PROTEINS: Structure, Function, and Genetics 15:436–444 (1993)

Computer Simulation of Antibody Binding Specificity

Matteo Pellegrini[1] and Sebastian Doniach[2]
Departments of [1]Physics and [2]Applied Physics, Stanford University, Stanford, California 94305-4090

ABSTRACT **A Monte Carlo algorithm that searches for the optimal docking configuration of hen egg white lysozyme to an antibody is developed. Both the lysozyme and the antibody are kept rigid. Unlike the work of other authors, our algorithm does not attempt to explicitly maximize surface contact, but minimizes the energy computed using coarse-grained pair potentials. The final refinement of our best solutions using all-atom OPLS potentials (Jorgensen and Tirado-Rives[8]) consistently yields the native conformation as the preferred solution for three different antibodies. We find that the use of an exponential distance-dependent dielectric function is an improvement over the more commonly used linear form.**

Key words: antigen-antibody recognition, docking algorithm, distance-dependent dielectric

INTRODUCTION

Understanding the molecular basis for specificity of receptor-substrate binding in general and immune specificity in particular is a problem of current interest in molecular biology. Over the last few years the structures of crystals of three complexes of antibodies bound to different epitopes on the surface of hen egg white lysozyme (HEL) have been solved to high resolution.[1–3] These provide a beautiful test case for our ability to model antibody-protein binding specificity, since there is evidence that the conformational changes which take place upon docking of the lysozyme to an antibody are small (on the scale of 1Å).[17,18] In this paper, we report on a new approach to modeling the specificity of antibody-lysozyme binding based on a rigid body docking strategy.

A number of authors have recently reported computer studies of docking.[4–7] In general, their algorithms search six-dimensional phase space (3 translations and 3 rotations) for the conformation that maximizes the contact area between the proteins. However, in all cases this leads to a large number of equally good candidates, among which is one similar to the native conformation found in the crystal. The problem of distinguishing the best one among these solutions was addressed by Cherfils et al.[4] and by Shoichet and Kuntz.[5] Both groups used interatomic potentials, including electrostatic terms, to refine the best solutions. Shoichet and Kuntz[5] also computed the buried surface area and the solvation free energy. Although the native solution was among the very best, neither group could disqualify all incorrect solutions. A few conformations, far from the native one, had energies very close to or even lower than the native one.

We have addressed this problem and have developed a strategy that consistently filters out non-native solutions and selects the native solution for all three antibody-lysozyme complexes. The algorithm we constructed to generate docking conformations uses successively more sophisticated forms of intermolecular potentials instead of explicitly trying to maximize surface contact. The best solutions generated in this manner are then refined using the OPLS potentials, created by Jorgensen and Tirado-Rives.[8] We find that the binding energies of the conformations generated at this stage of the search are very sensitive to the method used to simulate dielectric screening of the Coulomb potential. The most successful strategy we found was to introduce a distance-dependent dielectric constant leading to an exponentially screened Coulomb potential with a characteristic length of 3 Å, and a switching function from 9 to 10 Å. This potential, when applied to optimize the binding of our top ten solutions, yielded as the top answer a conformation within 1 Å rms of the native. A gap corresponding to roughly 20% of the top binding energy appeared between the best and second best solutions.

It is important to stress that our docking algorithm (like those of other authors) does not compute the binding affinity of the antigen to the antibody. Such a calculation would have to include effects due to changes in hydration and changes in entropy upon binding.[12,16] Our selection criteria are based only on comparison of binding energies which include electrostatic and van der Waals effects. One of our conclusions is that the electrostatic component of protein-protein interactions plays a significant role in determining immune specificity.

In the following sections, we give details of our

Received June 18, 1992; revision accepted October 5, 1992.
Address reprint requests to Dr. Sebastian Doniach, Department of Applied Physics, Stanford University, Stanford, CA 94305-4090.

TABLE I. Example of Hinge Points for the Three Complexes*

Complex	Atom type	Residue type	Residue number	Heavy or light chain	Distance to (Å) hinge point
HY10	CD2	TYR	33	H	4.8
	HE1	TRP	98	H	2.1
	HH	TYR	50	L	4.8
HY5	CH2	TRP	90	H	3.3
	HE1	TRP	33	L	4.7
	OE1	GLU	50	L	2.8
D1.3	CA	ASP	100	H	4.3
	OD2	ASP	100	H	1.6
	CE1	TYR	101	H	4.6

*The hinge points selected for three of our runs are given in terms of their distances from three reference atoms. As mentioned above, hinge points within ± 2.5 Å of the above ones will also allow the algorithm to function well.

searching strategy. In particular we will discuss the two novel components of our algorithm: (1) use of a set of intermediate binding energy criteria to reject unfavorable search paths, and (2) use of a phenomenologically screened Coulomb potential which smooths out barriers between local minima at the final stage of the search. We will also discuss whether our novel use of an exponentially screened dielectric function is generalizable to other problems, or whether it is a computational device specific to our situation.

METHODS

Docking Algorithm

Our docking algorithm is conceptually very simple. It can be viewed as a series of three filters, each of which selects only 10% of the conformations fed into it; the final solutions are then refined using all-atom OPLS potentials. We will first describe the method used to obtain the initial conformations, and then describe each filter in detail.

Selection of Initial Conformations

As discussed by Tramontano et al.[23] the complimentarity determining region (CDR) of an antibody is formed by six loops: L1, L2 and L3 are part of the variable domain of the light chain, and H1, H2 and H3 are part of the variable domain of the heavy chain. We select a region approximately 5 Å wide between the H3 and L3 loops. We allow this region to span an area between 2 and 7 Å from the surface of the antibody. From within this region we select a hinge point. For examples of hinge points used in our simulations, see Table I. Below we will discuss the dependence of the algorithm's performance on the location of the hinge point within this region. Note that the choice of hinge point, while localizing the binding site to be in the vicinity of the CDR, does not require any knowledge of the antigen's native docked conformation, and so it does not bias the simulation towards such a conformation.

We then attach each of a selected set of "fiducial" atoms (to be defined below) in the lysozyme molecule to this point, and rotate the lysozyme so that its center of mass is as far from the antibody as possible. For each fiducial atom we then rotate the lysozyme in 30 degree steps around the axis formed by the initial point and the center of mass of the antibody fragment. To make sure that we are not biasing our initial configurations we rotate the lysozyme by an arbitrary fixed angle around the same axis before applying the 30 degree rotations. We ran the algorithm for various values, between 0 and 30 degrees, of this angle. This yields approximately 10,000 initial conformations.

The selection of our initial conformation by this method has two important benefits. Because our hinge point is located at the approximate center of the antibody's antigen binding site, and only a few angstroms from the surface, we are sure that it must lie within the lysozyme when it is in its native docking conformation. Furthermore, since by the above procedure we create 10,000 initial conformations aligned with the center of mass away from the surface, we are sure that at least a few are close (within 10 Å rms) to the native conformation.

Fiducial Atoms

The set of fiducial atoms are selected from a list that attempts to choose predominantly atoms involved in surface interactions, such as those that form strong dipoles or are at the extreme end of a side chain. In all, this includes 58 atom types. We include the main chain oxygen and nitrogen, as well as the oxygens, nitrogens, and hydrogens from polar side chains. In nonpolar residues we typically select the carbon most distant from the alpha carbon. For a complete listing of fiducial atoms, see Table II. The

TABLE II. List of "Fiducial" Atoms Used in Coarse-Grained Pair Potentials*

Residue type	Side chain "fiducial atoms"
ALA	CB
ARG	NH1, HH11, HH12, NH2, HH21, HH22, NE
ASN	OD1, ND2, HD21, HD22
ASP	OD1, OD2
CYS	SG
GLN	OE1, NE2, HE21, HE22
GLU	OE1, OE2
HIS	ND1, NE2
ILE	CD1, CG1
LEU	CD1, CD2
LYS	CD, NZ, HZ1, HZ2, HZ3
MET	SD, CE
PHE	CD1, CD2, CZ
PRO	CG
SER	OG, HG
THR	OG1, HG1, CG2
TRP	CZ2, CZ3, NE1, HE1, CB
TYR	OH, CD1, CD2, HH
VAL	CG1, CG2

*List of side chain "fiducial" atoms used in our coarse-grained potentials. The main chain "fiducial" atoms include oxygen and nitrogen for all residues, and the alpha carbon for glycine.

"fiducial" atoms account for approximately half of the total atoms of lysozyme and the antibodies.

The "essential" hydrogens, those that are charged and participate in hydrogen bonds, are added to our Brookhaven Protein Data Bank (PDB) files using standard bond angles and lengths by the program SYBYL.[25] The hydrogens that are covalently bonded to carbons are not explicitly included, and the carbon is treated as an "extended" atom, following the convention used in CHARMM.[11]

Filters

We allow our initial conformations to relax using a Monte Carlo Metropolis algorithm[9] with a very simplified interatomic potential (only between fiducial atoms) designed to avoid steric clashes and maximize contact area. This potential is equally repulsive for all interatomic distances less than 2 Å, and attractive for distances less than five, and zero thereafter. The magnitude of the repulsive component is 3.0 Kcal/mole and −0.2 for the attractive one.

In our first stage we allow the 10,000 initial conformations to minimize for 50 Monte Carlo moves selected using the Metropolis algorithm from a range of 0.5 Å in translation and 2 degrees in rotation. We have found that one obtains more accurate results if each conformation is allowed to relax twice, starting from the same initial state but using different random number seeds, for 50 time steps, instead of once for 100 time steps. Therefore, initially we perform 20,000 relaxations of fifty time steps. The temperature is kept fixed throughout the run, and set to a value that approximates room temperature (kT = 0.6 Kcal/mole).

In the next stage we select the top 10% of these solutions, and minimize for another 50 time steps using coarse-grained statistical potentials. These are generalizations of potentials defined by Wilson and Doniach[10] that include side chain information (see also Sippl[19]). The potentials are referred to as "statistical" because they are derived from an analysis of pair correlations between "fiducial" atoms in the Brookhaven Protein Data Bank. Since PDB files do not include hydrogen atoms, our correlations are measured between all possible pairs of nonhydrogen atoms in our list of "fiducial" atoms. This leads to 800 potentials in all. The interactions between hydrogen and other atoms are not included in this stage, but are included in the next filter. The correlations are used to generate effective potentials by taking the natural logarithm of the resulting distribution functions and normalizing to zero at 10 Å interatomic distance. The total interaction energy for a given configuration is then set equal to the sum of potentials for all applicable interprotein fiducial atom pairs. Since these potentials are constructed from empirical distribution functions we believe that they incorporate, in an approximate way, effects due to electrostatic and hydrophobic interactions.

The energies of the top 10% of these solutions are then further minimized using the previous statistical potentials to which have been added hydrogen bonding terms from the OPLS potentials (in the OPLS potentials the hydrogen bonds are not given special treatment; they are represented by a Lennard-Jones 6, 12 potential). We weigh the value of the hydrogen bonds very heavily with respect to the statistical interactions: each hydrogen bond is arbitrarily multiplied by a factor of ten, while the statistical pair potentials are kept equal to their value described above. This allows this filter to primarily select conformations that are favorable in terms of hydrogen bonding.

Up to this point, all the pair potentials had been computed between "fiducial" atoms only. In the final stage we apply a full-atom description of the antigen-antibody complex. The final top ten solutions obtained by the above procedure were minimized for 1,000 MC steps using the full atom OPLS potentials. We paid particular attention to the method used for dielectric screening of the long range Coulomb term, as discussed below. Executing the entire algorithm requires approximately 30 hours of cpu time on a DECstation 5000/200.

Dielectric Screening of the Coulomb Potential

The nature of dielectric screening in proteins and the effects of the solvating water have been discussed by several authors.[15,20,21,26] In general it will consist of a term due to the reorientation of dipoles

both in the water and in the protein, an electronic polarizability term and a term resulting from ionic displacements in the water. In principle, if the electronic term can be taken care of by a homogeneous uniform dielectric constant, the effects of dipole reorientation and ionic displacements will automatically be taken care of if a full molecular dynamic simulation is performed, with resulting dipolar relaxation. However, this is prohibitive when a large number of docking configurations need to be examined. Therefore, a number of workers in the field have proposed distance-dependent dielectric functions which might be used as an ad hoc method to simulate the polarizability of proteins including the effects of the solvent, thus allowing rigid body estimates of Coulomb interactions between proteins or between different parts of a given protein (reviewed in reference 24). Distance-dependent dielectric functions which have been proposed include a linear form $\epsilon(r) = r$ and more complex forms, e.g., a form including an exponential term due to Mehler[14] (see also Warshel[22]). Other authors have developed additional pairwise energy terms that explicitly attempt to account for charge-solvent interactions;[26] however, for simplicity we have not yet included such terms in our simulation.

In addition to the problem of simulating the polarizability of proteins in solution, there is an additional computational problem resulting from the long range of the Coulomb potential, even when screened. This has been addressed by Brooks et al. in CHARMM[11] by introducing a switching function to cut off the potential. Alternately, one can group atoms to form neutral subunits which then interact through dipoles and higher order poles which decay faster than $1/r$. This technique is also addressed in CHARMM, although in this work we have used exclusively the simpler switching function approach.

We have found that the screening and cutoff methods used are critical to the success of the algorithm. For instance, we were unable to show that the native solution is best when we used a constant dielectric, regardless of its value, and discontinuously cut off Coulomb potential at 10 Å.

In our simulations we tried four methods, three from CHARMM and one developed independently. The methods used by CHARMM are:

(a) Constant dielectric

$$\Delta E = \sum_{i}^{\text{on } a} \sum_{j}^{\text{on } b} \frac{q_i q_j e^2}{r_{ij}\epsilon} \tag{1}$$

(b) Distance-dependent dielectric (linear)

$$\Delta E = \sum_{i}^{\text{on } a} \sum_{j}^{\text{on } b} \frac{q_i q_j e^2}{r_{ij}^2} sw\,(r_{ij}^2, r_{on}^2 r_{off}^2) \tag{2}$$

(c) Shifted dielectric

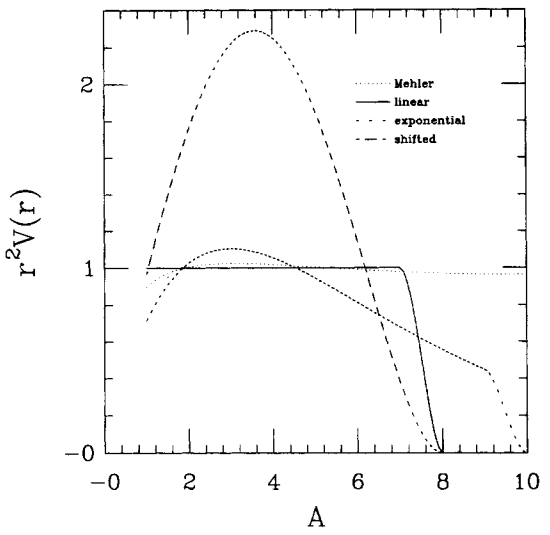

Fig. 1. Plot of $r^2V(r)$ for the four dielectric functions. Note that the switching function for the linear and shifted dielectric functions operates between 7 and 8 Å, while for the exponential dielectric function it operates between 9 and 10 Å (the Mehler dielectric function is not multiplied by a switching function). This is due to the fact that the inception and cutoff of the switching function were optimized for each dielectric function separately.

$$\Delta E = \sum_{i}^{\text{on } a} \sum_{j}^{\text{on } b} \frac{q_i q_j e^2}{r_{ij}} \left(1 - \frac{2r_{ij}^2}{r_{cut}^2} + \frac{2r_{ij}^4}{r_{cut}^4}\right). \tag{3}$$

We have found a fourth useful approximation:

(d) Exponentially decaying dielectric screening:

$$\Delta E = \sum_{i}^{\text{on } a} \sum_{j}^{\text{on } b} \frac{q_i q_j e^2}{r_{ij}} e^{-r_{ij}/3}\, sw(r_{ij}^2, r_{on}^2 r_{off}^2) \tag{4}$$

where r_{ij} is measured in Å. In all these cases the switching function is given by

$$sw(r_{ij}^2, r_{on}^2 r_{off}^2) = 1 \text{ when } r_{ij} \le r_{on} \tag{5}$$

$$sw(r_{ij}^2, r_{on}^2 r_{off}^2) = \frac{(r_{off}^2 - r_{ij}^2)^2 (r_{off}^2 + 2r_{ij}^2 - 3r_{on}^2)}{(r_{off}^2 - r_{on}^2)^3}$$

$$\text{when } r_{on} < r_{ij} \le r_{on} \tag{6}$$

$$sw(r_{ij}^2, r_{on}^2 r_{off}^2) = 0 \text{ when } r_{ij} > r_{on}. \tag{7}$$

In Figure 1 we plot $r^2V(r)$ as a function of r for the four different potentials. It can be seen that the major differences between exponential, linear and Mehler dielectric functions (to be defined later) occur after 6 Å while the shifted potential has a completely different form.

RESULTS

Selection of Preferred Docking Complex

We ran the three lysozyme-antibody complexes through the first three filters of our algorithm using the hinge points listed in Table I and nearby hinge points (within 3 Å rms). From these stages we ob-

TABLE III. Binding Energies Found in Sample Runs That Used an Exponentially Increasing Dielectric Function*

Complex	Lowest energy	rms distance	Second lowest energy	rms distance
Fab HyHEL5-lysozyme	−110	0.8	−77	14.0
Fab HyHEL10-lysozyme	−68	0.6	−49	17.2
Fab D1.3-lysozyme	−73	0.3	−59	17.0

*Binding energies calculated for the OPLS potentials with an exponentially screened Coulomb potential and a switching function at 9 Å. The rms values are calculated only for atoms within 15 Å of the antibody. This is used to avoid misleadingly large rms values that might be caused by the epitope acting as a pivot around which the lysozyme can rotate by small angles. The units of the OPLS potentials are Kcal/mole.

TABLE IV. Binding Energies Found in Sample Runs That Used a Linear Dielectric Function*

Complex	Lowest energy	rms distance	Second lowest energy	rms distance
Fab HyHEL5-lysozyme	−107	0.9	−67	14.3
Fab HyHEL10-lysozyme	−56	16.8	−39	1.8
	−64.7	0.9	−39	13.4
	−63.2	0.9	−35	18.2
Fab D1.3-lysozyme	−70	0.5	−55	16.7

*Binding energies for the OPLS potentials with a linear dielectric that has a switching function between 7 and 8 Å. In the case of HyHEL10, the OPLS potential with this dielectric function, starting from different but nearby hinge points (within 2 Å), was not able to consistently find a solution within 1 Å of the crystal.

TABLE V. Binding Energies Found in Sample Runs That Used a Shifted Dielectric Function*

Complex	Lowest energy	rms distance	Second lowest energy	rms distance
Fab HyHEL5-lysozyme	−42	0.9	−25	13.8
Fab HyHEL10-lysozyme	−20	0.7	−20	14.0
Fab D1.3-lysozyme	−61	0.7	−55	12.8

*Binding energies for the OPLS potentials with a shifted dielectric function that has a cutoff at 7 Å. In the case of HyHEL10, use of this dielectric function was not able to find an energy gap between the native solution and the next best solution with large rms deviation.

tained a list of the top ten docked conformations for each complex. These were then refined using the all-atom OPLS potentials with the four different phenomenological dielectric functions.

The results for the simulations are summarized in Tables III–V, for the various dielectric functions. For the case of a constant dielectric function we set the contributions from atoms separated by a distance greater than 10 Å to zero, but did not use a switching function. We did not include a table of results for this case since the results were very poor. This is due to the fact that the energy contribution from atoms separated by 10 Å is an order of magnitude larger than contributions from short distances. Thus, the binding energy is not a smooth function under this method, but varies wildly from one Monte Carlo move to the next.

To achieve a well-defined energy surface in docking space while limiting the number of atomic pairs computed, one needs to cut off the potentials smoothly. In Tables III–V we report some sample runs for the different dielectric functions. When we show only one run, this implies that all other runs with hinge points within 2 to 3 Å of this one obtained similar results.

As seen in the comparison of Tables III–V, the best method we found was to simulate the effects of the dielectric polarizability and solvent as leading to an exponentially decaying potential. The other two methods, linear and shifted dielectric functions, seemed to work well in the cases of the HyHEL5 and D1.3-lysozyme complexes, but did not consistently find the native as the top solution for the HyHEL10-lysozyme complex. For the case of HyHEL-10 and a

linear dielectric we report three runs with three different, but close (within 2 Å), hinge points. As can be seen, the algorithm failed to find the native in one of these runs, although when it did find a native-like conformation it turned out to be the one with lowest energy. On the other hand the shifted dielectric function consistently found a native-like solution but also found solutions that were far from the native (14 Å rms) but had equal energy (we have verified that with the exponentially increasing dielectric function the solution closest to the native is indeed lower in energy than these solutions). The exponentially increasing dielectric function instead found the native binding conformation consistently for all three antibodies, within a 1 Å rms deviation. It also demonstrated that for the top ten solutions found by our algorithm, the native one was energetically favorable and a gap of at least 20% of the native binding energy appeared between the top two solutions.

The use of an exponential dielectric function also leads to solutions that were closer, in terms of rms distance, to the native. For instance, the top solutions for the D1.3 antibody found using the exponential dielectric function was within 0.3 Å rms of the native, while that found using a linear dielectric function was within 0.5 Å. Thus, screening the Coulomb potential with an exponential dielectric function is more effective than the other methods both in its consistency, as seen above, and in its ability to select conformations that are closer to the one found in the crystal.

As stated above, we ran the algorithm several times to determine the dependence of the results on the initial hinge point and the initial angle of rotation. In the case of the D1.3 antibody-lysozyme complex, we selected three hinge points that varied from 2 to 7 Å from the surface, along the axis formed by the center of mass of the antibody and the first hinge point. In all cases we found results consistent with those stated above (i.e., a conformation within 1 Å rms of the native being at least 20% lower in energy than all other conformations). Similarly we allowed the initial angle of rotation (this is the rotation we apply to all our conformations before applying the 12, 30 degree rotations) to vary from 0 to 30 degrees, in 5 degree intervals, in the above simulations and found that its value had no effect on the results. Based on these tests we feel it is unlikely that we are biasing the initial conformations, and the subsequent search, towards the native conformation.

Hinge points selected outside the range described above yielded very poor results: among the top ten overall solutions none were close to the native, although a native-like solution was among the next ten. This suggests that one must be relatively careful in the selection of the hinge point, but with the rather large range of successful hinge points we found (2 to 7 Å from the surface, located between the L3 and H3 loops) this should not be a severe limitation. For the three complexes studied, a successful hinge point was found on our first attempt.

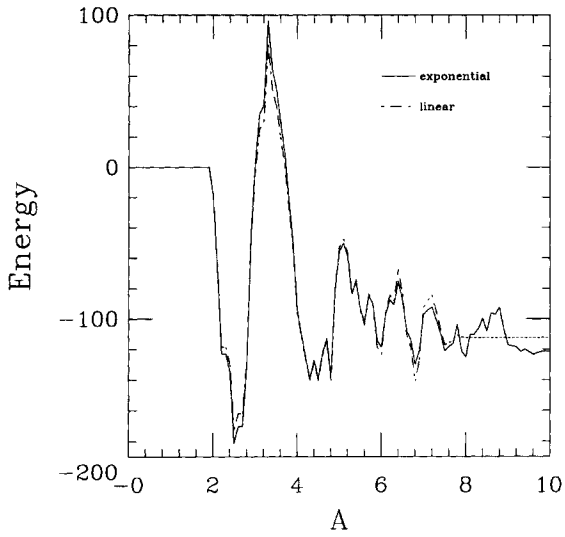

Fig. 2. Plot of the integrated binding energy as a function of distance:

$$E(r) = \int_0^r \sum_{ij} V_{ij}(r')dr'$$

where $V_{ij}(r)$ is the potential between the i^{th} atom in the lysozyme and the j^{th} atom in the antibody that are separated by a distance r. The integral is performed for both the exponential and linear dielectric functions. Note that the two begin to diverge only after 6 Å, and then by only 5%.

Comparison of Distance-Dependent Dielectric Functions

We have tried to understand why the exponentially increasing dielectric function performs better than the commonly used linear dielectric function. In Figure 2 we plot the integral of the total binding energy as a function of the distance between atomic pairs that have been included in the calculation. Both the linear and exponential dielectric functions have been cut off with a switching function. We note that in this case the two potentials lead to similar energies, and as expected, the small divergence between the two occurs mainly in the 6 to 10 Å range. Similarly, we compared the binding energies of the top 10 candidates of all three antibodies, after OPLS minimization, using both the linear and exponential dielectric functions and in all cases found that the energies of a given configuration calculated with the two dielectric functions did not differ by more than 10%. Since the energies of a given configuration measured by the two dielectric functions are not significantly different, the difference in performance

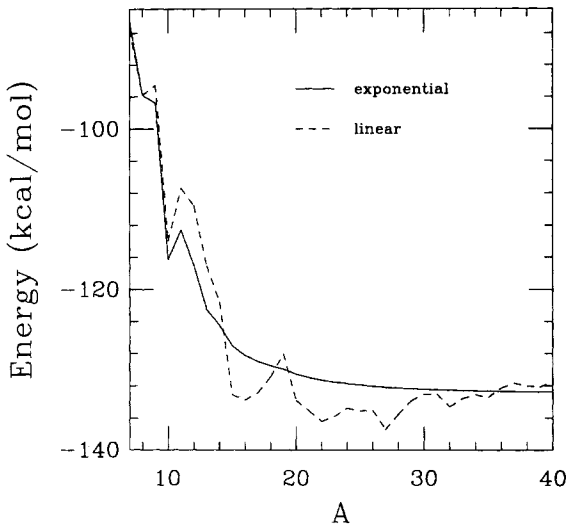

Fig. 3. Plot of the accumulated energy integrated with respect to distance versus distance (calculated without using a switching function to cut off the dielectric function):

$$E(r) = \int_0^r \sum_{all\ pairs} \delta E_{aa\ pair}(r')dr',$$

$\delta E_{aa\ pair}$ is given in equation (12). The energy is computed between atoms grouped by amino acid units, and the distance represents the length in Angstroms between the centers of mass of the amino acids.

must be due to differences in the algorithm's sampling of the phase space when using the two functions.

To get a better idea of this effect, we plot in Figure 3 the contributions to the total energy grouped by amino acid units, as a function of their separation, when no switching function is used:

$$\delta E_{aa\ pair}(r) = \sum_{\substack{aa1(i) \\ aa2(j)}} V(R_{aa1} - R_{aa2})\ \delta(r - |R_{aa1} - R_{aa2}|) \quad (8)$$

where *aa1(i)*, *aa2(j)* are atoms in amino acids 1 and 2 and $R_{aa1,2}$ are the centers of mass of the two amino acids. As shown in the figure, the linear and exponential dielectric functions lead to virtually indistinguishable total energies. The main difference appears to be that the exponential form tends to smooth out variations in energy versus distance. This has the effect of lowering barriers which get in the way of energy minimization using the Monte Carlo moves.

We have verified this fact by studying the basins of attraction under the two functions. We started the lysozyme in the native configuration of the D1.3-HEL complex. We then moved the lysozyme from this configuration by pulling it away from the surface, in steps of 1 Å. After 100 steps of relaxation at room temperature, the simulations conducted with the exponential dielectric function settled back into the native basin from as far away as 3 Å, while those conducted with the linear dielectric function got stuck in shallow local minima when started 1 Å away. Although after a lengthy relaxation the simulations using a linear dielectric function will eventually settle into the native basin, those that use the exponential dielectric function are significantly more efficient; in the runs above, when carried out for more than 100 steps, the exponential dielectric function requires two to three times fewer moves to settle into the native basin.

Thus, we believe that the improved performance we find using the exponential dielectric function results from its smoothing effect on the contributions to the interaction energy from atoms separated by distances greater than 6 Å, and the subsequent lowering of barriers. Because the linear dielectric function leads to noisy contributions due to the accumulated energy of interaction at long distances, we found that one had to switch it off between 7 and 8 Å to obtain the best results. On the other hand the exponential dielectric function performed best when switched off between 9 and 10 Å.

We found that the 3 Å decay length for the exponentially increasing dielectric used in our simulations was fairly critical. When extended to 4 Å, the simulations were no longer successful, and similarly for decay lengths below 2 Å. Note that the exponential form used here does not represent a Debye-Huckel screened potential at a particular value of ionic strength. Instead our dielectric function provides a very simplified model of effects due to electronic polarizability and dipolar relaxation within the protein and in the solvent.

DISCUSSION

Our algorithm has three features which differ from those of other authors addressing the problem: (1) the use of a hinge point near the center of the antigen-binding site that is used to determine the location of the initial conformations; (2) the use of a binding energy test based on coarse-grained pair potentials in the selection process for candidate docking conformations; and (3) the introduction of an exponentially increasing dielectric function in rigid body protein-protein interactions to simulate screening.

As discussed above, the use of a hinge point allows us to generate in an unbiased way initial conformations that include conformations close to the native. This has the advantage of allowing us to cut down the length of searches: we relax configurations for only 50 steps in the first three stages. We are also assured that our method will exclude all solutions that do not have a part of the lysozyme lying within the hinge point.

We believe part (2) provides an improvement over algorithms that select conformations by maximizing surface contact: our potentials tend to eliminate candidates which may be favorable in terms of contact area but unfavorable energetically. However, the algorithms of Shoichet and Kuntz[5] and Cherfils et al.[4] find solutions which are both favorable in terms of contact area and energetically. Why are these solutions not found by our algorithm? We conjecture that these solutions may be unfavorable in terms of a third criterion: size of basin of attraction. We assume that our strategy eliminates candidates with a narrow basin of attraction at an early stage in favor of ones with large basins of attraction. This is due to the fact that we allow our initial configurations to relax using a coarse-grained energy criterion for a limited search (fifty time steps) in the first stage, thus making it unlikely for the lysozyme to settle in a narrow basin. On the other hand, a search based on a criterion of maximizing surface contact may find a basin which is narrow on energy grounds but may appear much broader based on a contact criterion. Thus, we are lead to believe that the basins of attraction that we find when using our coarse-grained pair potentials more accurately reflect the properties of the real system than do the basins of attraction generated on geometric grounds.

Our use of an exponentially increasing distance-dependent dielectric function, part (3), also appears to be an improvement over the standard linear dielectric, at least in the case of these three complexes, in its ability to find a native-like configuration more consistently. As discussed above, this improvement seems not to be due to its ability of resolving the binding energy of the native relative to that of the false solutions, but rather due to the lowering of energy barriers between local minima. Thus, as we have shown, the lysozyme is able to sample phase space more efficiently than would be possible with the standard linear dielectric. At this stage we cannot say whether this feature is an artifact which is computationally useful, or a reflection of the topography of the real docking landscape found in the physical system.

We note that the use of a phenomenological distance-dependent dielectric function constitutes a substantial simplification of the real problem. We do not explicitly include water in our simulation, and we treat the protein as a rigid body. Thus, one should be wary of generalizing the use of distance-dependent dielectric functions of this type to other problems. However, our work provides an additional piece of evidence that suggests that these potentials are an effective representation of the antibody-lysozyme interaction in the case of our three complexes. In this context it is interesting to note that our distance-dependent dielectric function is similar to one used by Mehler and Solmajer to describe pK shifts within a protein.[14] Their function is:

$$\epsilon(r) = A + B/[1 + k \exp(-\lambda B r)]. \quad (9)$$

For small values of r ($r < 10$ Å) their dielectric also has an exponentially increasing behavior, with a length constant very close to our value of 3 Å. The success of their dielectric function in describing pK shifts and the ability to obtain similar results as simulations that include water explicitly,[13] suggests that our function might have wider applicability than that examined in this paper. However, we have not as yet attempted to apply our dielectric function to other problems.

Finally, the fact that the success of the algorithm is very sensitive to the treatment of long range (6 to 10 Å) interactions leads us to conclude that steric "fit", while probably a necessary condition for high binding affinity, is not a sufficient criterion for selectivity. Indeed, even the identity of the amino acids forming the surface epitope may not provide a complete selectivity criterion and mutation of residues buried in the antigen or antibody could lead to changes in binding specificity.

SUMMARY

We have constructed a docking algorithm that successfully finds the native conformation (within 0.3 Å rms) for three different antibodies to lysozyme and consistently finds it to be of substantially lower energy (20%) than any other docking solution generated. This is in contrast to the results of other docking algorithms published to date, which all found non-native conformations that were energetically comparable to, or even more tightly bound than solutions near the native one. We attribute the success of our algorithm to the fact that: (1) it generates starting conformations that in an unbiased way allow for efficient relaxation; (2) rather than searching for solutions on the geometric grounds of maximization of contact area between antibody and antigen, it screens solutions by filtering them through an energy selection criterion based on three different sets of coarse grained pair potentials; and (3) it uses an exponentially increasing dielectric function that, in the three cases examined, allows our algorithm to search phase space and locate the native conformation more consistently than the standard linear dielectric function.

ACKNOWLEDGMENTS

We are grateful to T.N. Bhatt and R. Poljak for use of the coordinates for the D1.3-lysozyme complex. S. Doniach wishes to thank P. Alzari for stimulating conversations which led to the investigations reported here. We thank the Dean of Research, Stanford University, and the Office of Naval Research for financial support. The use of equipment donated by Digital Equipment Corporation is gratefully acknowledged.

REFERENCES

1. Amit, A.G., Mariuzza, R.A., Phillips, S.E.V., Poljak, R.J. Three-dimensional structure of an antibody-antigen complex at 2.8 Å resolution. Science 233:747–753, 1986.
2. Sheriff, S., Silverton, E.W., Padlan, E.A., Cohen, G.H., Smith-Gill, S.J., Finzel, B.C., Davies, D.R. Three-dimensional structure of an antibody-antigen complex. Proc. Natl. Acad. Sci. U.S.A. 84:8075–8079, 1987.
3. Padlan, E.A., Silverton, E.W., Sheriff, S., Cohen, G.H., Smith-Gill, S.J., Davies, D.R. Structure of an antibody-antigen complex: Crystal structure of the HyHEL10 FAB-lysozyme complex. Proc Natl. Acad. Sci. U.S.A. 86:5938–5942, 1989.
4. Cherfils, J., Duquerroy, S., Janin, J. Protein-protein recognition analyzed by docking simulation. Proteins 11:271–280, 1991.
5. Shoichet, B.K., Kuntz, I.D. Protein docking complimentarity. J. Mol. Biol. 221:327–346, 1991.
6. Jiang, F., Kim, S. "Soft Docking": Matching of molecular surface cubes. J. Mol. Biol. 219:79–102, 1991.
7. Kasinos, N., Lilley, G.A., Subbarao, N., Haneef, I. A robust and efficient automated docking algorithm for molecular recognition. Protein Engineering 5:69–75, 1992.
8. Jorgensen, W.L., Tirado-Rives, J. The OPLS potential functions for proteins. Energy minimizations for crystals of cyclic peptides and crambin. J. Am. Chem. Soc. 110: 1657–1666, 1988.
9. Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., Teller, E. Equation of state calculation by fast computing machines. J. Chem. Phys. 21:1087–1092, 1953.
10. Wilson, C., Doniach, S. A computer model to dynamically simulate protein folding: Studies with crambin. Proteins 6:193–209, 1989.
11. Brooks, B.R., Bruccoleri, R.E., Olafson, B.D., States, J.S., Swaminathan, S., Karplus, M. CHARMM. A program for macromolecular energy minimization and dynamics calculations. J. Comp. Chem. 2:187–217, 1983.
12. Novotny, J., Bruccoleri, R.E., Saul, F.A. On the attribution of binding energy in the antigen-antibody complexes MCPC603, D1.3, and HyHEL-5. Biochemistry 28:4735–4749, 1989.
13. Solmajer, T., Mehler, E.L. Electrostatic screening in molecular dynamics simulations. Protein Engineering 4:911–917, 1991.
14. Mehler, E.L., Solmajer, T. Electrostatic effects in proteins: comparison of dielectric and charge models. Protein Engineering 4:903–910, 1991.
15. Wendoloski, J.J., Mathews, J.B. Molecular dynamics effects on protein electrostatics. Proteins 5:313–321, 1989.
16. Lee, F.S., Chu, Z.T., Bolger, M.B., Warshel, A. Calculations of antibody-antigen interactions: Microscopic and semi-microscopic evaluation of the free energies of binding of phosphorylcholine analogs to McPC603. Protein Engineering 5:215–228, 1992.
17. Davies, D.R., Padlan, E.A., Sheriff, S. Antibody-antigen complexes. J. Biol. Chem. 263:10541–10544, 1988.
18. Davies, D.R., Padlan, E.A., Sheriff, S. Antibody-antigen complexes Annu. Rev. Biochem. 59:439–473, 1990.
19. Sippl, M.J. Calculation of conformational ensembles with potentials of mean force. J. Mol. Biol. 214:859–83, 1990.
20. King, G., Lee, F.S., Warshel, A. Microscopic simulations of macroscopic dielectric constants of solvated proteins. J. Chem. Phys. 95:4366–4377, 1991.
21. Simonson, T., Perahia, D., Bricogne, G. Intramolecular dielectric screening in proteins. J. Mol. Biol. 218:859–886, 1991.
22. Warshel, A., Russell, S. Calculations of electrostatic interactions in biological systems and in solutions. Quart. Rev. Biophys. 17:283–422, 1984.
23. Tramontano, A., Chothia, C., Lesk, A. Framework residue 71 ia a major determinant of the position and conformation of the second hypervariable region in the V_H domains of immunoglobulins. J. Mol. Biol. 215:175–182, 1990.
24. Harvey, S. Treatment of electrostatic effects in macromolecular modeling. Proteins 5:78–92, 1989.
25. Tripos Associates, Inc. a subsidiary of Evans and Sutherland. 1699 S. Hanley Rd. Suite 303, St. Louis, Missouri 63144.
26. Gilson, M., Honig, B. The inclusion of electrostatic hydration energies in molecular mechanics calculations. Journal of Computer-Aided Molecular Design 5:5–20, 1991.