A Virtual Screening Approach Applied to the Search for Trypanothione Reductase Inhibitors

Dragos Horvath*

CEREP Lille, 1, rue Calmette, 59019 Lille Cedex, France

Received May 24, 1996[⊗]

A prediction algorithm of the binding affinity of ligands to trypanothione reductase (TR), the enzyme replacing glutathione reductase in the metabolism of trypanosomatidae, has been used for the "virtual screening" of a data base of 2500 molecular sketches and has detected several structures of putative TR ligands. Most of these compounds turned out to be micromolar inhibitors of TR, as predicted by the algorithm. While their inhibitory potencies are lower than those of previously reported compounds, one of the molecules reported here could represent the lead toward a structurally different class of TR inhibitors. The fully automated prediction algorithm converts the 2D molecular sketches into 3D ligand structures, explores the conformational space of the latter, and performs a grid-based, rigid-body docking of the resulting family of ligand conformations into the TR site, calculating enthalpic and entropic binding indexes and predicting the binding affinity. The docking model has also been used to obtain hints about the binding modes of TR ligands.

I. Introduction

The enzyme trypanothione reductase (TR) appears to be one of the most promising targets for trypanocidal drugs.[1,2] TR, the parasitic homologue of glutathione reductase (GR), is involved in the regulation of oxidative stress in the parasite cells. Both TR and GR are homodimeric FAD-dependent reductases.[3] It is generally considered[4−10] that the mutual selectivity of these enzymes toward their own substrates (trypanothione TS_2 and respectively glutathione GSSG) is based upon differences in the electrostatic and hydrophobic properties of their sites. Some basic residues of the GR site are replaced by acidic or hydrophobic residues in TR.[10] This has been exploited in order to design *specific* TR inhibitors, such as mimics of the natural substrate[11] or molecules containing an aromatic moiety linked to a (poly)amino chain by means of a hydrophobic spacer.[12]

Molecular modeling is a valuable tool in drug design. The low accuracy *pharmacophore approaches* are suitable for "virtual screening" of molecular data bases[13−15] due to their low computational cost. *De novo* ligand design techniques are used to generate structures that match a given pharmacophore pattern.[16−19] quantitative structure−activity relations (QSAR)[20−26] relate the activities of a series of known ligands to some calculable molecular properties. Ligand binding can be quantitatively described by *simulating* the enzyme−ligand−solvent complex. If the system has visited all of the major energy minima[27,28] during the simulation, statistical mechanics[29] can be used to calculate its thermodynamic properties. Free energy calculations by thermodynamic integration[30,31] or perturbation techniques[30,32] have been used to determine the relative binding energy differences upon a slight change in the structure of the ligand or the site. The use of a *grid description*[33,34] of the different potentials within the enzymatic site dramatically diminishes the computer cost of the docking calculations.

In this work, we report an original computational approach to predict the TR inhibitory potency of molecules in order to detect new putative inhibitors by virtual screening of molecular data bases. It consists in defining a set of simplified "virtual physical laws" supposed to govern the behavior of both the *free* and the *bound* ligands, so that binding can be characterized by means of the calculated "virtual" enthalpy and entropy indexes. A compromise must be reached between the *computational effort* and the *degree of realism of this set of rules* in order to make the method suitable for screening a large number of putative ligands. The binding indexes must not necessarily correspond to the real enthalpy and entropy as far as they can be used as explaining variables of the affinity, as in a QSAR approach.

The TR enzymatic system displays some peculiarities that were taken into account in our affinity model. The *flexibility* of most of the TR inhibitors[11,12] makes mandatory the use of *conformational analysis techniques*. For the same reason, the binding *entropy* contributions are expected to be important. Furthermore, the X-ray structures of both complexes of the TR[11] and GR[7,8] active sites witness a lot of weak and water-mediated site−substrate interactions. The *few changes*[35] in the positions of the side chains of the free[10] and complexed[11] TR sites suggests that a *rigid site model* might be a valid working hypothesis. Accordingly, a *rigid-body* docking procedure based upon a *grid description* of the site potentials has been adopted in this work. The ligand flexibility is accounted for without losing the simplicity of the rigid-body docking, by docking a *relevant family of rigid conformations* of the ligand.[36]

The adjustable parameters appearing in the model were evaluated by fitting the calculated to the experimental affinities of ligands.[12,35,37−43] As a test, the model has been used to predict the affinities of a different series of known ligands (Figure 1, Table 1). Finally, it has been applied to the search of new inhibitors and the affinities of molecules predicted to have a favourable binding free energy have been submitted to TR inhibition tests.

* Tel: (33)3.20.87.12.31. Fax: (33)3.20.87.12.33. E-mail: Dragos.Horvath@pasteur-lille.fr.

[⊗] Abstract published in *Advance ACS Abstracts,* June 15, 1997.

S0022-2623(96)00378-0 CCC: $14.00

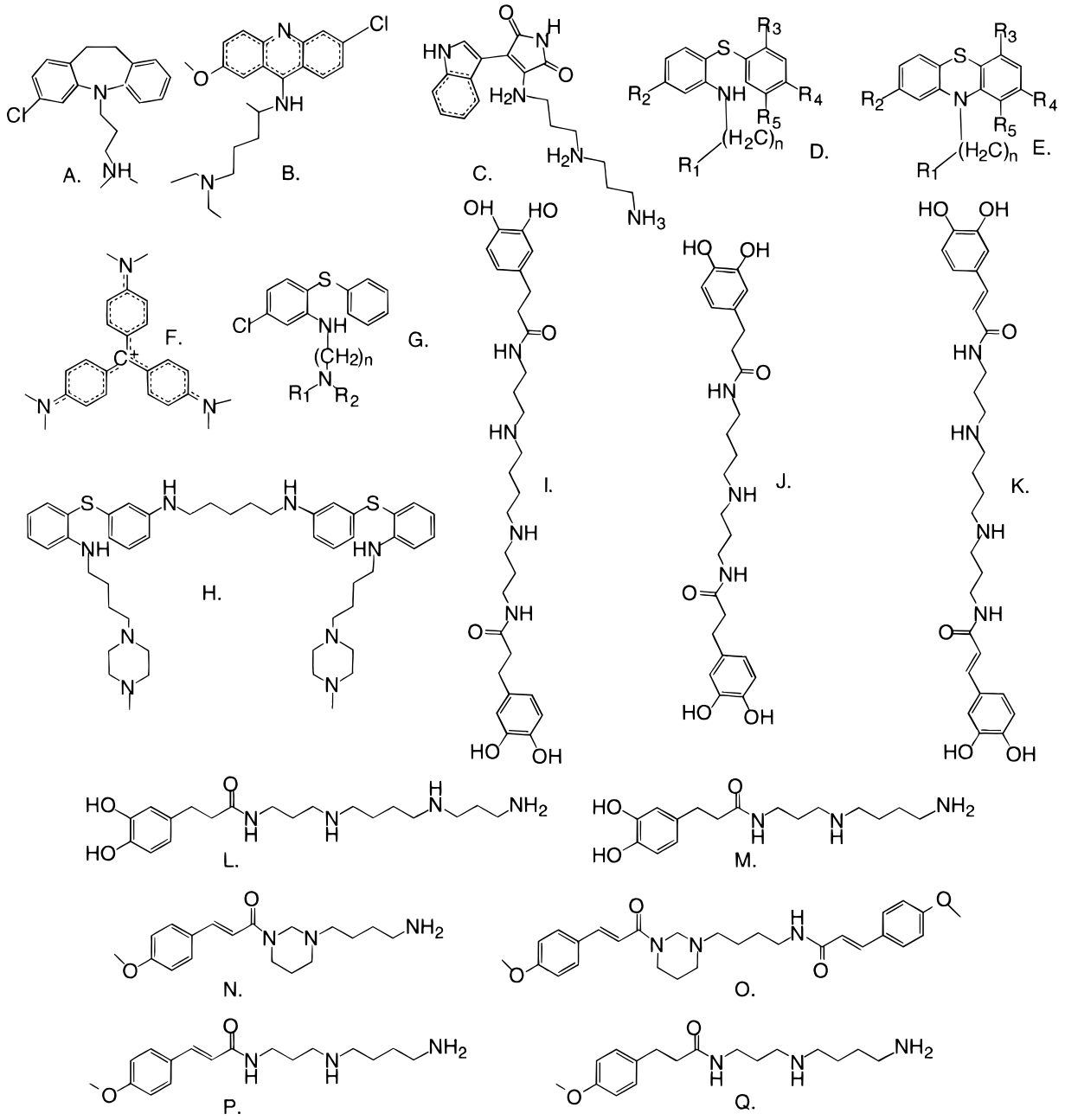

Figure 1. Generic structures of the molecules used to calibrate and test the docking model. The substituents R_i and the number of carbons of the hydrophobic spacer are given in Table 1.

II. Methods

1. Automated Generation of 3D Structures of Ligands. i. General Overview of Existing Methods. The prediction of the geometries of the stable conformers of small molecules has found a variety of solutions.[44,45] Knowledge-based rules can be used to restrain the considered values of a torsional angle in a specific chemical context.[46] The torsional angle driving procedures available in commercial modeling packages[47] add torsional constraint terms to the molecular Hamiltonian. Conformational sampling by molecular dynamics (MD) or Monte Carlo (MC) simulations do not offer the guarantee that all of the relevant minima have been visited. Biased sampling approaches such as "poling" techniques, which enhance crossing the energy barriers, show an improved sampling efficiency.[48]

ii. Working Hypotheses of the Conformational Search ("Sampling Axioms"). In our approach, systematic rotations around the most central torsional axes generate a very large number of crude geometries, out of which only a subset of *representative geometries, each converging toward a different energy minimum*, are filtered out and subjected to an energy minimization in vacuum. The selection is based upon the following rules, entitled the sampling axioms:

(a) Discard the geometries with overlapping nonbonded atoms.

(b) Discard the geometries that are *similar* to already sampled conformers.

(c) Give priority to the geometries *with the largest minimal distance* between two nonbonded atoms.

The rationale behind axiom b is to maximize the probability that two selected geometries will converge toward different minima. In consequence, a single enantiomer of a chiral conformation is kept, while the other will be regenerated by a mirroring operation during the docking (Methods, section 3.ii). Axiom c is meant to counteract the tendency of the vacuum minimization to fold the structures over themselves, favoring more "extended" geometries that may be more populated in solution.

The main advantage of this approach is that, at the step of "combinatorial explosion" of all possible rotamers, the energy calculations are replaced by the fast evaluation of the sampling criteria. An imposed number of 300−500 most diverse starting conformations are obtained within minutes.

iii. The energy minimization of the selected conformations is done by the Discover[47] program, with the CVFF[49,50] force field, using a steepest descent followed by a pseudo-

Table 1. Structures Used for the Calibration and the Testing of the Docking Model

Nr	Ref	R_1	R_2	R_3	R_4	R_5	n	ACT.	Ref.	Set.
1	C.	-	-	-	-	-	-	0.91	37	all
2	G.	$(CH_2)_2NH_2$	$(CH_2)_2NH_2$	-	-	-	3	1.50	35	all
3	A.	-	-	-	-	-	-	1.95	12	all
4	E.		Cl	H	H	H	3	2.71	38	all
5	D.		Cl	H	COMe	H	3	2.94	38	all
6	E.	NMe_2	Cl	H	H	H	3	3.00	38	all
7	G.	$(CH_2)_2NH_2$	H	H	H	H	3	3.00	35	all
8	B.	-	-	-	-	-	-	3.22	39	all
9	D.		Cl	H	H	H	3	3.30	40	none
10	D.		Cl	H	Me	H	3	3.33	38	all
11	D.		Cl	Cl	H	Cl	3	3.40	38	all
12	D.		Cl	H	H	H	2	3.68	40	all
13	D.	NEt_2	Cl	H	H	H	3	3.74	40	all
14	D.		Cl	H	H	H	3	3.78	40	none
15	D.		Cl	H	Cl	H	2	3.85	38	all
16	D.	NMe_2	Cl	H	H	H	2	3.91	40	all
17	D.		Cl	H	H	H	3	3.95	40	none
18	D.		Cl	H	Cl	H	3	4.04	38	all
19	D.		Cl	H	Cl	H	3	4.06	38	none
20	D.		Cl	H	Cl	H	3	4.09	38	D.
21	D.		H	H	Me	H	3	4.17	38	all
22	D.		Cl	H	H	H	2	4.34	38	all
23	D.	NMe_2	Cl	H	H	H	2	4.38	38	none
24	D.	NMe_2	H	H	Cl	H	2	4.38	38	none
25	D.		H	H	Cl	H	2	4.38	38	none
26	D.		Me	H	Me	H	3	4.38	38	all
27	D.		H	H	H	H	3	4.73	38	all
28	D.		Cl	H	F	H	3	4.74	38	D.
29	D.		H	H	Cl	H	2	5.08	38	all
30	D.		Me	H	F	H	3	5.35	38	none
31	D.	NEt_2	Cl	H	H	H	2	6.21	38	all
32	D.		Cl	H	H	H	2	6.55	38	all
33	D.		Cl	H	Cl	H	2	6.90	38	all
34	I.	-	-	-	-	-	-	0.59	41	C,D.
35	J.	-	-	-	-	-	-	2.01	41	B,C,D
36	K.	-	-	-	-	-	-	2.77	41	C,D.
37	L.	-	-	-	-	-	-	4.44	41	none
38	M.	-	-	-	-	-	-	4.68	41	none
39	N.	-	-	-	-	-	-	5.01	41	B,C,D
40	O.	-	-	-	-	-	-	5.01	41	none
41	P.	-	-	-	-	-	-	5.60	41	none
42	Q.	-	-	-	-	-	-	5.82	41	B,C,D
43	H.	-	-	-	-	-	-	1.00	42	none
44	F	-	-	-	-	-	-	1.60	43	none

[a] Template, molecular template from Figure 1; R_1−R_5, corresponding substituents; *n*, length of the spacer chain; ACT, natural log of the value (in μM) of the inhibition constant; Ref, literature reference; Set., presence of the molecule in the learning sets of different calibration trials (see Table 4).

Newton algorithm[51] until the maximum derivative reaches 0.1 kcal/Å. Minimum energy conformers differing by less than 0.05 kcal/mol with respect to both the *total* and the *Coulomb* energies are kept only once. The geometries with more than 5 kcal/mol of excess energy with respect to the best minimum are discarded. For reasons of computational expense, only the 30 most stable conformations are considered for docking.

iv. Modeling of the Protolytic Equilibria in Ligands. The effective charge distribution of polyamines, a class to which many TR inhibitors belong, is determined by the protolytic equilibria[52] in solution. Here we report a way to account for *an averaged* effect of the protolytic equilibria on the docking energies.

The equilibrium between an ammonium group and the uncharged amine can be accounted for by weighting the charge $Q = +1$ by the probability p of finding the group in a protonated state at pH = 7 (practically, the charges of all the atoms of the *ammonium charge group* need to be rescaled). The free energy governing the protolytic equilibria in function of the probabilities of protonation p_i can be written as

$$G_{\text{prot}} = -\sum_i h_i p_i + K\sum_{i\neq j}\frac{p_i p_j}{d_{ij}} \qquad (1)$$

where h_i is the specific protonation enthalpy of the group i and d_{ij} is the distance between the groups. The first term $(-h_i p_i)$ in (1) expresses the fact that isolated amino groups tend to accept a proton due to a favorable proton-binding enthalpy $-h_i$. If the proton-binding affinity were the only driving force of protonation, the minimum of G_{prot} should be at $p_i = 1$. In reality, $p_i < 1$ due to the mixing entropy terms. These will be ignored here, since isolated aliphatic amino groups at pH = 7 are *almost* completely protonated (mono- and diaromatic amines were always taken in their *un*protonated forms).

If there are several amino groups in the same molecule, the protonation of one will impede on the proton affinity of the second, due to an unfavorable Coulomb interaction between the two positive charges, according to the second term in the sum (1), where $p_i p_j$ is the fraction of molecules with both groups i and j in a protonated state.

Considering an average value $\langle h\rangle$ instead of h_i, $\Pi = \langle h\rangle/K$ becomes a *fittable* parameter that controls the protonation state of the ligands. The p_i values corresponding to a minimal G_{prot} are obtained by solving the linear system

$$\frac{1}{K}\frac{\partial G_{\text{prot}}}{\partial p_k} = -\Pi + \sum_{i\neq k}\frac{p_i}{d_{ik}} = 0 \qquad (2)$$

They may occasionally fall outside the range of (0,1) due to the neglect of mixing entropy terms, a situation which should be interpreted as either a total deprotonation or protonation. In a piperazine group, $p_n \approx 0.5...0.6$ and $d = 3$ Å between the two nitrogen atoms, which implies $\Pi \approx 0.2$ Å^{-1}.

2. Definition and Mapping of the Potentials in the TR Site. i. Definition of the Grid. A parallellipipedic box of $30 \times 40 \times 35$ Å^3 has been defined around the active site of the TR enzyme, which is available in the Brookhaven Protein Data Bank.[53] The grid spacing was taken 0.5 Å.[33]

ii. The Enzymatic Site. For reasons of computer efficiency, nonpolar CH_n groups were treated as "united atoms" and were assigned the parameters of the CVFF nonpolar carbon potential type.

(a) X-ray Geometries. A first-considered option was to calculate the potential grids using the X-ray geometries of either the free[10] or the complexed[11] TR.

(b) Average Geometries Considering the Movement of the Side Chains. An alternative approach was to subject the site to 100 ps of MD simulation in vacuum at 300 K, with a fixed backbone. Geometries were sampled every picosecond, and their corresponding potential values at every grid point were averaged, smoothing out the steep variations of the potentials near the site atoms, a major source of docking artifacts.

iii. The mapped potentials are evaluated at each point P of the grid, in function of its distances d_{iP} to the site atoms i.

(a) The Coulomb and van der Waals potentials were parameterized according to the CVFF force field:

$$V_{\text{Coul}}(P) = \sum_{i=1}^{\text{site atoms}} \frac{Q_i}{4\pi\epsilon_0(2d_{iP})d_{iP}} \tag{3}$$

$$V_{\text{vdW}}^{\text{rep}}(P) = \sum_{i=1}^{\text{site atoms}} \frac{A_i}{d_{iP}^{12}} \tag{4}$$

$$V_{\text{vdW}}^{\text{att}}(P) = -\sum_{i=1}^{\text{site atoms}} \frac{B_i}{d_{iP}^{6}} \tag{5}$$

(b) The continuum solvent model proposed by Gilson and Honig[54] approximates the change of the electrostatic energy of a charged atom i that approaches a low-dielectric (ϵ_{int}) atom j of volume v_j, which displaces the high-dielectric (ϵ_{sol}) solvent. Simultaneously, atom i plays the role of displacing atom with respect to the charge Q_j of atom j. This approach (similar to that in ref 32) allows, in contrast to more elaborated solvent models,[55] the definition of two *potentials* expressing the desolvation of the ligand by the site and vice versa:

$$V_{\text{desolv}}^{\text{sit}\rightarrow\text{lig}}(P) = \frac{1}{8\pi^2\epsilon_0}\left(\frac{1}{\epsilon_{\text{int}}} - \frac{1}{\epsilon_{\text{sol}}}\right)\sum_{i=1}^{\text{site atoms}} \frac{v_i}{d_{iP}^4} \tag{6}$$

$$V_{\text{desolv}}^{\text{lig}\rightarrow\text{sit}}(P) = \frac{1}{8\pi^2\epsilon_0}\left(\frac{1}{\epsilon_{\text{int}}} - \frac{1}{\epsilon_{\text{sol}}}\right)\sum_{i=1}^{\text{site atoms}} \frac{Q_i^2}{d_{iP}^4} \tag{7}$$

(c) The hydrophobic effect is usually represented as a term proportional to the surface area that is buried during the binding.[56] The area of an atom i of radius r_i buried by the sphere centered on atom j, of radius r_j, is a function of the intercenter distance $d_{ij} < r_i + r_j$:

$$b_{j\rightarrow i} = \pi r_i(r_i + r_j - d_{ij})\left(1 + \frac{r_j - r_i}{d_{ij}}\right) \tag{8}$$

These spheres correspond to "united" CH_n atoms, of equal radii $r = 2.5$ Å, so that eq 8 reduces to $b_{j\rightarrow i} = \pi r(2r - d_{iP})$. Ignoring the *multiple overlaps* between atoms and setting $\delta(i) = 1$ if atom i is a CH_n group and 0 otherwise, the definition of the hydrophobic potential becomes

$$V_{\text{hphob}}(P) = -\pi r \sum_{i=1}^{\text{site atoms}} \delta(i)(2r - d_{iP}) \tag{9}$$

3. The Energy Functions of the Model and Their Fittable Parameters: i. Molecular Hamiltonian or 3D QSAR Descriptors? In contrast with QSAR models based on empirical 2D or 3D descriptors, the terms used in this work have a physical foundation. Therefore, the *fittable* weighting coefficients that will be introduced in the expression of the docking energy can be considered either as *physical parameters* of uncertain value or as *QSAR parameters*, meant to account for the relative importance of the different energy contributions to the biological activity.[25]

(a) Electrostatic Term. The Coulombic term using a distance-dependent[57] dielectric constant $\epsilon_{\text{eff}} = 2r$ is not a rigorous description of charge-charge interactions in dielectrically heterogeneous media.[58] The coefficient χ may be regarded as a corrective term to be applied to this initial guess of ϵ_{eff}.

(b) Desolvation. Since the variance of the individual atomic volumes v_i in (6) is not relevant,[59] these have been replaced with an average volume $\langle v \rangle$ that can be taken out of the sum. In (14), we set $v_i = \langle v \rangle = 1$ and let λ scale up this average volume to its effective value.

(c) Hydrophobicity. The hydrophobic potential must be expressed in kcal/mol by multiplication with an appropriate coefficient η.

(d) Nonbonded Interactions. The *repulsive* van der Waals term has been adopted from CVFF without weighting, its role being to discard the geometries with site−ligand clashes. A factor ω has been used to scale the *attractive* term, as an attempt to compensate for the dispersive interactions with the solvent, which were *neglected* in the present work. The evaluation of the latter *potential of mean force* is a complex task,[60] but could be implicitly accounted for by continuum solvent models.[33]

(d) The protonation model controlled by a fittable coefficient Π has been described previously (section 1.iv, Methods).

In consequence, the site−ligand interaction energy becomes

$$E_{\text{S-L}} = E_{\text{S-L}}^{\text{rep}} + \omega E_{\text{S-L}}^{\text{att}} + \chi E_{\text{S-L}}^{\text{Coul}} + \lambda E_{\text{S-L}}^{\text{des}} + \eta E^{\text{hphob}}{}_{\text{S-L}} \tag{10}$$

Each of these terms can be written in function of the defined potentials

$$E_{\text{S-L}}^{\text{rep}} = \sum_{i=1}^{\text{ligand atoms}} A_i V_{\text{vdw}}^{\text{rep}}(i) \tag{11}$$

$$E_{\text{S-L}}^{\text{att}} = \sum_{i=1}^{\text{ligand atoms}} B_i V_{\text{vdw}}^{\text{att}}(i) \tag{12}$$

$$E_{\text{S-L}}^{\text{Coul}} = \sum_{i=1}^{\text{ligand atoms}} Q_i V_{\text{Coul}}(i) \tag{13}$$

$$E_{\text{S-L}}^{\text{des}} = \sum_{i=1}^{\text{ligand atoms}} Q_i^2\, V_{\text{des}}^{\text{sit}\rightarrow\text{lig}}(i) + v_i V_{\text{des}}^{\text{lig}\rightarrow\text{sit}}(i) \tag{14}$$

$$E_{\text{S-L}}^{\text{hphob}} = \sum_{i=1}^{\text{ligand atoms}} \delta(i) V_{\text{hphob}}(i) \tag{15}$$

Both the Coulombic and desolvation terms are calculated using the weighted atomic charges as derived from the protonation model.

ii. Energy Levels of the Free Ligands. An aposteriori solvation correction based upon the Gilson−Honig solvent model must be added to the energies of the ligand geometries that were minimized in vacuum. In spite of the failure[50] of this solvation potential to reproduce vacuum-to-water transfer energies, we argue that the model is able to explain the *interconformational variations* of the solvation energy. This hypothesis needs to be checked by comparisons with results obtained from more elaborate solvation potentials.

Denoting the Coulombic contribution by E_i^{c}, the desolvation contribution by E_i^{d} and the sum of all the other intramolecular force field terms by E_i^{nc}, the intramolecular total energy E_i^{free} of a free conformation i becomes

$$E_i^{\text{free}} = E_i^{\text{ne}} + \chi E_i^{\text{c}} + \lambda E_i^{\text{d}} \tag{16}$$

iii. Energy Levels of the Bound Ligands. Docking Energy. The site-ligand interaction energy in (10) is a function of the geometry of the docked conformer i, the relative position $\mathbf{r} = (r_x, r_y, r_z)$ of the mass centers of the two objects and the three angles $\alpha = (a_x, a_y, a_z)$ defining the rotation of the ligand with respect to the site. In function of the current conformation and of the initial positioning of this conformer in the site, the minimization of $E_{\text{S-L}}$ with respect to the 6 degrees of freedom will lead to the closest local minimum, defining an energy level of the docked ligand.

(a) Initial Positioning of the Ligands in the Site. An analysis of the TR site was performed in order to design different starting points from which the ligand will be allowed to evolve during optimization. These points should

(1) cover the whole accessible site thereby ensuring that the ligand will explore all the most relevant zones of the site

and find the optimal binding pocket when starting from at least one of these different initial positions

(2) be evenly spread across the site, thereby ensuring their distribution will not bias the relative populations of the minima, e.g. the ligand will converge more often toward binding locations in zones that were more densely covered with starting points.

A grid of 2 Å spacing has been generated around the active site of the enzyme. The grid points situated 4−7 Å from the nearest site atom have been considered as potential starting points. A Monte Carlo algorithm has been used to pick out a number of $N_S = 7$ starting points out of these candidates, in maximizing the distance between the two closest points of the selected distribution.

(b) The Docking Algorithm. The docking begins by placing a current conformer with its geometric center in one of the N_S starting points and ensuring a bump-free initial orientation. A combination of optimization methods (systematic rotations, translations, iterative searches, and eventually a conjugate gradients minimization) is applied, diminishing the risk of falling upon false minima. This is repeated from all the other starting points. In consequence, for each conformation $i = 1, ..., N_{free}$, the different optimizations carried out from each starting point $k = 1, ..., N_S$ will lead to a series of local energy minima E_n^{dock}, where $n = k + (i - 1)N_S$. Eventually, the whole procedure is repeated with the *mirror image* i' of the current conformation i in order to retrieve the enantiomeric geometries that were discarded at the 3D-building step, providing the energy levels $n' = N_S N_{free} + k + (i - 1)N_S$. The complete set of levels E_n^{dock}, $n = 1, ..., N_{dock}$ ($N_{dock} = 2N_S N_{free}$) characterize the bound ligand.

$$E_n^{dock} = E_i^{free} + \min_{r,\alpha} E_{S-L}(i/i';k'r,\alpha) \qquad (17)$$

4. The "Binding Indexes". i. Binding "Enthalpy" and "Entropy". While the free and bound energy levels thus defined do not form canonical ensembles, the equations of statistical mechanics are nevertheless a plausible choice for a *set of rules to derive some global docking scores* from these conformational score values.

(a) The stability index represents the difference between the most stable bound state and the most stable free state

$$\Delta H^* = \min_n E_n^{dock} - \min_i E_i^{free} \qquad (18)$$

(b) The binding enthalpy index can be defined considering a Boltzmann factor β

$$\Delta H = \frac{\sum_n E_n^{dock} e^{-\beta E_n^{dock}}}{\sum_n e^{-\beta E_n^{dock}}} - \frac{\sum_i E_i^{free} e^{-\beta E_i^{free}}}{\sum_i e^{-\beta E_i^{free}}} \qquad (19)$$

(c) The Binding Entropy Index. The binding affinity is a function of the global entropy difference of the families of populated conformations in the free and bound states. Rigorous calculations[61] of entropy differences are beyond the scope of this work. We define a $T\Delta S$ index

$$T\Delta S = \Delta H - \Delta G = \Delta H + \frac{1}{\beta} \ln \left(\frac{\sum_i e^{-\beta E_i^{dock}}}{\sum_j e^{-\beta E_j^{free}}} \right) \qquad (20)$$

only accounting for the entropy related to the number and Boltzmann weights of the populated states. We assume that the translational, rotational, and vibrational loss of disorder upon binding is constant for all the ligands of roughly the same size.

(d) An Empirical Estimation of Entropy. The accumulation of many energy levels close to the lowest level of a system, e.g. the existence of a multitude of populated states, implies a large entropy. Moreover, these energy levels contribute to the increment of the enthalpy H with respect to the lowest energy level H^*, due to their important Boltzmann weights. If, as this qualitative reasoning suggests, $(H - H^*)$ and as the TS term defined in (20) would correlate to a certain degree (a *too strong correlation* would imply the redundancy of one of the descriptors!), then $(H - H^*)$ might be used as an alternative measure of the entropy.

Table 2. Correlations between the Solvation Energies from Boundary Element Calculations (E^{BEM}) and the Gilson−Honig Term (E^{GH}) for the Sets of Conformations of Different Compounds

compd	y'_0	λ'	rms	r^2	rms(λ)	$r^2(\lambda)$
16	−33.9	1.19	0.53	0.64	0.84	
14	−30.2	0.69	0.48	0.39	1.02	
9	−112.3	2.07	1.11	0.62	1.12	0.61
27	−33.0	1.06	0.57	0.49	0.88	
8	−38.3	1.88	0.53	0.68	0.57	0.64
7	−101.1	1.88	1.05	0.65	1.10	0.61
2	−204.6	1.88	0.99	0.85	1.14	0.79
1	−113.8	1.51	1.04	0.79	1.55	0.53
6	−38.3	2.32	0.41	0.70	0.41	0.70
23	−110.7	2.07	0.46	0.91	0.50	0.90
4	−122.9	3.14	0.48	0.96	0.80	0.89
3	−35.64	1.51	0.24	0.86	0.31	0.75
13	−33.63	1.48	0.41	0.55	0.52	0.27[a]

[a] Columns 2−4 refer to the unconstrained regression equations $E^{BEM} = y'_0 + \lambda' E^{GH}$, where both the intercept y_0 and the RMS are in kcal/mol. Columns 5 and 6 display the correlation parameters at an imposed slope $\lambda = 2.30$.

5. Calibration of the Docking Model. i. Docking Parameters. The binding indexes defined in the previous section—the stability index in (18), binding enthalpy in (19), and binding entropy in (20)—are functions of the docking parameters (ω, χ, λ, η, Π). β has been taken $1/RT$ =1.67 mol/kcal (T = 300K).

ii. Affinity Parameters. A linear relation between the logs of the inhibition constants of the ligands and the previously defined indexes can be established and statistically validated if the latter are useful descriptors of the binding process (K_i denotes the inhibition constant of compound i).

$$\ln K_i = a\Delta H_i + b\Delta H_i^* + cT\Delta S_i + d \qquad (21)$$

The coefficients (a, b, c, d) will be referred as the affinity parameters (ideally $a = -c = 1.67$ kcal/mol, $b = 0$). While their optimal values will depend upon the peculiarities of the docking model, ΔH should in any case have a positive and $T\Delta S$ a negative coefficient. According to the hypothesis that $\Delta H - \Delta H^*$ is an *independent variable* acting as an *entropic descriptor*, ΔH and ΔH^* might be simultaneously used in the regression (21) in spite of their high degree of intercorrelation, if the obtained predictive power significantly improves with respect to the relations only in ΔH and $T\Delta S$.

iii. Fitting Procedure. A learning set of TR inhibitors of known affinities (Figure 1, Table 1) has been used to search for an optimal configuration of the fittable parameters (ω, χ, λ, η, Π, a, b, c, d) which minimizes the errors between calculated and measured ln K values.

6. Testing and Using the Model. Other inhibitors (Table 1, Figure 1) of known affinities, not used in the learning set, have been used for an affinity-prediction exercise in order to validate the calibration of the model. Eventually, the approach was applied to a library of 2500 molecules (partly available in ACD[62]). The molecules with $200 \le M_r \le 500$, with a calculated ln $K < 3$ and $\Delta H < -20$ kcal/mol, were subjected to TR inhibition tests.

The percentages of inhibition are given as $(r_0 - r_{inh})/r_0 \times 100$, where r_0 and r_{inh} are the rates of the enzymatic reduction of trypanothione disulfide by TR, measured in absence and in presence of the inhibitor.[63] An error of about 10% may affect the measured percentages of inhibition.

III. Results and Discussions

1. An Estimation of the Accuracy of the Desolvation Term. Table 2 reports the linear equations

Table 3. Efficiency of Our Conformational Sampling Algorithm Compared to a 100 ps Molecular Dynamics Run at 1000 K[a]

compd	ΔE_{min}	$\Delta(E_{min} + E_s)$	N_{init}	N	N_{MD}
1	−1.10	−1.20	326	33	8
2	−4.77	−2.26	304	19	13
7	−1.13	+0.49	426	39	14
4	+0.12	0.00	36	5	6
8	+2.71	−2.15	300	87	56

[a] The energy differences are in kcal/mol. ΔE_{min} is the difference between the vacuum energies of the best minumum found by our method and the one obtained by optimization of MD-sampled geometries. $\Delta(E_{min} + E_s)$ is the corresponding difference between the solvent-corrected energies. N_{int} represents the number of initial geometries that were subjected to minimization, while N and N_{MD} are the number of distinct minima found by our method and the MD simulation, respectively.

relating the solvation energies calculated with a modified[64] boundary element method (BEM) to the corresponding Gilson−Honig desolvation terms of the conformers of a given compound. For most of the molecules, fair to excellent correlations between the two solvation terms were found. In the two cases where $r^2 < 0.5$, the solvation energy shows little variance within the considered family of conformations of that compound. Considering only the six compounds with $r^2 > 0.7$, we obtain an average slope $\langle\lambda'\rangle$ of 2.07±0.56. This is in a very good agreement with the value of $\lambda = 2.30$ independently found by calibrating the model. Scaling the interconformational Gilson−Honig term differences by a factor of $\lambda = 2.30$ constitutes a reasonable approximation for the corresponding interconformational solvation energy differences. Furthermore, this λ value is a good estimate for the average atomic volume it was designed to account for (section 3.i, Methods).

2. The Efficiency of the Conformational Sampling Procedure. We have compared our sampling approach with a classical conformational analysis based upon a 100 ps (10^5 equilibration steps of 1 fs, with saving of the geometry at every 100th step) of a high-temperature (1000 K) MD simulation in vacuum.[47] For each compound, an alternative set of crude geometries, of the same size as the diversity-biased input set obtained by 2D-to-3D conversion, has been selected, in order of their increasing potential energy, among the geometries sampled by the MD trajectory and subjected to the same minimization procedure (section 1.iii, Methods).

The computer times needed to carry out the MD simulations were longer (typically 15−20 min) compared to the torsional angle driving that completed in 1...2 min on an Iris Indigo R4000.

In Table 3, we show some typical examples of the energy differences between the best minima obtained by our sampling approach with respect to the ones generated by the MD run, as well as the final number of accepted conformations. Maximizing the diversity of the starting geometries leads to both a larger number of minima and lower energy best minima for most of the ligands. However, this is a limited success since MD is a poor conformational sampling method compared to other approaches.[45,46,48]

The bias in favor of more extended (and therefore hopefully better solvated) conformations according to sampling axiom c (section 1.ii, Methods) appeared to be successful for compound **8**, for which the MD geometry set led to a better vacuum minimum, while the diversity-biased set found conformations that are more stable in solution. In contrast, it did not work for ligands **2** and **7**.

Table 4. Docking Parameters Resulting from Three Calibration Attempts (Denoted by A, B, and C) Making Use of Different Learning Sets

model	χ	η	λ	ω	Π
A	1.23	3.75	2.66	1.00*	0.21
B	1.15	3.94	2.29	0.96	0.27
C	1.15	3.62	2.31	0.96	0.22

[a] The hydrophobicity parameter η has been expressed in cal/A^2. All other parameters are dimensionless weighting factors. An asterisk (*) is used to denote a value that has been kept during this calibration run.

3. Calibration Results. i. Averaging of the Side Chain Geometries of the TR Site. A first important result is that only the potentials obtained by averaging over the MD trajectory of the active site led to good affinity predictions. Averaging would not have been a good strategy if the movement of site residues would have been strongly restrained upon complexation with a ligand. Therefore, our finding corroborates with the little differences observed between the side chain geometries of a free and complexed TR site.

ii. Optimal Parameters. Several optimization runs using different learning sets lead to those coefficients reported in Table 4 (docking parameters) and Table 5. (affinity parameters). The calibration set A was restricted to smaller ligands, while sets B and C progressively included larger and more flexible compounds.

(a) The docking parameters do not radically differ in function of the chosen learning set:

The van der Waals weighting factor ω, which was fixed to 1 in the run A, did not considerably drift away after fitting (sets B and C), although the Monte Carlo search was carried out in the range $\omega \in [0,1]$. This suggests that the van der Waals solvation contribution cannot be implicitly accounted for by weighting the van der Waals vacuum energies.

The desolvation term has been discussed in detail in section 1 of the Results section.

The hydrophobic coefficient is somewhat lower, but falls within the range usually found in the literature.[55,59] This is a consequence of the systematic overestimation of the buried area due to the ignoring of multiple overlaps.

The Coulombic weighting factor is close to 1, suggesting that the choice of the dielectric constant $\epsilon_{eff} = 2d$ was reasonable (the explored range was $\chi \in [0.5,1.5]$).

The protonation constant Π is the least stable docking parameter. The reported values correspond to a fraction of the diprotonated piperazine at pH = 7 of about 20% (A), 60% (B), and 30% (C). While the former and the latter values are in good agreement with the experiment, case B predicts a quite high rate of diprotonation. This important fluctuation of the total charge has little influence on the calculated binding indexes and affinity constants of most of the piperazine-containing ligands.

The observed tolerance of the model with respect to roughly 0.3 charge units on a dibasic ligand is probably due to the compensation between the Coulombic interactions, becoming more favorable, and the desolvation term, growing more unfavorable upon an increase of the global charge. Therefore, the quality of the model

Table 5. Established Relations between the (Natural) log of the Inhibition Constant and the Calculated Binding Indexes for Each of the Parametrization Schemes from Table 4[a]

	best relationship	rms	relationship in ΔH and $T\Delta S$	rms
A	$\ln K = 7.07 + 0.18\Delta H - 2.42T\Delta S$	0.78	$\ln K = 7.07 + 0.18\Delta H - 2.42T\Delta S$	0.78
B	$\ln K = 4.47 + 3.26\Delta H - 3.19\Delta H^* - 3.33T\Delta S$	1.01	$\ln K = 5.54 + 0.10\Delta H - 2.04T\Delta S$	1.12
C	$\ln K = 4.31 + 4.76\Delta H - 4.68\Delta H^* - 4.18T\Delta S$	0.95	$\ln K = 5.19 + 0.10\Delta H - 2.35T\Delta S$	1.30

[a] The BEST relations are the ones found by stepwise regression, retaining only the variables that are relevant for the model. Alternative equations only in ΔH and $T\Delta S$ are also shown. The cross-validated rms values refer to the calculated vs experimental values of ln K.

Table 6. Check of Whether the Binding Entropy Index as Evaluated by (20) Can Be Written as a Linear Combination of ΔH and ΔH^*[a]

model	established relationship $T\Delta S = a\Delta H - a^*\Delta H^*$	r^2 (CV)
A	$T\Delta S = 1.39\Delta H - 1.37\Delta H^*$	0.516
B	$T\Delta S = 1.25\Delta H - 1.24\Delta H^*$	0.553
C	$T\Delta S = 1.41\Delta H - 1.40\Delta H^*$	0.582

[a] A, B, and C refer to the different parametrization schemes listed in Table 4.

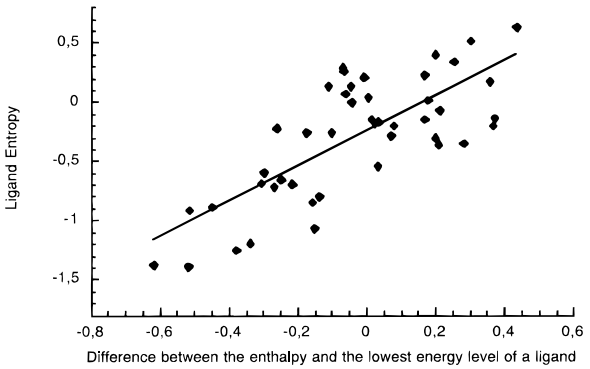

Figure 2. Covariance of the two different binding entropy indexes applied in the present study: $T\Delta S$ on the x-axis vs ($\Delta H - \Delta H^*$) on the y-axis (in kcal/mol) as resulting from parametrization scheme C (see Table 4).

displays a quite broad, but nevertheless explicit, maximum within a physically meaningful range of Π values.

(b) The affinity parameters significantly differ in function of the docking parameter configuration (Table 5). The largest discrepancies appear between the values obtained with the learning set A and respectively those of sets B and C, e.g. upon an increase of the diversity of the learning set. For parameter set A, the *best* expression of ln K is in ΔH and $T\Delta S$ only, while in the other two cases, ΔH^* is also involved.

While the correlation coefficient between ΔH and ΔH^* is as high as 0.995, the magnitude $\Delta H - \Delta H^*$ is completely independent with respect to both ΔH and ΔH^* ($r^2 \approx 0.2$). $\Delta H - \Delta H^*$ does not however, represent the "noise" due to the imperfect proportionality between ΔH and ΔH^* but appears to be a useful descriptor of the binding affinity (otherwise, the stepwise and cross-validated regression technique would have not entered ΔH^* into the model).

Expressing the $T\Delta S$ values of the ligands as linear combinations of ΔH and ΔH^* leads to correlations of fair quality (Table 6, Figure 2) which validate our hypothesis that $\Delta H - \Delta H^*$ may act as a descriptor of *entropic* nature. It is found that $T\Delta S \approx 1.2...1.4(\Delta H - \Delta H^*)$, independently of the docking parameter set. The correlation coefficients of 0.5 and 0.6 ensure that $T\Delta S$ and $\Delta H - \Delta H^*$ are not redundant variables.

The established regression equations of ln K in function of *only* ΔH and $T\Delta S$ are also reported in Table 5. In this case basically the same linear relation is obtained independently of the docking parameter set,

Table 7. Correlation Coefficients r^2 between the Sets of Binding Indexes ΔH and $T\Delta S$ Obtained with the Three Different Parametrizations Listed in Table 4[a]

	ΔH		$T\Delta S$	
	B	C	B	C
A	0.52	0.92	0.37	0.31
B	1.00	0.78	1.00	0.27

[a] These r^2 values result from linear regression calculations with fixed *unitary* slope and *null* intercept ($y = x$).

where the overall entropic contribution is about $-2T\Delta S$, close to the theoretical $-1.67T\Delta S$.

The large discrepancies in the coefficient of ΔH observed in the best regressions are mainly due to the very different contributions of $\Delta H - \Delta H^*$. The overall enthalpic term amounts to about $0.2\Delta H$, with a positive coefficient which proves to be much lower than expected.

(c) QSAR Model or Physical Binding Model? The obtained docking coefficients, close to values that can be backed by various theoretical or computational arguments, speak for the physical consistency of the model, while the values of the affinity coefficients suggest a QSAR-like affinity prediction model.

In spite of the fact that its weighting coefficient equals only 10% of the theoretical value, the enthalpy index ΔH accounts for roughly 25% of the explained variance, while $T\Delta S$ covers the rest (with parameter set A).

Table 7 compares the ΔH and $T\Delta S$ terms corresponding to each one of the docking parameter sets (A, B, C). It can be seen that the ΔH values are very well intercorrelated, while the entropy $T\Delta S$ displays a stronger dependence upon the used parameterization. Nevertheless, the overall contribution of about $-2T\Delta S$ suggests that while the $T\Delta S$ term for any particular ligand is affected by random errors, in general the calculated entropies may reflect quite well (up to a constant offset) the real binding entropies.

While the enthalpy index depends upon the separation between the bound and the free energy levels, the entropy index is only a function of the conformational energy differences within each of the ensembles of the bound and free states. An error in the substrate−ligand interaction energy that is roughly independent on the current ligand conformation (such as the neglect of the vdW and cavity-formation contributions to the solvation energy) would therefore mainly affect ΔH, constituting a first possible explanation for the low participation of ΔH in the model.

Alternatively, this could be explained on behalf of the enthalpy−entropy compensation as outlined in ref 65. The less tightly bound ligands (less negative enthalpy ΔH) still maintain a certain mobility in the site and lose only a fraction $T\Delta S^{\text{eff}}$ of the assumed translational−rotational entropic contribution, e.g. $T\Delta S^{\text{eff}} = \alpha\Delta H$. This partial loss of mobility could be implicitly accounted for by an appropriate weighting of the enthalpic contribution, since $\Delta H - T\Delta S^{\text{eff}} = (1 - \alpha)\Delta H$. Situations in which α is as large as 0.8−0.9 are outlined in ref 65,

Table 8. Predictive Power of the Obtained Models[a]

	all ligands		ligands with well-predicted affinities			ligands in learning set			
model	rms	r^2	no.	rms	r^2	no.	rms	r^2	outliers
A	1.29	0.18	38	0.81	0.58	23	0.67	0.79	30, 34, 39, 41, 42, 43
B	1.17	0.32	42	0.99	0.46	26	0.90	0.62	30, 34
C	1.12	0.38	41	0.85	0.62	29	0.87	0.69	20, 28, 34
D*	1.12	0.38	41	0.85	0.62	31	1.17	0.40	20, 28, 34

[a] "All ligands" refers to the full set of molecules shown in Table 1. Separate statistics are shown for the subsets of well-predicted compounds (for which the predicted affinity constants had the correct order of magnitude, e.g. $\Delta \ln K < 2$) and for the learning set molecules. no., number of compounds in that subset; rms, root-mean-squared error between predicted and experimental ln K values; r^2, the corresponding correlation coefficents according to (22). Entry D represents a further trial to refine the parameters by adding the outliers of model C to the learning set. Despite extensive MC search, no better parameter set has been found.

and therefore, the 10-fold reduced weighting coefficient of ΔH could be explained within this hypothesis.

While thermodynamic studies of the binding of TR inhibitors are not known to us, microcalorimetric measures of the binding enthalpies for glutathione and analogs to GR[4] show that these are as low as −25 kcal/mol and need to be compensated by a very unfavorable binding entropy contribution in order to reach the ≈−4 kcal/mol of binding free enthalpy. The predicted ligand−TR binding enthalpies in the present study are well within this order of magnitude.

4. Predictive Power of the Model. i. The General Applicability of the Docking Model. In contrast to QSAR models, which mostly apply to a series of related compounds, this approach works in principle with any compound of mass between 200 and 500 g/mol. While calibrated on the basis of polyamino ligands, it was able to predict well the affinity of a different species such as crystal violet **44**, which has been input as a single geometry obtained by MOPAC-AM1[66] minimization (its delocalized electron system cannot be described by the CVFF force field).

ii. Validity of the Models. According to the results in Table 8 and the plots of predicted vs. experimental affinities in Figure 3, the calibrated models were in most cases able to predict the correct order of magnitude of the inhibition constants of a wide variety of compounds. The RMS errors in the prediction of the binding free enthalpy are about 0.6−0.8 kcal/mol, a quite satisfactory result corresponding to the error range of the much more elaborated relative free energy calculations (FEP).[27]

The r^2 values reported in Table 8 are given by

$$r^2 = 1 - \frac{\sum (y_i^{\mathrm{calc}} - y_i^{\mathrm{exp}})^2}{\sum (y_i^{\mathrm{exp}} - \langle y_i^{\mathrm{exp}} \rangle)^2} \qquad (22)$$

where y is the explained variable (i.e. ln K). For molecules in the learning set, the calculated values y^{calc} are obtained from the *cross-validated* regression. For all the other molecules, y^{calc} is the predicted affinity.

Model A has been calibrated within a too narrow structural subspace and did not work for compounds of higher flexibility. The extension of the learning set led to the improved model B. The slight improvement of model C with respect to B is probably not due to the enlarged learning set, but to a better convergence of the optimization. Notably compound **34,** which was an outlier of model B, could not be better predicted after entering the learning set of model C. A prolonged search after adding the outliers of C to the learning set failed to find a better parametrization. The ligand **43**, which is still larger than any one of the latter entries in the learning set, has been well predicted by both models B and C.

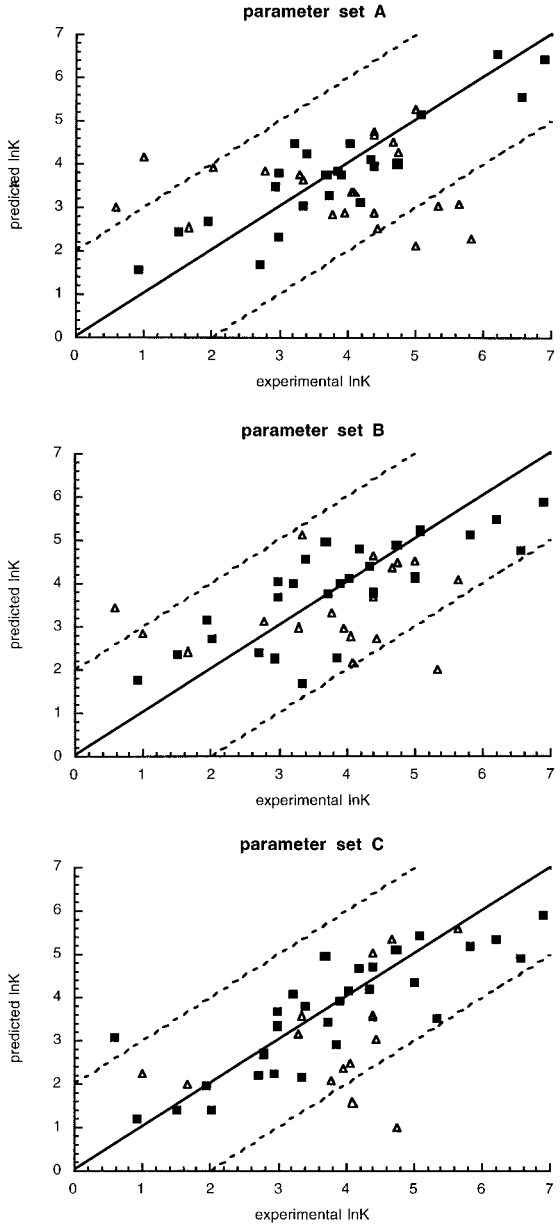

Figure 3. Plots of predicted vs experimental logs of inhibition constants for each set of parameters A, B, and C (Table 4). The molecules used in the learning set are plotted with filled squares while the other are shown as triangles. The dotted lines $y = x - 2$ and $y = x + 2$ are delimiting the "mispredicted" from the "well predicted" compounds for which the calculated inhibition constants has the correct order of magnitude.

The size of the learning set may appear somehow small compared to the number of 5 docking + 4 affinity parameters (out of which only 7 are used in model A).

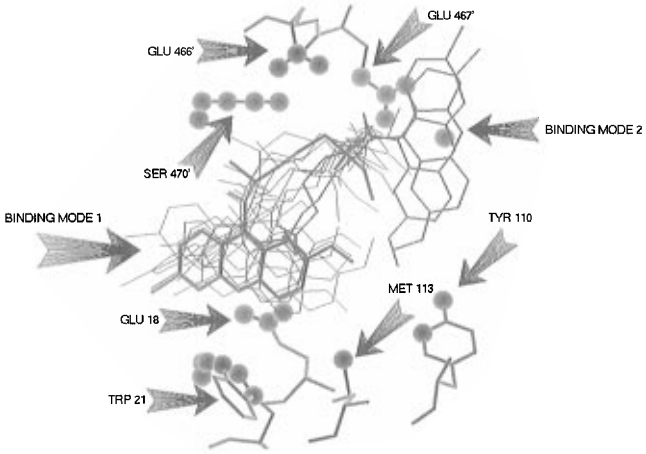

Figure 4. Predicted binding mode of mepacrine (compound **8**) in the TR site, obtained with the parametrization C from Table 4. The conformations within 2 kcal/mol with respect to the best minima (drawn in bold lines) are shown. CPK spheres have been drawn around the atoms of the site that are within 5 Å from the inhibitor molecule in its lowest energy conformation of binding mode 1. Specific TR residues are shown in "stick" representation.

A recent example[67] uses 51 complexes to fit 12 weighting factors. However, the size of the complete validation set (learning set + testing set), on hand of which the performance of the models is monitored, is of 44 compounds. The fact that part of the affinities included in the statistics were predicted for compounds that were not used for calibration should strengthen the confidence in the quality of the model.

The sometimes low r^2 values obtained for the whole set of molecules are partly due to the limited variance of the available experimental data. An extended range of affinity data offers much more confidence[67,68] (*better* r^2 values) in the obtained linear relations, at even *higher* RMS deviations than reported in this work. The difference between "strong" vs "weak" TR inhibitors is of less than 3 orders of magnitude, so that model,[67] with a reported RMS error of 1.14 units of log *K*, would hardly discriminate between them. Our approach has been proven able to do so, but the fact that the regression equations work well within the narrow affinity range used for calibration is not a guarantee for their applicability *outside* this range.

To our knowledge, no TR ligands with inhibition constants lower than 1 μM have been described in literature. The validation of our model outside the already addressed affinity range would therefore imply its generalization to other enzymatic systems. However, some of the introduced simplifications and notably the rigid site hypothesis on which this approach heavily relies are not of general validity. New terms, such as the entropy loss due to side chain rotation hindering,[69] may be required in order to properly describe other enzymatic systems. The application of the model to the highly homologous GR/GSSG system may be straightforward, but would not provide the desired extension of the available affinity range—the best GSSG-competitive GR ligands are micromolar[70] inhibitors as well.

5. Structural Information from the Docking Computations. The comparison of the predicted binding modes of mepacrine (compound **8**) with the recently determined X-ray structure of its TR complex[71] revealed that our docking approach is able to *qualitatively* reproduce the binding modes of this ligand. Figure 4 shows that, within 2 kcal/mol from the lowest energy level, two main binding modes were found by the docking approach.

Binding mode 1 is in good qualitative agreement with the X-ray result which places the aromatic moiety of mepacrine in the hydrophobic pocket defined by the residues W21, M103, and Y110, with the chlorine atom on the side of the W21. Most of the residues that are shown to participate at the binding in the X-ray structure can be found within 5 Å around the docked mepacrine molecules. However, the individual atom-to-atom contacts in the X-ray structure are not quantitatively respected. The charged amino group appears closer (<5 Å) to the glutamates 466′ and 467′, whereas in the X-ray structure a water-mediated interaction with E18 can be seen (in the docked geometry the distance between these groups is 8 Å).

Binding mode 2 comprises an alternative hydrophobic pocket, defined by Phe 396′, Thr 463′, Pro 462′, Val 58, Val 52, Leu 399′, Ile 339, and Ile 106 (not shown in Figure 4). The ammonium group appears to interact with E18. The energy levels corresponding to binding mode 2 are close to the best energies of binding mode 1, while the first binding mode is more populated, which suggests an entropic advantage with respect to the second. Such an explanation should nevertheless be considered with extreme caution.

6. New TR Inhibitors from Virtual Screening of Molecular Data Bases. A molecular data base of 2500 compounds has been screened by this algorithm in order to discover new TR inhibitors. UNIX *csh-scripts* control the procedure, calling the 2D-to-3D conversion program, the energy minimization program Discover,[47] and the docking program. The former and the latter programs are written in FORTRAN 77-Ratfor for Silicon Graphics.

The generation of the conformer families required about 2 weeks on two Silicon Graphics workstations (R4000/100 MHz and R3000/33 MHz). The docking of the 2500 molecules took 96 h on a R4400/175 MHz workstation. It is interesting to note that the use of faster docking approaches in "reciprocal space"[72,73] could accelerate this last step. However, since the bottleneck of this procedure is the conformational sampling step, the time savings resulting from the use of such a technique would be irrelevant.

A "hit list" of 13 molecules (Figure 5) has been selected and submitted to the testing.[63] The obtained results are shown in Table 9. It can be seen that 9 out of the 13 molecules caused a measurable inhibition of TR activity, at concentrations of 57 μM and lower. These are inhibitors of comparable potency to other molecules that were synthesized and published.

Most of the hits show the typical features of TR inhibitors, e.g. the presence of aromatic groups and of (poly)amino chains. Somehow different due to its heterocyclic aromatic moiety and the morpholine group, compound **9** matches nevertheless this typical TR− inhibitor pattern. The virtual screening procedure is able to recognize this pattern in a data base of highly diverse structures. This is a positive result that validates our calibration procedure. However, much faster techniques like neural networks[74] may perform this pattern recognition task on the basis of the molecular topology only. Our procedure, implying a time-consuming conformational analysis, should be therefore able to (a) distinguish between actives and nonactives *within*

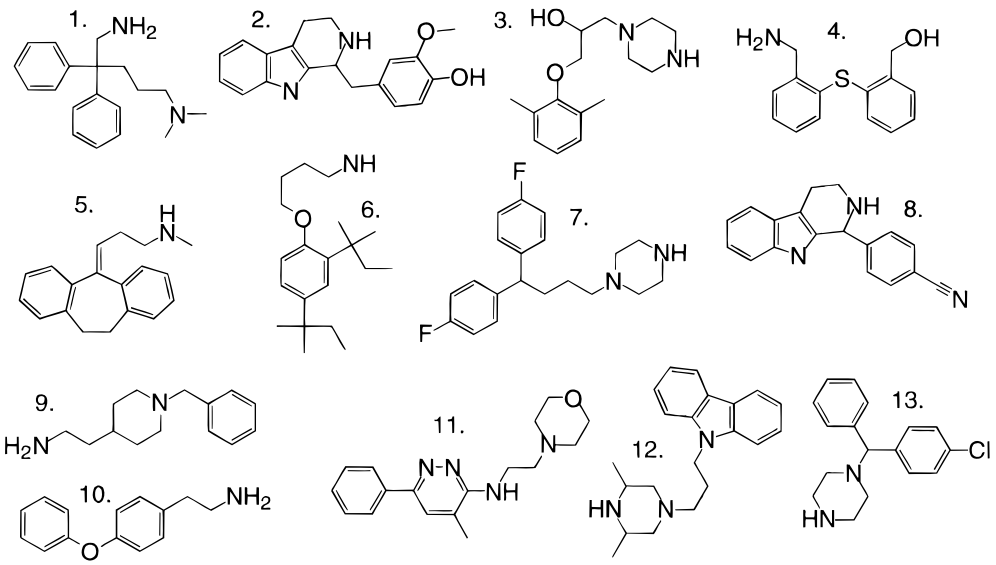

Figure 5. The structures of the "hits" found while screening a molecular data base for putative TR inhibitors. The predicted ln *K* values and the measured percentages of inhibition are listed in Table 9.

Table 9. Predicted Inhibition Constants and the Measured Percentages of Inhibition for the 13 "Hits" Obtained with the Parameter Set A of Table 4[a]

	inhibition percentage			predicted
no.	57 μM	28.5 μM	5.7 μM	K_i (μM)
1	36	9	0	14
2	27	13	0	7
3	7	0	0	14
4	31	0	0	12
5	25	0	0	5
6	0	0	0	15
7	18	9	0	14
8	0	0	0	2
9	0	0	0	5
10	0	0	0	5
11	12	16	18	14
12	23	15	10	85
13	40	30	0	3
ref	70···80	50···60		7(exp)

[a] Numbering (no.) is according to Figure 5. Percentages of inhibition are given as $(r_0 - r_{inh})/r_0 \times 100\%$, where r_0 and r_{inh} are the reduction rates of trypanothione disulfide by TR, in absence and in presence of inhibitor at the given concentrations. In parallel, the working reagents and solutions are tested against the reference inhibitor clomipramine "Ref." (A in Figure 1, $K_{ref} = 7$ μM).

the subset of molecules that match this pattern; this has been confirmed both by the calibration and test phase and by the high success rate of its predictions, and (b) discover *new* structural motifs that provide TR-inhibiting properties.

Molecules **2** and **7** are indeed qualitatively different, their novelty consisting in the absence of a flexible side chain carrying the ammonium group, which is included in a condensed ring system. Molecule **2** may represent a promising lead of a new series of related compounds. Quite analogous compounds have been reported in literature[75] to have trypanocidal activity, without however evidencing their action mechanism. Interestingly, molecule **8**, which has been predicted to be only slightly less active than **2**, had no measurable inhibitory potency under the given conditions. While we do not yet have an explanation for this, it is likely that the presence of the methylene bridge between the tricyclic system and the phenyl ring in A allows the latter to orient so as to minimize the sterical crowding around the ammonium group.

IV. Conclusions

The virtual screening approach reported here is based upon various approximations and hypotheses introduced in order to quantify the binding properties of ligands by means of calculated indexes. These indexes correspond to the virtual binding enthalpy and entropy of the ligand, according to the defined energy functions and sampling criteria. They have been shown to be useful, explaining variables of the logs of the inhibition constants.

Fittable parameters weighting the importance of the different interactions in the model were calibrated with respect to measured affinities of known ligands. While the optimal values of the parameters governing the site−ligand interactions have been found to be in agreement with expectations, the weighting factors of the enthalpic and entropic contributions to the affinity constant behave like empirical QSAR coefficients.

The algorithm was able to explain the affinities of 44 chemically different ligands and to detect new inhibitors. An interesting polycyclic molecule (**2**, Figure 5) could be related to a series of trypanocidal compounds that have not yet been tested with respect to TR-inhibiting properties. While none of the detected compounds are strong TR inhibitors of direct pharmaceutical interest, the latter structure may represent the *lead* of a new series of ligands.

Acknowledgment. D.H. acknowledges financial support from CEREP SA, the Pasteur Institute of Lille and Regional Council of the Nord-Pas de Calais District of France. We are indebted to Glaxo/Welcome France for having prepared the selection of ACD molecules and granted the permit to use the structures of Glaxo-Welcome compounds in this study. Valérie Lucas is acknowledged for having performed the inhibition tests of the selected compounds. Dr. Luise Krauth-Siegel (EMBL Heidelberg) is acknowledged for having provided the X-ray crystallographic results on the binding of mepacrine prior to their publication. Many thanks to André Tartar (Head of the SCBM Lab, Institut Pasteur de Lille), Christian Sergheraert, Elisabeth Davioud (enzymology team of SCBM), Daniel van Belle (UCMB Lab, ULB Brussels), Guy Lippens (NMR Lab, Institut

Pasteur de Lille), Eric Buisine (Molecular Modeling, Institut Pasteur de Lille), Mircea Diudea (Chemistry Department, University of Cluj/Romania) for encouragements and helpful discussions. Steven Brooks (Institut Pasteur de Lille) is acknowledged for carefully proofreading the manuscript.

References

(1) De Castro, S. L. The challenge of Chagas' disease chemotherapy: an update of drugs assayed against Trypanosoma Cruzi. *Acta Trop.* **1993**, *53*, 83−98.
(2) Etah, E. A. O; Smith, K.; Fairlamb, A. H. *Trypanothione detoxication systems in trypanosomatids*, Spring Meeting of the British Society for Parasitology, London, 1993.
(3) Cotgreave, I. A.; Moldeus, P.; Orrenius, S. Host biochemical defense mechanisms against oxidants. *Annu. Rev. Pharmacol. Toxicol.* **1988**, *28*, 189−212.
(4) Janes, W.; Schulz, G. E. Role of the charged groups of glutathione reductase in the catalysis of glutathione reductase: crystallographic and kinetic studies with synthetic analogues. *Biochemistry* **1990**, *29*, 4022−4030.
(5) Jockers-Scheruebl, M. C.; Schirmer, R. H.; Krauth-Siegel, R. L. Trypanothione Reductase from Trypanosoma Cruzi, catalytic properties of the enzyme and inhibition studies with trypanocidal compounds. *Eur. J. Biochem.* **1989**, *180*, 267−272.
(6) Pai, E. F.; Schulz, G. E. The catalytic mechanism of glutathione reductase as derived from X-ray diffraction analyses of reaction intermediates. *J. Biol. Chem.* **1983**, *258*, 1752−1757.
(7) Karplus, P. A.; Schulz, G. E. Refined structure of glutathione reductase at 1.54 Å resolution. *J. Mol. Biol.* **1987**, *195*, 701−729.
(8) Karplus, P. A.; Schulz, G. E. Substrate binding and catalysis by glutathione reductase as derived from refined enzyme-substrate crystal structures at 2 Å resolution. *J. Mol. Biol.* **1989**, *210*, 163−180.
(9) Kuriyan, J.; Kong, X.-P.; Krishna, T. S. R.; Sweet, R. M.; Murgolo, N. J.; Field, H.; Cerami, A.; Henderson, G. B. X-ray structure of trypanothione reductase from Crithidia fasciculata at 2.4 Å resolution. *Proc. Natl. Acad. Sci. U.S.A.* **1991**, *88*, 8764−8768.
(10) Hunter, W. N.; Bailey, S.; Habash, J.; Harrop, S. J.; Helliwell, J. R.; Aboagye-Kwarteng, T.; Smith, K.; Fairlamb, A. H. Active site of trypanothione reductase, a target for rational drug design *J. Mol. Biol.* **1992**, *227*, 322−333.
(11) Bailey, S.; Smith, K.; Fairlamb, A. H.; Hunter, W. N. Substrate interactions between trypanothione reductase and N^1-glutathionylspermidine disulfide at 0.28 nm resolution. *Eur. J. Biochem.* **1993**, *213*, 67−75.
(12) Benson, T. J.; McKie, J. H.; Garforth, J.; Borges, A.; Fairlamb, A. H.; Douglas, K.T. Rationally designed selective inhibitors of trypanothione reductase; phenothiazines and related tricyclics as lead structures. *Biochem. J.* **1992**, *286*, 9−11.
(13) Bruns, R. F.; Simmons, R. M. A.; Howbert, J. J.; Waters, D. C.; Threlkeld, P. C.; Gitter, B. D. *Virtual screening as a tool for evaluating chemical libraries* lecture at Exploiting Molecular Diversity − Small Molecule Libraries For Drug Discovery, Jan. 23−25, 1995, La Jolla, CA.
(14) Smelie, A. S.; Crippen, G. M.; Richards, W. G. Fast drug-receptor mapping by site-directed distances. A novel method of predicting new pharmacological leads. *J. Chem. Inf. Comput. Sci.* **1991**, *31*, 386−392.
(15) Martin, Y. C. 3D database searching in drug design. *J. Med. Chem.* **1992**, *35*, 2145−2154.
(16) Eisen, M. B.; Wiley, D. C.; Karplus, M.; Hubbard, R. E. HOOK: a program for finding new molecular architectures that satisfy the chemical and steric requirements of a macromolecule binding site. *Proteins* **1994**, *19*, 199−221.
(17) Bohacek, R. S.; McMartin, C. Multiple highly diverse structures complementary to enzyme binding sites: results of extensive application of a *de novo* design method incorporating combinatorial growth. *J. Am. Chem. Soc.* **1994**, *116*, 5560−5571.
(18) Boehm, H. J. The computer program LUDI: a new simple method for the de novo ligand design. *J. Comput.-Aided Mol. Des.* **1992**, *6*, 61−78.
(19) Gillet, V.; Johnson, A. P.; Mata, P.; Sike, S.; Williams, P. SPROUT: A program for structure generation. *J. Comput. Aided Mol. Des.* **1993**, *7*, 123−157.
(20) Oprea, T. I.; Waller, C. L.; Marshall, G. R. Three-dimensional quantitative structure-activity relationship of human immunodeficiency virus(I) protease inhibitors. 2. Predictive power using limited exploration of alternate binding modes. *J. Med. Chem.* **1994**, *37*, 2206−2215.
(21) Sternberg, M. J.; King, R. D.; Lewis, R. A.; Muggelton, S. Application of machine learning to structural molecular biology. *Philos. Trans. R. Soc. London B: Biol. Sci.* **1994**, *344*, 365−371.
(22) Benigni, R.; Cotta-Ramusion, M.; Giorgi, F.; Gallo, C. Molecular similarity matrices and quantitative structure-activity relationships: a case study with methodological implications. *J. Med. Chem.* **1995**, *38*, 629−635.
(23) Hocart, S. J.; Reddy, W.; Murphy, W. A.; Coy, D. H. Three-dimensional quantitative structure-activity relationships of somatostatin analogues. 1. Comparative molecular field analysis of growth hormone release-inhibiting potencies. *J. Med. Chem.* **1995**, *38*, 1974−1989.
(24) Diudea, M. V.; Ivanciuc, O. *Molecular Topology* (in Romanian), Comprex Editions; Cluj-Napoca: Romania, 1995.
(25) Ortiz, A. R.; Pisabarro, M. T.; Gago, F.; Wade, R. C. Prediction of drug binding affinities by comparative binding energy analysis. *J. Med. Chem.* **1995**, *38*, 2681−2691.
(26) Hansch, C. *Drug design*; Ariens, E. J., Ed.; Academic Press: New York, 1971; Vol. 16, p 271.
(27) van Gunsteren, W. F. *Methods for calculation of free energies and binding constants. Successes and problems*. In *Computer Simulations of Biomolecular Systems*; van Gunsteren, W. F., Wiener, S. ESCOM: Leiden, 1989; pp 27−59.
(28) Struthers, R. S.; Rivier, J.; Hagler, A. T. Design of peptide analogs: Theoretical simulation of conformation, energetics and dynamics. In *Conformationally Directed Drug Design; Peptides and Nucleic Acids as Templates or Targets*; Vida, J. A., Gordon, M., Eds.; American Chemical Society: Washington, DC, 1984; pp 239−261.
(29) McQuarie, D. A. *Statistical Mechanics*; Harper Collins Publishers: New York, 1976.
(30) Mezei, M.; Beveridge, D. Free energy simulations *Ann. N.Y. Acad. Sci.* **1986**, *482*, 1−23.
(31) Mezei, M. The finite difference thermodynamic integration, tested on calculating the hydration free energy difference between acetone and dimethylamine in water. *J. Chem. Phys*, **1987**, *86*, 7084−7088.
(32) Hodel, A.; Rice, L. M.; Simonson, T.; Fox, R. O.; Brunger, A. T. Proline cis-trans isomerization in staphylococcal nuclease: multisubstrate free energy perturbation calculations. *Protein Sci.* **1995**, *4*, 636−654.
(33) Stouten, P. F. W.; Froemmel, C.; Nakamura, H.; Sander, C. An effective solvation term based on atomic occupancies for use in protein simulations. *Mol. Simul.* **1993**, *10*, 97−120.
(34) Luty, B. A.; Zacharias, M.; Wasserman, Z. R.; Stouten, P. F. W.; Hodge, C. N.; McCammon, J. A. A molecular mechanics/grid method for evaluation of ligand-receptor interactions *J. Comput. Chem.* **1995**, *16*, 454−464.
(35) Baillet, S.; Buisine, E.; Horvath, D.; Maes, L.; Bonnet, B.; Sergheraert, C. 2-Aminodiphenylsulphides as inhibitors of trypanothione reductase. *Bioorg. Med. Chem.* **1996**, *4*, 891−899.
(36) Vajda, S.; Weng, Z.; Rosenfeld, R.; DeLisi, C. Effect of Conformational Flexibility and Solvation on Receptor-Ligand Binding Free Energies. *Biochemistry* **1994**, *33*, 13977−13988.
(37) Moutiez, M.; Lucas, V.; Davioud, E.; Tartar, A.; Sergheraert, C. Indolylmaleimide derivatives: synthesis and biological activities as new potential Trypanothione Reductase inhibitors; The IVth congress of COST-ACRIVAL antiparasitic chemotherapy; Grenada, Spain, June 5−6, 1995.
(38) Moutiez, M. Ph.D. Thesis, University of Lille II, France, 1995.
(39) Krauth-Siegel, R. L.; Lohrer, H.; Buecheler, U. S.; Schirmer, R. H. In *Biochemical Protozoology*; Coombe, G. H., North, M. J., Eds.; Taylor and Francis: London, 1991.
(40) Gomez, R. F.; Moutiez, M.; Aumercier, M.; Bethegnies, G.; Luyckx, M.; Ouaissi, A.; Tartar, A.; Sergheraert, C. 2-Aminodiphenylsulphides as new inhibitors of trypanothione reductase. *Int. J. Antimicrob. Agents* **1995**, *6*, 111−118.
(41) Ponasik, J. A.; Strickland, C.; Faerman, C.; Savvides, S.; Karplus, P. A.; Ganem, B.; Kukoamine A and other hydrophobic acylpolyamines: potent and selective inhibitors of Crithidia fasciculata trypanothione reductase. *Biochem. J.* **1995**, *311*, 371−375.
(42) Girault, S.; Baillet, S.; Lucas, V.; Davioud, E.; Tartar, A.; Sergheraert, C. Bis − Aminodiphenylsulphides as potent inhibitors of the TR from T. Cruzi; poster at the 5th European COST Conference on Antiparasitic Chemotherapy; Heidelberg, May 23−24, 1996.
(43) Moreno, S. N. J.; Carnieri, E. G. S.; Docampo, R. Inhibition of Trypanosoma cruzy trypanothione reductase by crystal violet. *Mol. Biochem. Parasitol.* **1994**, *67*, 313−320.
(44) Howard, A. E.; Kollman, P. A. An analysis of current methodologies for conformational search of complex molecules. *J. Med. Chem.* **1988**, *31*, 1669−1675.
(45) Leach, A. R. An algorithm to identify a molecule's most different conformations. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 661−670.
(46) Leach, A. R.; Prout, K.; Dolata, D. P. The application of artificial intelligence to the conformational analysis of strained molecules. *J. Comput. Chem.* **1990**, *11*, 680−693.
(47) Discover 2.9.0/3.1.0 User Guide, Jan. 1993, BIOSYM Technologies (Molecular Simulations Inc.), San Diego, CA.
(48) Smellie, A.; Teig, S. L.; Towbin, P. Poling: promoting conformational variation. *J. Comput. Chem.* **1995**, *16*, 171−187.
(49) Ermer, O. Calculation of molecular properties using force fields. Applications in organic chemistry. *Struct. Bonding* **1976**, *27*, 161−211.
(50) Hagler, A. T.; Lifson, S.; Dauber, P. Consistent force field studies of intermolecular forces in hydrogen bonded crystals. II. A benchmark for the objective comparison of alternative force fields. *J. Am. Chem. Soc.* **1979**, *101*, 5122−5130.

(51) Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; Vetterling, W. T. *Numerical Recipes: The Art of Scientific Computing*; Cambridge University Press: Cambridge, 1986.
(52) Albert, A.; Serjeant, E. P. *Ionization Constants of Acids & Bases*: Butler & Tanner: London, 1962.
(53) Protein Data Bank; Chemistry Department, Brookhaven National Laboratory, Upton NY 11973.
(54) Gilson, M. K.; Honig, B. The inclusion of electrostatic hydration energies in molecular mechanics calculations. *J. Comput.-Aided Mol. Des.* **1991**, *5*, 5−20.
(55) Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. Semianalytical treatment of solvation for molecular mechanics and dynamics. *J. Am. Chem. Soc.* **1990**, *112*, 6127−6129.
(56) Eisenberg, D.; McLahan, A. D. Solvation energy in protein folding and binding. *Nature* **1986**, *319*, 199−203.
(57) Mehler, E. L.; Eichele, G. Electrostatic effects in water-accessible regions of proteins. *Biochemistry* **1984**, *23*, 3887−3891.
(58) Jackson, J. D. *Classical Electrodynamics*; J. Wiley & Sons: New York, 1975.
(59) Horvath, D.; van Belle, D.; Lippens, G. Development and parametrization of continuum solvent models. II. An unified approach to the solvation problem. *J. Chem.Phys.* **1996**, *105*, 4197−4210.
(60) Brooks, C. L. III; Karplus, M.; Montgomery-Pettitt, B. *Proteins: A theoretical Perspective of Dynamics, Structure and Thermodynamics*; Advances in Chemical Physics vol. LXXI; John Wiley & Sons: New York, 1988.
(61) Brooks, C. L. III. Thermodynamics of ionic solvation: Monte Carlo simulation of aqueous chloride and bromide ions. *J. Phys. Chem.* **1986**, *90*, 6680−6684.
(62) ACD, Available Chemicals Directory, Copyright 1995, MDL Information Systems Inc., San Leandro, CA.
(63) Aumercier, M.; Meziane-Cherif, D.; Moutiez, M.; Tartar, A.; Sergheraert, C. A microplate assay to screen trypanothione reductase inhibitors. *Anal. Biochem.* **1994**, *223*, 161−164.
(64) Horvath, D.; van Belle, D.; Lippens, G.; Wodak, S. J. Development and parametrization of continuum solvent models. I. Models based on the boundary element method. *J. Chem. Phys.* **1996**, *104*, 6679−6695.
(65) Searle, M. S.; Williams, D. H.; Gerhard, U. Partitioning of Free Energy Contributions in the Estimation of Binding Constants: Residual Motions and Consequences for Amide-Amine Hydrogen Bond Strengths. *J. Am. Chem. Soc.* **1992**, *114*, 10697−10704.
(66) Stewart, J. J. P. MOPAC: a semiempirical molecular orbital program. *J. Comput.-Aided Mol. Des.* **1990**, *4*, 1−105.
(67) Head, R. D.; Smythe, M. L.; Oprea, T. I.; Waller, C. L.; Green, S. M.; Marshall, G. R. VALIDATE: A New Method for the Receptor-Based Prediction of Binding Affinities of Novel Ligands. *J. Am. Chem. Soc.* **1996**, *118*, 3959−3969.
(68) Boehm, H. J. The development of a simple empirical scoring function to estimate the binding constant for a protein-ligand complex of known three-dimensional structure. *J. Comput.-Aided Mol. Des.* **1994**, *8*, 243−256.
(69) Pickett, S. E.; Sternberg, M. E. Empirical scale of conformational entropy in protein folding. *J. Mol. Biol.* **1993**, *231*, 2674−2684.
(70) Zoellner, H. *Handbook of Enzyme Inhibitors, Part A*; VCH Verlag GmbH: Weinheim, 1993.
(71) Jacoby, E. M.; Schlichting, I.; Lantwin, C. B.; Kabsch, W.; Krauth-Siegel, R. L. Crystal structure of the Trypanosoma Cruzi trypanothione reductase−mepacrine complex. *Proteins* **1996**, *24*, 73−80.
(72) Katchalski-Katzir, E.; Shariv, I.; Eisenstein, M.; Friesem, A. A.; Aflalo, C.; Vakser, I. A. Molecular surface recognition: determination of geometric fit between proteins and their ligands by correlation techniques. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 2195−2199.
(73) Harrison, R. W.; Kourinov, I. V.; Andrews, L. C. The Fourier-Green's function and the rapid evaluation of molecular potentials. *Protein Eng.* **1994**, *7*, 359−369.
(74) So, S. S.; Karplus, M. Evolutionary optimization in quantitative structure-activity relationship: an application of neural networks. *J. Med. Chem.* **1996**, *39*, 1521−1530.
(75) Cavin, J. C.; Krassner, S. M.; Rodriguez, E. Plant-derived alkaloids active against Trypanosoma Cruzi. *J. Ethnopharmacol.* **1987**, *19*, 89.

JM9603781