Journal of Biomolecular NMR, 8 (1996) 261–272
ESCOM

J-Bio NMR 381

A transferred NOE study of a tricyclic analog of acyclovir bound to thymidine kinase

Jerzy Czaplicki[a,*], Thomas Bohner[b], Anne-Kathrin Habermann[c], Gerd Folkers[b] and Alain Milon[a]

[a]*Institut de Pharmacologie et de Biologie Structurale, CNRS, 118, route de Narbonne, F-31062 Toulouse, France*
[b]*Departement Pharmazie, ETH Zürich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland*
[c]*Institut für Pharmazie, Universität Leipzig, Brüderstrasse 34, D-04103 Leipzig, Germany*

Received 8 May 1996
Accepted 9 July 1996

Keywords: Acyclovir analogs; Thymidine kinase; Transferred NOE; Relaxation matrix

Summary

A purine derivative with an acyclic sugar analog, 3,9-dihydro-3-[(2-hydroxyethoxy)methyl]-6-ethyl-9-oxo-5*H*-imidazo[1,2-*a*]purine, was studied in the free state and in complex with herpes simplex virus thymidine kinase (HSV1 TK). Transferred NOE experiments, combined with a full relaxation matrix analysis of the substrate's spin system, resulted in a set of distance constraints for all proton pairs. These constraints were used in structure determination procedures based on simulated annealing and molecular dynamics simulations to obtain a family of structures compatible with the experimental NMR data. The results indicate that, although in both states the chains have the syn orientation with respect to the aromatic rings, in the free state the substrate's acyclic moiety is relatively disordered, while in the bound state only one specific conformation is preferred. Fluctuations can only be seen in the case of the terminal hydroxyl group, for which no NOE was recorded and hence no constraints were available.

Introduction

Thymidine kinase (TK) catalyzes the transfer of the γ-phosphate group from ATP to the 5'-OH group of thymidine in the presence of magnesium ions (ATP-thymidine-5'-phosphotransferase, EC 2.7.1.21). TK is a highly conserved enzyme, which occurs in all major organisms, with the notable exception of fungi. Its physiological role is to salvage thymidine into the DNA metabolism by converting it to thymidine monophosphate.

In eucaryotic cells at least two different forms of TK have been identified. Mammalian cytosolic TK is only present in the dividing cell in G1 to S stages, whereas the mitochondrial form, which is synthesized in the cytoplasm but predominantly found in mitochondria, is expressed constitutively.

Herpes viruses (HV) encode their own TK. Although both human TK isoforms differ in their substrate specificity (Eriksson et al., 1991), they are both more selective than HV TK; for example, they are not able to phosphorylate purine derivatives (Munch-Petersen et al., 1991). HV TK, however, accepts purine substrate analogs, even those carrying an acyclic sugar moiety such as acyclovir or ganciclovir (Keller and Elion, 1981). Several approaches in cancer therapy and the treatment of virus infections are based on this difference in substrate specificity.

Herpes simplex virus thymidine kinase (HSV1 TK) is the selective target for most antiviral drugs. In HV-infected cells, antivirals like acyclovir are converted to their monophosphates almost exclusively by the HV TK (Fyfe et al., 1978; Keller and Elion, 1981) and then primarily, if not exclusively, by cellular enzymes to the triphosphates (Miller and Miller, 1980), which are the active metabolites and serve as fraudulent nucleotides in nascent DNA strains and potent inhibitors of HV DNA polymerase (Reardon and Spector, 1989).

One main type of systemic gene therapy, the so-called VDEPT (virally directed enzyme prodrug therapy), is based on TK (Huber et al., 1991; Ohno et al., 1994). The gene for HSV1 TK, packaged in replication-defective retroviruses or adenoviruses, is introduced to tumor cells followed by administration of ganciclovir. The HV TK

*To whom correspondence should be addressed.

does not harm the proliferating cells directly, but it renders them susceptible to ganciclovir, a prodrug that is also harmless by itself but is transformed by the HV TK into a cytotoxic metabolite.

For the rational development of drugs, the interaction processes between the drug and its biological targets should be known. A prerequisite for this knowledge is the description of the structural properties of the drugs as well as those of the interacting biological sites. However, until now neither X-ray diffraction nor NMR spectroscopy have been used to study the TKs. A three-dimensional model of the active site of HSV1 TK has been created by the knowledge-based molecular design technique (Folkers et al., 1991). The model predicts that the enzyme consists of a central core β-sheet structure, connected by loops and α-helices that are very similar to the overall structure of other nucleotide-binding enzymes. The phosphate binding site contains a highly conserved glycine-rich loop at the N-terminus of the protein (amino acids 55–62 in HSV1 TK). The thymidine binding site is predicted to be located near this phosphate binding loop where the highly conserved FDRH sequence motif (amino acids 161–164) was suggested to form the bottom of the binding pocket.

Experimental evidence for the predicted active-site structure of HSV1 TK was provided by conformational and epitope mapping using synthetic peptide segments (Zimmermann et al., 1991). Site-directed mutagenesis as a further tool for experimental verification confirmed the importance of residues Phe^{161} and Asp^{162} for substrate binding (Black and Loeb, 1993; Fetzer et al., 1993), but suggested a rearrangement of the 3D model concerning the interchange of substrate and co-substrate binding site in relation to the phosphate binding loop (Michael et al., 1994).

Since the natural and synthetic ligands of TK are flexible molecules, prediction of the ligand binding site is particularly difficult, and would be made significantly easier by the knowledge of the bound structure of the ligand. In favorable cases this structure can be determined in NMR by the transferred NOE technique (Balaram et al., 1972; Clore and Gronenborn, 1982; Campbell and Sykes, 1991). One of the main conditions for the feasibility of the experiment is that the dissociation rate of the ligand is fast enough (10–100 s^{-1}). We have shown previously that thymidine itself does not produce observable transferred NOEs (Czaplicki et al., 1995), due to a dissociation constant that is too low ($k_D < 10^{-6}$ M). Acyclovir produced negative NOEs, indicating that transfer between bound and free states did take place ($k_D \geq 10^{-4}$ M), but the scarcity of the data precluded a detailed analysis of the structure. As a result of the search for a substrate with a dissociation rate at least matching that of acyclovir, but containing a higher number of observable protons and being more branched, we focused our attention on 3,9-dihydro-3-[(2-hydroxyethoxy)methyl]-6-ethyl-9-oxo-5*H*-imidazo[1,2-*a*]purine, a tricyclic analog of acyclovir (Borysk et al., 1991), which we shall call tricycle for brevity. In this work we report the results of transferred NOE experiments on tricycle, combined with a full relaxation matrix analysis and molecular dynamics simulations. Families of three-dimensional structures characterizing tricycle bound to TK have thus been obtained, all of which are compatible with the NMR data.

Materials and Methods

Thrombin, glutathione agarose, ATP agarose and other chemicals were purchased from Sigma (Buchs, Switzerland); [methyl-^{14}C] thymidine (50–62 mCi/mmol) was obtained from Amersham (Zürich, Switzerland). The plasmid pBR322-TK, containing the gene for HSV1 TK, was a gift from S. McKnight and the plasmid pGEX2T was purchased from Pharmacia (Dubendorf, Switzerland).

Expression and purification of HSV1 TK

The bacterial expression vector pGEX2T-TK was constructed as described earlier (Fetzer and Folkers, 1992). Purification was carried out by a two-step procedure described previously (Fetzer et al., 1994). Competent cells of the TK-deficient *E. coli* strain KY895 were transformed with pGEX2T-TK. Gene expression was induced by the addition of IPTG (isopropyl β-D-thiogalactopyranoside). HSV1 TK was expressed in *E. coli* as inducible thrombin cleavable fusion with glutathione S-transferase (GST). Bacteria were harvested by centrifugation, frozen, thawed, resuspended and lysed by mild sonication in the presence of lysozyme, PMSF (phenylmethylsulfonyl fluoride), DTT (DL-dithiothreitol) and Triton X-100. The supernatant containing large quantities of soluble GST-TK fusion protein was applied to glutathione affinity chromatography. The fusion protein was subsequently cleaved with thrombin directly on the column. The resultant flow-through was pre-incubated with thymidine and then applied to an ATP affinity column in order to separate HSV1 TK from thrombin. As a result, homogeneously pure recombinant enzyme was obtained (Fetzer et al., 1994).

The activity of HSV1 TK was determined by measuring the conversion of radiolabeled dT into phosphorylated product. Substrate and product were either separated by thin-layer chromatography on PEI-cellulose (polyethyleneimine cellulose) plates according to Fitt and coworkers (1976) or by ion exchange chromatography on DEAE-cellulose paper squares, first described by Furlong (1963).

A reaction mixture containing 50 mM Tris, pH 7.2, 5 mM ATP, 5 mM $MgCl_2$, 0.5 mg/ml BSA, 10 mM ^{14}C-thymidine and 250 ng enzyme in a total volume of 50 µl was incubated at 37 °C. Samples of 5 µl were removed every 10 min, boiled at 95 °C for 3 min and centrifuged.

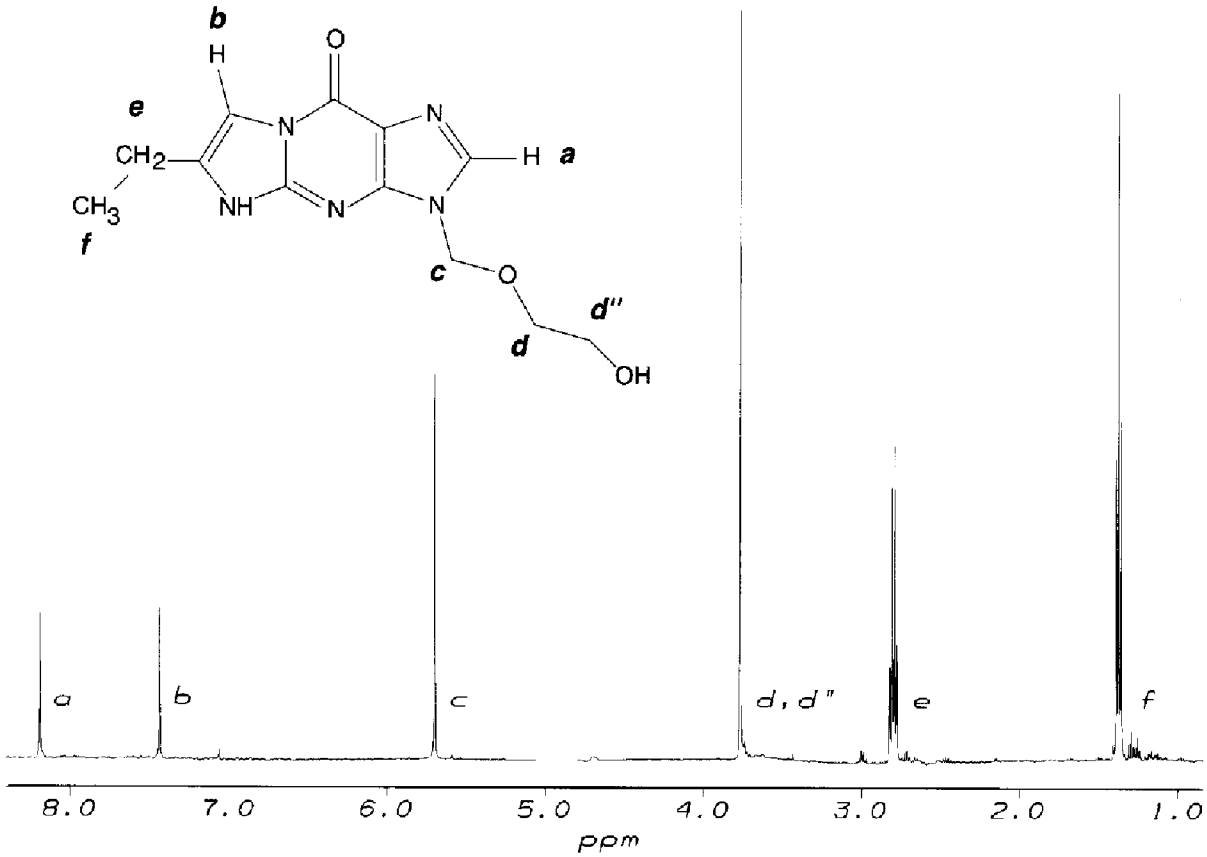

Fig. 1. One-dimensional ^1H spectrum of tricycle at 307 K. The solvent signal has been removed by convolution. The inset shows the assignments of all observable protons.

The supernatant was spotted onto PEI-cellulose plates, which had been prespotted with unlabeled thymidine, dTMP and dTDP. The plates were successively developed in 0.2 M LiCl for 2 min, 1 M LiCl for 6 min and 1.6 M LiCl for 22 min. The plates were either evaluated by autoradiography or by cutting out product and reactant spots, followed by determination of the radioactivity by scintillation counting.

NMR sample preparation

To determine the parameters characterizing the structure of tricycle in the free state, a 500 µl sample of 6 mM aqueous solution of the substrate containing 10% D_2O was used (using 100% D_2O resulted in a gradual decrease of signal intensities from aromatic protons over a few weeks, indicating their slow exchange). The optimal conditions for the observation of the transferred NOE signals have been determined previously (Czaplicki et al., 1995). The ratio of the ligand/enzyme molar concentrations was 100 :1. The NMR samples were prepared by mixing 170 µl of 0.12 mM enzyme solution containing 1 mM DTT, 50 mM glutathione, 0.0125% Triton X-100 at pH = 8.3, and 330 µl of the 6 mM solution of the ligand. All measurements were performed at 34 °C, which is close to physiological conditions. However, the temperature range in which the signals could be acquired is rather limited, as below 30 °C the exchange between the free and bound states becomes too slow to permit the observation of the transferred NOE effects, while above 37 °C the protein precipitates within one day, usually before the series of measurements is completed. Under the conditions specified above it was possible to acquire a complete set of data using the same sample throughout the whole of a single experiment (2 days) without any signs of its degradation.

Data acquisition and processing

All measurements were performed on a Bruker AMX-500 spectrometer. One-dimensional NOE difference spectra were acquired after selective inversion of individual substrate peaks by 180°, followed by variable mixing times. To achieve a selective inversion, a low-power CW decoupler pulse was applied at the peak offset frequency for 20 ms (which corresponds to a field strength $B_1 = 6$ mG). The relaxation delay was set to 8 s and the spectra were recorded on 8K data points over a spectral width of 5 kHz with 512 scans per FID. The solvent signal was continuously presaturated during the relaxation delay. Reference spectra with the irradiation frequency set to the edge of the spectral window were acquired at the beginning and end of the sequence, as well as each time the mixing time was changed. This protocol allowed not only

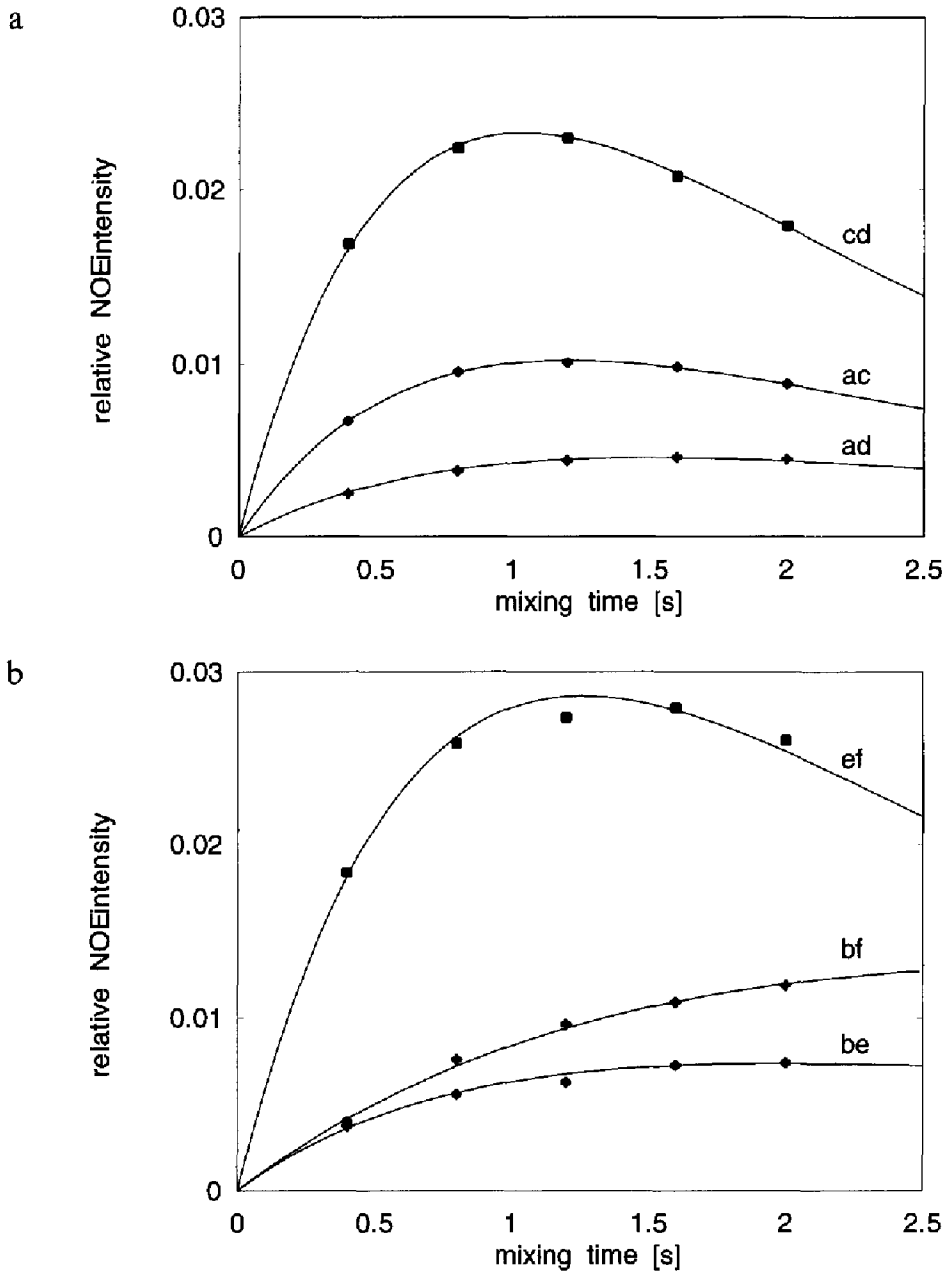

Fig. 2. Buildup curves for tricycle proton pairs in solution. Experimental points are depicted with symbols labeled in agreement with the proton assignments of Fig. 1, while the solid curves show the results of fitting (for details see text). For clarity, results for the internuclear vectors **ac**, **ad**, **cd** and **be**, **bf**, **ef** are shown separately in (a) and (b), respectively. Measured amplitudes correspond to the difference between the peak intensities in irradiated and reference spectra, divided by the intensity of an unperturbed single proton.

for averaging of the reference peak intensities, but also for monitoring eventual changes in the spectra. For each mixing time value, all observable groups of protons in tricycle were consecutively irradiated. In the free state, five mixing times were used: 0.4, 0.8, 1.2, 1.6 and 2 s, while in the case of the ligand/enzyme complex, data for eight mixing times were obtained: 0.2, 0.4, 0.6, 0.9, 1.2, 1.5, 2 and 3 s. Thus, the buildup curves were sampled throughout the whole of their time evolution, including the T_1 relaxation and spin-diffusion effects, which proved to be essential for the full relaxation matrix analysis.

Data were processed on an X32 station with the UXNMR program. Peak integrals were obtained with the Felix module of Biosym's NMRchitect software (Biosym, 1993). The resulting NOE areas were used as input data to the program developed in our laboratory, which reconstructs the relaxation matrix of a multistate and multicomponent spin system taking into account internal dynamics, spin equivalence and overlap. Although the multistate capabilities of the program were not utilized fully in this work, as the number of states for the tricycle/TK complex did not exceed two, we needed its ability to

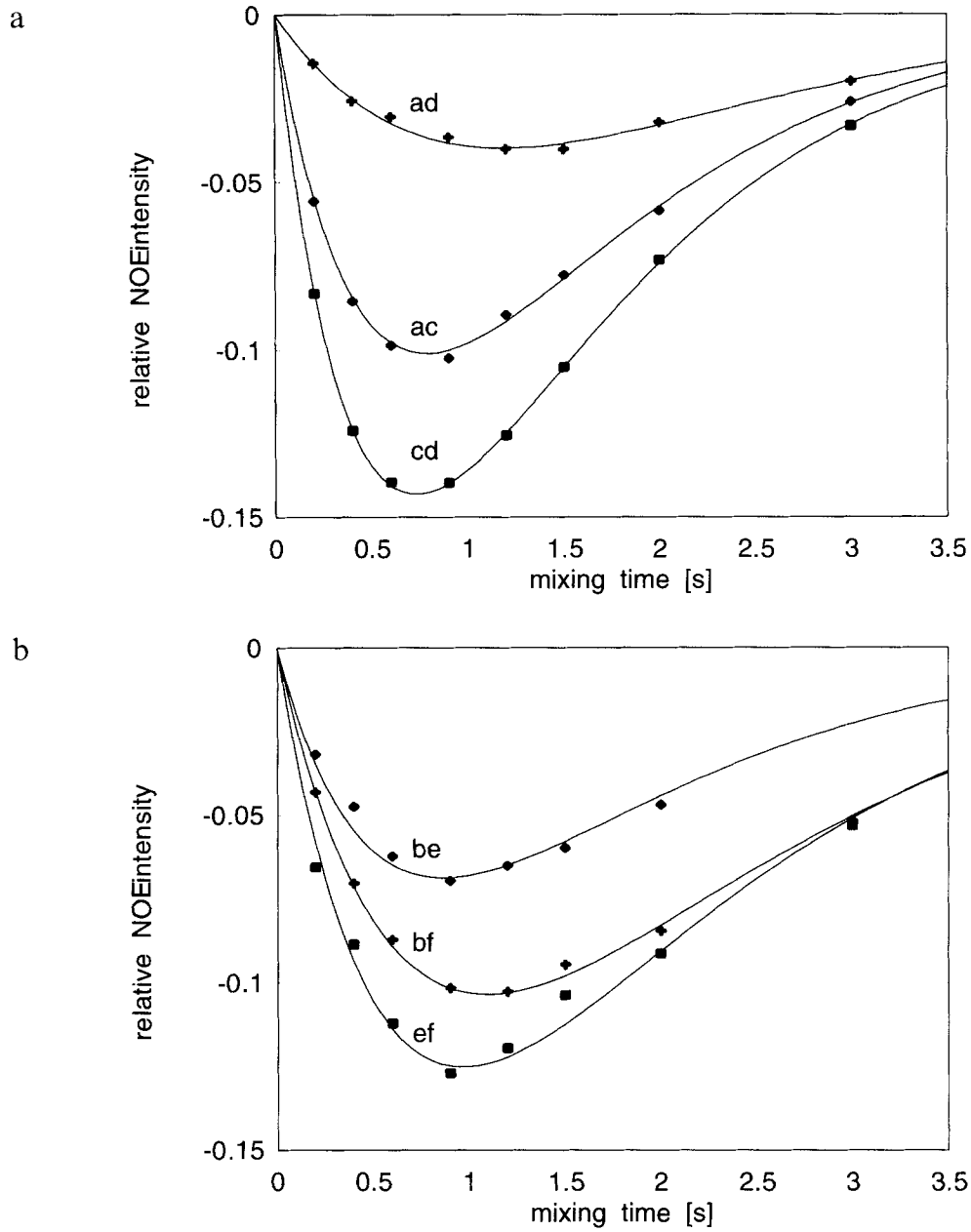

Fig. 3. Buildup curves for tricycle proton pairs in complex with TK. All details are identical to those described in the legend of Fig. 2, except for the signs of the NOEs, which are now negative.

handle overlapping groups of equivalent spins. The description of the program, called ANAGRAM (*A*nalysis of *N*OE *A*mplitudes by the *G*eneralized *R*el*A*xation *M*atrix Approach), will be published elsewhere (Czaplicki and Milon, 1996). The reconstructed relaxation matrix was used to derive distance constraints for the tricycle protons in the bound state. These constraints were in turn used by NMRchitect (v. 2.3) to create a family of structures compatible with the NMR data. The software was run on an Indigo Elan R3000 station. Standard protocols were used (Nilges et al., 1988) based on simulated annealing and molecular dynamics simulations. Families of structures were obtained by repeating the basic building blocks of the algorithm. Atomic coordinates were randomized and subjected to a quick initial energy minimization, followed by molecular dynamics simulations with the CFF91 force field in vacuum at 1000 K for 10 ps with a time step of 1 fs, during which the force constant of the distance constraints was gradually increased up to its full value of 25 kcal/mol/Å^2. During the next 10 ps the other force constants (covalent and nonbonded) were increased to their default values. Throughout the last 10 ps of the simulation, the temperature was decreased gradually to 307 K. The resulting structure was subjected to energy minimization. The above procedure was repeated a hundred times for both free and bound tricycle. Only those structures

that did not violate NOE distance constraints were used in further analysis.

Results and Discussion

Tricycle, whose one-dimensional NMR spectrum is shown in Fig. 1, is a molecule causing numerous problems in data analysis. Different mobility of the aliphatic chain with respect to the aromatic rings necessitates the use of site-specific effective correlation times. Lack of resolved signals from protons of methylene groups indicates their fast motion and hence averaging of interactions with their neighbors. Moreover, groups of protons appear at the same chemical shift, leading to spectral overlap and loss of coupling information. All of these effects, i.e., varying local mobility, averaging of interactions and overlap, have to be accounted for in the data analysis in order to obtain correct estimates of the internuclear distances.

In order to include the maximum of structural information in the acquired data, the experimental NOEs were measured over a vast range of mixing times, covering the initial buildup rate as well as the spin-diffusion effects. The results for tricycle in the free state are depicted in Fig. 2, while those for the tricycle/TK complex are shown in Fig. 3. The full set of 1D experimental data contains NOEs for proton pairs (i,j) and (j,i), which are equivalent to cross peaks in a 2D spectrum and hence should be identical (Neuhaus and Williamson, 1989). The numbers actually used in the fitting procedure (and shown in Figs. 2 and 3) resulted from averaging over these two cases, which differed as little as 3% for intense peaks at the maximum of their buildup, and as much as 35% in the case of weak peaks for very short or very long mixing times.

Relaxation matrix calculation in the free state

In the first stage of the present work, the relaxation matrix for tricycle in solution was reconstructed with the ANAGRAM program (J. Czaplicki and A. Milon, manuscript in preparation). The input data contained parameters permitting to generate diagonal (ρ_{ii}) and nondiagonal (σ_{ij}) relaxation matrix elements. The following relations were used to describe interactions between two nonequivalent protons (Edmondson, 1992; Liu et al., 1992):

$$\rho_{ii} = \sum_{j \neq i} \rho_{ij} + \rho_{ii}^{*} \qquad (1)$$

where the dipolar relaxation rates ρ_{ij} are given by:

$$\rho_{ij} = \frac{\gamma_i^2 \gamma_j^2 \hbar^2}{10 r_{ij}^6} [J_0(\omega) + 3J_1(\omega) + 6J_2(\omega)] \qquad (2)$$

and ρ_{ii}^{*} represent the relaxation leakage terms, i.e., interactions with other spins not included explicitly in the relaxation matrix, or nondipolar relaxation sources. Off-diagonal elements are computed as:

$$\sigma_{ij} = \frac{\gamma_i^2 \gamma_j^2 \hbar^2}{10 r_{ij}^6} [6J_2(\omega) - J_0(\omega)] \qquad (3)$$

The spectral density functions $J_n(\omega)$ are defined as:

$$J_n(\omega) = \frac{\tau_{ij}}{1 + (n\omega\tau_{ij})^2} \qquad (4)$$

where τ_{ij} are rotational correlation times characterizing the mobility of the interproton vectors r_{ij}, which were estimated on the basis of the T_1 and T_2 relaxation measurements. The other symbols have their usual meaning. External relaxation leakage terms ρ_{ii}^{*} were initially set to zero. The size of the relaxation matrix was 13×13 and was determined by the number of observable protons. The elements of the relaxation matrix corresponding to interactions between individual protons and methyl groups were calculated on the basis of the averaged spectral density functions (Tropp, 1980), of the form:

$$J_n(\omega) = \frac{4\pi}{5} \frac{\tau_{ij}}{1 + (n\omega\tau_{ij})^2} \sum_{m=-2}^{2} \left| \frac{1}{N} \sum_{i=1}^{N} \frac{Y_{2m}(\theta_{ij}, \varphi_{ij})}{r_{ij}^3} \right|^2 \qquad (5)$$

where N is the number of orientations (equivalent sites) of the methyl protons, and Y_{2m} are the spherical harmonics of the second order, describing relative orientations of protons i and j. This leads to the so-called $\langle r^{-3} \rangle^2$ averaging, justified in the case where the proton reorientations are faster than the overall tumbling motion of the molecule. In the case of interactions with unresolved protons of the methylene groups, the averaging of the relaxation matrix elements was accomplished by calculating an $\langle r^{-6} \rangle$ average, applicable when the motion rate is slower than the global rotation of the molecule:

$$J(\omega) = \frac{1}{N} \sum_{i=1}^{N} J_i(\omega) \qquad (6)$$

Finally, the spectral density functions describing intramethyl proton interactions are given by a combination of two terms:

$$J_n(\omega) = \left(\frac{1}{4}\right) \frac{\tau_{ij}}{1 + (n\omega\tau_{ij})^2} + \left(\frac{3}{4}\right) \frac{\tau}{1 + (n\omega\tau)^2} \qquad (7)$$

$$\frac{1}{\tau} = \frac{1}{\tau_{ij}} + \frac{1}{\tau_m} \qquad (8)$$

where τ_m is the correlation time characterizing methyl group rotations.

Several different algorithms for fitting were tried (Zangwill, 1969), such as conjugate directions, conjugate gradients and hybrid matrix. The most robust one was found to be the hybrid matrix method, and consequently this method was used throughout the rest of the work. Its only disadvantage is the necessity to use a starting structure, which may affect the results. In principle, the final structure is independent of the starting one when the complete set of NOEs is known, but in practice, because of spin equivalence and overlap, this is usually not the case. Hence, a set of initial structures of the ligand was generated by the NMRchitect software (Biosym, 1993) by minimizing the energy of the molecule without any constraints and randomizing the conformation of the chains with respect to the aromatic rings. In the hybrid matrix algorithm the calculated relaxation matrix was diagonalized, time-evolved and merged with the experimental data, with careful redistribution of the experimental NOEs among equivalent and overlapping spins. This was followed by back-calculation and back-substitution of those elements that should be held constant throughout the calculation (i.e., the *intra*group relaxation rates, in order to avoid distorting the structures of methyl and methylene groups). At each stage of iteration the R_{NOE} factor was calculated to monitor the progress of the calculation. This factor was defined as follows (Edmondson, 1992):

$$R_{NOE} = \frac{\sum_i |I_{exp} - I_{calc}|}{\sum_i |I_{exp}|} \qquad (9)$$

where I_{exp} and I_{calc} correspond to the experimental and calculated intensities, respectively.

In a series of runs the initial τ_{ij} values were varied for each internuclear vector, one at a time, in search of the minimum of R_{NOE}. Since a change of one element of the relaxation matrix affects the relaxation pathways of all its neighbors, after all correlation times had been varied it was necessary to repeat this procedure iteratively until no further decrease in R_{NOE} was visible. Usually a few iterations of this type sufficed to obtain a good agreement with the experimental data for all mixing times. The variations of correlation times were accomplished by the method of small steps within a specified interval. This is a rather time-consuming and inefficient method, but all our attempts to algorithmize this procedure failed in that during the runs the variables were set to values that were far from optimum or to non-physical (negative) values. In both of these cases convergence in the calculations could not be obtained.

The question that arises when dealing with a large set of variables is that of the uniqueness of the solution. Two or more different sets of variables may in principle produce a similar final R_{NOE} factor and overall fitting quality. Although the number of variables for a small molecule like tricycle is not really large (all inter and intra correlation times plus the methyl group parameters amounted to 15 variables), all kinds of motional problems are already encountered here. Therefore we did not rely entirely upon the value of the R_{NOE} factor, regarding it as a guideline rather than a strict determinant of the fitting quality. After completion of each iterative process, buildup curves for all proton pairs in the molecule were generated from the newly obtained relaxation matrix elements and their agreement with experimental data was carefully inspected. In many cases when the parameters were close to, but not exactly at the minimum, the R_{NOE} factor was low (~ 0.05) but expressed an average trend of curves rather than their individual characteristics, which is not surprising in view of the way it is defined (James, 1994). Typically, of six curves fitted simultaneously (see Fig. 2), up to four were observed to be in good agreement with the NOE data, while the remaining two were clearly showing noticeable deviations. This situation could not be remedied by a simple change of parameters pertaining to the protons whose curve was badly fitted, because of the overall NOE interdependence, explicitly introduced in the data by measurements at long mixing times. In fact, all parameters had to be varied until agreement for all individual curves was attained. Thus, the spin-diffusion effects have permitted us to differentiate between the local and global minima throughout the fitting process.

Since the measured T_1 relaxation times were similar to those calculated on the basis of the reconstructed relaxation matrix elements, and on the other hand the experimental and generated buildup curves were coinciding close-

TABLE 1
STRUCTURAL AND DYNAMICAL PARAMETERS OF TRICYCLE IN THE FREE STATE

Proton pair (ij)	τ_{ij} (ps)	σ_{ij} (s^{-1})·10^3
aa	–	–
ac	325	11.8
ad	150	2.4
ad"	100	1.8
bb	–	–
be	290	5.8
bf	150	4.0
cc	32	–
cd	28	15.0
cd"	13	0.3
dd	24	–
d"d"	22	–
ee	35	–
ef	18	10.5
ff	13	–

The table lists parameters characterizing tricycle in solution as determined by fitting to experimental data. The internuclear rotational correlation times τ_{ij} and intragroup correlation times τ_{ii} were the independent variables. Free rotation of methyl groups was assumed (the number of proton jumping sites was set to 12 (Liu et al., 1992)). The spin labeling (ij) corresponds to that of Fig. 1.

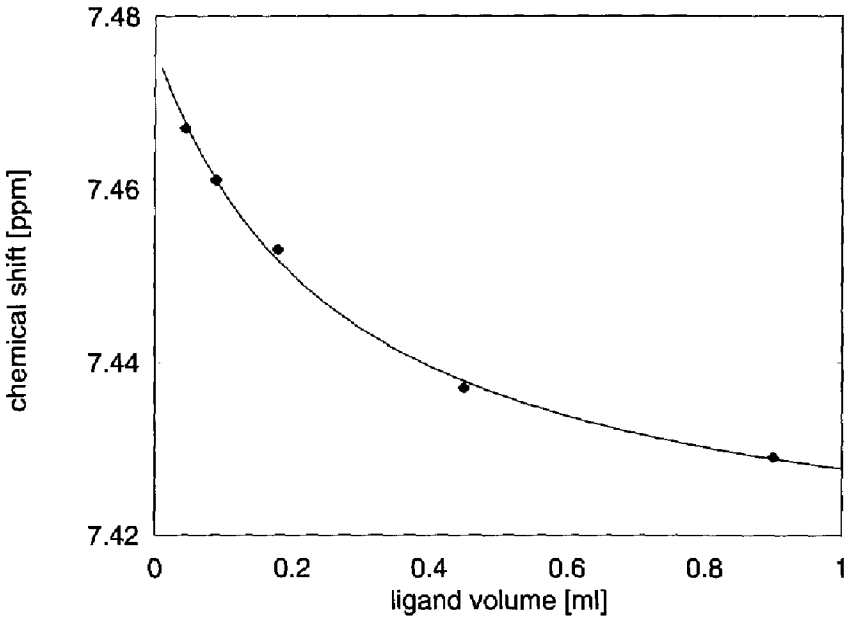

Fig. 4. Dependence of the chemical shift of proton b in tricycle (see Fig. 1) on the ligand volume in the sample. The solid line represents the results obtained from least-squares fitting. The values of the fixed parameters E_o, L_o and V_o are given in Materials and Methods. The curves measured for other protons were nearly identical. The average values of the fitted parameters are: dissociation constant $k_D \sim 10$ mM; chemical shift difference $|\delta_{free} - \delta_{bound}| \sim 2$ ppm.

ly (Fig. 2), it was not necessary to try to improve the fit by varying the external relaxation leakage terms ρ_{ii}^*. Apparently, the relaxation in the free state is dominated by the dipolar interactions. This is probably also true in the bound state, since the dipolar interaction is more efficient there, and the ρ_{ii}^* terms account for the dipolar relaxation induced by the protons of the protein (see below).

The above procedure was applied to different starting structures and identical results were obtained. The final R_{NOE} factor for tricycle in solution over all mixing times was 0.019, which can be seen in Fig. 2 as good agreement between the experimental points and calculated curves. Table 1 lists parameters obtained from fitting for tricycle in the free state.

Relaxation matrix in the bound state

Once the relaxation matrix of the ligand in the free state is known, it can be used in the determination of the relaxation matrix in the bound state by a direct substitution of its contents into the generalized relaxation matrix (London et al., 1992). Next, both submatrices are combined to produce an averaged matrix, determining the temporal evolution of NOE signals in the limit of fast ligand exchange between the free and bound states. To estimate the value of the dissociation constant k_D of the tricycle/TK complex we measured the chemical shifts of individual tricycle peaks as a function of the enzyme/ligand concentration ratio (Lian et al., 1994). A marked variation of δ_{obs} indicated the case of fast exchange, in which the observed value is an average of the chemical shifts in the free and bound states:

$$\delta_{obs} = p_f\delta_f + p_b\delta_b \tag{10}$$

where p_f and p_b are normalized spin fractions, corresponding to the free and bound states, respectively. The fraction of spins in the bound state is given by the ratio of the concentration of the complex to the total concentration of the ligand:

$$p_b = [EL]/[L] \tag{11}$$

Equations 10 and 11 can be combined with the definition of the dissociation constant k_D:

$$k_d = ([E]-[EL])([L]-[EL]) / [EL] \tag{12}$$

to obtain a nonlinear equation relating the observed chemical shift as a function of the total enzyme and ligand concentrations:

$$\delta_{obs} = \delta_f + \frac{\delta_b - \delta_f}{2[L]}\left\{([E]+[L]+k_D) - \sqrt{([E]+[L]+k_D)^2 - 4[E][L]}\right\} \tag{13}$$

Since [E] and [L] can be expressed in terms of the initial concentrations of enzyme E_o and ligand L_o by the relations:

$$[E] = E_0V_0 / (V_0+V_x) \tag{14}$$

and

$$[L] = L_0V_x / (V_0+V_x) \tag{15}$$

where V_o is the initial (and constant) volume of the enzyme, it follows that the observed chemical shift δ_{obs} is a function of only one parameter: the volume of the ligand V_x. An example of this dependence for proton b of tricycle is depicted in Fig. 4, along with the curve resulting from the least-squares minimization. The equilibrium constant k_D determined from the fit was found to be ~10 mM, which in the limit of the diffusion-controlled association constant of $10^7 \sim 10^8$ $M^{-1}s^{-1}$ translates to a k_{off} rate of $10^5 \sim 10^6$ s^{-1}, fully justifying the use of the fast exchange limit approximation.

Since upon addition of protein to the solution the viscosity is likely to change, affecting the correlation time of the ligand in the free state, we decided to vary the global rotational correlation time of the ligand while preserving the ratios of all individual correlation times in the molecule. Since the ratio of volumes of the ligand and enzyme used in preparing the sample was 2:1, no significant changes were expected. However, it seemed reasonable to include this possibility in the analysis. Although this step seemed well justified, we found that the best fit was obtained for virtually the same value of the correlation time as the one determined for the free state without the enzyme.

The procedure used to reconstruct the relaxation matrix of tricycle in the bound state was identical to the one described above. This time, however, it was necessary to vary the external relaxation leakage terms ρ_{ii}^*. Initially, they were all set to 0 and were modified whenever deemed necessary. Since the correlation time values affect mostly the initial slopes of the buildup curves and the influence of the ρ_{ii}^* terms is mainly visible when spin diffusion becomes significant, it was possible to separate the contributions from these two sets of parameters while fitting. The proceedings in this case were very similar to those given above, including the way to recognize local minima. The only notable difference with respect to the determination of the relaxation matrix in the free state was the fact that convergence could not be obtained for all starting structures used. For some of them the final values of the R_{NOE} factor were rather high, while for some others there was no convergence at all. This inability of the hybrid matrix method to reconstruct the final relaxation matrix elements from an arbitrary starting structure is typical for cases where incomplete NOE data are available. Most likely the convergence could be obtained upon inclusion of some bound-state NOEs in the analysis, but a rather low bound-state concentration (1%) and large line widths (representative of a TK macromolecule) precluded direct measurements in the bound state.

Since it is likely that in the bound state at least some of the observable ligand protons will be close to the protein protons, the question of protein-mediated spin-diffusion transfers comes into play. In a series of simulations we have shown (J. Czaplicki and A. Milon, manuscript in preparation) that when the ligand approaches the protein, so that some of the newly created (intermolecular) interproton distances are shorter than those within the ligand, it is impossible to recreate the corresponding buildup curves in their entirety by varying the available parameters (τ_{ij} and ρ_{ii}^*). When protein protons are close enough to the ligand to become dominant sources of relaxation, the deviations between measured and calculated curves are systematic and only the explicit inclusion of protein protons in the relaxation matrix can cure the problem. On the other hand, when the distances between ligand and protein protons are larger than those within a ligand, one can approximate experimental buildup curves with the above-mentioned parameters. The deviations, albeit still present, are negligible (in our particular case they translate into distance errors of less than 2%).

It must be noted, however, that it is possible to deduce the presence of additional protons only from an analysis of a complete buildup curve. Attempts to use the initial slope approximation will inevitably lead to incorrect distance estimates, with errors depending on the geometry of the interacting spin system.

The results of the fitting in the case of tricycle complexed with TK are shown in Fig. 3. The final value of the global R_{NOE} factor was 0.026. The values of variable parameters are presented in Table 2. The correlation times τ_{ij} for nearly all internuclear vectors are practically identical (125–150 ns), showing that the molecule is rigid.

TABLE 2
STRUCTURAL AND DYNAMICAL PARAMETERS OF TRICYCLE IN THE BOUND STATE

Proton pair (ij)	τ_{ij} (ns)	σ_{ij} (s^{-1})	r_{ij} (Å)	ρ_{ii}^* (s^{-1})
aa	–	–	–	2
ac	135	–18.61	2.8	–
ad	150	–3.38	3.6	–
ad"	150	–0.85	5.9	–
bb	–	–	–	39
be	150	–10.98	3.1	–
bf	50	–8.24	3.2	–
cc	135	–	–	0
cd	135	–18.52	2.9	–
cd"	135	–1.73	4.2	–
dd	135	–	–	23
d"d"	135	–	–	23
ee	125	–	–	0
ef	45	–6.87	3.2	–
ff	30	–	–	7

The table lists numerical values of variable parameters as determined for the tricycle/TK complex. Methyl group rotations were approximated by assuming a three-site model with the time constant τ_m characterizing the internal rotations of the methyl protons of 120 ps. Note that τ_{ij} is given in nanoseconds now, as opposed to the picoseconds listed in Table 1. Distances correspond to the best fit values, with maximum error $\Delta r_{ij} = \pm 0.1$ Å. Correlation time values are nearly identical (125–150 ns), showing that the molecule is rigid. The only exception is the mobile methyl group f (30–50 ns), for which the given distances represent the upper bound limits.

The only exception is the methyl group, whose mobility affects nearby protons. The given proton–methyl correlation times (30–50 ns) refer to the vector connecting a proton with the geometric center of the methyl group. While the rigidity of the molecule permits assignment of well-determined distances to most of the proton pairs, we know a priori that the distances involving methyl groups are not correct. This is because we do not know the geometric factors given by spherical harmonics in Eq. 5. Being a combination of trigonometric functions, their absolute value is <1, hence the apparent distance from the center of the methyl group calculated from Eq. 3 represents an upper limit estimate rather than a real distance. We are currently working on improving the accuracy of determination of distances involving methyl groups.

Distance error analysis

On the basis of the averaged relaxation matrix **R** it is possible to generate NOE matrices for all mixing times for which experimental data were acquired and to compare the two sets. The differences between the individual NOE intensities in the calculated and experimental spectra can be used to create the error matrix $\Delta\mathbf{R}$, following the relation:

$$\Delta\mathbf{R} = \frac{1}{\tau_m}\ln\left(\mathbf{A}_{calc}\mathbf{A}_{exp}^{-1}\right) \qquad (16)$$

where $\mathbf{A}_{calc}$ and $\mathbf{A}_{exp}$ represent the calculated and experimental matrices, respectively, corresponding to a given value of the mixing time τ_m. The above relation is valid in the case when all the elements of the $\Delta\mathbf{R}$ matrix are smaller than the corresponding elements of the **R** matrix. From the set of $\Delta\mathbf{R}$ matrices calculated for each mixing time, we obtained the standard deviations $\delta\sigma_{ij}$ of all elements σ_{ij} of the relaxation matrix **R**. The lower (r_{ij}^-) and upper (r_{ij}^+) bounds of the internuclear distances r_{ij} were calculated from:

$$r_{ij}^{\pm} = \frac{r_{ij}}{\sqrt[6]{1 \mp \left|\frac{\Delta\sigma_{ij}}{\sigma_{ij}}\right|}} \qquad (17)$$

where the maximal deviation $\Delta\sigma_{ij}$ was set to $3*\delta\sigma_{ij}$. The maximal distance errors obtained from experimental data did not exceed ± 0.1 Å. Thus, the obtained lower and upper distance bounds were used as restraints in the determination of the molecular structure.

Conformational analysis of the free state

Detailed analysis of the free state is a necessary intermediate step in the calculations. It permits reconstruction of the relaxation matrix of the free state, which will become a part of the generalized relaxation matrix. This free-state submatrix will subsequently be held constant throughout the following calculations, and thus allow for the recreation of the bound-state submatrix. Attempts to fit the whole generalized relaxation matrix to the limited experimental data are bound to fail, because the large number of variables render the system underdetermined. The usual outcome is either a lack of convergence or convergence to the nearest local minimum (usually unrelated to the system's real parameters).

It suffices to know the elements of the relaxation matrix in the free state to proceed further. It is not necessary to convert them to distances. In fact, in most cases such conversion does not offer any benefits, as only apparent distances can be obtained due to the high mobility of the molecule. Rapidly interconverting conformers cause averaging of the relaxation matrix elements (e.g. as $\langle r_{ij}^{-6}\rangle$ in the case of aromatic or methylene protons, or $\langle(r_{ij}^{-3})\rangle^2$ for methyl groups), so that usually no unique structure can be derived from such a matrix.

In order to verify this statement, we attempted to convert all the off-diagonal elements of the free-state relaxation matrix into distances and to obtain a unique structure of the ligand. As expected, in the final structures the distance violations were unacceptably large. When pseudoatoms were employed, families of structures were obtained, indicating a high degree of segmental mobility, manifesting itself, as expected, through the disorder of the side chains of the molecule.

Determination of the bound-state structure

In order to obtain a family of tricycle structures complexed with TK, the distance constraints were extracted from that part of the relaxation matrix which corresponded to the bound state. Since the mixture of free and bound experimental NOEs is measured in the free state, where no stereospecific assignments are available, only the sum of NOE intensities in each proton group is available. As a consequence, it is necessary to employ pseudoatoms in the process of structure determination in the bound state.

The distance constraints used in the structure determination protocol were created by setting the lower and upper limits to the values determined from Eq. 17. In the case of the methyl group, both of these limits were extended by 0.5 Å to account for the different orientations of this group (Liu et al., 1992). The constraints were considered satisfied if the distance violation was within 0.1 Å.

A total of 100 structures were generated with the above protocol, 33 of which were acceptable on the basis of their constraint violations. In terms of the symbols used in Fig. 1, in accepted structures only the distance between the proton b and the group e was 0.17 Å too large. Otherwise we note a very good agreement with the experimental data. The 33 structures were distributed nearly evenly among four symmetry-related families (preliminary results of the docking indicate that only one of the four con-

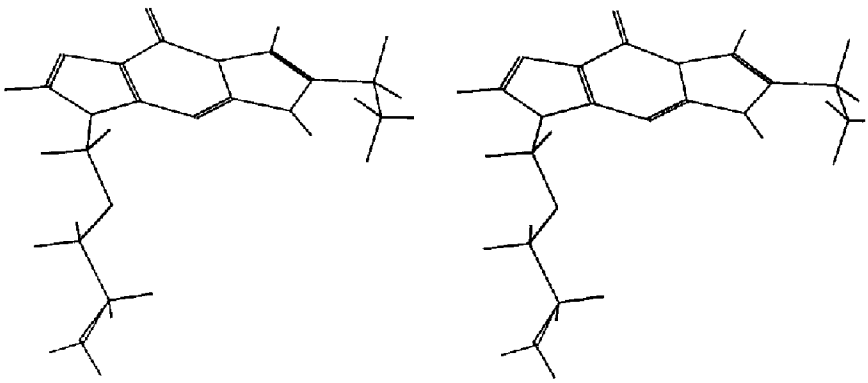

Fig. 5. Stereoview of the tricycle molecule in the conformation it assumes when bound to TK. The image is a superposition of nine different structures and represents one of four symmetry-related families of final structures, compatible with the experimental data. See the text for details.

formers fits inside the TK binding pocket; the docking itself will be reported elsewhere). One of them, consisting of nine superimposed conformations, is shown in Fig. 5. It can be seen that in this case one preferred conformation exists. The terminal OH group can be found in different conformations, because no constraints were available for it (the NMR signal for this proton could not be recorded due to fast exchange with the solvent), although it may have a well-defined orientation in the bound state. Our final check of the quality of the assessed structures consisted in reconstructing the relaxation matrix of the complex on the basis of the PDB files of the bound tricycle, and comparing the generated and experimental data. The results indicated overall agreement with experiment, with the exception of the region where the above-mentioned violation was reported.

Conclusions

We have been able to demonstrate the advantages of the full relaxation matrix analysis of complete buildup curves. A simple qualitative analysis performed by us in the preliminary part of this work (Czaplicki et al., 1995) showed that the NOE patterns in the free state and in the protein/ligand mixture were identical. However, a resulting observation that the structures of the ligand in the free and bound states are similar is highly unsatisfactory, particularly as we have shown here that there is no unique structure in solution, while there is one preferred conformation in the bound state. Using the more conservative approach of the initial slope approximation for determining internuclear distances would not work well in our case, because there is a substantial degree of intramolecular mobility, characterized by correlation times varying from site to site. We have obtained estimates of these times from measurements of the relaxation times T_1 and T_2, but the measured values correspond to the total relaxation in the spin system, while in the absence of detailed information on the spin geometry the estimates are made on the basis of the dominant contribution, i.e. the *intra*group term. As a consequence, incorrect correlation time estimates lead to incorrect distances without a clear indication of this problem in the final fitting quality. Similar results would be obtained by full relaxation matrix treatment of the initial parts of the buildup curves, because in the absence of spin-diffusion effects the program would compensate the erroneous estimates of correlation times by adjusting the distances and external relaxation leakage terms. Eventual disagreements between the experimental and generated data would thus be masked by arbitrary choice of the available fitting parameters. By way of contrast, a full relaxation matrix analysis of the complete buildup curves permits to overcome these problems by analyzing the interdependence of magnetization from all protons in the spin system measured at long mixing times. This method allows to visualize incorrect choices of local correlation times in the form of clear deviations between the experimental and calculated data. Moreover, an inability to reconstruct the entirety of some of the experimental curves by varying the fitting parameters indicates that in the proximity of the protons whose curve is badly fitted other protons are present, unaccounted for in the relaxation matrix and close enough to cause perturbations. For this effect to be observable, the distances between the enzyme protons and ligand protons must be shorter than those characterizing the ligand protons.

As a result of the series of the transferred NOE study on tricycle in solution and in complex with TK, we were able to determine the basic characteristics of families of the free and bound ligand structures. According to our expectations, the extended chain is disordered in the free state. On the other hand, the superposition of the bound structures clearly shows a preferred orientation, providing a basis for the docking of the ligand in the binding site of TK. The conclusions drawn from the analysis of the experimental data permit for the first time to actually model the structure of the binding site of TK. In modeling one can utilize the information inherent not only in the tricycle's bound structure itself, but also in the external relaxation leakage terms from the reconstructed relax-

ation matrix. These terms reflect the presence of the protein protons, which were not included explicitly in the relaxation matrix analysis. Once accomplished, the understanding of the nature of the binding site will permit the design of drugs with widely differing affinities for binding to TKs.

Acknowledgements

We thank Dr. Kurt Eger for the sample of tricycle used in this study, Dr. Martina Michael for the preparation of the protein samples and Dr. Didier Rognan for assistance in obtaining quality images of the final families of structures. We acknowledge the financial help of CNRS (IMABIO, Departement des Sciences de la Vie) and the Région Midi-Pyrénées in obtaining our NMR equipment.

References

Balaram, P., Bothner-By, A.A. and Dadok, J. (1972) *J. Am. Chem. Soc.*, **94**, 4015–4016.

Biosym Technologies, Inc., San Diego, CA, U.S.A., *NMRchitect User Guide*, v. 2.3, 1993.

Black, M.E. and Loeb, L.A. (1993) *Biochemistry*, **32**, 11618–11626.

Boryski, J., Golankiewicz, B. and De Clercq, E. (1991) *J. Med. Chem.*, **34**, 2380–2383.

Campbell, A.P. and Sykes, B.D. (1991) *J. Magn. Reson.*, **93**, 77–92.

Clore, G.M. and Gronenborn, A.M. (1982) *J. Magn. Reson.*, **48**, 402–417.

Czaplicki, J., Michael, M., Folkers, G. and Milon, A. (1995) *J. Chim. Phys.*, **92**, 1773–1776.

Edmondson, S.P. (1992) *J. Magn. Reson.*, **98**, 283–298.

Eriksson, S., Kierdaszuk, B., Munch-Petersen, B., Öberg, B. and Johansson, N.G. (1991) *Biochem. Biophys. Res. Commun.*, **176**, 586–592.

Fetzer, J. and Folkers, G. (1992) *Pharm. Pharmacol. Lett.*, **2**, 112–114.

Fetzer, J., Folkers, G., Müller, I. and Keil, G.M. (1993) *Virus Genes*, **7**, 205–209.

Fetzer, J., Michael, M., Bohner, T., Hofbauer, R. and Folkers, G. (1994) *Protein Expr. Purif.*, **5**, 432–441.

Fitt, P.S., Peterkin, P.I. and Grey, V.L. (1976) *J. Chromatogr.*, **124**, 137–140.

Folkers, G., Trumpp-Kallmeyer, S., Gutbrod, O., Krickl, S., Fetzer, J. and Keil, G.M. (1991) *J. Comput.-Aided Mol. Design*, **5**, 385–404.

Furlong, N.B. (1963) *Anal. Biochem.*, **5**, 515–522.

Fyfe, J.A., Keller, P.M., Furman, P.A., Miller, R.L. and Elion, G.B. (1978) *J. Biol. Chem.*, **253**, 8721–8727.

Huber, B.E., Richards, C.A. and Krenitsky, T.A. (1991) *Proc. Natl. Acad. Sci. USA*, **88**, 8039–8043.

James, T.L. (1994) *Methods Enzymol.*, **239**, 416–439.

Keller, P.M. and Elion, G.B. (1981) *Biochem. Pharmacol.*, **30**, 3071–3077.

Lian, L.Y., Barsukov, I.L., Sutcliffe, M.J., Sze, K.H. and Roberts, G.C.K. (1994) *Methods Enzymol.*, **239**, 657–699.

Liu, H., Thomas, P.D. and James, T.L. (1992) *J. Magn. Reson.*, **98**, 163–175.

London, R.E., Perlman, M.E. and Davis, D.G. (1992) *J. Magn. Reson.*, **97**, 79–98.

Michael, M., Fetzer, J. and Folkers, G. (1994) *Eur. J. Biochem.*, **226**, 219–226.

Miller, W.H. and Miller, R.L. (1980) *J. Biol. Chem.*, **255**, 7204–7207.

Munch-Petersen, B., Cloos, L., Tyrsted, G. and Eriksson, S. (1991) *J. Biol. Chem.*, **266**, 9032–9038.

Neuhaus, D. and Williamson, M. (1989) *The Nuclear Overhauser Effect in Structural and Conformational Analysis*, VCH Publishers, New York, NY, U.S.A.

Nilges, M., Clore, G.M. and Gronenborn, A.M. (1988) *FEBS Lett.*, **239**, 129–136.

Ohno, T., Gordon, D., San, H., Pompili, V.J., Imperiale, M.J., Nabel, G.J. and Nabel, E.G. (1994) *Science*, **265**, 781–784.

Reardon, J.E. and Spector, T. (1989) *J. Biol. Chem.*, **264**, 7405–7411.

Tropp, J. (1980) *J. Chem. Phys.*, **72**, 6035–6043.

Zangwill, W.I. (1969) *Nonlinear Programming: A Unified Approach*, Prentice-Hall, Englewood Cliffs, NJ, U.S.A.

Zimmermann, N., Beck-Sickinger, A.G., Folkers, G., Krickl, S., Müller, I. and Jung, G. (1991) *Eur. J. Biochem.*, **200**, 519–528.