PROTEINS: Structure, Function, and Genetics 37:88–105 (1999)

Exhaustive Docking of Molecular Fragments With Electrostatic Solvation

Nicolas Majeux,[1] Marco Scarsi,[1] Joannis Apostolakis,[1] Claus Ehrhardt,[2] and Amedeo Caflisch[1]*
[1]*Department of Biochemistry, University of Zürich, Zürich, Switzerland*
[2]*Novartis Pharma AG, Basel, Switzerland*

ABSTRACT **A new method is presented for docking molecular fragments to a rigid protein with evaluation of the binding energy. Polar fragments are docked with at least one hydrogen bond with the protein while apolar fragments are positioned in the hydrophobic pockets. The electrostatic contribution to the binding energy, which consists of screened intermolecular energy and protein and fragment desolvation terms, is evaluated efficiently by a numerical approach based on the continuum dielectric approximation. The latter is also used to predetermine the hydrophobic pockets of the protein by rolling a low dielectric sphere over the protein surface and calculating the electrostatic desolvation of the protein and van der Waals interaction energy. The method was implemented in the program SEED (solvation energy for exhaustive docking). The SEED continuum electrostatic approach has been successfully validated by a comparison with finite difference solutions of the Poisson equation for more than 2,500 complexes of small molecules with thrombin and the monomer of HIV-1 aspartic proteinase. The fragments docked by SEED in the active site of thrombin reproduce the structural features of the interaction patterns between known inhibitors and thrombin. Moreover, the combinatorial connection of these fragments yields a number of compounds that are very similar to potent inhibitors of thrombin. Proteins 1999;37:88–105.**

Key words: combinatorial ligand design; docking; thrombin; electrostatic solvation; SEED

INTRODUCTION

Computer-aided structure-based ligand design methods are useful tools for de novo design and lead modification.[1–3] The combinatorial strategy we have chosen for structure-based ligand design consists of three parts: the docking of small molecular fragments, the connection of the docked fragments by combinatorial principles to generate candidate ligands, and the estimation of relative binding affinities.[4] In this article we mainly address the first part of our strategy and propose a new approach for exhaustively determining positions and orientations of small-to-medium-sized molecular fragments in the binding site of a rigid protein and ranking them according to their binding energy. Polar fragments are positioned such that at least one hydrogen bond with optimal distance to a protein polar group is made. For the docking of apolar fragments a novel procedure has been developed to select in an accurate and efficient way the hydrophobic regions of the protein, i.e., those with low electrostatic desolvation and favorable van der Waals interactions with an uncharged probe sphere. Furthermore, our numerical continuum electrostatic methodology[5,6] and ad hoc look-up tables are employed to efficiently evaluate the protein and fragment desolvation upon binding and the screened electrostatic interaction. The approach is implemented in the computer program SEED (solvation energy for exhaustive docking).

The second part of our ligand design strategy deals with the connection of the docked molecular fragments and is performed by a procedure implemented in the program for computational combinatorial ligand design (CCLD).[7] Although CCLD performs an exhaustive search, it is efficient because of the precomputation of three lists containing the bonding fragment pairs, the overlapping fragment pairs, and the overlapping linker-fragment pairs. The third part of our combinatorial approach requires an efficient and accurate method to predict relative binding free energies and is not addressed in this report.

We show that the fragment docking method presented in this study determines in an efficient way all of the relevant binding modes in the binding site of thrombin for a series of fragments that are representative of known inhibitors. Furthermore, the SEED-CCLD approach generates candidate ligands which have the same interaction patterns with thrombin as a number of potent inhibitors.

MATERIALS AND METHODS

The docking approach implemented in the program SEED determines optimal positions and orientations of small-to-medium-sized molecular fragments in the binding site of an enzyme or receptor which will be hereafter referred to as receptor. The fragment types are docked in the order specified by the user. After each placement of a given fragment the binding energy is estimated. The successive fragment type is docked after all placement–energy evaluations of the preceding fragment type have been made. The procedure for the placement of the frag-

Grant sponsor: Swiss National Science Foundation (Nationalfonds); Grant number: 31-53604.98; Grant sponsor: Novartis Pharma Inc.
*Correspondence to: Amedeo Caflisch, Department of Biochemistry, University of Zürich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland. E-mail: caflisch@bioc.unizh.ch.
Received 9 February 1999; Accepted 30 April 1999

TABLE I. List of Input Parameters and Values Used for Thrombin†

Parameter	Value
Maximal angular deviation from ideal hydrogen bond geometry	50°
Number of additional vectors around each ideal hydrogen bond vector	100
Total number of polar vectors used	3000
Point density of SAS for generating apolar vectors on the receptor	1.0 Å^{-2}
Point density of SAS for generating apolar vectors on the fragment	1.0 Å^{-2}
Total number of apolar vectors used	100
Number of rigid body rotations of the fragment around each alignment vector	36
Factor for maximal number of clashes (it is multiplied by the number of atoms in the fragment)	2.0
Factor for severe overlap	0.6
Radius of the pseudo-sphere for evaluation of van der Waals neighbor list	10.0 Å
Size of the grid used for generation of non-bonding pair list (van der Waals and screened interaction)	1.0 Å
Grid boundary for solvation energy	8.0 Å
Grid size for solvation energy	0.5 Å
Number of points per sphere to generate SAS	500
Dielectric value of the solvent	78.5
Dielectric value of the solute	1.0
Threshold of van der Waals interaction between an apolar fragment and the receptor for evaluation of electrostatic contribution	−0.33 kcal/(mol atom)

†Besides the parameters listed here the user can specify a sphere (coordinates of the center and radius) for acceptance of the fragment.

ments is presented first. Then the estimation of the binding energy is described in detail. Particular emphasis is given to the continuum electrostatic approach used for both the selection of hydrophobic regions in the receptor and the evaluation of the electrostatic solvation energy. Finally, the sorting and clustering algorithms are briefly explained.

Fragment Docking

The binding site of the receptor is defined by a list of residues which are selected by the user. Fragments are considered polar if they have at least one H-bond donor or acceptor. Because of this definition some "polar" fragments can have considerable hydrophobic character (e.g., diphenylether). Therefore, they can also be docked by the procedure for apolar fragments if specified by the user. Table I lists the input parameters.

Docking of polar fragments

SEED docks polar fragments in positions of the receptor where at least one hydrogen bond with good geometry is made. First, predefined rules (Fig. 1) allow the distribution of vectors of unitary length on all H-bond groups of the fragment in a direction for an ideal H-bond geometry. For example, if a nitrogen atom is bound to two heavy atoms, one H-bond vector is generated in the direction of either the lone pair (Fig. 1a) or the NH bond (Fig. 1b). The same procedure is then used for the polar groups in the receptor binding site. Vectors for metal ions have to be provided by the user. For the receptor polar groups and metal ions an additional set of vectors is distributed uniformly on a spherical region around each of the ideal directions to increase the spatial sampling. To discard receptor vectors that point into a region of space occupied by other atoms of the protein and select preferentially vectors in the concave regions of the receptor a spherical probe is set on the vector extremity at a distance corresponding to the sum of the van der Waals radii of the acceptor or donor atom and the probe. The van der Waals interaction between the probe and all the receptor atoms is then evaluated except for the receptor hydrogen atom involved in the H-bond. The vectors which show less favorable van der Waals energies are discarded. Finally, the docking itself is achieved by matching a H-bond vector of the receptor with a H-bond vector of the fragment at a distance that depends on the atom types of donor and acceptor involved in the hydrogen bond. The fragment is then rotated around the H-bond axis to increase sampling.

Docking of apolar fragments

A novel method was developed for the docking of apolar fragments into hydrophobic regions of the receptor. First, a number of points are distributed uniformly on the solvent-accessible surface (SAS) of the fragment. The density of surface points for the fragment and the receptor (see below) are user-controlled parameters. Second, an automatic procedure defines the hydrophobic regions on the receptor. For this purpose a number of points are uniformly distributed on the SAS of the binding site. A low dielectric sphere is placed on each of these points, and the receptor desolvation energy (see below) and the probe/receptor van der Waals interaction are evaluated. The points on the receptor SAS are then ranked according to the sum of the two energy terms, and the best n (where n is an input parameter) are selected for docking. For both the fragment and the receptor, vectors are defined by joining each point on the SAS with the corresponding atom center. Finally, apolar fragments are docked by matching a vector of the fragment with a vector of the receptor at the optimal van der Waals distance. To improve sampling additional rotations of the fragment are performed around the axis joining the receptor atom and fragment atom.

For both polar and apolar fragments, the docking is exhaustive on a discrete space. The discretization originates from the limited number of preferred directions and rotations around them. Fragment symmetries are checked only once for every fragment type and are exploited to increase the efficiency in docking.

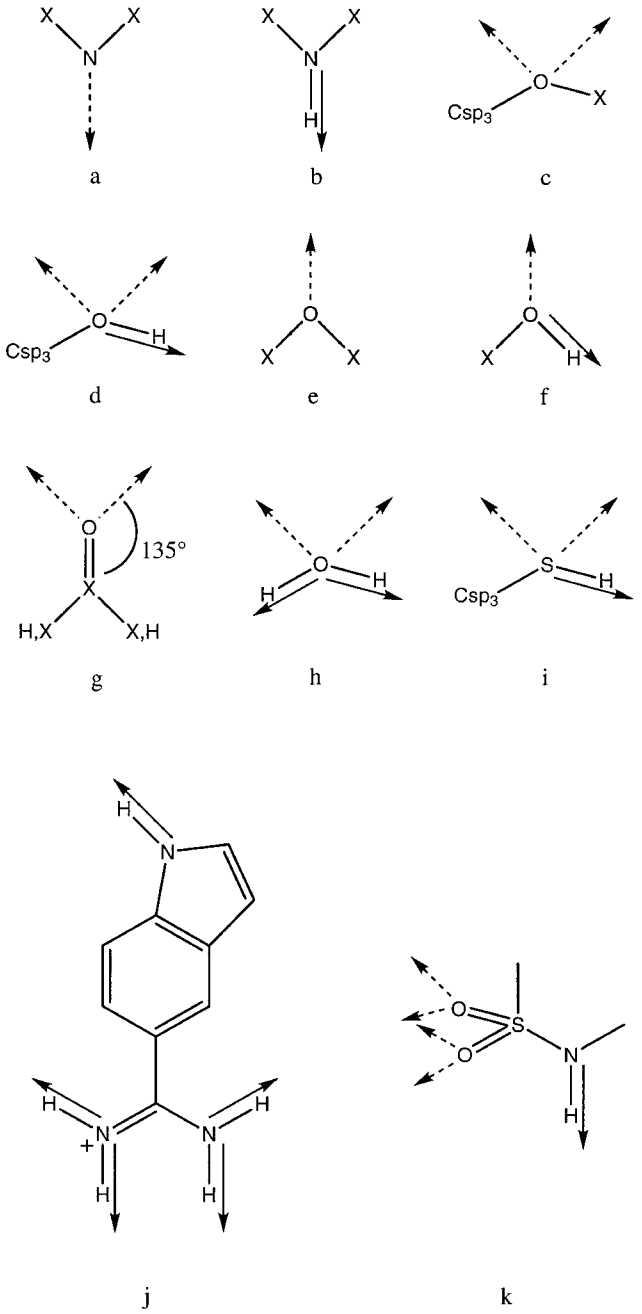

Fig. 1. Description of polar vectors for the fragment and for the receptor. X is a heavy atom. The broken arrow represents a vector of H-bond acceptor in the lone pair direction, and the full arrow represents a vector of H-bond donor. The geometry of c, d, h, and i is tetrahedral (angle of 109°). Examples: (**a**) imidazole, pyridine; (**b**) protein backbone, imidazole, indole; (**c**) ethers; (**d**) Ser and Thr side chains, sugars; (**e**) methoxybenzene; (**f**) Tyr side chain, phenol; (**g**) Asn, Gln, Asp, and Glu side chains, protein backbone, acetamide; (**h**) water; (**i**) Cys side chain; (**j**) 5-amidine indole; (**k**) *N*-methyl-methylsulfonamide.

Evaluation of Binding Energy

The procedure for detecting clashes between the fragment and the receptor is presented first. Then the methods used to evaluate the van der Waals interaction and the electrostatic energy (with and without solvation) are described. For the electrostatics the user can select either to neglect solvation (see "Electrostatic interaction without solvation" below) or to calculate it by a continuum approach (see "Continum electrostatic energy in solution" below) which is about six times slower. The total binding energy ($\Delta G_{binding}$) is then the sum of the intermolecular van der Waals and electrostatic energies which in the case of the continuum approach contains the desolvation of fragment and receptor. To increase efficiency apolar fragments are discarded without evaluation of the electrostatic contribution if the van der Waals interaction is less favorable than a threshold value.

In the current version of SEED both fragment and receptor are treated as rigid bodies. The nonelectrostatic contribution to solvation, if assumed to be proportional to the SAS, is negligible for small molecular systems[8] and for the docking of fragments having similar overall size. Changes in the intrasolute entropy of binding are also neglected. These are assumed to be sufficiently similar for the different binding modes.[9]

Bad contacts detection

A grid is defined over the whole receptor. This is followed by the generation of a list which contains for every cubic grid element the atoms which are enclosed by the surface of the cubic element. This list is calculated only once at the beginning of the program and is used to check bad contacts. A bad contact is defined as $r_{ij} < \alpha(R_i^{vdW} + R_j^{vdW})$ and a severe overlap is defined as $r_{ij} < \alpha\beta(R_i^{vdW} + R_j^{vdW})$, where r_{ij} is the distance between two atoms ($i \in$ fragment, $j \in$ receptor), and R_i^{vdW} and R_j^{vdW} are the corresponding van der Waals radii. The value of α is usually set to $2^{-1/6} \cong 0.89$, the distance at which repulsion balances attraction, and a β of 0.6 is employed normally. The distances between an atom of the fragment and the receptor atoms in the 27 grid cubes surrounding the fragment atom are calculated. This is repeated for every atom in the fragment. A grid spacing of twice the largest van der Waals radius is used to avoid missing clashes. For polar fragments the hydrogen atom involved in the H-bond is not checked for bad contacts. Fragments with one severe overlap or a number of bad contacts larger than an input parameter are discarded before evaluation of their energy.

Van der Waals interaction

A list of residue centroids is generated during the initial phase of the program and is used for an efficient estimation of van der Waals and screened electrostatic interactions. The atom closest to the geometrical center of the residue is selected as centroid for residues with zero formal charge, whereas the atom closest to the charge center is chosen for charged residues. The latter choice is more appropriate for the electrostatic interaction (see below). A three-dimensional (3D) grid is built over the receptor with a distance between neighbor grid points of 1 Å. Each centroid is assigned to the closest cubic element of the grid. Given a grid point *m*, all the grid points falling at a distance from *m* smaller than a given cutoff define a

pseudo-sphere associated to *m*. The neighbor list of a given fragment atom contains the atoms belonging to the receptor residues whose centroid is included in the pseudo-sphere centered on the grid point closest to the fragment atom. This increases the efficiency because it avoids calculating the distances between each fragment atom and all the receptor atoms when evaluating the interaction energy of a new fragment position. The van der Waals interaction energy is then computed between each atom of the fragment and the receptor atoms in the neighbor list according to

$$E_{ij}^{vdW} = \sqrt{\epsilon_i \epsilon_j}\left\{\left(\frac{R_i^{vdW} + R_j^{vdW}}{r_{ij}}\right)^{12} - 2\left(\frac{R_i^{vdW} + R_j^{vdW}}{r_{ij}}\right)^{6}\right\} \quad (1)$$

where ϵ_i is the minimum of the van der Waals potential between two atoms of type i at optimal distance of $2 \cdot R_i^{vdW}$.

Electrostatic interaction without solvation

In a homogeneous medium of dielectric constant ϵ the electrostatic interaction between a fragment and the receptor in kcal/mol is

$$E^{Coul} = \sum_{i \in fragment} q_i \Phi_i = \sum_{i \in fragment} q_i \left(332.0 \sum_{j \in receptor} \frac{q_j}{\epsilon r_{ij}^n}\right) \quad (2)$$

where q_i and q_j are the partial charges (in electronic unit charge), and r_{ij} is the distance in Å between the two atoms i and j. For n = 1 and n = 2, equation 2 corresponds to the Coulomb law and the distance-dependent dielectric model, respectively. This equation allows the factorization of the electrostatic potential of the receptor.[10–13] The values of Φ_i's are computed once over a grid containing the binding site plus a boundary and stored in a look-up table. The contribution of each fragment atom is computed by multiplying its partial charge with the potential of the receptor interpolated from the surrounding 8 points of the grid by the trilinear interpolation method.[14] Partial charges for the fragments were derived as described below.

Continuum electrostatic energy in solution

The main assumption underlying the evaluation of the electrostatic energy in solution of a fragment–receptor complex is the description of the solvent effects by continuum electrostatics.[5,6,15–23] The system is partitioned into solvent and solute regions, and two different dielectric constants are assigned to each region. In this approximation only the intra-solute electrostatic interactions need to be evaluated. This strongly reduces the number of interactions with respect to an explicit treatment of the solvent. Moreover, it makes feasible the inclusion of solvent effects in docking studies where the equilibration of explicit water molecules would be a major difficulty. The electrostatic effects of the solvent are relevant, and it has been shown that the continuum dielectric model provides an accurate description of molecules in solution.[6,8] The difference in electrostatic energy in solution upon binding of a fragment to a receptor can be calculated as the sum of the following three terms[7,16]:

- Desolvation of the receptor: electrostatic energy difference upon binding an uncharged fragment to a charged receptor in solution
- Screened fragment–receptor interaction: intermolecular electrostatic energy in solution
- Desolvation of the fragment: electrostatic energy difference upon binding a charged fragment to an uncharged receptor in solution

The definition of the solute volume, i.e., the low dielectric volume, is central in the evaluation of these energy terms with a continuum model. The solute–solvent dielectric boundary is described by the molecular surface (MS) of the solute.[24] A grid covering the receptor and the binding site is set up. In a first step the volume occupied by the isolated receptor is defined on the grid. Subsequently for every position of a docked fragment the volume enclosed by the MS of the fragment–receptor complex is identified. The volume occupied by a docked fragment is that part of the fragment–receptor complex that was not occupied by the isolated receptor. It consists of the actual volume of the fragment and the interstitial volume enclosed by the reentrant surface between fragment and receptor.

The screened fragment–receptor interaction and the fragment desolvation are evaluated according to the Generalized Born (GB) approximation.[5,6,19–23] The GB approach would be too time consuming for the evaluation of the desolvation of the receptor. Hence, an ad hoc procedure was developed for the receptor desolvation.

Desolvation of the receptor. According to the above assumptions the electrostatic energy in solution of the receptor can be expressed in terms of the electric displacement vector $\vec{D}(\vec{x})$ and of a location dependent dielectric constant $\epsilon(\vec{x})$ as an integral over three-dimensional space R^3 (refs. 6, 25):

$$E = \frac{1}{8\pi}\int_{R^3} \frac{\vec{D}^2(\vec{x})}{\epsilon(\vec{x})}\, d^3x. \quad (3)$$

Because $\vec{D}(\vec{x})$ is additive, for point charges it can be rewritten as a sum over all charges *i* of the receptor:

$$\vec{D}(\vec{x}) = \sum_i \vec{D}_i(\vec{x}). \quad (4)$$

Docking an uncharged molecular fragment in the receptor binding site only has the effect of modifying the dielectric properties of part of the binding site. Over the volume occupied by the fragment, the dielectric constant changes from the solvent value (ϵ_w) to the interior value (ϵ_p). In the limit in which the receptor electric displacement vector $\vec{D}$ does not change significantly upon fragment docking, the variation of the electrostatic energy of the receptor can be

written according to equation 3 as

$$\Delta E = \frac{\tau}{8\pi} \int_{V_{frag}} \vec{D}^2(\vec{x})\, d^3x \tag{5}$$

where $\tau = 1/\epsilon_p - 1/\epsilon_w$ and V_{frag} is the volume occupied by the fragment. On a 3D grid equation 5 becomes

$$\Delta E = \frac{\tau}{8\pi} \sum_{k \in V_{frag}} \vec{D}^2(\vec{x}_k) \Delta V_k. \tag{6}$$

Two approaches are possible to calculate the receptor electric displacement over a 3D grid. They both fulfill the condition of validity of equations 5 and 6 and have been implemented in SEED. In the first the electric displacement of every charge of the receptor can be represented by the Coulomb field[5,22]:

$$\vec{D}(\vec{x}) = \sum_i q_i \frac{(\vec{x} - \vec{x}_i)}{|\vec{x} - \vec{x}_i|^3}. \tag{7}$$

This is an analytical approximation of the total electric displacement. Alternatively, $\vec{D}$ can be calculated exactly for the isolated receptor by a finite difference solution of the Poisson equation and assumed not to change significantly upon fragment docking:

$$\vec{D}(\vec{x}) = -\epsilon(\vec{x}) \nabla \phi(\vec{x}) \tag{8}$$

where ϕ is the electrostatic potential solution of the Poisson equation. The second approach will be referred to in the following as the finite difference approximation for the evaluation of the receptor desolvation. The finite difference calculation of $\vec{D}$ for the isolated receptor (the program UHBD[26–28] is employed here) is performed only once at the beginning of the program run.

The receptor desolvation is computed from equation 6 together with equation 7 (Coulomb field approximation) or 8 (finite difference approximation).

Screened fragment–receptor interaction. The fragment–receptor interaction in solution is calculated via the GB approximation.[5,6,19–23] The interaction energy in solution between two charges embedded in a solute is

$$E_{ij}^{int} = \frac{q_i q_j}{\epsilon_p r_{ij}} - \frac{q_i q_j \tau}{R_{ij}^{GB}} \tag{9}$$

where

$$R_{ij}^{GB} = \sqrt{r_{ij}^2 + R_i^{eff} R_j^{eff} \exp\left(\frac{-r_{ij}^2}{4 R_i^{eff} R_j^{eff}}\right)} \tag{10}$$

and where q_i is the value of the partial charge i, and r_{ij} is the distance between charge i and j. R_i^{eff} is the effective radius of charge i, and it is evaluated numerically on a 3D grid covering the solute as described in ref. 5. It is a quantity depending only on the solute geometry and represents an estimate of the average distance of a charge from the solvent.

The intermolecular interaction energy is calculated as:

$$E^{int} = \sum_{\substack{i \in fragment \\ j \in list_i}} E_{ij}^{int} \tag{11}$$

where $list_i$ contains the receptor atoms belonging to the neighbor list of atom i. The electrostatic neighbor list includes all the receptor atoms of the van der Waals neighbor list (see above) and one atom for every charged residue whose centroid falls within a given cutoff (radius of the pseudo-sphere increased by 30%) of the binding site residues. The atom selected is the one closest to the center of charge. Supplementing the van der Waals neighbor list with a monopole approximation of distant charged residues dramatically reduces the error originating from the long range effects of electrostatics.

Desolvation of the fragment. The fragment intramolecular energy in solution is calculated with the GB formula as described in ref. 5:

$$E = \sum_{i \in fragment} E_i^{self} + \sum_{\substack{i>j \\ i,j \in fragment}} \left(\frac{q_i q_j}{\epsilon_p r_{ij}} - \frac{q_i q_j \tau}{R_{ij}^{GB}} \right) \tag{12}$$

where the two sums run over the partial charges of the fragment. Equation 12 differs from equation 11 owing to the presence of the *self-energy* term $\Sigma_i E_i^{self}$. This term is not zero only in the case of intramolecular energies. E_i^{self} is the *self-energy* of charge i and represents the interaction between the charge itself and the solvent. It is calculated as[5,22]

$$E_i^{self} = \frac{q_i^2}{2 R_i^{vdW} \epsilon_p} - \frac{q_i^2 \tau}{2 R_i^{eff}} \tag{13}$$

where R_i^{vdW} is the van der Waals radius of charge *i*.

The difference in the intramolecular fragment energy upon binding to an uncharged receptor in solution is

$$\Delta E = E^{docked} - E^{free} \tag{14}$$

where E^{docked} and E^{free} are the energies of the fragment bound and unbound to the receptor in solution, respectively. They are evaluated according to equation 12. For the unbound fragment (E^{free}) the effective radii are calculated considering as solute the volume enclosed by the molecular surface of the fragment. For the bound fragment (E^{docked}) the solute is the volume enclosed by the molecular surface of the receptor–fragment complex. E^{free} is evaluated only once per fragment type, whereas E^{docked} is recalculated for every fragment position in the binding site.

Validation. The approximations inherent to our continuum electrostatic approach were validated by comparison with finite difference solutions of the Poisson equation.

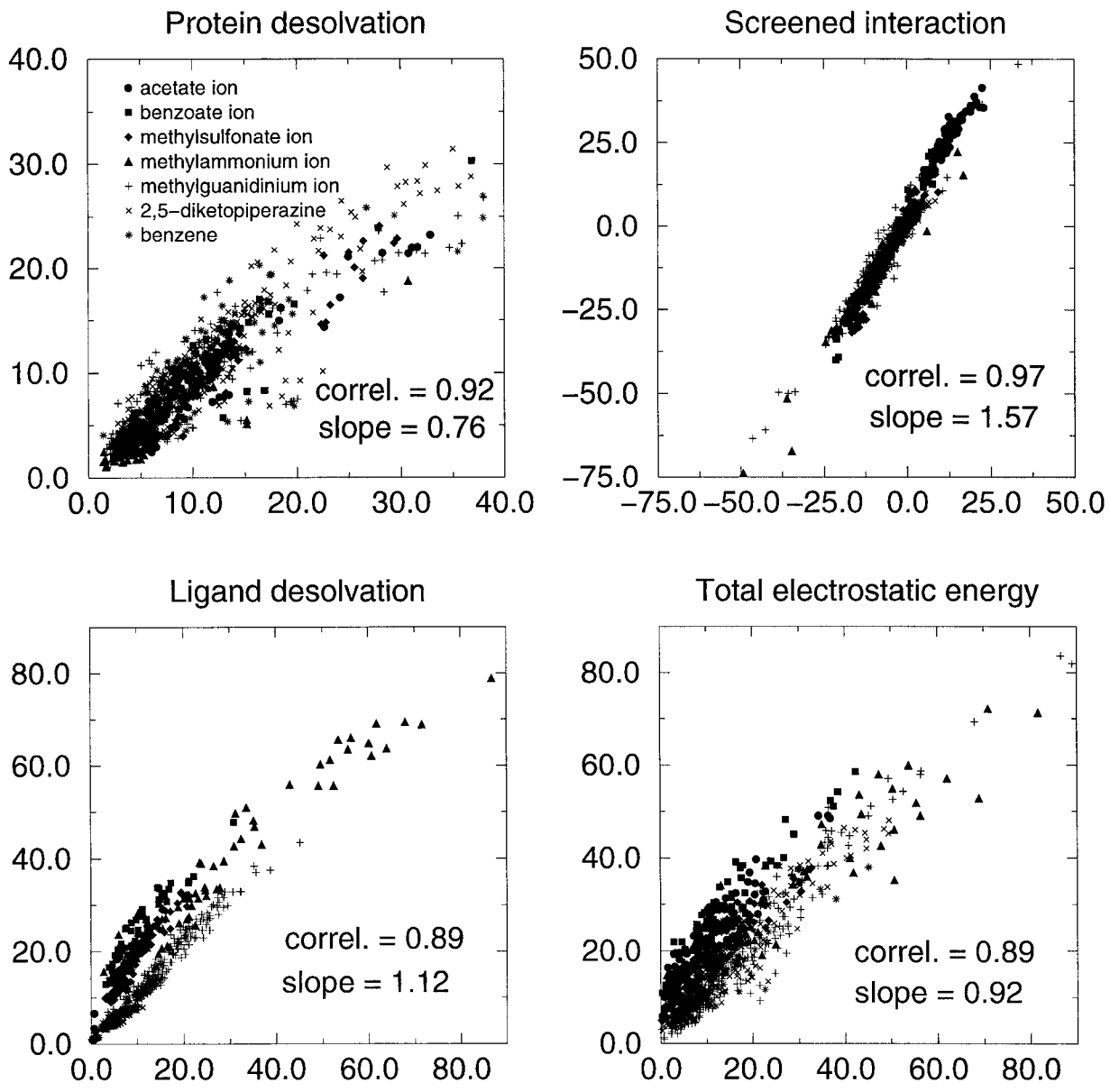

Fig. 2. Correlations in the electrostatic energies calculated by finite difference solution of the Poisson equation (x-axis) and SEED (y-axis). Values are plotted for 1,025 complexes of thrombin with small molecules. The total electrostatic energy is the sum of the protein desolvation, screened interaction, and ligand desolvation scaled by 0.76, 1.57, and 1.12, respectively. The finite difference calculations were performed with the program UHBD.[18,26,28] An interior dielectric of 1, solvent dielectric of 78.5, and grid size of 0.5 Å were used for both SEED and UHBD.

For this purpose, the three electrostatic energy terms were calculated with SEED and UHBD for a set of small molecules and ions distributed over the binding site of thrombin and at the dimerization interface of the HIV-1 aspartic proteinase (HIV-1 PR) monomer. The molecule set included acetate ion, benzoate ion, methylsulfonate ion, methylammonium ion, methylguanidinium ion, 2,5-diketopiperazine, and benzene. The total number of fragment–receptor complexes analyzed were 1,025 for thrombin (Fig. 2) and 1490 for the HIV-1 PR monomer. The results are shown in Table II for two different values of the solute dielectric constant (1 and 4). The agreement between the two methods is very good. Systematic errors (slope $\neq 1$) are independent of the receptor and the solute dielectric constant and consequently can be corrected by the use of appropriate scaling factors for the different energy terms (Fig. 2). The agreement between the electrostatic energies calculated by SEED and UHBD is worse when using an interior dielectric of 1 instead of 4. This is due to the lower correlation between desolvation energies of charged fragments (see Table II). The higher the ratio between the solvent and solute dielectric constants the more the polarization charge distribution induced at the dielectric discontinuity surface deviates from the values obtained with the

TABLE II. Comparison of Electrostatic Energies Calculated by SEED and UHBD†

	Thrombin		HIV-1 protease monomer	
	R[a]	α[b]	R[a]	α[b]
Solute dielectric constant = 1				
Receptor desolvation (Coulomb approx.)	0.92	0.76	0.95	0.80
Receptor desolvation (finite difference)	0.92	2.91	0.97	2.48
Screened interaction	0.97	1.57	0.97	1.39
Ligand desolvation	0.89	1.12	0.80	1.16
Total elec. energy (Coulomb approx. for receptor desolvation)	0.86 (0.89)	1.00 (0.92)	0.87 (0.90)	1.12 (0.95)
Total elec. energy (finite difference for receptor desolvation)	0.81 (0.88)	1.77 (0.96)	0.88 (0.90)	1.67 (1.06)
Solute dielectric constant = 4				
Receptor desolvation (Coulomb approx.)	0.93	0.78	0.96	0.83
Receptor desolvation (finite difference)	0.96	2.59	0.98	2.28
Screened interaction	0.95	1.28	0.98	1.31
Ligand desolvation	0.94	1.14	0.95	1.24
Total elec. energy (Coulomb approx. for receptor desolvation)	0.95 (0.94)	1.03 (0.93)	0.93 (0.94)	1.12 (0.94)
Total elec. energy (finite difference for receptor desolvation)	0.88 (0.94)	1.53 (0.96)	0.92 (0.95)	1.46 (0.99)

†Values in parentheses concern the sum of the three energy components, each scaled by the slope of the fitting line of the SEED to the UHBD energies.
[a]R: correlation coefficient for the fit of the SEED to the UHBD energies.
[b]α: slope of the fitting line of the SEED to the UHBD energies.

GB approximation, whereas in the case of $\epsilon_w = \epsilon_p$ the GB equation is exact. This effect is particularly evident for the desolvation of small charged ligands, where relatively high partial charges are directly exposed to the solvent, inducing a high polarization charge at the dielectric interface.

In the application to thrombin discussed below the Coulomb field approximation and a solute dielectric constant of 1 were used with scaling factors of 0.76 for the receptor desolvation, 1.57 for the screened interaction and 1.12 for the ligand desolvation (values of α from Table II).

Sorting and Clustering Procedures

The docking of a given fragment (with energy evaluation as described above) is followed by sorting and clustering. Within each fragment type, positions are first sorted according to binding energy. Positions whose binding energy is lower than a user-specified threshold value are then clustered using as similarity criterion between two fragment positions *A* and *B*[29,30]:

$$S(A, B) = \frac{S_{AB}}{\max(S_{AA}, S_{BB})}$$

$$\text{where } S_{XY} = \sum_{i \in X} \sum_{j \in Y} w_{t_i t_j} \exp(-\gamma r_{ij}^2) \quad (15)$$

where r_{ij} is the distance between two atoms ($i \in$ fragment position X, $j \in Y$), $w_{t_i t_j}$ is a user-controlled matrix whose coefficients reflect the similarity between element types (in most cases a unit matrix is used), and γ is a coefficient which acts on the broadness of the distribution of the positions. B is assigned to the cluster of A if S(A, B) is larger than a cutoff value δ with $0 \leq \delta \leq 1$. The clustering proceeds in two steps: (1) a first clustering with $\gamma = 0.9$ and $\delta = 0.4$ yields large clusters which contain almost overlapping as well as more distant fragments; (2) a second clustering with $\gamma = 0.9$ and $\delta = 0.9$ is done on each cluster found in (1) to eliminate fragments which are very close in space. A single clustering step with $\gamma = 0.9$ and $\delta = 0.9$ would generate too many small clusters. Hence, the first step is a real clustering, whereas the second step is done only to discard redundant positions.

System Set-up

The structure of thrombin was taken from the PDB database (code 1HGT).[31] The water molecules, the hirugen inhibitor, and the light chain of thrombin were removed. Hydrogen atoms were added with the molecular modeling program WITNOTP (A. Widmer, unpublished). Partial charges were assigned to thrombin and to the fragments with the MPEOE method[32–34] implemented in WITNOTP which reproduces the all-hydrogen MSI CHARMm22 parameter set.[35] Then the thrombin structure complexed with Nα-((2-naphthylsulfonyl)glycyl)-DL-*p*-amidinophenylalanylpiperidine (NAPAP, Fig. 3a) underwent a restrained CHARMM[36] conjugate gradient minimization. Atoms more distant than 9 Å from any NAPAP atom were kept rigid, and a 4.0 × r-dielectric model was used. The following residues were chosen to define the binding site: Leu41, Cys42, His57, Cys58, Tyr60A, Asn98, Leu99, Ile174, Asp189, Ala190, Cys191, Ser195, Val213, Ser214, Trp215, Gly216, Glu217, Gly219, Cys220, and Gly226. The van der Waals energies and radii were taken from the MSI CHARMm22 parameter set.[35]

The SEED default lengths of the hydrogen bonds for N-N, N-O, N-S, O-O, O-S, and S-S pairs of heavy elements were 3.1, 2.9, 2.9, 2.7, 2.9, and 2.9 Å, respectively. A H-bond distance of 2.6 Å was considered between the CHARMm22 atom types NC and OC, a length of 2.75 Å between NC or OC and all the other nitrogen and oxygen atom types. The van der Waals parameters of NP (peptide nitrogen) and CT (aliphatic carbon) were used in the evaluation of the van

der Waals energy for the selection of the polar and apolar vectors, respectively. A preselection of the positions of the fragments was made by discarding the fragments whose geometric center was outside a sphere of 10 Å centered on the oxygen of the primary amide of NAPAP.

RESULTS AND DISCUSSION

Thrombin Functionality Maps

In presenting the SEED results, both structural and energetic properties of the positions are analyzed. Moreover, a comparison of the functional group sites with the interaction patterns of the thrombin inhibitors PPACK[37,38] and NAPAP (racemic mixture K_i of 6.6 nM for human thrombin)[39,40] is given for nonpolar, polar, and charged fragments. Hereafter the cluster representatives obtained in the first clustering step will be referred to as "representatives" or "positions." Table III lists the functional groups docked by SEED and their electrostatic free energy of hydration calculated with equation 12.

Nonpolar groups

The cluster representatives of benzene, cyclopentane, cyclohexane, and naphthalene are distributed over the hydrophobic regions of the thrombin active site. Because their functionality maps and energy values are similar, only the positions of benzene and cyclopentane are analyzed in detail.

Benzene. The 15 cluster representatives with the best binding energy are listed in Table IV, and positions 1, 3, 6, and 8 are shown in Figure 3b. The two representatives with the best energy are in the S3 pocket and have strong van der Waals interactions (about −8 kcal/mol) and low electrostatic desolvation of the protein (about 4 kcal/mol). Benzene 1 is between the two rings of the naphthalene group of NAPAP, and benzene 2 is close to the first ring of naphthalene. Benzene 3 overlaps the proline ring of PPACK. The energy contributions of position 3 are comparable with those of the first two positions. The benzene representatives 4, 5, 7, 9, and 10 are in the S3 pocket; 5 overlaps the NAPAP naphthalene and 4, 7, 9, and 10 are slightly shifted. The most favorable van der Waals interaction is found for positions 6 (−13.5 kcal/mol) and 11 (−11.2 kcal/mol) where the benzene ring is sandwiched between the amide groups of residues 215–216 and 191–192 in the S1 pocket and occupies the same site as the aromatic ring of benzamidine in the NAPAP–thrombin complex. It partially desolvates the Asp189 side chain with a large electrostatic desolvation of thrombin (10.3 and 8.0 kcal/mol for benzenes 6 and 11, respectively). This is not compensated by favorable electrostatic interactions between the aromatic ring and the amide planes, as the former bears small partial charges of −0.13 (C) and 0.13 (H). It is worth noting that these two positions would be more favorable than the benzene representatives 1 to 5 in S3 and S2 if the electrostatic desolvation of the protein would be neglected. This is in agreement with previous calculations[7] and shows the importance of the electrostatic desolvation for a correct ranking. Benzenes 8 and 13 are located in the intermediate region between the S2 and S3 pockets.

Cyclopentane. The 12 representatives with the best energy are listed in Table IV, and positions 1, 2, 3, and 12 are shown in Figure 3c. Cyclopentanes 1, 4, and 5 are located in the S3 pocket. Cyclopentane 1 overlaps the PPACK benzene and the most distal part of the NAPAP naphthalene. Among these three representatives, cyclopentane 1 has the deepest position in S3 and shows the best van der Waals interaction and total binding energy. On the other hand, it pays a penalty of 3.6 kcal/mol for the electrostatic desolvation of thrombin due to the partial desolvation of the carbonyl group of residue Glu97A. Position 2 occupies the position of the PPACK proline. Position 3 lies in the middle between the S2 and S3 pockets. Position 12 is located in the S1 pocket close to the benzamidine ring of NAPAP. As for the corresponding representative of benzene it has the lowest van der Waals interaction of all cyclopentane positions (−14.0 kcal/mol). The penalty for the electrostatic desolvation of the protein is 10.0 kcal/mol, because cyclopentane 12 partially desolvates the peptide groups of residues 191–192 and 215–216 and the Asp189 carboxyl group located at the bottom of the S1 pocket. As in the case of benzene, the neglect of the electrostatic desolvation of the protein would result in a better ranking of cyclopentane 12 than those in the S3 and S2 pockets.

Cyclohexane. Representatives 1, 3, and 5 occupy the S3 pocket of thrombin, whereas cyclohexane 2 is in the intermediate region between S3 and S2. Position 4 is in the S2 pocket and overlaps the proline ring of PPACK. Cyclohexane 12 is located in S1 as the cyclopentane 12 and the benzene 6. Their individual energy contributions are analogous to those of the corresponding cyclopentane positions. The desolvation of the fragment is higher in the case of benzene (about 1.1 kcal/mol) than for cyclopentane or cyclohexane (about 0.1 kcal/mol), because the partial charges of aromatic groups are slightly larger ($q_C = -0.13$, $q_H = 0.13$) than those of aliphatic fragments ($q_C = -0.10$, $q_H = 0.05$).

Naphthalene. Representatives 1, 2, 4, and 5 are located in the S3 pocket, whereas position 3 occupies the intermediate region between S3 and S2. Position 5 shows the best fit with the naphthalene part of NAPAP with a RMSD of 1.68 Å. Naphthalene 21 is the best representative in S1. It has a binding energy of −1.5 kcal/mol with a protein desolvation energy of 14.3 kcal/mol and a van der Waals interaction of −14.9 kcal/mol.

From the analysis of the thrombin functionality maps of the nonpolar groups it is evident that hydrophobic fragments prefer to bind to the S3 and S2 pockets in agreement with the large amount of structural data on thrombin–inhibitor complexes.[40–43] The van der Waals interactions with the hydrophobic side chains of Leu99, Ile174, and Trp215 in the S3 pocket are slightly more favorable than the van der Waals interactions with Leu99, Tyr60A, and Trp60D in the S2 pocket.

Polar groups

Polar neutral groups are distributed over most of the hydrophilic regions of the active site. The positions of *N*-methyl-methylsulfonamide (NMMS, Fig. 1k) and phe-

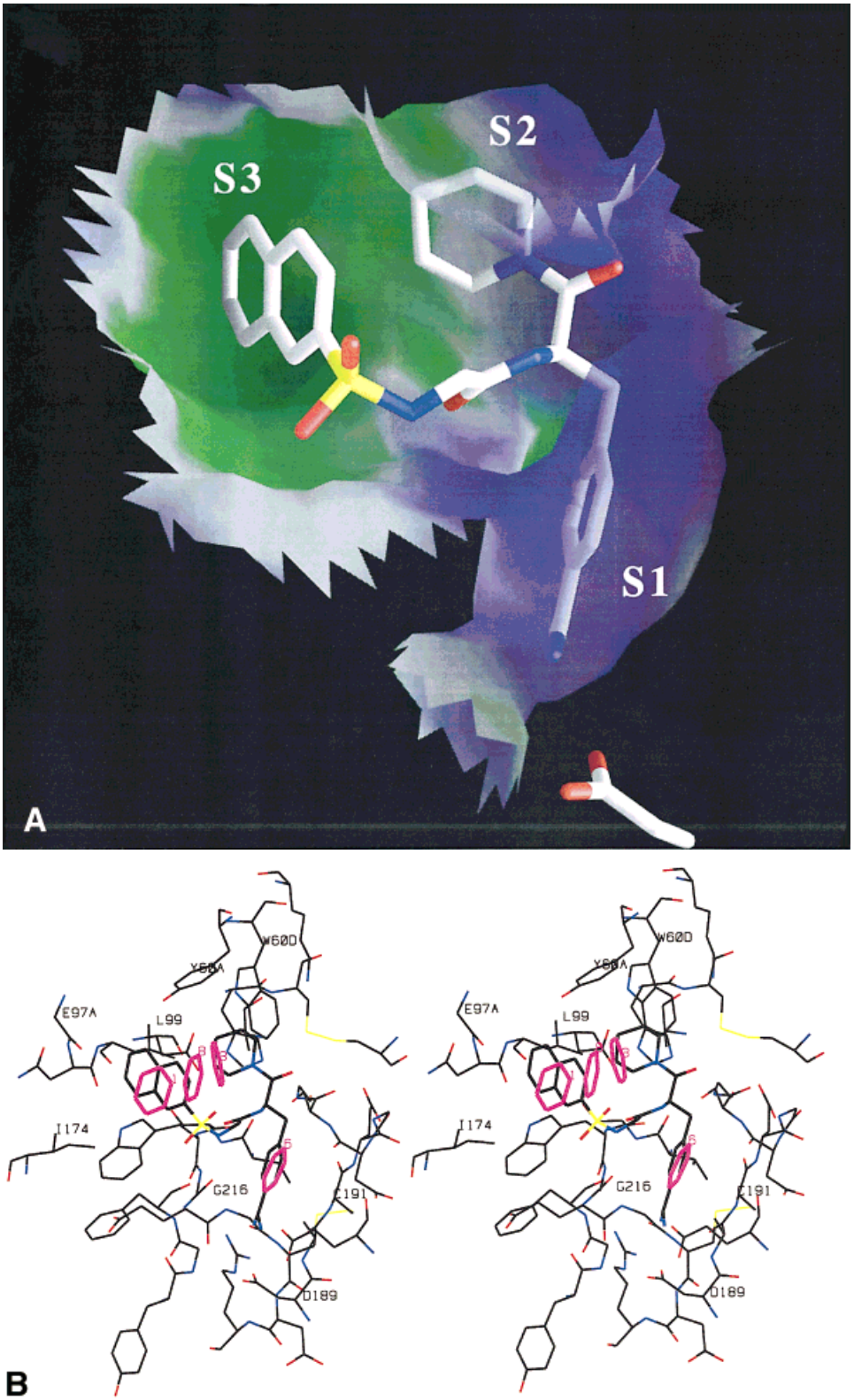

Fig. 3. **(a)** The molecular surface of the nonprime pockets of the thrombin active site is displayed together with NAPAP and the side chain of Asp189. Hydrophobic regions are green, and hydrophilic regions are blue. Figure made with GRASP.[66] **(b)** Stereoviews (relaxed-eyed) of the SEED cluster representatives of benzene (magenta) in the thrombin active site (thin lines colored by atom type, carbon black, nitrogen blue, oxygen red, sulfur yellow). The NAPAP inhibitor is also shown (medium lines colored by atom type), although it was removed during the SEED procedure. The SEED cluster representatives are labeled according to their binding energy rank within representatives of the same type. **(c)** Same as in b for cyclopentane. **(d)** Same as in b for phenol.

nol will be discussed in detail, whereas those of *N*-methylacetamide (NMA) and imidazole will be analyzed only briefly. Following criteria are used for the analysis of hydrogen bonds: a distance between donor and acceptor shorter than 3.6 Å and a donor-H · · · acceptor angle larger than 110°.

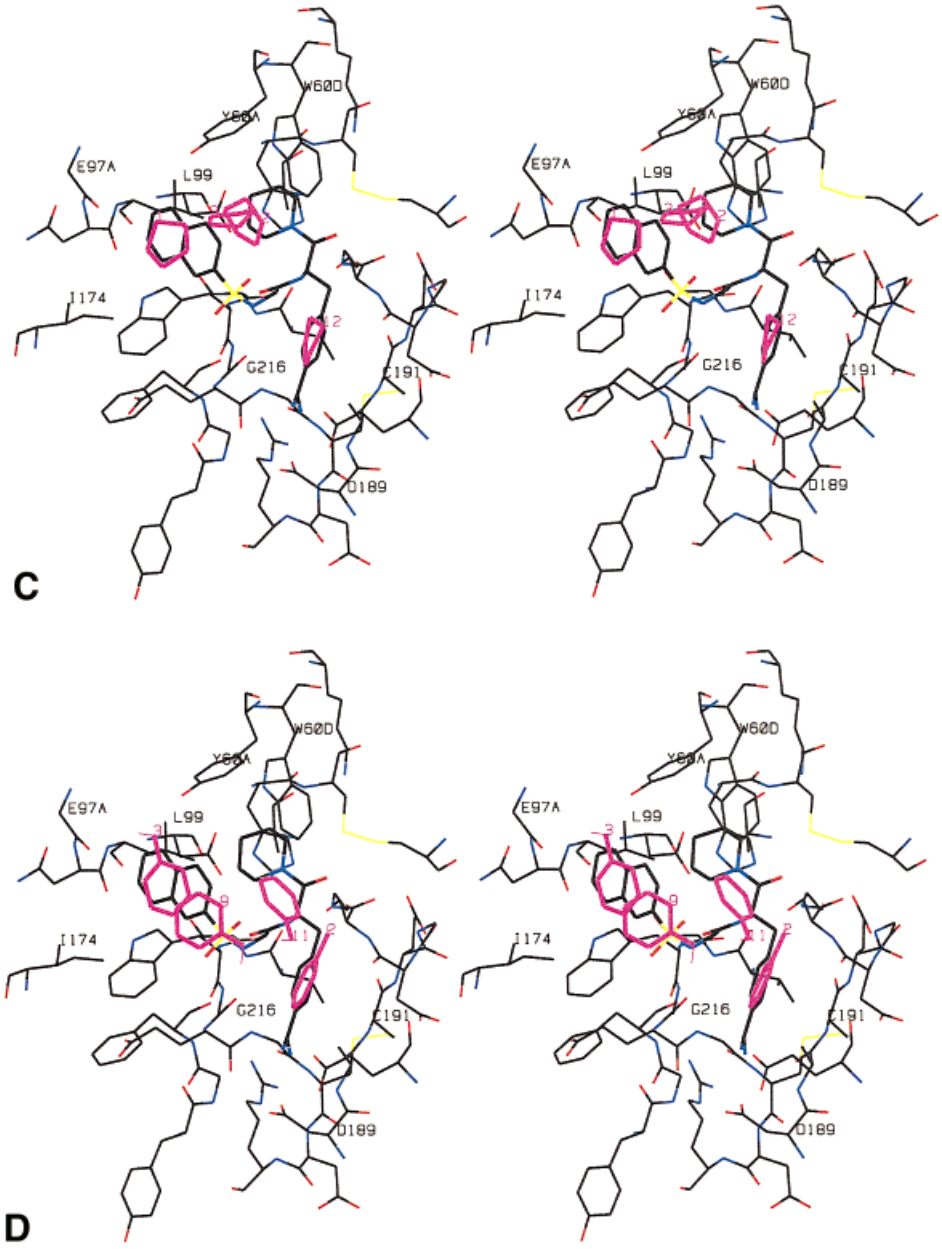

Figure 3. (Continued.)

N-Methyl-methylsulfonamide. Representatives 1 and 3 are located on the top of the S1 pocket. They donate a hydrogen bond to the CO of Gly216 and accept from 216NH. The latter corresponds to the hydrogen bond between the carbonyl group of NAPAP and 216NH. The oxygen atom of NMMS 1 and 3 not involved in a hydrogen bond is partially exposed to the solvent. Position 1 is more buried in S1 than position 3, which is reflected in a more favorable van der Waals interaction (−9.3 and −7.2 kcal/mol, respectively; Table V) and a larger desolvation of the ligand (9.6 and 8.8 kcal/mol). The difference of about 5 kcal/mol between the respective electrostatic interaction energies originates from the longer hydrogen bond with 216NH for position 3. Moreover, in NMMS 3 both hydrogen bonds are more exposed to the screening effect of the solvent. NMMS 2 is buried in S1 and has a large desolvation of the protein (10.7 kcal/mol) because its methyl groups point toward the bottom of the S1 pocket. It donates a hydrogen bond to 191CO and accepts from 216NH. NMMS 18 almost overlaps the corresponding group of NAPAP. It donates a hydrogen bond to 216CO, whereas its oxygen atoms are exposed to solvent. The poor ranking and slightly unfavorable binding energy (0.3 kcal/mol) suggest that the NAPAP sulfonamide may not be involved in optimal interactions with thrombin. In this context it is important to note that low nanomolar inhibitors of throm-

TABLE III. Functional Groups Used in the Thrombin Application†

	Electrostatic solvation free energy[a]	$\Delta G_{binding}$		
		Lowest	2nd	3rd
Nonpolar groups				
Benzene	−1.4	−4.2	−3.1	−2.9
Cyclopentane	−0.0	−7.1	−5.5	−5.3
Cyclohexane	−0.0	−6.3	−4.7	−4.6
Naphthalene	−2.0	−6.1	−5.1	−4.6
Polar groups				
N-Methyl-methylsulfonamide	−13.8	−7.4	−4.7	−1.7
N-Methylacetamide	−6.3	−8.8	−7.5	−4.9
Phenol	−4.5	−8.2	−5.2	−4.0
Imidazole	−3.2	−3.8	−3.1	−2.6
Charged groups				
Benzamidine	−65.4	−29.0	−23.9	−10.7
5-Amidine indole	−67.2	−23.7	−21.2	−2.9
Methylguanidinium	−61.3	−27.7	−17.8	−15.0
Acetate ion	−70.1	9.5[b]	10.3	11.3
Piperazine dication	−243.2	11.2[b]	12.8	13.5

†All energy values are in kcal/mol.
[a]Calculated with equation 12.
[b]For the acetate ion and the piperazine dication there are no favorable positions in the nonprime region of the thrombin active site.

TABLE IV. Cluster Representatives of Apolar Fragments†

	Intermolecular		Electrostatic desolvation			
Rank[a]	vdWaals	Elect	Receptor	Fragment	$\Delta G^{b}_{binding}$	Site
Benzene						
1	−8.7	−0.5	3.9	1.1	−4.2	S3
2	−7.3	−0.7	3.8	1.1	−3.1	S3
3	−7.7	−0.0	3.6	1.2	−2.9	S2[c]
4	−7.2	−0.2	3.5	1.1	−2.9	S3
5	−7.1	0.0	3.3	1.1	−2.7	S3
6	−13.5	−0.6	10.3	1.2	−2.6	S1
7	−7.8	0.2	3.8	1.2	−2.6	S3
8	−6.2	0.2	2.6	1.0	−2.5	S3–S2
9	−7.9	0.7	3.7	1.2	−2.3	S3
10	−7.8	0.4	4.1	1.0	−2.3	S3
11	−11.2	−0.1	8.0	1.3	−2.0	S1
12	−6.7	0.1	3.5	1.2	−1.9	S3
13	−6.6	0.4	3.2	1.1	−1.9	S3–S2
14	−6.8	−0.8	4.6	1.1	−1.8	S2
15	−5.7	−1.0	3.9	1.0	−1.8	S3
Cyclopentane						
1	−11.2	0.4	3.6	0.1	−7.1	S3
2	−9.3	0.6	3.1	0.1	−5.5	S2[c]
3	−8.2	0.6	2.2	0.1	−5.3	S3–S2
4	−8.1	0.5	2.4	0.1	−5.0	S3
5	−8.6	0.5	3.3	0.1	−4.8	S3
6	−9.0	0.5	3.9	0.1	−4.4	S3
7	−7.3	0.8	2.6	0.1	−3.8	S3–S2
8	−8.9	0.8	4.6	0.1	−3.4	S2
9	−7.4	0.5	3.4	0.2	−3.4	S3
10	−7.7	0.4	3.9	0.1	−3.3	S3
11	−7.2	0.2	3.8	0.1	−3.1	S3
12	−14.0	0.7	10.0	0.1	−3.1	S1

†All energy values are in kcal/mol.
[a]Ranked among the clusters of the same functional group type according to total binding energy. Representatives with rank in bold are shown in Figure 3b,c.
[b]Sum of the values in the four preceding columns, i.e., intermolecular and electrostatic desolvation energies.
[c]Position of the PPACK proline ring.

TABLE V. Cluster Representatives of Polar Fragments[†]

Rank[a]	Intermolecular vdWaals	Intermolecular Elect	Electrostatic desolvation Receptor	Electrostatic desolvation Fragment	$\Delta G^{b}_{binding}$	Site and H-bond partners[c]
***N*-methyl-methylsulfonamide**						
1	−9.3	−16.0	8.2	9.6	−7.4	S1; 216CO,216NH
2	−10.6	−15.1	10.7	10.2	−4.7	S1; 216NH,191CO
3	−7.2	−11.5	8.2	8.8	−1.7	S1; 216CO,216NH
4	−3.6	−5.5	2.2	5.4	−1.6	S3; $Tyr^{60A}O^{\eta}$
5	−4.8	−7.0	3.7	6.9	−1.2	S3; 216CO,219NH
6	−6.6	−9.9	7.5	7.9	−1.0	S1; 216CO,219NH
7	−7.1	−11.7	7.7	10.1	−1.0	S1; 214CO,216NH
8	−3.8	−5.0	2.9	5.2	−0.8	S3; $Tyr^{60A}O^{\eta}$
9	−7.3	−12.0	9.7	8.8	−0.8	S1; 216CO
10	−3.4	−5.0	2.7	5.1	−0.7	S3; $Tyr^{60A}O^{\eta}$
18	−5.7	−5.3	3.9	7.3	0.3	S3[d]; 216CO
Phenol						
1	−11.0	−10.6	9.7	3.7	−8.2	S1; 214CO
2	−12.4	−7.8	11.3	3.8	−5.2	S1; $Ser^{195}O^{\gamma}$
3	−9.6	−2.7	4.6	3.8	−4.0	S3; $Tyr^{60A}O^{\eta}$,97A CO
4	−9.7	−7.2	9.3	3.7	−3.9	S1; $Ser^{195}O^{\gamma}$
5	−4.3	−5.3	3.0	2.9	−3.7	S3; 216CO,219NH
6	−9.0	−7.4	9.4	3.8	−3.2	S1; 214CO
7	−3.7	−5.2	3.3	2.8	−2.8	S3; $Tyr^{60A}O^{\eta}$
8	−4.8	−4.8	3.9	2.9	−2.7	S3; 216CO
9	−5.2	−4.4	3.8	3.2	−2.6	S3; 216CO,216NH
10	−6.0	−4.0	4.5	3.0	−2.4	S3[e]; $Tyr^{60A}O^{\eta}$
11	−7.1	−6.1	7.2	3.7	−2.3	S2; 214CO

[†]All energy values are in kcal/mol.
[a]Ranked among the clusters of the same functional group type according to total binding energy. Representatives with rank in bold are shown in Figure 3d.
[b]Sum of the values in the four preceding columns, i.e., intermolecular and electrostatic desolvation energies.
[c]For positions with more than one H-bond, the first H-bond partner in the list is the one which was used by SEED for docking.
[d]Almost superimposed to the corresponding group of NAPAP.
[e]Aromatic ring close to the PPACK benzene.

bin exist that do not contact Gly216CO, e.g., a series of derivatives of 5-amidine indole.[44]

Phenol. In the positions 1, 6, and 11 phenol donates a hydrogen bond to the main chain CO of residue 214 (Fig. 3d and Table V). The position of the aromatic ring is deep in S1, at the top of S1 and in S2, respectively. Phenols 2 and 4 are located in the S1 pocket and donate to the hydroxyl oxygen of Ser195, whereas phenol 3 is in the S3 pocket, donates to the main chain CO of Glu97A, accepts from the OH of Tyr60A. The aromatic ring of phenol 2 overlaps the benzamidine ring of NAPAP. Phenol 1 has a very favorable electrostatic interaction energy (−10.6 kcal/mol) and a large desolvation of the protein (9.7 kcal/mol). On the contrary phenol 2 has a less favorable electrostatic interaction energy (−7.8 kcal/mol) owing to the screening effect of the solvent and has a larger desolvation of the protein (11.3 kcal/mol). The benzene ring of phenol 4 is shifted toward the solvent with respect to phenol 2. Representatives 5, 8, and 9 donate to the CO group of Gly216. Phenol 5 makes an additional hydrogen bond with NH of Gly219. Phenol 7 and 10 are in S3 and donate to the hydroxyl oxygen of Tyr60A.

N-Methylacetamide. Representatives 1 and 2 donate to the CO of Ser214 and have the best electrostatic interaction energies (−10.0 and −8.5 kcal/mol, respectively). The oxygen is not engaged in hydrogen bonds with thrombin. The NH of position 2 is close to the backbone nitrogen of PPACK arginine (distance of 0.37 Å). In representatives 3, 5, 6, 7, and 10, NMA donates to the hydroxyl group of Tyr60A. Positions 8, 9, and 13 donate to the main-chain carbonyl of residue Gly216. Nevertheless, no position was found in the same orientation as the NAPAP amide. The SEED energy evaluation of NAPAP NMA yields an unfavorable binding energy of 4.5 kcal/mol. The oxygen in NAPAP acts as an acceptor for Gly216NH (the distance N-O is equal to 3.2 Å), and the NH group is not involved in any hydrogen bond with the protein. Since the primary amide group of NAPAP is partially exposed to the solvent the electrostatic interaction energy is very weak (−1.0 kcal/mol). This may indicate that the amide group of NAPAP has mainly the function of a linker for the optimal placement of two apolar moieties in S3 and S2.

Imidazole. Imidazole has a map similar to that of phenol. There are positions of imidazole in the S3 and S2 pockets and also in S1. They are involved in hydrogen bonds with the Ser195 hydroxyl oxygen, the main chain CO group of residues 214, 216, and 219 in S1.

As already mentioned in previous work,[7] general conclusions about preferential thrombin sites for polar groups cannot be easily drawn. They depend on their size, the

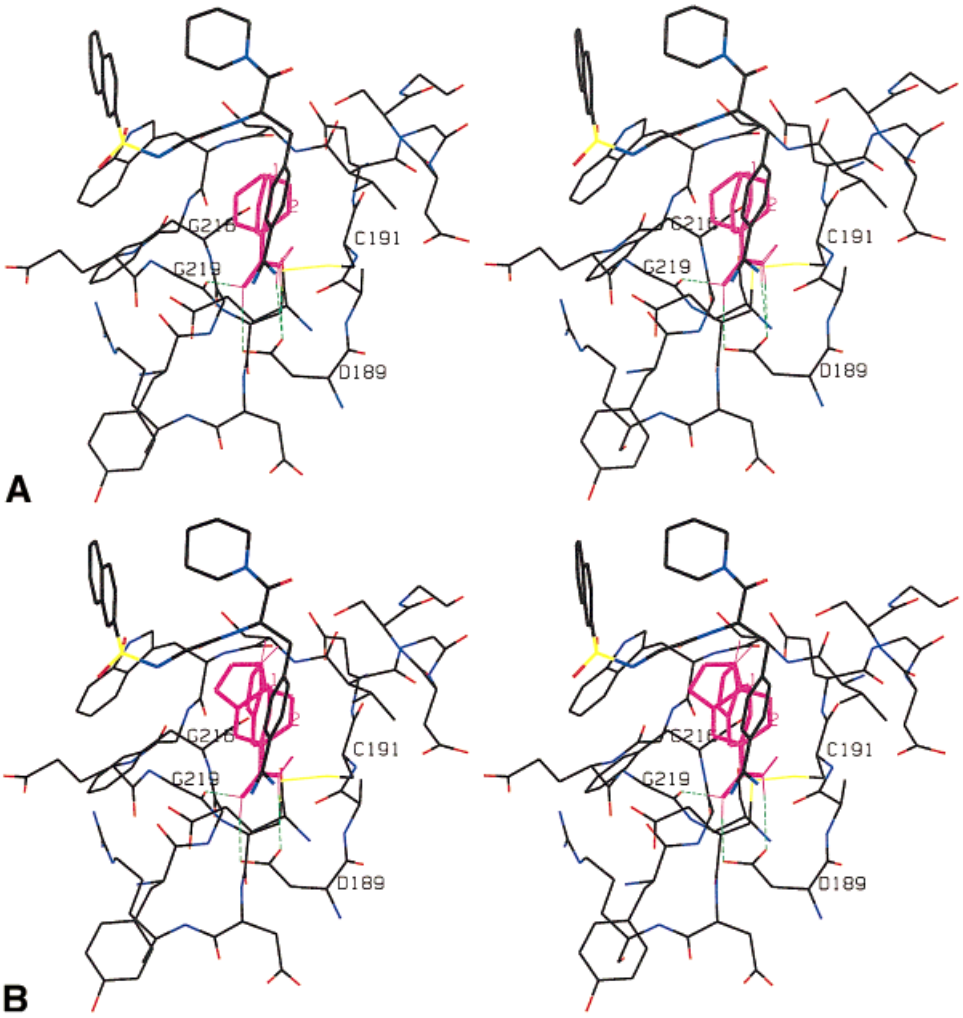

Fig. 4. Same as Figure 3b. Hydrogen bonds between protein and SEED cluster representatives are shown as green dashed lines. **(a)** Benzamidine. **(b)** 5-Amidine indole.

particular arrangement of charges, and also on a compromise between strong hydrogen bonds and good van der Waals interactions (e.g., phenol representative 3). Polar groups with favorable binding energy are often involved in hydrogen bonds with the main chain groups 214CO, 216NH, and 216CO, and with the hydroxyl groups of Tyr60A and Ser195.

Charged groups

Benzamidine and 5-amidine indole (Fig. 1j) are discussed in detail. For both fragments two conformations were used with the plane containing the amidine moiety tilted by +25° and −25° with respect to the aromatic ring plane. Methylguanidinium ion, acetate ion, and piperazine dication are analyzed briefly.

Benzamidine. The amidine moieties of the first two representatives overlap the NAPAP amidine in the S1 pocket (Fig. 4a and Table VI). The RMS deviation of all heavy atoms with respect to the benzamidine group of NAPAP is 0.85 Å and 1.07 Å, respectively. In both cases one nitrogen atom donates to the CO of residue Gly219 and to one carboxyl oxygen of the side chain of Asp189. The other benzamidine nitrogen donates to the second carboxyl oxygen of Asp189. These three hydrogen bonds which occur in the buried region at the bottom of S1 lead to a very strong electrostatic interaction energy (about −70 kcal/mol). Benzamidine binds to human thrombin with a K_i of 300 μM.[41] At room temperature this corresponds to a binding free energy of about −5 kcal/mol which is much smaller in magnitude than the values obtained by SEED. Two possible reasons for this discrepancy are the neglect of the solute entropy and the use of a solute dielectric of 1, which yields excessively strong intermolecular energy.

Positions 3 and 4 are less buried than 1 and 2, and make only one hydrogen bond in the S1 pocket with the carbonyl oxygen of Gly219. The amidine part of benzamidine is located relatively close to the carboxylic part of Asp189, but the separation is such that the hydrogen bonds are broken. It is worth noting the large gap in binding energy between representatives 1, 2 and 3, 4 (Table VI), which originates mainly from the difference in the electrostatic interaction. Position 5 has its benzene part pointing toward the bottom of the S1 pocket. It acts as a donor in a hydrogen bond with the main chain oxygen of Ser214. It is

TABLE VI. Cluster Representatives of Charged Fragments[†]

Rank[a]	Intermolecular vdWaals	Intermolecular Elect	Electrostatic desolvation Receptor	Electrostatic desolvation Fragment	$\Delta G^{b}_{binding}$	Site and H-bond partners[c]
Benzamidine						
1[d]	−14.0	−70.0	11.3	43.7	−29.0	S1; 219CO,Asp$^{189}O^{\delta1}$,$O^{\delta2}$
2[e]	−9.2	−69.9	11.6	43.5	−23.9	S1; 219CO,Asp$^{189}O^{\delta1}$,$O^{\delta2}$
3[d]	−9.5	−58.0	12.7	44.1	−10.7	S1; 219CO
4[e]	−3.8	−58.6	12.0	44.8	−5.5	S1; 219CO
5[d]	−11.8	−25.1	12.8	19.6	−4.6	S1; 214CO
5-Amidine indole						
1[d]	−10.2	−72.1	13.4	45.2	−23.7	S1; 219CO,Asp$^{189}O^{\delta1}$,$O^{\delta2}$
2[e]	−9.1	−71.9	13.7	46.1	−21.2	S1; 219CO,Asp$^{189}O^{\delta1}$,$O^{\delta2}$
3[e]	−4.7	−59.0	14.4	46.5	−2.9	S1; 219CO
4[d]	−4.0	−60.4	15.5	46.2	−2.7	S1; 219CO
5[e]	−4.7	−2.5	3.1	1.7	−2.4	S3; $Tyr^{60A}O^{\eta}$

[†]All energy values are in kcal/mol.
[a]Ranked among the clusters of the same functional group type according to total binding energy. Representatives with rank in bold are shown in Figure 4a,b.
[b]Sum of the values in the four preceding columns, i.e., intermolecular and electrostatic desolvation energies.
[c]For positions with more than one H-bond, the first H-bond partner in the list is the one which was used by SEED for docking.
[d]First conformation.
[e]Second conformation.

interesting to note that the second NH_2 is exposed to solvent; hence benzamidine 5 has a much less favorable electrostatic interaction and pays a smaller fragment desolvation than the first four positions.

5-Amidine indole. Representatives 1 and 2 are involved in the same three hydrogen bonds as benzamidine 1 and 2 in the S1 pocket (Fig. 4b) and show a strong electrostatic interaction energy (about −72 kcal/mol; Table VI). The indole moiety in position 1 is parallel to the benzene ring of NAPAP benzamidine, whereas it is rotated by about 50° in position 2. The indole NH in positions 1, 2, and 3 points toward the inner part of the protein, whereas it points toward the solvent for position 4. Representatives 3 and 4 make only one hydrogen bond to the carbonyl oxygen of residue Gly219 but still have strong electrostatic interaction energies due to the proximity of the amidine moieties with the Asp189 carboxylic group. Again, a large gap occurs between positions 1, 2, and positions 3, 4. Representative 5 is located close to the S3 pocket, and its amidine moiety is exposed to the solvent.

Methylguanidinium. Representative 1, whose binding energy is −27.7 kcal/mol, donates to the carbonyl group of Gly219 and to the two side chain oxygens of Asp189. It overlaps the arginine side chain of PPACK with a RMS deviation of 0.64 Å for the heavy atoms. The second position donates to the carbonyl groups of Ser214, Phe227, and Ala190 in the S1 pocket. Its binding energy is −17.8 kcal/mol. Position 20 is the first representative outside the S1 pocket and is located in the region between the S2 and S3 pockets with a binding energy of −2.0 kcal/mol. It makes a single hydrogen bond with the hydroxyl oxygen of Tyr60A.

Acetate. The first representative is located at the bottom of the S3 pocket. It acts as an acceptor in an hydrogen bond with the main-chain NH of residue Gly219. The electrostatic interaction energy is −4.8 kcal/mol, whereas the desolvation energy of the fragment is 13.5 kcal/mol. The total energy is positive with a value of 9.5 kcal/mol. Most of the following representatives have their positions in the same place apart from representative 32, which acts as an acceptor in a hydrogen bond with the hydroxyl O of Tyr60A.

Piperazine. All of the positions of the piperazine dication have an unfavorable binding energy mainly because of the large desolvation penalty of the fragment. Representative 1 donates to the hydroxyl oxygen of Tyr60A and has a binding energy of 11.2 kcal/mol (van der Waals −3.5 kcal/mol, screened electrostatic interaction −11.8 kcal/mol, thrombin desolvation 2.4 kcal/mol, and piperazine desolvation 24.0 kcal/mol). There is a representative in the S1 pocket involved in hydrogen bonds with the carboxy of Asp189 and the mainchain CO of Ser214. It has a very unfavorable binding energy (86.9 kcal/mol, rank 75) because of the large piperazine desolvation (161.2 kcal/mol), which is balanced only partially by the intermolecular electrostatic term (−106.2 kcal/mol).

Two main conclusions can be drawn from the analysis of the positions of the charged functional groups. First, the best positions of the +1 charged groups (but not the piperazine dication) have optimal hydrogen bonds with the Asp189 side chain in the S1 pocket. The low binding energy results from a weak shielding of the charge–charge interaction and a not-too-large desolvation of the carboxylate oxygens of Asp189. Second, binding to polar groups on the protein surface may lead to an unfavorable total binding energy because of the high desolvation penalty.

Most of the representatives of the apolar, polar, and charged groups found in a previous study[7] are reproduced in this work. Only two exceptions arise: a too-high desolvation of the protein in the region of Lys60F hinders the docking of N-methylacetamide close to the CO group of Leu41, and a large desolvation of the methylammonium functional group disfavors binding to the main chain oxygens of residues Ser214 and Gly216 or close to the

hydroxyl oxygen of Ser195. Hence, for small and rigid fragments SEED is able to dock exhaustively and evaluate a binding energy which includes electrostatic solvation effects. This represents a major improvement with respect to previous programs such as MCSS[45] which evaluates energies in vacuo.

Computation times

Depending on the fragment type SEED requires between 0.5 and 1.5 hour CPU time of an SGI R10000 processor for the docking in a medium-size binding site (10 Å sphere) with evaluation of the continuum solvation. For the application to thrombin presented here the CPU time ranged from 25 minutes (phenol) to 56 (benzene and benzamidine) and 98 minutes (methylguanidinium).

CCLD Results

As a further test of the usefulness of SEED for ligand design, the fragments docked by SEED were then connected by the program CCLD.[7] A maximum of ten members within each cluster were used for cyclohexane (270 positions), benzene (200), *N*-methyl-methylsulfonamide (200), *N*-methylacetamide (100), and naphthalene (100). For benzamidine all the members of the first four clusters in the S1 pocket were employed (52 with the first conformation and 41 with the mirror image conformation) and a total of 100 positions for methylguanidinium. The following values in Å were used for the distance (d) between linkage atoms for each connection type[7]: $d < 0.65$, 0-bond (the two atoms are fused); $1.4 < d < 1.62$, 1-bond (a single covalent bond); $2.10 < d < 3.0$, 2-bond (a single atom linker); $3.20 < d < 4.50$, 3-bond. The dihedral angles were checked around Csp^3-Csp^3 and Csp^2-Csp^3 bonds. A deviation of 40° was tolerated from the minimal values of the dihedral angles (+60°, −60° and 180° for Csp^3-Csp^3; −90° and +90° for Csp^2-Csp^3). The candidate ligands with less than 20 or more than 60 atoms were discarded. CCLD generated 390 candidate ligands in 33 minutes on a R10000 processor. Four interesting hits which ranked as 2nd, 37th, 42nd, and 90th are shown in Figure 5c–f.

Hits 1 and 2 (Fig. 5c,d) are similar to Argatroban (Fig. 5b), which is a reversible inhibitor of thrombin with a K_i of 19 nM.[40,46] They have nonpolar groups in S3 and S2 and a guanidinium in S1. The sulfonamide NH of compounds **1**, **2**, and Argatroban donates a hydrogen bond to the backbone CO of Gly216. Moreover, the carbonyl group in **2** and Argatroban accepts from the NH group of Gly216. Additional hydrogen bonds with respect to Argatroban are present in **1** and **2** namely between an SO_2 oxygen and the NH group of Gly219 and between the guanidinium and the main chain CO of Gly219. Furthermore, the amide group close to the guanidinium in compound **1** donates to the carbonyl of Ser214. Argatroban has less polar interactions with thrombin than hits **1** and **2**, but its double ring moiety fills the S3 pocket better than the cyclohexyl ring of compounds **1** and **2**. Hits **3** and **4** (Figs. 5e,f) have a benzamidine in the S1 pocket and a benzene in S2. The benzamidine moiety of compound **3** donates to the two oxygens of the Asp189 side chain and to the carbonyl group of Gly219. In the S3 pocket the hydroxyl substituent of cyclohexane donates to the main chain CO of Glu97A. Compound **4** is similar to 4-TAPAP (Fig. 5a), a reversible inhibitor of thrombin[40,47] whose racemic mixture has a K_i of 640 nM. In 4-TAPAP and hit **4** the benzamidine is involved in a salt bridge with the Asp189 side chain, the sulfonyl part accepts from the NH group of Gly219, and one NH donates to the carbonyl group of Gly216. The interaction with NH of Gly216 is missing in **4**, but there are two additional hydrogen bonds, with the CO groups of Ser214 and Gly219. This last hydrogen bond also occurs in the NAPAP thrombin complex. The naphthalene ring of **4** fills the S3 pocket as NAPAP.[48]

These representative examples and visual analysis of others SEED–CCLD hits indicate that the present approach generates candidate ligands with interaction patterns similar to known thrombin inhibitors. It is also clear that the SEED–CCLD hits presented above are much closer to known inhibitors than the MCSS–CCLD hits presented in our previous work.[7] This is mainly due to the fact that the fragments are docked with a more accurate energy evaluation in SEED and not exclusively in local minima of the *vacuum* force field.

CONCLUSIONS

We have presented an automatic approach for docking small-to-medium-sized molecular fragments to the binding site of a rigid protein. A comprehensive electrostatic method based on the continuum approximation allows the accurate and efficient evaluation of binding energies by taking into account the electrostatic contribution to the potential of mean force of the high dielectric solvent. The accuracy was assessed by a comparison with finite difference solutions of the Poisson equation for more than 2,500 complexes of small molecules or ions with thrombin and the HIV-1 PR monomer. The electrostatic energies in solution calculated with the two methods are in very good agreement with correlation coefficients larger than 0.9 in most cases. The continuum electrostatic approach in SEED requires about 0.1 seconds for the evaluation of the three electrostatic contributions of a ligand–receptor complex, i.e., the protein and fragment desolvation energies upon binding and the screened electrostatic interaction. For the same system the solution of the finite difference Poisson equation needs between 2 and 5 minutes. The gain in CPU-time of more than three orders of magnitude is due to our fast continuum electrostatic approach[5] and its optimal implementation into SEED.

A number of very efficient docking programs have been published recently. Most of them use either a scoring function with a crude approximation of solvation[49–51] or a *vacuum* energy derived from a molecular mechanics force field.[52,53] The program DOCK, which pioneered the use of geometric criteria to select ligands which best complement the shape of the receptor site,[54,55] has recently been supplemented by the evaluation of ligand desolvation.[56,57] To efficiently screen large databases of compounds, DOCK assumes that every ligand desolvates the receptor equally and that the ligand is completely desolvated upon binding.

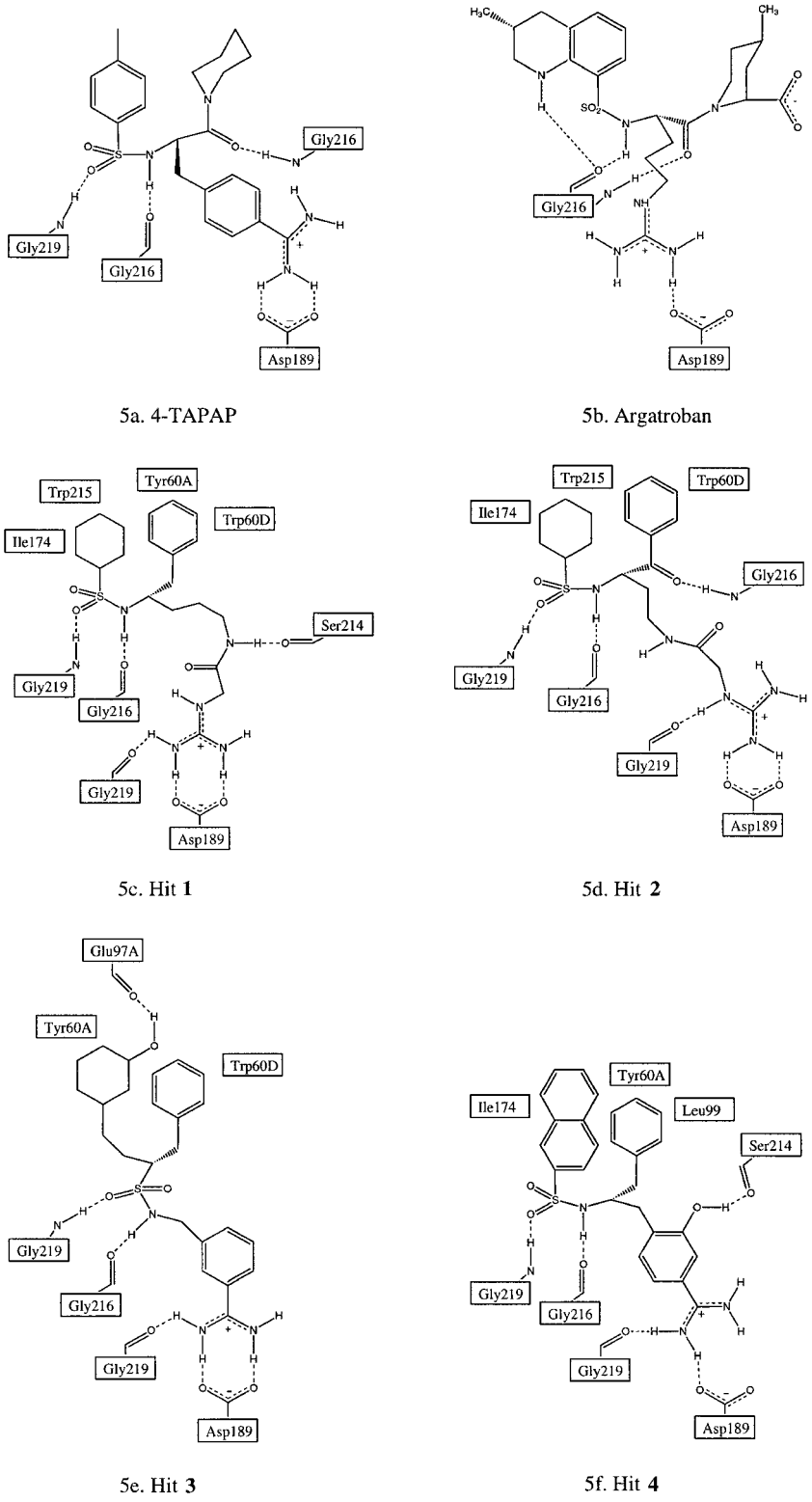

Fig. 5. Schematic representations of the interactions between thrombin and (**a**) 4-TAPAP, (**b**) Argatroban, and (**c–f**) CCLD hits **1** to **4**.

The continuum electrostatic approach implemented in SEED does not make these assumptions.

There are two main advantages in SEED with respect to the multiple copy simultaneous search method (MCSS), which is a force field–based approach for determining optimal positions and orientations of functional groups in a protein binding site.[45,58–61] These are the inclusion of electrostatic solvation and the determination of all favor-

able binding modes. The effects of the solvent are completely neglected in MCSS, which calculates the protein–fragment interactions with a *vacuum* potential.[36] This choice in MCSS was based on the principle that fast methods are necessary to perform effective searches of the binding site and that good candidate ligands subsequently can be ranked in terms of their binding energy. A possible difficulty is that minimized positions may be missed or misplaced due to the lack of a solvation correction during the MCSS minimization. The best energy minima without solvation do not always turn out to be those of the most interest.[60] Particularly problematic is the docking of apolar fragments which, without inclusion of solvation, are positioned in both hydrophilic and hydrophobic regions of the binding site.[7] This problem is solved in SEED by the prioritization of apolar regions on the protein surface according to low electrostatic desolvation and favorable van der Waals interactions, as well as the efficient use during docking of a protein desolvation look-up table. Moreover, SEED can find favorable positions which do not correspond to local minima of the energy function (e.g., a favorable region with relatively flat potential energy in between two well-pronounced minima). This is an advantage with respect to MCSS because not all of the molecular fragments, even in potent inhibitors, have optimal interactions with the protein.[60]

Fragments are docked as rigid bodies by SEED. For larger ligands with rotatable bonds, conformational flexibility can be taken into account by docking different conformations, as was done with benzamidine and 5-amidine indole in the present study. Programs are available for the automatic generation of diverse low-energy conformations of small molecules.[62,63] In the case of large ligands with many rotatable bonds one could use SEED to find optimal positions for the rigid moieties and then use other techniques which allow for full ligand and eventually also protein flexibility.[64]

An application of SEED to human thrombin has been presented. The approximation of a rigid protein is acceptable for thrombin, which mostly assumes the same conformation in complexes with different low-molecular-weight inhibitors.[40,48,65] SEED reproduces the crucial structural and energetic aspects of the interactions between known inhibitors and thrombin. The fragments docked by SEED were used as input for the ligand design program CCLD, which constructs larger compounds by combinatorial principles. The SEED–CCLD approach yielded a number of candidate thrombin inhibitors with the salient features of known potent thrombin inhibitors such as NAPAP, Argatroban, and 4-TAPAP, as well as additional interactions.

The SEED–CCLD strategy uses combinatorial principles to construct candidate ligands. Possible applications are for de novo design, as documented here for thrombin, lead optimization, and the selection of synthons for combinatorial chemistry.

ACKNOWLEDGMENTS

We thank S. Ahmed, N. Budin, and P. Ferrara for helpful discussions and A. Widmer (Novartis Pharma Inc., Basel) for the molecular modelling program WITNOTP which was used for visual analysis of the results and for preparing Figures 3b–d and 4. We also thank Prof. J.A. McCammon for providing the UHBD program, which was used for all the finite difference calculations. The program SEED (for SGI or PC running the Linux operating system) as well as the library of fragments in mol-2 format is available for not-for-profit institutions from A. Caflisch.

REFERENCES

1. Veerapandian P, editor. Structure-based drug design. New York: Marcel Dekker Inc.; 1997.
2. Kubinyi H. Structure-based design of enzyme inhibitors and receptor ligands. Curr Opin Drug Design Discov 1998;1:4–15.
3. Caflisch A, Wälchli R, Ehrhardt C. Computer-aided design of thrombin inhibitors. News Physiol Sci 1998;13:182–189.
4. Caflisch A, Karplus M. Computational combinatorial chemistry for de novo ligand design: review and assessment. Perspect Drug Discov Design 1995;3:51–84.
5. Scarsi M, Apostolakis J, Caflisch A. Continuum electrostatic energies of macromolecules in aqueous solutions. J Phys Chem A 1997;101:8098–8106.
6. Scarsi M, Apostolakis J, Caflisch A. Comparison of a GB solvation model with explicit solvent simulations: potentials of mean force and conformational preferences of alanine dipeptide and 1,2-dichloroethane. J Phys Chem B 1998;102:3637–3641.
7. Caflisch A. Computational combinatorial ligand design: application to human α-thrombin. J Comput Aided Mol Design 1996;10: 372–396.
8. Marrone TJ, Gilson MK, McCammon JA. Comparison of continuum and explicit models of solvation: potential of mean force for alanine dipeptide. J Phys Chem 1996;100:1439–1441.
9. Caflisch A, Fischer S, Karplus M. Docking by Monte Carlo minimization with a solvation correction: application to an FKBP-substrate complex. J Comput Chem 1997;18:723–743.
10. Goodford PJ. A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. J Med Chem 1985;28:849–857.
11. Goodsell DS, Olson AJ. Automated docking of substrates to proteins by simulated annealing. Proteins 1990;8:195–202.
12. Meng EC, Shoichet BK, Kuntz ID. Automated docking with grid-based energy evaluation. J Comput Chem 1992;13:505–524.
13. Luty BA, Wasserman ZR, Stouten PFW, Hodge CN, Zacharias M, McCammon JA. A molecular mechanics/grid method for evaluation of ligand-receptor interactions. J Comput Chem 1995;16: 454–464.
14. Press WH, Teukolsky SA, Vetterling WT, Flannery BP. Numerical recipes in Fortran. Cambridge: Cambridge University Press; 1992.
15. Warwicker J, Watson HC. Calculation of the electric potential in the active site cleft due to α-helix dipoles. J Mol Biol 1982;157: 671–679.
16. Gilson MK, Honig BH. Calculation of the total electrostatic energy of a macromolecular system: solvation energies, binding energies, and conformational analysis. Proteins 1988;4:7–18.
17. Bashford D, Karplus M. pK_a's of ionizable groups in proteins: atomic detail from a continuum electrostatic model. Biochemistry 1990;29:10219–10225.
18. Davis ME, Madura JD, Luty BA, McCammon JA. Electrostatics and diffusion of molecules in solution: simulations with the University of Houston Brownian dynamics program. Comp Phys Comm 1991;62:187–197.
19. Still WC, Tempczyk A, Hawley RC, Hendrickson T. Semianalytical treatment of solvation for molecular mechanics and dynamics. J Am Chem Soc 1990;112:6127–6129.
20. Hawkins GD, Cramer CJ, Trulhar DG. Pairwise solute descreening of solute charges from a dielectric medium. Chem Phys Lett 1995;246:122–129.
21. Hawkins GD, Cramer CJ, Trulhar DG. Parametrized models of aqueous free energies of solvation based on pairwise descreening of solute atomic charges from a dielectric medium. J Phys Chem 1996;100:19824–19839.

22. Schaefer M, Karplus M. A comprehensive analytical treatment of continuum electrostatics. J Phys Chem 1996;100:1578–1599.
23. Qiu D, Shenkin PS, Hollinger FP, Still WC. The GB/SA continuum model for solvation. A fast analytical method for the calculation of approximate Born radii. J Phys Chem A 1997;101:3005–3014.
24. Richards FM. Areas, volumes, packing, and protein structure. Annu Rev Biophys Bioeng 1977;6:151–176.
25. Jackson JD. Classical electrodynamics. New York: John Wiley & Sons; 1963. p 124.
26. Davis ME, McCammon JA. Solving the finite difference linearized Poisson-Boltzmann equation: a comparison of relaxation and conjugate gradient methods. J Comput Chem 1989;10:386–391.
27. Davis ME, McCammon JA. Calculating electrostatic forces from grid-calculated potentials. J Comput Chem 1990;11:401–409.
28. Davis ME, McCammon JA. Dielectric boundary smoothing in finite difference solutions of the Poisson equation: an approach to improve accuracy and convergence. J Comput Chem 1991;12: 909–912.
29. Kearsley SK, Smith GM. An alternative method for the alignment of molecular structures: Maximizing electrostatic and steric overlap. Tetrahedron Comput Methodol 1990;3:615–633.
30. Klebe G, Mietzner T, Weber F. Different approaches toward an automatic structural alignment of drug molecules: applications to sterol mimics, thrombin and thermolysin inhibitors. J Comput Aided Mol Design 1994;8:751–778.
31. Bernstein FC, Koetzle TF, Williams GJB, Meyer EF Jr, Brice MD, Rodgers JR, Kennard O, Shimanouchi T, Tasumi M. The protein data bank: a computer-based archival file for macromolecular structures. J Mol Biol 1977;112:535–542.
32. Gasteiger J, Marsili M. Iterative partial equalization of orbital electronegativity. A rapid access to atomic charges. Tetrahedron 1980;36:3219–3288.
33. No KT, Grant JA, Scheraga HA. Determination of net atomic charges using a modified partial equalization of orbital electronegativity method. 1. Application to neutral molecules as models for polypeptides. J Phys Chem 1990;94:4732–4739.
34. No KT, Grant JA, Jhon MS, Scheraga HA. Determination of net atomic charges using a modified partial equalization of orbital electronegativity method. 2. Application to ionic and aromatic molecules as models for polypeptides. J Phys Chem 1990;94:4740–4746.
35. Momany FA, Klimkowski VJ, Schäfer L. On the use of conformationally dependent geometry trends from ab initio dipeptide studies to refine potentials for the empirical force field CHARMM. J Comput Chem 1990;11:654–662.
36. Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M. CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. J Comput Chem 1983;4: 187–217.
37. Kettner C, Shaw E. D-Phe-Pro-Arg-CH_2Cl: a selective affinity label for thrombin. Thromb Res 1979;14:969–973.
38. Bode W, Mayr I, Baumann U, Huber R, Stone SR, Hofsteenge J. The refined 1.9-Å crystal structure of human α-thrombin: interaction with D-Phe-Pro-Arg chloromethylketone and significance of the Tyr-Pro-Pro-Trp insertion segment. EMBO J 1989;8:3467–3475.
39. Stürzebecher J, Markwardt F, Voigt B, Wagner G, Walsmann P. Cyclic amides of Nα-arylsulfonylaminoacylated 4-amidinophenylalanine, tight binding of thrombin. Thromb Res 1983;29:635–642.
40. Brandstetter H, Turk D, Hoeffken HW, Grosse D, Stuerzebecher J, Martin DP, Edwards BFP, Bode W. Refined 2.3 Å X-ray structure of bovine thrombin complexes formed with the benzamidine and arginine-based thrombin inhibitors NAPAP, 4-TAPAP and MQPA. J Mol Biol 1992;226:1085–1099.
41. Hilpert K, Ackermann J, Banner DW, Gast A, Gubernator K, Hadvary P, Labler L, Müller K, Schmid G, Tschopp T, van de Waterbeemd H. Design and synthesis of potent and highly selective thrombin inhibitors. J Med Chem 1994;37:3889–3901.
42. Tapparelli C, Metternich R, Ehrhardt C, Cook NS. Synthetic low-molecular weight thrombin inhibitors: molecular design and pharmacological profile. TIPS 1993;14:366–376.
43. Lyle TA. Small-molecule inhibitors of thrombin. Perspect Drug Discov Design 1993;1:453–460.
44. Iwanowicz EJ, Lau WF, Lin J, Roberts DGM, Seiler SM. Derivatives of 5-amidine indole as inhibitors of thrombin catalytic activity. Bioorganic Med Chem Lett 6:1339–1344, 1996.
45. Miranker A, Karplus M. Functionality maps of binding sites: a multiple copy simultaneous search method. Proteins 1991;11: 29–34.
46. Kikumoto R, Tamao Y, Tezuka T, Tonomura S, Hara H, Ninomiya K, Hijikata A, Okamoto S. Selective inhibition of thrombin by (2R,4R)-4-methyl-1-[N2-[(3-methyl-1,2,3,4-tetrahydro-8-quinolinyl) sulfonyl]-L-arginyl)]-2-piperidinecarboxylic acid. Biochemistry 1984;23:85–90.
47. Stürzebecher J, Walsmann P, Voigt B, Wagner G. Inhibition of bovine and human thrombins by derivatives of benzamidine. Thromb Res 1984;36:457–467.
48. Banner DW, Hadvary P. Crystallographic analysis at 3.0-Å resolution of the binding to human thrombin of four active site-directed inhibitors. J Biol Chem 1991;266:20085–20093.
49. Rarey M, Kramer B, Lengauer T, Klebe G. A fast flexible docking method using an incremental construction algorithm. J Mol Biol 1996;261:470–489.
50. Eldridge MD, Murray CW, Auton TR, Paolini GV, Mee RP. Empirical scoring functions: I. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. J Comput Aided Mol Design 1997;11:425–445.
51. Baxter CA, Murray CW, Clark DE, Westhead DR, Eldridge MD. Flexible docking using tabu search and an empirical estimate of binding affinity. Proteins 1998;33:367–382.
52. Jones G, Willett P, Glen RC, Leach AR, Taylor R. Development and validation of a genetic algorithm for flexible docking. J Mol Biol 1997;267:727–748.
53. Lorber DM, Shoichet BK. Flexible ligand docking using conformational ensembles. Protein Sci 1998;7:938–950.
54. Kuntz ID, Blaney JM, Oatley SJ, Langridge R, Ferrin TE. A geometric approach to macromolecule-ligand interactions. J Mol Biol 1982;161:269–288.
55. DesJarlais RL, Sheridan RP, Dixon JS, Kuntz ID, Venkataraghavan R. Docking flexible ligands to macromolecular receptors by molecular shape. J Med Chem 1986;29:2149–2153.
56. Shoichet BK, Stroud RM, Santi DV, Kuntz ID, Perry KM. Structure-based discovery of inhibitors of thymidylate synthase. Science 1993;259:1445–1450.
57. Shoichet BK, Leach AR, Kuntz ID. Ligand solvation in molecular docking. Proteins 1999;34:4–16.
58. Caflisch A, Miranker A, Karplus M. Multiple copy simultaneous search and construction of ligands in binding sites: application to inhibitors of HIV-1 aspartic proteinase. J Med Chem 1993;36:2142–2167.
59. Miranker A, Karplus M. An automated method for dynamic ligand design. Proteins 1995;23:472–490.
60. Grootenhuis PDJ, Karplus M. Functionality map analysis of the active site cleft of human thrombin. J Comput Aided Mol Design 1996;10:1–10.
61. Joseph-McCarthy D, Hogle JM, Karplus M. Use of the multiple copy simultaneous search (MCSS) method to design a new class of picornavirus capsid binding drugs. Proteins 1997;29:32–58.
62. Kolossvary I, Guida WC. Low mode search. An efficient, automated, computational method for conformational analysis: Application to cyclic and acyclic alkanes and cyclic peptides. J Am Chem Soc 1996;118:5011–5019.
63. Klebe G, Mietzner T. A fast and efficient method to generate biologically relevant conformations. J Comput Aided Mol Design 1994;8:583–606.
64. Apostolakis J, Plückthun A, Caflisch A. Docking small ligands in flexible binding sites. J Comput Chem 1998;19:21–37.
65. Stubbs MT, Bode W. Crystal structures of thrombin and thrombin complexes as a framework for antithrombotic drug design. Perspect Drug Discov Design 1993;1:431–452.
66. Nicholls A, Sharp KA, Honig B. Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. Proteins 1991;11:281–296.