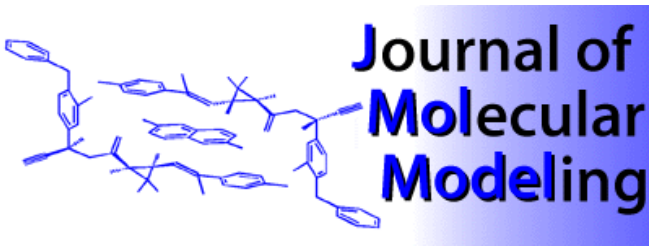
J. Mol. Model. **1997**, *3*, 261 – 286



A Concerted Molecular Design Method by Temperature-Guided Monte Carlo Search / Potential Applications for *ab initio* Ligand Design

Jeffrey H. Rothman*[,1], Romano T. Kroemer

Physical and Theoretical Chemistry Laboratory, Oxford University, South Parks Road, Oxford, UK OX1 3QZ

[1]present address: Department of Chemistry, Havemeyer Hall Box #3154, Columbia University, New York, NY 10027-6948 (jeff@still3.chem.columbia.edu)

Received: 3 February 1997 / Accepted: 2 July 1997 / Published: 21 July 1997

Abstract

This work demonstrates the conceptual feasibility of a method for systematic and concerted design of a molecular structure with a favorable interaction toward a prescribed target. Since a semi-empirical QM hamiltonian is used to calculate energies, the method described here allows the continual atomic change and reassessment of molecular structure with respect to its own internal energy and its target interaction energy. Modelling the ligand molecular structure on an atomistic basis, in contrast to a fragment basis, permits the construction of higher energy transitional structures which bridge between viable ligand structures. This allows a most thorough way to accomplish meticulous configurational sampling. The use of two examples, the creation of an ammonia molecule to interact with a formic acid target, and a formic acid molecule to form a dimer with a target, more clearly illustrates the features of a temperature guided Monte Carlo simulation which was developed to allow thorough configurational sampling. This is accomplished by appropriately varying the temperature of the MC simulation. A series of random starting configurations underwent this systematic MC sampling procedure to locate various structural and interaction energy minima. The results of these simulations demonstrate thorough configurational sampling and convergence.

Keywords: Monte Carlo simulation, automated drug design, molecular recognition, temperature dependent Monte Carlo, molecular design

Introduction

Background

A central problem in current drug design research is to find an inhibitor starting with a known macromolecular binding site. Many procedures have appeared recently which accomplish this in a knowledge-based and automated fashion. Two fundamentally different *de novo* algorithms have been used. One type docks molecular fragments, selected from a predefined library, into the binding site as isolated units and then connects them to form molecules. The programs MCSS [1]/HOOK, [2] GROW, [3] and LUDI [4] are examples of this strategy. A second approach *sequentially* grows molecules in the binding site in stepwise fashion. Examples

* *To whom correspondence should be addressed*

such as SPROUT, [5] LEGEND, [6] GROUPBUILD, [7] and GROWMOL [8] utilize this strategy. The program, LEGEND, fills the binding site with atoms by van der Waals potential and then connects favorable combinations by force field methods. In this manner the ligand interaction and structure are optimized separately. In GROWMOL, atoms and substructure fragments are constructed in the binding site by force field geometries and fit to the target via binding complementarity scores. [8] Its algorithm applies a MC-mediated construction strategy at each step. Both approaches have sequential features and are force field/functional group-based. However, we describe an *QM non-sequential* method using a Monte Carlo (MC) approach. [12]

Central Idea

Reasons for QM approach: Because of the complexity in computational details MM- two body potential functions are widely used. An important feature of these effective pairwise potential functions is that the many-body polarization effects are incorporated in an average manner into the parameters fitted to reproduce experimental fluid continuum properties. [9–11] Unfortunately these parametrisations do not fare so well in nonuniform environments. One of the main difficulties in *de novo* ligand design is the reliability of the calculated nonbonded interaction energies. Nonbonded interactions are much more sensitive to anomeric effects, and electron density displacements are on the same order as interaction distances. In fact, the authors of GROWMOL state that the binding complementarity score is not meant to accurately represent contributions to the free energy of binding. [8] In order to improve the reliability, using semi-empirical QM would be a marked improvement. An additional advantage demonstrated by using QM is the ability to form structures without the need to predispose atoms to a fixed type (as required in force field substructures). This would also allow a much less restricted construction than the substructure/atom-type based force field.

Reasons for non-sequential construction: In all of these current methods the *de novo* creation of the ligand to fit a binding site is executed in a serial fashion. The only apparent allowance occurs in GROWMOL where a sequentially created atom may fuse with a previously created atom in the chain to permit cyclic structures. [8] However, even in this program, no provision is allowed for a sequentially later construction to be an alteration in an earlier construction. This is due to the limitations inherent in force field dependent ligand construction. Prior constructs predetermine those that follow in a given design sequence, ie. later constructs are in response not only to the binding site but also to earlier constructs. Thus a continuously metamorphic/non-serial construction would favor a more extensive structural exploration of regiochemical and configurational space. Similar structure perturbation is also used in DLD [15] which allows a particular atom to change hybridization/connectivity, and atomic number. This is accomplished through variable occupancy (ie. spacially overlapping populations of atoms). However, although the pseudo-energy scoring may be conducive towards convergence, interaction energies would not be well represented. QM methods are much more suited for this type of ligand development since many possible transitional structures that are created would usually not be included in force field substructure libraries.

Operation: This study initially suggests that a ligand be designed to fit a binding site by continually modifying its structure in a randomly driven manner. These changes are accomplished by randomly choosing to create, destroy, move, or change the atomic number of an atom. In any of these cases the initial atom position, changes in position or atomic number are chosen randomly within boundary parameter guidelines. Changes in structure are continually sampled in varied regions and assessed with respect to ligand structural and target binding energies.

Parameter Limits: Although an unrestricted random search would be the most thorough in theory, the practicalities of CPU limitations dictate the use of boundary parameters that direct the process towards more meaningful changes with minimal compromise towards the thoroughness of the search. These fixed spatial parameters include the ligand box boundaries, the maximum move displacement, minimum interatomic ligand-target atom distance, and minimum interatomic ligand-ligand atom distance to improve the chances of a modification being accepted. A maximum interatomic ligand-ligand atom distance has also been incorporated in order to prevent decreased interatomic orbital overlap causing a minimally interacting neutral unbonded atom with weak and unresponsive interaction energies. Since interactions from non-overlapping atoms do not at all reflect their electrostatic and van der Waals properties when bonded, their existance is avoided.

Overview

In order to achieve this parallel ligand construction the scheme would require a continuous sampling of all regions comprising the ligand space to attain a constant reassessment of the ligand-binding site interaction. Two general types of interactions must be assessed during the ligand design procedure. One type, the interactions between the binding site and the proposed atomic arrangement in the ligand region (ligand molecule), determine the binding interaction energy. The second type, interactions among the ligand atoms, determine the internal molecular geometry and energy of the ligand itself. A crucial requirement of this ligand design simulation is the ability to sample the series of atom positions of generated ligand configurations that simultaneously minimize both of these interaction energies without overbias. In other words, the internal energies guide the ligand atom positions into a viable molecular structure while the ligand binding energies

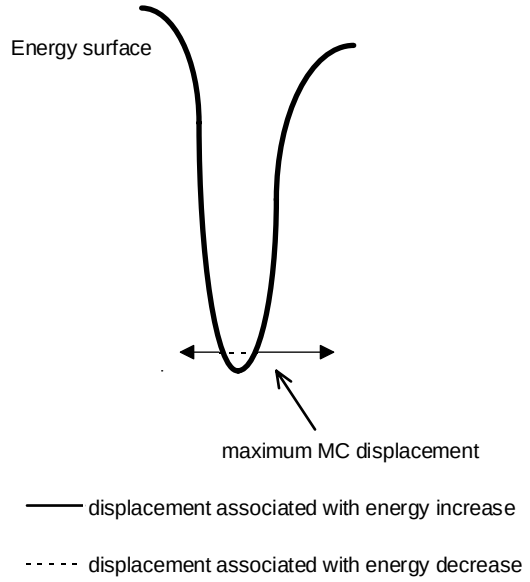


Figure 1a. *Available energy surface within range of MC maximum displacement near energy minimum. The higher proportion of higher energy surface deems a temperature increase to allow further exploration.*

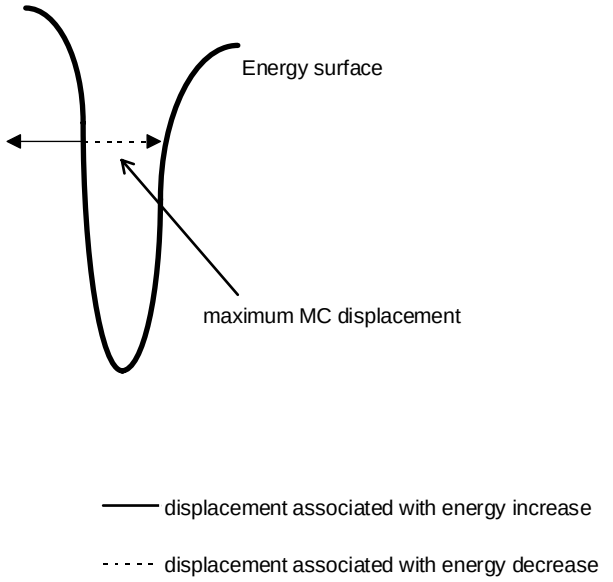


Figure 1b. *Available energy surface within range of MC maximum displacement on slope contour. The higher proportion of lower energy surface deems a temperature decrease to more effectively proceed towards lower energy potential.*

guide the ligand atom positions into favorably interacting structures. The intersection contains the viable solution configurations.

Ultimately, the proposed method for generating this extensively sampled series of ligand configurations would be to run a grand canonical MC-controlled simulation of a series of atoms that would be allowed to change position, atomic number, and to be created or destroyed. Initially the ligand is modelled as free atomic potentials in space, which will eventually coalesce into a molecular structure through repeated MC generated configurations. In the interest of purposeful convergence, the molecular formula of the simulated ligand must correspond to a complete molecule (fulfilling the octet rule) for all configurations. Since speed of calculation is of concern in MC procedures, molecular mechanics empirical methods might be the obvious choice. However, the accuracy of assigned partial charges, force constants, and polarization coefficients may become misleading especially since the model uses molecular atom types in a precariously unbonded fashion. Due to the difficulties and lack of parameter continuity present between geometry and adjacent atom dependent atom types in library functions such as in MM, it is necessary to calculate energies required for MC by semi-empirical QM methods to allow energy convergence.

Aside from the energy calculation method utilized, the internal and binding energies must be appropriately weighted in the Boltzmann term so as to afford a maximum number of both well bound and unstrained ligand structures. A similar approach is used in the GROWMOL program where a Boltzmann weighting factor is utilized to bias the selection of atoms with high complementarity score. This permits structures to be constructed in which groups of atoms with favorable interactions can be connected by atoms which show less target binding complementarity. [8] In this manner molecular construction need not be obstructed by an unfavorable binding complementarity score. The Boltzmann weighting factor in GROWMOL is determined by the user and run at constant value throughout the program, however the method described here continuously varies the weighting factors according to acceptance ratios of their respective structural and interaction energy trajectory profiles.

Variable Boltzmann-search temperature/parallel Metropolis algorithm

The decision to accept or reject a proposed MC-change in ligand structure in these simulations is not obvious. Changes are to be analyzed according to the interaction energy between ligand and target atoms, and the internal bonding energy of the ligand itself. Each energy value is assessed separately by Metropolis algorithm. Difficulties arise when one energy change value is accepted and not the other. A cautious approach would require both energy change conditions to be satisfied before allowing an overall acceptance. This leads to slow but steady convergence upon an ‘acceptable structure’ whereas the requirement of only one satisfactory

change in energy to allow an overall acceptance leads to destructive interference in obtaining (convergence) minimal energy values.

Control of the MC process mainly relies upon adjustment of the Boltzmann-search temperature which guides the sensitivity of the acceptance criteria. While using fixed maximum displacements, the temperatures of each Metropolis process are periodically and separately adjusted so that their acceptance/trial ratios are maintained around a set value. In this manner the temperatures are a function of the energy space that is being sampled, and regulation of the acceptance ratios is then controlled by the temperatures. Accordingly, acceptances are controlled by the locally sampled energy space. When the acceptance ratio becomes lower than the set value (eg. the system has already sampled an energy minimum (Figure 1a) and MC movements correspond to higher energy contours), the temperature is raised to allow escape and facilitate configurational exploration. When the acceptance ratio becomes higher (eg. the system energy is on a slope contour (Figure 1b) where decreasing energies are accessible to at least half of the MC movements), the temperature is lowered to favor energy decrease. For this approach to work effectively the maximum displacement value must be large enough so that it significantly spans the edges of the possible energy minima near enough to their nadir. On account of the differences in well width the maximum atomic displacement for ligand energy (bonded interactions) explorations is smaller than the maximum ligand translation for interaction energy (nonbonded interactions) explorations. In addition the acceptance ratio thresholds should have appropriate upper and lower bounds to adequately detect the location of a trajectory. In this manner the simulation allows the capability of finding and climbing out of energy minima to explore configurational space effectively. At first glance this approach would appear to suffer from a negative feedback feature in which it is simultaneously trying to jump out of energy wells that it is also trying to locate. However, this is overcome with the delayed temperature adjustment response by using a sufficient acceptance ratio sampling rate (ie. sample size).

For each separate energy calculation an acceptance of a new configuration energy may be summarized by the Metropolis MC algorithm,

$$Acpt_{k+1} = \begin{cases} Acpt_k + 1, & \min(1, \exp(-\Delta V_{nm}/kT_i) \geq ran(0,1) \\ Acpt_k, & \min(1, \exp(-\Delta V_{nm}/kT_i) < ran(0,1) \end{cases}$$

where k is the trial number, $Acpt_k$ is the number of accepted configuration transitions, and ΔV_{nm} is the (interaction or ligand) energy difference between the proposed n-state and initial m-state. The acceptance ratio, *Acpt_ratio*, can now be defined as,

$$Acpt\_ratio = \frac{Acpt_k - Acpt_{k-s}}{s}, \quad \mathrm{mod}\,(k,s) = 0$$

where s is the sample size of trials to be scrutinized. At this point the simulation (ligand or interaction energy trajectory) temperature, T_i, is systematically set every s-number of trials as

$$T_i = \begin{cases} T_{i-1} t_{incr}, & Acpt\_ratio_i < lo\_threshold \\ T_{i-1} / t_{incr} & Acpt\_ratio_i > up\_threshold \\ T_{i-1}, & lo\_threshold \leq Acpt\_ratio_i \leq up\_threshold \end{cases}$$

where $i=k/s$,

t_{incr} is the multiplicative incremental temperature adjustment, *up_threshold* is the optimum upper acceptance ratio value, and *lo_threshold* is the optimum lower acceptance ratio value. The sample size, s, for the acceptance ratio (ie. the number of trials per ratio assessment) and the incremental temperature adjustment, t_{incr}, influences the delay of the energy profile response so that this feedback control information is damped. This is necessary since the acceptance ratio is recursively correlated to the temperature, and without a suitable delay destructive feedback would occur. In this pulsed manner the Boltzmann-search interaction temperature successfully guides the interaction energies in and out of potential wells on the energy surface. The general scheme of this procedure is outlined in Figure 2.

Each trajectory, ligand energy and interaction energy, is examined separately and independently by their own acceptance ratios, temperatures, and threshold values. For a configurational change from state-m to state-n to occur, the Metropolis algorithms of both energy changes must permit an acceptance.

Method Details

Initially, it is necessary to prove the feasibility of this novel approach by demonstrating a successful pilot study where only some of the above mentioned types of random changes are operational to limit the configurational space and computational burden. For this purpose, a simulation was devised to show that a nitrogen and three hydrogen atoms, initially and randomly positioned around a target formic acid molecule, would arrange themselves into an appropriately inter and intramolecularily oriented ammonia molecule. The number of atom types are fixed, but they are allowed to mutate one into another in a pairwise fashion. Each atom is allowed to move separately with a (0.3Å) maximum displace-

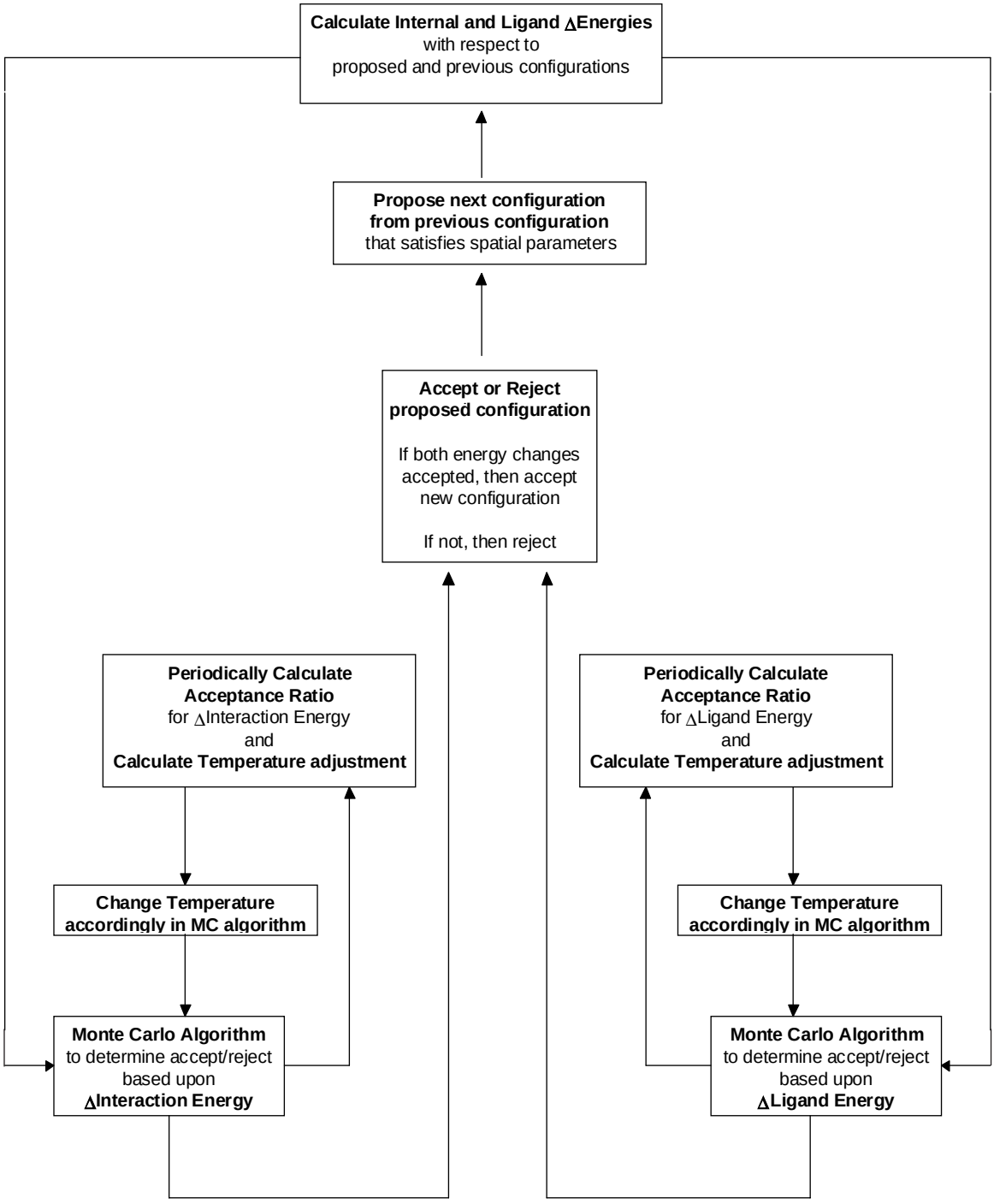


Figure 2. *General scheme describing the Temperature Search MC procedure.*

ment, and the entire ligand can move with a (0.8Å) maximum translation or rotate in two orthogonal planes. Spatial parameters include the ligand box boundaries which are confined within 10.0Å of the formic acid target, a maximum (2.5Å) and minimum (0.8Å) ligand-ligand interatomic distance, and minimum (1.8Å) ligand-target interatomic distance. In this manner unacceptable energy trials are avoided by preventing unfavorably close nonbonded interactions with target and bonded interactions between ligand atoms themselves. A maximum ligand-ligand interatomic distance guarantees that any particular atom will be near bonding distance to at least one other ligand atom.

Due to the existence of relatively long 'transition-bond' distances in the ligand, all energies are calculated by semi-

Table 1. *Intermolecular H-bonding distances and Ligand and Interaction Energies of optimized general orientations and best simulation configurations (av.-average H-bond length of the trifurcated H-bond).*

General Orientations	H_3N:–HC- -formic (Å)	H_3N:–HO- -formic (Å)	NH_3–O= =formic (Å)	NH_3– O- -formic (Å)	Ligand Energy (kcal/mol)	Interaction Energy (kcal/mol)
Config-A	2.69		2.21		-308.65	-3.58
Config-B	2.41			2.36	-308.65	-2.70
Config-C		2.26	2.22		-308.65	-4.10
Simulation-Trial step						
1-926		2.85		2.32	-304.29	-2.05
2-1101			2.66av		-300.41	-3.26
2-373		2.14	2.26		-285.77	-4.17
3-754			2.10		-260.24	-3.04
4-493	2.61				-294.55	-1.80
4-1052	2.74			2.63	-289.23	-1.85
5-1487		2.60			-240.81	-2.21
6-1327	2.67				-293.20	-1.67
6-579		2.75		2.45	-286.32	-2.06
7-1403	2.63			2.77	-293.68	-2.02
8-1131			2.42		-299.39	-2.65
8-1208			2.62av		-299.02	-3.21
9-325	3.32				-305.71	-1.12
10-1088		2.41			-307.08	-2.16

empirical unrestricted HF QM methods (MOPAC6.0). [13] The total energy of the fixed geometry target molecule is calculated once at the beginning of the simulation while the total energy of the entire system (target+ligand) and total energy of the ligand are calculated at each step of the simulation. The ligand-target interaction energies, E_{inter} , are calculated by subtracting the individual target, E^{tot}_{target}, and ligand, E^{tot}_{ligand}, total energies from the total energy of the entire system, E^{tot}_{entire}, simply as,

$$E_{inter} = E^{tot}_{entire} - E^{tot}_{ligand} - E^{tot}_{target} ,$$

where the total energies are just the sum of the electronic and core-core repulsion energies. The internal ligand energies, E_{ligand}, are calculated by subtracting the monatomic total energies, $E^{tot}_{atom(i)}$from the total ligand energy as

$$E_{ligand} = E^{tot}_{ligand} - \sum_i n_i E^{tot}_{atom(i)} ,$$

where i is atomic number and n_i the number of atoms of i-atomic number. Only a constant is subtracted from the total ligand energy throughout the simulations presented here, however this is included as a reminder for future simulations where the number of atoms are varied. The Boltzmann-search temperatures are each periodically adjusted after 10 Metropolis energy trials. All trajectories are limited to 1500 steps where all attempted and accepted moves including their associated energies and temperatures are accumulated. Throughout each simulation the five lowest internal ligand energy and ligand-target interaction energy structures are saved.

Results and discussion

Each simulation begins with the random placement of the set of atoms required to form an NH_3 molecule. Although the random atomic positions satisfy spatial parameters such as the minimum and maximum inter-atomic distances and box boundaries, they still typically correspond to a very high

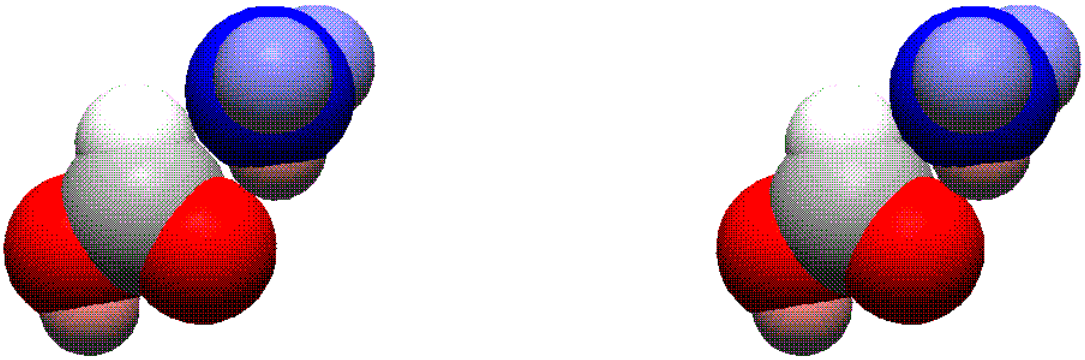


Figure 3a. *Configuration A: corresponding to H_3N:---HC-formic acid and NH_3---O=formic acid H-bonds*

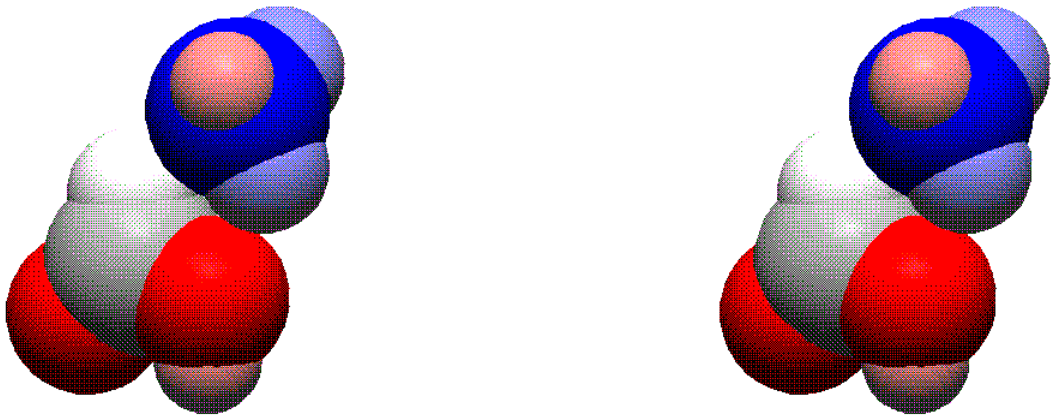


Figure 3b. *Configuration B: corresponding to H_3N:---HC-formic acid and NH_3---O-formic acid H-bonds.*

Figure 3c. *Configuration C: corresponding to H_3N:---HO-formic acid and NH_3--- O=formic acid H-bonds*

Figure 4a. *Trajectory snapshot (trial-457 simulation-1.1) corresponding to the N-H/H_2 local minimum:*
N-H fragment: N-H---O=formic acid 2.19Å
H-H---N-H fragment 2.34Å
H-H bond length 0.89Å

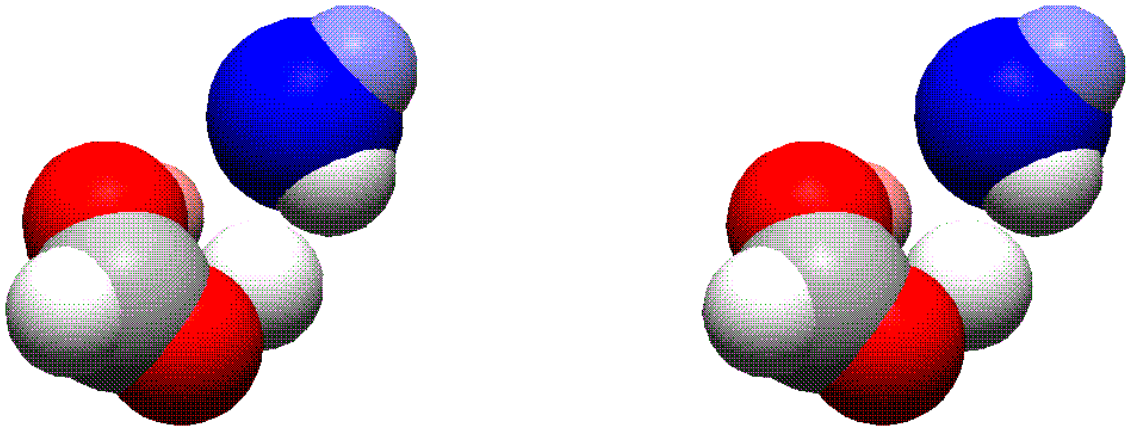


Figure 4b. *A trajectory snapshot (trial-569 simulation-2.3) depicting the concerted development of NH_3 ligand through target interaction and internal ligand interaction*
H_2N---HO-formic acid 2.86Å
H_2N-H---O=formic acid 2.10Å
H_2N---H 2.72Å

initial temperature. In order to facilitate continuity the simulation is started initially at a high temperature corresponding to 3000 K for both the internal ligand energy and interaction energy Metropolis partition functions, and then self regulation of Boltzmann-search temperature becomes operative. Lower initial simulation temperatures such as 600 K would more often result in early delay of trajectories in local potential energy wells.

Initial outline of energy minima

In order to determine the effectiveness of this MC search method, the optimum ligand and interaction energies and their respective orientations were obtained for assessment of trajectory efficiency. General ammonia ligand structure and orientations to the formic acid in vacuo were determined beforehand from all chemically viable arrangements. Three general orientations of the ammonia-formic acid interaction were found to be stable and were assessed. As seen in Table 1, these entries, configurations A-C (Figures 3a-c), are two H-bond systems of the possible H_3N:–HC-formic acid, H_3N:---HO-formic acid, NH_3---O=formic acid, and NH_3---O-formic acid interactions. The other chemically viable pair combinations were unstable and converge into the three listed orientations. These were all evaluated by the same energy calculation parameters used in the MC search. The H-bonding distances and interaction energies of these semi-empirical unrestricted HF geometry optimized systems are listed in Table 1. Optimal ammonia internal ligand energy is a found

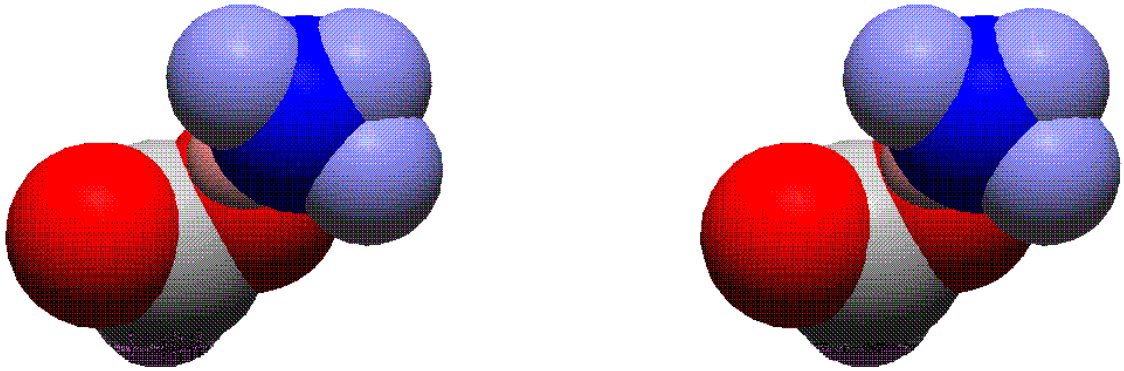


Figure 4c. *A trajectory snapshot (trial-1487 simulation-2.5) depicting the concerted development of NH_3 ligand through target interaction and internal ligand interaction*
H_2-NH---O=formic acid 4.19Å
H_2-(H)N---HO-formic acid 2.56Å
(H---H)NH 1.22Å
H_2---NH 0.88Å, 1.22Å

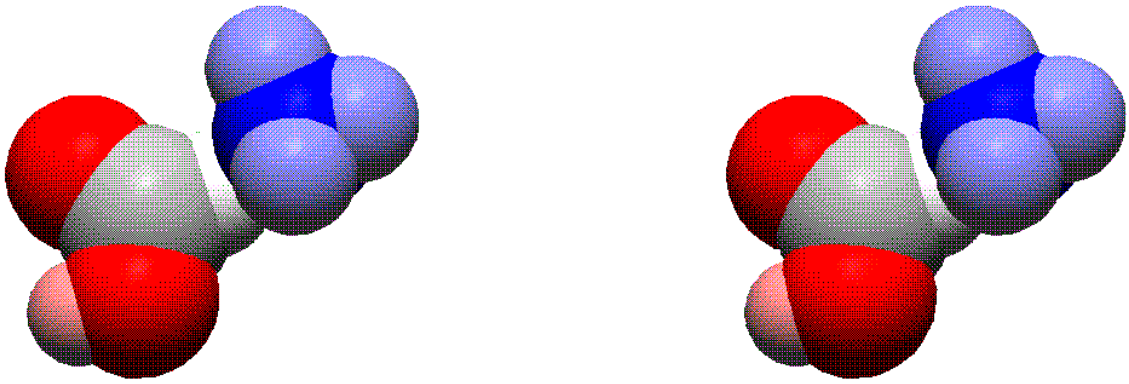


Figure 4d. *A trajectory snapshot (trial-1327 simulation-2.6) depicting the concerted development of NH_3 ligand through target interaction and internal ligand interaction*
H_2-(H)N---H-formyl 2.67Å
(H---H)NH 1.37Å
H_2---NH 0.96Å, 1.00Å

to have a minimum of -308.66 kcal/mol and N-H bond lengths are appropriately 1.00Å for these systems.

Pilot runs and identifcation of initial difficulties

In order to test the search capabilities and sensitivity of ligand assembly, interatomic distance limits must be set to allow effective yet non-restrictive trajectory guidance. Based upon the equilibrium N-H bond and H-bond distances a pilot set of simulations were run with a 1.0Å minimum intra-ligand interatomic distance, a 2.5Å maximum intra-ligand interatomic distance to at least one other ligand atom, and a 1.8Å minimum ligand-target distance parameters. It would seem plausible that maintaining the atoms at closer distances should aid the atomic trajectories in finding the optimum configuration more efficiently. Thus a second pilot set of simulations was run with maximum intra-ligand interatomic distance and minimum ligand-target distance parameters decreased to 2.0Å and 1.4Å, respectively. In order to expedite the energy calculations for each proposed step the SCF convergence criteria was set at 0.1 kcal/mol. Each simulation was executed with different random number sets to allow different starting configurations and to assess the path independence of the search thoroughness and convergence towards a solution configuration. Of these 20 simulations 14 simulation trajectories (seven from each simulation series) successfully converged upon and sampled the optimal configurations and energies.

Four of the simulations became detained in a local ligand energy minimum of ~ –200 kcal/mol and did not sample or converge towards the optimum configurational energies

within 1500 trials. These local minima correspond to generation of the N-H nitrene and H_2 and their unrestricted translational separation (Figure 4a). This is due to the statistically driven tendency to form the less ordered diatomic structures rather than a single pyramidal 4-atom molecule. In an example of such a detained simulation (Plot 1.1/see appendix) the search temperature associated with the internal ligand configuration jumps to 2500 K at ~400 trial steps and then 3000 K at ~1100 trial steps in an attempt to escape this local ligand energy minimum. Although most of the trajectories temporarily sample this state, eventual nitrene insertion into the H-H bond produces the energetically optimum NH_3 coordination.

In one simulation of each series the trajectory becomes restrained within a configuration that is more susceptible to inaccurately calculated energy values due to insufficiently converged SCF calculations. These artifactually occuring errors in interaction energy arise due to the comparitively uneven total ligand and total system energy calculations which are susceptible to insufficient SCF convergence. Both energy surfaces in this region (Plot 1.2) appears to form an artificially wide and flat nadir thus producing a high MC-acceptance ratio. The algorithm then sets colder Boltzmann-search temperatures (Plot 1.2) with the aim of finding a minimum since the fixed maximum displacement and translation values do not significanly span this region. Evidence for such poor convergence is shown by the inordinately low interactions energies (~-30 kcal/mol) produced by the separate ligand and target total energies subtracted from the energy of the entire system. Similar behavior is even evident in nine of the trajectories that successfully converge upon and sample the optimum configurations, however these trajectories succeed in escaping these regions of poorly calculated energy configurations.

Design improvements

In order to circumvent these fragmentations of the ligand molecule into less ordered species, additional restrictions were next incorporated which required each ligand atom to be connected to any other ligand atom through a chain network of distances within the allowed 2.0Å maximum intra-ligand interatomic distance. This condition would more directly arrest the dissociation of the ligand atoms into separate units. As seen in the series of simulations that incorporate these conditions, Plots 2.1-2.10, none of these ten trajectories remains trapped within the energy minima corresponding to the N-H and H_2 configuration.

Setting a much narrower SCF convergence criterion of 0.01 kcal/mol decreases the likelihood of artifactual energies due to poorly converged SCF energy calculations as initially postulated. At this level of self-consistency, only one simulation becomes susceptible to a configuration where its trajectory is restrained due to poorly converged energy calculations. Running another simulation with a narrower SCF convergence criterion of 0.0001 kcal/mol of the same starting configuration (random number set) allows successful sampling of the optimal configurations and energies without any deterred movement (Plot 2.7). Nevertheless, the fact that nine out of ten simulations (Plots 2.1-2.10) successfully converge upon optimum configurations and energies within 1500 MC trials (steps) with only 0.01 kcal/mol SCF convergence criterion, shows much promise.

Minima Search Capabilities

For each simulation run, the five configurations with the best overall ligand and interaction energies were recorded. The best examples of each general type of ammonia-formic acid interaction of each simulation are shown in Table 1. Of particular note is that a trifurcated H-bond configuration from NH_3 to the O=formic acid was located and sampled in simulations-2 and -8 (Plots 2.2,2.8) as trials 2-1101 and 8-1208 (Table 1), respectively. This configuration easily falls into the single H-bond NH_3---O=formic acid orientation (configuration-C) upon MM steepest descent energy minimization thus demonstrating the sensitivity of the MC-temperature search sampling method. Simulation-2 (Plot 2.2) additionally located a configuration (trial 373) incorporating H_3N:---HO-formic acid and NH_3---O=formic acid interactions (Table 1) while simulation-8 (Plot 2.8) also located a configuration (trial 1131) of a single NH_3---O=formic acid H-bond (Table 1). The interaction energy of -4.17 kcal and H-bond distances of 2.14Å and 2.26Å of trial 2-373 are similar to that of the -4.10 kcal and 2.26Å and 2.22Å H-bond distances of the same optimized configuration-C (Table 1). Similarily for trials 3-754 and 8-1131, their interaction energies and H-bond distances are comparable to that of optimized configuration-A, and the results of trials 4-1052 and 7-1403 comparable to that of optimized configuration-B (Table 1). Further substantiation of the sampling efficiency of this MC search routine is also shown by simulations-4 and -6 (Plots 2.4,2.6) (Table 1) each locating and sampling more than one general type of configuration within only 1500 trial steps. Although no single simulation is capable of sampling all of the available interaction energy minima within the alotted trial length, the collective results of these simulations certainly do. In this manner it may be more effective to run a number of simulations due to the various starting configurations rather than a single simulation of extreme length.

Simulations Allowing Multiple Product Formation

In addition to the ammonia fomation/formic acid-binding interaction simulation, a study was performed that would aim to create a formic acid molecule to interact as a dimer with an existing formic acid molecule. As in the previous series of simulations the constituent set of atoms (1C, 2O, 2H in this case) are randomly distributed around the formic acid target molecule, and the trajectories controlled by the same distance parameters, Monte Carlo thresholds, convergence criteria, and number of trial steps. However, unlike the previ-

ous example which only constructed an ammonia or a nitrene/diatomic hydrogen pair, this simulation series can allow the formation of many iso-energetic structures to that of the desired formic acid. In fact the randomly positioned atom sets condense to either the CO_2/H_2 or CO/H_2O van der Waals pairs, or the dihydroxy carbene as well as formic acid. Each of these trajectories incorporate orientations that sample significantly strong interaction energy space (Plots 3.1-3.10).

Of the ten trials, two simulations (Plots 3.3,3.8) created a formic acid molecule that interacted with the formic acid target. In simulation-3.3 the formic acid molecules interact via their formal carbonyl oxygens and formyl-hydrogens as a dimer (Figure 5a) while in simulation-3.8 ligand atoms assemble the formic acid at a much later stage of the simulation whereby only a single carbonyl oxygen to formyl-hydrogen intermolecular interaction is accomplished. Transition to formic acid structure is best indicated by the bond formation between the formyl and hydroxyl radicals of simulation-3.8 (Figure 5b) and carbene decomposition transition structures created during the simulations-3.1,3.5,3.7 (Figure 5c). The CO/H_2O van der Waals pairs created in simulations-3.4,3.9,3.10 (Figures 5d,e) and the CO_2/H_2 van der Waals pairs created in simulations-3.2,3.6 (Figures 5f,g) all demonstrate dipolar interactions with the formic acid target molecule.

Trajectory Control

As seen in the figures of the simulation trajectories of the ligand and interaction energies and their accompanying Boltzmann-search temperatures, the contours of the interaction energy closely follow that of their temperatures for all simulations (Plots 2.1-2.10, 3.1-3.10). Upon careful inspection it is evident that indeed the energy trajectories respond to their temperature guidance by noting a short lag-period between the initial temperature change and the energy response. This is consistent with the pulsed delay-feedback characteristic of the search.

The ligand energies behave similarily, however when they are sampling their lowest energy minimum, they are not as responsive to their Boltzmann-search temperature due to the steepness of the energy wells of bonded atoms in comparison to the relatively shallow contour of the non-bonded interaction potential surface. This can be observed by the series of pulsed jumps to high temperature in the attempt to drag the ligand atom-set out of the well to explore others. In simulation-2.5 (Plot 2.5) the ligand configuration becomes temporarily detained in the nitrene-H_2 configuration whereby the high temperature pulse of 1500 K between 1100-1300 trial steps causes a slight rise in ligand energy peaking at ~trial step 1300 and eventually falling towards the -300 kcal region corresponding to the NH_3 configuration. In simulation-2.3 (Plot 2.3) the ligand configuration trajectory is similarily detained in the nitrene-H_2 configuration. Analogously, the temperature remains at ~2000 K until 450 trial steps until the ligand energy trajectory starts to fall towards the NH_3 configuration minimum. In simulation-3.8 a temperature increase from 1400 K to 3000 K during the 600-1000 trial step interval (Plot 3.8) moves a formyl and hydroxyl radical configuration at -460 kcal into a formic acid molecule at -520 kcal.

When the ligand atom set locates and samples the ammonia configuration, in all simulations the temperature responds by generally increasing again or remaining high near the initial temperature if the ligand energies drop quickly at the start. This is also most evident in the simulations were HCOOH (Plots 3.3,3.8), CO_2/H_2 (Plots 3.2,3.6), and CO/H_2O (Plots 3.4,3.9,3.10) are formed. The temperature rise attempts to redirect the trajectory out of the steep ligand energy well. In this situation the acceptance ratio threshold for temperature adjustment is not set low enough for the given atomic MC maximum displacement to allow the trajectory to exit a

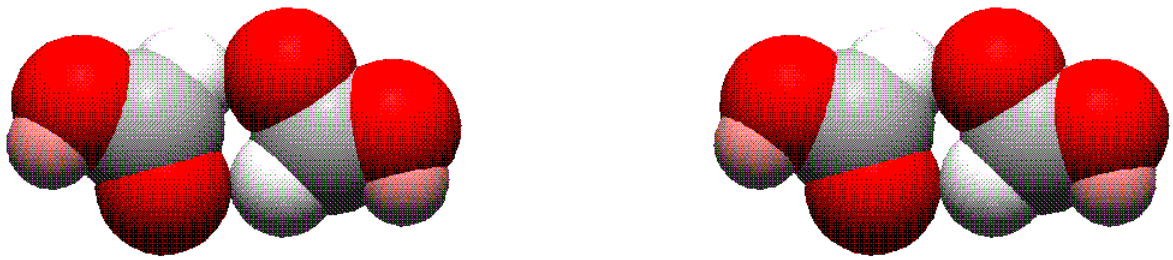


Figure 5a. *A trajectory snapshot (trial-296 simulation-3) depicting the final concerted development of HCOOH ligand assembly as a dimer with the formic acid target*
formic=O---H-formyl target 2.25Å
formyl-H---O=formic target 2.03Å

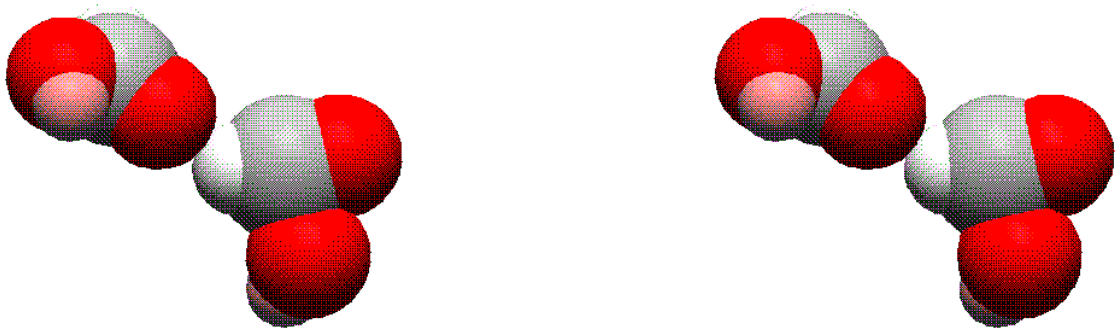


Figure 5b. *A trajectory snapshot (trial-584 simulation-8) depicting the concerted development of HCOOH ligand through target interaction and internal ligand interaction; Note hydroxyl radical approaching formyl radical*
formyl-C---O=formic target 3.13Å
formyl-H---O=formic target 2.20Å
formyl-C---OH 1.97Å

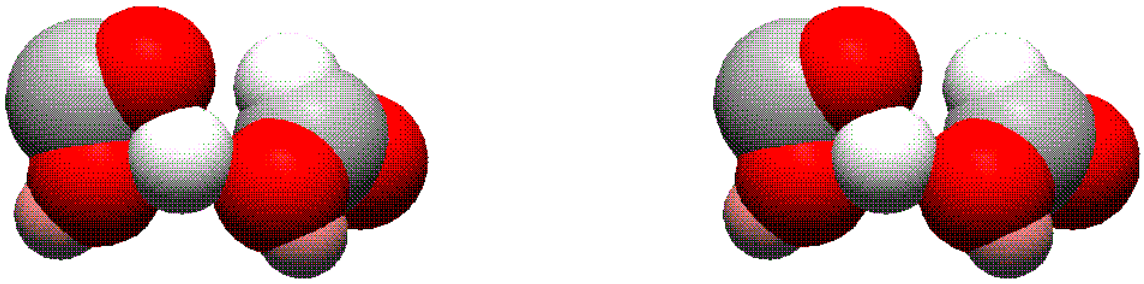


Figure 5c. *A trajectory snapshot (trial-544 simulation-5) depicting the concerted development of HCOOH ligand through target interaction and internal ligand carbene rearrangement interaction; Note hydroxyl hydrogen shifting position to acquire more carbon overlap*
shifting-H---C 2.25Å
shifting-H---O-formic target 1.82Å
shifting-H---O-C 1.74Å

NH_3 configuration (N, 3H simulation), and in the same respect the iso-energetic formic acid, CO_2/H_2 or CO/H_2O configurations (C, 2O, 2H simulation). In this manner it is possible to adjust the search routine to not escape well minima of a particular width and depth.

Concerted Optimization of Ligand and Interaction Energies

Concerted development of the ligand and interaction energies with respect to the ligand structure and its target orientation during simulation are portrayed in Figures 4b-d. A carbonyl dipole stabilized NH_2 radical is seen acquiring a hydrogen atom which is in turn stabilized by the carbonyl oxygen (Figure 4b). Insertion of diatomic hydrogen onto an N-H nitrene can be seen where the amine nitrogen interacts with the target molecule carboxyl proton (Figure 4c) or the formyl hydrogen (Figure 4d). These trajectory snapshots are indicative of the simultaneous assembly of the ammonia molecule and its vanguard non-bonded interactions with the formic acid target. As shown especially in simulations-2.3,2.5 (Plots 2.3,2.5) the ligand-target interaction energies start to decrease a few hundred trial steps before the ligand energies start to decrease to that of the NH_3 configuration. In simulations-2.2,2.4 (Plots 2.2,2.4) the ligand-target interaction energies decrease almost synchronously with the ligand energies.

Similar arrangements are demonstrated in **Figure 5b** where the formic acid target interacts with the hydrogen of the formyl radical subsequent to the hydroxyl radical ap-

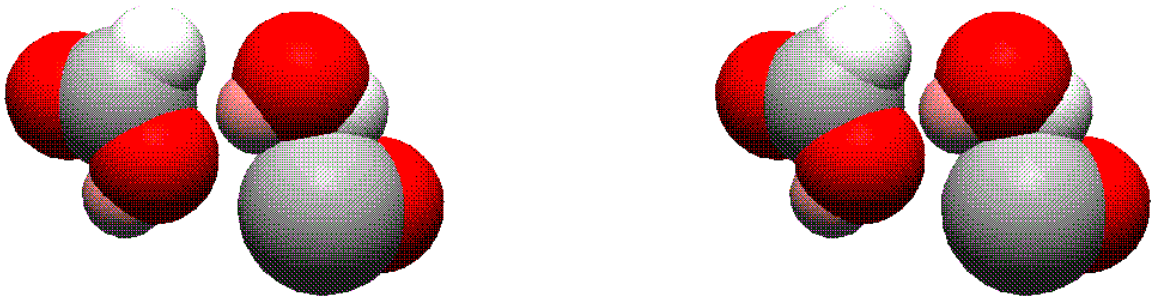


Figure 5d. *A trajectory snapshot (trial-885 simulation-4) depicting the concerted development of a close CO/H_2O ligand pair through target interaction and internal ligand interaction; Note H_2O in close contact with CO and its proton near target- formic hydroxyl-O*
CO---HOH 2.08Å
OC---OH_2 2.58Å
HOH---O-formic target 2.40Å

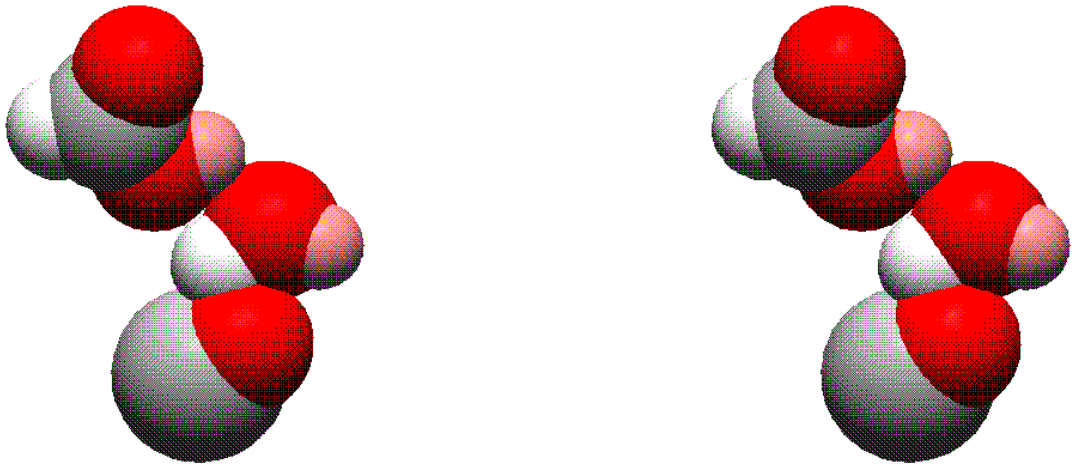


Figure 5e *A trajectory snapshot (trial-1084 simulation-4) depicting the concerted development of a close CO/H_2O ligand pair through target interaction and internal ligand interaction; Note H_2O in close contact with CO and its proton near target- formic hydroxyl-O, and its water-O near formic acid proton*
CO---HOH 1.61Å
HOH---O-formic target 2.41Å
H_2O---HO-formic target 2.06Å

proach. External guidance imparted by the target molecule towards ligand formation in simulation-3.8 can also be seen by the formal carbonyl of the formic acid target interaction with the hydrogen of the formyl radical ligand component (Figure 5b) while in simulation-3.5 the hydroxyl moiety of the carbene transition structure shows a beneficial interaction with the formic hydroxyl and formyl-hydrogen of the target (Figure 5c). In the simulations that create the CO_2/H_2 (Plots 3.2,3.6), and CO/H_2O (Plots 3.4,3.9,3.10) van der Waals pair ligands, the target interactions accomodate a transitional ligand structure that approaches formic acid (Figures 5d-g). Consequently, in simulations-3.2, 3.3, 3.7, 3.8, 3.9, 3.10 (Plots 3.2, 3.3, 3.7- 3.10) a concurrent decrease in ligand-target interaction energy and ligand energy is most easily seen. These examples demonstrate the operation of the nonbonded interactions as guidance for the ligand construction which is a crucial feature of this study. Linking the construction of a viable molecular structure with the development of significant target interaction potential through their energy trajectories is the crucial feature and foundation of this study.

Figure 5f. *A trajectory snapshot (trial-431 simulation-6) depicting the concerted development of a close CO_2/H_2 ligand pair through target interaction and internal ligand interaction; Note formyl C-O dipole interaction with CO_2.*
OCO---C-formyl target 3.18Å
O_2C---O-formic target 4.02Å
H_2---OCO (closest) 1.87Å
H---H 0.97Å

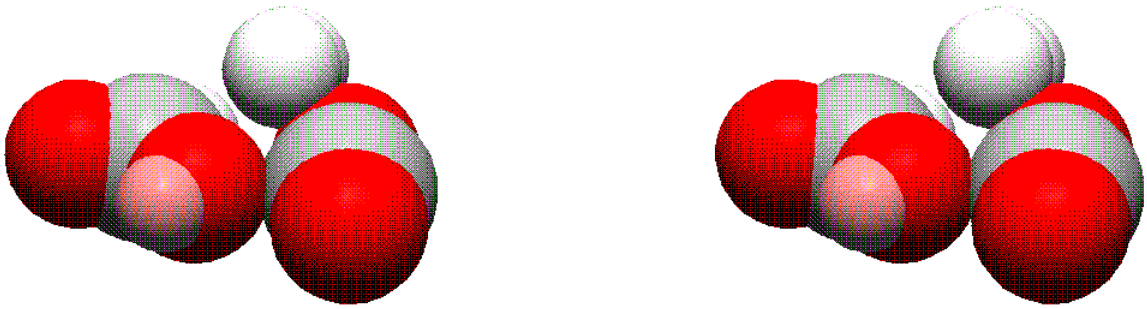


Figure 5g. *A trajectory snapshot (trial-827 simulation-6) depicting the concerted development of a close CO_2/H_2 ligand pair through target interaction and internal ligand interaction; Note formyl C-O dipole interaction with CO_2.*
OCO---C-formyl target 3.40Å
O_2C---O-formic target 2.62Å
H_2---OCO (closest) 1.87Å
H---H 0.80Å

Conclusions

The goal of producing an effective design algorithm which operates in a concerted manner at the semi-empirical HF level is within reach, however systems of practical significance ~50 atoms would require an inconveniently large amount of CPU time to complete a simulation of consequential (meaningful) length. Including higher dimensions of variability such as modulating numbers of atoms and atom types would pose even further computational burden on attaining convergence for location of meaningful configurations. (The use of energy gradient directing methods rather than a strictly Monte Carlo type approach should possibly be more purposeful and efficient if it were effectively installed). Fortunately, the advent of semi-empirical HF packages that use localized molecular orbitals such as MOZYME [14] offers considerable relief from the CPU dependence of the necessary energy calculations. (By using localized molecular orbitals instead of matrix methods, the time required can be made almost proportional to the size-N of the system rather than N [3].) This would make these simulations more viable in practice.

The presented parallel Metropolis algorithm offers versatility in that more guiding values may be assessed and incorporated by just adding additional MC evaluation processes. For example to ensure ligand binding specificity, simultaneously maximizing unfavorable binding to a second

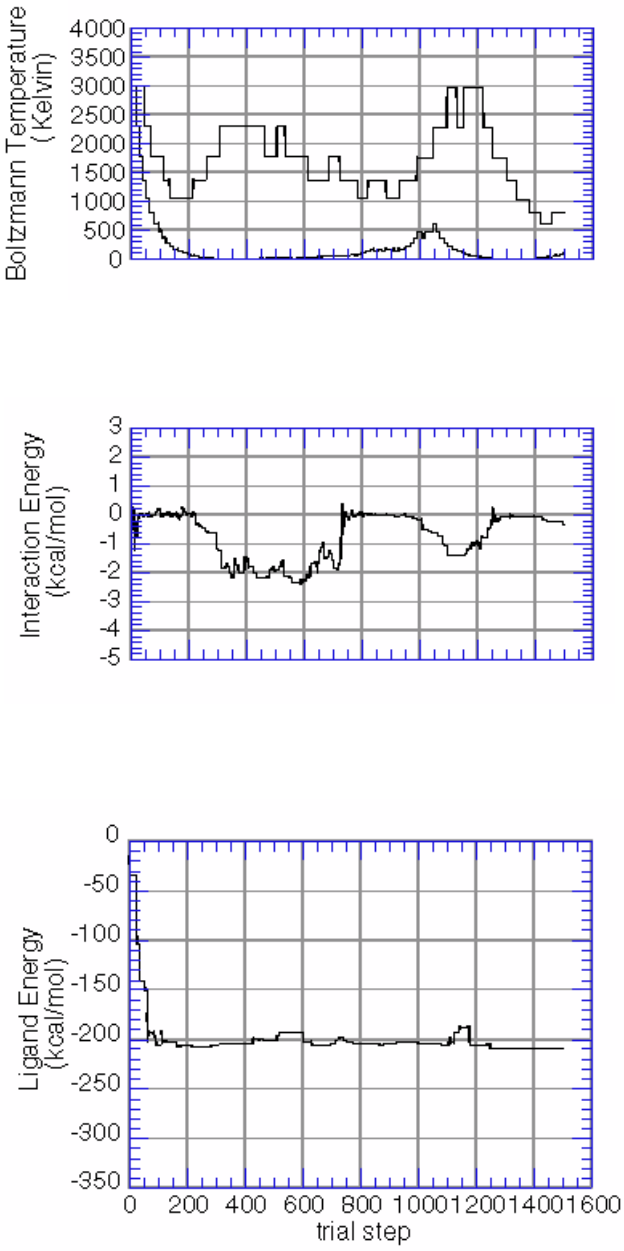
target site may be incorporated by including an additional MC evaluation for maximizing a second interaction energy. Of course other Metropolis partition functions may be used to include variables specific to other ensembles.

The dependence of convergence efficiency upon maximum displacement and translation parameters and acceptance ratio thresholds has not been rigorously assessed in this study. The fact that these original, yet unoptimized parameters are sufficient for low-dimensional ligand development explorations/simulations lends much credence for its development potential. Although the example system is relatively uncomplicated, obtaining successful and meaningful results is clearly not trivial. Nevertheless, the simulation results demonstrate the feasibilty of this systematic temperature-guided MC search method and its appropriateness for the problem presented. Its utility would be best represented in development of small structures at a target binding surface. Development and optimization of this method on low-dimensional searches must have been first accomplished in order to demonstrate the proof of concept and its capabilities for more complex simulation studies which are currently underway and show equally promising results.

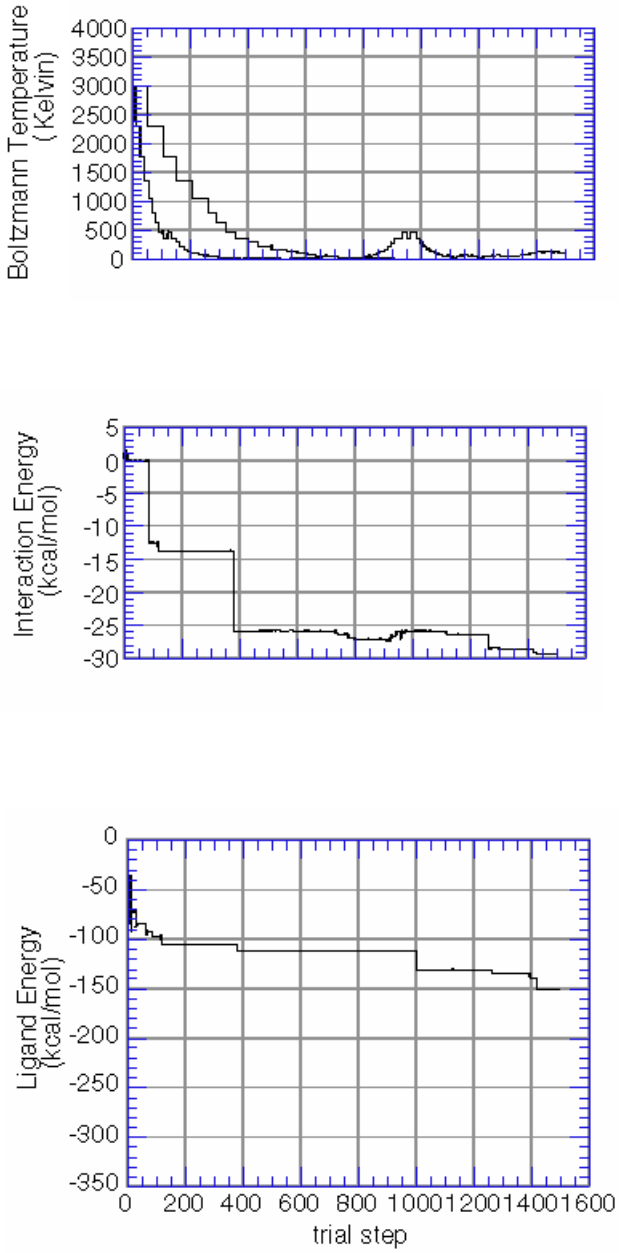
Acknowledgement: J.H.R. is supported by a Hitchings-Elion postdoctoral fellowship from the Burroughs-Wellcome Fund.

References

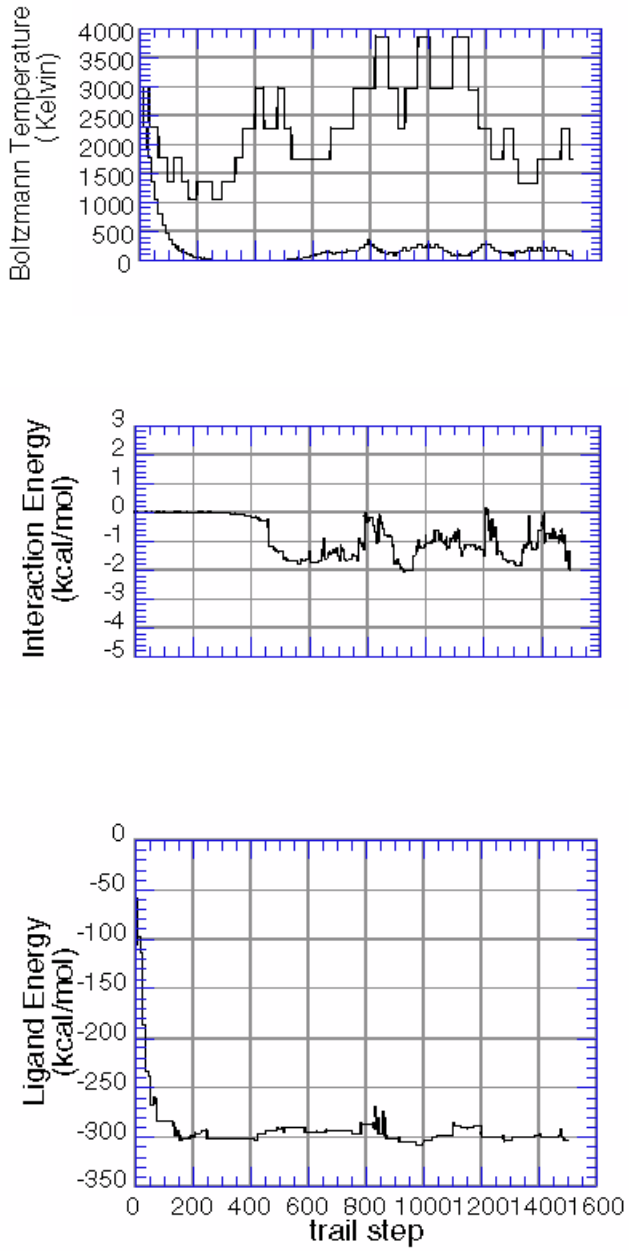
1. Caflisch, A.; Miranker, A.; Karplus, M. *J. Med. Chem.* **1993**, *36*, 2142-2167.
2. Eisen, M.B.; Wiley, D.C.; Karplus, M.; Hubbard, R.E. *Proteins: Struct. Funct. Genet.* **1994**, *19*, 199-221.
3. Moon, J.B.; Howe, W.J. *Proteins: Struct. Funct. Genet.* **1991**, *11*, 314-328.
4. Bohm, H.-J. *Comput.-Aided Mol. Des.* **1992**, *6*, 61-78.
5. Mata, P.; Gillet, V.J.; Johnson, A.P.; Lampreia, J.; Myatt, G.J.; Sike, S.; Stebbings, A.L. *J. Chem. Inf. and Comp. Sci.* **1995**, *35*, 479-493.
6. Nishibata, Y.; Itai, A. *Tetrahedron* **1991**, *47*, 8985-8990.
7. Rotstein, S.H.; Murcko, M.A. *J. Med. Chem.* **1993**, *36*, 1700-1710.
8. Bohacek, R.S.; McMartin, C. *J. Am. Chem. Soc.* **1994**, *116*, 5560-5571.
9. Jorgensen, W.L.; Tirado-Rives, J. *J. Am. Chem. Soc*. **1988**, *110*, 1657-66.
10. Pranata, J.; Wierschke, S.G.; Jorgensen, W.L. *J. Am. Chem. Soc.* **1991**, *113*, 2810-2819.
11. Weiner, S.J.; Kollman, P.A.; Case, D.A.; Singh, U.C.; Ghio, C.; Alagona, G.; Profeta, S.; Weiner, P. *J. Am. Chem. Soc.* **1984**, *106*, 765-776.
12. Metropolis, N.; Rosenbluth, A.W.; Rosenbluth, M.N.; Teller, A.H.; Teller, E.J. *J. Chem. Phys.* **1953**, *21*, 1087-1092.
13. Stewart, J.J.P. *J. Comp. Aided Mol. Design* **1990**, *4*, 1.
14. Stewart, J.J.P. *Stewart Computational Chemistry* 15210 Paddington Circle, Colorado Springs, CO 80921-2512.
15. Miranker, A.; Karplus, M. *Proteins* **1995**, *23*, 472-490.



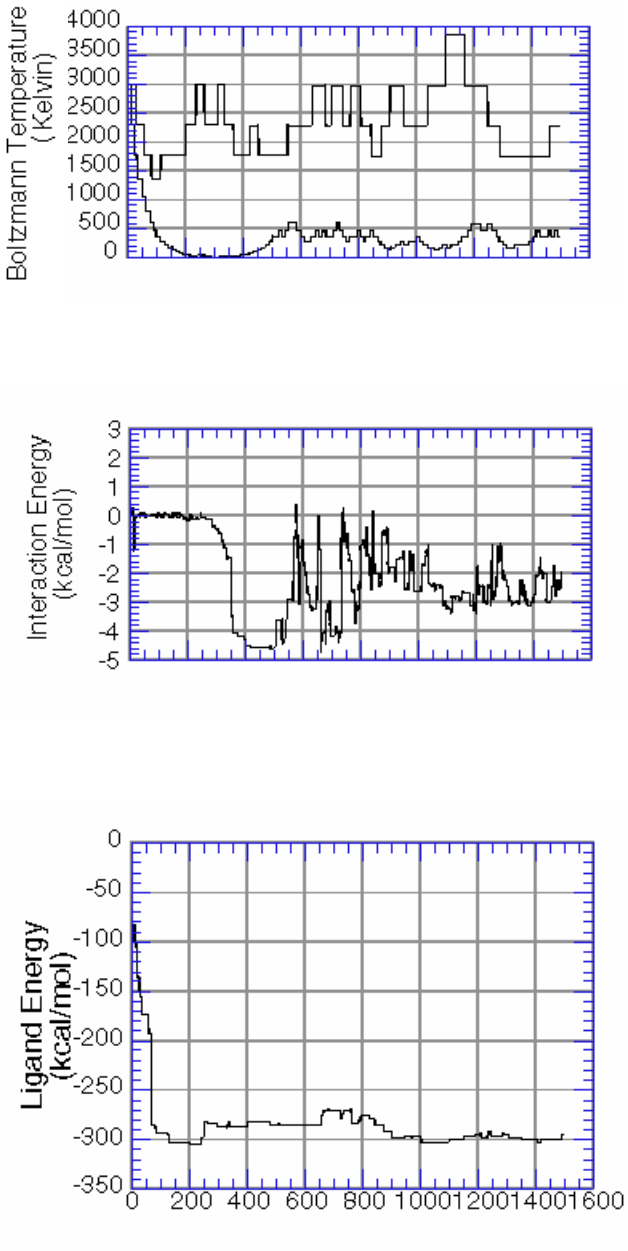
Plot 1.1. *A pilot simulation detained in an N-H/H_2 minimum.*



Plot 1.2. *A pilot simulation restrained due to insufficiently converged SCF convergence criterion set at 0.1 kcal/mol.*

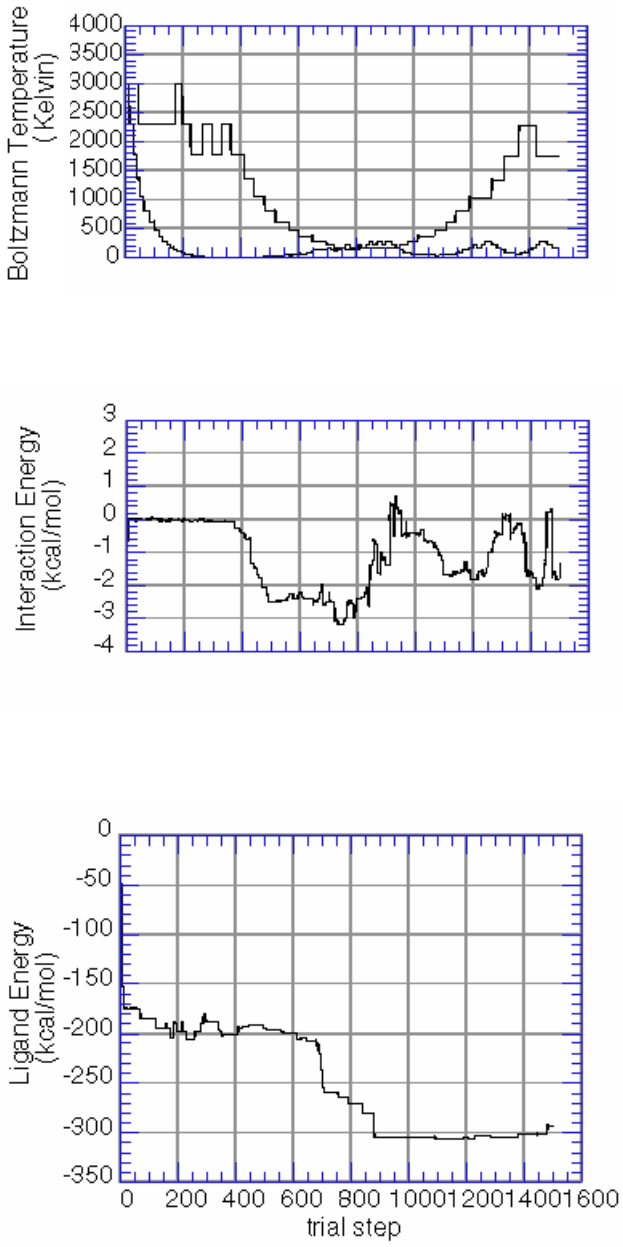


Plot 2.1.

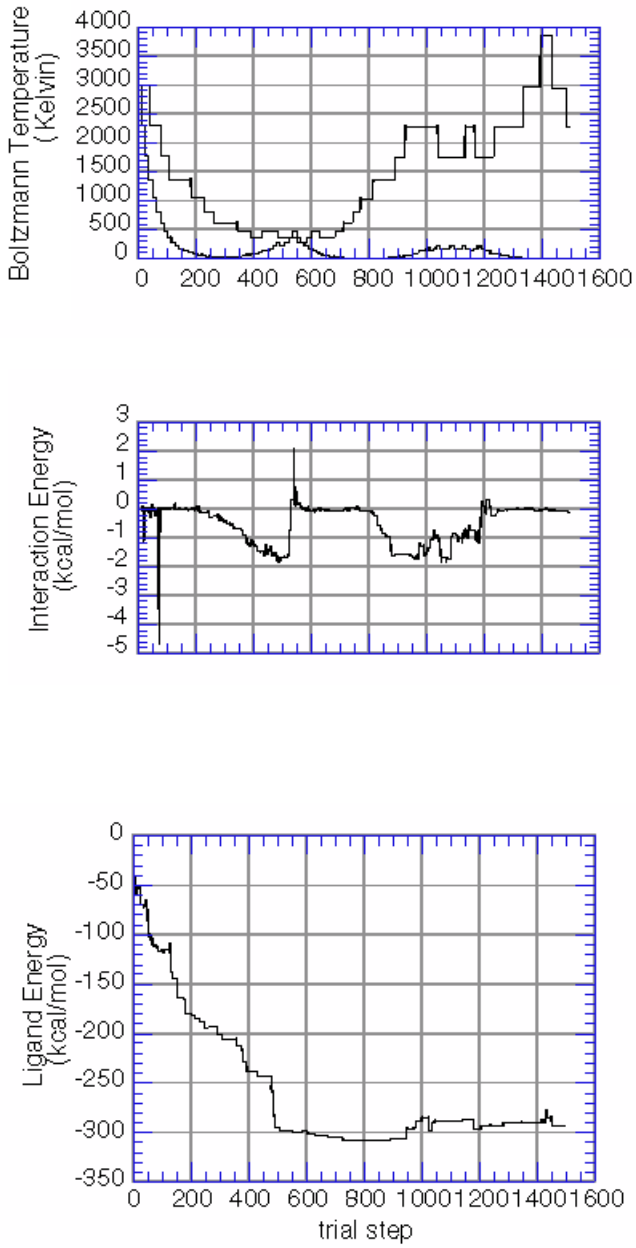


Plot 2.2.

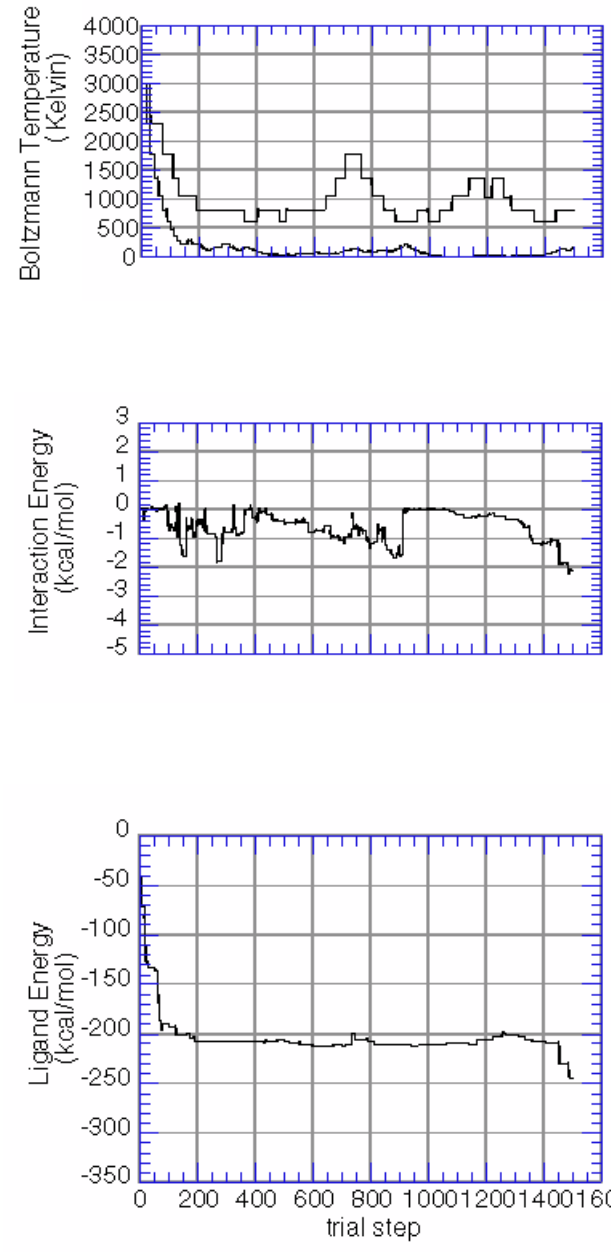
Plots 2.1-2.10. *MC temperature-guided simulations to form ammonia-formic acid complex with distance network constraints and run with 0.01 kcal/mol SCF convergence criterion (simulation 7 required 0.0001 kcal/mol SCF convergence criterion)*



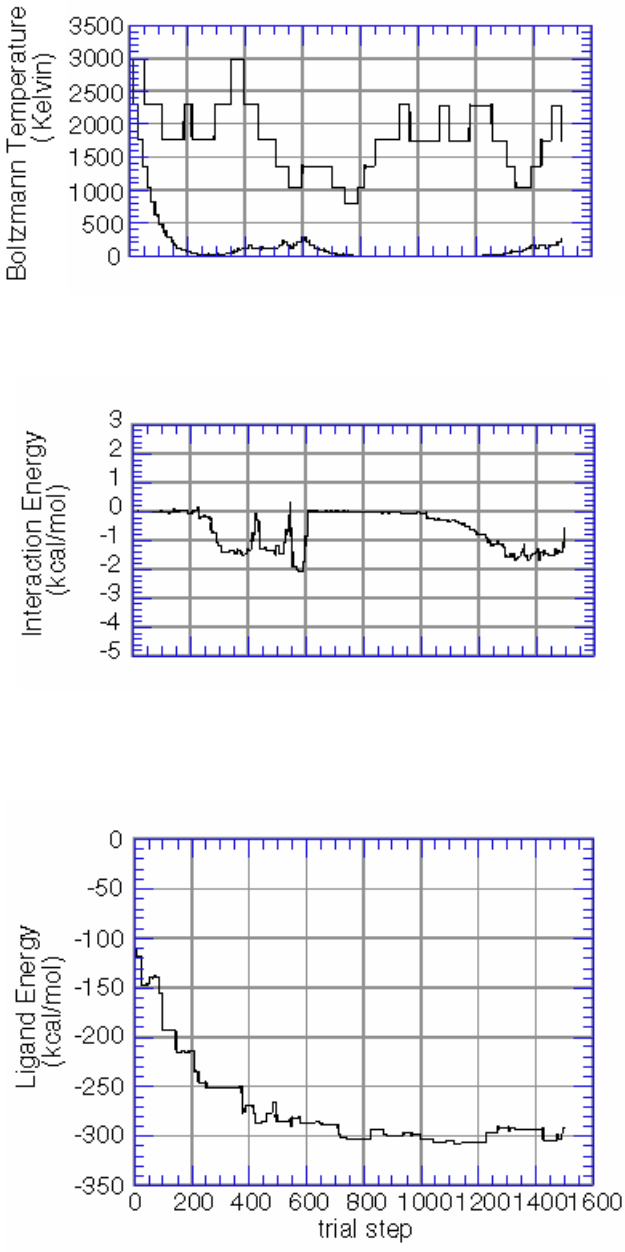
Plot 2.3.



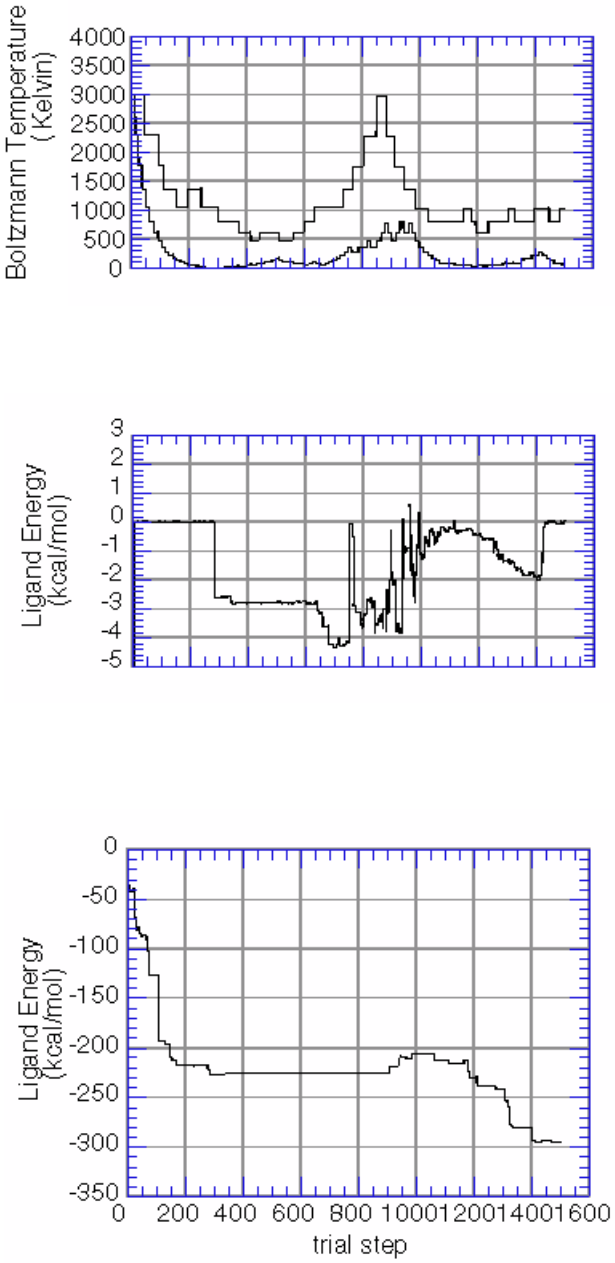
Plot 2.4.



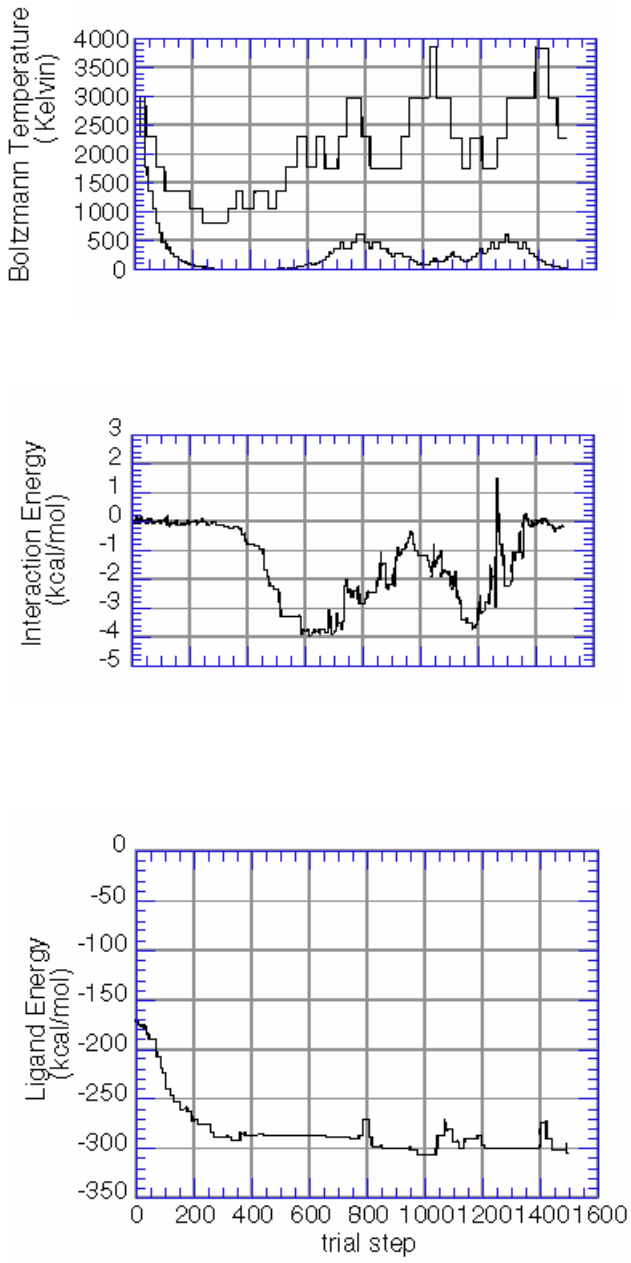
Plot 2.5.



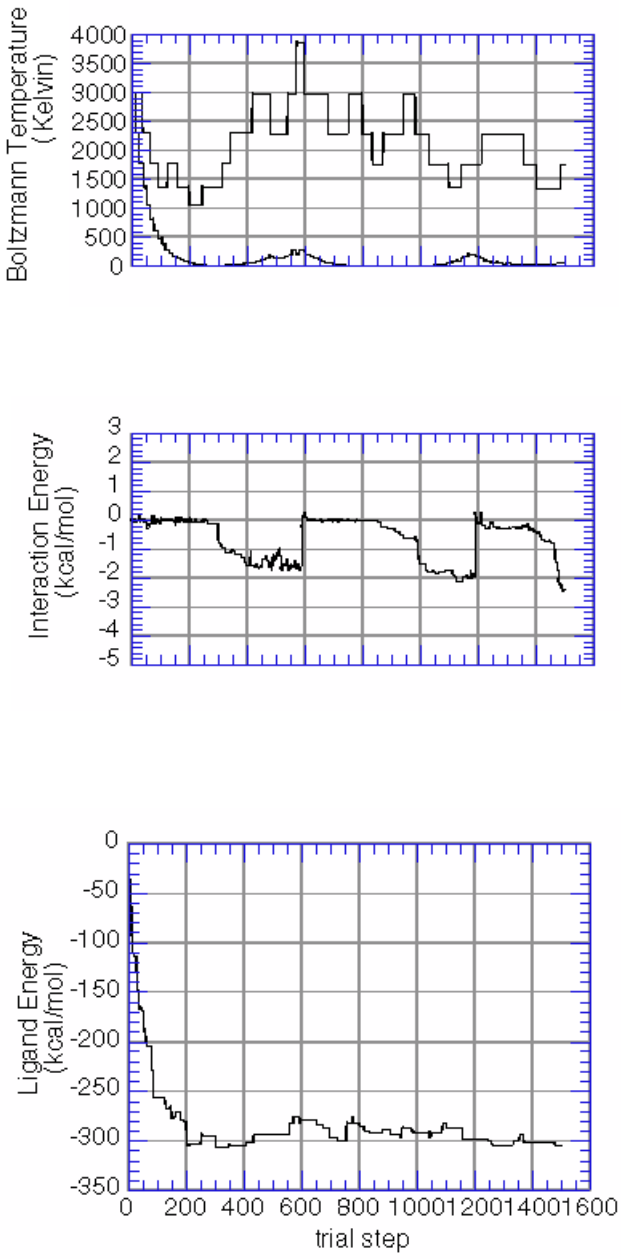
Plot 2.6.



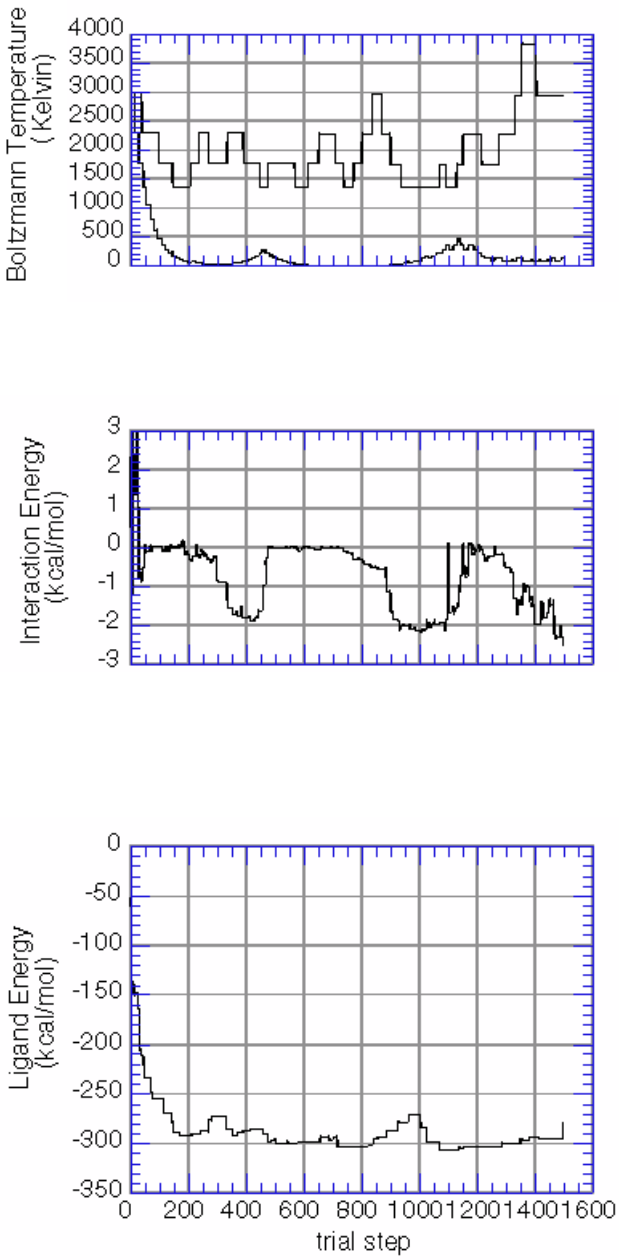
Plot 2.7.



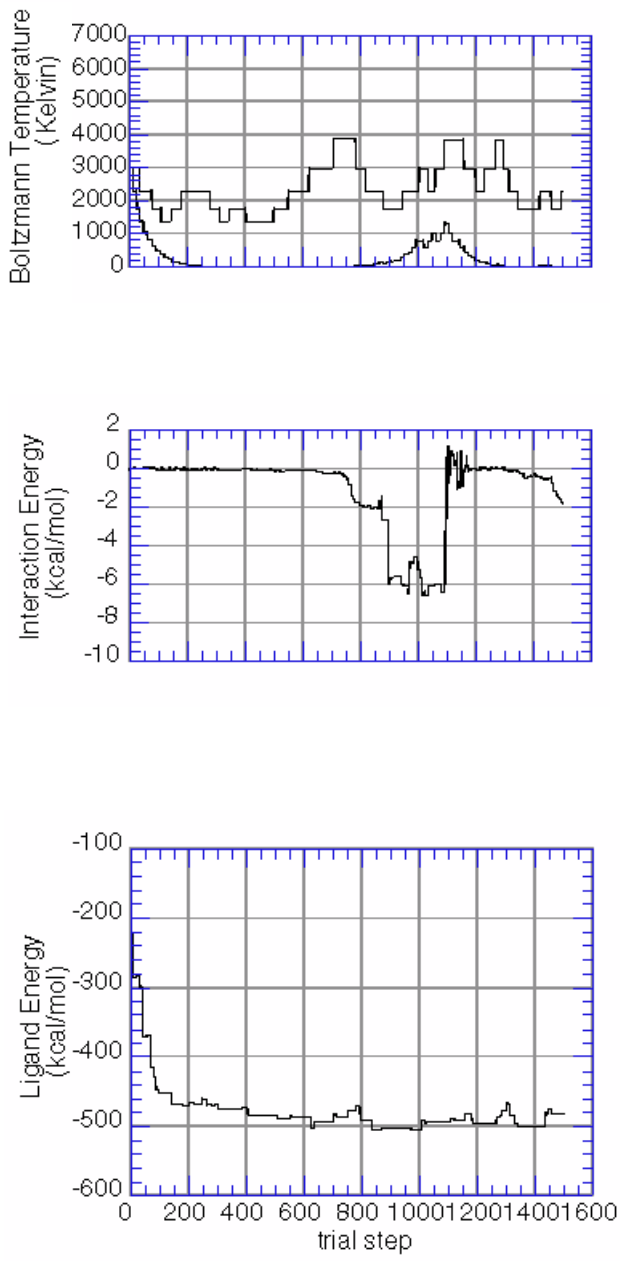
Plot 2.8.



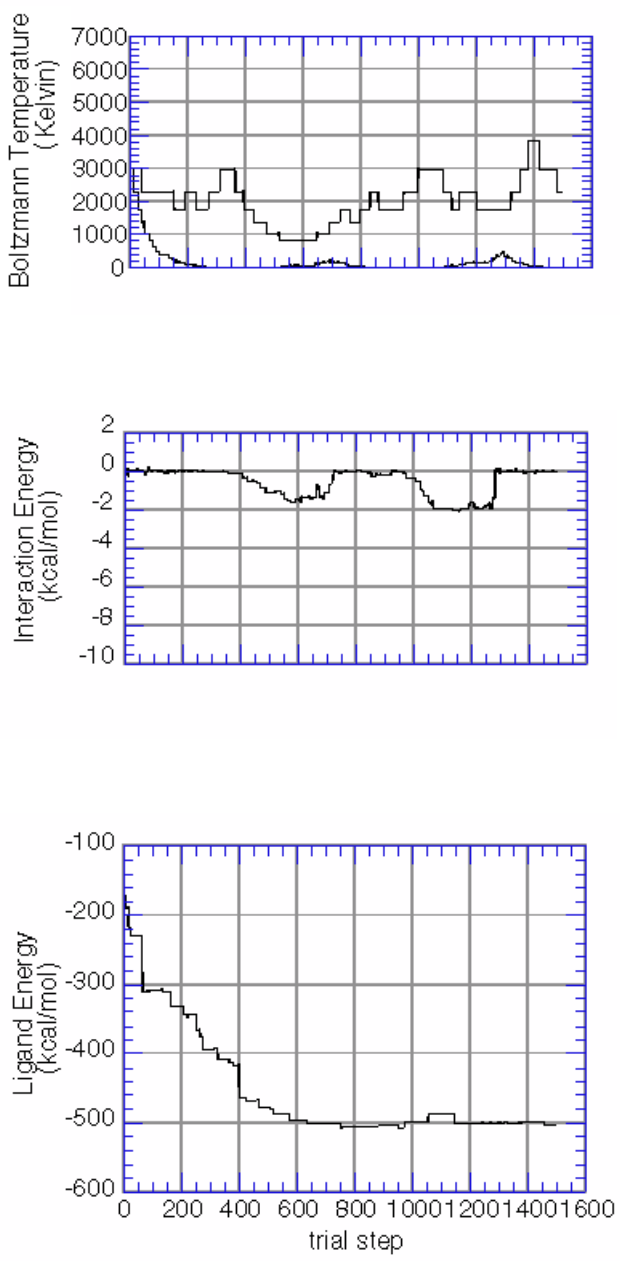
Plot 2.9.



Plot 2.10.

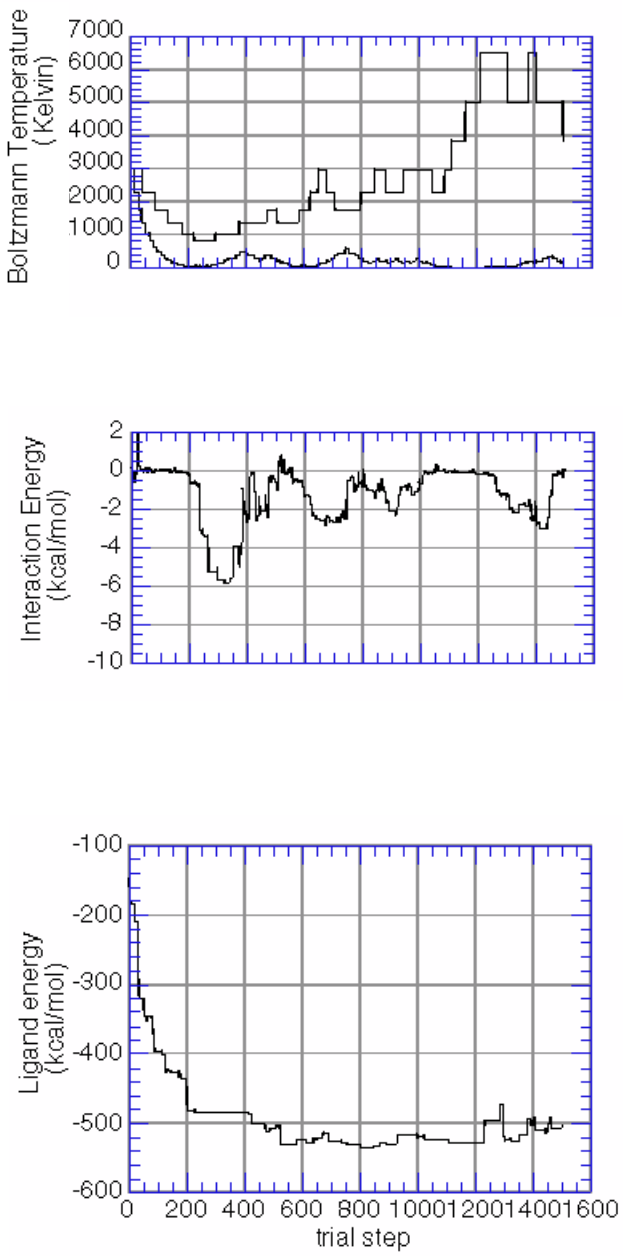


Plot 3.1.

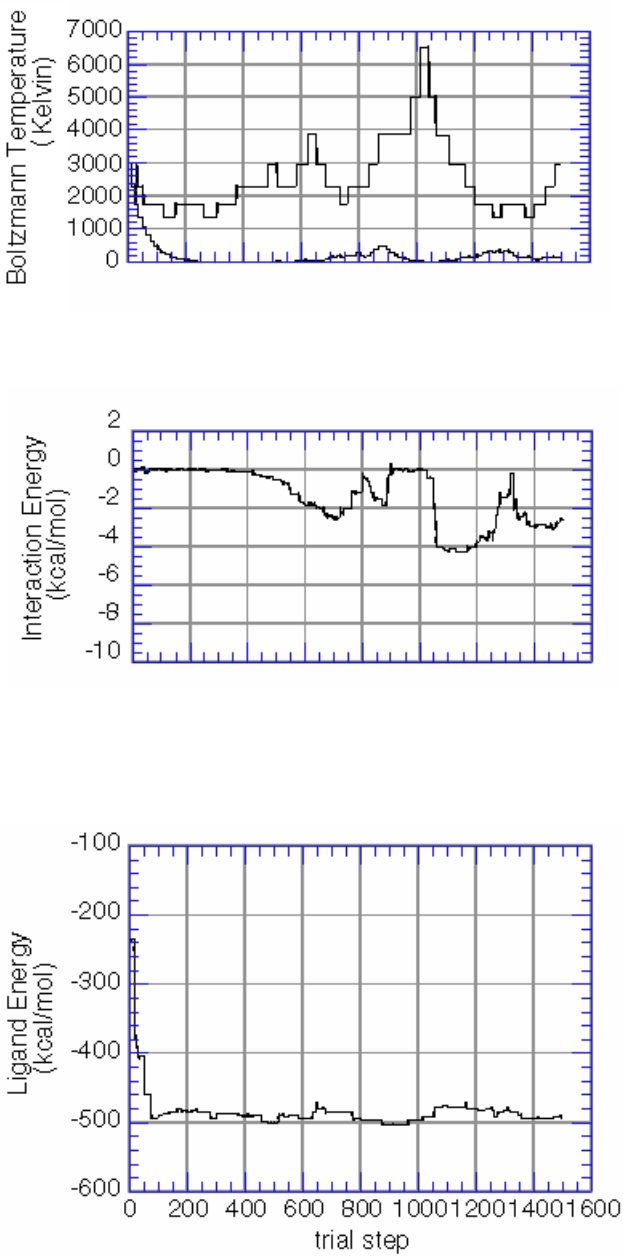


Plot 3.2.

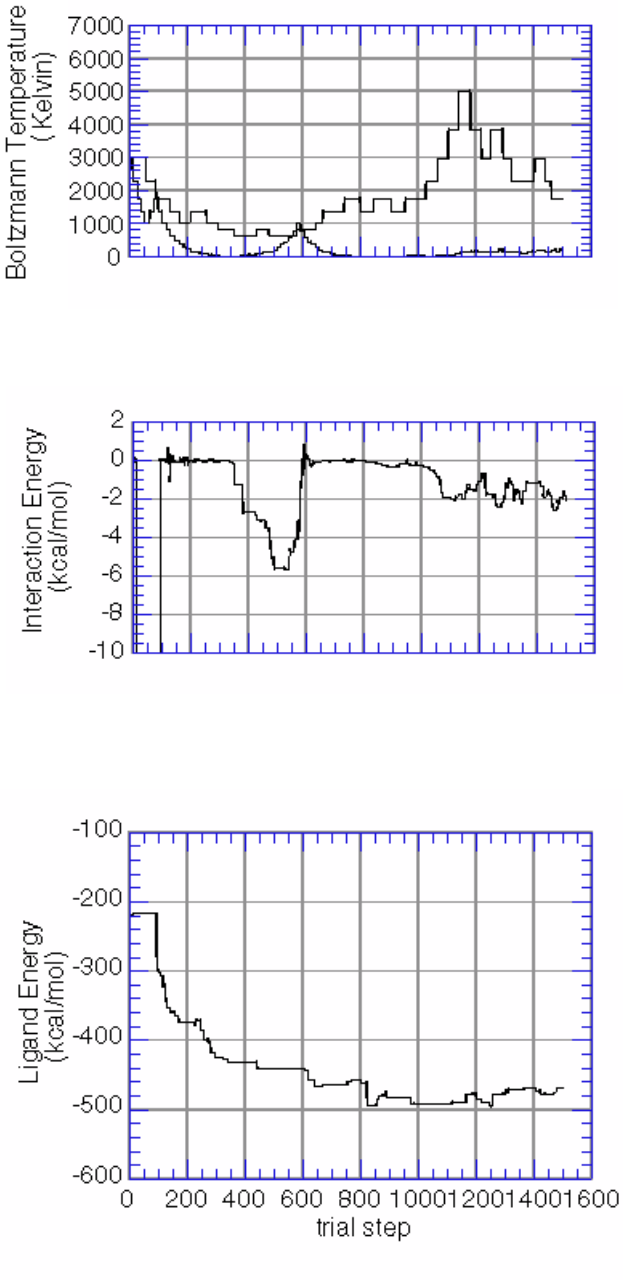
Plots 3.1-3.10. *MC temperature-guided simulations to form formic acid dimer complex with distance network constraints and run with 0.01 kcal/mol SCF convergence criterion*



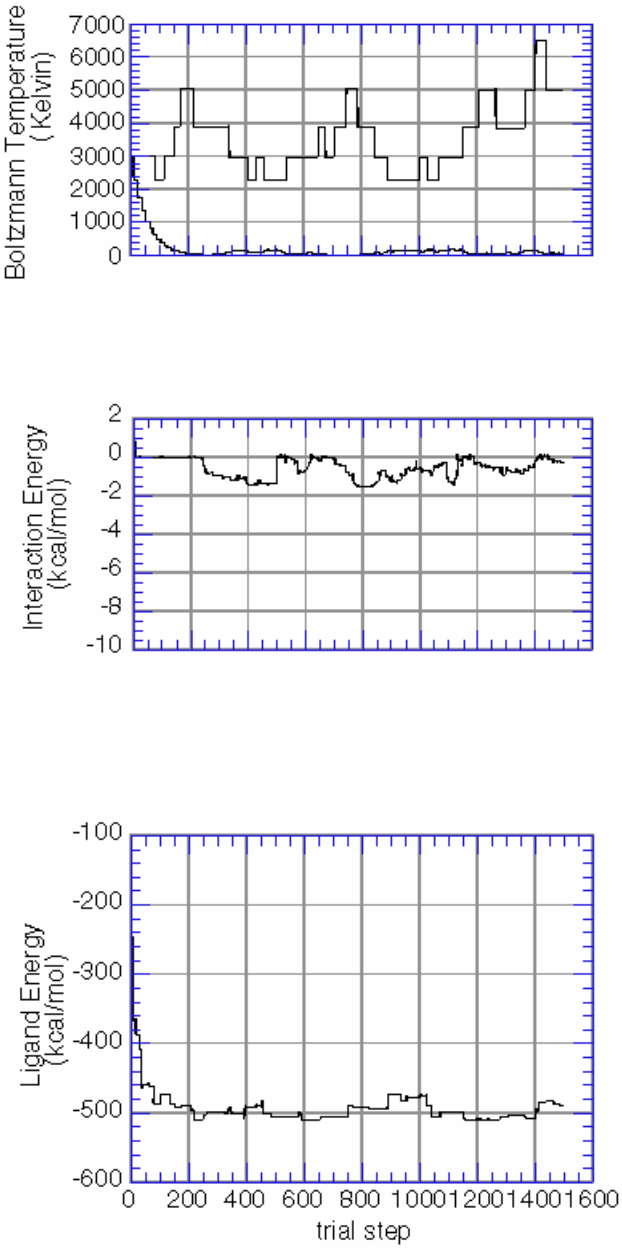
Plot 3.3.



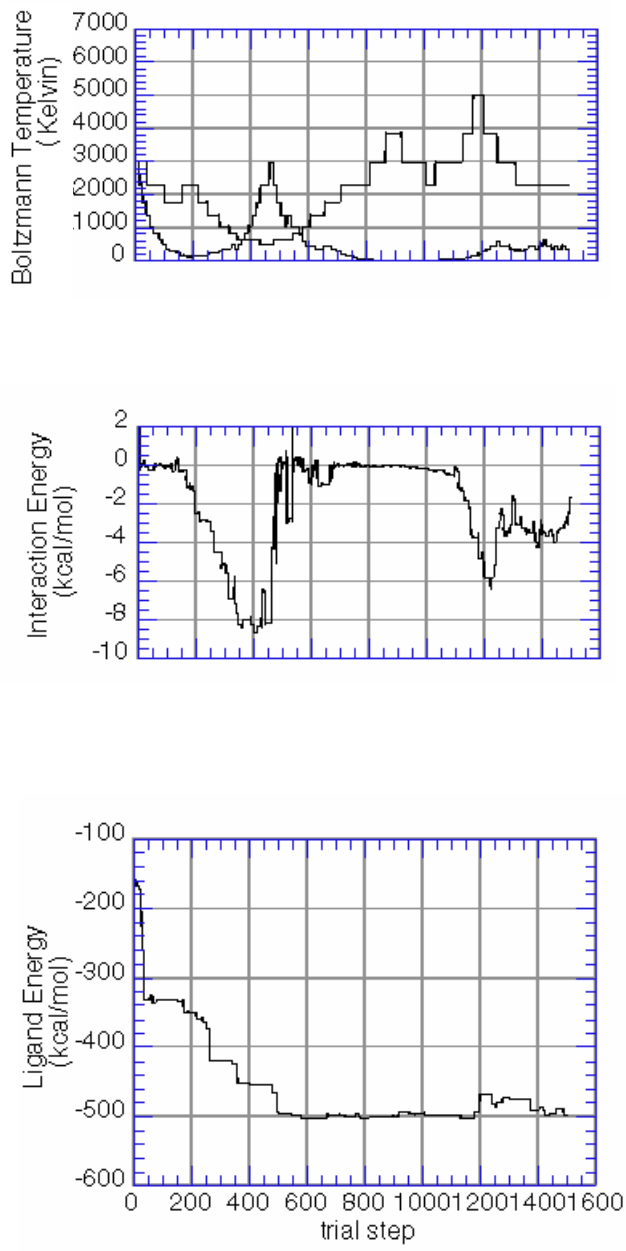
Plot 3.4.



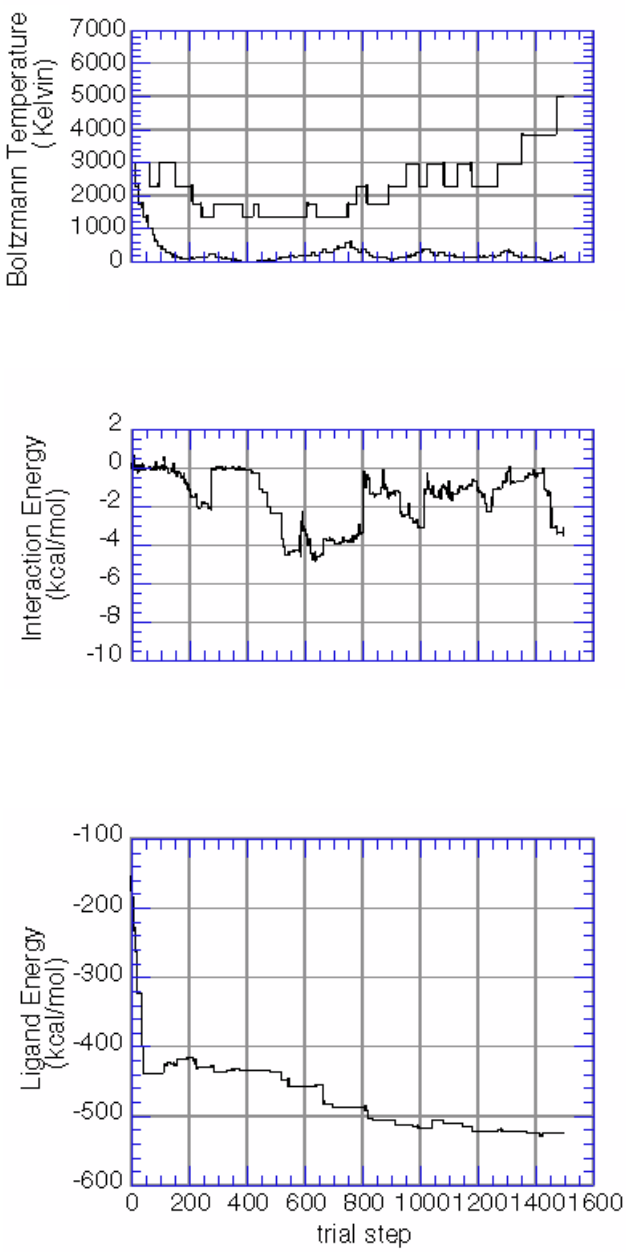
Plot 3.5.



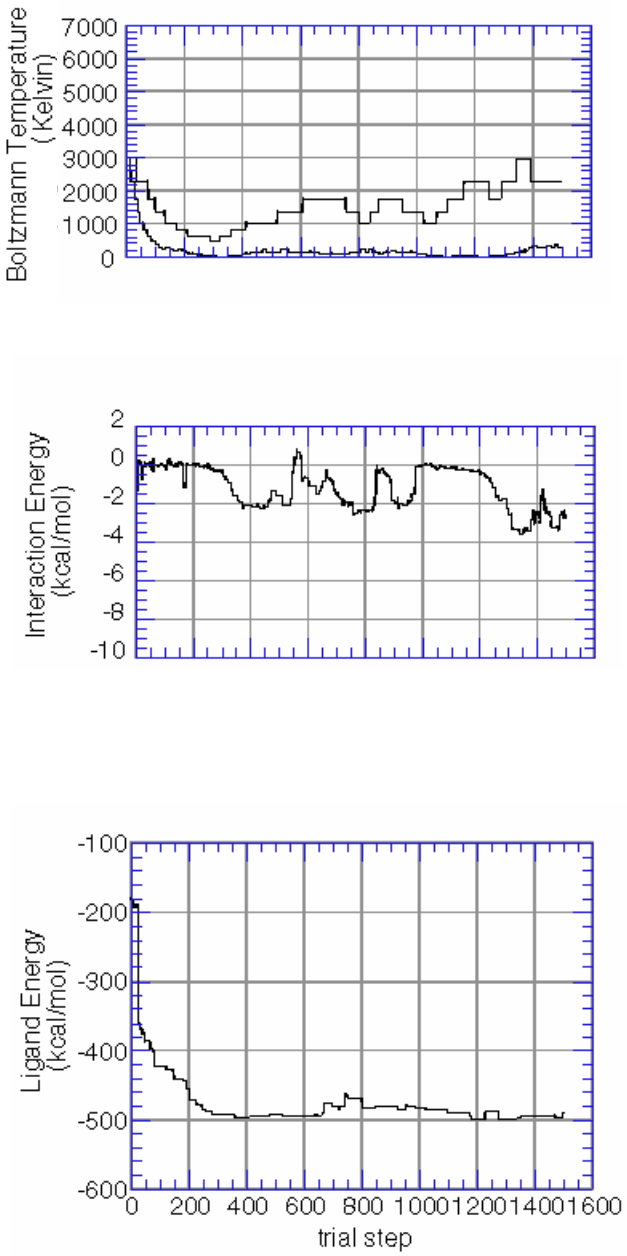
Plot 3.6.



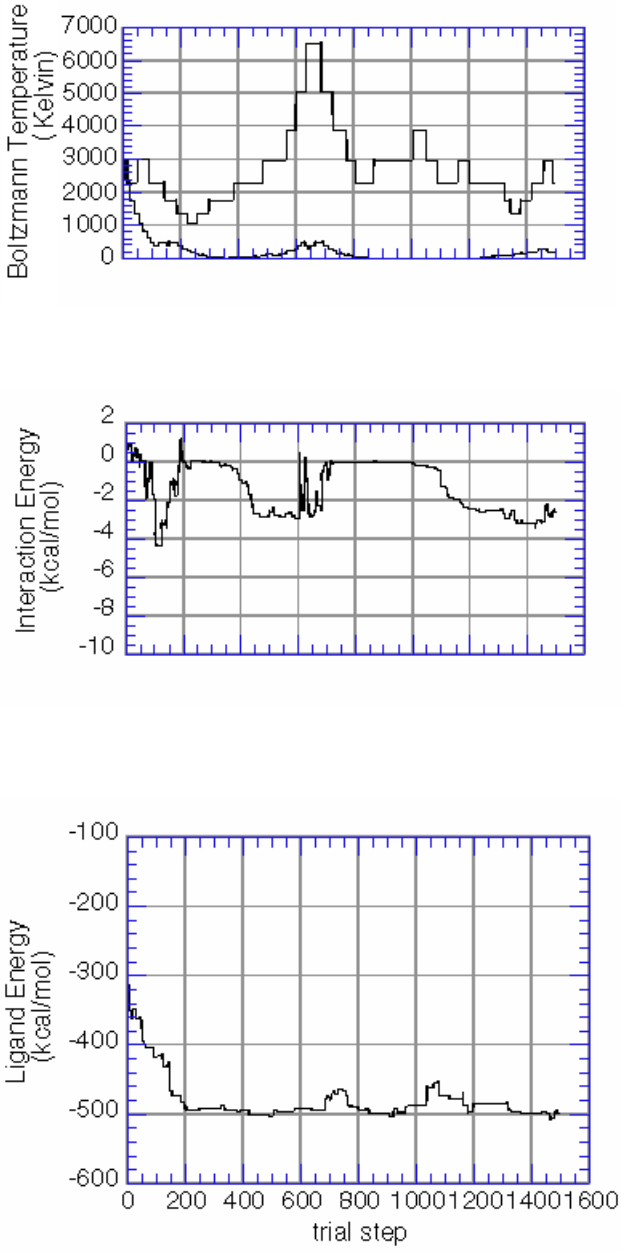
Plot 3.7.



Plot 3.8.



Plot 3.9.



Plot 3.10.