De Novo Design of Ligands for Metal Separation

DOE Grant FG07-96ER14693
Final Report – Garland R. Marshall, P. I.
September 14, 2001

While the report is entitled "Final", this report is, in fact, a progress report of an ongoing collaboration between the P.I., Prof. Anders Carlsson of the Department of Physics and Prof. Jay Ponder of the Department of Biochemistry and Molecular Biophysics to develop a theoretically based potential function for molecular mechanics that can be incorporated into research-quality force fields for computer-aided design.

Prof. Carlsson has concentrated his expertise of the development of a theoretical basis of a generalized functional form for transition metals and initial calibration for metals (Cu, Zn, etc.) commonly seen in biological systems before extensions to lanthanides and actinides.

Prof. Ponder and Dr. Rohit Pappu have focused on the global minimization problem with crucial extensions to the diffusion equation approach of Piela et al.. We hope to marry the diffusion equation approach, that has been shown to be capable of correctly smoothing the potential surface of biological systems as complicated as a transmembrane helix dimer, with a Gaussian version of the ligand-field functional form from Prof. Carlsson in the near future.

The P.I. has focused on metal-ligands such as siderophores and the design and evaluation of metal-ligand complexes with conventional force fields to provide a basis for comparison with the improved model. In addition, the P.I. has supported an effort to provide a graphical-user interface (GUI) to TINKER, the command-line, molecular-mechanics toolkit, that Prof. Ponder distributes freely over the internet. We are implementing all of our research tools as TINKER subroutines and a user-friendly GUI would make this tool much more accessible by the scientific community. The first version of such a GUI has been completed.

Progress has been episodic as well as problematic. Several episodes of insubordination resulted in dismissal of a research associate from the project despite the uniquely strong theoretical background and excellent progress that the individual was making. Recruitment of an appropriate replacement significantly delayed progress.

Each of the three faculty have written a short summer of their progress during the tenure of this Grant.

I. Quantum-Mechanically Derived Force Fields for Coordination-Bonded Transition Metals

(Prof. Anders Carlsson, Department of Physics, Washington University)

Our efforts have been aimed at both developing functional forms for the transition-metal force fields, and parameterizing these force fields on the basis of quantum-mechanical total-energy calculations.

A. Development of functional forms. The analysis began with a ligand-field Hamiltonian derived using the angular-overlap model (AOM). This formulation is suitable for ionically bound transition metals. The second moment of the ligand-field Hamiltonian, or the trace of its square, was used as a measure of the ligand-field stabilization energy. For sigma-bonded ligands, this yielded an expression for the ligand-field stabilization energy as a sum of angular interaction terms between ligands, where the functional form of the angular dependence is completely determined by the ligand-field Hamiltonian. Tests of the accuracy of this form were performed on a series of model clusters, and it was found that the ligand-field stabilization energy is reproduced with about 90 percent accuracy. The analysis was subsequently extended to pi-bonded systems. In this case, the angular dependence of the interaction is different, but otherwise the form of the energy function is the same. The accuracy is about the same as for the sigma-bonded case. We have also extended the analysis to treat cases that are not in the ionic limit, i. e. where covalent-bonding effects are important. Accuracy similar to the ionic case was found for the angular energy functions.

We performed some very simple application of this force field to blue-copper proteins. Using ballpark parameters derived from, for example, typical values for the ligand-field stabilization energy, we found that the structures observed in the native protein are very close to being an energy minimum for our force field.

B. Parametrization of the force field. This is based on total-energy calculations for small clusters. At this point, we have been focusing on clusters of transition metals surrounded by four ammonia molecules or four water molecules, where the nitrogens or oxygens are neighbors of the transition metal. In order to evaluate the magnitude of the angular forces, we calculate the energy difference between the square and tetrahedral structures of these clusters. The calculations are performed using the VASP (Vienna Ab-Initio Simulation Package) method, which uses a plane-wave basis set. This greatly simplifies the procedure of evaluating whether or not the calculations are converged with respect to basis set. A supercell geometry is used, and the results are monitored to make sure that the supercell is large enough.

For Cu++ and low-spin Ni++, we find that the square geometry is preferred by an energy in the range of 10-20 kcal/mol. This is basically in line with known ligand-field stabilization energies. For Zn++, which has no ligand-field stabilization energy, we had expected to find a small electrostatic energy favoring the tetrahedral structure. However, we found the energy favoring the tetrahedral structure is over 20 kcal/mole, much too large to be explained by electrostatic considerations. Instead, we feel that a partial covalency effect is responsible. We have derived a simple description of this effect in terms of angular ligand-ligand interactions.

Using these parameterizations, we have performed energy-minimization calculations for small complexes of a transition metal surrounded by four imidazoles. The energy-minimized structures agree with the observed structures, except that the imidazole planes are rotated with respect to the axis connecting the nitrogen to the transition metal. We

have also looked at the energetics of antisite defects in Cu-Zn superoxide dismutase (hypothetical defects in which the Cu and Zn ions are interchanged). We find that the energies of these defects are large enough they would never be observed. This is in distinction to most force fields in use today. With these force fields, which do not include the ligand-field stabilization energy, Cu and Zn ions are very similar and thus the antisite defect energy will be rather small.

We are at present working with Dr. Jay Ponder on incorporating the metal-ligand force field into his TINKER code in a graceful fashion.

II. Diffusion- Equation Approach to Global Minimization and TINKER Molecular Mechanics Toolkit.

(Prof. Jay Ponder, Department of Biochemistry and Molecular Biophysics)

Conformational search procedures are concerned with the generation of sample structures from the low energy regions of conformational hyperspace available to a specified molecule or complex. The procedures can be divided into two broad classes. One group consists of methods for obtaining statistical mechanical results from the potential energy surface (see Allen and Tildesley, 1987). These methods need to sample the available conformations in a thorough, statistical mechanically valid fashion. An prominent example of this type of search is the widely-used, and much debated, free energy perturbation technique. The requirement of adequate sampling is currently thought to be a major factor limiting application of free energy simulations to small, relatively homogenous systems. Even very small ligands in aqueous solution can require hundreds of picoseconds of molecular dynamics simulation to arrive at converged results (Pearlman, 1994). Use of free energy methods to model mutations or binding events in large, heterogeneous protein systems has been the subject of much debate (Yun-yu, *et al.*, 1993, and many subsequent papers).

The second class of conformational search, and on one we studied with DOE funding, consists of the rapid generation of a panel of reasonable trial conformations which can then be used to screen for any desired property. Ideally, all reasonable portions of conformational space should be searched. Examples would include docking procedures in which the protein conformation must vary in order to achieve an optimal fit between two structures (Meng, *et al.*, 1994). The tertiary template algorithms for inverse protein folding as implemented in the PROPAK program suite (Ponder and Richards, 1987b, 1987c, 1989) represent a heuristic approach to constrained combinatorial search. The systematic search and macrosearch methodology, developed at Washington University in Prof. Marshall's group, is another variant of directed conformational search that is very effective when applied to appropriately constrained pharmacological problems. Of great importance for the design of conformational search procedures is the fact that most minimization methods move continuously lower in energy until a local minimum is reached. A search strategy must avoid being trapped in only nearby local minima by defining a path that is able to overcome energy barriers. Alternatively, search can be based upon repeated projection to low energy structures consistent with some set of reasonable geometric or energetic constraints, as in distance geometry, energy embedding and the potential energy surface (PES) smoothing methods discussed at great length in this proposal.

Global optimization represents the extreme, ultimate goal of unconstrained conformational search. When applied to small systems, a global optimizer is typically asked to find the single lowest function value. For large systems, one usually is content with one or a few very good local minima. Tremendous amounts of effort have been expended on the "global optimization problem" in the context of econometrics and various aspects of applied math (Torn and Zilinskas, 1989, Floudas and Pardalos, 1992, and Horst and Tuy, 1993). Global searches of peptide and protein conformation also have a long history, much of it centered in the research group of H. A. Scheraga (see Scheraga, 1992 and subsequent reviews). This "multiple minimum" problem has been effectively solved for very small peptides such as the five-residue enkephalin. For larger protein problems, the only reported solutions are for highly regular or symmetrical structures.

The most general search methods of this second class are aimed more toward scanning the available space than at producing a statistical result. Many of these procedures have come in recent years from the small molecule community (Leach, 1991). The most successful current methods tend to be based in internal coordinates and use modifications of Monte Carlo-torsional sampling (Senderowitz, *et al.*, 1995). In searching for the global minimum energy conformation of cycloheptadecane, standard simulated annealing protocols in Cartesian coordinates are distinctly inferior to other methods (Saunders, *et al.*, 1990).

However, the performance of this entire class of methods on large, folded macromolecules is uniformly poor. The algorithms tuned for flexible small molecules often do not scale well when extrapolated to biopolymers the size of proteins. Many of these algorithms fail due to an inability to generate a search path consistent with the large excluded volume of condensed folded structures. For general macromolecular conformational searches, simulated annealing techniques such as those embodied in the X-PLOR package are widely used (Brünger, 1992). A number of alterations have been suggested for increasing the efficiency of simulated annealing searches: increasing the mass of light atoms, increasing the time step (Zhang and Schlick, 1993) and using softer repulsion terms (Laughton, 1994).

Indeed, this paradigm is repeated across all of numerical analysis and computer science - a heuristic, deterministic search is always superior to a stochastic one provided an appropriate heuristic can be found. The development of even partially successful conformational searches for larger peptides and proteins would greatly enhance our ability to build useful protein receptor models, or to screen ligand-receptor interactions based upon receptor conformation in addition to ligand conformation. The present search procedures fall far short of providing this optimal capability.

Global optimization is an important issue in complex systems such as glasses, clusters and large biomolecules. Issues of global optimization are connected to the interest in locating the lowest energy configuration of large molecules. A suite of techniques have emerged over the past ten years, all of which exhibit varying degrees of success in application to an established set of test problems. These methods present plausible strategies to overcome the multiple minimum problem. The governing principle behind such strategies is summarized in the work of Zwanzig, *et al*. (1992) who showed using a simple mathematical model that Levinthal's (1968) paradox can be circumvented if a search procedure incorporates small and reasonable energy biases against unfavorable conformations.

The state-of-the-art in global optimization methods can be classified into four major categories: (1) deterministic methods, (2) stochastic methods, (3) heuristic methods and (4) smoothing methods. Deterministic methods include (a) space covering techniques such as branch-and-bound search (Horst and Tuy, 1987), (b) systematic search methods and (c) trajectory and generalized descent methods (Griewank, 1981). These methods are useful for small problems or for larger systems with well-established constraints, but in general will fail due to an exponential increase in the size of the space to be searched.

Stochastic methods include (a) Bayesian statistical models (see Torn and Zilinskas, 1989, Chapter 5), (b) Monte Carlo techniques, (c) simulated annealing (Kirkpatrick, *et al.*, 1983) and (d) multicanonical algorithms (Berg, 1996). An underlying theme in stochastic search procedures is the use of Monte Carlo sampling methods enhanced by the Metropolis criterion (Metropolis, *et al.*, 1953). Some notable Monte Carlo techniques include the Monte Carlo-Minimization method (Li and Scheraga, 1987), a revision of MCM referred to as "basin hopping" (Wales and Doye, 1997), acceptance ratio methods such as Dynamically Optimized Monte Carlo (Bouzida, *et al.*, 1993), and reweighting methods (Janke, 1994).

A noteworthy stochastic method is simulated annealing which is an important paradigm for global optimization on rugged energy landscapes. In simulated annealing the system is coupled to a heat bath that is initially at some high temperature. At this temperature the system is characterized by rapid transitions between high and low lying minima. The temperature is then slowly lowered according to a prescribed cooling schedule and the system is allowed to equilibrate at each level using either a Metropolis Monte Carlo criterion or molecular dynamics simulated annealing (MDSA; Brünger, *et al.*, 1987). Decrease in temperature is associated with an increased likelihood of being in low energy states and reduced likelihood of jumping out of minima. The approach is analogous to the slow cooling or annealing of a system through the transition region between the liquid and solid phases. The success of simulated annealing is largely determined by the cooling schedule which is closely related to the size of the largest barrier on the potential energy surface and the separation of between the global minimum and lower lying conformations. While simulated annealing has been of limited use in global optimization of protein and other biomolecular conformations, MDSA methods are the *de facto* standard in refinement of X-ray and NMR structures of biomolecules (Brünger, *et al.*, 1997).

The algorithmic state-of-the-art in simulated annealing is Adaptive Simulated Annealing (ASA) that involves an adjustment in the maximum deviation of optimization variables for acceptance ratios of greater than 25% combined with importance sampling methods (Wang and Pachter, 1997). Another variant of Monte Carlo simulated annealing derives from generalized statistics or Tsallis statistics (Tsallis, 1988). The classical Boltzmann machinery can be recovered as a special case of this generalized formalism. The generalized simulated annealing method (Tsallis and Stariolo, 1996) has been revised to satisfy detailed balance (Andricioaei and Straub, 1996) and adapted to reduce to a steepest descent algorithm at low temperatures.

Several of the above mentioned stochastic search algorithms have been applied to a subset of conformational problems derived from a set including varying sizes of Lennard-Jones clusters and water clusters, oligopeptides such as Met- and Leu-enkephalin, polyalanine and glucagon, and organic molecules like cycloheptadecane.

Heuristic algorithms are based on a reduction of the problem into smaller subsets for optimization and assembling the sum of optimized parts to obtain the final answer.

These methods include (a) the build-up procedure (Vasquez and Scheraga, 1985), (b) constrained systematic search algorithms (Beusen, *et al.*, 1996), (c) scanning methods which replace exhaustive enumeration with Monte Carlo methods (Ripoll, *et al.*, 1991), (d) generalized chain growth methods such as Pruned Enriched Rosenbluth Methods (Frauenkron, *et al.*, 1998), (e) the random kick method (Saunders, 1987) and (f) genetic algorithms (Goldberg, 1989) derived from parallels between biological evolution and optimization problems. A large number of these heuristic methods are based on simplifying assumptions about the importance of short-range interactions over longer-range interactions. These methods lead to iterative generation of large yet manageable numbers of conformations from which the lowest energy conformer is located. An important shortcoming of most heuristic methods is the overemphasis of short-range effects over long-range interactions as determinants of low energy conformations.

An emerging concept in global optimization is a mathematical transformation to smooth the multidimensional potential energy surface of a molecule. The resulting surface is less rugged and is much easier to search for minimum energy conformations. It has been shown previously that short range potentials generate large numbers of local minima, while increasing the range of these potentials via softening reduces the number of local minima (Hoare and McInnes, 1983). This concept has been used to deform conformational energy functions in a deterministic fashion according to the Diffusion Equation Method (DEM) of Scheraga and coworkers (Piela, *et al.*, 1989, and succeeding papers). Straub's group (Ma and Straub, 1994) has developed methods for annealing the classical density distribution, called Gaussian Density Annealing (GDA). They use a Hartree approximation to the many body distribution function, which leads to largely deformed potential energy surfaces for large temperatures. The equations of motion in terms of the reciprocal temperature for a packet center and width are solved for decreasing temperatures, *i.e.*, annealing. The GDA method reduces to a diffusion equation method (DEM) for the original potential function in the limit of identical packet widths for all of the atoms. Smoothing of potential energy surfaces can be extended to smoothing of the Gibbs distribution function; the theme of packet annealing methods from Shalloway and coworkers (Church, *et al.*, 1993).

The objective in potential smoothing is to smear out some of the roughness of a rugged potential energy landscape, hopefully without altering the basic structure of the landscape. Traditional search procedures will not confront local minima or energy traps on the deformed surface. If the procedure is reversed then one can potentially find a way back into a region that lies near the global minimum of the original undeformed surface. The success of smoothing algorithms is contingent upon (a) there being a connection between the deformed surface and the undeformed surface, and (b) the presence of underlying structure to the rough undeformed surface. Smoothing is similar to a many-to-few mapping, and in this way the multiple minimum problem is made tractable.

One of the best known benchmark problems for conformational search involves the determination of the low energy conformations of the highly flexible cycloheptadecane (Saunders, *et al.*, 1990) This system continues to attract attention and serve as a test for newly developed search methods (Ngo and Karplus, 1997). While not a particularly large molecule, this system presents a difficult challenge due to its great flexibility and the close energy spacing of the lower lying minima. Extensive analysis via a variety of search methods has located exactly 263 minima within 3.0 kcal/mole of the purported global minimum. **{**Since the full spectrum of energy minima for this molecule

has not been described in the literature, we undertook its generation. A NMLS-based basin-hopping algorithm was used to generate local minima converged to an RMS gradient per atom of 0.000001 kcal/mole on the MM2 potential energy surface traditionally used in studies of this problem. Elimination of duplicate minima resulted in 20,469 unique minimum energy structures with an MM2 energy distribution as shown in Figure 5. Even with the use of truncated Newton minimization, generation of the full distribution required nearly 2 weeks of CPU time on a 250 MHz DEC Alpha workstation. From the decay in the rate at which new minima were located over the course of the calculation, we estimate that essentially all of the available local minima have been located for this system.} In agreement with the original study by Saunders, *et al.*, we find the global minimum at an MM2 energy of 19.0937 kcal/mole. A second minimum lies only 0.01 kcal/mole above the global minimum, and these two structures are separated by about 0.4 kcal/mole from the third best and subsequent structures. We find 10 minima within 1.0 kcal/mole of the global minimum, and 68 within 2.0 kcal/mole.

Application of our current PSS algorithm to cycloheptadecane is dramatically effective at locating many of the lowest energy structures for this flexible system. In all tests we used a maximum deformation of t_d = 25.0 Å^2 at which point only one minimum remains on the smoothed surface. Slight variations in the reversal protocol (n_d = 100 to 150, and s = 2 or 3) coupled with variation in the number of modes searched during NMLS (values from 5 to 16) resulted in the procedure finding different low energy structures. Five consecutive runs with parameters selected from the above ranges, gave energy minima ranked 2, 4, 7, 8 and 9 out of the full distribution of 20,469 known minima. All of these minima are within 1 kcal/mole of the global minimum. Depending on the modes searched during NMLS each run required as little as 20 minutes of CPU time on the same machine used to generate the full distribution of structures.

The global minimum for cycloheptadecane and the second lowest energy structure which was located by our PSS method. We were unable to cause the algorithm to locate the absolute global minimum for cycloheptadecane using reasonable search parameters. However, the global minimum is only 0.01 kcal/mole lower in energy than minimum 2, and based on its MM2 vibrational frequencies is entropically disfavored relative to all of the minima located by our procedure. In fact on the potential energy surface corresponding to the very small deformation value of t = 0.00024 Å^2, minimum 2 from the undeformed surface is already the "global minimum".

We applied our Potential Smoothing Search (PSS) method to obtain the minimum energy conformation of the TM helix dimer of glycophorin A (GpA). The structure of the GpA helix dimer was recently solved by solution NMR spectroscopy (MacKenzie, *et al.*, 1997). This structure shows an average crossing angle of -40° (about a mutually parallel orientation) between the two helices and a dimeric interface with no inter-helical hydrogen bonds. In our application of the PSS algorithm to obtain the minimum energy conformation for the GpA dimer, we ignored all electrostatic interactions. This is a reasonable first approximation since the sequence of the 18-residue helix monomer -TLIIFGVMAGVIGTILLI- does not contribute any dimerstabilizing hydrogen bonds. All calculations were performed using the TINKER modelling package (Ponder, 1997) and the deformable OPLS (DOPLS) force field described earlier. Three different helix packing calculations of increasing generality were performed in our application of the PSS method to predict the structure of the GpA helix dimer. These were: (1) packing

rigid helices obtained from the NMR structure; (2) packing idealized rigid helices which have backbone (ϕ, ψ) angles set to canonical values of (-60°,-45°) and side chain torsion angles (χ) set to values chosen from a backbone dependent rotamer library (Dunbrack and Karplus, 1993); (3) packing of semi-rigid helices with canonical (ϕ, ψ) angles and flexibility allowed for selected side chain torsions.

Packing rigid helices from the NMR structure: Coordinates for the individual helices were obtained from the consensus NMR structure. The PSS method finds a minimum energy conformation with an inter-helical Gaussian van der Waals energy of -29.56 kcal/mole on the undeformed DOPLS surface. The crossing angle is Ω = -52.2° and the closest distance of approach between the two helices is d = 6.36 Å. The value of Ω for the PSS structure is different from the average crossing angle seen in the NMR structures which is -40°. The contact distance is close to the values observed experimentally and structures generated from a global search simulated annealing calculation (Adams, *et al.*, 1996). The rotation angles about the helical axes, relative to the NMR structure, are α = -0.4° and β = 6.2° for the A and B helices respectively, indicating the contact interface between the two helices is very similar. We computed the *rms* deviation for an α-carbon superposition of the PSS structure and each of the 20 NMR structures. The smallest value for the *rms* deviation is 0.64 Å. We also show that the PSS structure is in fact the global minimum by characterizing the undeformed PES through a series of systematic two body grid searches over all the rigid body degrees of freedom. This exhaustive grid search results in a non-gaussian distribution of energies for the 5834 unique minima. It is qualitatively similar to the corresponding cycloheptadecane figure, except for the few lowest minima which are better separated from the large, approximately gaussian central peak. The global minimum from the grid search is identical to the structure found from the PSS calculation. Since the PSS calculation is completely independent of starting structures, it is considerably more efficient and general than any extensive search procedure. The reliability of grid searches are largely determined by the completeness of the set of sampled conformations. A typical PES has a large number of minima even in the vicinity of the global minimum. The potential smoothing part of the PSS algorithm circumvents the extensive sampling requirement by smearing out rough features of the PES. The local search part of the algorithm corrects for crossings at bifurcations and the global minimum can be found reproducibly from arbitrary starting structures.

***Ab initio* prediction of the dimer conformation using rigid, idealized helices:** The calculation described above uses helix monomers from the NMR structure and the result of the PSS calculation could be biased by the choice of NMR internal coordinates for the helices. A true test of the PSS algorithm would be to use idealized helices and sidechain conformations from a rotamer library to investigate the possibility of an *ab initio* prediction. We first generated model structures for each of the helix monomers. A two body grid search identified 4105 unique local minima for the packed helices. The distribution of energies for the unique minima found from the grid search is gaussian with a low energy tail. The global minimum has an inter-helical van der Waals energy of -31.84 kcal/mole, Ω = -52.8° and d = 6.58 Å. The rotation angles are α = -2.8° and β = -3.8°. A PSS calculation using the eigenvectors of the rigid body Hessian as search directions for the local search finds the global minimum from completely arbitrary starting orientations for the two idealized helices. This is impressive since the PES is dotted by a large number of minima, many of which are energetically close to the global

minimum. However, conformational energy is not an adequate parameter to distinguish between predicted structures. Table 1 shows a set parameters for the ten lowest energy structures found from the grid search. Comparison of the appropriate structural parameters indicates that the lowest energy conformer is in fact the one that is closest to the NMR structure and this is the structure that the PSS calculation predicts. Results from Table 1 also establish the roughness of a typical PES, *i.e.*, structures that are far apart can have similar conformational energies. Our original hypothesis that the structure of interest is the global energy minimum is also justified in this calculation. Furthermore we have established a method that successfully locates the global minimum irrespective of the starting orientation for the idealized helices.

Table 1: Comparison of conformational energies and structural parameters for the set of ten lowest energy conformers obtained from a two-body grid search using model built idealized helices.

Interhelical vdW Energy (kcal/mole)	Crossing Angle Ω	Contact Distance d (Å)	Helix A Rotation Angle α	Helix B Rotation Angle β	Relative Shift s (Å)	Smallest *rms* C_α superposition on NMR structures
-31.84	-52.81	6.58	-2.75	-3.75	0.00	0.74
-31.31	-159.91	7.49	69.54	68.54	0.00	8.28
-30.37	-165.69	7.33	48.54	47.54	0.00	7.96
-30.23	-135.86	6.62	25.22	24.22	0.00	6.00
-29.75	154.84	7.64	19.14	18.14	0.00	10.02
-28.72	144.14	7.38	3.02	79.97	4.47	9.76
-28.19	-144.63	6.97	57.09	56.09	0.00	6.65
-28.04	-124.14	6.30	46.42	45.42	0.00	7.22
-27.25	-50.82	6.83	103.85	67.89	0.68	4.19
-26.97	159.26	7.70	22.86	21.86	0.00	9.84

***Ab initio* prediction of the dimer conformation using semi-rigid idealized helices:** Of the 18 residues in the helix monomer sequence, only leucine, phenylalanine and methionine do not exhibit a clear preference for helix dependent rotamers. The most general calculation would be to let the χ-angles of these residues be flexible. The optimization protocol chooses a set of ideal values for the rigid body coordinates from global motions. Side chain torsions based on local motions are coupled to the rigid body degrees of freedom to find the global minimum. We choose the six "rigid body" degrees of freedom for each helix, (χ_1,χ_2) angles of Leu-75, Leu-89, Leu-90, Phe-78 and the (χ_1,χ_2,χ_3) angles of Met-81 as optimization parameters (*21*). A PSS optimization over the 34 degrees of freedom finds a structure with an interhelical energy of -31.53 kcal/mole. Note that this structure has a higher interhelical energy than for the rigid idealized helix minimum, but a lower total energy including torsional terms. The helix parameters of Ω = -47.9° and d = 6.78 Å. The rotation angles are α = -6.2° and β = -7.0°. The smallest RMS of this structure from an α-carbon superposition on the set of NMR

structures is 0.59 Å. While their is a small difference in crossing angle between this particular NMR model structure and our global minimum, in all other respects the two structures are essentially identical. It should be stressed here that this represents a completely general ab initio prediction free of any biases toward the NMR conformation.

Another important aspect of a PSS-style algorithm, besides accuracy, is computational efficiency. A method based on molecular dynamics simulated annealing (MDSA) searches has also been used to generate an *ab initio* prediction that is in very good agreement with the experimental data (Adams, *et al.*, 1996). The model structure of reference and the NMR structure show a difference of 0.8 Å *rms* for the backbone atoms of residues 74-91. The method used is based on exhaustive two-body searches coupled to MDSA runs for a seriesof conformations. Molecular dynamics simulations were preformed at 600K and 300K for 5,000 steps at each temperature followed by energy minimization. The MDSA runs use conformations generated from the two-body searches as starting points. An important feature of these MDSA simulations is the large scale sampling of conformational space achieved through global search methods. A PSS method is different from an MDSA type simulation in that it obviates the need for multiple simulations from different starting conformations. This is because the results of a PSS calculation are always independent of the starting conformation. The local search aspect of PSS can be generalized to yield a set of structures in addition to the global minimum. Methods aimed at this type of generalization are a specific aim of the current proposal.

A useful way to compare the efficiency of simulation methods is to distinguish between the number of function and gradient evaluations required. A single function call implies the calculation of both the energy and gradient. For the helix dimer studied here a typical MDSA simulation, as implemented by Adams *et al.*, would require upwards of 10,000 function calls for each run. The global search component of their calculation will require approximately 500 independent MDSA runs for a total computational cost of at least $500 \times 10{,}000 = 5$ million function evaluations. The PSS method *with fully flexible side chains and with exhaustive local searches* coupled to the smoothing protocol on average requires between 50 to 100 thousand function calls. Thus a "brute force" PSS is already 50-100 times faster for locating the Glycophorin A helix dimer than a simulated annealing global search. If only a few directions are chosen, by means of a heuristic selection procedure, as discussed in the future research plan, a further reduction in computational time is feasible. We estimate that an additional reduction in function evaluations of 5-10 fold should be achievable.

The accuracy of the predicted results and tunable computational efficiency makes a search enhanced potential smoothing method attractive for global optimization and conformational searching applied to typical problems such as X-ray and NMR structure refinement, docking of ligands to active sites and prediction of antibody loop conformation. We are currently pursuing further improvements to the smoothing algorithms and generalizations to include electrostatics and continuum solvation models needed for treatment of other biomolecular search and/or docking problems.

III. Exploration of Available Transition-Metal Molecular Mechanics Force Fields and Generation of Test Cases for Comparison

(Prof. Garland R. Marshall, Department of Biochemistry and Molecular Biophysics)

Reaka, Ho and Marshall (manuscript in preparation) have examined 11 crystal structures of pentaazacrown ethers with different substituent patterns and complexed with 3 different metals (Mn, Fe, Cd) to compare the relative orientations of side chains with those seen in parent cyclopepntapeptides or other structures of interest such as β-turns. The CADD tool FOUNDATION[42] was used to find overlap of the vectors corresponding to side-chain orientations between ideal β-turn conformations and the crystal structure of the metal complexes. In a simple example, the Mn(II) complex shown orients the side-chain substituents almost exactly as those seen for the *i*, *i*+1, and *i*+2 residue of an ideal type I β-turn. As a peptidomimetic of this turn, it suffers from the fact that only three of the four side chains of the β-turn are oriented correctly. Nevertheless, if only those three side chains that correctly overlap are involved in receptor recognition, then the Mn-complex should show activity. These complexes are of particular interest as they readily crystallize and the orientation of sidechain is very dependent on the metal complexed. Thus, they can be used as experimental probes of the validity of the force field used.

The unsubstituted 1,4,7,10,13-pentaazacyclopentadecane prefers one of two alternate conformations depending on metal and axial ligand when complexed with a metal. First is a pentagonal bipyramidal array with the five nitrogens of the MAC nearly planar. The alternative conformation thought to be responsible for SOD activity in the Mn(II) complexes shows a folded ligand in which one of its secondary amines occupies an axial site forming a pseudo-octahedral complex similar in geometry to the oxidized Mn(III) complex. This hypothesis is supported by the observation that bis(*trans*)-cyclohexano derivatives that stabilize the folded pseudo-octahedral conformation show maximal superoxide dismutase activity while those that stabilize the planar structure are not enzymatically active (Riley, 2001). For our purposes, the important aspect of the work is the growing experimental knowledge base of the effects of chiral substitution on stabilizing either of the two conformers. Different metals have different van der Waals radii and, therefore, require different distances between the metal and the inner-sphere nitrogens. The average distance seen in comparable crystal structures of azacrown-metal complexes is 2.271 Å for Fe(III)-N [152], 2.328 Å for Mn(II)-N, and 2.411 Å for Cd(II)-N distances. These average distances are for the dichloro complexes; in the case of the bis(nitrato) complexes, the average distances are 2.283 Å for Mn(II)-N, and 2.355 Å for Cd(II)-N, a consistent 0.8 to 0.10 Å increase in the Cd distances. Thus, each metal flexes the MAC ring differently resulting in different fixed orientations of the side-chain positions for each conformer. Considering the distances between the Cα and the side chain (the lever), relatively small changes in orientation (0.05 Å at 5 positions around the ring) can cause significant changes in side-chain location.