported by the polar nature of both the

binding site (RNA molecules do not have non-polar cavities) and of the ligands (except for the methyl groups, the ligands are polar molecules). No correlation was observed between pK_a and pK_D . However, ΔG_{sol} was weakly correlated to pK_D ($R^2 = 0.512$, standard error of estimate 1.054). This correlation, while not robust, points out that soluble compounds have higher affinity for TCT8-4. It should be noted that, except for caffeine, the lowest ΔG_{sol} is observed at the 3 compounds mentioned (Jenison et al. 1994) to have poor solubility: 7-methylxanthine, theobromine and 1,3-dimethyl uric acid.

The lack of correlation between substance descriptors (e.g., LogP , ΔG_{sol}) and pK_D warranted the use of the CoMFA method (based on atomic-level descriptors). The model system of 10 xanthine derivatives was analysed with the CoMFA-GOLPE procedure. Using Principal Component Analysis (Wold et al. 1987), we found that 3 principal components explain 63.14% of the variance. The initial CoMFA run, using 2112 variables, yields the following 3-component model: $R^2 = 0.956$, standard error of estimate 0.281, $Q^2 = 0.341$ and standard error of prediction 1.09 (using the leave-one-out procedure). This model, while suggesting common trends in the series, lacks robustness. Therefore, we submitted this model to a variable selection procedure using the Fractional Factorial Design strategy (implemented in GOLPE). The model retained 48 out of 2112 variables, and the 3-component model explains 87.1% of the variance. This model yields improved robustness ($R^2 = 0.932$, standard error of estimate 0.352, $Q^2 = 0.731$ and standard error of prediction 0.699 with leave-one-out cross-validation). The increase in internal consistency is also reflected in the internal predictivity, as determined by leave-three-out cross-validation: $Q^2 = 0.598$, standard error of prediction 0.855. The predicted activities of this latter model are listed in Table 1. While these do not reflect the best predictive model (leave-one-out values are slightly better), they are likely to reflect better the predictive ability of the model, in the absence of an external (test) set (Oprea and García 1995). Out of 10 compounds, two are predicted with an error higher than 1 log unit (or 0.7 kcal/mol in terms of binding energy): CP-theophylline and 7-methylxanthine. The first one is the only compound to have a CP substituted at N1, which makes it a chemical outlier, whereas for 7-methylxanthine, the measured pK_D is just an estimate, due to the lack of solubility of this compound (Jenison et al. 1994).

Graphical analysis of CoMFA fields revealed the following trends: a favorable electrostatic interaction, located around N7-C8-N9, which is likely to be occupied by negative charges in the binding site. This accounts for the poor affinity of 1,3-dimethyl uric acid. Areas of reduced steric bulk tolerance are located around N7, accounting for the poor binding affinity of 7 methyl substituted compounds (theobromine, 7-methylxanthine and caffeine). Areas of beneficial steric interaction are located around N3 (explaining the overpredicted affinity of 3-methylxanthine) and considerably less at N1 (theophylline is underpredicted by this model). Since no chemical variation occurs at C6 and C2 (except for hypoxanthine), CoMFA fields do not reveal the importance of electrostatic interactions at these two positions. This does not constitute a problem since these atoms were used in the superimposition of the ten compounds (and are part of the pharmacophore).

RNA Folding

The binding between theophylline and TCT8-4 is specific, hence the intercalating mode of binding in the duplex region of the RNA molecule is ruled out. The possibility that the UCU hairpin loop is involved in the binding site was ruled out because of the lack of sequence homology and length in this loop. The two possibilities for the theophylline binding site in TCT8-4 are: (1) the totally conserved CCU bulge and (2) the highly conserved 6 bases symmetric internal loop. Based on proton nuclear magnetic resonance (NMR) spectra (Jenison et al., 1994), the stoichiometry of binding is one theophylline per RNA molecule. We favor the notion of a unique binding site for theophylline in the RNA molecule. The CCU bulge is completely homologous among the set of RNA molecules that bind theophylline. In addition, the appearance of two imino proton resonances at 11 and 15 ppm upon binding of theophylline was interpreted as a G forming a nonstandard base pair and a C protonated at the imino group (Janison et al., 1994). Both interpretations are consistent with the collapse of the 6 bases symmetric bulge into a stem region when theophylline is bound to the RNA. Therefore, we conclude that the CCU bulge is part of the binding site.

Other parts of the RNA molecule interacting directly with the drug, or in close proximity of the binding site include a three base-pair helix (r(CCA)-r(UGG)) and a six base-pair helix

(r(GCAUCG)-r(UGAUGC)). The CCU loop and the two flanking helices constitute the part of the RNA molecule that was modeled in detail. Using the reduced set of coordinates described in the Methods section, 40 parameters are required for the description of the drug- RNA complex (28 parameters for the CCU loop, 6 parameters for the six base pairs helix and 6 parameters for theophylline).

We have studied the end-to-end distances for all three-base spans in the crystal structure of tRNA^{phe} (TRNA04 in PDB). We found that the average end-to-end distance is 16Å, ranging of 10Å to 20Å. We used this average end-to-end distance (16Å) as a constraint for the three-base loop. The optimum distance for phosphate group (3.5Å for O3'-C5' of two neighboring sugars) was used as another constraint. These allowed us to position and orient the two helices with respect to each other. Many different arrangements for the two helices satisfy both the imposed constraints as shown in Fig. 3. This indicates that the RNA molecule is highly flexible at the CCU bulge region in the unbound state. This observation is consistent with the notion that binding of theophylline stabilizes a conformation that is otherwise flexible. Structural constraints deduced from the pharmacophoric map were used as a guideline for modeling the drug-RNA complex.

Fig. 4 shows four different possible H-bond patterns between theophylline and the CCU loop. The structure of the four different binding modes was modeled with the modified MC algorithm. The result of a typical run is shown in Fig. 5. The initial conformation, selected from a pool of random structure is shown on the left. The equilibrated structure (final conformation) is shown on the right. The equilibrated structures of different runs were ranked according to their total energies (the conformational energy plus the constraint energy). The lowest energy structure of a particular set of runs was selected for that specific binding mode. The structures of the four different binding modes are plotted in Fig. 6.

Atomic Model Refinement

The structures corresponding to the four different binding modes were subject to further refinement using AMBER (Pearlman et al., 1991). Based on 5000 steps of unconstrained minimization, structure II showed the lowest energy and was chosen for further analysis.

This choice was also motivated by the agreement with previously identified constraints. In particular, structure II has the 1-methyl group pointing outwards such that the CP linker can be used to connect CP-theophylline to the selection column in the SELEX procedure. In this conformation, the CP linker is pointing above the bicyclic ring. In addition, the N-7 position is buried in the binding pocket and involved in a hydrogen bond. These results are consistent with the negative selection against caffeine, theobromine and 7-methylxanthine, which have the N-7 hydrogen replaced by a methyl group. This methyl group causes unfavorable steric and electrostatic interactions if the binding shows similar overall structural properties.

In the TCT8-4 sequence, the 5'-CG end of the RNA is closed by a UCU loop (Jenison et al., 1994). During unconstrained minimization, the base pair at the end tends to fold into the major groove. However, this hardly affects the structural properties of the binding pocket. Nevertheless, to avoid these problems we constrained the terminal base pair. Another problem we encountered is that two of the three hydrogen bonds between theophylline and RNA bases are partly lost or substituted by hydrogen bonds with RNA-backbone atoms. This is attributed to the significantly reduced electrostatic interactions when using a physically-motivated dielectric constant 80. For dielectric constant 1, charge interactions contribute considerably to hydrogen-bond energies. Therefore, we have applied distance constraints for the hydrogen bonds between theophylline and the RNA bases.

Fig. 7 shows a space-filling representation of the structure after 50,000 steps of minimization (rms-deviation of the forces $0.006 \text{ Kcal mol}^{-1} \text{ \AA}^{-1}$). Theophylline is tightly packed into the binding pocket. It is stacked with an adjacent guanine of the long helix. On the N3-C4-N9 side, theophylline is packed against an adenine of the short helix. Fig. 4 shows the hydrogen-bond interactions of O2, O6 and H7 with the CCU bulge. Also shown is the connecting ribose-phosphate backbone. The hydrogen H7 is interacting with a cytosine. If caffeine, 7-methylxanthine or 3,7-dimethylxanthine are to replace theophylline, no space will be available for the methyl group at the 7 position. This is consistent with the drop in binding affinity observed for these compounds (see Table 1). Also, the positively-charged hydrogen H8 is interacting with a nitrogen of a base-paired cytosine (not shown), leaving no

room for an oxygen at the 8 position, as in 1,3-dimethyl uric acid. This is consistent with the electrostatic CoMFA fields (previously described) and may explain the lack of binding affinity for this ligand. Thus, the energy-minimized structure of the drug-RNA complex can explain the weak binding of derivatives with large substitutes at the 7 and 8 positions under the assumption of similar binding modes. The two methyl groups are facing the opening and are partly exposed to solvent. Addition of the CP linker (used in the SELEX procedure) to the C1 carbon is possible without large structural changes. The minimized structure is therefore in agreement with the important pharmacophoric and SELEX-related constraints.

Conclusion

Results from different molecular modeling techniques converged in obtaining the 3D structural model of an RNA molecule in the absence of extensive structural information. Small-molecule methods such as pharmacophoric mapping and structure-activity relationships of known ligands were used to provide structural constraints for RNA folding. The putative binding site for theophylline was identified by analyzing the available RNA sequences using secondary structure prediction methods and known experimental data. These preliminary results were instrumental in providing the structural details used to construct the 3D structure. Atomic refinement of the structural model was obtained by tailoring force-field parameters for theophylline. The final model of the TCT8-4 RNA molecule is in agreement with experimental observation, and may provide further insight on the molecular mechanisms of RNA-theophylline specificity.

Acknowledgement

We thank Dr. Martijn Huynen for discussions on RNA secondary structure prediction. Dr. David Weininger is acknowledged for permission of using the Daylight software and access to the MedChem94 database. Prof. Sergio Clementi (Perugia, Italy) and Tripos Inc. (St. Louis, MO) have provided us with software. This work was supported by U.S. Department of Energy.

Reference

- Baroni, M. et al. (1993) *Quant. Struct.- Act. Relat.* **12**, 9-20.
- Bock, L. C. et al. (1992) *Nature* **355**, 564-566.
- Besler, B.H., Merz, K.M., and Kollman, P.A. (1990) *J. Comp. Chem.* , **11**, 431-439.
- Cramer, C. and Truhlar, D. (1992) *J. Comput. Aided Mol. Des.* **6**, 629-666.
- Cramer III, R. D. et al. (1988) *Quant. Struct.- Act. Relat.* **7**, 18-25.
- Cramer III, R. D. et al. (1988) *J. Am. Chem. Soc.* **110**, 5959-5967.
- Cruciani, G. et al. (1993). "3D-QSAR in Drug Design. " H. Kubinyi (Ed.), 551- 564, ESCOM, Leiden.
- Daylight programs and the MedChem94 database are available from Daylight Chemical Information Systems, 18500 Von Karman Ave Suite 450, Irvine, CA 92715 (1994)
- Ellington, A. D. and Szostak, J. W. (1990) *Nature* **346**, 818-822.
- Frissh, M.J. et al. (1992) Gaussian92 (Gaussian, Inc., Pittsburgh, PA).
- Goodford, P. J. (1985) *J. Am. Chem. Soc.* **28**, 849-856.
- Jaeger, J.A.,etal., (1989) *Methods in Enzymology*, **183**, 281-306.
- Jellinek, D. et al. (1993) *Proc. Natl. Acad. Sci. USA* **90**, 11227- 11231.
- Jenison, R. D. et al. (1994) *Science* **263**, 1425-1429.
- Leo, AJ (1993) *Chemical Reviews* **93**, 1281-1306.
- Metropolis, N. et al. (1953) *J. Chem. Phys.* **21**, 1087-
- Oprea, T. I. and A. E. Garcia (1995) *J. Med. Chem.* submitted.
- Pearlman, D. A. et al. (1991) *Amber 4.0* Technical report, University of California, San Francisco.
- Sassanfar, M. and Szostak, J. W. (1993) *Nature* **364**, 550- 553.
- Stewart, J. J. P. (1990) *J. Comput. Aided Mol. Design* **4**, 1- 105.
- SYBYLTM and CoMFA are available from Tripos, Inc., 1699 S Hanley Rd, Suite 303, St. Louis, MO 63144. (1994)
- Tuerk, C. and Gold, L. (1990) *Science* **249**, 505-510.
- Tuerk, C. et al. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 6988-6992.
- Soumpasis, D. M. and Tung, C.-S. (1988) *J. Biomol. Struct. and Dyn.* **6**, 397-420.

Tung, C.-S. (1993) "*Computation of Biomolecular Structures*" D. M. Spoumpasis and T. M. Jovin (Eds) 87-97, Springer-Verlag, NY.

Tung, C.-S. et al. (1994) *J. Biomol. Struct. and Dyn.* **11**, 1327- 1344.

Vinter, J. (1994) *J. Comput. Aided Mol. Des.* **8**, 653-668.

Weiner, S. J. et al. (1984) *J. Am. Chem. Soc.* **106**, 765- 784.

Weiner, S. J. et al. (1986) *J. Comput. Chem.* **7**, 230-252.

Wold, S. et al. (1987) *Chemometrics and Intelligent Laboratory Systems* **2**, 37-52.

Wold, S. et al. (1993) "*3D-QSAR in Drug Design.* " H. Kubinyi (Ed.), 523-550, ESCOM, Leiden.

Zuker, M. (1989) *Science*, **244**, 48-52.

Table 1. Chemical structures, calculated and experimental properties of xanthine derivatives used in the competitive assay with TCT8-4RNA. For each property, m is measured, c is calculated, and p is predicted. See explanations in text.

Compound	R ₁	R ₂	R ₃	R ₄	R ₅	LogP(m)	pKa(m)	ΔG _{SOL} (c)	pK _D (m)	pK _D (p)
Theophylline	CH ₃	CH ₃	H	H	---	-0.02	8.81	-44.315	0.495	-0.377
CP-theophylline	propionate	CH ₃	H	H	---	-0.12	---	-24.165	0.032	-1.024
Xanthine	H	H	H	H	---	-0.73	7.53	-22.384	-0.930	-0.775
1-Methylxanthine	CH ₃	H	H	H	---	-0.27	7.70	-20.982	-0.954	-0.954
3-Methylxanthine	H	CH ₃	H	H	---	-0.72	8.10	-20.705	-0.301	0.404
7-Methylxanthine	H	H	CH ₃	H	---	-0.89	8.33	-18.441	-2.778	-3.814
Theobromine	H	CH ₃	CH ₃	H	---	-0.78	10.00	-16.554	-2.778	-2.898
1,3-Dimethyluric acid	CH ₃	CH ₃	H	O	H	-0.52	---	-19.337	-3.041	-3.452
Hypoxanthine ^a	H	---	H	H	---	-1.11	1.91	-23.797	-1.690	-1.196
Caffeine	CH ₃	CH ₃	CH ₃	H	---	-0.07	0.60	-15.680	-3.544	-2.544

^aHypoxanthine has no oxygen substituted at C2.

Table 2. Force -field parameters for theophylline and caffeine.

Bond parameters:

Bond	K_b	l_0
H -N4	440.00	1.01
N4-C4	440.00	1.38
N4-C7	520.00	1.40
C4-H4	340.00	1.08
C4-N5	520.00	1.33
N5-C6	350.00	1.40
C6-N6	350.00	1.40
C6-C7	520.00	1.35
N6-C8	340.00	1.47
N6-C5	440.00	1.43
C8-H8	340.00	1.09
C5-O5	570.00	1.22
C5-C7	440.00	1.48
N6-C4	440.00	1.38
N6-C7	520.00	1.40

Angle parameters:

Angle	K_a	θ_0
H -N4-C4	70.00	127.2
H -N4-C7	70.00	126.9
N4-C4-H4	35.00	121.7
N4-C4-N5	70.00	113.4
N4-C7-C6	70.00	105.4
N4-C7-C5	70.00	131.8
C4-N5-C6	70.00	102.6
C4-N4-C7	70.00	105.9
H4-C4-N5	35.00	124.9
N5-C6-N6	70.00	127.5
N5-C6-C7	70.00	112.7
C6-N6-C8	70.00	117.7
C6-N6-C5	70.00	123.9
C6-C7-C5	70.00	122.8
N6-C8-H8	70.00	109.5
N6-C5-O5	80.00	121.7
N6-C5-N6	70.00	114.8
N6-C6-C7	70.00	119.9
C8-N6-C5	70.00	118.3
H8-C8-H8	35.00	109.5
C5-N6-C5	70.00	123.9
N6-C5-C7	70.00	114.7
O5-C5-C7	80.00	123.6
C4-N6-C7	70.00	105.2
C4-N6-C8	70.00	127.6
7-N6-C8	70.00	127.2
N6-C7-C6	70.000	105.7
N6-C4-H4	35.00	121.2
N6-C4-N5	70.00	114.0
N6-C7-C5	70.00	131.6

Table 2. continued

Dihedral angle parameters:

Dihedral	N	V ₀	f ₀	n
X -N4-C7-X	4	6.60	180.00	2.00
X -N4-C4-X	4	6.70	180.00	2.00
X -C4-N5-X	2	20.00	180.00	2.00
X -C6-C7-X	4	16.30	180.00	2.00
X -C5-C7-X	4	4.40	180.00	2.00
X -N5-C6-X	2	5.10	180.00	2.00
X -C6-N6-X	4	6.60	180.00	2.00
X -N6-C5-X	4	6.60	180.00	2.00
X -N6-C8-X	6	0.00	0.00	3.00
X -X -N4-H		1.0	180.00	2.00
X -X -C4-H4		0.0	180.00	2.00
X -X -N6-C8		0.0	180.00	2.00
X -X -C5-O5		10.5	180.00	2.00

Hydrogen bond parameters:

Atom Pair	A	B
H N5	7557.0	2385.0
H2 N5	7557.0	2385.0
H3 N5	4019.0	1409.0
HO N5	7557.0	2385.0
HS N5	14184.0	3082.0
HW N5	7557.0	2385.0

Lennard-Jones parameters:

Atom	R	ε
N	1.85	0.09
C	1.85	0.12
O	1.6	0.20

Table 2. continued

Partial Charges (in esu):

Atom	Theophylline	Caffeine
N1	-0.5319	-0.2899
CN1	-0.1794	-0.2398
HM1	0.1231	0.1285
C2	0.9673	0.7519
O2	-0.4432	-0.5580
N3	-0.7823	-0.4728
CN3	0.0936	-0.1226
HM3	0.0807	0.1029
C4	0.6902	0.7898
O4	-0.6128	-0.5480
C5	-0.1845	-0.4364
C6	0.6337	0.4883
HM7	---	0.1710
CN7/H7	0.4137	0.4261
N7	-0.5025	0.0855
H8	0.2974	0.2317
N9	-0.6139	-0.5961

Force field parameters are defined by the energy functions $E_b = K_b (l - l_0)^2$, $E_{ba} = K_a (\theta - \theta_0)^2$, $E_\phi = V_0 / N (1 - \cos(n\phi - \phi_0))$ and $E_{HB} = A/r^{12} - B/r^{10}$ for bond lengths, bond angles and dihedral angles and hydrogen bonds respectively. Parameters are defined as in Weiner et al, 1986.

Table 3. Calculated properties for N7 (left) and N9 (right) tautomers of xanthine derivatives.

Compound	R ₁	R ₂	LogP(m)	ClogP	HoF(g)	HoF(s)	ΔG _{SOL} (c)
Theophylline-N7	CH ₃	CH ₃	-0.02	-0.06	4.781	-39.534	-44.315
Theophylline-N9	CH ₃	CH ₃	---	-0.06	8.245	-12.057	-20.303
CP-theophylline-N7	propionate	CH ₃	-0.12	-0.99	-91.683	-115.848	-24.165
CP-theophylline-N9	propionate	CH ₃	---	-0.99	-87.908	-112.860	-24.952
Xanthine-N7	H	H	-0.73	-0.70	-8.316	-30.700	-22.384
Xanthine-N9	H	H	---	-0.70	-4.113	-28.671	-24.557
1-Methylxanthine-N7	CH ₃	H	-0.27	-0.34	-1.448	-22.430	-20.982
1-Methylxanthine-N9	CH ₃	H	---	-0.34	2.094	-20.176	-23.079
3-Methylxanthine-N7	H	CH ₃	-0.72	-0.69	-2.203	-22.908	-20.705
3-Methylxanthine-N9	H	CH ₃	---	-0.69	1.053	-20.298	-21.351
Hypoxanthine-N7	H	---	-1.11	-1.20	44.453	20.456	-23.797
Hypoxanthine-N9	H	---	---	-1.20	43.326	17.992	-25.334
Hypoxanthine-N3	---	H	---	-0.59	50.389	23.146	-27.243
Hypoxanthine-N3N9	---	H	---	-0.59	57.309	26.456	-30.853

Note: Four possible tautomers of hypoxanthine have been investigated. CLogP is the estimated LogP calculated using PCModules (from Daylight CIS). HoF(g) and HoF(s) are the AM1 (gas phase) and AM1-SM2 (solvent) heats of formation.

Figure Legends

Figure 1. The pharmacophoric pattern of theophylline, obtained using the XED procedure.

Figure 2. Schematic representation of a three bases single-stranded loop. Seven vectors were used to represent the structure of the loop such that it can easily fit to a specified end-to-end distance. In this reduced representation, 28 parameters are required to completely describe the structure of the three bases loop.

Figure 3. Different arrangements of the two helices in TCT8-4. The three base-pair helix is fixed in space as shown in the lower part of the figure. The six base-pair helix is allowed to sample different positions and orientations such that the two connecting distances (between the two helices) are constrained to $3.5\text{\AA}$ (the length of a phosphate group) and $16.6\text{\AA}$ (the averaged length of a three bases single-stranded loop) respectively.

Figure 4. Four H-bonding patterns between theophylline and the CCU loop. Based on the pharmacophoric map, the three strong H-bond donors/acceptors on theophylline are O2, O6 and N9. These three were chosen to form H-bonds with the CCU loop.

Figure 5. Structure of the CCU loop, the two flanking helices and the theophylline. The left image shows a selected random structure of the group while the right image shows the structure after 5,000 cycles of Monte Carlo simulation. The MC simulation not only positioned the theophylline inside the binding cavity but also reduced the cavity size to improve the interaction between the drug and the RNA molecule.

Figure 6. Four different mode of binding between the theophylline and the RNA molecule. In this figure, theophylline is plotted as CPK model while the RNA is plotted as stick model. For clarity, only residues closed to the binding pocket were shown. Each of the four corresponds to the lowest energy MC equilibrated structure out of a pool of 100 selected random structures.

Figure 7. Space-filling representation of theophylline (black) bound to RNA (gray) after energy minimization. The labeled bases are part of the CCU bulge. The C1-O2-C3 face is partly exposed to solvent.

Figure 8. Interactions of theophylline (black) with the CCU bulge and a base-paired cytosine in the energy-minimized structure. Bases are shown in gray, the ribose-phosphate backbone connecting the CCU bulge is shown in dark gray.

Fig 7

Fig. 3

NAMDT2

NAMOT2

(I)

(II)

(III)

(IV)

NAMOT2

Fig. 6.