*Journal of Computer-Aided Molecular Design*, **12:** 229–240, 1998.
*KLUWER/ESCOM*



# Structure-based design of ligands for protein basic domains: Application to the HIV-1 Tat protein

Anton V. Filikov & Thomas L. James*
*Department of Pharmaceutical Chemistry, University of California, San Francisco, CA 94143-0446, U.S.A.*

Received 25 February 1997; Accepted 1 August 1997

*Key words:* calorimetry, drug design, molecular docking, molecular mimicry, protein structure, thermodynamics

**Summary**

A methodology has been developed for designing ligands to bind a flexible basic protein domain where the structure of the domain is essentially known. It is based on an empirical binding free energy function developed for highly charged complexes and on Monte Carlo simulations in internal coordinates with both the ligand and the receptor being flexible. HIV-1 encodes a transactivating regulatory protein called Tat. Binding of the basic domain of Tat to TAR RNA is required for efficient transcription of the viral genome. The structure of a biologically active peptide containing the Tat basic RNA-binding domain is available from NMR studies. The goal of the current project is to design a ligand which will bind to that basic domain and potentially inhibit the TAR-Tat interaction. The basic domain contains six arginine and two lysine residues. Our strategy was to design a ligand for arginine first and then a superligand for the basic domain by joining arginine ligands with a linker. Several possible arginine ligands were obtained by searching the Available Chemicals Directory with DOCK 3.5 software. Phytic acid, which can potentially bind multiple arginines, was chosen as a building block for the superligand. Calorimetric binding studies of several compounds to methylguanidine and Arg-/Lys-containing peptides were performed. The data were used to develop an empirical binding free energy function for prediction of affinity of the ligands for the Tat basic domain. Modeling of the conformations of the complexes with both the superligand and the basic domain being flexible has been carried out via Biased Probability Monte Carlo (BPMC) simulations in internal coordinates (ICM 2.6 suite of programs). The simulations used parameters to ensure correct folding, i.e., consistent with the experimental NMR structure of a 25-residue Tat peptide, from a random starting conformation. Superligands for the basic domain were designed by joining together two molecules of phytic acid with peptidic and peptidomimetic linkers. The linkers were refined by varying the length and side chains of the linking residues, carrying out BPMC simulations, and evaluation of the binding free energy for the best energy conformation. The dissociation constant of the best ligand designed is estimated to be in the low- to mid-nanomolar range.

## Introduction

Human immunodeficiency virus type 1, (HIV-1), encodes a transactivating regulatory protein, called Tat, which regulates expression of all viral genes by increasing production of mature, full-length viral RNA [1, 2]. Tat acts by binding to a specific RNA target termed the transactivation response element (TAR); TAR consists of a short RNA stem-loop structure found at the 5′-ends of all nascent lentiviral transcripts. Binding of Tat to TAR is thought to lead to the recruitment of cellular proteins and induce viral transcription [3]. Critical features of both the HIV-1 Tat protein and HIV-1 TAR have been delineated [4–11].The specificity of binding is principally determined by residues in a short, highly basic domain of Tat which contains six arginines (GRKKRRQRRR). A ligand which binds tightly to the basic domain of Tat should prevent the Tat protein from binding to TAR RNA, i.e., prevent

* To whom correspondence should be addressed.

the viral genome from transcription. Design of such a ligand would provide a potential drug lead.

Structure-based ligand design is a paradigm for current drug design efforts, as available protein structures from X-ray crystallographic and NMR studies are proliferating. Recently there have been published several structural studies of different fragments of Tat proteins, from various different viruses, alone and complexed with RNA [12–17]. As our lab has determined the structure of a 25-residue Tat peptide capable of inducing HIV-1 replication *in vivo*, we are inclined to use that as a starting point for ligand design [15, 16]. The peptide is a hybrid of the 10-amino acid HIV-1 Tat basic domain with a 15-amino acid segment of the core regulatory domain of another lentiviral Tat protein, i.e. from equine infectious anemia virus (EIAV). A large fraction of the 25-mer peptide assumes a structure in aqueous solution with the lysine- and arginine-rich HIV-1 basic domain being α-helical and the EIAV core domain being separated from the basic domain by a turn and characterized by a nascent helix as well.

There are numerous examples in the literature of designing ligands based on three-dimensional protein structure [see Refs. 18, 19 for rewievs], but all of them deal with rigid globular proteins with a convex surface or binding pocket consisting predominantly of hydrophobic residues. The methods used for these cases cannot be employed for the 25-mer Tat peptide because of the following:

- there is no hollow or pocket of any kind on the surface of the Tat peptide – it is actually an α-helix with a bend [15];
- there is no hydrophobic patch on the surface of the basic domain – on the contrary, it is highly positively charged;
- an NMR study shows that the basic domain of the Tat is structured when incorporated into a hybrid 25-mer [15] but it is flexible in the HIV-1 Tat protein alone [14].

The challenge of the present problem is not unique to the design of a ligand to bind to the Tat basic domain, nor is the method of ligand design we have devised. The method used here, or variants, could be applicable to ligand design for other molecules without convenient hydrophobic pockets and which exhibit flexibility. For example, in addition to basic or acidic domains on proteins, the method could be applied to RNA loops.

The approach employed in the current work was to design a ligand for an arginine first and then to build a superligand for the HIV-1 Tat basic domain by joining the arginine ligands together with peptidic and peptidomimetic linkers. Ligands to be considered for arginine binding were selected from a list generated by searching the Available Chemicals Directory (ACD) with DOCK 3.5 software. Modeling of the conformations of the complexes with the superligand and the basic domain both being flexible has been made possible by carrying out Biased Probability Monte Carlo (BPMC) simulations in internal coordinates (ICM suite of programs [20–22]). The simulations with the complexes used parameters which had ensured correct folding of the 25-mer Tat peptide to its experimental NMR structure from a random starting conformation. The linkers used to build the superligands have been refined by varying the length and the side chains of the constitutive residues. The variants were evaluated by carrying out BPMC calculations followed by calculation of the binding free energy for the conformation with the best energy. Free energies of binding of the superligands were estimated on the basis of an empirical method we developed. Calorimetric measurements of the binding of several similarly highly charged compounds to methylguanidine and small peptides containing arginyl and lysyl residues were performed. These experimental data were used as the basis for predicting binding energies of the superligands.

## Materials and methods

*Database search.* The DOCK [23, 24] software package (version 3.5) was used for searching the ACD and docking the ligands onto the guanidinium group of arginine. The default parameters and force field scoring were used. Instead of a set of sphere centers as commonly used by DOCK as putative ligand atom positions, five points were used for this purpose, the coordinates of which were calculated by placing points in front of the hydrogen atoms of the guanidinium group at a distance typical for a hydrogen bond (1.8 Å).

Independently, the ISIS/Base program, Revision 1.2 (MDL Information Systems, Inc.), was used to search the ACD for compounds which contain two acid groups separated by one, two or three bonds. A number of compounds containing two acidic groups separated by one or two bonds were found, and many were subsequently used in the present study.

*Conformational simulations.* Biased Probability Monte Carlo simulations of conformations of the compounds and their complexes were carried out with the ICM 2.6 suite of programs [20–22]. The idea of the biased probability procedure is to sample with larger probability those regions of the conformational space which we know *a priori* are, on the average, highly populated and to sample with less probability regions known to be less populated. The statistical information about the torsion angle distribution in proteins was derived from the known three-dimensional protein structures and is used in the BPMC sampling procedure [20, 21]. The bias is used for peptides, all the other molecules are sampled completely randomly.

Atomic solvation parameters for oxygens in phosphate and phosphonate groups were taken to be equal to those in the carboxylic group, and a solvation parameter for phosphorus was assumed to be zero. ICM library charges were used for the peptides. The charges for all the other molecules were calculated with Sybyl 6.2 (Tripos, Inc.) using the Gasteiger-Hückel method. The phosphates and phosphonates are not completely deprotonated at neutral pH, especially if they are spatially near one another as in a molecule such as phytic acid [25]. Nevertheless, upon formation of a complex with a highly positively charged ligand, it is reasonable to expect a higher degree of deprotonation. Because of this uncertainty with the formal charges, we carried out all calculations twice: for $-1.5$ and for $-2.0$ formal charge on phosphate and phosphonate groups. Formal charges for all the other acidic groups were each assumed to be $-1$. The ECEPP/3 [26] force field was used for the BPMC simulations. The torsion angles of the six-membered ring of the phytic acid molecule were kept fixed to maintain a 'chair' conformation for the ring in all simulations.

*Isothermal titration calorimetry.* Binding free energies were determined experimentally via titrations of possible ligands with methylguanidine and with the peptides. These calorimetric experiments were performed with the 1–3 mM samples in the microcalorimetry tube at 25 °C in 30 mM potassium phosphate buffer at pH 7.5 with an OMEGA instrument from MicroCal (Amherst, MA) [27]. Heats of dilution for solutions of ligands were determined in separate titrations of ligand into buffer. For all titrations, the heat per injection remained constant throughout such dilution experiments, and a mean value for this heat was subtracted from heats measured in the binding titration. Titration curves were fit by using ORIGIN software (MicroCal) utilizing a model for a single class of binding site.

*Compounds for study.* Experimental thermodynamic studies were carried out with the following compounds. Phytic acid, methylenediphosphonic acid, 3′-azido-3′-deoxythymidine, 1-methylguanidine hydrochloride, 1,2-ethanedisulfonic acid, 3-methylenecyclopropane-*trans*-1,2-dicarboxylic acid, and the peptides Lys-Lys-Lys and Lys-Lys were obtained from Sigma Chemical Company. Other compounds used were anthraquinone-1,8-disulfonic acid (TCI America), *N*,*N*,*N*′,*N*′-ethylenediaminetetra(methylenephosphonic acid) (ICN Biomedicals Inc.), ethylenediphosphonic acid (Johnson Matthey Co.), and Arg-Arg-$CONH_2$ (custom synthesis by the Biomolecular Resource Center at UCSF).

## Results

The goal of this work was to design ligands to bind to the basic domain of the HIV-1 Tat protein, which contains six arginines and two lysines. The approach was to identify ligands potentially capable of binding the guanidino group of arginine, test some of these ligands, and from some of the most promising of these, construct a superligand using peptide-like linkers. Construction of the superligand utilized an NMR solution structure of the Tat basic domain and entailed Monte Carlo simulations to minimize the structure of the ligand/Tat complex and thus aid in design of the linkers. To assess the binding energy of the superligand to the basic domain, an empirical method was developed which is based on the measured binding energies of the smaller ligands to methylguanidine and di- and tripeptides containing Arg and Lys.

*Biased Probability Monte Carlo simulations of conformations of the compounds and their complexes.* In order to choose a BPMC simulation energy function and parameters which insure reasonable modeling of conformations of the complexes, we carried out simulations of the structure of the 25-mer hybrid Tat peptide itself, since there was an NMR structure available for comparison [15]. The sequence of the peptide is as follows:

**YHCQLCFLRSLGIDY** - - **GRKKRRQRRR**

core EIAV domain — basic HIV-1 domain

BPMC simulations were performed with all internal degrees of freedom flexible using a distance-independent Coulomb electrostatic term, standard van der Waals, hydrogen bonding and torsion energy terms and an atomic solvation method to account for solvent effects. It was found that ICM library values for peptide parameters and a dielectric constant value of 4 produced satisfactory results, *viz.* Three simulations with this set of parameters were carried out, each starting from a different random conformation obtained by randomizing the coordinates of the NMR structure; the root-mean-square deviation (rmsd) for the backbone atoms compared to the NMR structure was greater than 7.7 Å for each of the three starting structures. Each sampling consisted of *ca.* $10^7$ conformations, requiring 12 days on a SGI Indigo R4000, 100 MHz. All three runs converged to basically the same fold, an α-helix with a bend, the best energy conformation being rather similar to the experimental one (Figure 1). The rmsd between this structure of lowest energy and the experimental structure for all backbone atoms was 3.79 Å. Most of the differences between the experimental and the simulated structures are concentrated in the EIAV domain: the rmsd for residues 1 to 9 is 2.65 Å, and for residues 1 to 15, it is 3.77 Å. The fragment between residues 10 and 15, inclusive, is unstructured in the simulated conformation and accounts for most of the difference. That region was also the least well-defined in the experimental structure [15]. The conformations of the backbone of the basic domain are the most similar (Figure 1). The rmsd for the backbone atoms of the basic domain (16 to 25 residues) was 1.09 Å and the backbone fold of the residues 16 to 23 is reproduced essentially at the level of experimental structure accuracy: rmsd = 0.93 Å. The orientations of the side chains are not consistent with the averaged experimental structure, which is certainly reasonable; although the peptide backbone is structured, the side chains are expected to be flexible.

The parameters which gave reassuring results for refolding the Tat peptide alone should insure consistency with the experimental structure for all the subsequent simulations of complexes formed between the low molecular weight ligands and Tat peptide receptors. We note that other simulations using various alternative parameter values were unsuccessful in promoting proper peptide folding.

*Design of ligands for arginine.* The basic domain contains six arginines. The strategy employed in the current work was to design a ligand for an arginine first and subsequently to build a superligand for the HIV-1 Tat basic domain by linking the arginine ligands together.

The ACD was searched with the DOCK 3.5 suite of programs to find ligands for the guanidinium group of arginine. All the best scoring ligands contained two or more negatively charged acidic groups to interact with the guanidinium group. Phosphate- or phosphonate-containing compounds scored best, since they are more highly charged than the other acidic groups. A good ligand must be both specific and of high affinity, i.e., it should contain at least two highly charged groups joined by a fairly rigid linker, ensuring the correct orientation of the groups toward the guanidinium group. Several compounds containing two acidic groups separated by one or two bonds were found by searching the ACD with the ISIS/Base program.

A number of the best scoring compounds from the DOCK and ISIS/Base searches were redocked onto the guanidinium group via Monte Carlo minimization with the ligand being flexible, which resulted in rescoring of the ligands. Some of the best scoring compounds following flexible docking are shown in Table 1.

There are several molecules in the ACD which can potentially serve as a ligand for the guanidinium group of arginine, but a number of even better ligands can be designed by attaching two phosphate or phosphonate groups to a fairly rigid linker. The linker can be a three-, four-, or five-membered ring, a two-ring moiety, or similar to a molecular cleft as described in [31].

Phytic acid was one of the best scoring ligands in both the original DOCK search and after Monte Carlo minimization. Phytic acid contains six phosphates attached to a cyclohexane ring, i.e., it could potentially bind up several guanidinium groups. This is one reason why it was chosen as a building block for our subsequent ligand design efforts. The other reason is that phytic acid and its derivatives are known as highly biologically active compounds and bind several proteins with high affinity (see, for instance [32]).

*Development of a method of binding affinity prediction.* In order to design a ligand, a method is needed to assess the affinity of its binding to the receptor. Perturbation methods do not work for highly charged molecules, so we have developed an empirical method which is able to predict the affinity at least semiquantitatively. There are several empirical

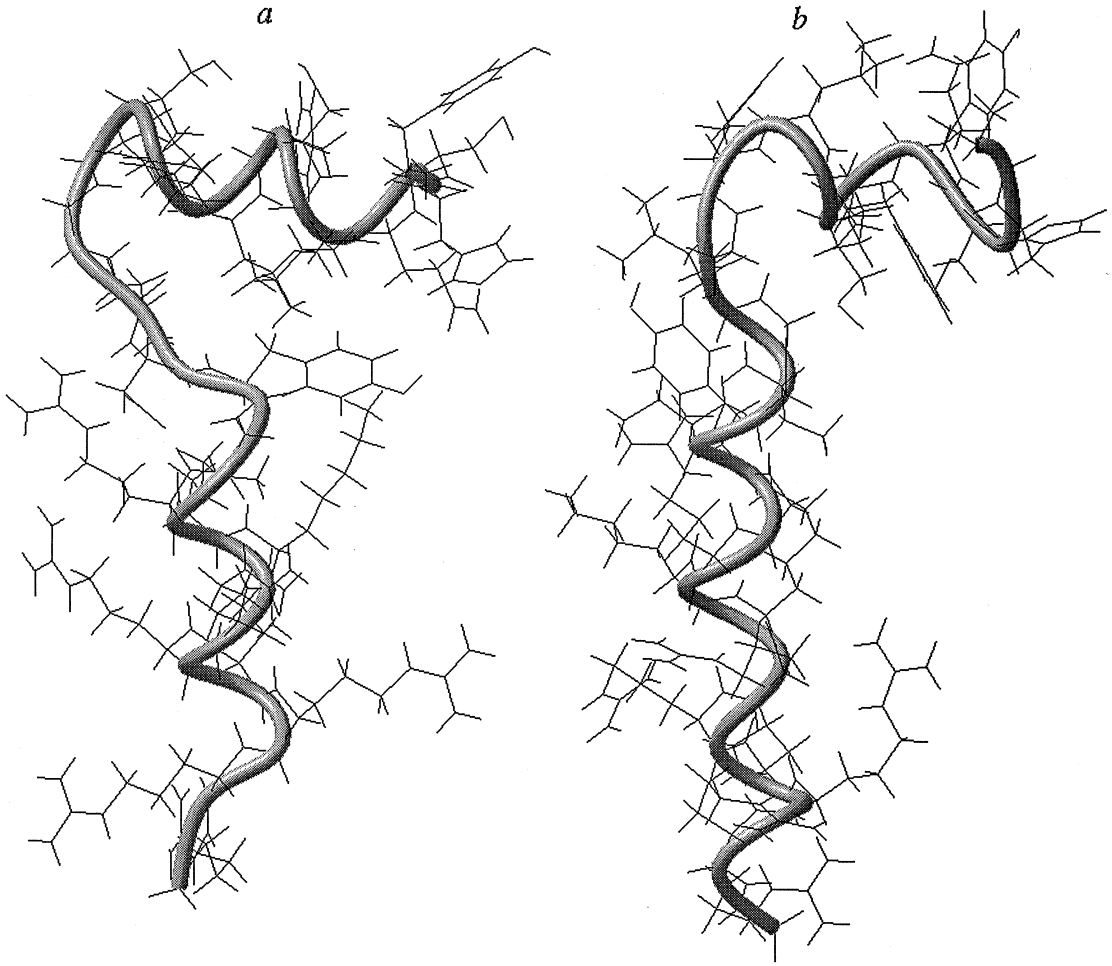


*Figure 1.* (a) The best energy structure of the hybrid 25-mer Tat peptide obtained with a Biased Probability Monte Carlo simulation from a random starting conformation (three runs with sampling of ca. $10^7$ conformations each). (b) An experimental NMR structure of the peptide. Note that the basic domain is in the lower part of the figure.

methods described in the literature [28–30]. However, those empirical methods were developed and tested for ligand-receptor interactions which were predominantly hydrophobic, with little hydrophilic contribution. A method was needed for the particular system studied here, i.e., for the Tat peptide. For the systems we are studying, i.e., complexes of highly charged acidic compounds with arginine, lysine and Arg/Lys-containing peptides, intracomplex electrostatic interactions and desolvation effects contribute much more and therefore require a more precise treatment.

*Theoretical estimates.* Most of the empirical methods described in the literature calculate the free energy of binding on the basis of additive empirical free energy terms evaluating $\Delta G$, the difference of Gibbs free energies between the noncovalent complex and its parts. The free energy terms added are usually the following [28–30]: $\Delta G_{el}$ is an electrostatic interaction term; $\Delta G_{hb}$ is a hydrogen bond term; $\Delta G_{tr/rot}$ is a term accounting for three translational and three rotational degrees of freedom lost by the ligand on binding; $\Delta G_{cratic}$ is a cratic entropy term; $\Delta G_{isomer}$ is a conformational isomerization term which accounts for the change of the molecular internal energies on binding; $\Delta G_{tors}$ is a term which accounts for the loss of bond configurational entropy due to freezing of some torsion angles upon complex formation; and $\Delta G_{solv}$ is a solvent effect term which is usually calculated as a product of a difference between the solvent accessible surface area of the complex and that of the free molecules by a surface tension coefficient (a surface

*Table 1.* Experimental binding free energies of some ligands to methylguanidine (a model for the arginine guanidinium group), Arg-Arg-$CONH_2$, Lys-Lys, and Lys-Lys-Lys peptides measured by isothermal titration calorimetry and the energies as predicted by Equation (3)

| Ligand | Receptor | $\Delta G^o_{kcal/mol}$ | | |
|---|---|---|---|---|
| | | experiment | Eq. (3)[a] | Eq. (3)[b] |
| Anthraquinone1,8-disulfonic acid | Methylguanidine | –2.2±0.5 | –1.1±0.3 | –1.0±0.3 |
| Phytic acid | Methylguanidine | –1.9±0.2 | –1.3±0.2 | –1.3±0.2 |
| Ethylenediphosphonic acid | Methylguanidine | –1.2±0.2 | –1.2±0.3 | –1.3±0.3 |
| N,N,N′,N′-ethylenediamine-tetra(methylenephosphonic acid) | Arg-Arg-$CONH_2$ | –3.6±0.8 | –2.8±0.5 | –2.8±0.5 |
| Methylenediphosphonic acid | Methylguanidine | –1.4±0.03 | –1.4±0.3 | –1.4±0.3 |
| Phytic acid | Arg-Arg-$CONH_2$ | –3.5±0.5 | –3.4±0.6 | –3.5±0.6 |
| Phytic acid | Lys-Lys | –3.7±0.4 | –3.6±0.6 | –3.6±0.6 |
| Phytic acid | Lys-Lys-Lys | –3.2±0.5 | –4.5±0.7 | –4.5±0.7 |
| 3-Methylene-cyclopropane-trans-1,2-dicarboxylic acid | Methylguanidine | > –0.5 | –1.2±0.3 | –1.1±0.3 |
| 1,2-Ethanedisulfonic acid | Methylguanidine | > –0.5 | –1.1±0.3 | –1.1±0.3 |
| 3′-Azido-3′-deoxythymidine | Methylguanidine | > –0.5 | –0.9±0.3 | –0.9±0.3 |

[a] The energies, predicted by Eq. (3), derived for formal charges on phosphates and phosphonates equal to –1.5. The equation coefficients are: $c = -60 \pm 20$, $(1 - a - b) = 74 \pm 11$.

[b] The energies, predicted by Eq. (3), derived for formal charges on phosphates and phosphonates equal to –2. The equation coefficients are: $c = -80 \pm 30$, $(1 - a - b) = 110 \pm 16$.

tension penalty for cavity formation). $\Delta G_{vdW}$, the van der Waals interaction energy, is usually neglected in calculations of binding free energies because it is approximately compensated by the energy gain due to the loss of van der Waals interactions with the solvent on complex formation.

The terms $\Delta G_{tr/rot}$ and $\Delta G_{cratic}$ are generally considered constant. We can therefore designate $\Delta G_{const} = \Delta G_{tr/rot} + \Delta G_{cratic}$, and $\Delta G_{var}$ as the sum of all the other terms listed. The constant terms have been estimated in the literature as 7–11 kcal/mol for $\Delta G_{tr/rot}$ [30] and 2 kcal/mol for $\Delta G_{cratic}$ [29]. The estimate for $\Delta G_{tr/rot}$ is probably reasonable for the gas phase. In solution a reaction leading to complex formation is a complicated process: there are usually solvent molecules bound to the ligand and/or to the receptor. Their release upon complex formation may compensate or overcompensate the loss of six degrees of freedom by the ligand. The changes incurred upon binding may be only partially accounted for by $\Delta G_{solv}$. In our opinion, the constant additive term can best be derived empirically from experimental data for similar complexes, as it probably varies somewhat for different complexes.

Calculation of $\Delta G_{isomer}$ is computationally demanding. The conformational spaces of the free molecules and the complex have to be searched very extensively to ensure that the global minima are reached. Then $\Delta G_{isomer}$ can be calculated as a difference of two fairly large numbers. The estimate is obviously very approximate. Here we utilize the simplest method of estimation: we assume that for similar compounds the term is proportional to the variable part of the binding free energy:

$$\Delta G_{isomer} = a \Delta G_{var}, \tag{1}$$

where coefficient *a* will be incorporated into a linear regression analysis of theoretically calculated $\Delta G_{var}$ versus experimental $\Delta G_{binding}$ data for a series of pertinent complexes (*vide infra*). As we will see, this approach delivers a reasonable correlation, which justifies the proportionality assumption to a certain extent.

The usual way to account for the $\Delta G_{tors}$ term, i.e., the conformational entropy loss, is to elucidate which torsions are on the receptor-ligand interface and add a corresponding value (0.6 kcal/mol or so) to the binding free energy for each of them. This approach overesti-

mates the contribution; many of these torsion angles are already partially frozen in the free compounds and many of them are only partially frozen in the complex. On the other hand, the term is very significant. It can make a major contribution for flexible compounds. Although this kind of estimation for the $\Delta G_{tors}$ term is often used in the literature, in our opinion, it is too rough even for our semiquantitative assessment. Another possible way to account for this conformational entropy loss is to assume that, at first approximation, it is proportional to the receptor-ligand interface surface area. The proportionality coefficient can be estimated from experimental data.

Here we utilize a simpler approximate method. As for $\Delta G_{isomer}$, we assume that for similar compounds the term is proportional to the variable part of the binding free energy:

$$\Delta G_{tors} = b\Delta G_{var}. \quad (2)$$

If the assumption is correct, the contribution of $\Delta G_{tors}$ to binding will be through the coefficient $b$. This seems to be reasonable as a rough approximation: the bigger the receptor-ligand interface, the more torsions become frozen, and the stronger the interaction, the higher the degree of fixation of the torsions. The coefficients $a$ (Equation (1)) and $b$ will be incorporated into linear regression of theoretically calculated $\Delta G_{var}$ versus experimental $\Delta G_{binding}$ data (*vide infra*). Again, as we will see, this approach provides a reasonable correlation.

The electrostatic interaction term ($\Delta G_{el}$) can be calculated by a number of methods including Coulomb electrostatics, distance-dependent electrostatics or Poisson electrostatics. For the current study, $\Delta G_{el}$ is calculated via Coulomb electrostatics.

The typical calculation of the $\Delta G_{solv}$ term as a penalty for cavity formation can give a reasonable result for neutral and predominantly nonpolar molecules, but it is probably an oversimplification for hydrophilic and highly charged molecules. In an extreme case of molecules consisting of only charged atoms, say $Na^+$ and $Cl^-$, there is obviously no penalty for cavity formation. In contrast, when they dissociate, the penalty is negative. A more reasonable estimate can be obtained by accounting for the interaction with the solvent via atomic solvation coefficients [20, 21, 33]. This approach quantitatively takes into account the highly heterogeneous nature of the atoms present in the compounds considered in this study.

The $\Delta G_{hb}$ and $\Delta G_{vdW}$ terms are calculated by ICM2.6 with the ECEPP/3 force field [26] . Although the van der Waals interaction energy is usually neglected in calculations of binding free energies, we consider it important to include it, because this particular set of terms with $\Delta G_{vdW}$ included worked well for simulating the Tat peptide folding (*vide supra*). The very good agreement of the theoretical and experimental conformations provides a convincing reason for using this set of terms for calculation of the free binding energy at least for similar highly charged receptors. Thousands of interatomic interactions calculated with this combination of terms lead to simulating the correct fold of the 25-mer peptide. This means that relative contributions of interatomic interactions are calculated correctly but, in order to use the set of interactions for binding energy calculations, we need to calibrate the absolute values with experimental binding data.

*Relationship between binding energy and constitutive terms.* From the above description of terms, it is apparent that we can write the binding free energy as

$$\Delta G_{binding} = \Delta G_{const} + \Delta G_{var} = \Delta G_{const} + \Delta G_{el} + \Delta G_{solv} + \Delta G_{hb} + \Delta G_{vdW} + a\Delta G_{var} + bG_{var} = [c + \Delta G_{el} + \Delta G_{solv} + \Delta G_{hb} + \Delta G_{vdW}] / [1 - (a + b)] \quad (3)$$

The constants $c$ and $[1 - (a + b)]$ will be obtained from a set of experimental $\Delta G_{binding}$ measurements together with calculated values of $[\Delta G_{el} + \Delta G_{solv} + \Delta G_{hb} + \Delta G_{vdW}]$ for a series of simpler, but similar, set of ligand-receptor interactions.

*Experimental measurements.* In order to derive an empirical binding free energy function, able to predict affinities for basic domain ligands, we performed experimental measurements of affinities to methylguanidine and to Arg-Arg-amide, Lys-Lys and Lys-Lys-Lys peptides for smaller commercially available compounds. These receptor/ligand pairs can model the interaction of bigger charged acidic ligands with the charged basic domain: the ligands contain two or more charged acidic groups and the receptors essentialy comprise constitutive parts of the basic domain. Methylguanidine rather than arginine or argininamide was chosen as the initial model for the peptidic arginyl guanidinium group in order to avoid possible interaction with the hydrophobic moiety of the arginine side chain which would mask the interaction with the guanidinium group. The results of calorimetric measurements of $\Delta G_{binding}$ for these receptor/ligand pairs are listed in Table 1.

*Derivation of an empirical relationship through linear regression of the theoretical estimates of $\Delta G_{var}$ vs. experimental measurements.* Figure 2 displays the correlation of the theoretical estimates of $\Delta G_{var}$ calculated via Equation (3) and the experimental data obtained. The correlation is reasonable enough (correlation coefficient $R^2 = 0.88$) to be used as a tool for semiquantitative prediction of free binding energies for other complexes of similar compounds. A linear regression of the theoretical data, versus the experimental data, by a weighted (by relative experimental error) least squares method gives the following values for the constants $c$ (in kcal/mol) and $[1 - (a + b)]$: –60±20 and 74±11 with formal charges on phosphates and phosphonates equal to –1.5, and –80 ±30 and 110 ±16 with formal charges on phosphates and phosphonates equal to –2.

*Design of ligands for the Tat basic domain.* The basic domain contains six arginines. The strategy employed in the current work was to design a ligand for a guanidino group first and then to build a ligand for the basic domain by linking the arginine ligands together.

As a first step in the design, two molecules of phytic acid were docked onto the 25-mer hybrid Tat peptide [15] by means of a BPMC conformational search in internal coordinates using ICM2.6. Initially the conformation of the peptide was kept rigid until the two phytic acids found their binding sites on the basic domain. Then the side chain torsion angles of the basic domain (residues 16 to 25) were unfixed during subsequent BPMC calculations, allowing adjustment of the side chains. This resulted in the formation of two well-determined phytic acid binding sites approximately on opposite sides of the basic domain. One site encompasses four arginines, the other consists of two arginines and two lysines.

Next, the molecules of phytic acid were connected via a peptide linker of approximately suitable length, -Gly-Gly-Gly-Gly-, and another BPMC run was carried out with the ligand torsions and the basic domain side chain torsions being flexible. The expected binding free energies of the best energy conformation were calculated using the constants extracted from the linear regression analysis presented above with Equation (3).

A number of modifications of the linker were tested by running a limited BPMC search with each variant, consisting of about 2 to 4 million conformations, with 12 to 14 loose distance restraints which kept the ligand close to the desired binding site and significantly confined the conformational space to be searched. Table 2 lists some of the linker modifications tested. Binding free energies for the best energy conformations of the ligands designed were calculated using Equation (3). Results are given in Table 2. The goal was to get a linker as short as possible while still ensuring proper orientation of the two phytic acids and the target side chain groups, i.e., encouraging good affinity. A number of different types of amino acids constituting the linker were tested (Table 2) with the aim of improving affinity by filling nearby cavities with a moiety of a suitable geometry, polarity and charge. Since peptides are highly susceptible to hydrolysis *in vivo*, three linkers with modified backbones (peptide mimetics) were also tested (Table 2).

The best energy conformation of a complex of the Tat peptide with the best affinity ligand, #14 in Table 2, is depicted in Figure 3. This ligand has a peptidomimetic linker. It is also the shortest ligand designed in terms of the number of bonds, but it clearly permits the designed ligand to wrap around the basic domain with a nice molecular fit enabling several energetically favorable interactions. The first phytic acid moiety is located at the binding site formed by K18, K19, Q22 and R25. There are four hydrogen bonds formed between the phosphate oxygens of this moiety and the $\epsilon$-amino groups of K18 and K19. We count it as a hydrogen bond if the distance between an H-bond donor and acceptor is $<$2.8 Å. The R25 guanidino group lies within H-bonding range near two of the phosphoryl groups on that phytic acid moiety as well, but with the current structure, the angle may preclude classical H-bonding; it would appear that charge neutralization is well served by this arrangement at the very least. In other words, all five of the phosphoryl groups on this phytic acid interact with basic side chains. The other phytic acid moiety binds to R17, R20, R23 and R24. The phosphate oxygens form two hydrogen bonds with R23 and R24 and two bifurcated hydrogen bonds with R20 and R21. So, four of the five phosphoryl groups interact with basic side chains. The linker clearly also contributes to the stability by forming the following hydrogen bonds: the backbone carbonyls form one regular and two bifurcated bonds with R17, R21 and R24, and the side chain carboxyls form two bifurcated hydrogen bonds with R17 and R24. There are 19 hydrogen bonds formed between the ligand and the Tat peptide with an average length of 2.26 Å.

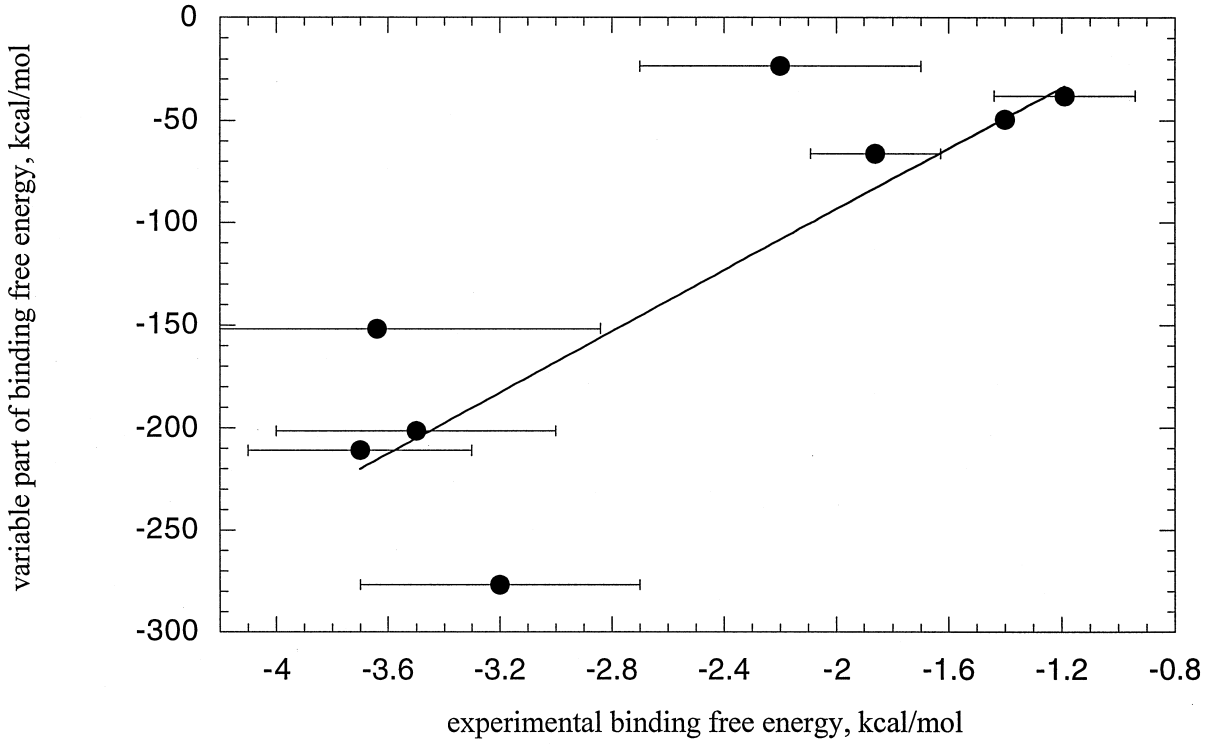


*Figure 2.* Correlation of the theoretical estimates of $\Delta G_{var} = \Delta G_{el} + \Delta G_{hb} + \Delta G_{vdW} + \Delta G_{solv}$ from Equation (3) and the experimental data. Formal charges on phosphates and phosphonates are equal to –1.5. The line is a linear regression fit weighted by a relative experimental error.

**Discussion and conclusions**

Extant structure-based ligand design strategies are largely based on finding ligands to fit into hydrophobic pockets on proteins. To target the Tat protein with hopes of disrupting the Tat/TAR interaction necessary for HIV-1 replication, we are presented with the basic domain of the protein as the logical target site. Recently, we have determined the solution structure of a biologically active Tat peptide in which the basic domain is α-helical, suggesting that the basic domain may assume this structure either with the Tat protein in solution or when it binds to TAR. Using this structure as the basis for computer-aided ligand design presents us with the challenge of targeting a strongly charged protein moiety with no obvious pockets for a ligand. Here we have presented a strategy to approach such problems in general, and we have demonstrated it in the specific case of the Tat basic domain.

Basically, the idea is to choose small molecules which will bind to particular residues in the target and then to link the liganding molecules. This overall strategy is certainly not novel, but our implementation is. Furthermore, the means of implementation can be generalized for ligand design entailing other targets. The small molecules we chose on the basis of a DOCK search of the ACD to obtain a list of compounds capable of binding the guanidino group of an arginyl residue, followed sequentially by optimization of the binding by flexible docking Monte Carlo calculations and experimental calorimetry measurements of the most promising of the small molecule ligands to methylguanidine and some di- and tripeptides of lysine and arginine. The experimental binding data were used together with computed binding free energy terms to predict the binding free energy of ligands designed with two of the small liganding molecules targeted for separate, but proximate, sites on the target protein. Various linkers were tested using the two liganding moieties, with Monte Carlo optimization of the structure of the entire complex. The linking moieties were chosen such that the linker could interact with the target peptide as well to improve stability.

The ligand design described here targeted the arginine residues essential for the HIV-1 Tat interaction with TAR RNA. The development fairly early selected

*Table 2.* Free energies of binding calculated with Equation (3) for the best energy conformations of the complexes of the TAT peptide for various ligands designed for the basic domain

| # | Ligand [ab] | $\Delta G_{binding}$ (kcal/mol) calculated with a formal charge of –1.5 | $\Delta G_{binding}$ (kcal/mol) calculated with a formal charge of –2 |
|---|---|---|---|
| 1 | Two molecules of phytic acid (considered as one ligand), no linker $C_6H_6P_6O_{24}$+ $C_6H_6P_6O_{24}$ | –5.8±0.9 | –5.7±0.9 |
| 2 | $C_6H_6P_6O_{24}$-(gly)-(gly)-(gly)-(gly)-$C_6H_6P_6O_{24}$ | –8.6±1.3 | –8.0±1.2 |
| 3 | $C_6H_6P_6O_{24}$-(gly)-(gly)-(asp)-(gly)-$C_6H_6P_6O_{24}$ | –9.2±1.4 | –8.5±1.3 |
| 4 | $C_6H_6P_6O_{24}$-$PO_3$-CO-(D-ala)-(glu)-(gly)-$C_6H_6P_6O_{24}$ | –8.0±1.2 | –7.6±1.1 |
| 5 | $C_6H_6P_6O_{24}$-$PO_3$-CO-(D-leu)-NHCHCO-(D-phe)-$C_6H_6P_6O_{24}$ <br> $CH_2$ <br> $CH_2$ <br> $PO_4$ | –9.3±1.4 | –8.3±1.3 |
| 6 | $C_6H_6P_6O_{24}$-PO3-CO-(gly)-NHCHCO-NHCHCO-$C_6H_6P_6O_{24}$ <br> $CH_2$ $PO_4$ <br> $CH_2$ <br> $PO_4$ | –8.6±1.3 | –7.9±1.2 |
| 7 | $C_6H_6P_6O_{24}$-CO-(gly)-NHCHCO-NHCHCO-$C_6H_6P_6O_{24}$ <br> $CH_2$ $PO_4$ <br> $CH_2$ <br> $PO_4$ | –8.6±1.3 | -7.9±1.2 |
| 8 | $C_6H_6P_6O_{24}$-CO-(gly)-NHCHCO-NHCHCO-$C_6H_6P_6O_{24}$ <br> $CH_2$ $PO_3$ <br> CH <br> ⋀ <br> $O_3P$ $PO_3$ | –8.2±1.2 | –7.9±1.2 |
| 9 | $C_6H_6P_6O_{24}$-CO-(gly)-NHCHCO-NHCHCO-$C_6H_6P_6O_{24}$ <br> $PO_3$ <br> $CH_2$ <br> $PO_3$ | –7.9±1.2 | –7.9±1.2 |
| 10 | $C_6H_6P_6O_{24}$-(gly)-(ala)-(asp)-(gly)-$C_6H_6P_6O_{24}$ | –8.3±1.3 | –8.2±1.2 |
| 11 | $C_6H_6P_6O_{24}$-(gly)-(gly)-(asp)-(ser)-$C_6H_6P_6O_{24}$ | –9.1±1.4 | –7.8±1.2 |
| *Linkers with a modified peptide backbone (peptidomimetics):* | | | |
| 12 | $C_6H_6P_6O_{24}$-($COCH_2NH$)-(COCHNH)-(COCHCO)-$C_6H_6P_6O_{24}$ <br> $CH_2$ $CH_2$ <br> COO OH | –9.1±1.4 | –8.2±1.2 |
| 13 | $C_6H_6P_6O_{24}$-($COCH_2NH$)-(COCHNH)-(COCHCO)-$C_6H_6P_6O_{24}$ <br> COO $CH_2$ <br> COO | –9.2±1.4 | –8.5±1.3 |
| 14 | $C_6H_6P_6O_{24}$-($COCH_2NH$)-(COCHNH)-(COCHCO)-$C_6H_6P_6O_{24}$ <br> COO COO | –9.7±1.5 | –9.0±1.4 |

[a] All acidic groups are shown deprotonated.

[b] -NHCHCO- represents an L-amino acid residue, and -NHCHCO- represents a D-amino acid residue.

guanidinium-oriented ligands containing at least two phosphate groups with further development of Tat peptide ligands based on phytic acid as a viable ligand for individual guanidinium groups in the basic domain. However, a variety of other ligands for the guanidinium group, including highly specific ligands designed on the basis of molecular cleft type compounds [31], could be used as building blocks for the superligand. Furthermore, the approach developed here could also be used to target three arginines, or two arginines

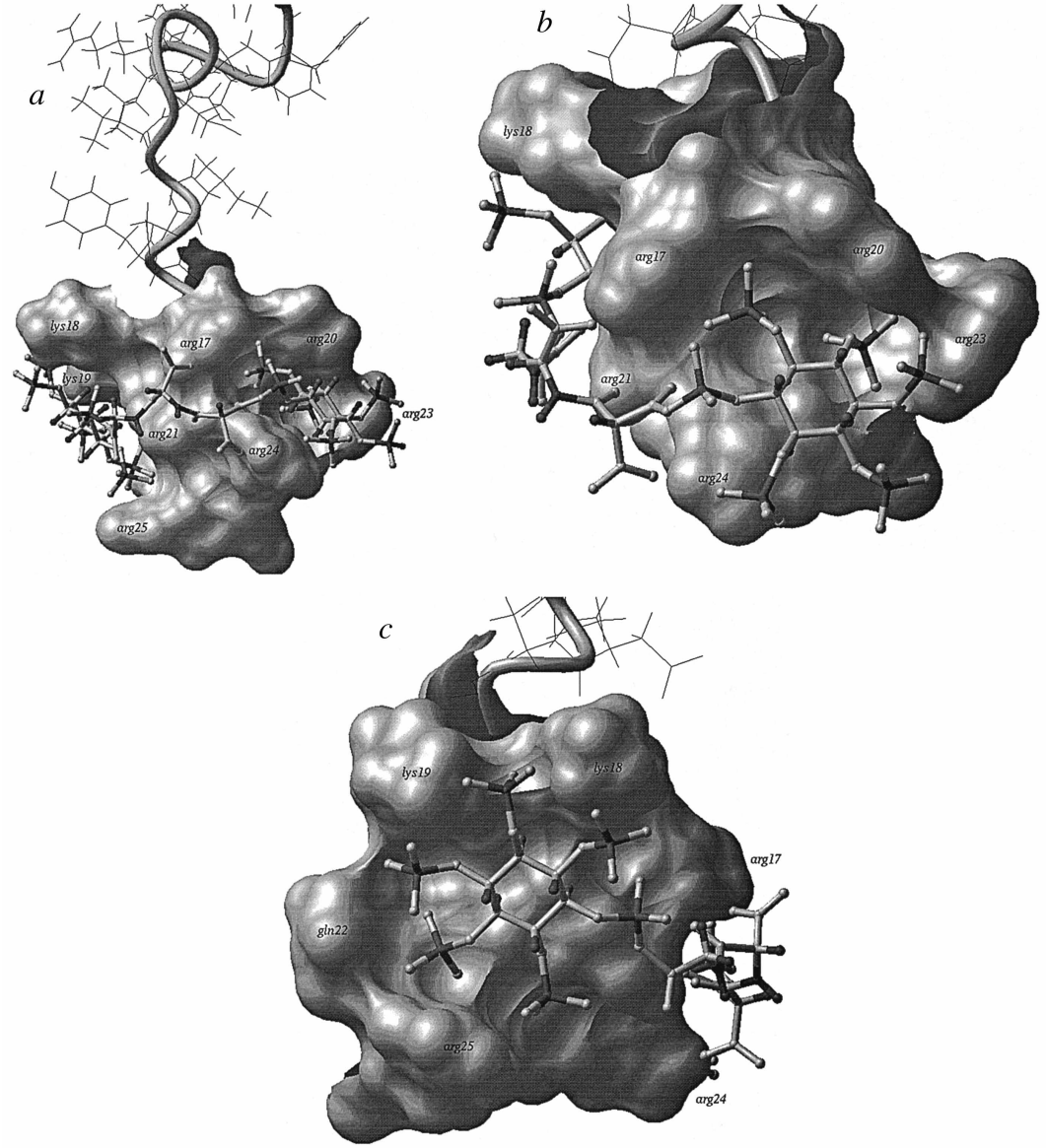


*Figure 3.* (a–c) The best energy conformation of a complex of the 25-mer hybrid Tat peptide with the ligand designed with best affinity (#14 in Table 2).

and a lysine, in the basic domain. Scaffolds other than phytic acid could readily be used; indeed some phosphates could possibly be replaced with other acidic groups. An alternative approach could be the use of Schiff base-forming compounds, which could react with arginine or lysine, with a suitable linker.

In the current work we confined the design only to ligands which would not distort the spatial structure of the basic domain very much. Of course, another approach is possible, i.e., design of ligands which would interact not only with the surface groups but with hydrophobic parts of the side chains and would be able to completely unfold the basic domain and form a complex using stronger hydrophobic interactions.

## Acknowledgements

This work was supported by NIH grant GM39247. We thank Dr. R. Abagyan (NYU) for valuable advice and making the ICM2.6 software available to us, Drs. I.D. Kuntz and C. Oshiro (UCSF) for expert advice in using

DOCK 3.5, Drs. R.H. Shafer and P. Scaria (UCSF) for instruction and use of the OMEGA microcalorimeter, and Drs. P.A. Kollman and N.B. Ulyanov (UCSF) for helpful advice. The facilities of the UCSF Computer Graphics Laboratory, supported by NIH Grant RR01081, are gratefully acknowledged.

## References

1. Kao, S.Y., Calman, A.F., Luciw, P.A., and Peterlin, B.M., Nature, 330 (1987) 489.
2. Marciniak, R.A. and Sharp, P.A., EMBO J., 10 (1991) 4189.
3. Carroll, R., Peterlin, B.M. and Derse, D., J. Virol., 66 (1992) 2000.
4. Calnan, B.J., Tidor, B., Biancalana, S., Hudson, D. and Frankel, A.D., Science, 252 (1991) 1167.
5. Derse, D., Carvalho, M., Carroll, R. and Peterlin, B.M., J. Virol., 65 (1991) 7012.
6. Hamy, F., Asseline, U., Grasby, J., Iwai, S., Pritchard, C., Slim, G., Butler, P.J., Karn, J. and Gait, M.J., J. Mol. Biol., 230 (1993) 111.
7. Hauber, J., Malim, M.H. and Cullen, B.R., J. Virol., 63 (1989) 1181.
8. Karn, J., Dingwall, C., Finch, J.T., Heaphy, S. and Gait, M.J., Biochimie, 73 (1991) 9.
9. Loret, E.P., Georgel, P., Johnson Jr., W.C. and Ho, P.S., Proc. Natl. Acad. Sci. USA, 89 (1992) 9734.
10. Puglisi, J.D., Chen, L., Frankel, A.D. and Williamson, J.R., Proc. Nat. Acad. Sci. USA, 90 (1993) 3680.
11. Weeks, K.M., Ampe, C., Schultz, S.C., Steitz, T.A. and Crothers, D.M., Science, 249 (1990) 1281.
12. Puglisi, J.D., Chen, L., Blanchard, S. and Frankel, A.D., Science, 270 (1995) 1200.
13. Aboul-ela, F., Karn, J. and Varani, G., J. Mol. Biol., 253 (1995) 313.
14. Willbold, D., Rosin-Arbesfeld, R., Sticht, H., Frank, R. and Rösch, P., Science, 264 (1994) 1584.
15. Mujeeb, A., Parslow, T.G., Yuan, Y.-C. and James, T.L., J. Biomol. Struc. Dyn., 13 (1996) 649.
16. Mujeeb, A., Bishop, K., Peterlin, B.M., Turck, C., Parslow, T.G. and James, T.L., Proc. Natl. Acad. Sci. USA, 91 (1994) 8248.
17. Ye, X., Kumar, R.A. and Patel, D.J., Chem. Biol., 2 (1995) 827.
18. Bohacek, R.S., McMartin, C. and Guida, W.C., Med. Res. Rev., 16 (1996) 3.
19. Verlinde, C.L. and Hol W.G., Structure, 2 (1994) 577.
20. Abagyan, R., Totrov, M. and Kuznetsov, D., J. Comput. Chem., 15 (1994) 488.
21. Abagyan, R. and Totrov, M., J. Mol. Biol., 235 (1994) 983.
22. ICM2.6 is available from Molsoft, L.L.C., Metuchen, NJ 08840.
23. Meng, E.C., Shoichet, B.K. and Kuntz, I.D., J. Comput. Chem., 13 (1992) 505.
24. Shoichet, B.K., Bodian, D.L. and Kuntz, I.D., J. Comput. Chem., 13 (1992) 380.
25. Mernissiarifi, K., Schmitt, L., Schlewer, G. and Spiess, B., Anal. Chem., 67 (1995) 2567.
26. Momany, F.A., McGuire, R.F., Burgess, A.W. and Scheraga, H.A., J. Phys. Chem., 79 (1975) 2361.
27. Wiseman, T., Williston, S., Brandts, J.F. and Lin, L.N., Anal. Biochem., 179 (1989) 131.
28. Verkhivker, G., Appelt, K., Freer, S.T. and Villafranca, J.E., Protein Eng., 8 (1995) 677.
29. Novotny, J., Bruccoleri, R.E. and Saul, F.A., Biochemistry, 28 (1989) 4735.
30. Page, M.I. and Jencks, W.P., Proc. Natl. Acad. Sci. USA, 68 (1971) 1678.
31. Bell, T.W. and Liu, J., Angew. Chem. Int. Ed. Engl., 29 (1990) 923.
32. Lemmon, M.A., Ferguson, K.M., O'Brien, R., Sigler, P.B. and Schlessinger, J., Proc. Natl. Acad. Sci. USA, 92 (1995) 10472.
33. Wesson, L. and Eisenberg, D., Protein Sci., 1 (1992) 227.