Protein Folding and Peptide Docking : A Molecular Modeling and Global Optimization Approach

J. L. Klepeis, M. G. Ierapetritou and C. A. Floudas[1]

Department of Chemical Engineering, Princeton University, Princeton, NJ 08544-5263, USA

Abstract–Global optimization approaches are proposed for addressing both the protein folding and peptide docking problems. In the protein folding problem, the ultimate goal involves predicting the native protein conformation. A common approach, based on the thermodynamic hypothesis, assumes that this conformation corresponds to the structure exhibiting the global minimum free energy. However, molecular modeling of these systems results in highly nonconvex energy hypersurfaces. In order to locate the global minimum energy structure on this surface, a powerful global optimization method, αBB, is applied. The approach is shown to be extremely effective in locating global minimum energy structures of solvated oligopeptides. A challenging problem related to protein folding is peptide docking. In addressing the peptide docking problem, the task is not only to predict a macromolecular-ligand structure but to also rank the binding affinities of a set of potential ligands. Many methods have used qualitative descriptions of the macromolecular–ligand complexes in order to avoid the need to perform a global search on the nonconvex energy hypersurface. In this work, a novel decomposition based approach that incorporates quantitative, atomistic–level energy modeling and global optimization is proposed. This approach employs the αBB global optimization method and is applied to the prediction of peptide docking to the MHC HLA-DR1 protein.

INTRODUCTION

Recent advances in genetic engineering have heightened the interest in research related to predicting native protein folding and docking conformations. The ability to predict these structures is of great theoretical interest, especially in the fields of biophysics and biochemistry. Moreover, the applications of such knowledge also promise to be exciting. For example, the ability to predict these structures would greatly increase our understanding of hereditary and infectious diseases and aid in the interpretation of genome data. In addition to these advances, the ability to understand peptide docking would likely revolutionize the process of de novo drug design.

The use of computational techniques and simulations in addressing the protein folding and peptide docking problems became possible through the introduction of qualitative and quantitative methods for modeling these systems. The development of realistic energy models also established a link to the field of global optimization, where, based on Anfinsen's hypothesis, the quantity to be optimized is the free energy of the system. However, because of the computational complexity associated with these problems, only the most efficient global optimization strategies will be successful.

This work addresses the protein folding and peptide docking problems, including the effects of solvation, through the use of a deterministic global optimization algorithm. This branch–and–bound based global optimization algorithm, known as αBB, is applicable to a large class of nonlinear optimization problems that have twice–differentiable functions [1, 2, 3, 4, 5, 6].

MODELING

Potential Models

Many models have been developed using a classical description of molecules in terms of atomic bonds and effective interactions. In general, these models, also known as force fields, are expressed as summations of empirically derived potential functions. Thermodynamic data from small molecules and spectroscopic data are used to derive the parameters describing the relative strengths of particular interatomic interactions. In most cases, these force fields are atom centered potentials from which the total molecular energy is computed as a sum over all pairwise interactions.

In this work, the ECEPP/3 (Empirical Conformational Energy Program for Peptides) potential model is utilized [26]. In this force field, it is assumed that the covalent bond lengths and bond angles are fixed at their equilibrium values. It has been observed that variations in bond lengths and bond angles depend

[1]Author to whom all correspondence should be addressed.

mostly on short range interactions; that is, those between the side chain and backbone of the same residue. Under this assumption, all residues of the same type have essentially the same geometry in various proteins. Therefore, a chain of any sequence can be generated using the fixed geometry specific to each type of amino acid residue in the sequence.

Based on these approximations, the conformation is only a function of the dihedral angles. That is, ECEPP/3 accounts for energy interaction terms that can be expressed solely in terms of the dihedral angles. The total conformational energy is calculated as the sum of the nonbonded (NB), hydrogen bonded (HB), electrostatic (ES) and torsional (TOR) contributions, as given by the following expression:

$$\begin{aligned} E &= \sum_{i,j \in NB} \epsilon_{ij} \left[\left(\frac{r^o_{ij}}{r_{ij}} \right)^{12} - \left(\frac{r^o_{ij}}{r_{ij}} \right)^{6} \right] \\ &+ \sum_{i,j \in HB} \epsilon_{ij} \left[5 \left(\frac{r^o_{ij}}{r_{ij}} \right)^{12} - 6 \left(\frac{r^o_{ij}}{r_{ij}} \right)^{10} \right] \\ &+ \sum_{i,j \in ES} \frac{q_i q_j}{D r_{ij}} + \sum_{k \in TOR} \frac{A_k}{2} (1 \pm \cos n_k \theta_k) \end{aligned} \tag{1}$$

A full discussion of these terms, and the appropriate parameters, can be found elsewhere [26].

Solvation Models

A complete description of the total energy of a polypeptide must also include its interactions with the solvent. Explicit methods can be used by actually surrounding the polypeptide with solvent molecules and calculating solvent–peptide and solvent–solvent interactions using potentials similar to those previously described. Although these methods are conceptually simple, explicit inclusion of solvent molecules greatly increases the computational time needed to simulate the polypeptide system. Therefore, most simulations of this type are limited to restricted conformational searches.

Simpler methods for estimating solvent free energies have been developed using continuum models, which use a simplified representation of the solvent environment by neglecting the molecular nature of the water molecules. For this study, solvation contributions are included implicitly using empirical correlations with both surface area and volume. The main assumption of these models is that, for each functional group of the peptide, a hydration free energy can be calculated from an averaged free energy of interaction of the group with a layer of solvent known as the hydration shell. In addition, the total free energy of hydration is expressed as a sum of the free energies of hydration for each of the functional groups of the peptide; that is, an additive relationship is assumed. These solvation contributions can be described by the following general equation:

$$E_{HYD} = \sum_{i=1}^{N} (S_i)(\sigma_i) \tag{2}$$

In Equation (2), an additive relationship for N individual functional groups is assumed. (S_i) represents either the solvent–accessible surface area, (A_i), or the solvent–accessible volume of hydration layer, (VHS_i), for the functional group, and (σ_i) are empirically derived free energy density parameters.

In this work, solvent–accessible surface areas are calculated using the MSEED [31] program, which employs algorithms developed by Connolly [11]. MSEED eliminates many unnecessary computations by considering only those convex faces that are on the accessible surface. Rigorous implementation of Connolly's method requires the calculation of interior surface areas, which are ultimately found to be zero. A full description of the MSEED algorithm is given elsewhere [31].

Once the solvent–accessible surface areas have been calculated, these values must be multiplied by the appropriate (σ_i) parameters, as shown in Equation (2). There are a number of models available, including JRF, OONS, and SRFOPT, which provide estimates for these parameters based on interactions between water and the functional groups of peptides. It has been shown that minimum energy solvated conformations predicted by the JRF model provide the best correspondence to native (crystallographic) structures when compared with other models [34]. These parameters were derived from NMR studies of low energy solvated configurations of 13 tetrapeptides. Because it was developed from minimum energy conformations of peptides, the JRF parameter set has been shown to produce undesirable perturbations during local minimizations if the solvation energy contributions are added at every iteration. Therefore, the surface–accessible solvation energies are only included at local minimum conformations.

For volume based hydration energies, the RRIGS (Reduced Radius Independent Gaussian Sphere) approximation is used to efficiently calculate the exposed volume of the hydration shell [9]. This method artificially reduces the van der Waals radii of all atoms other than atom i when calculating (VHS_i). These reductions effectively decrease the contribution of the double overlap terms, leading to a cancellation of the error which results from neglecting the triple and higher overlap terms. In addition, the characteristic density of being inside the overlap volume of two intersecting spheres is not represented as a step function but as a Gaussian function, which provides continuous derivatives of the hydration potential. Therefore, the solvation energy contributions can easily be added at every step of local minimizations.

Peptide Docking

The complexity associated with the prediction of peptide docking conformations complicates the task of modeling these interactions. First, the binding site of the target globular protein must be correctly characterized. This task usually requires experimental structure determination of the binding site. Such information is invaluable because it can be used to approximate rigid binding sites, which greatly reduces the translational space that must be explored in a conformational search. The second task is to select potential ligands, dock these ligands to the active site, and assign a "score" to each complex. These "scores" may then be used to rank binding affinities for a set of ligands.

The fundamental feature of the peptide docking problem is the development of accurate scoring functions. Many methods have relied on qualitative modeling of the peptide docking interactions. In the case of a rigid binding approximation, the use of shape complementarity has had some limited success [20]. These algorithms model the ligand and macromolecule according to their surface topology, and attempt to identify which complexes exhibit the best "fit". Here scoring functions are based on the complementarity of the molecules, which, in most cases, is related to their solvent accessible surface areas [11]. The obvious strength of these methods is that they can be made computationally efficient and used to screen large databases of potential ligands. However, studies comparing computational results to experimentally derived, native complexes indicate that many non-native low energy structures are identified.

On the other hand, it has been demonstrated that exact modeling of binding free energies provides results in nearly exact quantitative agreement with experimental results [15]. In contrast to the rigid description of docking, these methods allow for flexibility of both the ligand and receptor molecules. However, for general peptide docking problems, thermodynamic integration and free energy perturbation methods are computationally infeasible with current computing power.

A more universal approach, applicable to flexible ligands, is to base energy calculations on potential energy models. In this study, a full quantitative model is used by employing the the ECEPP/3 force field. The proposed binding energy function also accounts for solvation energy, which is calculated using the MSEED solvent–accessible surface area model.

PROBLEM FORMULATION

Protein Folding

For protein folding, the energy minimization problem can be formulated as a nonconvex nonlinear global optimization problem in which the energy, E_{Fold}, must be globally minimized with respect to the dihedral angles of the protein. The energy, E_{Fold}, represents the total potential energy function, E^{Unsol}, plus the free energy of solvation, E^{Sol}. For accessible volume shell hydration (RRIGS) this is the exact formulation because both energetic and gradient contributions can be added at each step of the minimization. However, in the case of surface–accessible hydration (MSEED and JRF parameters), the potential energy function is minimized before adding the hydration energy contributions. In other words, gradient contributions from solvation are not considered. This approach is represented by the following equation:

$$E_{Total} = E_{Min}^{Unsol} + E^{Sol} \tag{3}$$

Peptide Docking

The peptide docking methodology is more complex when compared to the protein folding formulation. To begin with, the dimensionality is inherently larger due to the need to consider translational and rotational degrees of freedom. For this study, the problem is simplified somewhat by assuming rigid binding sites, although the ligands are considered to be fully flexible. In addition, a novel decomposition scheme is proposed for modeling complex formation between HLA-DR1 and binding peptides. The key ideas of the decomposition approach are: (i) to consider the binding at each pocket separately, (ii) to study the binding of each naturally occurring amino acid to each pocket, and (iii) to create a rank ordered list of the bound amino acids for each pocket based on an energetic criterion that reflects binding affinity. This approach is justified by experimental observations which conclude that the binding specificity of the HLA-DR1 molecule is mainly determined by the binding characteristics of its five pockets [32].

The details of the peptide docking approach are as follows:

(1) Using experimental information [32], the HLA-DR1 pocket is characterized by a set of fixed atomic coordinates which describe the relevant residues in a given pocket.

(2) For each naturally occurring amino acid, a mathematical model is formulated that represents all the energetic atom-to-atom interactions. These interactions are classified as (i) inter-interactions between the atoms that define the HLA-DR1 pocket and the atoms of amino acid, and (ii) intra-interactions between the atoms within the amino acid. Potential energy contributions are modeled by the ECEPP/3 force field, and solvation energy is calculated using the solvent accessible surface area model (MSEED and JRF parameters).

(3) Each subproblem results in a global optimization problem in which the total energy, E_{Dock}, must be globally minimized with respect to the dihedral angles of the ligand, and the translational and rotational degrees of freedom. The inclusion of solvation is analogous to the protein folding formulation for the surface accessible solvation model. The formulation also includes additional constraints for the placement of the ligand backbone to ensure that the residue fits within the structure of an overall binding peptide.

(4) Once the global minima are identified, an energetic-based criterion is used to compare the binding affinities of each naturally occurring amino acid in a given pocket. This measure, which is denoted as ΔE_{Dock}, corresponds to the difference of the global minimum energy for the bound complex, E_{Bound}, and the global minimum energy of the unbound amino acid, $E_{Unbound}$:

$$\Delta E_{Dock} \ = \ E_{Bound} - E_{Unbound} \qquad (4)$$

This criterion quantifies the tendency of an amino acid to bind with the pocket of the HLA-DR1 molecule so that the complex that exhibits the minimum value of ΔE_{Dock} corresponds to the amino acid with the highest binding affinity to the HLA-DR1 pocket.

GLOBAL OPTIMIZATION

The problems associated with both protein folding and peptide docking can be formulated as global optimization problems in which global minimum energy structures must be identified. A large number of techniques have been developed to search the nonconvex conformational space. Many methods employ stochastic search procedures, while others rely on simplifications of the potential model and/or mathematical transformations. The major limitation is that there is no guarantee for convergence to the global minimum energy structure. A number of recent reviews have focused on global optimization issues for these systems [13, 28, 30, 33].

In this work, the global optimization approach αBB has been extended to identifying global minimum energy conformations of solvated peptides and peptide docking complexes. The development of this branch and bound method was motivated by the need for an algorithm that could guarantee convergence to the global minimum of nonlinear optimization problems with twice–differentiable functions [12]. The application of this algorithm to the minimization of potential energy functions was first introduced for microclusters [21, 22], and small acyclic molecules [23, 24]. The αBB approach has also been extended to constrained optimization problems [2, 3, 4, 6]. In more recent work, the algorithm has been shown to be successful for isolated peptide systems [7, 25].

Minimization of Energy using αBB

The αBB global optimization algorithm effectively brackets the global minimum solution by developing converging lower and upper bounds. These bounds are refined by iteratively partitioning the initial domain. Upper bounds on the global minimum are obtained by local minimizations of the original energy function, E. Lower bounds belong to the set of solutions of the convex lower bounding functions, which are constructed by augmenting E with the addition of separable quadratic terms. The lower bounding function (L) of the energy hypersurface can be expressed in the following manner:

$$L \ = \ E + \sum_{i=1}^{N_\phi} \alpha_{\phi,i} \left(\phi_i^L - \phi_i\right) \left(\phi_i^U - \phi_i\right) \qquad (5)$$

The ϕ_i^L and ϕ_i^U values correspond to lower and upper bounds, for the given domain, on each of the ϕ_i variables. In the protein folding problem the ϕ_i correspond to the independent torsion angles (dihedral angles) of the peptide, whereas in peptide docking this variable list must also include the relative translation vector and Euler angles. The $\alpha_{\phi,i}$ represent nonnegative parameters which must be greater or equal to the negative one–half of the minimum eigenvalue of the Hessian of E over the defined domain. These parameters can be estimated by the solution of an optimization problem or by using the concept of the measure of a matrix [1, 2, 5, 23]. The overall effect of these terms is to overpower the nonconvexities of the original nonconvex terms by adding the value of 2α to the eigenvalues of the Hessian of E. The convex lower bounding functions, L, possess the following important properties which guarantee global convergence [24] : (i) L is a valid underestimator of E; (ii) L matches E at all corner points in the current box constraints; (iii) L is convex in the current box constraints; (iv) the maximum separation between L and E is bounded and proportional to α and to the square of the diagonal of the current box constraints, which ensures that ϵ_f feasibility and ϵ_c convergence tolerances can be reached; (v) the underestimators constructed over supersets of the current set are always less tight than the underestimator constructed over the current box constraints for every point within the current box constraints.

Once solutions for the upper and lower bounding problems have been established, the next step is to modify these problems for the next iteration. This is accomplished by successively partitioning the initial domain into smaller subdomains. The default partitioning strategy involves successive subdivision of the original hyper–rectangle by halving on the midpoint of the longest side (bisection).

COMPUTATIONAL STUDIES

The implementation of these approaches within the αBB framework has resulted in two global optimization packages : GLO-FOLD [19] for protein folding and GLO-DOCK [29] for peptide docking. These packages are flexible and provide a variety of options through the interfacing of several local optimization and energy modeling programs. In the case of protein folding, a comparative study of the two aforementioned solvation approaches has been performed for the set of 20 naturally occurring residues and a number of oligopeptides, including Ac–Ala$_4$–Pro–NHMe, met–enkephalin, leu–enkephalin, and decaglycine [16, 17]. For the peptide docking problem, computational and experimental results are available for peptide binding in pocket 1 of HLA-DR1 [8]. The next subsection presents results for the met–enkephalin example, and the second subsection summarizes the results obtained for the binding in pocket 1 of HLA-DR1.

Protein Folding : Met–Enkephalin

Met–enkephalin (H–Tyr–Gly–Gly–Phe–Met–OH) is an endogenous opioid pentapeptide found in the human brain, pituitary, and peripheral tissues. Its biological function involves a large variety of physiological processes, most notably the endogenous response to pain. The peptide consists of 24 dihedral angles and a total of 75 atoms, and has played the role of a benchmark molecular conformation problem. The energy hypersurface is extremely complex with the number of local minima estimated on the order of 10^{11}. The unsolvated global minimum energy conformation has been shown to exhibit a type II' β–bend along the N–C' peptidic bond of Gly^3 and Phe^4 [7].

Experimental results have indicated that met–enkephalin in aqueous solution does not possess an unique structure [14]. In general, the experimentally determined aqueous conformations were found to exhibit characteristics of extended random–coil polypeptides with no discernible secondary structure. When considering the effects of hydration, the competition for backbone hydrogen bonding (with water), which contributes to the bending of the unsolvated conformation, should result in a more extended structure. These qualitative arguments have been confirmed by the analysis of hydrated met–enkephalin using the MSEED model. For the global minimum energy structure, residues near the N–terminus are almost fully extended, although there is slight bending near the C–terminus. This bending is stabilized by the formation of 2.10 $\AA$ hydrogen bond between the CO of the second glycine residue and the NH proton of the methionine residue. In addition, the structure displays a large 17.00 $\AA$ separation between the centroids of the Phe and Tyr aromatic rings.. A plot of the conformation corresponding to the global minimum energy of -283.76 kcal/mol is given in Figure 1. Locating this solution required 1033 iterations and 5,082 seconds (HP-C110).

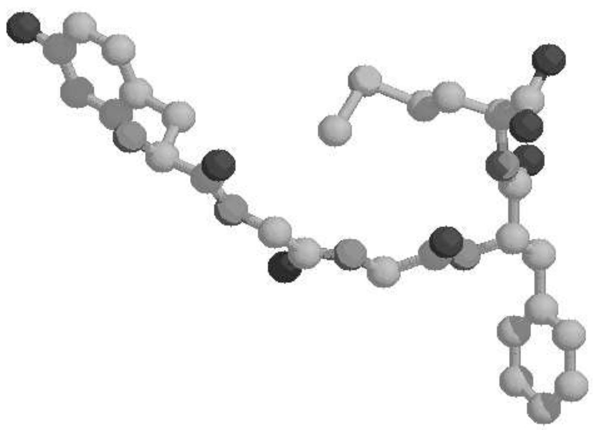

Figure 1: Global minimum energy conformation using MSEED model for hydration.

The RRIGS method also predicts a more extended structure than the global minimum structure reported for the unsolvated case [7]. In fact, although a slight bend occurs near the N–terminus, the structure possesses no hydrogen bonds ($<$ 2.3 $\AA$). In addition, unlike the MSEED structure, there exists close proximity of the Tyr and Phe aromatic rings. The centroids of these rings are separated by 4.16 $\AA$, which is slightly closer than the preferential aromatic–aromatic interaction distance of 4.5 to 7 $\AA$ [10]. Furthermore, the aromatic rings are essentially in a parallel, as opposed to the more common orthogonal, orientation. This suggests an attempt to substantially reduce the hydrophobic exposure of the aromatic side chains. The global minimum conformation, with an energy of -50.01 kcal/mol, was located in 1058 iterations and 8,695 seconds (HP-C110). A plot of this structure is given in Figure 2.

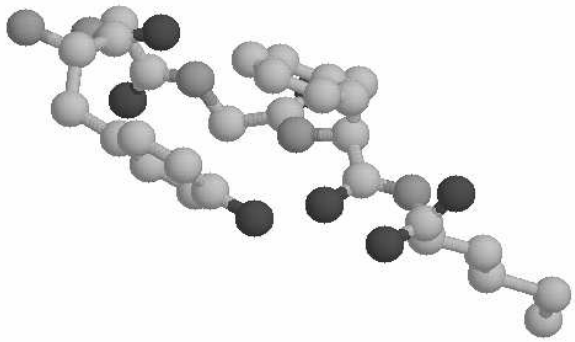

Figure 2: Global minimum energy conformation using RRIGS model for hydration.

It is also interesting to compare energy evaluations at corresponding global minimum solutions. This information is given in Table 1. It is apparent that the MSEED model predicts large stabilizing hydration free energies. In addition, these contributions tend to dominate the prediction of the global minimum structure. Specifically, energy evaluations at the RRIGS and unsolvated solutions indicate a substantial increase in overall energy, which can be directly correlated to the increase in hydration free energy. In contrast, this correlation does not hold for the RRIGS model. In fact, the RRIGS model, like the MSEED model, predicts a more stabilizing hydration free energy at the MSEED solution. However, nonbonded interactions are less favorable at this solution, resulting in an overall energy increase. In addition, although the solvation free energy becomes less stabilizing at the unsolvated solution, an increase in the number of favorable nonbonded interactions causes the overall energy to be near the global minimum solution.

It should also be noted that the αBB algorithm is able to identify low energy conformers, along with the global minimum energy conformation. Table 2 lists five local minimum energy conformations within 0.5 kcal/mol of the RRIGS global minimum energy. The structures are related to the global minimum energy

Table 1: Comparison of energies for met–enkephalin. Functions evaluations, using both the MSEED and RRIGS models, are performed at the RRIGS, MSEED and UNSOL (unsolvated [7]) global solutions. The total energy, E_{TOT}, is provided along with the contributions from hydration, E_{HYD}, nonbonded interactions, E_{NB}, electrostatic interactions, E_{ES}, and torsion, E_{TOR}.

Energy Term	MSEED at Global of MSEED	MSEED at Global of RRIGS	MSEED at Global of UNSOL	RRIGS at Global of RRIGS	RRIGS at Global of MSEED	RRIGS at Global of UNSOL
$\mathbf{E}_{TOT}$	-283.77	-139.35	-170.88	-50.01	-41.63	-47.49
$\mathbf{E}_{HYD}$	-288.83	-130.75	-159.17	-41.41	-46.69	-35.78
$\mathbf{E}_{NB}$	-19.13	-31.47	-35.26	-31.47	-19.13	-35.26
$\mathbf{E}_{ES}$	23.29	21.84	21.46	21.84	23.29	21.46
$\mathbf{E}_{TOR}$	0.90	1.03	2.09	1.03	0.90	2.09

conformation as evidenced by their similar conformational codes [35]. Such information has important ramifications for more detailed free energy calculations [18].

Table 2: Low energy conformers (within 0.5 kcal of global minimum energy) for RRIGS model. Total energies and conformational codes [35] are given.

Conformer	$\mathbf{E}_{TOT}$	**Code**
1	-49.97	BC*G*AG
2	-49.89	BC*H*EG
3	-49.67	BC*H*EB
4	-49.61	BC*H*EA
5	-49.57	BC*GEF

Peptide Docking : HLA-DR1 Pocket 1

Crystallographic studies have shown that HLA-DR1 binding is accommodated by five polymorphic pockets on the surface of the HLA-DR1 molecule. Of these, pocket 1 is the largest and deepest pocket, with an estimated contact area of 200 $\mathring{A}^2$. It has also been postulated that the binding in this pocket acts as an "anchor" for the overall binding peptide. In modeling the rigid structure of this binding site, the crystallographic structure of an influenza virus bound to HLA-DR1 was used [32]. Specifically, all residues in the pocket within a distance of 5 $\mathring{A}$ from the experimental binding residue were considered in the atomistic energy modeling. This pocket definition consists mainly of hydrophobic residues, which account for its preference to accommodate other hydrophobic residues [32].

Computational results for pocket 1 are summarized in Figure 3, in which a rank ordered list is developed based on the ΔE_{Dock} energy criterion. Experimental results, shown in Figure 4, are based on a series of competitive binding assays involving analogs of the HA peptide (306-318) and the DRB1*0101 molecule [27]. In order to simulate different binding residues, analog peptides were synthesized in which the Y(308) residue of the HA peptide (306-318) was substituted with 11 different amino acids. The relative binding affinities were derived from the reciprocal of 50% inhibitory concentration (IC50) for each analog peptide.

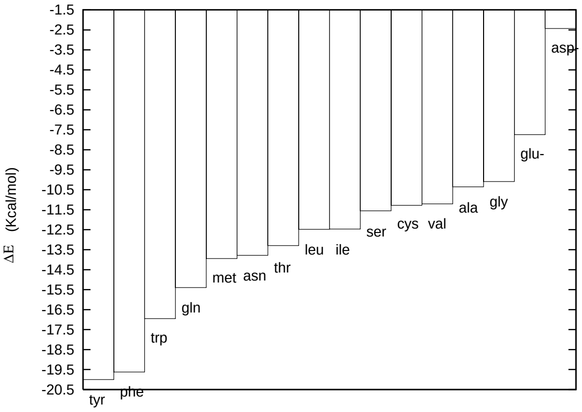

Figure 3: ΔE_{Dock} (kcal/mol) of the naturally occurring amino acids.

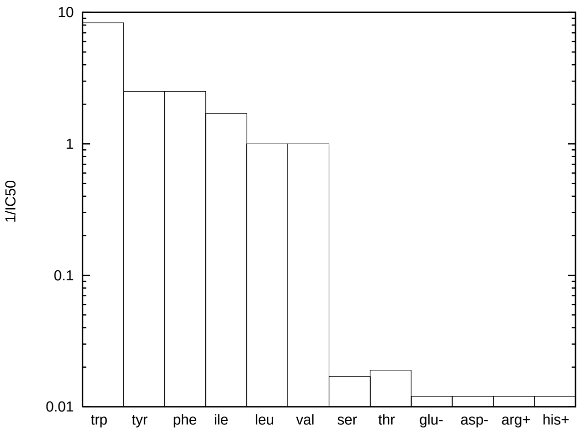

Figure 4: Experimental data for the naturally occurring amino acids.

Based on results from the experimental binding assays, three residue groupings can be identified. The first group includes the amino acids Tyr, Phe, and Trp, which exhibit the highest binding affinity to DR1. The second group includes the amino acids Ile, Leu, and Val, and can be characterized by an intermediate level of affinity to DR1. The third group consists of low level affinity amino acids. This group includes the charged residues Asp-, Glu-, Arg+, His+ and the neutral Ser and Thr residues.

The theoretical results, shown in Figure 3, are in excellent agreement with those obtained by experiment. The hydrophobic residues, Tyr, Phe and Trp occupy the top positions of the rank ordered list. This result is also qualitatively supported by this pocket's preference to accommodate hydrophobic side chains. The amino acids Leu, Ile, and Val are characterized by ΔE_{Dock} that correspond to 8th, 9th and 12th position on the ordered list, respectively. Similarly, binding assays resulted in intermediate affinities for these amino acids. The prediction of low binding affinities for charged residues is also in agreement with experimental data.

For both serine and threonine, intermediate binding affinities would be expected due to weak interactions between their hydroxyl groups and the hydrophobic pocket. Although not as apparent as experimental results, the theoretical predictions qualitatively support these observations.

Theoretical and experimental agreement was also found when comparing the structures of the bound complexes. For example, Figure 5 shows the tyrosine conformation for the HA peptide binder, Y(308), and the predicted global minimum energy conformation of tyrosine in pocket 1. Almost identical orientations, with a 1.28 $\mathring{A}$ RMS deviation, are observed.

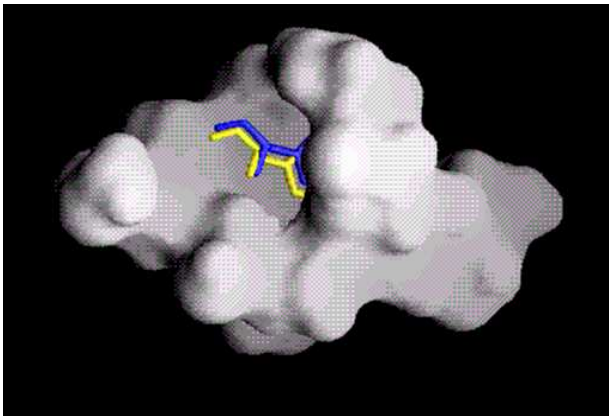

Figure 5: Comparison of tyrosine amino acid binding to Pocket 1.

CONCLUSIONS

In this study, the related problems of protein folding and peptide docking were formulated as global optimization problems. In the case of protein folding, solvation effects were included in the context of a global search, and results for met–enkephalin were presented and compared for two solvation models. For peptide docking, a novel decomposition approach, applicable to general docking problems, was presented and results for pocket 1 of HLA-DR1 were discussed. In both cases, the computational results were shown to be consistent with experimental studies. In addition, the deterministic branch and bound algorithm, αBB, was shown to be an effective framework for locating global minimum energy structures.

Acknowledgments–The authors gratefully acknowledge financial support from the National Science Foundation, the Air Force Office of Scientific Research and the National Institutes of Health.

REFERENCES

[1] C. S. Adjiman, I. P. Androulakis and C. A. Floudas, Global optimization of MINLP problems in process synthesis and design, Comput. Chem. Eng. **21**, (1997), S445-S450.

[2] C. S. Adjiman, S. Dallwig, C. A. Floudas and A. Neumaier, A global optimization method,αBB, for general twice-differentiable NLPs - I. Theoretical advances, Comput. Chem. Eng., (accepted for publication), (1997).

[3] C. S. Adjiman, I. P. Androulakis and C. A. Floudas, A global optimization method, αBB, for general twice-differentiable NLPs - II. Implementation and computational results, Comput. Chem. Eng., (accepted for publication), (1997).

[4] C. S. Adjiman, I. P. Androulakis, C. D. Maranas and C. A. Floudas, A global optimization method, αBB, for process design, Comput. Chem. Eng., **20**, (1996), S419-S424.

[5] C. S. Adjiman and C. A. Floudas, Rigorous convex underestimators for general twice-differentiable problems, J. Global Optim., **9**, (1996), 23-40.

[6] I. P. Androulakis, C. D. Maranas and C. A. Floudas, αBB: A global optimization method for general constrained nonconvex problems, J. Global Optim., **7**, (1995), 337-363.

[7] I. P. Androulakis, C. D. Maranas and C. A. Floudas, Global minimum potential energy conformations of oligopeptides, J. Global Optim., **11**, (1997), 1-34.

[8] I. P. Androulakis, N. N. Nayak, M. G. Ierapetritou, D. S. Monos and C. A. Floudas, A predictive method for the evaluation of peptide binding in pocket 1 of HLA-DRB1 via global minimization of energy interactions, Proteins, **29**, (1997), 87-102.

[9] J. D. Augspurger and H. A. Scheraga, An efficient, differentiable hydration potential for peptides and proteins, J. Comput. Chem., **17**, (1996), 1549-1558.

[10] S. K. Burley and G. A. Petsko, Aromatic-aromatic interaction: A mechanism of protein structure stabilization, Science, **229**, (1985), 23-28.

[11] M. L. Connolly, Analytical molecular surface calculation, J. Appl. Cryst., **16**, (1983), 548-558.

[12] C. A. Floudas, Deterministic global optimization in design, control, and computational chemistry, In IMA Volumes in Mathematics and its Applications, **93**, Springer-Verlag, (1997), 129-184.

[13] C. A. Floudas, J. L. Klepeis and P. M. Pardalos, Global Optimization Approaches in Protein Folding and Peptide Docking, In DIMACS Series in Discrete Mathematics and Theoretical Computer science, (to appear), (1998).

[14] W. H. Graham, E. S. Carter II and R. P. Hicks, Conformational analysis of met-enkephalin in both aqueous solution and in the presence of sodium dodecyl sulfate micelles using multidimensional NMR and molecular modeling, Biopolymers, **32**, (1992), 1755-1764.

[15] P. D. J. Grootenhuis and P. A. Kollman, Crown ether-neutral molecule interactions studied by molecular mechanics, normal mode analysis, and free energy perturbation calculations. Near quantitative agreement between theory and experimental binding free energies, J. Am. Chem. Soc., **111**, (1989), 4046-4051.

[16] J. L. Klepeis, I. P. Androulakis, M. G. Ierapetritou and C. A. Floudas, Predicting solvated peptide conformations via global minimization of energetic atom-to-atom interactions, Comput. Chem. Eng., (accepted for publication), (1997).

[17] J. L. Klepeis and C. A. Floudas, A comparative study of global minimum energy conformations of solvated peptides, (submitted), (1997).

[18] J. L. Klepeis and C. A. Floudas, Free energy calculations for peptides using a deterministic global optimization algorithm, (submitted), (1997).

[19] J. L. Klepeis, X. V. Nguyen and C. A. Floudas, GLO-FOLD : A package for global optimization using αBB in protein folding, Princeton University, Princeton, NJ, (1997).

[20] I. D. Kuntz, J. M. Blaney, S. J. Oatley, R. Landgridge and T. E. Ferrin, A geometric approach to macromolecule-ligand interactions, J. Mol. Biol., **161**, (1982), 269-288.

[21] C. D. Maranas and C. A. Floudas, A global optimization approach for Lennard-Jones microclusters, J. Chem. Phys., **97**, (1992), 7667-7677.

[22] C. D. Maranas and C. A. Floudas, Global optimization for molecular conformation problems, Ann. Oper. Res., **42**, (1993), 85-117.

[23] C. D. Maranas and C. A. Floudas, A deterministic global optimization approach for molecular structure determination, J. Chem. Phys., **100**, (1994), 1247-1261.

[24] C. D. Maranas and C. A. Floudas, Global minimum potential energy conformations of small molecules, J. Global Optim., **4**, (1994), 135-170.

[25] C. D. Maranas, I. P. Androulakis and C. A. Floudas, A deterministic global optimization approach for the protein folding problem, In DIMACS Series in Discrete Mathematics and Theoretical Computer Science, **23**, Amer. Math. Soc., (1996), 133-150.

[26] F. A. Momany, R. F. McGuire, A. W. Burgess and H. A. Scheraga, Energy parameters in polypeptides. VII. Geometric parameters, partial atomic charges, nonbonded interactions, hydrogen bond interactions, and intrinsic torsional potentials for the naturally occurring amino acids, J. Phys. Chem., **79**, (1975), 2361-2381.

[27] D. Monos, A. Soulika, E. Argyris, J. Gorga, L. Stern, V. Magafa, P. Cordopatis, I. P. Androulakis and C. A. Floudas, HLA diversity : Functional and medical implications, Proc. Intl. Histocomp. World. Conf., **12**, (1997).

[28] A. Neumaier, Molecular modeling of proteins and mathematical prediction of protein structure, SIAM Rev., **39**, (1997), 407-460.

[29] X. V. Nguyen, M. G. Ierapetritou and C. A. Floudas, GLO-DOCK : A package for global optimization using αBB in peptide docking, Princeton University, Princeton, NJ, (1997).

[30] P. M. Pardalos, D. Shalloway and G. Xue (editors), Global minimization of nonconvex energy functions : Molecular conformation and protein folding, DIMACS Series in Discrete Mathematics and Theoretical Computer Science, **23**, Amer. Math. Soc., (1996).

[31] G. Perrot, B. Cheng, K. D. Gibson, J. Vila, K. A. Palmer, A. Nayeem, B. Maigret and H. A. Scheraga, MSEED: A program for the rapid analytical determination of accessible surface areas and their derivatives, J. Comp. Chem., **13**, (1992), 1-11.

[32] L. Stern, J. Brown, T. Jardetsky, J. Gorga, R. Urban, L. Strominger and D. Wiley, Crystal structure of the human class II MHC protein HLA-DR1 complexed with an influenza virus peptide, Nature, **368**, (1994), 215-221.

[33] M. Vásquez, G. Némethy and H. A. Scheraga, Conformational energy calculations on polypeptides and proteins, Chem. Rev., **94**, (1994), 2183-2239.

[34] R. L. Williams, J. Vila, G. Perrot and H. A. Scheraga, Empirical solvation models in the context of conformational energy searches : Application to bovine pancreatic trypsin inhibitor, Proteins, **14**, (1992), 110-119.

[35] S. S. Zimmerman, M. S. Pottle, G. Némethy and H. A. Scheraga, Conformational analysis of the 20 naturally occurring amino acid residues using ECEPP, Macromolecules, **10**, (1977), 1-9.